Intelligent use of mobile positioning data to measure domestic tourism in Peru
Caso: Campaña Vamos Pal Norte, Junio de 2017.

Authors: Walter Vizarreta [a*], Leyge Villanueva[a]
[a] Comisión de Promoción del Perú para la Exportación y el Turismo, Calle Uno Oeste 50, Edificio Mincetur, Pisos 13 y 14, San Isidro, Peru.

Address: Calle Uno Oeste 50, Edificio Mincetur, Pisos 13 y 14, San Isidro. Peru. Tel.: + 511 616 7300 Extensions: 1307 / 1215; E-mail address: wvizarreta@promperu.gob.pe; lvillanueva@promperu.gob.pe

Abstract
The measurement of domestic tourism with detail and specific information usually implies a market research that requires periodic data (monthly, quarterly or semiannually) and the methodology of calculation and allocation is, in the majority of cases, per year. Therefore, domestic tourism figures can only be obtained within the annual statistics and the results are issued after a long period of time (up to six months).

Looking for information on domestic travel behavior in the shortest time possible, Foreign Trade and Tourism Specialized Technical Agency (PromPerú) started a Big Data project as a management strategy, in order to explore different sources of information and identify domestic tourist movements.

In order to learn the mobilizing effect of travel promotion campaigns in one specific market segment - Peruvians living in districts of Lima with the largest population belonging to the low (C) socioeconomic level (San Juan de Lurigancho, Chorrillos, Independencia and San Martín de Porres) -, PromPerú sought to use mobile positioning data.

This study allowed to track the behavior of internal transfers within Peru, for thirty days, in order to learn the destination visited, stay and demographic profiles of the selected target population in the shortest possible time.

To do this, PromPerú used data extracted from Telefónica's network systems: information about each mobile user and its location. This information was anonymized and extrapolated according to the population distribution of official and variable statistics such as billing, contracted plan, number of teams, among others.

This information creates the opportunity to observe the almost immediate effects of important events in tourism, respecting data protection laws.

Introduction
This document presents an alternative approach for obtaining and analyzing domestic tourism indicators for Lima (Peru), based on mobile positioning data used on the monitoring of the travel promotion campaign *Vamos Pal Norte*, implemented by Peru Export and Tourism Promotion Board (PromPerú) from June 1st to the 30th, 2017.

A total change began with the arrival of databases of macro data or massive data also known as Big Data, specifically on the ways of storing and processing data in addition to technological developments, which expands the possibilities for exploiting the value of new sources of information, from different sources and in different formats.

The Statistical Office of the European Union (2014) shows the possibilities and challenges of using mobile positioning as a new source of data, based on a research that set the relevance of this type of data to generate statistical information about domestic and inbound tourism.

In 2016, Statistics Finland work closely with Finnish mobile network operators to convert the Call Detail Records (CDR) and Data Detail Records (DDR) into specific monthly aggregated data such as the number of trips and their length per country of destination to obtain tourism statistics. Meanwhile, BPS Statistics Indonesia started using mobile positioning data from the top operator in Indonesia in border areas to measure the arrival of cross border visitors (Esko, 2018).

Likewise, in 2016 PromPerú took part on the 103rd session of the UNWTO Executive Council held in Spain, which established the progress in the use of Big Data for tourism. Henceforth, PromPerú began the project Big Data in order to use new sources of information to measure tourism and strengthen tourism statistics in Peru.

To that effect, this research shows the results of the use of mobile positioning information as an alternative indicator to monitor visitor's movement patterns, both at interregional and intraregional levels, from the 1st to the 30th of June 2017. This analysis allows to answer the following questions: Which places have the national travelers visited? For how long have the national travelers stayed at the area under evaluation? With the advantage that the mobile positioning data are available and can be analyzed right after the closing of the traveling promotion campaign; unlike traditional quantitative researches, that consider the information collection process after the campaign was launched, and which results to be analyzed are gathered after about ninety days, due to the time that processing data can take.

Therefore, knowing the opportunities that mobile positioning data represents to measure indicators in tourism sector, through the analysis of this type of data related to the time the *Vamos Pal Norte* campaign was scheduled, it allows to identify the feasibility and restrictions that arise when using this tool.

The growing prominence of cell phones has generated dependency on people either for entertainment, to be informed or to get in contact. In that sense, it can be inferred that where the cell phone stays overnight usually, it is the place where the individual lives. According to statistics about information and communication technology from Lima Metropolitana, during 2016, 93.3% of households had at least one family member with a cell phone (Instituto Nacional de Estadística e Informática, 2016). Furthermore, according to the "Habits, uses and attitudes towards mobile networks" report, 48% of Peruvian citizen own a cellphone, i.e. 15.1 million of Peruvian citizens. Similarly, 81% of the population between 12 and 70 years old from urban areas own a cellphone (Ipsos Group S.A., 2017).

During 2017, Peru registered more than 37 million of active mobile lines and achieved a penetration of 127 lines for every 100 inhabitants. Furthermore, the mobile network operators Telefonica and Claro, achieved 38,2% and 32,1% market share respectively. Meanwhile, new operators achieved 29.7% of the telephone mobile market (Organismo Supervisor de Inversión Privada en Telecomunicaciones, 2018). (See Annex 3: Mobile Service Indicators 2017).

During the first quarter of 2017, the weather changes known as Coastal El Niño caused one of the biggest natural disasters of the last decade in Peru, and affected mainly northern regions of Peru. In this context, PromPerú developed a plan to tackle the decrease of tourists' flows nationwide, as a result of natural disasters. As part of this plan, the campaign Muévete Perú was designed. The launch of the campaign happened during Easter, with the hashtag #ElTurismoAyuda. Later the promotion of trips for the long weekend of May 1st was carried out; and the campaign with the hashtag *#VamospalNorte* was

launched to promote tourism in the following regions: Amazonas, Ancash, Cajamarca, Lambayeque, La Libertad, Loreto, San Martín, Piura and Tumbes. (See annex 4: Stages of the campaign *Muévete Perú*).

PromPerú have carried out market research for more than ten years, to keep track of national tourists travel behavior, from those who live in Lima, Arequipa, Trujillo, Chiclayo, Piura and Huancayo and belong to Hight (A), Medium (B) and Low (C) socio economic levels.

Considering the mobile service indicator of the Peruvian telecommunication market and the availability of Telefonica *Big Data* service, PromPerú decided to contract the *Smart Steps* service in order to know the flow of Peruvian who live in Lima Metropolitana to destinations outside their usual environment during the campaign *#Vamospalnorte.*

Literature review

According to the definition of Toscón (2013) the term *Big Data* is related to three concepts: volume, variability and velocity. Furthermore, the concept refers to managing large amounts of data that "is analyzed at high speed and can present complex variability in its structure´s composition" (Toscón, 2013, p. 95). Therefore, managing these large amounts of data represents a challenge for organizations because the data needs to be processed quickly applying new technologies to meet market demands in order to create ideas that lead to better strategic decisions.

The use of *Big Data* in the tourism sector is very valuable because of its high potential to be used as a predictor of tourist flows for specific destinations (Ramos Lobo & Pedregal Mateos, 2015). Tourist flows outside the usual environment, either inside or outside the country of residence, stand out among relevant indicators for tourism.

It should be highlighted that the list of *Big Data* sources is extensive (Beresewicz, Lehtonen, Reis, Di Consiglio, & Karlberg, 2018). In this sense, the data sources that are considered very valuable for the tourism sector are: data from cell phones (detailed call records, location of mobile phone) data from credit card usage (transactions), data from travel websites (comments from TripAdvisor) and data from social networks (comments from Twitter, Facebook, among others). This diversity of sources represents a positive panorama about this new way to manage large amounts of data.

On the other hand, mobile positioning refers to traceability or tracking of geographical coordinates of the location of phones in the mobile network (Ahas, Aasa, & Silm, 2007). As it was anticipated, the location of mobile phones are a *Big Data* source. This information was already fundamental for mobile operators, since it is necessary to route the device´s owner phone calls, to manage their networks effectively, and to manage resources where they are more needed.

In 2012, Telefonica Corporation, through its global business unit, Dynamic Insight, launched Smart Steps to provide anonymous, aggregated[1] and extrapolated[2] data that allows to know the movement of cell phones that represent people who move toward a specific place and in a specific period of time (Zamora, 2016). Therefore, this technological tool processes, by using Big Data tools, billions of events per day generated in Telefónica's mobile network, 24 hours a day, 365 days

[1] With the goal of complying with imposed restrictions by the eighth article that refers to secrecy and inviolability of communications, the law 27336 for development of functions and faculties of the Supervisory Agency for Private Investment in Communications (OSIPTEL by its initials in Spanish) and the law 29733 for Personal Data Protection; Telefonica does not provide data that refer to groups smaller than ten people. In that sense the restriction about the minimum number is relative to the number of trips. For this work the numbers provided are results from average calculations.

2 Since only data from Movistar´s travelers is acquired, Telefonica compares its numbers with official figures about population provided by the National Institute of Statistics and Information (INEI by its initials in Spanish), to extend the behaviors to the target group from the study area. To reduce biases in the distribution of Movistar´s customers and the population distribution provided by INEI, Telefonica extrapolates social strata by areas, gender, age and socioeconomic level. This procedure generates a weight for the combination of strata for each studied user.

a year. Additionally, mobile data can be analyzed considering users profiles, based on the information provided by a Telefónica's customer when buying a phone line (Zamora, 2016). Since 2016, Telefonica has the Big Data service unit called Last Universal Common Ancestor - LUCA that offers solutions such as Smart Steps service that combines mobility data, volumes, frequency of trips and demographic profile of people (Telefonica Data Unit, 2018). For this purpose, the corporation uses diverse infrastructure to get information from Telefónica's antennas toward clusters of servers with different large data processing programs.

The use of mobile positioning to generate data raises expectations among companies, especially from the tourism sector. Below are listed the benefits and barriers of using this tool. The following table lists some of the benefits and barriers identified as a result of this research:

Table 1: *Benefits and barriers of using mobile positioning data*

Benefits	Barriers
It is an effective method to collect data which raises interest within the tourism research community.	It does not consider information about travel behavior such as travel motivations, planning, among others.
It can be complemented by traditional data collecting methods.	It must consider legal aspects of privacy law and data protection.
It helps to reduce costs of collecting data for decisions regarding tourism campaigns.	The operator can charge a fee for using technology to process data.
Samples can be bigger than samples covered by traditional surveys.	It is vulnerable to communication market characteristics such as cost of calls, ways of using mobile phones, lost connections, among others.

Note. Source prepared by the authors.

It has to be highlighted that the use of mobile positioning data and traditional methods are complementary. The mobile network operator can guarantee to their customers that the data that it is going to be used will be anonymized, and at the end, the customer can request data by territories of entry and exit and/or specific periods of time based on the transfers of users that are of interest to their customers.

Smart Steps Methodology

The Smart Steps process to obtain mobile positioning data is developed in seven steps, according to the information provided by C. Zapata (personal communication, August, 22nd, 2018):

Phase 1: mobile phone users interact using their mobile devices with Telefonica´s network when they travel, by text messaging, call or internet connection. These events as well as the user identifier are collected by Telefonica´s antennas. Furthermore, data is complemented by CRM information. (See annex 1: Structure of information sources and annex 2: Smart Steps: Process of obtaining data from mobile positioning).

Phase 2: In order to protect the user's identity, the user identifier that accompanies the event is a code generated by an anonymization algorithm. These data are encoded or anonymized by the *Smart Steps* platform that uses a hash[3] function to generate an unrepeated code for each user, preventing to obtain the user's identity from the generated code. The algorithms used in the anonymization process have two main characteristics: on one hand, they prevent to reverse the process; therefore it is not possible under no circumstances revealing the users' or travelers´ identity, for the purposes of this research. On the other hand, the identifiers for each mobile line are constant over time, which allows to avoid data duplication, as well as tracking the mobility of movements. Once the

[3] A cryptographic hash function, usually known as hash, is a mathematical algorithm that transforms any arbitrary block of data into a new series of characters with a fixed length. No matter the length of the input data, the value of the output hash will always have the same length (Kaspersky Lab Daily, 2104).

research was conducted with these travelers, the records are added in groups of at least ten travelers in order to follow current legislation.

From this point onward, it is not possible to extract a user's identity, only the codes are available.

Phase 3: Raw data are stored in Telefónica's Cloud platforms. These data are not structured.

Phase 4: To transform these raw data into a source of information for analysts, several algorithms are implemented using *Sparks* and *Hadoop* to transform large quantities of data. The result of these processes are relational databases that contain stays, movements and points of interest for different travelers. This information is ready to be used by analysts.

Phase 5: At this point databases previously obtained, are exploited and transformed into tables with required data from the customer, in this case, PromPerú. To do this, tools such as *Python* or *PostgreSQL* are used.

Phase 6: Data from these tables are changed and aggregated to provide the customer with different insights. In the case of tourism, information is offered as: time ranges, days of week, socioeconomic levels and gender. At this point of the process tools such as Python are used.

Phase 7: The obtained insights are reflected into deliverables. These deliverables can be of several types: plain text files, spreadsheets or visualizers. At this point in the process tools such as *Python, Excel* and *Tableau* are used.

Methodology

The study area for data positioning in the mobile network was defined by PromPerú and divided into sectors by Telefonica based on call records, text messages and other ways of data usage. This traffic is monitored during a specific period of time that allows to identify if the people in the analysis are tourists or residents.

The indicators that were taken into consideration to determine the study area of data position were the following:

- According to the domestic traveler profile (PromPerú, 2017), 84% of domestic trips were trips for vacations and had Lima Metropolitana as the departure city. Therefore, Lima Metropolitana is rank as the most important city of origin for domestic tourism.

- According to "Socioeconomic Levels Report" (*Asociación Peruana de Empresas de Investigación de Mercados*, 2017) 42.2% of the population of Lima belongs to the low (C) socioeconomic level.

Based on this information it was decided to focus this research on observing the traffic of mobile positioning data of the residents from the most populated districts belonging to low (C) socioeconomic level, i.e Chorrillos, Independencia, San Juan de Lurigancho and San Martin de Porres districts. This data was acquired by *Smart Steps* methodology.

Study universe: call records, text message and other ways of data usage (events) from users of Telefonica´s mobile network, from 20 years or older, that traveled to destinations outside their usual environment, from June 1st to June 30th, 2016 and from June 1st to June 30th, 2017.

For the research, Telefonica reported an average of 20 thousand events (call records, text messages and other ways of data usage) per day. (See annex 5: Lima Metropolitana: districts that make up the range of geographic origin and that were considered for this research).

Length of stay: from 1 to 6 nights.

Geographic spectrum of the place of origin:
Chorrillos, Independencia, San Juan de Lurigancho (most populated district of Peru) and San Martin de Porres.

Geographic spectrum of the destination:

- **Level 1:** trips to provinces of Lima region (Lima, Barranca, Cajatambo, Canta, Cañete, Huaral, Huarochirí, Huaura, Oyón and Yauyos).
- **Level 2:** trips to other regions of Peru (it does not include Lima region neither El Callao province).

Results or findings

The main findings obtained with mobile position data about domestic transfer are listed below:

Trips outside Lima regions (interregional transfers): on June, 2017; there were 223,1 thousand trips registered from Lima to other regions within the country; 44,6 thousand more trips than on June, 2016 (+25%). Most of the trips had San Juan de Lurigancho district as point of departure (See annex 6: Number of trips made by residents from Chorrillos, Independencia, San Juan de Lurigancho and San Martin de Porres districts to destination outside Lima region).

In general, it was noticed that residents of Lima who traveled the most belong to the low (C) socioeconomic level (+18,4 thousand more trips than on June, 2016). Furthermore, the number of trips from residents belonging to the lowest (DE) socioeconomic level, is about five thousand for both periods of analysis. This is an important figure that was not noticed before this research (See annex 7: Distribution of trips by socioeconomic level and age range made by residents from Chorrillos, Independencia, San Juan de Lurigancho and San Martin de Porres districts to destination outside Lima region).

During the *Vamos Pal Norte* campaign residents from San Juan de Lurigancho traveled mostly to Junín (21,6 thousand trips). On the other hand, trips to northern destinations (that included Amazonas, Ancash, La Libertad, Lambayeque, Loreto, Piura, San Martín and Tumbes regions) that were promoted by *Vamos Pal Norte* campaign, were increased in 76% in contrast to trips that took place during June, 2016 from the same district. (See annex 8: Number of trips made by residents of San Juan de Lurigancho district to destination outside Lima region).

Mobilizations within the Lima region (intraregional transfers)

In June, 2017 54,4 thousand trips were registered from Lima Metropolitana to different provinces of Lima region; i.e. 21,6 thousand more trips than on June, 2016. Those trips were mostly from residents of San Juan de Lurigancho (See annex 9: Number of trips made by residents from Chorrillos, Independencia, San Juan de Lurigancho and San Martin de Porres districts to provinces witin Lima region).

It should be noticed that when residents of San Juan de Lurigancho made short trips within the Lima region, they traveled mostly to Cañete province (5,8 thousand trips) and Huarochiri province (5,2 thousand trips). (See annex 11: Mobilizations of residents of the district of San Juan de Lurigancho to Lima´s provinces).

In general, it was observed that residents who traveled the most belong to the low (C) socioeconomic level (+10,8 thousand more trips than on June, 2016). Although, the number of trips by residents belonging to the lowest (DE) socioeconomic levels is at least one thousand, more than 50% of those trips were made by young people from 20 to 29 years old. (See annex 10: Distribution of trips by socioeconomic level and age range made by residents from Chorrillos, Independencia, San Juan de Lurigancho and San Martin de Porres districts to Lima provinces).

Conclusions

The mobile positioning data were very useful to measure trips made by residents from Chorrillos, San Juan de Lurigancho, San Martin de Porres and Independencia districts. It was found that most intraregional trips departed from San Juan de Lurigancho district and travelers went mostly to Junín region, and residents from the city of Lima who

traveled the most belonged to the low (C) socioeconomic level.

Furthermore, the use of large volume of data or Big Data has allowed to measure, for the first time, the movement of population belonging to the lowest (DE) socioeconomic level; segments of the population that were not taken into consideration by traditional market research because of costs and accessibility limitations.

On the other hand, the use of mobile positioning data and its follow-up analysis have allowed to generate knowledge about the scope of trips made by residents of Lima regarding interregional and intraregional territory, being an alternative tool to assess the effectiveness of specific campaigns of PromPerú. In the case of *Vamos Pal Norte* campaign, which was evaluated by this research, it was found that residents from San Juan de Lurigancho district (lowest (DE) socioeconomic level) increased their travels by 76%, in contrast to the number of travels registered during June, 2016, from the same district.

However, further research should be carry out in other socioeconomic levels such as High (A) and Medium (B), to contrast if this alternative source allows to reach them, and the results could be contrasted with official indicators of domestic trips provided by the *"Encuesta Nacional de Viajes de los Residentes*" – ENVIR, MINCETUR (from its initials in Spanish).

In this sense, it is necessary to do further research based on mobile positioning data to measure the effects not only of specific campaigns, but also of other cases. In further researches it could established if the mobile positioning data source serves also to evaluate population from other socioeconomic strata High/Medium (A and B) or if it is useful just for strata Low/Lowest (C, D and E).

Additionally, it could be considered if the analysis of this type of information is feasible for all the Peruvian regions (urban areas) taking into consideration the antenna locations and access to the mobile broadband service throughout Peru (3G or 4G). Finally, another aspect that should be cover for future researches is the validation of temporal and spatial representativeness for these type of studies.

ANNEX 1
Structure of information sources

Tabla CRM

Nombre	Descripción	Tipo de variable
SuscripcionID	Identificador del teléfono	cadena
ClienteCD	Identificador del cliente	cadena
Edad	Edad del cliente	cadena
Segmento	Segmento	cadena
LineaNegocio	Line de negocio	cadena
Cod_Seg_postpago	Indica si es pospago	cadena
Cod_Seg_prepago	Indica si es prepago	cadena
Ingreso_pre	Ingresos prepago	real
Ingreso	Ingresos	real

Tabla de CDR's

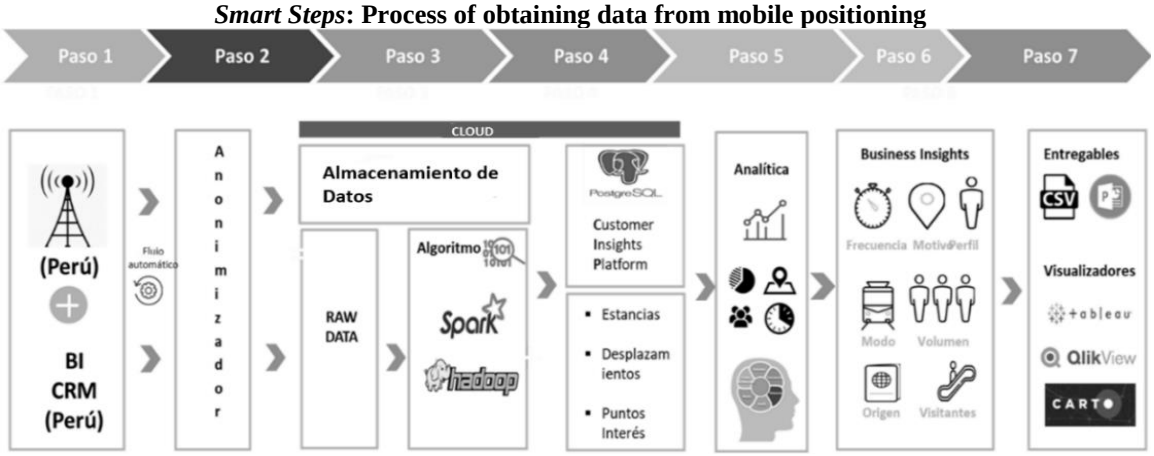
Nombre	Descripción	Tipo de variable
SuscripcionID	Identificador del telefono	cadena
Fecha	Fecha	date
Hora	Hora	hh:mm:ss
Tecnología	Red	cadena
CeldaCD	Identificador de la celda	cadena
LAC	Identificador LAC	cadena
TAC	Identificador TAC	numérica
Sentido	Evento entrante o saliente	cadena
Tipo_Evento	Voz ó SMS	cadena

Tabla Catálogo de Celdas

Nombre	Descripción	Tipo de variable
Id_Cell	Identificador de la celda	cadena
Id_Sector	Identificador del sector	date
Latitud	Latitud del mástil	Real
Longitud	Longitud del mástil	Real
Azimud	Azimut de la celda	cadena
LAC	Identificador LAC	cadena
Banda Frecuencia	Frecuencia	numérica
Altura	Altura de la celda	
Tecnologia	Evento entrante o saliente	cadena

Source: Telefónica (2017).

ANNEX 2
Smart Steps: Process of obtaining data from mobile positioning



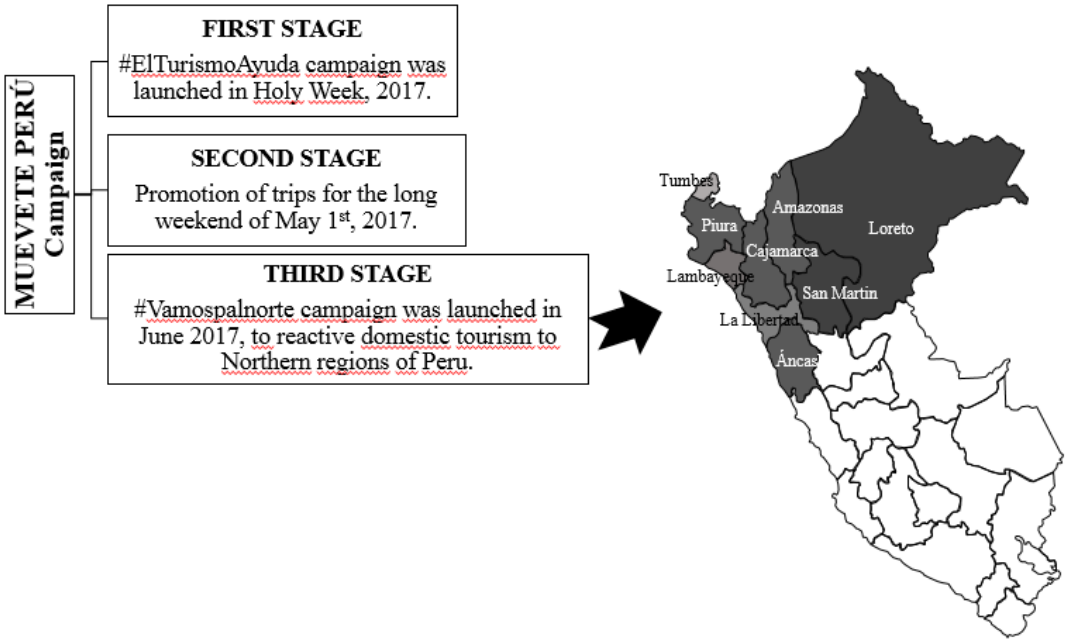
Base Transceiver station | Source: Telefonica (2018).

ANNEX 3
Mobile Service Indicators 2017

Companies	Lines in Service by company
Telefónica del Perú S.A.A.	14.865.510
América Móvil Perú S.A.C.	12.506.551
Entel Perú S.A.	6.371.929
Viettel Perú S.A.C.	5.084.880
Incacel Movil S.A.	86.516
Total mobile lines (Peru)	**38.915.386**
National Teledensity	127.2 lines

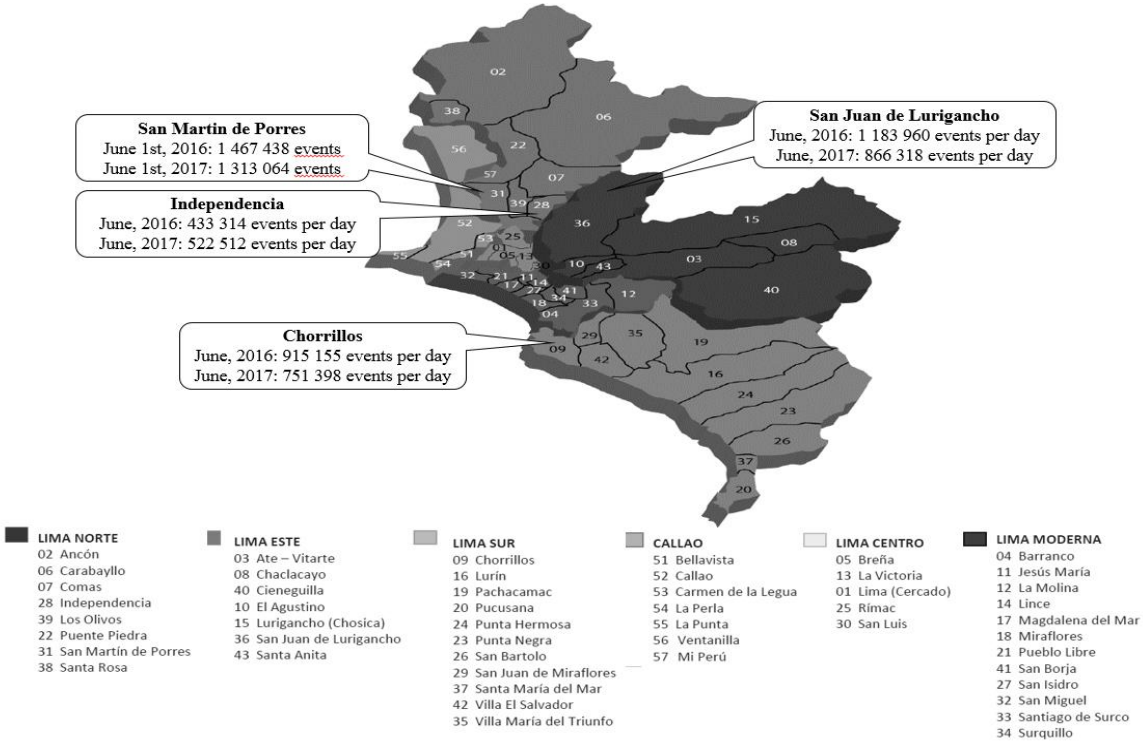
Source: Osiptel / National teledensity: Mobile lines per 100 inhabitants.

ANNEX 4
Stages of the campaign *Muévete Perú*



Source PromPerú (2017).

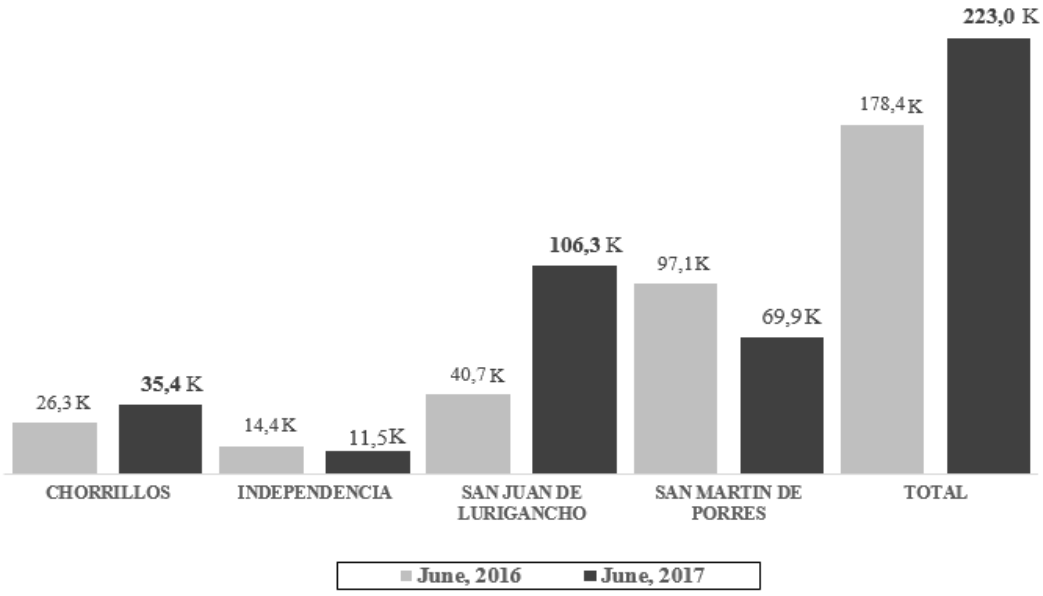
ANNEX 5
Lima Metropolitana: districts that make up the range of geographic origin and that were considered for this research



Source PromPerú (2017)

ANNEX 6

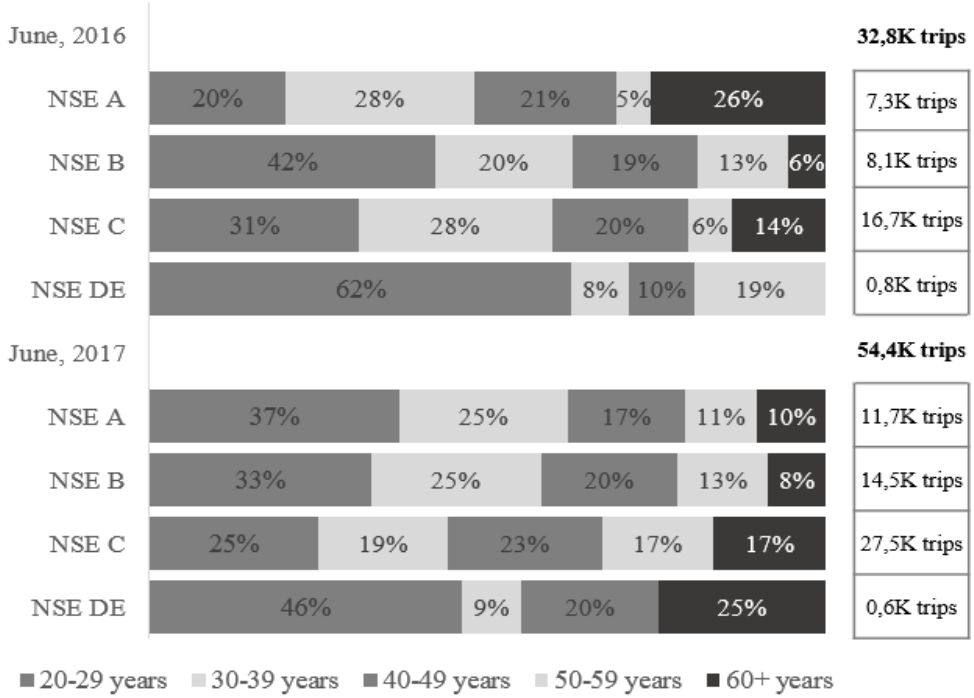
Number of trips made by residents from Chorrillos, Independencia, San Juan de Lurigancho and San Martin de Porres districts to destination outside Lima region (Constitutional Province of Callao is not included) (in thousands = K)



Source PromPerú (2017)

ANNEX 7

Distribution of trips by socioeconomic level and age range made by residents from Chorrillos, Independencia, San Juan de Lurigancho and San Martin de Porres districts to destination outside Lima region (Constitutional Province of Callao is not included) (in thousands = K)



Note: High Socioeconomic Level (NSE A), Medium Socioeconomic Level (NSE B), Low Socioeconomic Level (NSE C) and Lowest Socioeconomic Level (NSE DE).

Source PromPerú (2017)

ANNEX 8

Number of trips made by residents of San Juan de Lurigancho district to destination outside Lima region (Constitutional Province of Callao is not included) (in thousands = K)

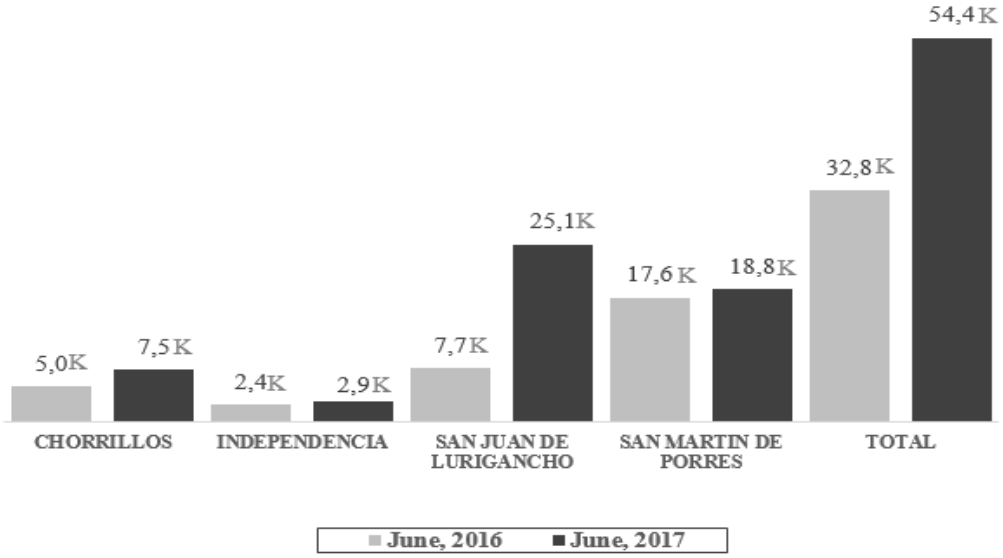
Regions	June 2016	June 2017	Variation	Ranking 2016	Ranking 2017
Junín	4,13 K	21,61 K	424%	3°	1°
Ancash (*)	**6,23 K**	**10,73 K**	**72%**	**1°**	**2°**
La Libertad (*)	**3,98 K**	**8,03 K**	**102%**	**4°**	**3°**
Lambayeque (*)	**4,35 K**	**7,72 K**	**77%**	**2°**	**4°**
Ica	2,25 K	7,42 K	230%	7°	5°
Ayacucho	1,35 K	6,46 K	378%	9°	6°
Huánuco	2,25 K	6,33 K	181%	8°	7°
Cajamarca	**3,15 K**	**5,85 K**	**86%**	**5°**	**8°**
Pasco	0,6 K	4,33 K	621%	18°	9°
Cusco	1,2 K	4,02 K	235%	10°	10°
Piura (*)	**3,15 K**	3,91 K	**24%**	**6°**	**11°**
Arequipa	1,2 K	3,58 K	199%	11°	12°
Huancavelica	0,38 K	3,44 K	817%	20°	13°
San Martin (*)	**0,9 K**	**2,43 K**	**170%**	**15°**	**14°**
Apurímac	1,05 K	2,17 K	107%	12°	15°
Ucayali	0,53 K	2,09 K	299%	19°	16°
Puno	1,05 K	1,58 K	50%	13°	17°
Amazonas (*)	**0,75 K**	**1,21 K**	**61%**	**16°**	**18°**
Loreto (*)	**0,98 K**	**1,05 K**	**7%**	**14°**	**19°**
Tacna	0,75 K	0,95 K	27%	17°	20°
Tumbes (*)	**0,15 K**	**0,73 K**	**383%**	**21°**	**21°**
Moquegua	0,15 K	0,54 K	257%	22°	22°
Madre de Dios	0,15 K	0,10 K	-34%	23°	23°
Total	40,7 K	106,3 K			

(*) Northern regions	**23625**	**41651**	**76%**

Note: Destinations promoted by the *Vamos Pal Norte* campaign.
Source PromPerú (2017)

ANNEX 9

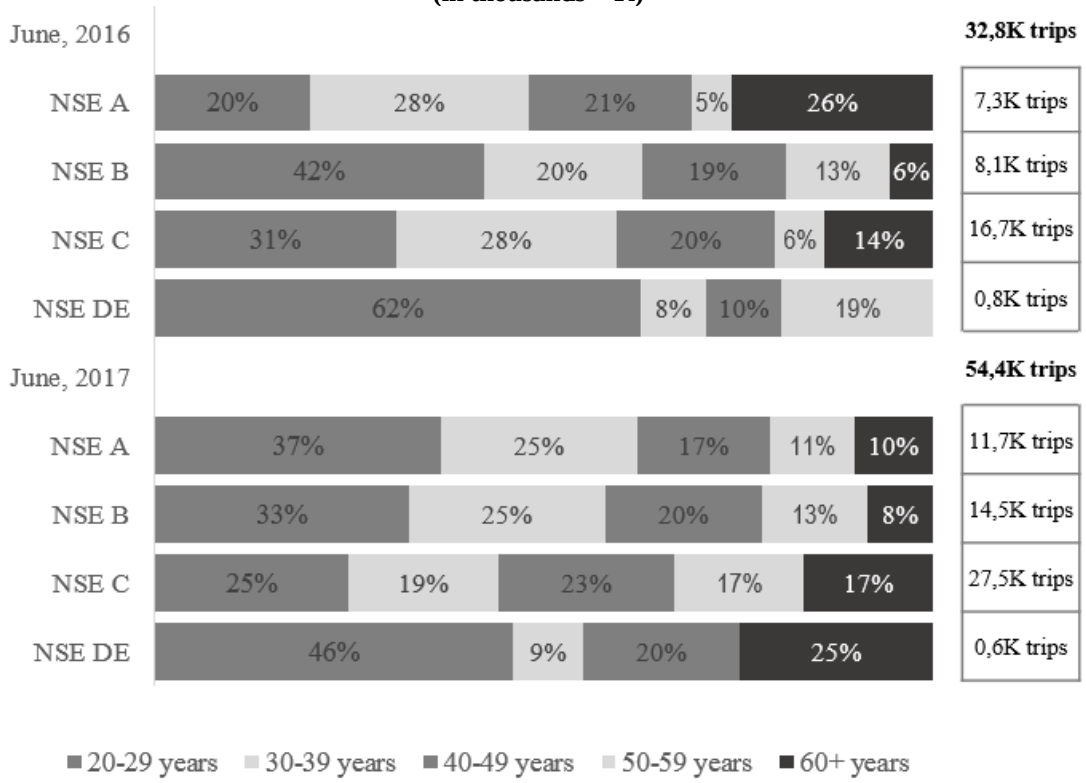
Number of trips made by residents from Chorrillos, Independencia, San Juan de Lurigancho and San Martin de Porres districts to provinces witin Lima region (Constitutional Province of Callao is not included)(in thousands = K)



Source PromPerú (2017)

ANNEX 10

Distribution of trips by socioeconomic level and age range made by residents from Chorrillos, Independencia, San Juan de Lurigancho and San Martin de Porres districts to Lima provinces (Constitutional Province of Callao is not included) (in thousands = K)

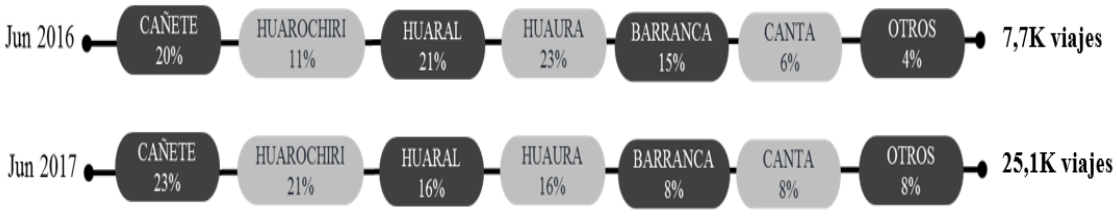


Note: High Socioeconomic Level (NSE A), Medium Socioeconomic Level (NSE B), Low Socioeconomic Level (NSE C) and Lowest Socioeconomic Level (NSE DE).

Source PromPerú (2017)

ANNEX 11

Mobilizations of residents of the district of San Juan de Lurigancho to Lima´s provinces (Constitutional Province of Callao is not included).



Source PromPerú (2017)

Glossary

CDR (Call Detailed Record) Data record that details all de information about active events (user-event data) generated by the mobile networks.

CRM: Card index that details all the information that identifies a client.

Cells catalog: Card index that details the characteristics of cells that configure the network.

Cloud is a term used to describe network service which can be accessed through any internet connection. The term cloud usually refers to data storage outside physical devices.

Spark is a programming language specially designed for high integrity systems.

Hadoop is an open source software to store, process and analyze large amounts of data. It supports hundreds of terabytes, petabytes or bigger measures.

Python is a programming language which philosophy emphasizes a readable code.

PostgreSQL is an open source relational database system, since it includes object orientated characteristics such as inheritance mechanism, types of data, functions, restrictions, triggers, rules and transactional integrity. It is free and it is distributed under the licensed BSD.

Tableau is an intelligent software to analyze data with an excellent visualization and presentation dashboards. It is considered as one of the best software for visual presentations of data by many, and it has high qualifications because of its ease of use since it is very suitable to Microsoft Excel.

References

Ahas, R., Aasa, A., & Silm, S. (2007). Mobile positioning data in tourism studies and monitoring: case study in Tartu, Estonia. Information and communication technologies in tourism, 119-128.

Asociación peruana de empresas de investigación de mercados, (2017) Niveles Socioeconómicos. Retrieved from http://www.apeim.com.pe/wp-content/themes/apeim/docs/nse/APEIM-NSE-2017.pdf

Beresewicz, M., Lehtonen, R., Reis, F., Di Consiglio, L., & Karlberg, M. (2018). An overview of methods for treating selectivity in big data sourcing. Statistical Office of the European Union. Luxembourg: European Union. Retrieved from http://www.sipotra.it/wp-content/uploads/2018/07/An-overview-of-methods-for-treating-selectivity-in-big-data-sources.pdf

Esko, S. (2018). Mobile phone data use for tourism statistics in Indonesia and Estonia. Madrid.

Instituto Nacional de Estadística e Informática. (2016). Encuesta Nacional de Hogares. Retrieved from https://webinei.inei.gob.pe/anda_inei/index.php/catalog/543/datafile/F2

Ipsos Group S.A. (2017). Hábitos, usos y actitudes hacia la telefonía móvil.

Kasperby Lab Daily. (10 de Abril de 2014). Retrieved from https://kaspersky.com/que-es-un-hash-y-como-funciona/2806/

Organismo Supervisor de Inversión Privada en Telecomunicaciones. (2018). Líneas en servicio por empresa al 30 de enero de 2018. Obtenido de https://www.osiptel.gob.pe/articulo/24-lineas-en-servicio-por-empresa

PromPerú (2017). Perfil del Vacacionista Nacional. Retrieved from https://promperu.gob.pe/TurismoIN/sitio/PerfVacacionistaNac.

Ramos Lobo, J., & Pedregal Mateos, B. (2015). Big data y turismo cinematográfico ¿un factor de predicción? El caso de juego de tronos de Osuna. VIII Jornadas de investigación en turismo: Impulso al desarrollo económico a través del turismo, (págs. 233-256). Sevilla.

Telefonica Data Unit. (2018). LUCA Tourism.

Toscón, M. (2013). Introducción: Big Data. Pasado, presente y futuro. (F. Telefónica, Ed.) Telos. Cuadernos de comunicación e innovación, 47-50.

Zamora, J. (2016). Tomar mejores decisiones con el Big Data. Harvard Deusto Business Review, 6-14.