Jose Luis Ribero Duque
Data Engineer

Education
- Universidad de Antioquia: Phisicist
- University of San Diego California: Big Data Specialist

(E, T, L) Project

Description of the Problem:

For the local 2018 government elections a political team need create a data base with all the possible data from the electors of a particular town. We will use this database to send a political message from the political team. The Data was recollected from divers sources between these handmade data bases all with different structures.

Project Objective: To send a message via Whats App

Objetives:

- To obtain all the possible population information taking divers sources and databases
- Filter, structure and preparing the databases, cleaning and fixing the errors
- Unified all the databases dropping the multiplied data
- Filter the unified database dropping the contacts with no ACTIVE WhatsApp account
- To Send a specific message via WhatsApp Web to all the conctacs in the filtered unified database

Development:

Extract

The first phase of project was recollect all the possible information from different sources like web pages, local physical databases and information taken from people close to the political team. Exploring the information obtained we noticed that almost all the information is with no format, disordered with errors and empty fields. We digitalized the physical data and we transformed it to .csv format.

Transformation

In this point we have three data bases with no format, disordered with empty fields.
- db_cocorna_enlaces.csv
- db_cocorna_people.csv
- db_cocorna_consolidate.csv

We use *filtering.py* for filter, clean and organize each database with a specified format with no empty fields. Once this first step was made we obtained three new databases
- Cocorna_phone_numbers_people.csv
- Cocorna_phone_numbers_enlaces.csv
- Cocorna_phone_numbers_consolidated.csv

After filtering and cleaning we used *data_integration.py* for integrate the three databases in a database with no duplicated records and well formted.
- Cocorna_phone_numbers_integrated.csv

Load

After integration we load the database to WhatsApp Web application for filter the the phone numbers with an ACTIVE account. This step produced a new database called
- WhatsApp_contacts.csv

This last database will be used to send the messages to the voters.

We used a code made in Python that load the database, open a Google Chrome window search for WhatsApp Web, takes one by one the contacts in the database and send the message to them.