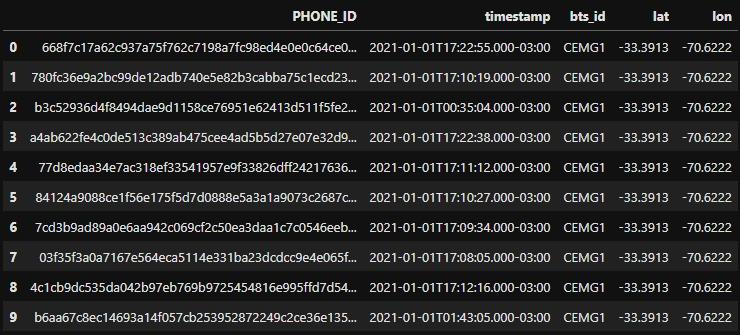
**Reto: Privacy and Security**

**¿Is the data anonymous?**

*The data provided by the Socio Formador looks like the image from above, the additional data information and the meanings and some details about the information can be found in the documentation (dataset description). In the previous image it is possible to distinguish different columns, the one of greatest importance in the anonymization process is the phone\_id, which was already encrypted by the training partner's computer security methods, this is of great importance because the organization already contains a series of processes security, that is, it has its own defined guidelines and protocols.*

*The fact of having information about a cell phone tends to be sensitive information, for which, although it is not possible to encrypt the other columns about the position of the connection, the time, or the connection antenna. On the other hand, the id\_phone can be encrypted, since the identifier does not provide information other than the unique identifier itself, so in this case, the training partner carried out a series of algorithms to "mask" any inference about the id\_phone information. On the other hand, the only point that is no longer possible to mask or hide from the dataset is the fact that the identifier can be repeated, as a movement through time, the reason for this is in accordance with the main point of the dataset, which is to generate traceability in the movement of mobile phone users during the indicated day.*

*Something that is also important to add is the terms in which the dataset was provided, since the date of the data is precisely anonymized, the reason for this is that, by having an indicated time relationship, it is possible to filter very sensitive information when related to another database based on time, an example would be to generate the traceability of a person and converge information from their environment that happened during their journeys such as accidents, crimes, purchases, use of services or other sensitive information and likely to be obtained in internet, information that can reach the point of identifying the person. Therefore, the dataset does not belong to the day indicated in the timestamp column, although the hour does have a point of meaning to get closer to reality, in this way we can obtain behavior patterns and analysis of the variables that we intend to generate prediction results. Therefore, the masked variables were:*

* *phone identifier*
* *connections date*

**Normatives to follow in the project**

***Normatives in México about data security and privacy***

*In Mexico there are certain statutes and characteristics that must be given to the data so that they do not represent a risk to one or more individuals, in the same way, certain methods are used to mask the information, with this, the data will never have a way of relating to other databases.*

*Ley Federal de Protección de Datos Personales en Posesión de los Particulares*

*Since July 5, 2010, a law was established in the country that aims to generate a transformation in personal data so that they are regulated for their handling by private organizations. This law enters into jurisdiction when the company can have the following characteristics:*

* *The person responsible for management is located in Mexico.*
* *When the information is processed by software that is processed in a database located in Mexico.*

*Under the Data Protection Law there are two individuals who stand out as figures in the processing of personal data:*

* *While the person responsible decides on the modification or any other set of decision-making on the treatment of the data.*
* *The person in charge is the person (other than the person responsible) who is dedicated to processing personal data following the orders of the person in charge.*

*Something important to mention is that, in order for a person in charge to be involved, there must be a contract between both parties so that each and every one of the conditions and obligations of both parties are taken into account, also that there is that degree of confidentiality.*

*Basic principles in the protection of personal data in Mexico*

1. *Legality: All data processing must be in accordance with each of the laws and regulations stipulated in the Data Protection Law, the Regulation and any other regulation issued in the country's legislation.*

1. *Consent: There must be the approval of the owners for the use of their data at any time that is necessary, the owner must also agree to transfer them.*
2. *Information: The owner must be informed of the information related to the processing of their data and obtain the approval of the owners for the use of their personal data. The information may be related to: what data will be collected and the purposes of the treatment. These details should go in the privacy notice.*

1. *Quality: In the use of personal data, these must be complete without discarding, exact at the time they were obtained, correct and updated in accordance with the purposes for which they were collected.*

1. *Purpose: The processing of personal data must coincide with the purposes established in the privacy notice.*

1. *Loyalty: Always treat personal data in such a way that the protection of the interests of the owner are in its entirety, a priority,*

1. *Proportionality: Only the data that is necessary and in accordance with the purposes of what is established in the privacy notice may be used.*

1. *Responsibility: Always answer for all the changes made to the personal data collected and everything that will influence future decisions.*

*Obligations for private organizations*

*The Data Protection Law has established a series of obligations and responsibilities that must be fulfilled by private companies that deal with this data. Among them are:*

1. *Have a privacy notice: This mentioned document is intended to demonstrate everything that encompasses the requirement of the Data Protection Law, where the owner is informed of all the details and procedures to be carried out with their personal data. A privacy notice must contain the clear identity of the person in charge, the data to be used, the purpose of the procedure, the rights of the owner and if there are other organizations with which the information will be shared.*

1. *Have the consent of the owner: There are two types of consent by the owner, while the consent is express when the owner contains the initiative to manifest by any means of unequivocal signs. When it comes to financial or mostly sensitive personal data, it is necessary to obtain the express consent of the owners. Consent is tacit when, by having the privacy notice visible, the owner does not express any opposition to the processing of data.*

1. *Allow access, rectification, cancellation and opposition of the holder: It consists of allowing the holders to exercise their rights. Those can be:*

***Access***

*Every owner has the right to know the details about the handling of their data.*

***Rectification***

*The holders can request the correction or modification of their data, in the situation that affects their morality or is inappropriate.*

***Cancellation***

*When the data processing procedure is not as expected, the owner may request the deletion of their data at any time.*

***Opposition to data processing***

*Users may object to the processing of their personal information.*

*Lastly, data holders must respond to requests for revocation of consent or limitation of use. The person in charge must always designate a person or a personal data department to deal with requests for ARCO rights.*

***Main aspects to consider in data protection in Chile***

* *The processing of personal data will be subject to the statutes of the Chilean National Personal Data Protection Regime for both public and private organizations.*
* *Any person can process personal data if they do so in accordance with current law and for all purposes that are permitted in a legal framework and that respect the rights of the owners.*
* *The collection of personal data must be reported to the owners of the data in a mandatory manner and the purpose of that collection. This process should never disobey the rights and obligations established by law. The results in the research motives should never give indications where the person surveyed can be identified. Owners can always object to the use of their personal data.*
* *The processing of personal data can only be done when the legal provisions allow it or when the owner expresses consent to its use. The owner must know what process their data will have and what is intended to be communicated to the public. Even if the authorization is in writing, it can always be revoked.*
* *It is not necessary to have authorization in the processing of personal data that are in sources accessible to the public, when they are of an economic, financial, banking or commercial nature, that contain a list of information that is limited to only showing information without identification. of the holders or that are necessary for a direct response communication or of an urgent nature.*
* *Personal data must be deleted when the storage of the same lacks importance, legal basis or when its time of use has expired. Personal data may be blocked when its accuracy cannot be determined, erroneous, inaccurate or incomplete data may also be modified.*
* *When data processing is through a mandatory process, it will be necessary to keep each and every one of the conditions of use in writing, leaving both parties informed to guarantee compliance.*
* *Unless the data is publicly accessible, it must always be used for the reason in which it was collected, the information of these data must always be updated and its veracity must be guaranteed as long as the identity of the owner is known.*
* *It cannot be processed, sensitive data can only be done when the law authorizes it, a consent process is carried out with the owner or the benefits are for the same.*

**Process for working on the dataset.**

*In order to guarantee that third parties obtain the provided dataset, our team has it stored in a shared folder in Google Drive, to which only the following people have access:*

***Team members:***

* *Carlos Alberto Hurtado Sanchez*
* *Kevin Joan Delgado Perez*
* *Weng Hu*
* *Gerardo Pena Perez*

***Teachers:***

* *Benjamin Valdes Aguirre*
* *Eduardo Daniel Juarez Pineda*
* *Ishmael Solís Moreno*

*Along with the initial dataset, we also store auxiliary datasets and files. Because these files were generated from the original set, they also contain sensitive information, so it is necessary that only those involved in the project have access to them. The associations between data that we are making do not put the anonymity of the data at risk, since we have not mixed with personal unique identifiers. We have limited ourselves to relating the antennas with the name of the commune in which they are located.*

**Security tools used in the project**

*Initially there was the dataset provided by the training partner, but taking into account the amount of risk of putting it in contact with the Internet network, it is necessary to mask some data, although this data is already anonymized as explained at the top of this document. The first decision of the team was to save the dataset locally, only for the established use with the objectives already proposed, later a google drive repository was created with the purpose of saving all the documentation and organization of the work team there. It was initially configured so that only accounts from the Tec de Monterrey organization could view it but searching specifically, later it was configured so that only the work team and professors had access to that folder. Since only documentation files were generated in this folder, your risk is minimal so the settings were correct.*

*At the same time as the documentation, the generation of datasets such as the antenna datasets by commune were obtained from the Geopandas library, but it was carried out as an abstraction towards a JSON type file that worked as an intermediate method between obtaining data through the framework and data storage to a CSV table file, this method requires a key which was delivered to the team members to access that data abstraction, if they did not have it, there was no way to obtain the information to process, the advantage of all this is that the generation of datasets can be accessed through security methods and that, on the other hand, they work as a code optimization since they are easier to process than a file by tables, this whole procedure speeds up data cleaning, organization and processing to generate more information. Finally another advantage would be the easy backtracking to repeat a phased processing, this is important for iteration of cleanup procedures.*

***References:***

***Presidencia, de la. (2020, 26 agosto). Ley 19628 (28-ago-1999) M. Secretaría General de la Presidencia | Ley Chile. Biblioteca del Congreso Nacional de Chile. www.bcn.cl/leychile.*** [***https://www.bcn.cl/leychile/navegar?idNorma=141599***](https://www.bcn.cl/leychile/navegar?idNorma=141599)

***Viollier, P. (2017). El estado de la protección de datos personales en Chile. Derechos Digitales América Latina.*** [***https://www.derechosdigitales.org/wp-content/uploads/PVB-datos-int.pdf***](https://www.derechosdigitales.org/wp-content/uploads/PVB-datos-int.pdf)

***DocuSign, C. de. (2021, 20 octubre). Protección de datos personales en México: conoce el marco legal. DocuSign.*** [***https://www.docusign.mx/blog/proteccion-de-datos***](https://www.docusign.mx/blog/proteccion-de-datos)

***LEY FEDERAL DE PROTECCIÓN DE DATOS PERSONALES EN POSESIÓN DE LOS PARTICULARES. (2010). CÁMARA DE DIPUTADOS DEL H. CONGRESO DE LA UNIÓN.*** [***https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf***](https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf)

***(2021, 13 octubre). Ley Federal de Protección de Datos: ¿qué es y cuándo es aplicable? DocuSign.*** [***https://www.docusign.mx/blog/proteccion-de-datos-personales***](https://www.docusign.mx/blog/proteccion-de-datos-personales)