Agenda

Business Case

Bike-sharing, it's a program in which many bicycle stations are set up through a city and people can rent a bike to use for a certain time frame and return it at a different station.

An IESE study examines the performance of 13 bike-sharing initiatives in Europe shows that for every €1 spent, cities see €1.72 in economic and health benefits. These numbers were arrived at by evaluating both the direct economic impact of the bicycle programs, in terms of job creation, revenues and knock-on economic effects on local businesses, and in terms of the health benefits derived from increasing physical activity and shifting traffic patterns.

In the other side of the world, in Minnesota, a study -mentioned by NextBike- reported an additional $1.20 a week spent by customers after a bike share station was installed nearby. Annually, this could mean around $62.40 extra spent per customer. Furthermore, 70% of businesses in Washington D.C. identified a positive impact of bike sharing on the neighbourhood, with 20% claiming there was direct positive impact on revenues.

According to MetroBike, a bikeshare consultancy company in North America with 15+ years of experience, as of May 2018 more than 1.600 bike-sharing programs were in operation worldwide, providing more than 18 million bicycles for public use… To mention some of the top 10 bike-sharing programs in the world:

- CitiBike: It has around 14.000 bikes and 900 stations across NYC.
- Bixi: It has more than 6.000 bikes and 610 stations spread out across Montreal.
- Bike Share Toronto: Offers 24/7 access to 6.850 bikes and 625 stations across 200 km2 of the city.

Data Collection

The data sources that are part of the solution developed are listed below:

- Bixi Montreal trip history and stations
- Bike share Toronto ridership data
- Bike share Toronto station information
- Montral Road collisions
- Toronto Police Service (collisions)

Data Management

Extraction

- To get the bike-sharing data of Montreal (Bixi site), as no API was found, the technique of web scrapping was used in order to retrieve all the hyperlinks related to the trip history data available since 2014. The downloaded data also needed to be extracted and gathered under the same folder as there were several csv files inside .zip files, so *unziping* them created different sudfolders.
- To get the bike-sharing and collisions data of Toronto, the API provided by the city's open data site was consumed, this required a bit of a workaround as its format wasn't straightforward and some objects were nested within other objects.

Transformation

The data retrieved whether by scrapping or API was collected into pandas dataframes, so any basic transformation as formatting a column or adding another one to get the weekday was done at that level while the data was store in a dataframe.

Another type of transformations were done at the database level, to compare performance and to test the development, loading the data without any transformation at the dataframe level (prior inserting into the database) took ~5min to download and import all 63 files related to Montreal and Toronto, but doing it with the transformations consumed 4 times more processing.

Load

To load the data from a dataframe into a table in the database, the to_sql function was used.

Data Visualization

De la data almacenada se puede obtener las siguientes métricas:

- Distribución de los viajes por tipo de usuario.
- Estaciones más usadas.
- Rutas más usadas.
- Día de la semana con más demanda.
- Distribución de la demanda del servicio durante el tiempo.
- Estaciones cercanas a donde históricamente se han registrados más accidentes.