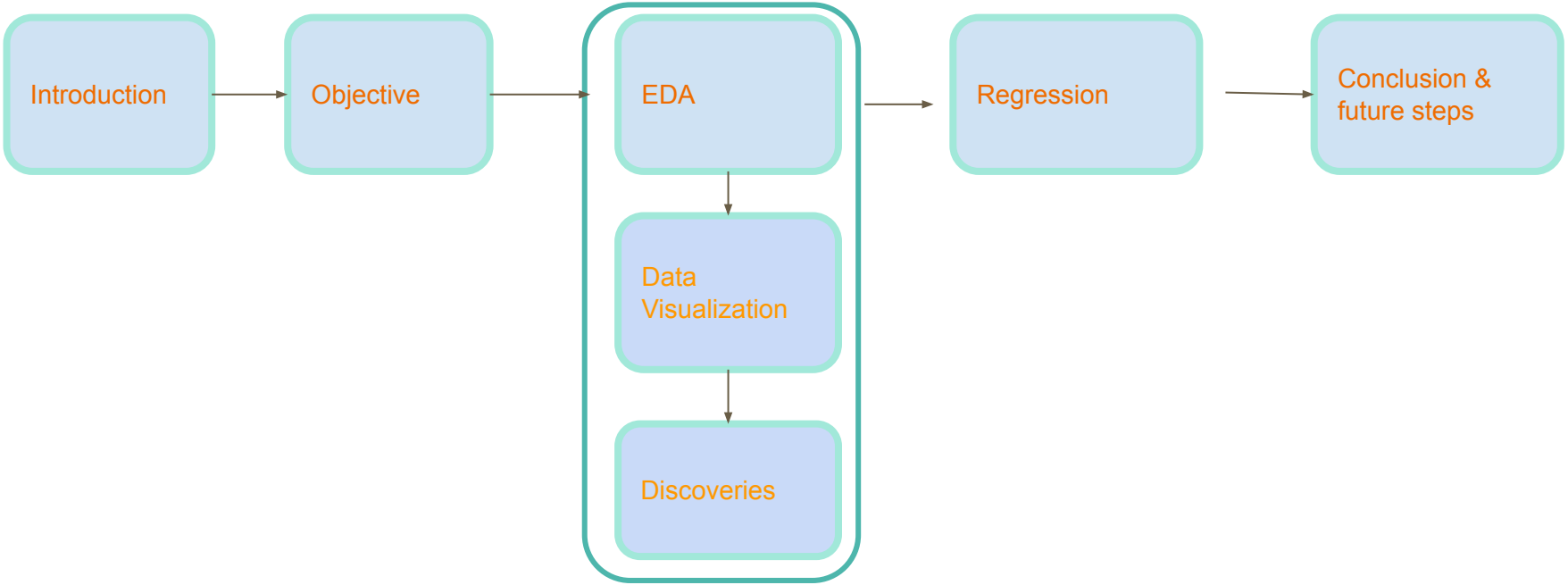
CodeOp Challenge

Presented by: Neha Sharma

Outline:


Introduction
Objective
EDA
Data Visualization
Discoveries
Regression
Conclusion & future steps

Introduction

Average monthly rent & rent per surface for Barcelona City

Source: Open Data BCN
Theme: Territory
Category: Housing

Average Monthly Rent and Rent per surface

Introduction (I):

- Average monthly rent(euro/month)
- Average rent per surface (euro/m2 per month)
- For Barcelona City
- 6 datasets from year 2014 to year 2019

Objective

Objective (I):

- Visualize data for insight information.
- Build a Linear Regression model to predict the rent prices.

Exploratory Data Analysis

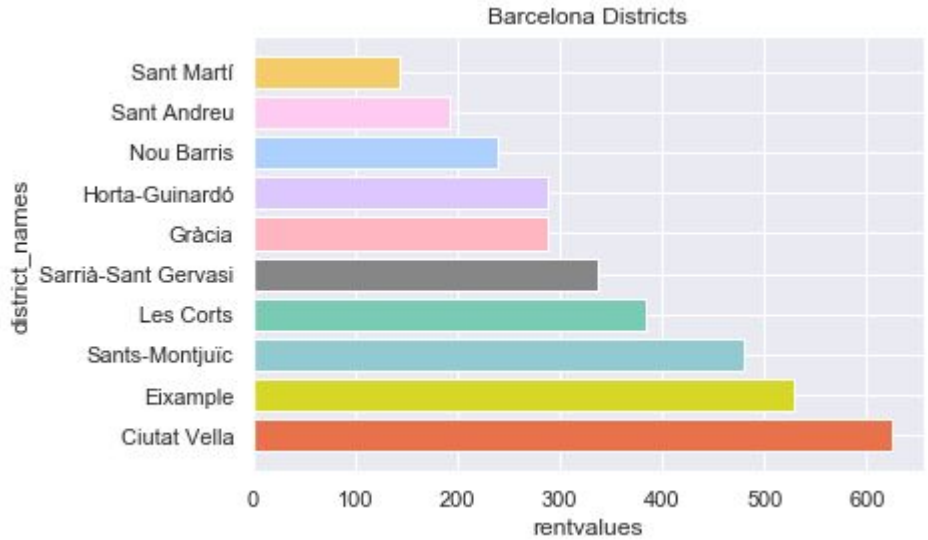
EDA:

- 6 datasets (2014 - 2019):
  - 584 rows and 8 columns
- Dataset after concatenating all year datasets:
  - 3504 rows and 8 columns
- We have 2 types of rent prices:
  - Average rent per month: Dataset1 (1752,8)
  - Average rent per surface per month: Dataset2 (1752,8)

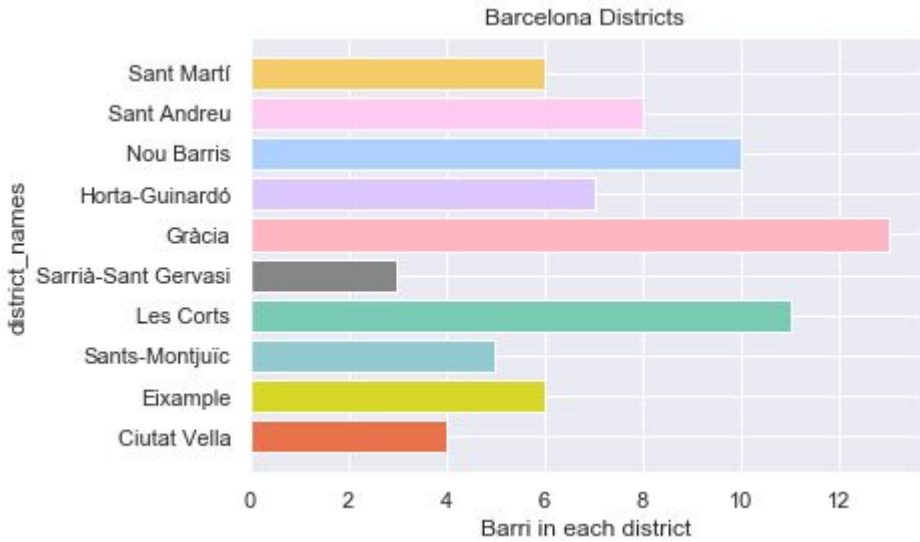
- Columns:
  - "Any" : year (2014 - 2015), datatype: integer
  - "Trimestre" : Data Quarter, datatype: integer
  - "Codi_Districte": District code, datatype: integer
  - "Nom_Districte": District name, datatype: object/string
  - "Codi_Barri": Neighbourhood code, datatype: integer
  - "Nom_Barri": Neighbourhood name, datatype: object/string
  - "Lloguer_mitja": average monthly rent or rent per surface, datatype: object/string
  - "Preu": Rent, datatype: Float

Data Visualization:

- District names: 10 districts

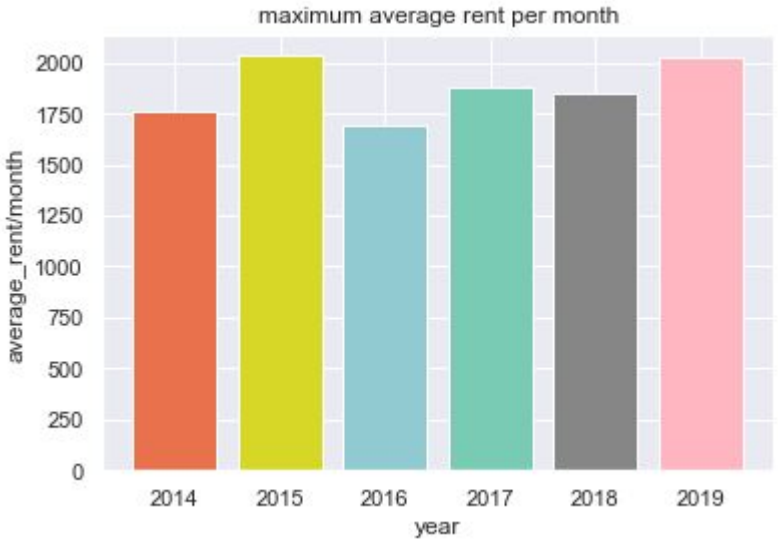


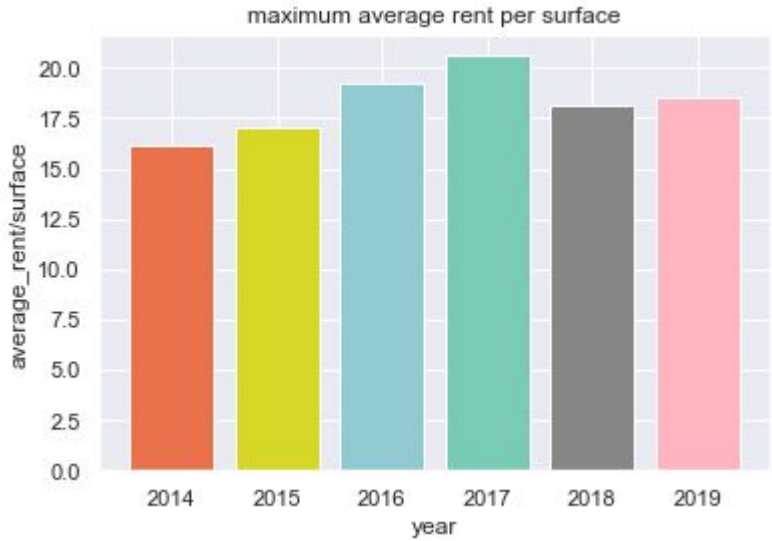
- Number of neighbourhoods in each District: 73 Barris



Data Visualization:

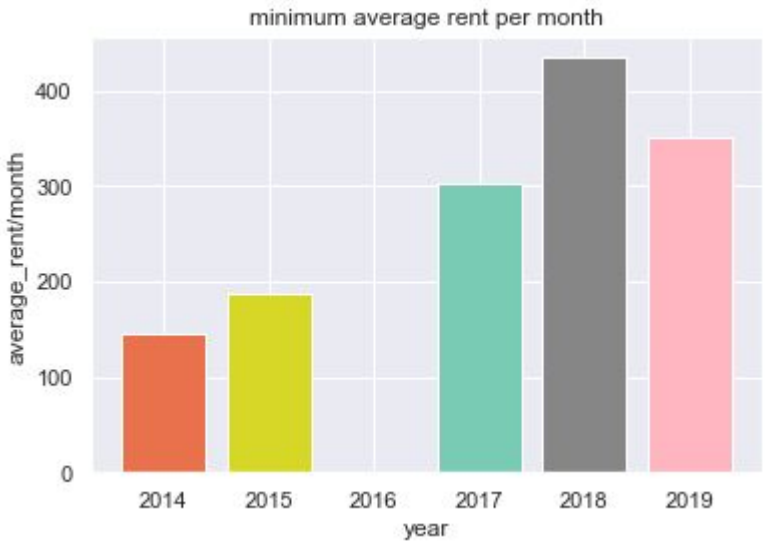
- Maximum rent by year:
  - Average rent per month
  - Average rent per surface

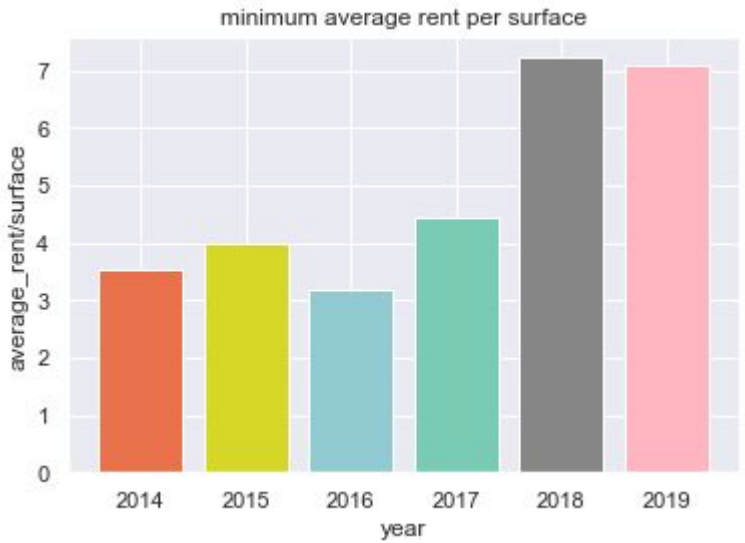




Data Visualization:

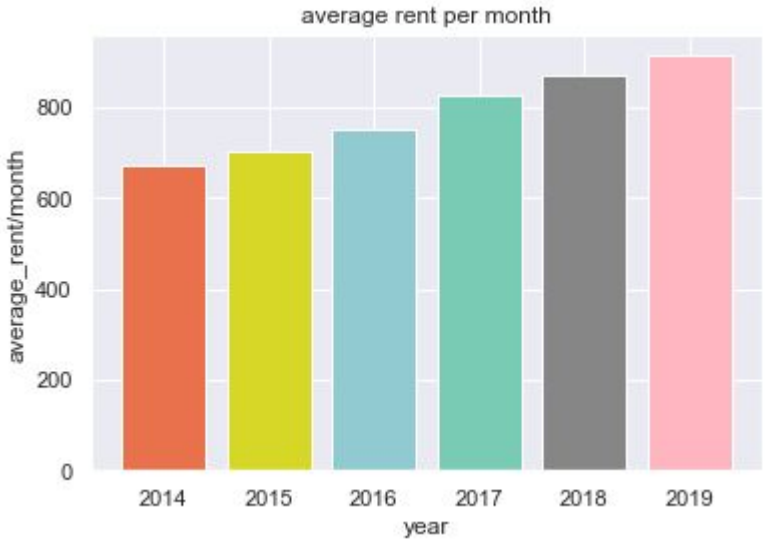
- Minimum rent by year:
  - Average rent per month
  - Average rent per surface

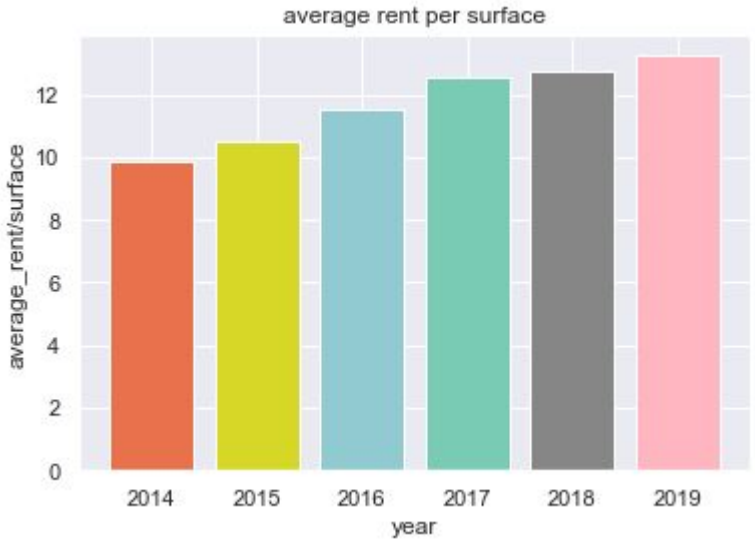




Data Visualization:

- Average rent by year:
  - Average rent per month
  - Average rent per surface

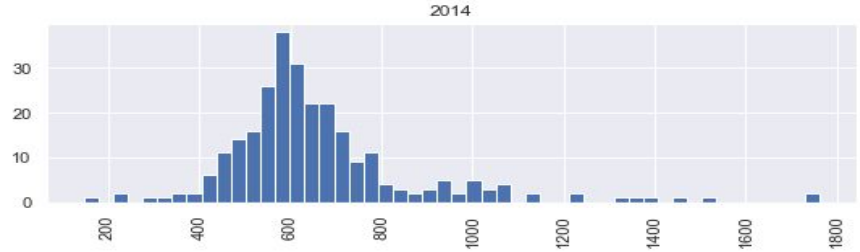


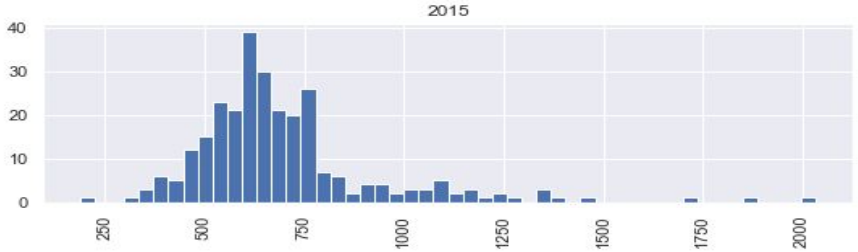


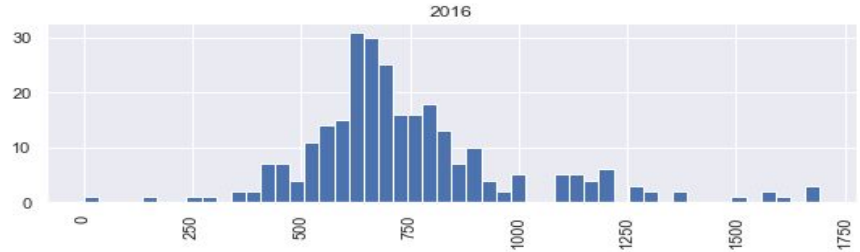
Discoveries:

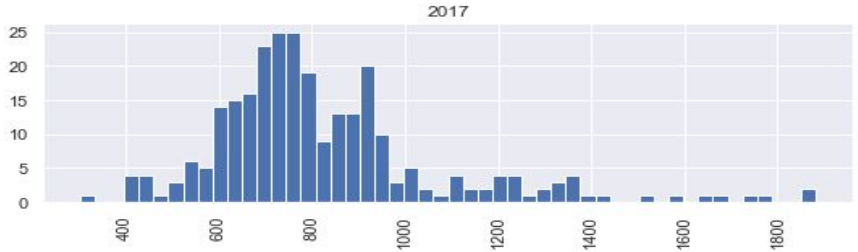
- Average maximum rent per month District and neighbourhood:
  - District: Sarria-Sent Gervasi
  - Barri: Les Tres Torres
  - Year 2014 to 2019
  - 2034.00 euros in 2015
- Average maximum rent per surface district and neighbourhood:
  - District: Sarria-Sent Gervasi
  - Barri: Les Tres Torres
  - Year 2014 to 2019
  - 20.61 euros in 2017
- Average minimum rent per month district and neighbourhood:
  - District: Ciutat Vella
  - Barri: Baro de Viver
  - Year: 2014 to 2019
  - 0 euros in 2016 *due to missing values
- Average minimum rent per surface district and neighbourhood:
  - District: Ciutat Vella
  - Barri: Baro de Viver
  - Year: 2014 to 2019
  - 3.18 euros in 2016

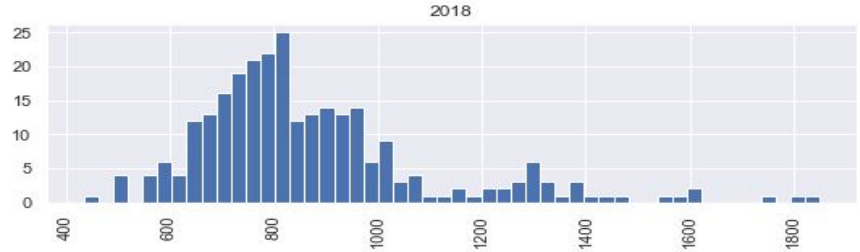
Data Visualization:

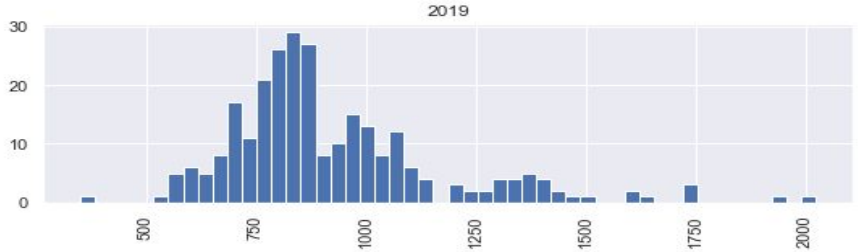




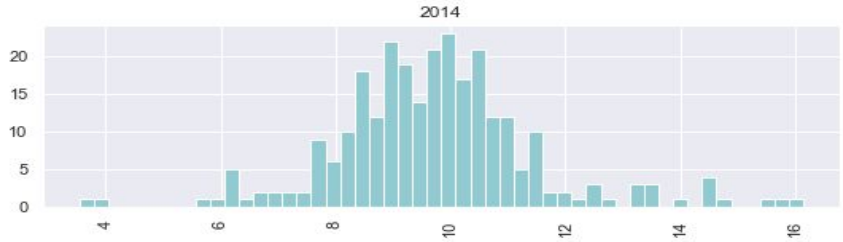


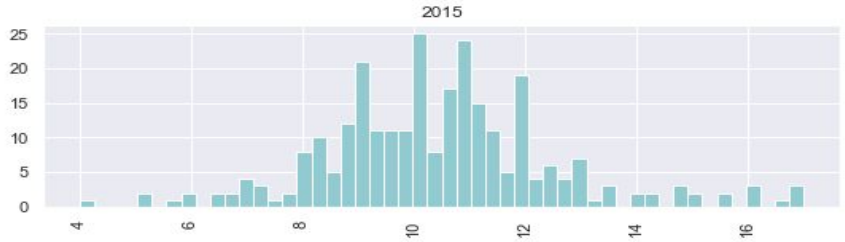


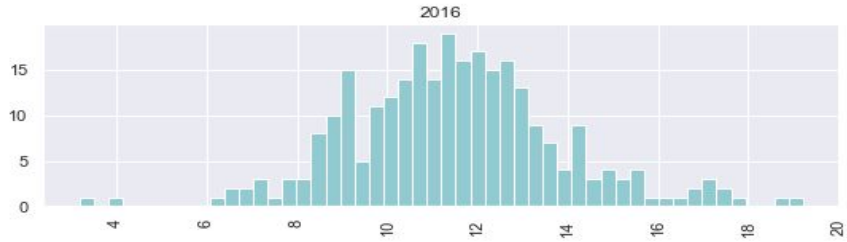


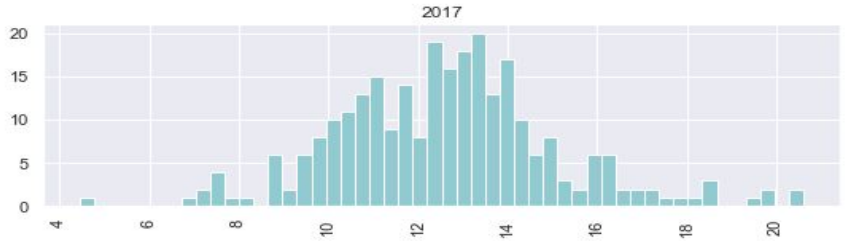


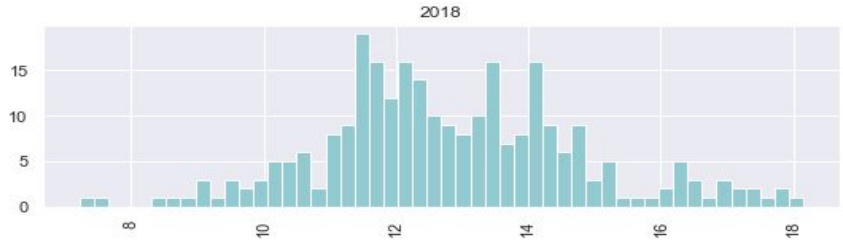
Data visualization:

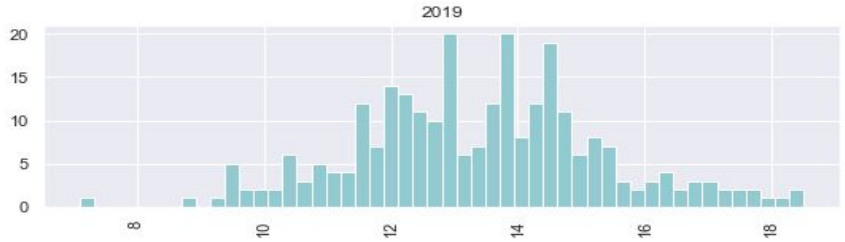












Discoveries:

- Outliers
- Missing values in “Preu” column (229)

Linear Regression

Linear Regression:

- Regression problems are a type of supervised learning problems - i.e. problems in which the target values are observable - where the target variable is continuous.
- Linear models assume an equal (linear) impact of our feature on our target for any value of our feature. And it can be expressed in this general way: $y = f(x) = \alpha + \beta x$
- Where *y* is our target variable, **x** is our vector of features, α is the intercept and β is the coefficient that explains the effect of feature x on our target variable *y*.

Evaluation:

Evaluation	Value
MAE	385.289
MSE	177010.176
RMSE	420.725
Accuracy	0.0050

Conclusion & Future Steps

Conclusion:

- Average rent per month increased by 40% approx. in 6 years
- Average rent per surface increased by 30% approx. in 6 years
- 70-80% of salary(minimum salary condition) goes into rent while 30% of salary should go for rent.
- District: Sarria-Sent Gervasi , Barri: Les Tres Torres are the expensive District and neighbourhood
- District: Ciutat Vella, Barri: Baro de Viver are the cheapest locality.
- There is no correlation between features to see the prediction.

Future Steps:

- Collect more data to see why different Districts has different rent prices.
- Collect more data to correlate features and make predictions
- Use of different Regression models.

Thank You

Questions?