

DATA Analytics

Gasolineras de México

Made by:

Carlos Casio Pérez

Eduardo Ayala Saldaña

Paulina Islas Huajuca

Rodrigo Guarneros Gutiérrez

Samuel Cerón Martínez

Introduction

Historically mexican population has been sensitive to fuel prices to the extent that it has even been a driver of political campaigns and discourse.

Also, in recent times, customer service has become paramount in diverse areas related to the purchase of goods and services,

Finally, ourselves as fuel consumers, with the liberalization of prices from the energy reform, we wanted to see if added features influence consumption and ratings in Service Stations

We wanted to see if the data reflects our personal observations.

Available Data and Database

We wanted to have a more humane analysis to see if there are any factors that may impact the price or the ratings with the following data.

Georeferenced Service Stations Listings:

This included a record of Service Stations (SS) in XML format, with data such as the SS unique id, business name, address, and geographic coordinates.

Prices and Fuel types:

We got the listing of fuel prices by type (regular, premium, diesel) and by SS, with data such as SS unique id linked to the Georeferenced Service Stations Listings.

After merging and cleaning up the mentioned databases, we aimed to compare between the government databases based on individual reports by the Service Stations with the Google user appreciation. Taking advantage of the information that **Google API** provides. Such as address, type of SS, amount of raters, ratings and additional features.

Project Hypothesis

- Understand if there is any relation between the price and the rating.
- Understand if there is any relation between the price and the geographic location.
- Understand if there is any relation between the ratings and the geographic location.

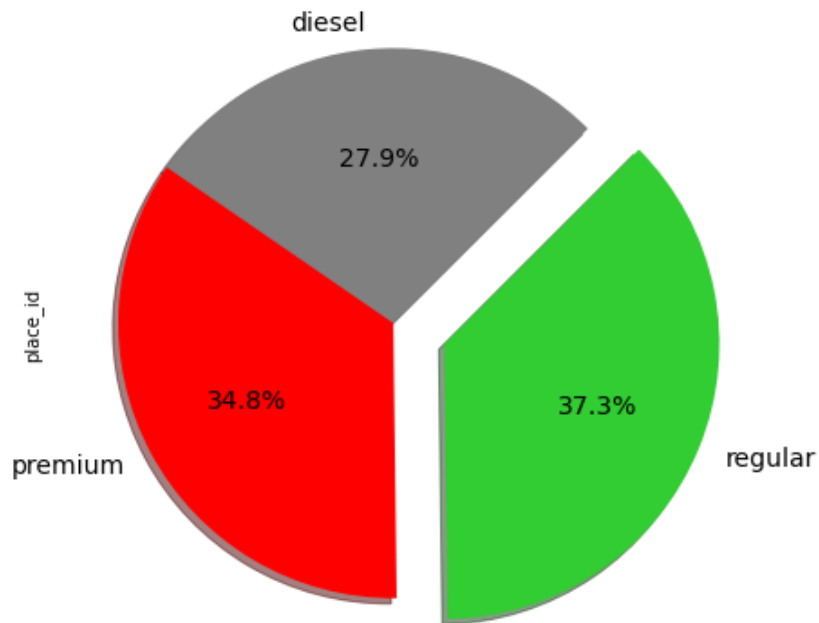
Data Cleanup

For this exercise we performed the following actions:

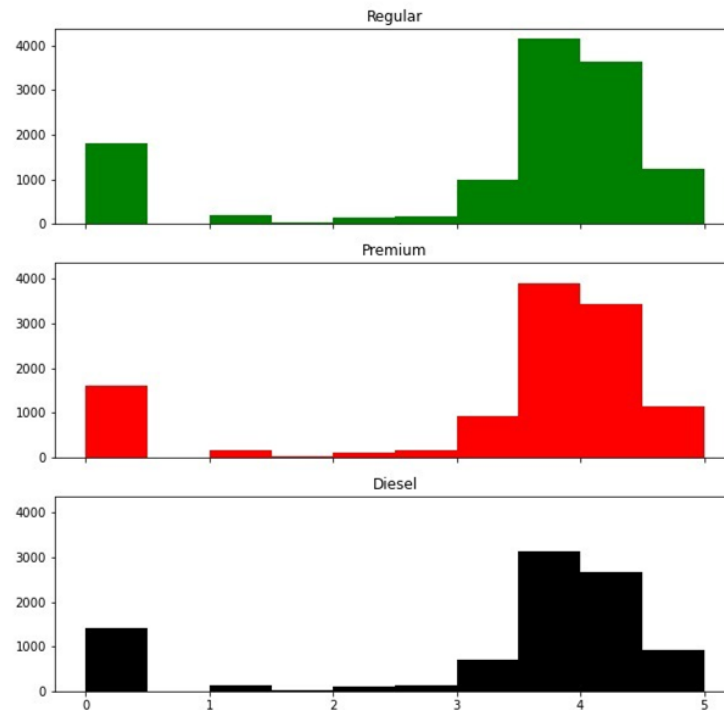
- Understand the data's dimension.
- Merge the two databases.
- Drop missing values and duplicates.
- Classify information by fuel type.
- Clean up incorrect coordinates.
- Execute API calls to get features and ratings by Service Station.
- Concatenate the APIs results we got after using each team member's API keys.

Fuel distribution and Ratings

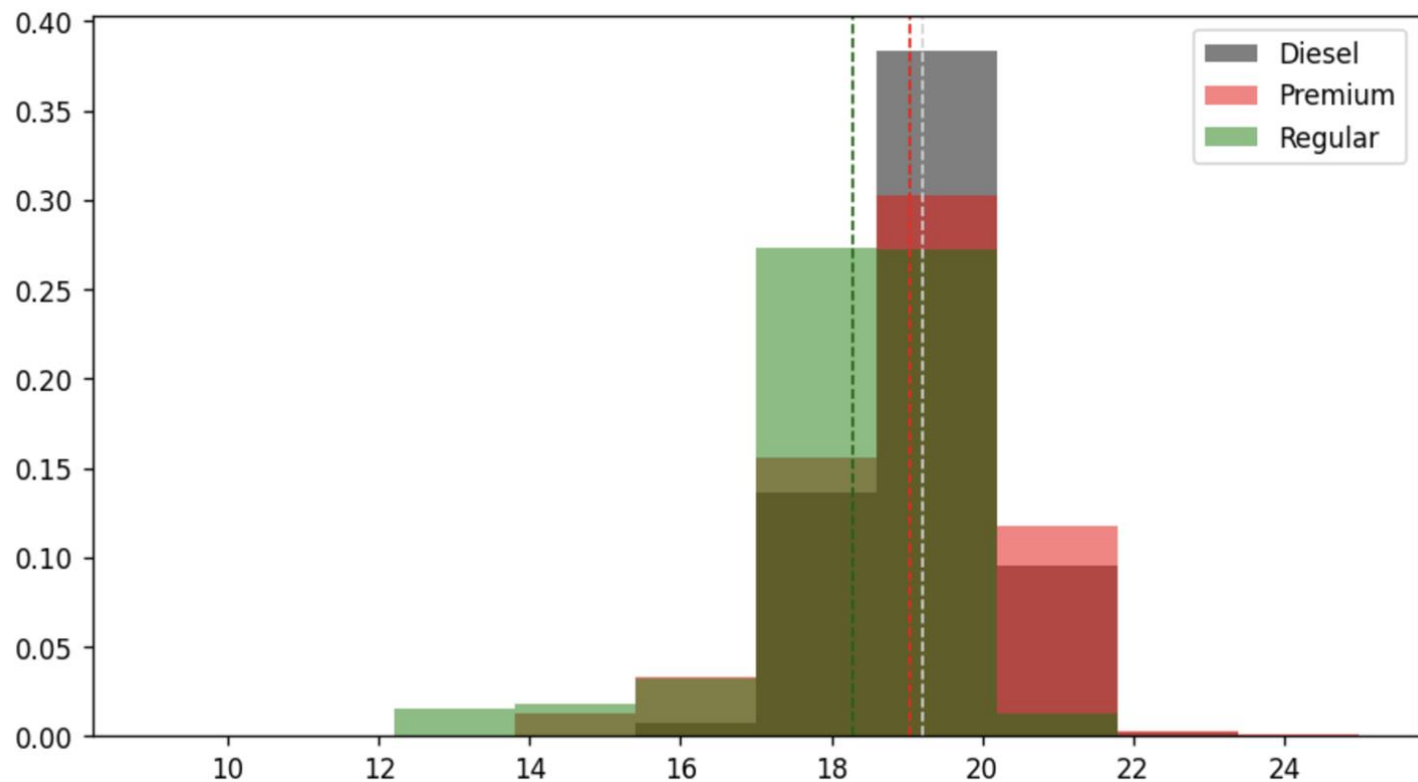
Percentage of gas type by each gas station



Rating distribution for each gas type



Fuel Price Distribution

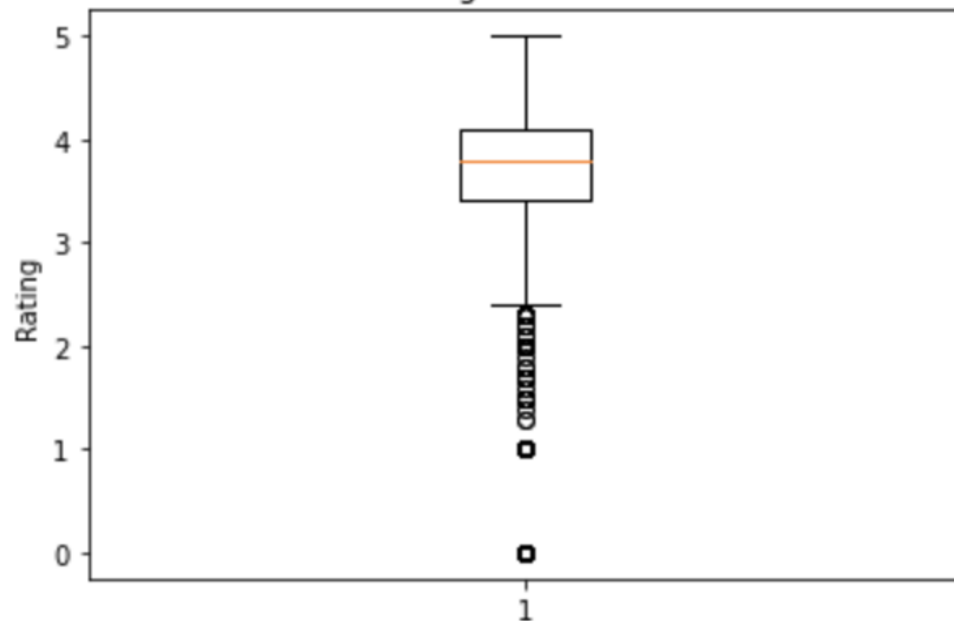


User Ratings

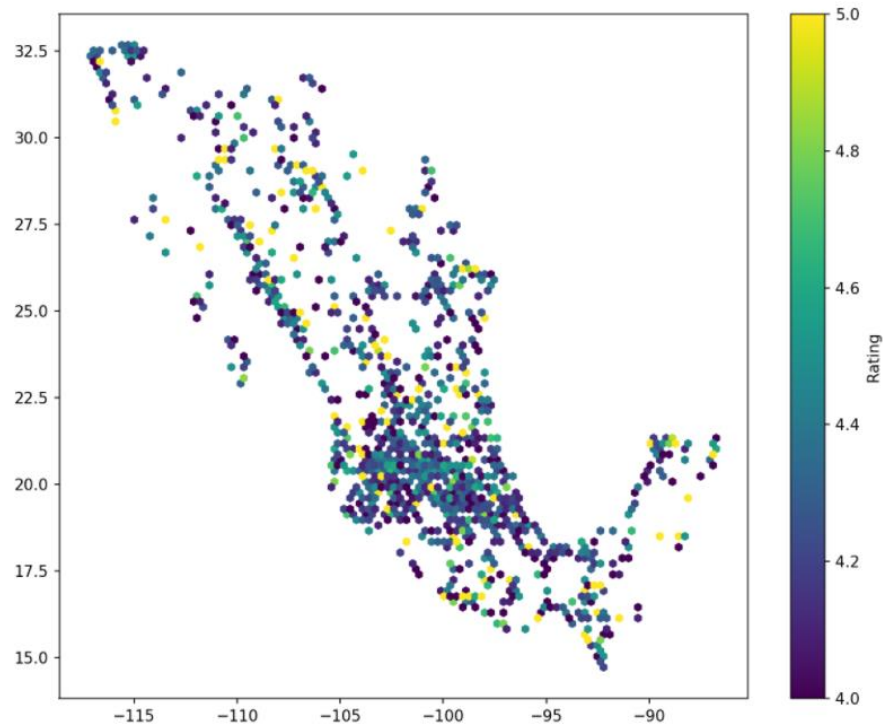
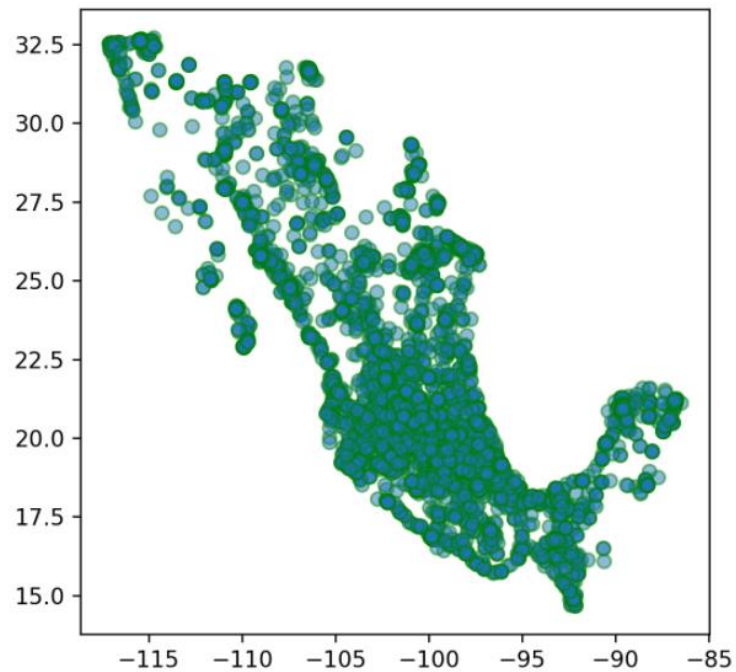
User Ratings

| Ratings | |
|-----------|------|
| <0.99 | 188 |
| 1-1.99 | 111 |
| 2-2.99 | 589 |
| 3-3.99 | 6257 |
| 4-4.49 | 2498 |
| 4.49-5.00 | 978 |

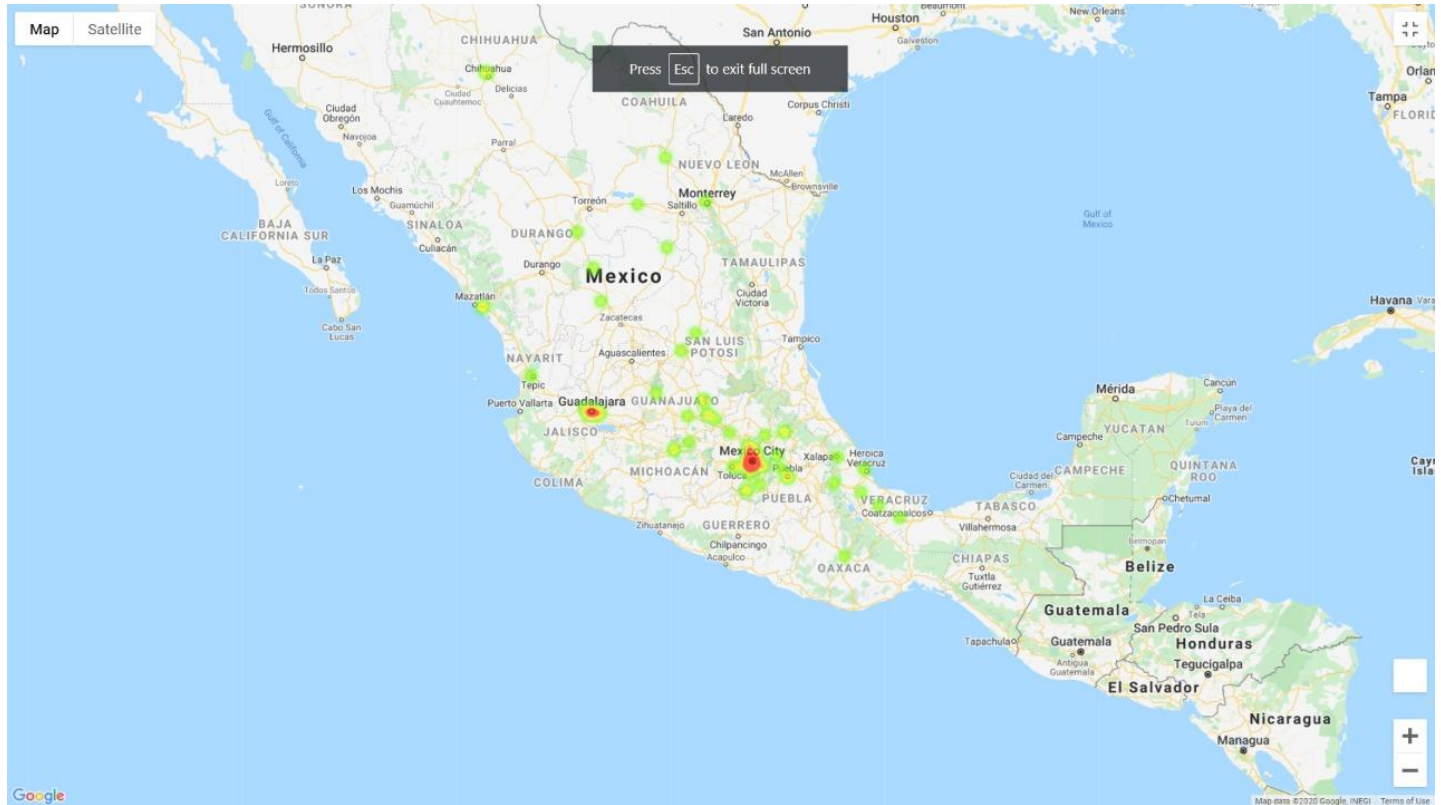
Rating Distribution



Ratings Scatterplot



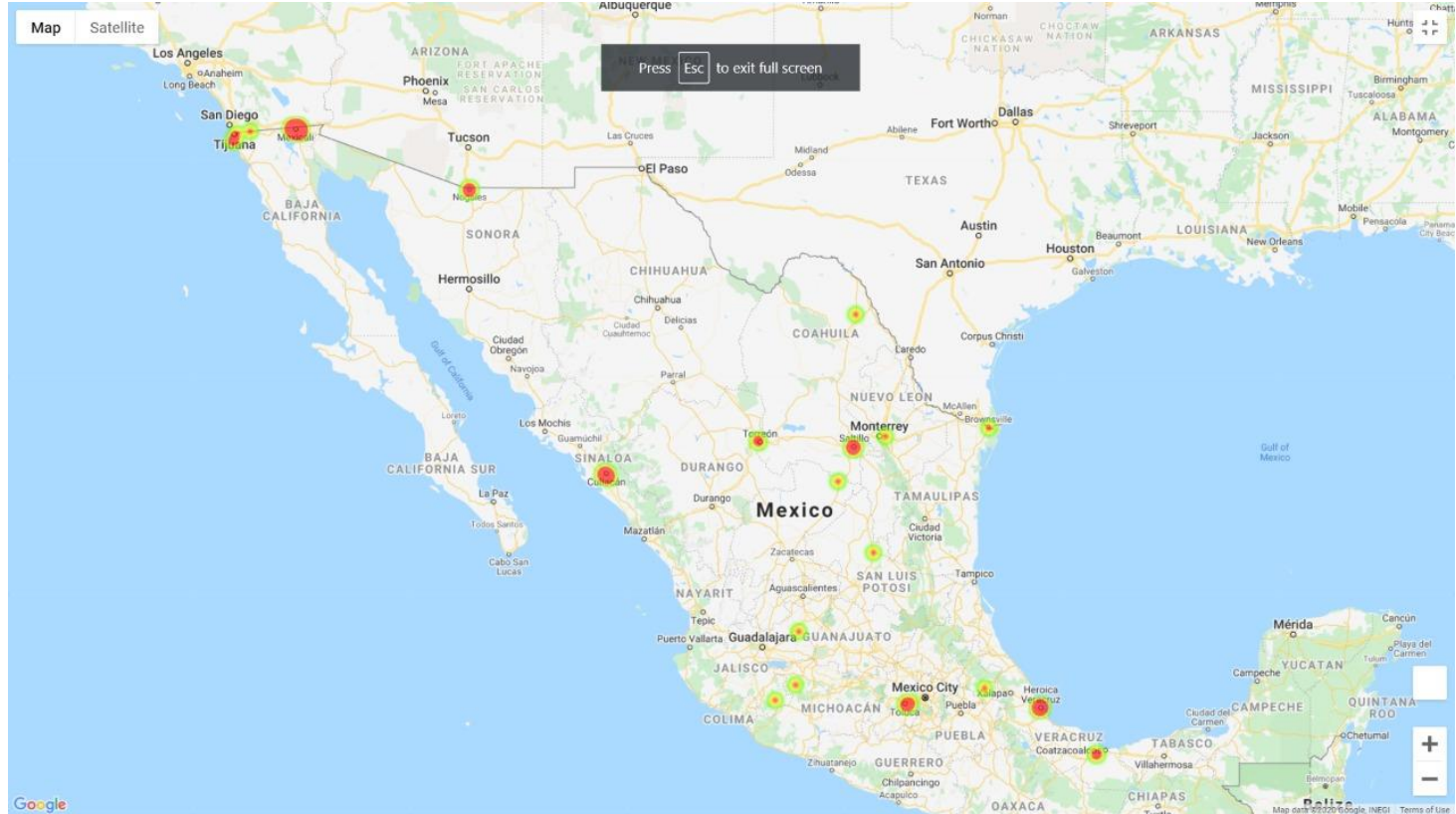
Heatmap Top 100 Best Rated



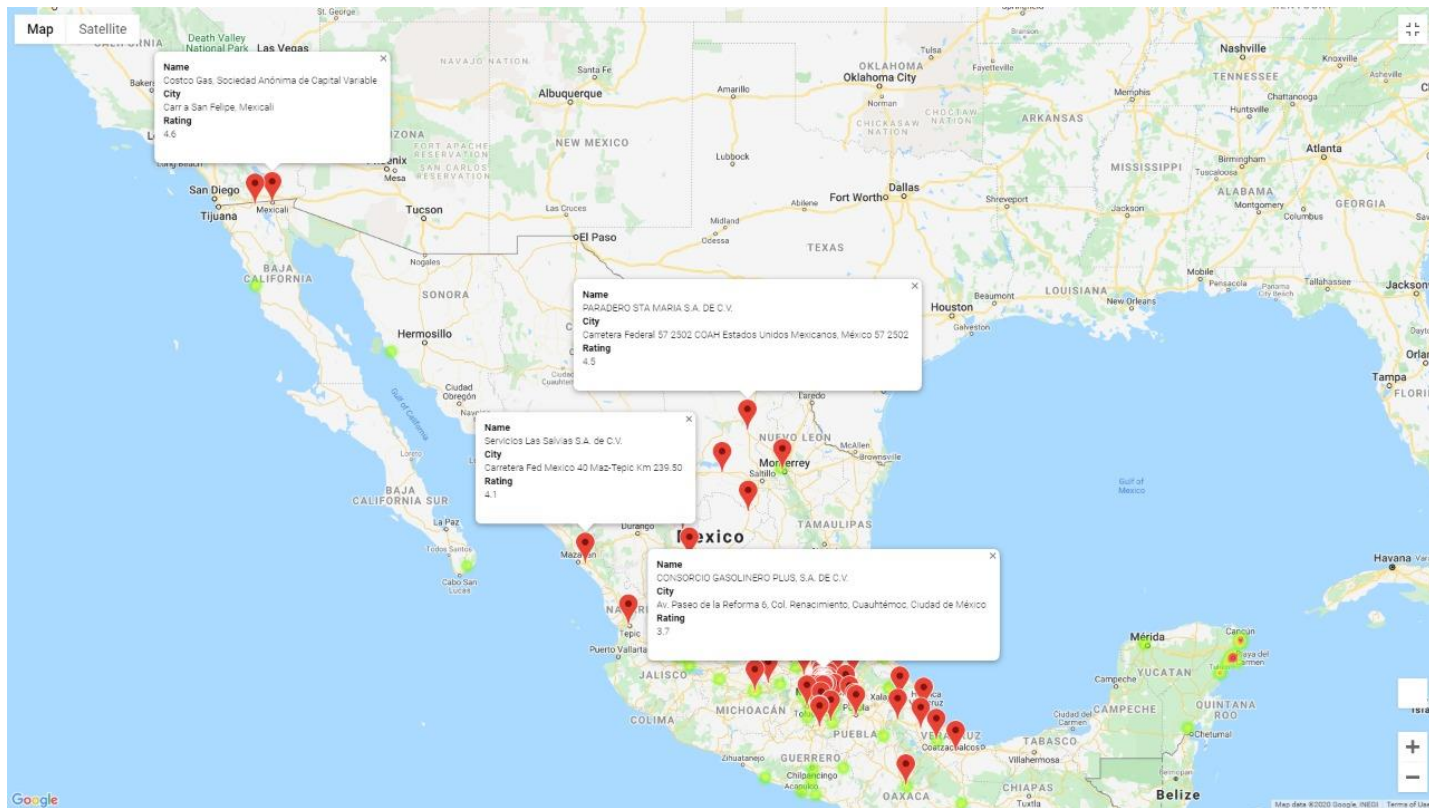
Heatmap Top 100 Worst Rated



Heatmap 100 Cheapest Places



Markers Top 100 Rated



Final thoughts...

- There is no relation between the user ratings and the fuel price.
- There's a relation between the rating and the features or added services of each of the Service Stations.
- There is not enough measurable data to determine specifically which aspect of the added services drives up the rating.
- Fuel is cheaper near the border.
- Data required to be reported by the government is insufficient to procure a thorough analysis.
- A more complete analysis requires gathering additional data from other sources, such as: mobility information, fuel distribution centers, urbanizations vs rural areas, socioeconomic status, among others.