Toward Decision Support for Telecom External Data Monetization

Colot, Christian; Baecke, Philippe; Linden, Isabelle

Published in:
Big Data

DOI:
10.1089/big.2021.0152

Publication date:
2022

Towards Decision Support for Telecom external data monetization: a study of the value of network- and personality-based metrics for third party businesses

Christian Colot,[1,2*] Philippe Baecke,[3] Isabelle Linden[1]

[1]Department of Business Administration, University of Namur,
Rue de Bruxelles 61, 5000 Namur, Belgium

[2]Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg,
Avenue John F. Kennedy 29, Luxembourg

[3]Vlerick Business School, Reep 1, 9000 Ghent, Belgium

[*]To whom correspondence should be addressed;
E-mail: christian.colot@uni.lu

February 1, 2022

Keywords: Decision Support; Big data; Personality Metrics; Data Monetization

Abstract: The big data revolution has led to unprecedented opportunities for data sharing between industries. Telephone companies offer specific data involving rich information not only about the customer behaviour but also regarding his/her relationship with other customers and with third party businesses. The paper addresses the following research question: might telecom data help to improve the prospect selection of third party businesses ? By answering this question, we expect to offer support for two specific investment decisions: on the one hand, the decision of the telecom operator to invest in the new market of the external data monetization for third party business, on the other hand, the decision of third party businesses to buy such customer profiling extracted from telecom Call Data Records.

Using complex data treatments and more than one million models, the paper addresses the challenges and opportunities in collecting and analyzing telecom data from two European telephone companies for improving the prospect selection processes of 36 third-party businesses. This improvement relies on new features extracted from the CDR among which behavioural variables considered as Personality Proxy variables and network-based variables. The results highlight that Personality Proxy variables are useful to support smaller niche businesses. For these businesses these variables are predominant and they can be directly implemented. Additionally, the study shows that network analysis-based variables have the potential to be more beneficial to large companies since the value of network analysis continuously increase with the number of third-party business clients identified.

1 Introduction

With the advent of big data, companies and academia are increasingly searching for ways to leverage these data to create business value [1, 2]. According to [3], 65% of decision makers believe that their companies could become non-competitive if they do not exploit big data. This exploitation can take two forms: using big data to develop the company's own activities or selling/licensing big data to third-parties that may be part of a digital ecosystem [4]. Numerous studies focus on the first type of data exploitation, whereby methods are developed to improve the company's own business activities based on internal data. For example, big data can be used for risk mitigation [5], marketing analytics [6] and safety [7] among other uses.

The commercialization of big data is another means to create additional revenues, although this type of exploitation remains in its infancy. This possibility of monetizing data could become critical to business success. Indeed, 63% of decision makers believe that selling company data to third parties could become as essential to a company as its current business activities [3]. This percentage differs by industry, with the highest percentage (87%) found among managers of telephone companies.

Telephone companies own very specific data, commonly called Call Data Records (CDRs). These mobile data include details on every call, text messages and sometimes also the location where the communication took place between a customer and other users. The source is not only specific but also highly representative. Indeed, mobile devices are now pervasive and fixed lines are becoming obsolete: while fixed telephone penetration rate has steadily declined from 18.8% to 13.0% for the last ten years, mobile device penetration rate has more than doubled during the same period to reach 103.5% in 2017 [8].

From this raw data source, rich information and knowledge can be extracted [9, 10]. First, these data allow the company to map the network surrounding a customer, which yields valuable insight into the customer's behaviour. A customer is typically more likely to exhibit behaviour similar to the behaviour of those who frequently interact with him/her. Previous research has identified several reasons for this phenomenon. The network may directly influence the customer through word of mouth [11]. Alternatively, network influence can be indirect, because connected people usually share similar features [12] or are exposed to the same external shocks, such as local publicity. Second, information related to the personality traits of the customer can be extracted from telecom data. For instance, telecom data are used by [13] to measure the diversity of contacts, gregariousness, regularity, spatiality and use intensity. In an experimental setting, they show that significant relationships exist between mobile phone usage and the personality traits of the customer. Third, compared to traditional fixed line data, telecom data are typically specific to only one user. They are not only generated whatever the location of the user, but they might also record this location.

The elements presented above suggest that customer profiling through CDRs' analysis could provide valuable information for third-party businesses. In particular, they could find in this information support to their prospect selection process. However, the data collection and information extraction process has a cost and raises ethical and security issues. On top of the volume of data, a cost is related to complexity of data wrangling required to extract this information. This complexity is mainly due to the conception of complex features that will be described in section 3.1. such as the regularity of call habits of a customer from one week to another.

Furthermore, the storage of this data should also be taken into account. This is not only due to the massive amount of data carried by mobile data but, more importantly, due to the high security level required to protect these sensitive data against any outbreak. Indeed, the collection of CDR data hosted by the telecom operators could have massive security implications for the concerned customers if they fall into the hands of adversaries [1]. For instance, they could compromise their home security (since CDR records can reveal hours/days of absence), and could even contribute to industrial espionage activities. For more detailed information on this key issue, the interested reader may refer to [14] which contains a thorough discussion on the practical use and exploitation of CDR data in an ethical manner.

Moreover, with the emergence of new data protection regulations, such as the General Data Protection Regulation in Europe [15], the commercial use of this data on an individual level will require to collect the explicit consent from the consumers to be allowed to use the data for third-party prospecting purposes. What is more, this consent should also be obtained for their contacts if the data transmitted by the telecom operator might lead to identify these individuals as well. In this context, consumers will make a privacy calculus in which they balance the perceived benefits with the perceived risks of providing consent [16]. In other words, collecting consumer consent require strategic investments in personalised services, convenience improvements and financial rewards to improve the perceived benefits and communication investments to increase the trust in the service, decrease privacy concerns and perceived risk. It would be useful to assess the value of this data before making these strategic investments.

As discussed by [4], such an investment can not be realised only on the intuition of a potential value. This leads to the definition of the research question addressed in this paper: might telecom data help to improve the prospect selection of third party businesses ? By answering this question, we expect to offer support to two specific investment decisions: on the one hand, the decision of the telecom operator to invest in the new market of the external data monetization for third party business, on the other hand, the decision of third party businesses to buy such customer profilings extracted from CDRs.

[1] We thank an anonymous reviewer for this remark and the related examples.

Using complex data treatments and more than one million models, the current paper shows the value of telecom data from two European telephone companies for improving the prospect selection processes of 36 businesses. This value is obtained by including advanced proxies of human factors in the decision process. We propose a framework to distinguish between two components of these human factors: human interactions and mobile behaviour based on personality proxies. Together, these two dimensions lead to the identification of four categories of data: "Individual Demographics", "Individual Mobile Behaviour (personality proxies)", "Network Business Adoption" and "Network Based Mobile Behaviour (personality proxies)". Our results highlight that Personality Proxy variables are useful to support smaller niche businesses. For these businesses these variables are predominant and they can be directly implemented. Alternatively, network analysis-based variables have the potential to be more beneficial to large companies since the value of network analysis continuously increase with the number of third-party business clients identified.

To the best of our knowledge, this study is the first to consider the value of telecom data for improving acquisition decision processes of other businesses in real-life settings. We propose a framework to distinguish between two components of these human factors: human interactions and mobile behaviour based on personality proxies. To generalize these insights, we investigate the value of these data for three Data Contexts (DCs) and 36 third-party businesses in 11 distinct business domains using a deep analysis. The study is structured as follows. In the next section, we present the related works and propose a new framework that disentangles different features derived from telecom data. Section 3 introduces the data used in this research, data featurization and modelling design of the study. Section 4 reports our main results. Section 5 discusses the implications of these results. In the last section, conclusions are drawn and future research directions are suggested.

2 Related works

2.1 Internal data monetization

Several studies such as [17, 18, 19] have been published in which telecom data are used to support the decision processes of a telecom company. Telecom data, which store each interaction of the customer with others, are a unique source of data that are interesting from a marketing perspective. Traditionally, these data have been used to improve marketing models by retrieving the global activity of the customer and tracking its evolution [20]. One technique for obtaining these data is based on recency, frequency and monetary (RFM) segmentation [21], whereby the following three main features are extracted: monetary value (average revenue); the frequency of communication; and the time since the most recent communication. Recently, studies—such as the work by [22]—have described the opportunity to transform these records into a social network that depicts individuals as nodes and communications as edges. In these studies, the customers surrounding a single customer are investigated; and their global behaviour may be correlated with the behaviour of the focal customer.

These studies address various marketing issues faced by telephone companies. For example, customer churn has received significant attention in the literature [23, 24]. Customer churn may occur when a customer is influenced by his/her network to leave the telecom company or may exhibit the same tendencies as others in the network because they share certain common traits, i.e., homophily [12]. For the same reasons, network analysis of telecom data helps to improve customer acquisition through referrals [25, 26] or to grow customer value [27, 28].

2.2 External data monetization

As mentioned earlier, a common characteristic of all studies cited above is that telecom data are used to improve the business of the same company that generates the data. However, these records could also be used to promote other businesses. This manner of exploiting telecom data has received less attention in the literature. [29] suggests methods to monetize mobile data but does not consider proxies of personality traits and more importantly does not offer concrete results. Hence, this study will evaluate the value of telecom data collected in the core activity of a specific company, including proxies of personality traits, that could later be used to benefit a third-party business.

Beyond the telecom industry, [30] proposes a comprehensive overview of the big data value chain from data generation to data monetization. This study concludes that, although there is a big interest in data monetisation from the business side, this topic has hardly been discussed in the literature. In particular, [31] suggests in their thematic review that the application of an external data monetization should be clarified specifically for each industry generating significant data assets such as the telco industry. The study conducted by [32] is the only one that shows this value in an empirical setting. That study investigates the added value of data from an email provider (i.e. Yahoo) for an offline retail department store. In the study, a communication network was created based on 2-sided interactions through instant messages and emails. This network was used to better estimate the probability that potential customers would make purchases at the department store. The authors find superior performance from acquisition models when network data are added to socio-demographic data; for example, someone who has four contacts with other customers who made purchases in the last six months is 35% more likely to become a customer in the following six months. [32] consider only one business case, however, which raises the issue of whether the conclusions can be generalized. Moreover, the network structure in that study is based on data from emails and instant messages. Other types of data, such as telecom data from a telecom company, could be a better proxy for a customer's social network.

2.3 Personality traits

A study by [13] discusses new opportunities of telecom data. Based on an experimental setting in which participants were equipped with ad hoc mobile phones for one year, they extracted data similar to call data records. Next, they derived metrics from these records based on research in personality psychology. These metrics were classified into a widespread 5 factor model of personality, also called OCEAN (i.e., openness, conscientiousness, extraversion, agreeableness and neuroticism), derived from [33] and measured through a questionnaire. Although this study is based on a small sample of a specific population category (i.e., 69 students of the same university), this telecom data-based approach is considered more promising by the authors compared to other studies where personality traits are derived from social media data. The study reports a higher mean increase in accuracy of 42% above random and the authors point out that telecom data are already routinely collected by telephone companies while social media data require to install a tracking application on each mobile device. While [13] is the only study that links telecom data to personality traits, a few more studies have been conducted based on social media data. [34] illustrate that there exist meaningful links between website-related Facebook likes, Facebook profile features and the same 5 OCEAN personality traits as discussed in [13]. Also, [35] were able to predict the big five personality traits based on text analysis on the language used in Facebook status messages. Another telecom data related study [36] uses an experimental setting to determine the extent to which telecom data can predict spending behaviours. However, this study includes only 52 people living in the same building and utilizes supplementary information that is not typically available to telecom companies (such as bluetooth scans and survey responses).

2.4 Our study

The results presented above leverage the value of personality proxies variables and network variables. However, in most cases, businesses do not have this information. As this can be extracted from telecom data, there is a potential opportunity for external data monetization. Indeed, [37] suggests that new insights often come from new data available, mobile data in particular. However, from the telecom company point of view, this investment might be risky as a survey of [38] showed that 40% of investments in big data made by telephone companies had not a positive impact on the profit or had even a negative return. Furthermore, [4] warns that external data monetization is the most difficult way to monetize data. Hence, the article advices decisions makers to start first by assessing if there is a real potential in the data for this purpose. This is why, in this paper, we investigate if is there any potential in the telecom data from a third party businesses point of view.

In particular, our study builds on the above-discussed studies in several ways. First, instead of using an experimental setting, as [13] and [36] did, this study empirically investigates the value of telecom data for third-party businesses based on data from two European telecom companies. This study also goes beyond the inclusion of telecom data by using network analysis-based variables that include the buying or non-buying behaviour of neighbours, as in [32]. Additionally, mobile behaviours linked to psychological traits of the customer, which are created based on the telecom data, will be taken into account, including spatial data. Finally, to allow the findings to be generalized, the value of these data will be investigated for 36 business in three different Data Contexts.

Based on the literature referenced above, this study will classify telecom company data into four categories within the following two dimensions: network-related variables and mobile behaviour-based personality proxies (see Table 1). Network-related variables are variables created based on the network structure extracted from telecom data. This network structure is created in a manner similar to that of [32] but is extracted from mobile interactions instead of email interactions. Mobile behaviour-based personality proxies are variables extracted from telecom data that measure the regularity, diversity, spatiality, gregariousness and use intensity. These metrics are reported to be significantly correlated with the 5 factors of the personality model discussed in [13]. Consequently, in the first category, "Individual Demographics", all socio-demographic variables of the customers can be positioned. These are node-specific variables traditionally used in prospect selection models, unrelated to psychology or to any linkage between customers. Second, "Network Business Adoption" includes the buying behaviour of a customer's network. These first two categories are not directly related to personality traits. In contrast, "Individual Mobile Behaviour (personality proxies)" contains personality-related variables that do not require the identification of counter-parties, such as regularity, gregariousness, spatial behaviour and use intensity. "Network Based Mobile Behaviour (personality proxies)" also contains variables linked to psychological traits but these are drawn from the network. More specifically, this category reflects the diversity of the customer's contacts, as previously identified by [13]. The variables in each category of this study are discussed in more detail in the next section.

		Personality trait related ?	
		No	Yes
Network Related ?	No	« Individual Demographics » variables Internal customer data (6), local statistics (25)	« Individual Mobile Behaviour (personality proxies) » variables Regularity (11), gregariousness (5), spatial behavior (4) and use intensity (2) extracted from CDR data
	Yes	« Network Business Adoption » variables Business Adoption of neighbours extracted from CDR data (9)	« Network Based Mobile Behaviour (personality proxies) » variables Diversity of contacts extracted from CDR data (6)

Table 1: Variable cartography of telecom company data (the numbers refer to the numbers of features used in our study)

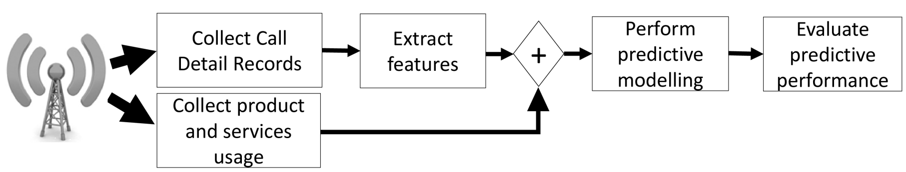

Figure 1: Framework of the methodological approach

3 Method

The following section discusses the methodological approach to perform this study. This approach is illustrated in figure 1 and it includes the collection of data (subsection 3.1), the specific datasets analysed in this study (subsection 3.2.), the extraction of features (subsection 3.3) and the predictive modelling, this part including the evaluation of the performance (subsection 3.4.).

3.1 Data collection

From a general perspective, the first step of the methodological approach of this study consists in collecting relevant data from telephone companies. In particular, there are two types of data which are needed: Call Data Records and product and services usages. CDRs are rows which are generated for each communication activity operated by the telephone company. The exact content of the row depends on the telephone company database management system, but it typically includes the id of the counterpart, the date of the event, the duration (if relevant) and, in some cases, the location of the mobile phone at the moment of the event. The usage of products and services of third party businesses can also be detected trough CDRs, as long as these products and services are operated trough mobile phone communications. Furthermore, some additional products and services usages might be detected based on data ecosystems where the telephone company is active. The specific data collection process of this study are further described hereafter.

3.2 Data Description

In particular, this study explores the value of telecom data for third-party businesses in three different data contexts (DCs). The first DC is based on data of a smaller European telecom company that has young prepaid customers as target segment (28 years old on average). A large number of third-party business activities are detected based on the CDR data of the customer. However, this telecom company did not have spatial data available which slightly limits the creation of Individual Mobile Behaviour personality proxies. In addition, although clear payments or orders are identified in the telecom data, this provides only a partial customer view. Some unidentified customers might also have transactions with the business since they bought the product or service without using their mobile device. The second DC is designed to handle this last issue. It is based on data of the same telecom company, but the third-party business activity is identified in the customer relationship management system through a partnership between the telecom operator and a financial institution. This has the advantage that it correctly identifies almost all consumers who are active clients at the financial institution. The last DC is based on data of a second telecom company. This is a larger company that serves the overall population (47 years old on average) and includes both prepaid and postpaid consumers. This company could also provide spatial data, which allows the creations of all Individual Mobile Behaviour personality proxies. In this DC, the third-party business activities are identified based on telecom data, similar to the first DC. However, since this company targets an overall, less mobile savvy, population, on average a lower percentage of business clients are identified than in DC 1.

Socio-demographic data, transactional data and Call Data Records were collected from these companies in 2013 for the first telecom operator and 2017 for the second telecom operator. The CDR data include caller id, counter party id, time stamp, type of communication (call or text message), the length of the call and, only for the third data context, the location where the communication took place. The data were collected from a sample of customers with regular and normal mobile phone usage. Furthermore, customers that have extremely low or high calling behaviour (i.e., those falling within the top or bottom 5%) are not included in the sample. The final selection contains 98,570 customers for the two first DCs and 904,947 customer for the third DC. For these three DCs, 36 unique business activities are identified, which are classified into 11 industries. These business activities are represented into the three DCs as follows: 26 cases for DC 1, one case for DC 2 and 21 cases for DC 3. Notice that 11 cases are the same across DC 1 and 3. Appendix A provides more explanation on the creation of these dependent variables.

3.3 Data featurization

This research will investigate the value of telecom data for third-party businesses. More specifically, we will investigate the added value of these data in a customer acquisition context. A model will be developed to estimate whether a customer is interested in purchasing from a third-party business. Six months of data are used: the first three months of data are allocated to the creation of explanatory variables, and the last three months of data are used to observe business activities at one of the third-party businesses. To explain which customers have the highest probability of using these business activities, explanatory variables will be created based on data from the three months preceding the dependent period. As presented in Table 1, these variables can be divided into four categories. Descriptive statistics about these variables are available in appendix B.

"Individual Demographics" includes all variables typically used in a traditional customer acquisition model. Because customer acquisition occurs at an early stage in the customer journey, only limited data are available to differentiate potential buyers from other individuals [39]. This study will consider the gender, age and postal code of the customer, as well as two aggregated geographical levels that are augmented with local statistics on taxable income, buildings, household composition and age of the area in which the customer lives (see Table 2). None of these variables are based on network relationships, nor are they related to psychology. Because these are the traditional input data used to support customer acquisition, this model will be considered the benchmark.

The "Network Business Adoption" category contains 9 variables extracted from telecom data (see Table 3). These variables are created based on the social network that can be extracted from the telecom data. In this network,

Sub-category	Feature
Local statistics	
Income info	Total net income divided by number of tax reports
	% of tax reports with positive net income
Building type	% of dwellings by building type (4 categories)
Land use	Proportion of land registry surface area
	Mean taxable value by hectares
Structure of population	% of people by marital status (4 categories)
	% of men
	% of people by age category (9 categories)
	% of family by number of people (3 categories)
Customer data	Age of the customer
	Postal code
	3 aggregated geographical levels
	Sex of the customer

Table 2: "Individual Demographics" variables

a customer, also called a node, can be linked through an edge with another customer (or node) if there is sufficient interaction between them. This link suggests certain behavioural similarities between the customer and his/her network. There are different reasons for these similarities; they may be the result of a direct influence during interactions between customers, or they may stem from an indirect influence linked to certain inherent similarities (i.e. homophily) or mutual exposure to exogenous shocks. These effects can be included using network-based variables that account for the behaviour of neighbours towards the third-party businesses. It is important to remark that from all four variable categories; the Network Business Adoption variables are the most difficult to operationalise. Network Business Adoption variables are specific to each third-party business and necessitate additional data about the customer's activity at each third-party business. Hence, the creation of these variables requires a closer collaboration between the mobile operator and the third-party business. Besides these technical difficulties, the creation of these variables is also more challenging from a legal perspective. The construction of these variables requires not only the consent of the focal customer, but also the consent of all linked customers.

Following [40], two types of network-based variables are retrieved for each customer: ego-based and clique-based. Ego-based variables are built by mapping the network around a particular customer (called the ego customer). The focus is on the first degree neighbourhood (i.e., the direct contacts of the ego customer) as suggested by [41], because taking neighbourhoods of higher orders into account does not increase the explanatory power of the model. With respect to clique-based variables, using the definition of graph theory, the network is here a group, composed of the focal customer and a part of his/her first degree neighbourhood, namely all neighbours who interact with one another. Note that a customer can be part of multiple cliques. These two types of network-based approaches are interesting to include in the study because the ego-based approach analyses the diversity of behaviours around the customer while the clique approach highlights the influence of small homogeneous groups to which the customer belongs.

It is important that during the creation of the network, only edges that represent true relationships are included. Therefore, a threshold should be defined to identify communications that represent true relationships between customers. This threshold should not be set too low because it would assume relationships between customers that do not actually exist. Conversely, if the threshold is too high, the number of influences is significantly reduced [28]. Taking this trade-off into account, the present study defines an edge between two customers if the communication is bidirectional and includes at least 5 contact activities during the first month of the observation period. This definition is consistent with [28].

Variables from the “Individual Mobile Behaviour” category are also extracted from the telecom data. These 18 metrics do not require the creation of a network structure. This category comprises metrics that reflect the person-

Sub-category	Feature
Ego network	Average % Network Business Adoption among neighbours of the customer
	Average % Network Business Adoption among neighbours of the customer weighted by total communication
	Dummy with a value of 1 if at least one neighbour has adopted the Business under consideration and a value of 0 otherwise
	Number of neighbours of the customer who have adopted the Business under consideration
	Total communication time spent with neighbours who have adopted the Business under consideration
Cliques	Average % Business Adoption among cliques to which the customer belongs (customer excluded)
	Average number of Business Adoptions among cliques to which the customer belongs (customer excluded)
	Maximum % Business Adoption among cliques to which the customer belongs (customer excluded)
	Maximum number of Business Adoptions among cliques to which the customer belongs (customer excluded)

Table 3: "Network Business Adoption" variables

Sub-category	Feature
Gregariousness	Portion of all calls that were initiated by the customer
	Portion of all text messages that were initiated by the customer
	% calls occuring between 11 p.m. and 5 a.m.
	% incoming calls answered
	% sms answered within one hour
	Total number of call interactions
Spatial behaviour	Average daily radius of gyration
	Average daily distance traveled
	Number of places visited
	Entropy of places visited: large values indicate regular visits to many places
Use intensity	Total number of call interactions
	Total number of text message interactions
Regularity	6 coefficients of an autoregressive model for each customer based on the CDR observation period (3 months). Dependent variable: number of calls during a 6-hour period (each day being divided into 4 periods). Independent variables: lags of dependent variable (e.g., psy_ar4 coefficient associated with the number of calls made during the same period on the previous day)
	Average time between two call events
	Average time between two text messages
	Variance of time between two call events
	Variance of time between two text message events
	Metric assessing the extent to which the customer is regular in his/her pattern of calls from one week to another

Table 4: "Individual Mobile Behaviour (personality proxies)" variables

ality of a customer, as suggested by [13]. More specifically, metrics linked to regularity, gregariousness, spatial behaviour and use intensity are retrieved (see Table 4). In particular, metrics linked to regularity include six coefficients of an autoregressive model for each customer. Consequently, the implementation of this feature for our specific case leads to the creation of 1 003 517 models i.e. the sum of customer sets of the two telephone companies.

The last category, "Network Based Mobile Behaviour", contains six personality-related metrics designed to reflect the diversity of the customer's contacts. Deriving these personality-related metrics entails the creation of each customer's network based on telecom data. As shown in Table 5, this category includes

Sub-category	Feature
Diversity	Number of different call contacts divided by number of call interactions
	Number of different text message contacts divided by number of text message interactions
	Entropy of call contacts: large values indicate regular contact with many contacts
	Entropy of text message contacts: large values indicate regular contact with many contacts
	Number of different contact persons (call)
	Number of different contact persons (sms)

Table 5: "Network Based Mobile Behaviour (personality proxies)" variables

both features that capture the number of different contacts and features that assess the extent to which a customer has regular contact with numerous people. These features are also based on [13].

After the creation of the variables described above, the data are transformed to increase their usability. First, missing values of quantitative variables are replaced by the mean and outliers (outliers 1% and 99%) are pruned to their corresponding borders. The treatment of outliers not only avoids an overweight of extreme values but also mitigates potential numerical issues linked to large values. Indeed, even if the classifier used, i.e. logistic regression described in the next section, is theoretically not sensitive to scaling, predictors with large values in logistic regression might practically affect the numeric precision of their corresponding coefficients if no treatment is done. Second, to reduce the dimensionality of the dataset, categorical variables, such as postal code, are replaced by their corresponding weights of evidence (WOEs) during the modelling process. WOE is a state-of-the-art technique used to transform categorical variables into quantitative variables, which is especially useful for variables that contain many different values [42, 43]. After transformation, each original categorical variable is replaced by only one variable instead of n-1 dummy variables (n being the number of categories).

3.4 Modelling design

To investigate if there is any real value in the telecom data for external data monetization, an appropriate methodology is needed. The following section describes the methods used to develop our models and retrieve performance metrics. For each model, we use a tenfold cross-validation process to ensure the stability of our results. The training set (90% of observations) is split into two tiers. The first tier, which includes 30% of the training set, is used to compute the WOEs. These WOEs are then applied to the categorical variables in the second tier (which comprises 70% of the training observations). Next, a logistic regression model with a backward feature selection process (p-value threshold: 0.01) is applied on the second set. This feature selection process handles potential multicollinearity issues and ensures to reject the hypothesis that the corresponding coefficients might be zero. This leads to a resulting logistic model where the assumption of the linearity with the log odds holds in the population.

This model estimates the relationship between the explanatory variables obtained from the telecom company and third-party business activity in the following manner:

$$P(Y_{i,j}|X_i, X_{i,j}) = \frac{1}{1 + e^{-(\beta_{0,j} + \beta_{1,j} * X_i + \beta_{2,j} * X_{i,j}}}) \tag{1}$$

where $P(Y_{i,j}|X_i, X_{i,j})$ contains the estimated probability for prospect i for the third-party business j assessed through maximum likelihood, X_i is a column vector of order 59 containing the values of the variables of "Individual Demographics", "Individual Mobile Behaviour" and "Network Based Mobile Behaviour" categories for prospect i, $X_{i,j}$ is a column vector of order nine containing the values of the variables "Network Business Adoption" category for prospect i which depends on the specific third-party business j, $\beta_{0,j}$ is the intercept, $\beta_{1,j}$

is a vector of coefficients associated to X_i, and $\beta_{2,j}$ is a vector of coefficients associated to $X_{i,j}$. More advanced techniques than logistic regression, such as ensemble methods, could be used to increase performance; however, because the aim of this study is to compare the value of different telecom data-related features, a more interpretable regression-based technique is better suited to evaluate the results. Indeed, [44] highlights that even companies innovating in analytics still use regression techniques when understandable techniques are needed.

The selected model is applied to the test set based on which the area under the ROC curve (AUC) is computed. The AUC is frequently used as a performance metric for binary classification techniques. It reflects the probability that a randomly chosen positive case is better classified than a randomly chosen negative case. The average AUC across 10 runs is retained as a final performance metric. For each business, at least five models are estimated: one for each of the four categories of input variables, as well as one full model that includes all categories. The benchmark model represents the targeting efficiency that would be obtained with traditional socio-demographic and regional variables (i.e., "Individual Demographics" variables). Next, this study will investigate the improved explainability of third-party customer behaviour that is achieved by adding the corresponding category(ies) of variable(s). The results are further explained by calculating the relative importance of each variable in the models using the method developed by [45]. This method uses uncorrelated transformations of the original variables to decompose the explained variance of the model into contributions associated with each variable.

4 Results

4.1 Data Context 1: multiple third-party business activities identified through telecom data for a young prepaid segment

Table 6 presents the main results by business category for DC 1. The third column, "% Business", represents the average percentage of people having activities with this business category. The fourth column, "AUC base", contains the average AUC of the corresponding base models using "Individual Demographics" variables. The following columns report the mean increase in AUC in percentage points by adding the corresponding variable category(ies). Individual results by business activity are available in the Appendix C. Compared to other marketing models (e.g., churn, cross-sell), the performance of an acquisition model is typically lower because of the limited availability of data [39]. However, with an average AUC of 75.52% for all benchmark models, these models perform already well. Figure 1, which presents the relative importance of the top 20 contributing variables for all full models, supports this finding: 12 of the top 20 variables are base variables.

Adding psychological and network features derived from telecom data lead to an average increase in AUC of 3.99%. This global result clearly shows the value of telecom data for identifying potential customers in other industries. However, this increase is not of the same order for all industries. Industries with a low base model performance typically show a higher increase in performance when the telecom data are included. A correlation of -53.73% (p-value: 0.0046) is observed between the base model AUC and the improved AUC of the full model.

Considering the AUC improvements separately by variable category, several

Industry	Nb Business	% Business	AUC base	Delta AUC: Network Business Adoption	Delta AUC: Individual Mobile Behaviour	Delta AUC: Network Based Mobile Behaviour	Delta AUC: All Personality Proxies*	Delta AUC: All
App&soft	3	0.60%	71.72%	0.79%	1.47%	1.40%	1.86%	2.17%
Charity	1	1.10%	71.07%	1.11%	1.03%	0.01%	1.19%	2.21%
Entertainment	3	0.87%	73.86%	1.02%	4.04%	3.49%	4.82%	5.52%
Information	2	1.57%	77.11%	1.01%	6.15%	6.75%	7.15%	7.40%
Parking	2	2.92%	71.48%	0.91%	1.54%	1.84%	2.30%	3.03%
Public Transport	1	14.61%	73.21%	2.78%	1.30%	1.08%	1.47%	3.64%
Radio	4	0.97%	75.25%	1.98%	3.54%	3.29%	4.14%	5.05%
Smartphone	1	10.96%	57.22%	3.09%	4.64%	3.27%	5.45%	7.19%
Taxi	5	0.67%	90.30%	0.74%	1.30%	1.28%	1.63%	1.88%
Television	4	1.00%	68.89%	2.50%	2.29%	2.08%	2.68%	4.31%
All	26	1.97%	75.52%	1.46%	2.64%	2.47%	3.17%	3.99%

Table 6: DC 1 - Main results (* The All Personality Proxies column displays the delta AUC using both individual and Network Based Mobile Behaviour variables)

differences across industries can be observed. In general, “Individual Mobile Behaviour” variables lead to the most significant AUC increase (2.64%), followed closely by “Network Based Mobile Behaviour” variables (2.47%). Nevertheless, the combination of the three categories still achieves significantly better performance (3.99%). This result shows that each variable category adds value to the other categories.

This added value varies across categories. To analyse the differences, it is useful to combine the personality proxy variables (i.e. Individual Mobile Behaviour and Network Based Mobile Behaviour) into one aggregated category, which will be labelled “All Personality Proxies”, and then to compare this aggregated category with the Network Business Adoption category. This segmentation is implemented for two reasons. First, from a marketing perspective, personality proxy variables may indicate the need to contact the potential customer directly through an offer tailored to his/her profile. In contrast, network business variables may indicate the need to contact the prospect indirectly through his/her network. Second, from an operational perspective, Personality Proxy variables

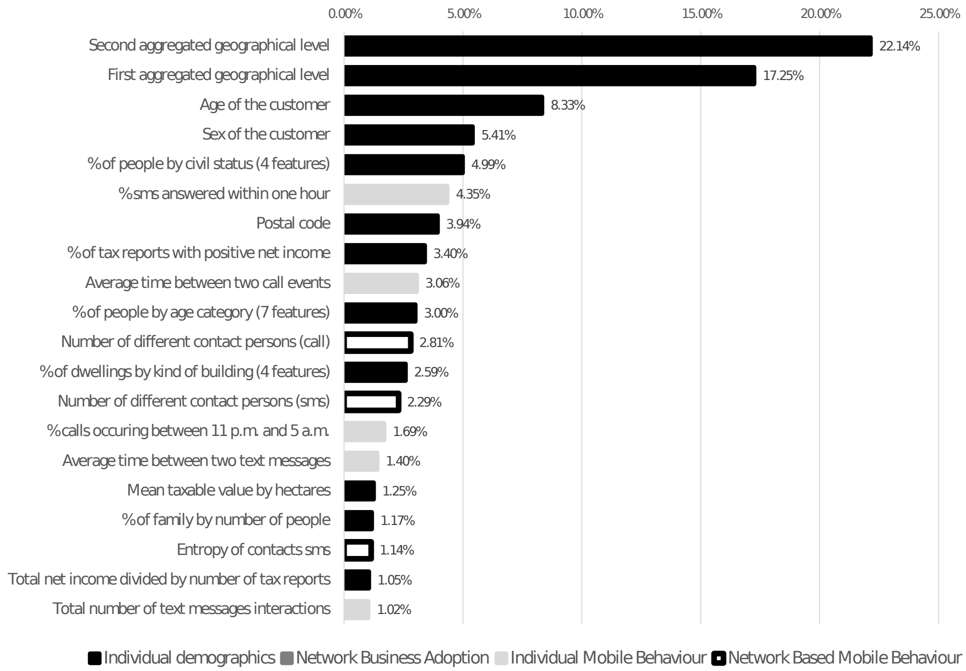

Figure 2: DC 1 : Relative importance of the top 20 contributing variables across all full models

are independent of the third-party business involved and thus technically and legally easier to create and monetize. Given these differences, the following approach is suggested. First, the value of all personality variables should be assessed, because these variables can be retrieved at a lower cost. Next, the potential incremental value of network business variables should be computed.

Taking all mobile behaviour-based Personality Proxies into account results in an average increase of 3.17%. Hence, 0.82% of the overall increase is specific to Network Business Adoption variables. This result is in line with the relative importance of each variable category over all full models, which is depicted in Figure 2. This figure shows that network business adoption variables account for only 2.72% of the explained variance of these models, whereas personality proxy variables account for 22.76% of the variation.

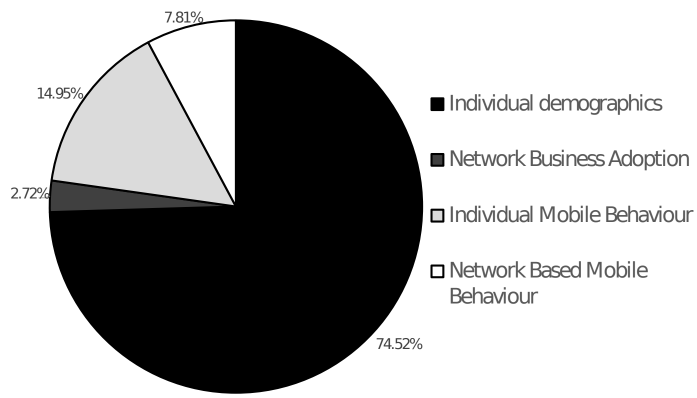

Figure 3: DC 1 : Relative importance of variable categories over all full models

Figure 3 shows the relative weight of Network Business Adoption variables in the full model by industry. A clear distinction can be made between industries in which this type of variable has an impact and those in which it does not. Specifically, the importance of Network Business Adoption variables is at least 11.85% for the Public Transport and Smartphone industries, whereas it is less than 3.80% for other industries. This result is consistent with Table 6, which shows significant differences in predictive performance for these industries when comparing the model with network-based variables (i.e., the full model) to the model without network-based variables (i.e., the All Personality Proxies model).

Next, the data were used to search for distinctive factors that differentiate businesses where Network Business Adoption variables have an impact and those in which they do not. Interestingly, a positive correlation of 39.91% (p-value: 0.0434%) can be observed between the percentage of customers using third-party products or services (i.e., % of business) and the AUC increase resulting from the inclusion of Network Business Adoption variables. This correlation shows that for businesses with a high penetration rate the customer network adds more value.

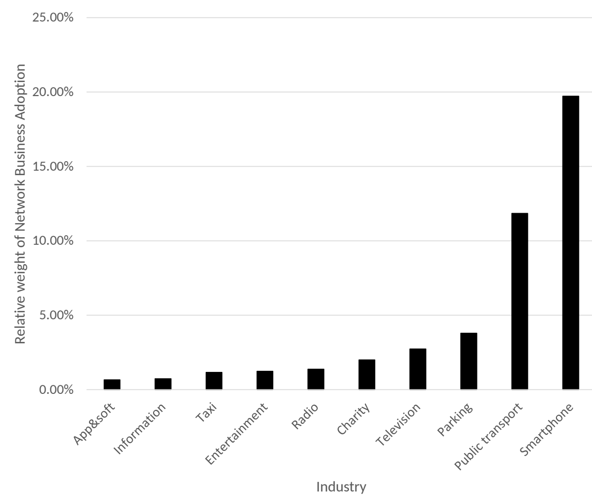

Figure 4: DC 1: Relative importance of Network Business Adoption variables

4.2 Data Context 2: Finance third-party business activity for a young prepaid segment

Table 7 shows the main results for DC 2 with the same meaning of columns as table 6. In this case, the business activity represents if a customer is client at a large financial institution. This activity is identified in the customer relationship management system through a partnership with the institution. Hence, all consumers who are active clients at the financial institution could be identified. This is visible in the % business column, which is clearly higher than the business activities identified in DC 1. In addition, this number is in line with the average market share of the financial institution. Further, table 7 shows that the base model performance is particularly low (56.56%). Since this is a general business, traditional Individual Demographical variables are not able to distinguish clients from others.

The inclusion of mobile derived features leads to an AUC increase of 9.77%. This result is consistent with the results of the first DC as low base models

Industry	Nb Business	% Business	AUC base	Delta AUC				
				Network Business Adoption	Individual Mobile Behaviour	Network Based Mobile Behaviour	All Personality Proxies*	All
Finance	1	25.70%	56.56%	8.16%	1.29%	1.81%	2.68%	9.77%

Table 7: DC 2 - Main results (* The All Personality Proxies column displays the delta AUC using both Individual and Network Based Mobile Behaviour variables)

are correlated with high overall improvements. However, a closer inspection of this table shows that 7.09% of this global improvement is only due to Network Business Adoption which is high in comparison to the results of DC 1. This is also confirmed in Figure 5, which shows that the relative importances of the Personality Proxy variables is quite equal to the averages of all business activities in DC 1, while the importance of the Network Business Adoption variables is with 66,51% a lot higher than the average percentage in DC 1 (i.e. 2,72%). This aligns with the insight from DC 1 that businesses with a high percentage of customers, in this case 25.70%, benefit more from Network Business Adoption variables.

Figure 4 presents the specific variables that drive this increase. We see that the importance is particularly high for both the weighted and the unweighted percentage of neighbours that are also client at the financial institution. Further, the total communication time spend with neighbours who are client at the financial institution has an impact. This indicates that not only the number of neighbouring clients, but also the strength of the relationships has an impact on the social influence or homophily effect. Further, on top of these ego-based network variables, also a clique-based variable adds value. This means that small homogenous groups have additional impact on top of their individual influence.

However, as previously mentioned, the Network Business Adoption variables are the most difficult to operationalise since it requires a close collaboration

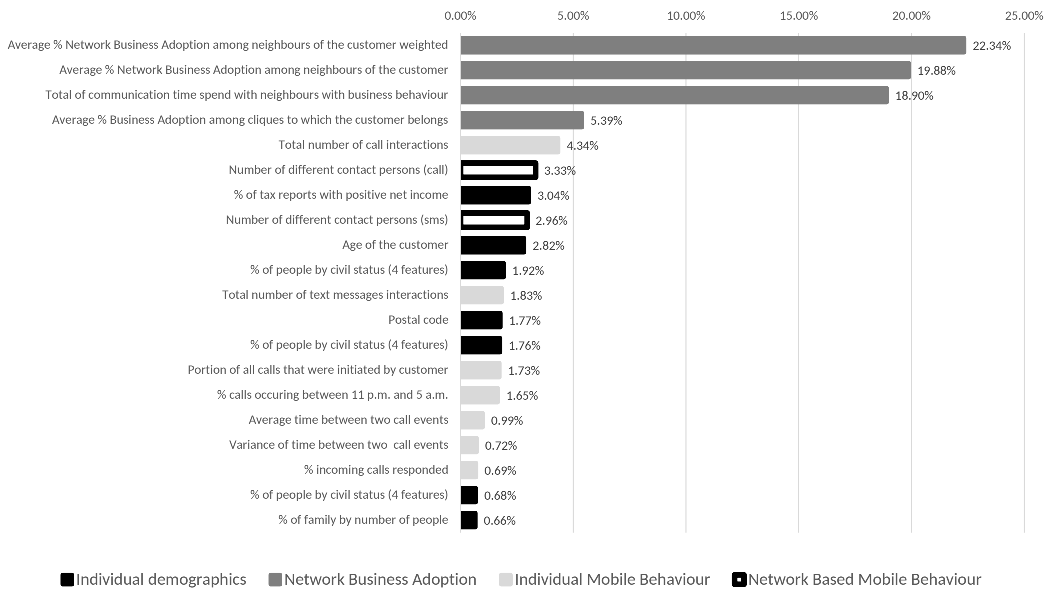

Figure 5: DC 2 : Relative importance of the top 20 contributing variables across all full models

between the telecom operator and the third-part business. In addition, with the emergence of data protection regulations, besides the explicit consent of the focal customer also the consent of linked customers is required to operationalise this approach. Probably an organisation will not be able to collect the consent of all customers. To assess the consequences of this partial identification, we conducted a simulation study in which the business activity identification of the financial institution is randomly cleared for a fraction of customers.

Figure 6 presents the result of this simulation for a percentage of customers varying from 0.5% to the natural percentage of this business (i.e. 25.70%). This graph shows that if only a limited number of customers can be identified (in this case less than 6%), the Personality Proxy variables are more valuable. The model performance based on Individual Demographic and Personality Proxy variables slightly increases when more business clients can be identified since

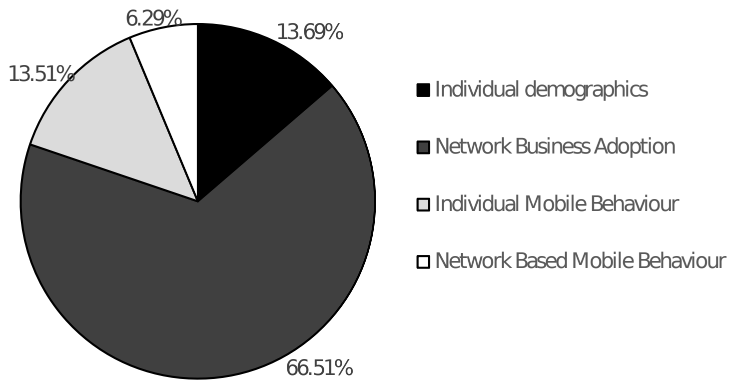

Figure 6: DC 2 : Relative importance of variable categories over all full models

this helps the model to better estimate the coefficients of the independent variables. However, this performance impact is much stronger for Business Network Adoption variables. In other words, if sufficient business activity exists and this can be legally used, Business Network Adoption variables have a lot of potential.

4.3 Data Context 3: multiple third-party business activities identified through telecom data for an overall population

The main difference between DC 3 and DC 1 is that the telecom operator has a more general population and relies on four additional spatial behaviour variables to calculate the Individual Mobile Behaviour variables that are a proxy for personality traits. In order to increase the comparability between DC 1 and DC 3, Table 8 reports the main results by business category for DC 3 based on the same input variables as DC 1 (see Table 6), thus, without the spatial variables. Next, table 9 presents a similar table, but the results are based on models augmented with the four additional spatial variables. Individual results by business activity are available in the Appendix D. With an average AUC of

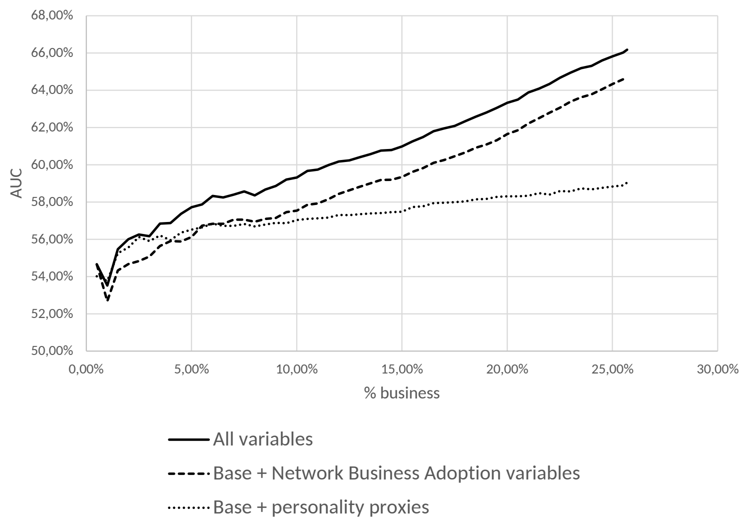

Figure 7: DC 2 - Result simulation where varying fraction of customers is not identified

83.31%, the base models perform already very well in identifying third-party businesses customers. The incremental value of telecom data is still high: on average 4.21% (without spatial variables) which is similar to the result reported in DC 1 (i.e. 3.99%). The addition of spatial data increases this incremental value with another 0.76% to 4.97% (see table 9). This is also reflected in a larger relative importance of Individual Mobile Behaviour variables compared to DC 1 (22.88% for DC 3 in Figure 8 vs. 14.95% for DC 1 in Figure 2). Figure 7 presents that half of the top 20 most important variables are Mobile Behaviour Based Personality Proxy variables from which three of the four created spatial behavioural variables are included. When comparing the full model in table 9 and 8, we can identify the four business categories that benefit the most from the addition of Mobile Behaviour variables, namely Smartphone (delta AUC: 2.46%); Parking (delta AUC: 1.83%); Television (delta AUC: 1.39%) and Entertainment (delta AUC: 1.00%). In general, the percentage of clients identified is smaller than DC 1: 1.07% for DC 3 compared to 1.97% for DC 1. This can be explained by the more diversified consumer base which is less mobile savvy than the younger consumer base in DC 1. As a result, in line with the insights of DC 1 and DC 2, the average added value of the Network Business Adoption variables is limited (i.e. 0.61% in Figure 8). However, this differs per industry and is again highly correlated with the percentage business (87.31%, p-value: <0.0001), which is consistent with results of DC 1. In addition, Figure 9 presents that the same industries as in DC 1 show the highest relative importance for Business Network Adoption variables, namely Smartphone industry (4.82%), Parking (1.65%) and Public Transport (1.53%).

The link between a low Base model and a high overall AUC increase resulting from telecom data augmentation is also confirmed (correlation: -66.71%, p-value: 0.001%). Further, industries with a high penetration rate are significantly correlated with a low base model performance (correlation: -46.16%, p-value: 0.0352). Especially for these organisations telecom data might be valuable. This last result should be carefully interpreted though. This negative correlation was also observed for DC 1 but was not significant (correlation: -32.76%, p-value: 0.1023), which could be due to the fact that correlations were calculated on a statistically limited number of business.

5 Discussion

When we aggregate the results of the three data contexts interesting insights could be revealed. First, businesses for which demographical variables are insufficient to identify clients well, resulting in a lower base model performance, benefit more from telecom data. This result is in line with the study of [32], which shows that the value of an email based consumer network was higher when fewer relevant customer data are available. These businesses with a low base model performance tend to be more general businesses, having a higher

Industry	Nb Business	% Business	AUC base	Delta AUC				
				Network Business Adoption	Individual Mobile Behaviour	Network Based Mobile Behaviour	All Personality Proxies*	All
App&soft	1	1.52%	60.01%	0.65%	3.34%	4.08%	4.73%	4.73%
Entertainment	3	0.28%	77.50%	0.24%	7.63%	5.88%	8.40%	8.45%
Information	4	1.23%	78.33%	0.72%	4.95%	4.59%	5.57%	5.67%
Parking	1	2.20%	81.05%	1.09%	2.23%	2.83%	3.47%	4.00%
Public Transport	2	2.04%	89.67%	0.70%	1.50%	0.69%	1.55%	1.85%
Radio	2	0.50%	86.71%	0.42%	1.72%	1.63%	2.07%	2.34%
Smartphone	1	6.26%	65.52%	2.05%	4.80%	3.12%	4.92%	5.69%
Taxi	6	0.25%	94.97%	0.06%	1.16%	0.65%	1.28%	1.28%
Television	1	0.10%	74.52%	0.56%	9.94%	5.61%	9.84%	9.84%
All	21	1.07%	83.31%	0.50%	3.64%	2.87%	4.07%	4.21%

Table 8: DC 3 - Main results, excluding spatial behaviour variables (* The All Personality Proxies column displays the delta AUC using both individual and Network Based Mobile Behaviour variables)

number of customers.

Second, given the data contexts of our studies, variables related to personality traits usually prevail on Network Business Adoption variables. This is an interesting result since it contradicts the focus of current academic research. Most marketing studies that include telecom data, only consider the use of network analysis to create Network Business Adoption variables, but ignore the potential value of Personality Proxy variables [23, 25, 27, 46, 28, 22, 26, 47]. However, the predominant performance of Mobile Behaviour Based Personality Proxies needs to be put into perspective. This is the case if the percentage of identified business clients is limited. This small percentage can occur either because the business is a niche business or because of limitations in matching consumers due to data technical or privacy regulation reasons. Unlike Personality Proxy variables, the value of Business Adoption Network variables is strongly linked to the percentage of identified business clients: the more the better. This can be explained by two reasons. First, from a social influence perspective, [48] maintain that for efficient network influence, a person should have a minimum number of previously influenced contacts. Second, even if the Business Adoption Network

Industry	Nb Business	% Business	AUC base	Delta AUC				
				Network Business Adoption	Individual Mobile Behaviour (incl. spatial behaviour)	Network Based Mobile Behaviour	All Personality Proxies* (incl. spatial behaviour)	All (incl. spatial behaviour)
App&soft	1	1.52%	60.01%	0.65%	3.49%	4.08%	4.82%	4.83%
Entertainment	3	0.28%	77.50%	0.24%	8.74%	5.88%	9.41%	9.45%
Information	4	1.23%	78.33%	0.72%	5.93%	4.59%	6.38%	6.47%
Parking	1	2.20%	81.05%	1.09%	4.76%	2.83%	5.37%	5.83%
Public Transport	2	2.04%	89.67%	0.70%	2.52%	0.69%	2.53%	2.78%
Radio	2	0.50%	86.71%	0.42%	2.57%	1.63%	2.72%	2.97%
Smartphone	1	6.26%	65.52%	2.05%	7.36%	3.12%	7.45%	8.15%
Taxi	6	0.25%	94.97%	0.06%	1.32%	0.65%	1.43%	1.43%
Television	1	0.10%	74.52%	0.56%	11.25%	5.61%	11.23%	11.23%
All	21	1.07%	83.31%	0.50%	4.52%	2.87%	4.84%	4.97%

Table 9: DC 3 - Main results, including spatial behaviour variables (* The All Personality Proxies column displays the delta AUC using both individual and Network Based Mobile Behaviour variables)

variables would only capture homophily, from a variable construction perspective, [49] support that some patterns can only be identified with a large amount of data. In particular, [42] have set-up a simulation, similar as in our study, in which the value of a behavioural similarity model was assessed depending on different amounts of data. They conclude that that as more seed customers become available, more focal customers will receive a non-trivial score. In other words, these network-based variables are constructed based on what they call fine-grained behaviour data, for which there continues to be substantial value to increasing the data size. Consequently, companies with a high market penetration, such as the financial institution in DC 2, can potentially benefit a lot from network analysis and the creation of Network Business Adoption variables. Nevertheless, this will require a close collaboration with the telecom operator and strategic investments to collect explicit consent from many customers to process their data for this purpose.

Third, this study provides interesting insights concerning the Network Business Adoption variable. Based on the financial institution case discussed in DC 2, we find that within this category, ego-based variables are most important.

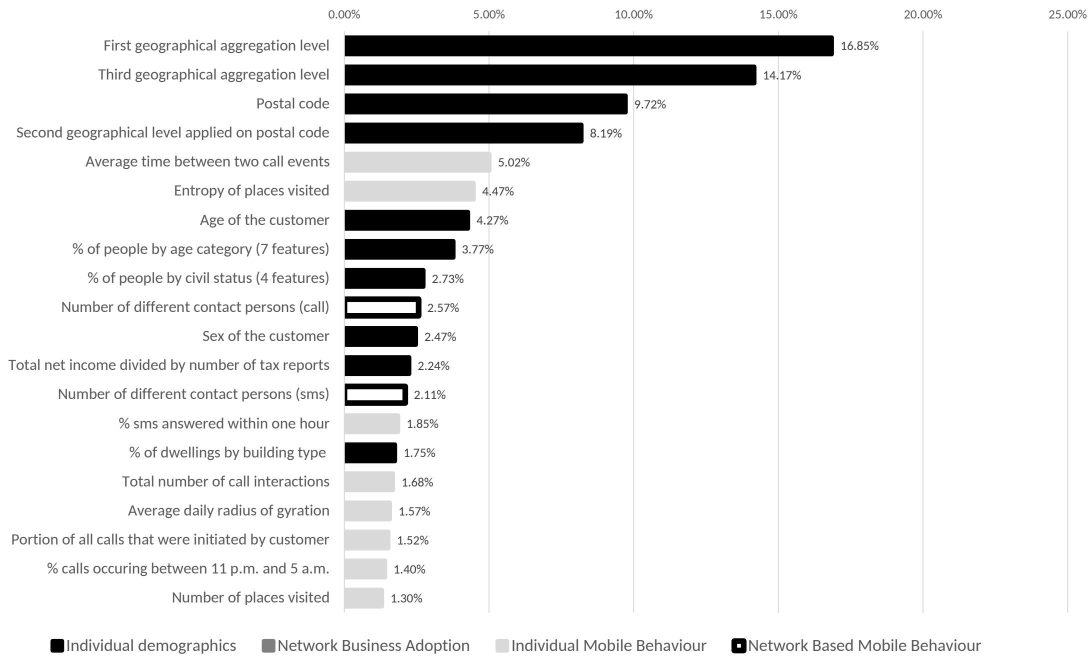

Figure 8: DC 3 : Relative importance of the top 20 contributing variables across all full models

This type of variables is also mostly discussed in the literature (e.g., [32]. However, clique-based variables, as proposed by [40], can measure additional effects on top of the ego-based variables. This can be linked to the study by [50], which is also based on data from a German retail bank. They point to the social enrichment theory to explain the lower churn rate and higher customer lifetime value of referred customers. In essence, the social enrichment theory implies that the relationship with a firm is enriched due to the presence of another linked customer. This can create functional benefits such as the discussion and education of specific products and services [51]. Moreover, this social bounding mechanism could increase the trust in the company. All this is particularly relevant for financial institutions. The fact that a clique-based variable has a positive significant influence on top of ego-based network variables may indicate

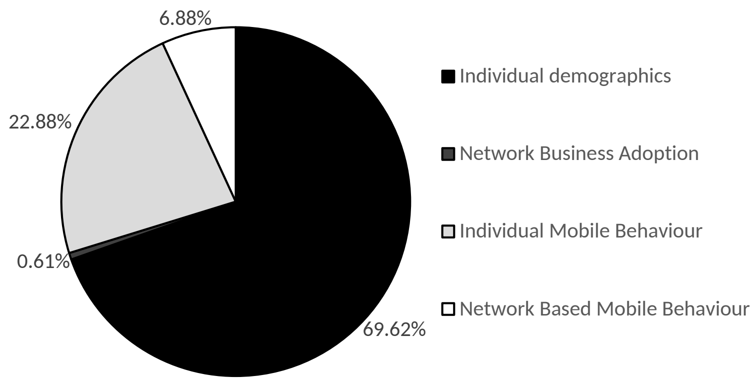

Figure 9: DC 3 : Relative importance of variable categories over all full models

that this bounding mechanism is even stronger when the linked customers are also connected.

Fourth, spatial data are key to develop better Personality Proxy variables through Mobile Behaviour and as a result improve the identification of third-party businesses. More specifically, the Smartphone, Parking, Television and Entertainment industry benefit most from this. The fact that Parking is on the second position illustrates that the Individual Mobile Behaviour variables are only proxies for Personality traits. In this case, the improved predictive performance is most likely due to the direct link between mobility and the need for this business activity. Nevertheless, this direct link does not exist for the other three business categories. In these cases, the relationship could be mediated by personality traits [13]. Although the spatial behaviour variables contribute to the prediction of all 5 personality factors, [13] detected a particularly strong link with Neuroticism. This personality factor represents the tendency towards feeling unpleasant emotions, going from calm to emotional. In an experimental study, [52] investigate the correlation between the big five personality traits and the determinants of computer choice, where especially Neuroticism showed the strongest correlations in the study. Assuming that the purchase process of

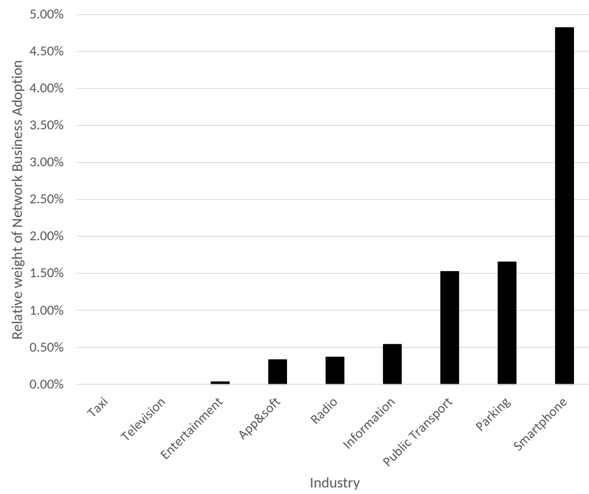

Figure 10: DC 3: Relative importance of Network Business Adoption variables

smartphones is similar to computers, this study supports the supposition that the increased predictive performance resulting from spatial behaviour variables might be due to the mediation of personality traits. In another survey-based study, [53] associate the big five personality traits with impulse buying tendency. Also in that study, neuroticism has the highest influence, besides conscientiousness and extraversion, that have a smaller significant impact. A lot of the business activities in the Television and Entertainment category are closely linked to impulse behaviour. They require to pay using a text message in order to win a lottery prize, chat or receive “astrology prediction” about the future (see Appendix A). Hence, also this study supports the assumption that personality traits are mediating the relationship between spatial behaviour variables and business activity.

Fifth, this study compares the added value of telecom data across multiple industries. Figure 10 positions each business category based on the relative importance weights of Network Business Adoptions and the Personality Proxy variables in predicting third-party business activities. Further, the average percentage of people having business activities in the category is represented by the bubble size in the graph. DC 2 (i.e. the financial institution) is not included for visibility reasons (for this DC, we refer to Figure 5). To increase the comparability, the relative importance of the Personality Proxy variables is calculated excluding the spatial behaviour variables, which were not available in DC 1. A colour coding distinction is made between business categories from DC 1 and DC 3. Also, please note that the underlying businesses within the same business category can differ between the two data contexts (see Appendix A). Despite these differences, interesting insights could be observed from Figure 10. Again, the larger the percentage of identified customers represented by the bubble size, the greater the impact of Network Business Adoption variables. However, this does not only determine the impact of Network Business Adoption variables. For example, in DC 1, more clients are identified for the Public Transportation business activity (i.e. 14.61%) than the Smartphone business activity (i.e. 10.96%). Nevertheless, the importance of Network Business Adoption variables is much larger for Smartphones (i.e. 19.73%) than Public Transportation (i.e. 11.85%). This signals that also the type of business has an impact on the value of these variables. Smartphones are more hedonic products with a lot of visibility. Hence, we can expect that more social influence, captured by the Network Business Adoption variable, will take place for this business activity than for a functional product such as the use of public transportation. On the other hand, the graph also clearly illustrates that the importance of Personality Proxies does not depend on the percentage of customers identified.

From the perspective of the third-party businesses, these results may suggest the development of a tailor-made data acquisition strategy based on a cost-benefit assessment. On the one hand, leveraging variables in an acquisition model generate basic costs of treatment whatever their categories. However,

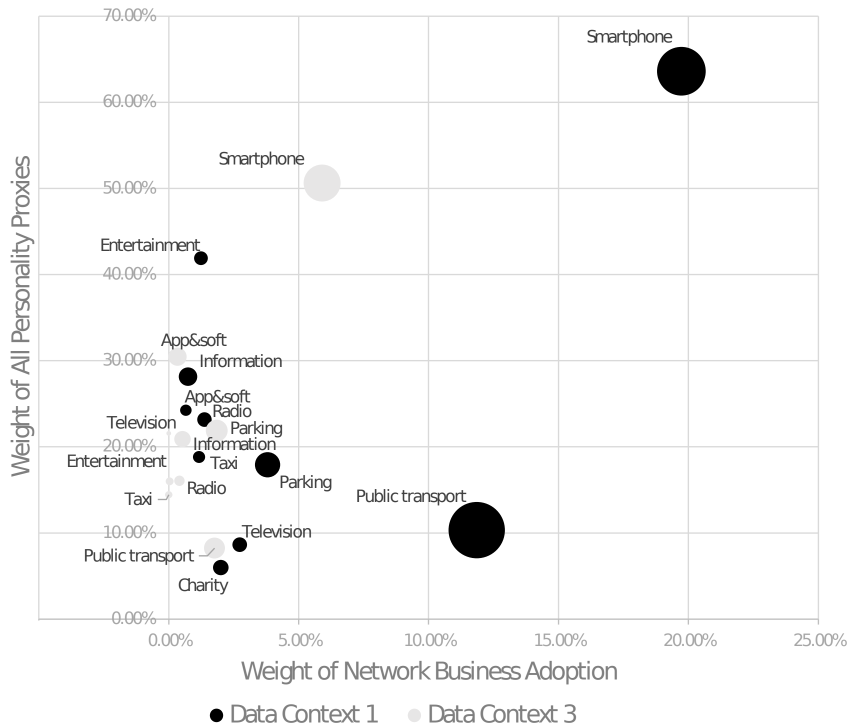

Figure 11: DC 1 and DC 3: Relative importance of Network Business Adoption variables vs. All Personality Proxy variables (spatial behaviour excluded)

while Individual demographic variables do not require additional specific treatments, this is not the case for the other types of variables. For Network Business Adoption variables, this additional treatment is mainly driven by the needed of a deeper collaboration between the telephone company and the third-party businesses. With respect to Personality Proxy variables, the specific treatment of these variables relates to their sensitive nature. It should offer additional guarantees of data protection against any undesired disclosure or misuse compared to other variables. On the other hand, the four categories of variables do not lead to the same acquisition model improvements. In general, depending on the added value of telecom data, three scenarios could be identified. For certain businesses, such as the cases of the Taxi industry, the use of Individual

Demographic data achieves highly accurate acquisition models (AUC base for Taxi: 90.30% for DC 1 and 94.97% for DC 3). Personality Proxy variables add some value (delta AUC for Taxi: 1.63% for DC 1 and 1.28% for DC 3), but this improvement should be compared to the cost of acquiring the data. For other businesses, such as the Information industry (but also App&Soft), utilizing the base variables alone generates low to average acquisition models (AUC base for Information: 77.11% for DC 1 and 78.33% for DC 3), whereas the addition of Personality Proxy variables substantially improves the model, nearly reaching the maximum improvement of the full model (delta AUC for Information: 7.15% for DC 1 and 5.57% for DC 3). In other words, the Network Business Adoption variables do not add a lot of value on top of this. For a third type of businesses, such as the Smartphone business category, utilizing Individual Demographical variables alone also generates low to average acquisition models (AUC base for Smartphone: 57.22% for DC 1 and 65.52% for DC 3). Personality Proxy variables have some value for these industries (delta AUC personality for Smartphone: 5.45% for DC 1 and 4.92% for DC 3), but the combination of both personality traits and network-related business adoption variables generates an even more substantial improvement (delta AUC all for Smartphone: 7.19% for DC 1 and 5.69% for DC 3). Consequently, an incremental approach is recommended. Managers should first try to develop an acquisition model with the base variables. If this model does not perform adequately, the inclusion of personality variables should be considered next because these variables may deliver a superior performance compared to network analysis-based variables. Finally, if the model is still inadequate, Network Business Adoption variables should be considered. Especially organisations with a high market penetration could benefit from this last approach to improve their prospect selection campaigns.

6 Conclusion

The study reported in this paper aims to evaluate the presence of a potential value for third party businesses in telecom data. To support this hypothesis, the paper addresses the challenges and opportunities in collecting and analyzing such data. In particular, a deep analysis including complex data treatments and more than 1 million models has been led on the big data collected from two European telecom companies, for 36 business opportunities classified into 11 industries. Based on the literature review, this study proposes a framework that organises extracted features from telecom data in two dimensions: network variables and personality proxies. Network variables are features that represent the influence of others in the customer's network. Personality traits measure the diversity of contacts, gregariousness, regularity of mobile use and spatial behaviour. These Mobile Behaviour variables should relate to the five factor model of personality. Based on these two dimensions, four categories of variables could be identified: Demographic, Network Business Adoption, Individual Mobile Behaviour and Network Based Mobile Behaviour variables. This study is the first to use this framework to empirically investigate the added value of telecom data for third-party decision processes.

In term of managerial implications, the results of the analyses confirm the expectations of managers that telecom data can be valuable to third-party decision processes [3]. The results show that all telecom data categories improve a traditional model which selects prospects based on demographic variables. In general, Personality Proxy variables (i.e. Individual Mobile Behaviour and Network Based Mobile Behaviour variables) predominate in this improvement. These Personality Proxy variables are in particular useful to support smaller niche businesses. Besides the fact that the impact of the variables is less related to the business size, they can be directly implemented since they are not specific to the third-party business. This insight contributes to previous studies that investigate the value of telecom data, which usually focus exclusively on the extraction of a consumer network. Large companies with a high market penetration however, could opt for a deeper collaboration by creating Network Business Adoption variables.

As discussed in the introduction, the implementation of the external data monetization strategy proposed here requires significant investments from the telecom company and will have a cost for the third-party business. The results established in this paper will support the strategy definition of the telecom company by offering the identification the most relevant target businesses. In parallel, they offer important insights to the potential buyers to estimate the potential return on this kind of investment.

The study presented here paves the way for several complementary future research. Indeed, this study has enhanced the existence of a value of the proposed traits for third parties, in particular in their customer acquisition process. The next natural step towards the development of the target decision support systems will be the development of a monetisation model of the benefits and cost assessment models both for the telecom Data owner and the third party buyer. Then, will come the conception of an architecture integrating these model with ours.

Beside this specific project, other extensions of the presented results would benefit from more investigation. First, the focus of this study is on improving the decisions regarding third-party businesses. Future research could study the value of personality trait data extracted from telecom data for improving the internal decision process of telecom companies, for example, by predicting churn. In addition, these Mobile Behaviour Personality Proxies could also be used to investigate the impact of different communication channels. Lastly, this study points to the legal need to collect explicit consent for their data to be used. Future research could investigate in more detail the best approach from a legal and consumer privacy calculus perspective to collect this consent. In particular, this approach should shed some light on the acceptable usages of this data. Indeed, as highlighted in this study, assessing personality traits of customers has business value but it should not lead to potential misuses such as political ad targeting[2] [54].

Author Disclosure Statement

No conflicts to disclose.

Funding Information

No funding information to declare.

References

[1] Bart Baesens, Ravi Bapna, James R Marsden, Jan Vanthienen, and J Leon Zhao. Transformational issues of big data and analytics in networked business. *MIS quarterly*, 40(4), 2016.

[2] Foster Provost and Tom Fawcett. Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1):51–59, 2013.

[3] EMC2 Capgemini. Big & fast data: The rise of insight-driven business, 2013.

[4] Barbara H Wixom and Jeanne W Ross. How to monetize your data. *MIT Sloan Management Review*, 58(3), 2017.

[2] We thank an anonymous reviewer for this remark.

[5] Mohsen Bahrami, Burcin Bozkaya, and Selim Balcisoy. Using behavioral analytics to predict customer invoice payment. *Big data*, 8(1):25–37, 2020.

[6] Lili Zhang, Jennifer Priestley, Joseph DeMaio, Sherry Ni, and Xiaoguang Tian. Measuring customer similarity and identifying cross-selling products by community detection. *Big Data*, 2020.

[7] Sobhan Sarkar, Sammangi Vinay, Rahul Raj, Jhareswar Maiti, and Pabitra Mitra. Application of optimized machine learning techniques for prediction of occupational accidents. *Computers & Operations Research*, 106:210–224, 2019.

[8] ITU. Measuring the information society report 2017, 2017.

[9] Christian Colot and Isabelle Linden. From mobile data towards better customer knowledge: proposals for an information framework. *Procedia Computer Science*, 52:75–82, 2015.

[10] Christian Colot, Isabelle Linden, and Philippe Baecke. A survey on mobile data uses. *International Journal of Decision Support System Technology (IJDSST)*, 8(2):29–49, 2016.

[11] David Godes and Dina Mayzlin. Using online conversations to study word-of-mouth communication. *Marketing science*, 23(4):545–560, 2004.

[12] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27:415–444, 2001.

[13] Yves-Alexandre de Montjoye, Jordi Quoidbach, Florent Robic, and Alex Sandy Pentland. Predicting personality using novel mobile phone-based metrics. In *Social computing, behavioral-cultural modeling and prediction*, pages 48–55. Springer Verlag, Berlin, 2013.

[14] Talia Gerstle, Maya Hardimon, Galen Hunt, Iona Main, Liza Paudel, Daniel Pontón, Sri Harshita Rallabhandi, Kaitlin Roh, Maria Romano, Emily Tenenbom, and Julia Tuttle. Assessing the use of call detail records (cdr) for monitoring mobility and displacement, 2021.

[15] Council of European Union The European Parliament. On the protection of natural persons about the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). *Official Journal of the European Union*, 679:1–88, 2016.

[16] Yuan Li. Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems*, 54(1):471–481, 2012.

[17] Christian Colot, Philippe Baecke, and Isabelle Linden. Alternatives for telco data network: the value of spatial and referral networks for churn detection. *Information Systems Management*, pages 1–19, 2021.

[18] Christian Colot, Philippe Baecke, and Isabelle Linden. Leveraging fine-grained mobile data for churn detection through essence random forest. *Journal of Big Data*, 8(1):1–26, 2021.

[19] Sophia Daskalaki, Ioannis Kopanas, M Goudara, and N Avouris. Data mining for decision support on customer insolvency in telecommunications business. *European Journal of Operational Research*, 145(2):239–255, 2003.

[20] Shin-Yuan Hung, David C Yen, and Hsiu-Yu Wang. Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3):515–524, 2006.

[21] Ron Kahan. Using database marketing techniques to enhance your one-to-one marketing initiatives. *Journal of Consumer Marketing*, 15(5):491–493, 1998.

[22] Irit Nitzan and Barak Libai. Social effects on customer retention. *Journal of Marketing*, 75(6):24–38, 2011.

[23] Aimée Backiel, Bart Baesens, and Gerda Claeskens. Predicting time-to-churn of prepaid mobile telephone customers using social network analysis. *Journal of the Operational Research Society*, 67, 2016.

[24] Wouter Verbeke, Karel Dejaeger, David Martens, Joon Hur, and Bart Baesens. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218(1):211–229, 2012.

[25] Ina Garnefeld, Andreas Eggert, Sabrina V Helm, and Stephen S Tax. Growing existing customers' revenue streams through customer referral programs. *Journal of Marketing*, 77(4):17–32, 2013.

[26] Iris Roelens, Philippe Baecke, and Dries F. Benoit. Identifying influencers in a social network: The value of real referral data. *Decision Support Systems*, pages –, 2016. Online before print.

[27] Shawndra Hill, Foster Provost, and Chris Volinsky. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, 21(2):256–276, 2006.

[28] Liye Ma, Ramayya Krishnan, and Alan L Montgomery. Latent homophily or social influence? an empirical analysis of purchase within a social network. *Management Science*, 61(2):454–473, 2014.

[29] H Cao, WS Dong, LS Liu, CY Ma, WH Qian, JW Shi, CH Tian, Yu Wang, David Konopnicki, Michal Shmueli-Scheuer, et al. Solomo analytics for telco big data monetization. *IBM Journal of Research and Development*, 58(5/6):9–1, 2014.

[30] Abou Zakaria Faroukhi, Imane El Alaoui, Youssef Gahi, and Aouatif Amine. Big data monetization throughout big data value chain: a comprehensive review. *Journal of Big Data*, 7(1):1–22, 2020.

[31] Payam Hanafizadeh and Mohammad Reza Harati Nik. Configuration of data monetization: A review of literature with thematic analysis. *Global Journal of Flexible Systems Management*, 21(1):17–34, 2020.

[32] Sharad Goel and Daniel G Goldstein. Predicting individual behavior with social networks. *Marketing Science*, 33(1):82–93, 2014.

[33] Robert R McCrae and Oliver P John. An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, 1992.

[34] Michal Kosinski, Yoram Bachrach, Pushmeet Kohli, David Stillwell, and Thore Graepel. Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, 95(3):357–380, 2014.

[35] Gregory Park, H. Andrew Schwartz, Johannes C. Eichsteadt, Margaret L Kern, Michal Kosinski, David J. Stillwell, Lyle H. Ungar, and Martin E P Seligman. Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6):1–25, 2015.

[36] Vivek K Singh, Lindsay Freeman, Bruno Lepri, and Alex Sandy Pentland. Predicting spending behavior using socio-mobile features. In *2013 Interna-*

tional Conference on Social Computing (SocialCom), pages 174–179. IEEE, 2013.

[37] Eric T Bradlow, Manish Gangwar, Praveen Kopalle, and Sudhir Voleti. The role of big data and predictive analytics in retailing. *Journal of Retailing*, 93(1):79–95, 2017.

[38] Jacques Bughin. Reaping the benefits of big data in telecom. *Journal of Big Data*, 3(1):14, 2016.

[39] Philippe Baecke and Dirk Van den Poel. Data augmentation by predicting spending pleasure using commercially available external data. *Journal of Intelligent Information Systems*, 36(3):367–383, 2011.

[40] Véronique Van Vlasselaer, Tina Eliassi-Rad, Leman Akoglu, Monique Snoeck, and Bart Baesens. Gotcha! network-based fraud detection for social security fraud. *Management Science*, 2016.

[41] Sofus A Macskassy and Foster Provost. Classification in networked data: A toolkit and a univariate case study. *Journal of Machine Learning Research*, 8(May):935–983, 2007.

[42] David Martens, Provost Foster, and Jessica Clark. Mining massive fine-grained behaviour data to improve analytics. *MIS quarterly*, 2016. Online before print.

[43] Julie Moeyersoms and David Martens. Including high-cardinality attributes in predictive models: A case study in churn prediction in the energy sector. *Decision Support Systems*, 72:72–81, 2015.

[44] Jasmien Lismont, Jan Vanthienen, Bart Baesens, and Wilfried Lemahieu. Defining analytics maturity indicators: A survey approach. *International Journal of Information Management*, 37(3):114–124, 2017.

[45] Scott Tonidandel and James M LeBreton. Determining the relative importance of predictors in logistic regression: An extension of relative weight analysis. *Organizational Research Methods*, 13(4):767–781, 2010.

[46] Oliver Hinz, Bernd Skiera, Christian Barrot, and Jan U Becker. Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing*, 75(6):55–71, 2011.

[47] Wouter Verbeke, David Martens, and Bart Baesens. Social network analysis for customer churn prediction. *Applied Soft Computing*, 14:431–446, 2014.

[48] Duncan J Watts and Peter Sheridan Dodds. Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4):441–458, 2007.

[49] Enric Junqué de Fortuny, David Martens, and Foster Provost. Predictive modeling with big data: is bigger really better? *Big Data*, 1(4):215–226, 2013.

[50] Christophe Van Den Bulte, Emanuel Bayer, Bernd Skiera, and Philipp Schmitt. How Customer Referral Programs Turn Social Capital into Economic Capital. *Journal of Marketing Research*, 2017.

[51] Leonardo Bursztyn, Florian Ederer, Bruno Ferman, and Noam Yuchtman. Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions. *Econometrica*, 82(4):1273–1301, 2014.

[52] Jeffrey S. Nevid and Amy Pastva. "I'm a Mac" versus "I'm a PC": Personality Differences between Mac and PC Users in a College Sample. *Psychology and Marketing*, 31(1):31–37, 2014.

[53] Svein Ottar Olsen, Ana Alina Tudoran, Pirjo Honkanen, and Bas Verplanken. Differences and Similarities between Impulse Buying and Variety Seeking: A Personality-based Perspective. *Psychology and Marketing*, 33(1):36–47, 2016.

[54] Joanne Hinds, Emma J Williams, and Adam N Joinson. "it wouldn't happen to me": Privacy concerns and perspectives following the cambridge analytica scandal. *International Journal of Human-Computer Studies*, 143:102498, 2020.

Appendix A

Definition of dependent variables

Industry	Case	DV category	Definition of event leading to 1 for Dependant variable (for DC 1 and DC 3 at least one event within the 90 days time frame) else 0	Present in DC1	Present in DC2	Present in DC3
App&soft	1	Real Transaction	free of charge call to the phone number to activate a software on a desktop	X		
	2	Real Transaction	international text message sent to activate an application on a mobile phone	X		
	3	Real Transaction	premium rate text message sent to get access to a training platform for driving licence during 2 or 3 days (depending on amount paid)	X		
	4	Real Transaction	premium rate text message to use an application			X
Charity	1	Real Transaction	premium rate text message sent to support a charity cause	X		
Entertainment	1	Real Transaction	premium rate text message sent to send a message on a chat forum	X		X
	2	Real Transaction	text message to subscribe to a paid subscription delivering weekly quizzes along with the possibility of winning a lottery prize (undetermined period)	X		
	3	Real Transaction	premium rate text message sent to participate to a lottery prize	X		
	4	Real Transaction	premium rate text message sent to participate to a lottery prize			X
	5	Real Transaction	premium rate text message sent to get some predictions about future (astrology)			X
Finance	1	Partnership	Customers of this Finance case are identified in the Customer Relationship Management System of one telephone company		X	
Information	1	Real Transaction	premium rate call sent to request coordinates of someone	X		X
	2	Real Transaction	premium rate call sent to request coordinates of someone	X		X
	3	Real Transaction	premium rate call sent to request coordinates of someone			X
	4	Real Transaction	premium rate call sent to request coordinates of someone			X
Parking	1	Real Transaction	text message sent to start/stop using a parking place, monthly invoice afterwards on phone bill	X		
	2	Real Transaction	text message sent to start/stop using a parking place, monthly invoice afterwards on phone bill	X		X
Public transport	1	Real Transaction	premium rate text message sent to buy a public transport for one hour	X		X
	2	Order	Call to booking centre of a public transport			X
Radio	1	Real Transaction	premium rate text message sent to ask for a specific song	X		
	2	Real Transaction	premium rate text message sent to get a chance to win a ticket concert	X		
	3	Real Transaction	premium rate text message sent to warn other road users about special traffic conditions	X		
	4	Real Transaction	Premium rate call to get a ticket concert trough a radio station	X		
	5	Real Transaction	Premium rate text message to participate to activities of the radio			X
	6	Real Transaction	Premium rate text message to participate to an event of the radio			X
Smartphone	1	Real Transaction	international text message sent to activate a common application of a specific smartphone provider	X		X
Taxi	1	Order	call to booking center of a taxi company	X		X
	2	Order	call to booking center of a taxi company	X		X
	3	Order	call to booking center of a taxi company	X		X
	4	Order	call to booking center of a taxi company	X		X
	5	Order	call to booking center of a taxi company	X		X
	6	Order	call to booking center of a taxi company			X
Television	1	Real transaction	premium rate text message to win a lottery prize trough a TV station	X		
	2	Real transaction	premium rate text message to win a lottery prize trough a TV station	X		
	3	Real transaction	premium rate text message to win a lottery prize trough a TV station	X		
	4	Real transaction	premium rate text message to win a lottery prize trough a TV station	X		
	5	Real transaction	premium rate text message to win a lottery prize trough a TV station			X

Appendix B

Descriptive statistics

Data contexts 1 and 2 (telecom operator 1):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Individual Demographics	All cases	age	98570	4.0000	131.0000	28.4467	10.4996
	All cases	Total net income divided by number of tax reports*					
	All cases	% of tax reports with positive net income*					
	All cases	% of dwellings by building type*					
	All cases	Proportion of land registry surface area*					
	All cases	Mean taxable value by hectares*					
	All cases	% of people by marital status*					
	All cases	% of men*					
	All cases	% of people by age category*					
	All cases	% of family by number of people*					
Network Business Adoption	App&soft 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	0.5000	0.0018	0.0275
	App&soft 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0037	0.0559
	App&soft 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	0.5000	0.0025	0.0354
	App&soft 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0051	0.0715
	App&soft 1	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0025	0.0435
	App&soft 1	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0025	0.0458
	App&soft 1	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0043	0.0658
	App&soft 1	Number of neighbours with behaviour	63248	0.0000	1.0000	0.0043	0.0658
	App&soft 1	Total of communication time spend with neighbours with business behaviour	63248	0.0000	1967.5000	0.7817	23.4361
	App&soft 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0183	0.0891
	App&soft 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.0391	0.1886
	App&soft 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0286	0.1242
	App&soft 2	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.0601	0.2614
	App&soft 2	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0145	0.0954
	App&soft 2	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0143	0.1022
	App&soft 2	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0315	0.1746
	App&soft 2	Number of neighbours with behaviour	63248	0.0000	5.0000	0.0344	0.1998
	App&soft 2	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2392.0000	3.9498	43.6209
	App&soft 3	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	0.5000	0.0032	0.0358
	App&soft 3	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0069	0.0751
	App&soft 3	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	0.5000	0.0050	0.0489
	App&soft 3	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0104	0.1013
	App&soft 3	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0038	0.0502
	App&soft 3	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0038	0.0545
	App&soft 3	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0078	0.0882
	App&soft 3	Number of neighbours with behaviour	63248	0.0000	2.0000	0.0079	0.0890
	App&soft 3	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2253.0000	1.2065	27.3039
	Charity 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0128	0.0725
	Charity 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0272	0.1516
	Charity 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0194	0.0981
	Charity 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0404	0.2037

* Descriptive statistics on ‘Local statistics’ sub-category variables within ‘Individual demographics’, postal codes and aggregated geographical levels are not disclosed due to the Non Disclosure Agreement. They have no missing value.

Descriptive statistics

Data contexts 1 and 2 (telecom operator 1):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Network Business Adoption	Charity 1	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0113	0.0850
	Charity 1	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0112	0.0917
	Charity 1	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0241	0.1534
	Charity 1	Number of neighbours with behaviour	63248	0.0000	3.0000	0.0248	0.1601
	Charity 1	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2236.0000	3.4877	46.9961
	Entertainment 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0041	0.0406
	Entertainment 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0087	0.0854
	Entertainment 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0060	0.0540
	Entertainment 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0128	0.1138
	Entertainment 1	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0041	0.0526
	Entertainment 1	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0040	0.0551
	Entertainment 1	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0085	0.0918
	Entertainment 1	Number of neighbours with behaviour	63248	0.0000	2.0000	0.0086	0.0940
	Entertainment 1	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2864.0000	1.2847	30.6074
	Entertainment 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0160	0.0854
	Entertainment 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0337	0.1776
	Entertainment 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0227	0.1091
	Entertainment 2	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0473	0.2265
	Entertainment 2	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0147	0.1004
	Entertainment 2	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0146	0.1061
	Entertainment 2	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0281	0.1653
	Entertainment 2	Number of neighbours with behaviour	63248	0.0000	3.0000	0.0294	0.1766
	Entertainment 2	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2751.0000	4.2945	52.0192
	Entertainment 3	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0066	0.0542
	Entertainment 3	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0141	0.1132
	Entertainment 3	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0094	0.0698
	Entertainment 3	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0194	0.1436
	Entertainment 3	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0063	0.0658
	Entertainment 3	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0063	0.0706
	Entertainment 3	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0125	0.1111
	Entertainment 3	Number of neighbours with behaviour	63248	0.0000	2.0000	0.0128	0.1156
	Entertainment 3	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2934.5000	2.2793	40.9617
	Finance 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.2685	0.3104
	Finance 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.8333	0.5667	0.6535
	Finance 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.3420	0.3598
	Finance 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	4.0000	0.7230	0.7778
	Finance 1	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.2615	0.3684
	Finance 1	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.2611	0.3933
	Finance 1	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.4084	0.4915
	Finance 1	Number of neighbours with behaviour	63248	0.0000	8.0000	0.5399	0.7701
	Finance 1	Total of communication time spend with neighbours with business behaviour	63248	0.0000	3065.0000	83.6670	231.1119

Descriptive statistics

Data contexts 1 and 2 (telecom operator 1):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Network Business Adoption	Information 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0027	0.0366
	Information 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0056	0.0742
	Information 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0035	0.0435
	Information 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0071	0.0873
	Information 1	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0042	0.0571
	Information 1	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0042	0.0604
	Information 1	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0067	0.0813
	Information 1	Number of neighbours with behaviour	63248	0.0000	2.0000	0.0068	0.0847
	Information 1	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2012.0000	1.4063	30.4393
	Information 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	#VALEUR!	#VALEUR!
	Information 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0652	0.2352
	Information 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0449	0.1483
	Information 2	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.0937	0.3088
	Information 2	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0276	0.1326
	Information 2	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0275	0.1432
	Information 2	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0566	0.2311
	Information 2	Number of neighbours with behaviour	63248	0.0000	3.0000	0.0603	0.2541
	Information 2	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2746.5000	8.9117	74.1545
	Parking 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0065	0.0545
	Parking 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0135	0.1132
	Parking 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0089	0.0685
	Parking 1	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0069	0.0701
	Parking 1	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0068	0.0742
	Parking 1	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0128	0.1124
	Parking 1	Number of neighbours with behaviour	63248	0.0000	3.0000	0.0131	0.1166
	Parking 1	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2582.5000	2.1144	37.4563
	Parking 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0543	0.1495
	Parking 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.5000	0.1144	0.3108
	Parking 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0774	0.1917
	Parking 2	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.1617	0.4006
	Parking 2	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0539	0.1868
	Parking 2	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0534	0.1992
	Parking 2	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.1031	0.3041
	Parking 2	Number of neighbours with behaviour	63248	0.0000	5.0000	0.1132	0.3516
	Parking 2	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2808.5000	17.0327	106.2282
	Public transport 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.2105	0.2855
	Public transport 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.3333	0.4467	0.6060
	Public transport 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.2754	0.3449
	Public transport 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	5.0000	0.5861	0.7539
	Public transport 1	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.1778	0.3141
	Public transport 1	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.1782	0.3379

Descriptive statistics

Data contexts 1 and 2 (telecom operator 1):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Network Business Adoption	Public transport 1	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.3011	0.4588
	Public transport 1	Number of neighbours with behaviour	63248	0.0000	9.0000	0.4132	0.7439
	Public transport 1	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2934.5000	61.7293	205.4456
	Radio 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0112	0.0690
	Radio 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.7500	0.0237	0.1444
	Radio 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0164	0.0913
	Radio 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.0341	0.1897
	Radio 1	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0105	0.0828
	Radio 1	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0106	0.0902
	Radio 1	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0217	0.1457
	Radio 1	Number of neighbours with behaviour	63248	0.0000	3.0000	0.0223	0.1520
	Radio 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0033	0.0410
	Radio 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0067	0.0833
	Radio 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0042	0.0485
	Radio 2	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0086	0.0987
	Radio 2	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0043	0.0579
	Radio 2	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0045	0.0619
	Radio 2	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0068	0.0823
	Radio 2	Number of neighbours with behaviour	63248	0.0000	2.0000	0.0070	0.0851
	Radio 2	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2400.5000	1.5835	36.0717
	Radio 3	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0116	0.0720
	Radio 3	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.2500	0.0246	0.1507
	Radio 3	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0163	0.0922
	Radio 3	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.0341	0.1915
	Radio 3	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0132	0.0967
	Radio 3	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0133	0.1037
	Radio 3	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0244	0.1544
	Radio 3	Number of neighbours with behaviour	63248	0.0000	3.0000	0.0252	0.1615
	Radio 3	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2855.0000	5.0869	62.9870
	Radio 4	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0218	0.0936
	Radio 4	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.0465	0.1985
	Radio 4	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0340	0.1303
	Radio 4	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	4.0000	0.0713	0.2740
	Radio 4	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0166	0.1011
	Radio 4	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0166	0.1099
	Radio 4	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0371	0.1891
	Radio 4	Number of neighbours with behaviour	63248	0.0000	4.0000	0.0392	0.2056
	Radio 4	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2130.0000	5.5147	59.5675
	Smartphone	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.2191	0.4250
	Smartphone	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.1458	0.2570
	Smartphone	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.3038	0.5371

Descriptive statistics

Data contexts 1 and 2 (telecom operator 1):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Network Business Adoption	Smartphone	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.1092	0.2774
	Smartphone	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.1972	0.3979
	Smartphone	Number of neighbours with behaviour	63248	0.0000	6.0000	0.2257	0.4903
	Smartphone	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2916.0000	33.4160	144.4460
	Smartphone	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.1043	0.2046
	Taxi 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0063	0.0525
	Taxi 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0137	0.1124
	Taxi 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0104	0.0756
	Taxi 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0220	0.1585
	Taxi 1	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0061	0.0620
	Taxi 1	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0061	0.0674
	Taxi 1	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0134	0.1151
	Taxi 1	Number of neighbours with behaviour	63248	0.0000	4.0000	0.0145	0.1296
	Taxi 1	Total of communication time spend with neighbours with business behaviour	63248	0.0000	1789.0000	1.7772	29.7326
	Taxi 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0055	0.0519
	Taxi 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0113	0.1065
	Taxi 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0075	0.0649
	Taxi 2	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0154	0.1330
	Taxi 2	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0051	0.0590
	Taxi 2	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0049	0.0625
	Taxi 2	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0098	0.0984
	Taxi 2	Number of neighbours with behaviour	63248	0.0000	4.0000	0.0105	0.1098
	Taxi 2	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2073.5000	1.4495	27.6333
	Taxi 3	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0036	0.0411
	Taxi 3	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0073	0.0833
	Taxi 3	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0050	0.0521
	Taxi 3	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0102	0.1048
	Taxi 3	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0051	0.0617
	Taxi 3	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0051	0.0646
	Taxi 3	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0087	0.0931
	Taxi 3	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2752.5000	1.2942	27.5880
	Taxi 4	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0090	0.0688
	Taxi 4	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.0185	0.1411
	Taxi 4	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0115	0.0816
	Taxi 4	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.0236	0.1693
	Taxi 4	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0125	0.0981
	Taxi 4	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0124	0.1022
	Taxi 4	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0202	0.1405
	Taxi 4	Number of neighbours with behaviour	63248	0.0000	4.0000	0.0221	0.1612
	Taxi 4	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2838.5000	3.0883	41.4237
	Taxi 5	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0053	0.0490

Descriptive statistics

Data contexts 1 and 2 (telecom operator 1):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Network Business Adoption	Taxi 5	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0110	0.0998
	Taxi 5	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0081	0.0657
	Taxi 5	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0166	0.1341
	Taxi 5	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0054	0.0578
	Taxi 5	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0051	0.0611
	Taxi 5	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0121	0.1092
	Taxi 5	Number of neighbours with behaviour	63248	0.0000	3.0000	0.0128	0.1186
	Taxi 5	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2481.0000	1.5218	29.8459
	Television 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0067	0.0536
	Television 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0139	0.1114
	Television 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0094	0.0692
	Television 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0194	0.1421
	Television 1	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0069	0.0693
	Television 1	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0070	0.0750
	Television 1	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0135	0.1155
	Television 1	Number of neighbours with behaviour	63248	0.0000	2.0000	0.0138	0.1195
	Television 1	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2129.0000	2.4129	38.3016
	Television 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0030	0.0349
	Television 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0062	0.0724
	Television 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0042	0.0453
	Television 2	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0034	0.0484
	Television 2	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0033	0.0511
	Television 2	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0065	0.0801
	Television 2	Number of neighbours with behaviour	63248	0.0000	2.0000	0.0065	0.0812
	Television 2	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2808.5000	1.1344	31.3534
	Television 3	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0097	0.0650
	Television 3	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0202	0.1347
	Television 3	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0146	0.0877
	Television 3	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.0000	0.0299	0.1793
	Television 3	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0112	0.0876
	Television 3	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0112	0.0941
	Television 3	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0217	0.1456
	Television 3	Number of neighbours with behaviour	63248	0.0000	3.0000	0.0225	0.1539
	Television 3	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2934.5000	3.6005	49.8433
	Television 4	Average of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0295	0.1092
	Television 4	Average of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	2.2000	0.0627	0.2304
	Television 4	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	23722	0.0000	1.0000	0.0446	0.1480
	Television 4	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	23722	0.0000	3.0000	0.0936	0.3119
	Television 4	Average business behaviour of neighbours of the customer	63248	0.0000	1.0000	0.0226	0.1172
	Television 4	Average business behaviour of neighbours of the customer weighted by total communication	63248	0.0000	1.0000	0.0223	0.1270

Descriptive statistics

Data contexts 1 and 2 (telecom operator 1):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Network Business Adoption	Television 4	Dummy with value 1 if one neighbour has the behaviour else 0	63248	0.0000	1.0000	0.0497	0.2173
	Television 4	Number of neighbours with behaviour	63248	0.0000	4.0000	0.0527	0.2379
	Television 4	Total of communication time spend with neighbours with business behaviour	63248	0.0000	2277.5000	7.1539	66.6828
Individual Mobile Behaviour	all cases	Lag 1 autoregressive model of nb calls	98536	-1.1459	0.2290	-0.0801	0.0835
		Lag 4 autoregressive model of nb calls	98536	-1.2147	0.6636	-0.0873	0.0850
		Lag 8 autoregressive model of nb calls	98536	-1.8791	1.1201	-0.0494	0.0748
		Lag 12 autoregressive model of nb calls	98536	-1.3689	1.1143	-0.0429	0.0716
		Lag 24 autoregressive model of nb calls	98536	-0.7668	1.1819	0.0391	0.0479
		Lag 18 autoregressive model of nb calls	98536	-0.7263	0.6347	-0.0490	0.0728
		Portion of all calls that were initiated by customer	98529	0.0000	1.0000	0.5286	0.1002
		Portion of all text messages that were initiated by customer	98516	0.0000	1.0000	0.4934	0.0704
		Average time between two call events	98533	0.0000	52812.9083	600.2177	953.4754
		Average time between two text messages	98485	0.0000	101970.5500	299.4425	1193.6892
		Total number of call interactions	98570	0.0000	4322.0000	377.1592	308.9746
		Total number of text messages interactions	98570	0.0000	8679.0000	1727.4576	1530.9366
		% calls occuring between 11 p.m. and 5 a.m.	98544	0.0000	0.8000	0.0520	0.0626
		% incoming calls responded	98462	0.0000	1.0000	0.8833	0.0820
		% sms answered within one hour	98474	0.0000	1.0000	0.5658	0.1759
		Variance of time between two call events	98533	0.0000	2789169838.6000	4391957.4032	32295605.5280
		Variance of time between two text messages events	98485	0.0000	2527732029.4000	2719637.7490	28068599.5530
		Metric assessing to which extend the customer is regular in his/her pattern of call from one week to another	98536	0.0071	2.2122	0.7108	0.2107
Network Based Mobile Behaviour		Contacts to interaction ratio text	98529	1.0000	368.0000	9.9282	7.5457
		Contacts to interaction ratio calls	98516	1.0000	3072.0000	44.2874	44.0738
		Entropy of contacts call	98529	0.0000	5.5095	2.1595	0.7069
		Entropy of contacts sms	98516	0.0000	6.5517	2.2129	0.7123
		Number of different contact persons (call)	98570	0.0000	551.0000	38.7601	22.8250
		Number of different contact persons (sms)	98570	0.0000	1256.0000	39.7487	23.0592

b) Categorical variables

Category	Case	Variable and value	Frequency	Percent
Individual Demographics*	all cases	Sex -Female	35357	35.87
		Sex - Male	63213	64.13
		Sex -Missing	0	0.00
		Postal code*		
		Aggregated geographical level 1*		
		Aggregated geographical level 2*		
		Aggregated geographical level 3*		

* Descriptive statistics on 'Local statistics' sub-category variables within 'Individual demographics', postal codes and aggregated geographical levels are not disclosed due to the Non Disclosure Agreement. They have no missing value.

Descriptive statistics

Data context 3 (telecom operator 2):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Individual Demographics*	Customer data	age	583143	4.0000	47.0347	159.0000	16.3431
	All cases	Total net income divided by number of tax reports*					
	All cases	% of tax reports with positive net income*					
	All cases	% of dwellings by building type*					
	All cases	Proportion of land registry surface area*					
	All cases	Mean taxable value by hectares*					
	All cases	% of people by marital status*					
	All cases	% of men*					
	All cases	% of people by age category*					
	All cases	% of family by number of people*					
Network Business Adoption	App&soft 4	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0120	0.0728
	App&soft 4	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0249	0.1504
	App&soft 4	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0143	0.0829
	App&soft 4	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0297	0.1713
	App&soft 4	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0089	0.0749
	App&soft 4	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0091	0.0816
	App&soft 4	Dummy with value 1 if one neighbour has the behaviour else 20	590626	0.0000	1.0000	0.0203	0.1411
	App&soft 4	Number of neighbours with behaviour	590626	0.0000	3.0000	0.0206	0.1438
	App&soft 4	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1212.0000	1.2935	16.8890
	Entertainment 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0007	0.0177
	Entertainment 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0014	0.0369
	Entertainment 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0007	0.0195
	Entertainment 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0016	0.0407
	Entertainment 1	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0007	0.0204
	Entertainment 1	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0006	0.0218
	Entertainment 1	Dummy with value 1 if one neighbour has the behaviour else 14	590626	0.0000	1.0000	0.0015	0.0388
	Entertainment 1	Number of neighbours with behaviour	590626	0.0000	5.0000	0.0016	0.0418
	Entertainment 1	Total of communication time spend with neighbours with business behaviour	590626	0.0000	920.5000	0.0851	4.2164
	Entertainment 4	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0061	0.0527
	Entertainment 4	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0128	0.1089
	Entertainment 4	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0075	0.0608
	Entertainment 4	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0155	0.1260
	Entertainment 4	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0035	0.0461
	Entertainment 4	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0037	0.0513
	Entertainment 4	Dummy with value 1 if one neighbour has the behaviour else 18	590626	0.0000	1.0000	0.0088	0.0933
	Entertainment 4	Number of neighbours with behaviour	590626	0.0000	3.0000	0.0089	0.0956
	Entertainment 4	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1135.5000	0.5434	10.6486
	Entertainment 5	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	0.6667	0.0014	0.0247
	Entertainment 5	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0029	0.0515
	Entertainment 5	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0017	0.0285
	Entertainment 5	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0035	0.0594

* Descriptive statistics on ‘Local statistics’ sub-category variables within ‘Individual demographics’, postal codes and aggregated geographical levels are not disclosed due to the Non Disclosure Agreement. They have no missing value.

Descriptive statistics

Data contexts 3 (telecom operator 2):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Network Business Adoption	Entertainment 5	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0006	0.0175
	Entertainment 5	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0006	0.0197
	Entertainment 5	Dummy with value 1 if one neighbour has the behaviour else 13	590626	0.0000	1.0000	0.0017	0.0409
	Entertainment 5	Number of neighbours with behaviour	590626	0.0000	2.0000	0.0017	0.0412
	Entertainment 5	Total of communication time spend with neighbours with business behaviour	590626	0.0000	672.0000	0.1045	4.6664
	Information 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0289	0.1140
	Information 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0602	0.2364
	Information 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0349	0.1313
	Information 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0728	0.2732
	Information 1	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0171	0.0995
	Information 1	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0181	0.1121
	Information 1	Dummy with value 1 if one neighbour has the behaviour else 11	590626	0.0000	1.0000	0.0426	0.2019
	Information 1	Number of neighbours with behaviour	590626	0.0000	4.0000	0.0454	0.2220
	Information 1	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1334.5000	3.3603	28.8152
	Information 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0099	0.0701
	Information 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0205	0.1445
	Information 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0109	0.0753
	Information 2	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0226	0.1554
	Information 2	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0097	0.0834
	Information 2	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0101	0.0897
	Information 2	Dummy with value 1 if one neighbour has the behaviour else 9	590626	0.0000	1.0000	0.0176	0.1314
	Information 2	Number of neighbours with behaviour	590626	0.0000	4.0000	0.0181	0.1373
	Information 2	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1643.0000	1.3156	17.8533
	Information 3	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0014	0.0256
	Information 3	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0029	0.0526
	Information 3	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0016	0.0278
	Information 3	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0032	0.0569
	Information 3	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0017	0.0354
	Information 3	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0018	0.0377
	Information 3	Dummy with value 1 if one neighbour has the behaviour else 10	590626	0.0000	1.0000	0.0032	0.0562
	Information 3	Number of neighbours with behaviour	590626	0.0000	2.0000	0.0032	0.0565
	Information 3	Total of communication time spend with neighbours with business behaviour	590626	0.0000	809.5000	0.2086	6.4857
	Information 4	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	0.7500	0.0015	0.0254
	Information 4	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	1.5000	0.0030	0.0523
	Information 4	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0018	0.0301
	Information 4	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0038	0.0620
	Information 4	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0011	0.0267
	Information 4	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0011	0.0289
	Information 4	Dummy with value 1 if one neighbour has the behaviour else 12	590626	0.0000	1.0000	0.0026	0.0507
	Information 4	Number of neighbours with behaviour	590626	0.0000	2.0000	0.0026	0.0511
	Information 4	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1315.0000	0.1771	6.0291

Descriptive statistics

Data contexts 3 (telecom operator 2):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Network Business Adoption	Parking 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0185	0.0947
	Parking 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0385	0.1963
	Parking 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0214	0.1053
	Parking 2	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0445	0.2182
	Parking 2	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0154	0.1008
	Parking 2	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0160	0.1097
	Parking 2	Dummy with value 1 if one neighbour has the behaviour else 15	590626	0.0000	1.0000	0.0319	0.1756
	Parking 2	Number of neighbours with behaviour	590626	0.0000	6.0000	0.0337	0.1914
	Parking 2	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1173.0000	2.1308	22.2675
	Public transport 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0247	0.1132
	Public transport 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0512	0.2347
	Public transport 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0270	0.1215
	Public transport 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0561	0.2524
	Public transport 1	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0229	0.1266
	Public transport 1	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0235	0.1353
	Public transport 1	Dummy with value 1 if one neighbour has the behaviour else 16	590626	0.0000	1.0000	0.0409	0.1980
	Public transport 1	Number of neighbours with behaviour	590626	0.0000	4.0000	0.0444	0.2241
	Public transport 1	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1476.5000	2.7689	25.2807
	Public transport 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0004	0.0132
	Public transport 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0008	0.0271
	Public transport 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0004	0.0148
	Public transport 2	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0009	0.0314
	Public transport 2	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0004	0.0159
	Public transport 2	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0004	0.0170
	Public transport 2	Dummy with value 1 if one neighbour has the behaviour else 5	590626	0.0000	1.0000	0.0007	0.0262
	Public transport 2	Number of neighbours with behaviour	590626	0.0000	2.0000	0.0007	0.0269
	Public transport 2	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1116.5000	0.0469	3.4310
	Radio 5	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0126	0.0749
	Radio 5	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0262	0.1553
	Radio 5	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0157	0.0885
	Radio 5	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0326	0.1835
	Radio 5	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0061	0.0562
	Radio 5	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0065	0.0660
	Radio 5	Dummy with value 1 if one neighbour has the behaviour else 19	590626	0.0000	1.0000	0.0174	0.1309
	Radio 5	Number of neighbours with behaviour	590626	0.0000	4.0000	0.0182	0.1393
	Radio 5	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1147.5000	1.3129	18.4960
	Radio 6	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0055	0.0494
	Radio 6	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0115	0.1032
	Radio 6	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0067	0.0576
	Radio 6	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0140	0.1206
	Radio 6	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0024	0.0353

Descriptive statistics

Data contexts 3 (telecom operator 2):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Network Business Adoption	Radio 6	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0026	0.0418
	Radio 6	Dummy with value 1 if one neighbour has the behaviour else 17	590626	0.0000	1.0000	0.0073	0.0850
	Radio 6	Number of neighbours with behaviour	590626	0.0000	3.0000	0.0074	0.0876
	Radio 6	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1212.0000	0.5725	12.5696
	Smartphone 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0557	0.1571
	Smartphone 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.1159	0.3258
	Smartphone 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0651	0.1760
	Smartphone 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.1355	0.3657
	Smartphone 1	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0378	0.1512
	Smartphone 1	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0391	0.1662
	Smartphone 1	Dummy with value 1 if one neighbour has the behaviour else 8	590626	0.0000	1.0000	0.0839	0.2772
	Smartphone 1	Number of neighbours with behaviour	590626	0.0000	6.0000	0.0902	0.3097
	Smartphone 1	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1348.0000	5.9535	37.8084
	Taxi 1	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0009	0.0222
	Taxi 1	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0020	0.0455
	Taxi 1	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0010	0.0236
	Taxi 1	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	3.0000	0.0021	0.0490
	Taxi 1	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0012	0.0305
	Taxi 1	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0013	0.0323
	Taxi 1	Dummy with value 1 if one neighbour has the behaviour else 3	590626	0.0000	1.0000	0.0021	0.0462
	Taxi 1	Number of neighbours with behaviour	590626	0.0000	3.0000	0.0022	0.0484
	Taxi 1	Total of communication time spend with neighbours with business behaviour	590626	0.0000	973.0000	0.1372	5.4565
	Taxi 2	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0011	0.0240
	Taxi 2	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0023	0.0492
	Taxi 2	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0012	0.0258
	Taxi 2	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0025	0.0527
	Taxi 2	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0011	0.0267
	Taxi 2	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0011	0.0285
	Taxi 2	Dummy with value 1 if one neighbour has the behaviour else 2	590626	0.0000	1.0000	0.0023	0.0479
	Taxi 2	Number of neighbours with behaviour	590626	0.0000	5.0000	0.0024	0.0513
	Taxi 2	Total of communication time spend with neighbours with business behaviour	590626	0.0000	971.0000	0.1397	5.4957
	Taxi 3	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0009	0.0213
	Taxi 3	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0018	0.0432
	Taxi 3	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0010	0.0227
	Taxi 3	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0020	0.0463
	Taxi 3	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0013	0.0304
	Taxi 3	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0013	0.0324
	Taxi 3	Dummy with value 1 if one neighbour has the behaviour else 0	590626	0.0000	1.0000	0.0023	0.0482
	Taxi 3	Number of neighbours with behaviour	590626	0.0000	2.0000	0.0024	0.0495
	Taxi 3	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1476.5000	0.1686	6.5072
	Taxi 4	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0027	0.0374

Descriptive statistics

Data contexts 3 (telecom operator 2):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Network Business Adoption	Taxi 4	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0057	0.0767
	Taxi 4	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0030	0.0405
	Taxi 4	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0063	0.0831
	Taxi 4	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0037	0.0514
	Taxi 4	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0038	0.0548
	Taxi 4	Dummy with value 1 if one neighbour has the behaviour else 1	590626	0.0000	1.0000	0.0068	0.0822
	Taxi 4	Number of neighbours with behaviour	590626	0.0000	4.0000	0.0072	0.0891
	Taxi 4	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1124.5000	0.4652	10.4221
	Taxi 5	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0005	0.0158
	Taxi 5	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0010	0.0319
	Taxi 5	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0005	0.0167
	Taxi 5	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0011	0.0337
	Taxi 5	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0006	0.0212
	Taxi 5	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0006	0.0226
	Taxi 5	Dummy with value 1 if one neighbour has the behaviour else 7	590626	0.0000	1.0000	0.0011	0.0336
	Taxi 5	Number of neighbours with behaviour	590626	0.0000	2.0000	0.0011	0.0342
	Taxi 5	Total of communication time spend with neighbours with business behaviour	590626	0.0000	762.0000	0.0755	4.0717
	Taxi 6	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0006	0.0169
	Taxi 6	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0011	0.0341
	Taxi 6	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0006	0.0182
	Taxi 6	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0013	0.0369
	Taxi 6	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0006	0.0210
	Taxi 6	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0006	0.0223
	Taxi 6	Dummy with value 1 if one neighbour has the behaviour else 4	590626	0.0000	1.0000	0.0011	0.0328
	Taxi 6	Number of neighbours with behaviour	590626	0.0000	2.0000	0.0011	0.0339
	Taxi 6	Total of communication time spend with neighbours with business behaviour	590626	0.0000	804.0000	0.0712	3.8663
	Television 5	Average of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0012	0.0235
	Television 5	Average of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0026	0.0485
	Television 5	Maximum of cliques mean behaviour (customer excluded) where the customer belongs to	111516	0.0000	1.0000	0.0015	0.0269
	Television 5	Maximum of cliques number of behaviours (customer excluded) where the customer belongs to	111516	0.0000	2.0000	0.0031	0.0559
	Television 5	Average business behaviour of neighbours of the customer	590626	0.0000	1.0000	0.0006	0.0191
	Television 5	Average business behaviour of neighbours of the customer weighted by total communication	590626	0.0000	1.0000	0.0007	0.0215
	Television 5	Dummy with value 1 if one neighbour has the behaviour else 6	590626	0.0000	1.0000	0.0017	0.0412
	Television 5	Number of neighbours with behaviour	590626	0.0000	2.0000	0.0017	0.0414
	Television 5	Total of communication time spend with neighbours with business behaviour	590626	0.0000	1060.0000	0.1292	5.8264
Individual Mobile Behaviour	All cases	Lag 1 autoregressive model of nb calls	901758	-3.9858	0.7120	-0.0796	0.0870
	All cases	Lag 4 autoregressive model of nb calls	901758	-3.9507	33.5054	-0.1176	0.1785
	All cases	Lag 8 autoregressive model of nb calls	901758	-32.4488	120.7712	-0.0670	0.4808
	All cases	Lag 12 autoregressive model of nb calls	901758	-16.9541	138.2246	-0.0619	0.5307
	All cases	Lag 18 autoregressive model of nb calls	901758	-5.4944	28.3307	0.0258	0.1145

Descriptive statistics

Data contexts 3 (telecom operator 2):

a) Quantitative variables

Category	Case	Variable	N	MIN	MAX	MEAN	STD
Individual Mobile Behavior	All cases	Lag 24 autoregressive model of nb calls	901758	-0.9427	0.9458	-0.0626	0.0781
	All cases	Entropy of places visited: large values indicate regular visits to many places	904010	0.0000	6.1368	1.7815	0.8906
	All cases	Portion of all calls that were initiated by customer	902028	0.0000	1.0000	0.5654	0.1939
	All cases	Portion of all text messages that were initiated by customer	734518	0.0000	1.0000	0.4637	0.1729
	All cases	Average time between two call events	899639	0.0833	128897.4333	2239.7106	4514.6037
	All cases	Average time between two text messages	707889	0.0000	129204.5333	3150.4999	6718.9759
	All cases	Average daily distance traveled	904010	0.0000	3285.5448	69.6894	126.9736
	All cases	Average daily radius of gyration	904010	0.0000	292.7032	25.7177	34.0485
	All cases	Total number of call interactions	904947	0.0000	4541.0000	205.2273	233.5884
	All cases	Number of places visited	904010	1.0000	32433.0000	532.3920	597.4054
	All cases	Total number of text messages interactions	904947	0.0000	28671.0000	222.1078	563.0464
	All cases	% calls occuring between 11 p.m. and 5 a.m.	902029	0.0000	1.0000	0.0187	0.0435
	All cases	% incoming calls responded	887817	0.8333	1.0000	1.0000	0.0004
	All cases	% sms answered within one hour	719389	0.0000	1.0000	0.4992	0.2715
	All cases	Variance of time between two call events	899639	0.0000	4151071343.3000	24698423.1320	108659155.7600
	All cases	Variance of time between two text messages events	707889	0.0000	4183069055.2000	69061668.1440	216016503.1900
	All cases	Metric assessing to which extend the customer is regular in his/her pattern of call from one week to another	901758	0.0001	2.5096	0.5825	0.1715
Network Based Mobile Behaviour	All cases	Contacts to interaction ratio text	902028	1.0000	1316.0000	6.7423	7.5095
	All cases	Contacts to interaction ratio calls	734518	1.0000	27264.0000	52.4555	146.7768
	All cases	Entropy of contacts call	902028	0.0000	6.6762	1.9791	0.7634
	All cases	Entropy of contacts sms	734518	0.0000	6.0606	0.8013	0.6132
	All cases	Number of different contact persons (call)	904947	0.0000	1679.0000	30.4851	28.1807
	All cases	Number of different contact persons (sms)	904947	0.0000	544.0000	3.9951	4.4665

b) Categorical variables

Category	Case	Variable and value	Frequency	Percent
Individual Demographics*	all cases	Sex -Female	35357	36.48
		Sex - Male	63213	62.14
		Sex -Missing	0	1.38
		Postal code*	[redacted]	[redacted]
		Aggregated geographical level 1*	[redacted]	[redacted]
		Aggregated geographical level 2*	[redacted]	[redacted]
		Aggregated geographical level 3*	[redacted]	[redacted]

* Descriptive statistics on 'Local statistics' sub-category variables within 'Individual demographics', postal codes and aggregated geographical levels are not disclosed due to the Non Disclosure Agreement. They have no missing value.

Appendix C

Individual results by business activity for DC 1

Industry	Case	% Business	AUC base	Delta AUC: Network Business Adoption	Delta AUC: Individual Telco Usage	Delta AUC: Network Telco Interaction	Delta AUC: All Personality	Delta AUC: All
App&soft	1	0.27%	66.54%	0.14%	2.09%	1.72%	1.86%	1.86%
	2	1.15%	78.63%	1.91%	2.31%	2.16%	3.73%	4.66%
	3	0.37%	69.99%	0.32%	0.00%	0.32%	0.00%	0.00%
Charity	1	1.10%	71.07%	1.11%	1.03%	0.01%	1.19%	2.21%
Entertainment	1	0.42%	71.41%	0.48%	8.76%	7.12%	10.27%	9.87%
	2	1.48%	76.55%	1.00%	1.55%	2.35%	2.63%	3.49%
	3	0.70%	73.60%	1.59%	1.82%	0.99%	1.57%	3.20%
Information	1	0.51%	88.64%	0.17%	1.74%	2.00%	2.00%	2.00%
	2	2.63%	65.57%	1.85%	10.56%	11.51%	12.30%	12.80%
Parking	1	0.67%	71.94%	1.04%	1.59%	2.05%	2.21%	2.87%
	2	5.18%	71.02%	0.78%	1.49%	1.63%	2.38%	3.18%
Public transport	1	14.61%	73.21%	2.78%	1.30%	1.08%	1.47%	3.64%
Radio	1	0.93%	73.68%	1.12%	2.34%	2.26%	3.24%	3.92%
	2	0.44%	89.71%	0.65%	1.12%	1.46%	1.44%	1.44%
	3	1.26%	69.73%	2.67%	5.29%	3.43%	5.45%	6.72%
	4	1.24%	67.89%	3.48%	5.40%	6.01%	6.43%	8.14%
Smartphone	1	10.96%	57.22%	3.09%	4.64%	3.27%	5.45%	7.19%
Taxi	1	0.56%	87.46%	1.22%	2.81%	2.23%	3.16%	3.42%
	2	0.46%	92.95%	0.21%	0.37%	0.53%	0.55%	0.65%
	3	0.51%	91.02%	0.49%	0.95%	1.19%	1.34%	1.34%
	4	1.37%	92.43%	0.28%	0.62%	0.84%	1.03%	1.22%
	5	0.44%	87.63%	1.52%	1.73%	1.62%	2.07%	2.76%
Television	1	0.68%	73.96%	3.32%	1.22%	0.38%	1.07%	4.13%
	2	0.36%	67.16%	1.86%	2.05%	1.86%	2.05%	2.05%
	3	1.18%	63.66%	3.40%	2.47%	1.33%	2.42%	5.39%
	4	1.79%	70.77%	1.40%	3.40%	4.73%	5.17%	5.67%
All		1.97%	75.52%	1.46%	2.64%	2.47%	3.17%	3.99%

Appendix D

Individual results by business activity for DC 3

Industry	Case	% Business	AUC base	Delta AUC: Network Business Adoption	Individual Telco Usage	Network Telco Interaction	All Personality	All	Individual Telco Usage - spatial behaviour	All Personality - spatial behaviour	All - spatial behaviour
App&soft	4	1.52%	60.01%	0.65%	3.49%	4.08%	4.82%	4.83%	3.34%	4.73%	4.73%
Entertainment	1	0.13%	75.20%	0.20%	11.95%	8.57%	12.45%	12.45%	10.35%	11.42%	11.42%
	4	0.58%	81.17%	0.19%	0.63%	0.45%	1.16%	1.28%	0.40%	0.79%	0.95%
	5	0.13%	76.14%	0.34%	13.64%	8.63%	14.60%	14.60%	12.14%	12.99%	12.99%
Information	1	2.57%	78.89%	1.55%	7.88%	6.80%	8.39%	8.60%	7.38%	7.89%	8.14%
	2	1.71%	79.62%	1.06%	7.19%	6.51%	7.63%	7.77%	6.94%	7.32%	7.47%
	3	0.42%	77.55%	0.06%	3.74%	2.25%	4.33%	4.33%	2.22%	3.03%	3.03%
	4	0.25%	77.27%	0.22%	4.92%	2.80%	5.18%	5.18%	3.26%	4.05%	4.05%
Parking	2	2.20%	81.05%	1.09%	4.76%	2.83%	5.37%	5.83%	2.23%	3.47%	4.00%
Public transport	1	4.01%	84.65%	1.33%	4.50%	0.96%	4.52%	5.03%	2.49%	2.57%	3.16%
	2	0.07%	94.69%	0.07%	0.55%	0.41%	0.54%	0.54%	0.50%	0.53%	0.53%
Radio	5	0.70%	86.83%	1.14%	2.55%	1.69%	2.78%	3.10%	1.67%	2.11%	2.46%
	6	0.30%	86.59%	(0.29%)	2.60%	1.58%	2.67%	2.84%	1.77%	2.03%	2.22%
Smartphone	1	6.26%	65.52%	2.05%	7.36%	3.12%	7.45%	8.15%	4.80%	4.92%	5.69%
Taxi	1	0.23%	95.72%	0.14%	0.99%	0.80%	1.14%	1.14%	0.99%	1.14%	1.14%
	2	0.17%	94.54%	0.05%	2.94%	0.43%	2.91%	2.91%	2.02%	2.05%	2.05%
	3	0.24%	93.64%	(0.01%)	0.94%	0.83%	1.13%	1.13%	0.98%	1.19%	1.19%
	4	0.62%	94.75%	0.02%	0.98%	0.75%	1.20%	1.20%	0.90%	1.16%	1.16%
	5	0.12%	94.87%	0.08%	1.15%	0.51%	1.26%	1.26%	1.09%	1.19%	1.19%
	6	0.11%	96.32%	0.10%	0.92%	0.57%	0.93%	0.93%	0.95%	0.97%	0.97%
Television	5	0.10%	74.52%	0.56%	11.25%	5.61%	11.23%	11.23%	9.94%	9.84%	9.84%
All		1.07%	83.31%	0.50%	4.52%	2.87%	4.84%	4.97%	3.64%	4.07%	4.21%