Properties' rent & sell analysis
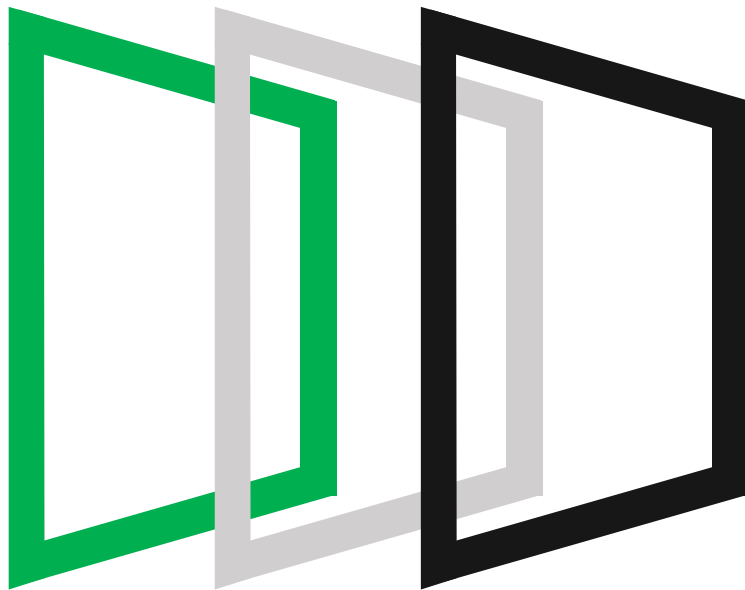
What would be the best location to invest into purchasing a house in Mexico?

Paulo Gerardo Ramos Orozco

July, 2019

Content

Context & Objective

01

First it is necessary to review the data and set that as a baseline for the project, also define the objective and possible scope...

Data Extraction

Extraction of two public data sets from Google Big Query through SQL query language

Challenge: The storage allowance is 1 GB in Google Drive and 90 MB in local. Therefore, it was necessary to split the queries in parts to extract the data

Data Sets Description

Rent_Data201_All : 633907 rows & 27 columns. CSV format, encoding UTF-8

Sell_Data201_All : 2762514 rows & 27 columns. CSV format, encoding UTF-8

Used Tools

- **Google Big Query SQL**
- **Google Drive**
- **Jupyter Notebook**
- **Python libraries :** Pandas, Matplotlib, Sklearn, Stat, Numpy, OS, Glob
- **Power Bi**
- **Git**
- **Excel**

…the objective is…

To find the best location to invest in Mexico when buying a property and renting, using different variables like; square meters, zone, prices, etc.
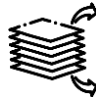
Doing..

Explore and validate the data bases to identify principal variables

Clean the data bases from non existing values and zeros

Transform and join the data bases to get the identical properties from both data sets

Generate visual analysis to identify which are the locations with higher rent and lower initial investment

Conclusions and main observations

Further potential analysis

Data Pre-processing

02

..then, the analysis continues with the data exploration, cleaning and transformation...

Data exploration

Main challenges:

➢ Big amount of information
➢ Irrelevant data for the project's scope
➢ Identify similar data for both data sets
➢ Find data structure
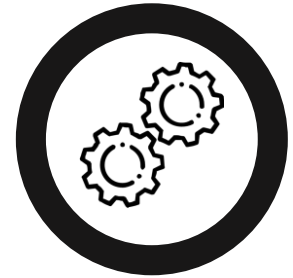➢ Discover errors in the data

Data cleaning

Main challenges:

✓ Standardize data types
✓ Clean outliers with quartiles
✓ Remove zeros on prices
✓ Remove "NaN" values
✓ Clear duplicates

Transformation

Main Challenges:

❑ Data sets with different sizes
❑ Variables validation using linear regression
❑ Join tables for a master file

…a join was made to find the exact match between rent & sell data. The final table has a total of 26900 rows and 9 columns…

property_type	place_name	state_name	lat	lon	currency	surface_total_in_m2	sell_price	rent_price
apartment	Quintana Roo	Quintana Roo	21.161908	-86.851528	MXN	98	1790000	16500
house	Colima	Colima	19.234996	-103.727212	MXN	200	2000000	8500
store	Colima	Colima	19.244782	-103.724963	MXN	435	1350000	10000
house	Puebla	Puebla	19.056652	-98.228841	MXN	364	8140000	30000
apartment	Puebla	Puebl	19.053259	-98.226545	MXN	150	3400000	17000
apartment	Puebla	P	19.035937	-98.223369	MXN	96	2327000	12000
apartment	Puebla		19.035937	-98.223369	MXN	96	2100000	12000
apartment	Centro		17.989	-92.942795	MXN	100	1900000	13000
house	Cuautla		18.840925	-98.934196	MXN	130	1800000	5000
house	Cuautla		18.840925	-98.934196	MXN	130	1550000	5000
house	Cuautla		18.840925	-98.934196	MXN	130	890000	5000
house	Cuautla	Morelos	18.840925	-98.934196	MXN	130	1500000	5000
house	Cuautla	Morelos	18.840925	-98.934196	MXN	130	1400000	5000
house	Cuautla	Morelos	18.850763	-98.943855	MXN	130	950000	7000
house	Cuautla	Morelos	18.850763	-98.943855	MXN	130	1200000	7000
house	Cuautla	Morelos	18.850763	-98.943855	MXN	130	1250000	7000
house	Cuautla	Morelos	18.864874	-98.925911	MXN	400	2000000	6800
house	Cuautla	Morelos	18.864874	-98.925911	MXN	400	2700000	6800
house	Cuautla	Morelos	18.864874	-98.925911	MXN	400	5300000	6800

These columns matched between the 2 data sets

These 2 columns where merge from the data sets

***Source: join_data_Mexico_since_2015.csv

Data Analysis

03

The first visual analysis was on this map where it can be observe how most of the data is concentrated in the center of the country...

....followed is a visual analysis is the bar & line chart which compares average rent & sell prices in Mexico...

Average Sell_Price vs Rent_Price

Legend:
- Average Sell_Price
- Average Rent_Price
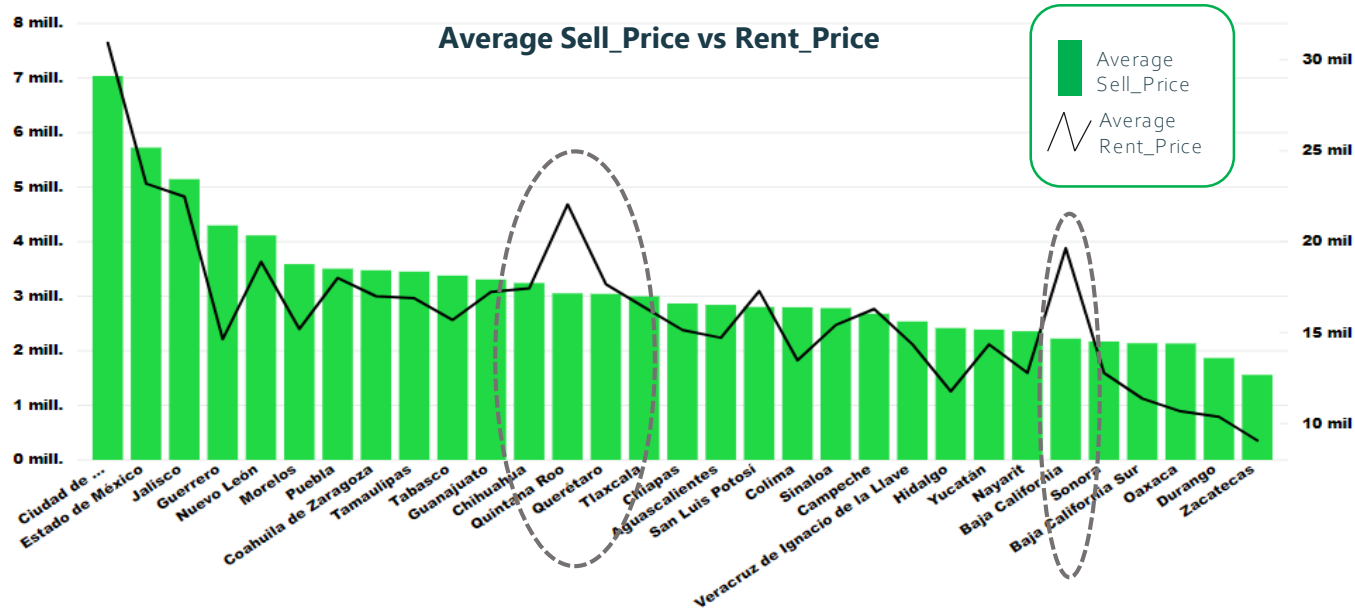
In this graph the average for each 32 states in Mexico for sell & rent prices can be observed.

This allows an initial evaluation per state

The following states have low buying costs and high rent prices:

- Chihuahua
- Quintana Roo
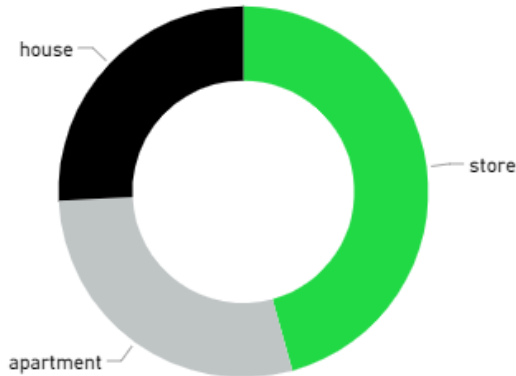- Queretaro
- Tlaxcala
- Baja California

The main observation is that the group of states in the circles have low sell prices and high rents

***Source: join_data_Mexico_since_2015.csv

… next is the average sell & rent prices by property type in all the country..

Average Sell_Price
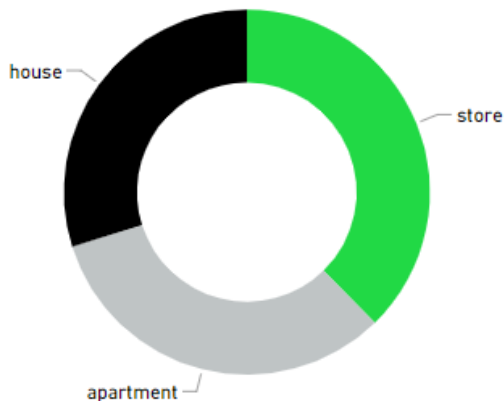
Average Rent_Price

For the average sell price, "store" holds the highest price in the market with 38%, followed by "apartment" with 32% and finally "house" with 30%.

For the rent price, "store" has the highest average price with 46%, followed by "apartment" with 28% and finally "house" with 26%

The average sell prices are similar for all property types, but there is a significant higher rent prices for "stores", making it the property type with the fastest ROI

***Source: join_data_Mexico_since_2015.csv

… to validate the variables with more influence in the rent and sell price, a linear regression was performed…

$$Y_{Price\ r\&s} = B_0 + B_{1_{S\_m2}} + B_{2_{State1}} + B_{n_{.....}} + B_{34_{State32}} + \varepsilon$$

Response variable Coefficient Independent variables Error

ANOVA (Linear Regression)

	coef	std err	t	P>|t|	[0.025	0.975]
const	6523.9140	118.049	55.265	0.000	6292.542	6755.286
surface_total_in_m2	47.2473	0.164	288.927	0.000	46.927	47.568
Aguascalientes	-1513.7605	283.224	-5.345	0.000	-2068.871	-958.650
Baja California	2249.0992	423.675	5.309	0.000	1418.707	3079.492
Baja California Sur	-2156.2375	453.724	-4.752	0.000	-3045.524	-1266.951
Campeche	307.3672	256.218	1.200	0.230	-194.814	809.548
Chiapas	-1272.6985	194.685	-6.537	0.000	-1654.277	-891.120
Chihuahua	1340.3112	212.237	6.315	0.000	924.332	1756.290
Ciudad de México	1.513e+04	158.229	95.604	0.000	1.48e+04	1.54e+04

The state names were converted to categorical variables, in order to run the linear regression and compare the significance on each state (zone) and surface (square meters)

It can be observed that the states (zones) have a higher influence on the price than the surface, being "Ciudad de Mexico" the one with highest significance in the model

…having a linear regression the next step is to validate the correlation between price and surface…
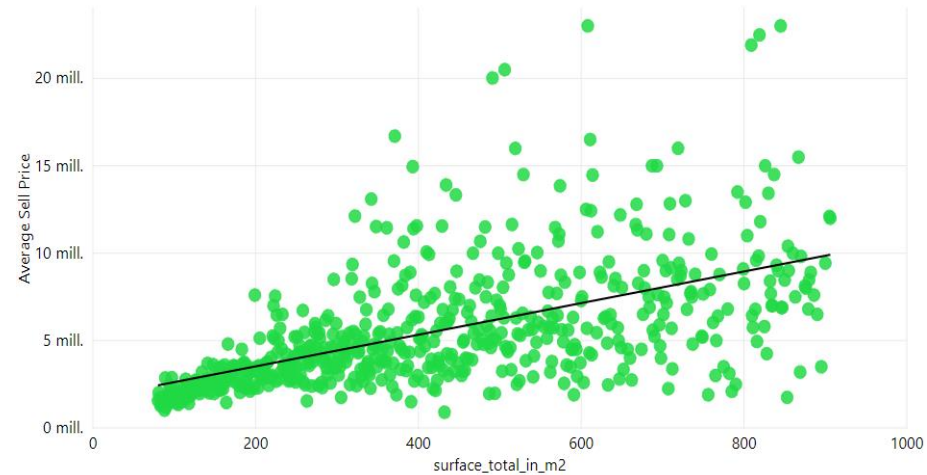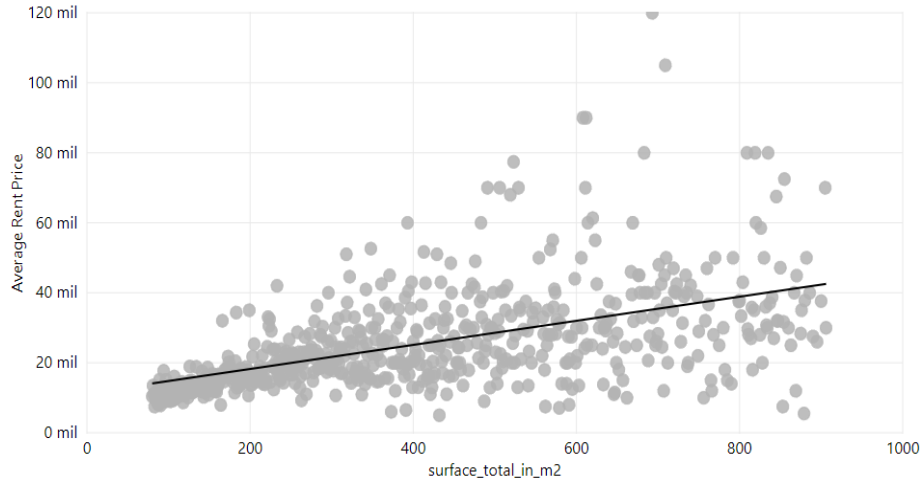

At country level the correlation between price and surface is clearly positive, meaning that as bigger the surface is, higher the prices

...finally the following charts zoom into the 5 states mentioned before to identify which are the best investment locations.

Chihuahua

Tlaxcala

Average Sell_Price

Average Rent_Price
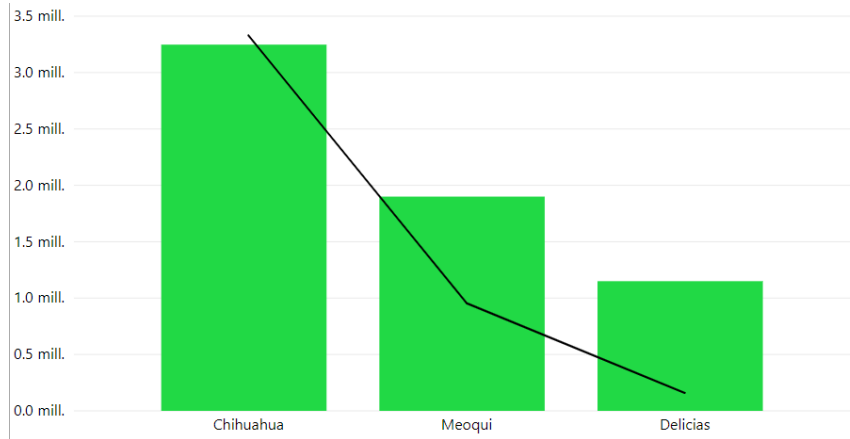
The best place to invest in Chihuahua is Chihuahua place, ROI = 185 months
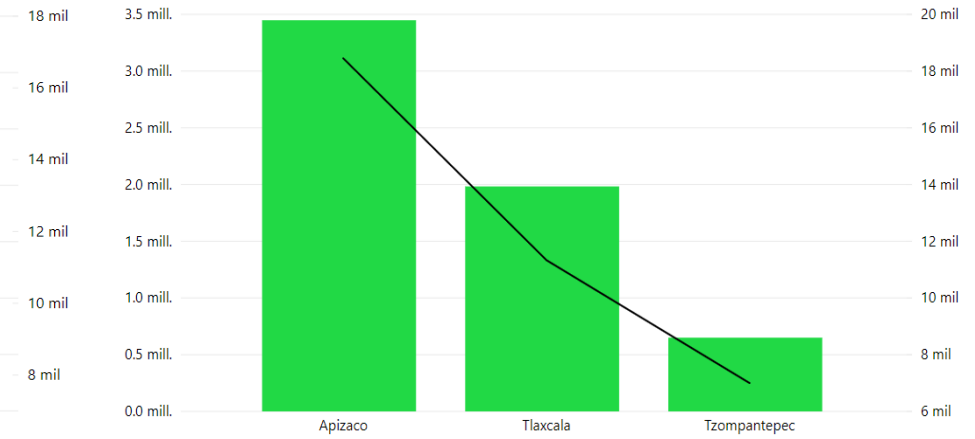
The best place to invest in Tlaxcala is Tzompantepec, ROI = 92 months

03. Data Analysis

Quinta Roo

Baja California

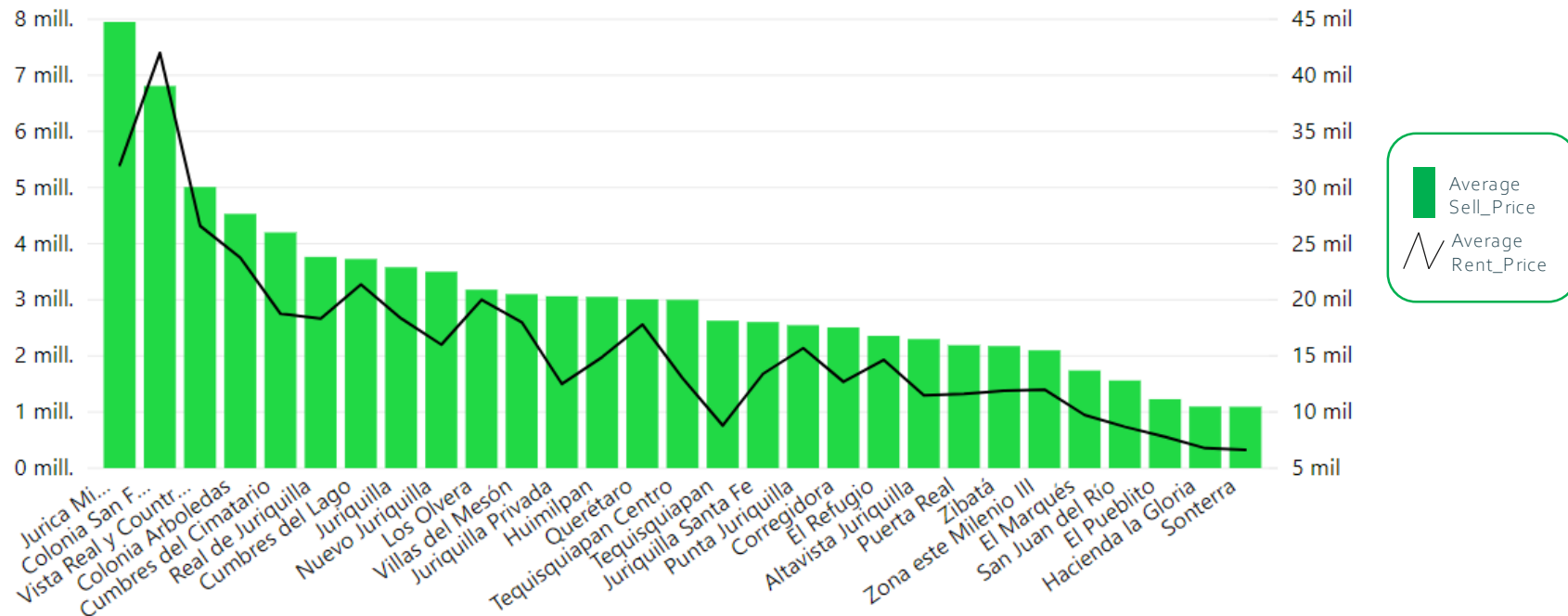Average Sell_Price
Average Rent_Price
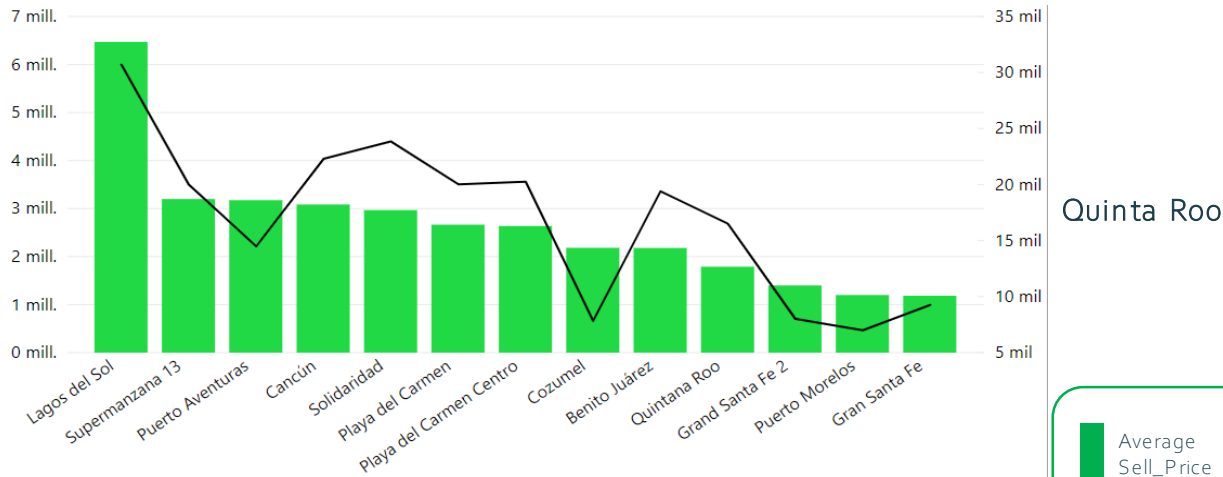
The best place to invest in Quintana Roo is Benito Juarez, ROI = 112 months
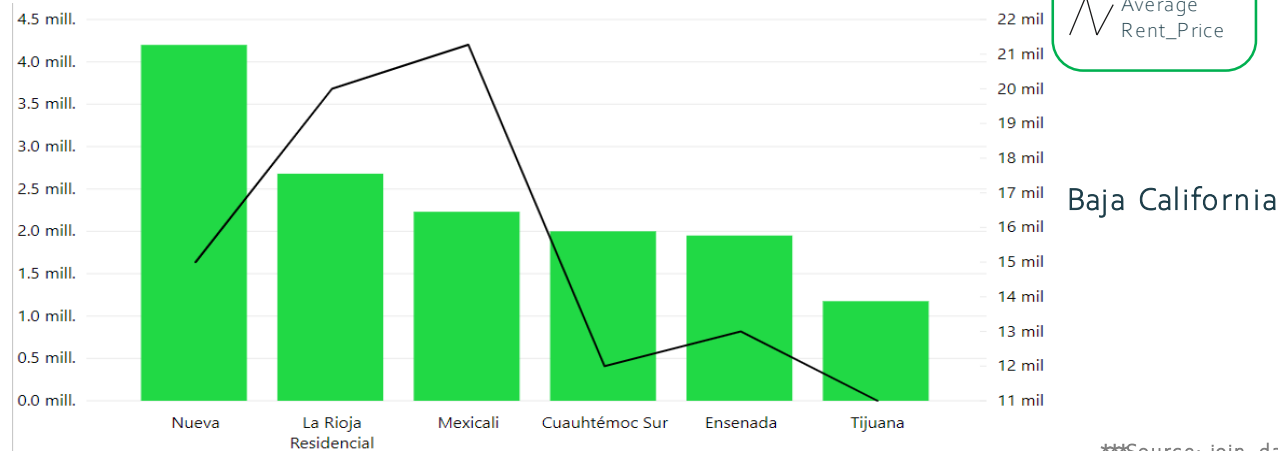
The best place to invest in Baja California is Mexicali, ROI = 105 months

***Source: join_data_Mexico_since_2015.csv

16

Querétaro

The best place to invest in Queretaro is Colonia San Francisco, ROI = 163 months

Recommendations & Further analysis

04

After a long path of data analytics the following recommendation were obtained in order to get the most of a real state investment and a further analysis is suggested to

Recommendations

- ✓ When making real state business, location is the most important variable

- ✓ The top 5 states to invest in Mexico are: Chihuahua, Tlaxcala, Queretaro, Quintana Roo, and Baja California

- ✓ Tlaxcala is the best option to invest being Tzompantepec the best location within the state

- ✓ The property type which will give the fastest ROI is "Store", followed by "apartment"

*All this is considering only price factor

Further Analysis

More variables can be considered: **economics, life quality, crime rate, weather conditions, population, pollution**.

To get these variables, **web scrapping** can be performed from different public sources in the country

A **ML model with classification** can be implemented in order to get more value out of the data.

Thank you

Time Invested in the Project:

Data Extraction (BigQuery SQL) : 0.5 Hours
Data Transformation (Jupyter, Python, excel) : 1.5 Hour
Data Visualization (Power Bi) : 2 Hours
Validation & review : 1 Hour

Total hours : 5