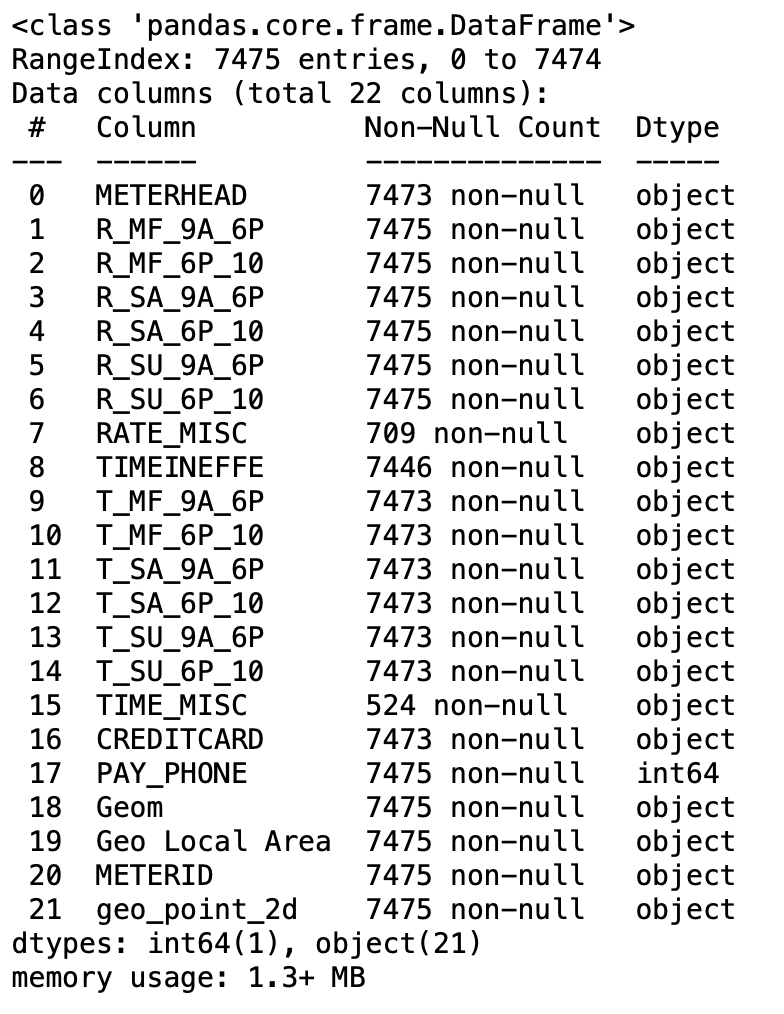
**1) Introduction and Discovery**

*Title : An Analysis about parking meters in Vancouver City*

The dataset is about parking meters in the city of Vancouver, British Columbia, Canada. I look to delve into the data to gather visible insights about paying to park in the city. I then try to classify the meters into three different localities and also try to predict the average price a person would spend for parking using different classification and regression machine learning models. Vancouver is a very expensive city to live in, in addition to the already sky-rocketing expenditure, the parking expenditure needed a few insights. I assume that the locality plays a vital role in determining the amount spent for parking.

**2) Data Preparation**

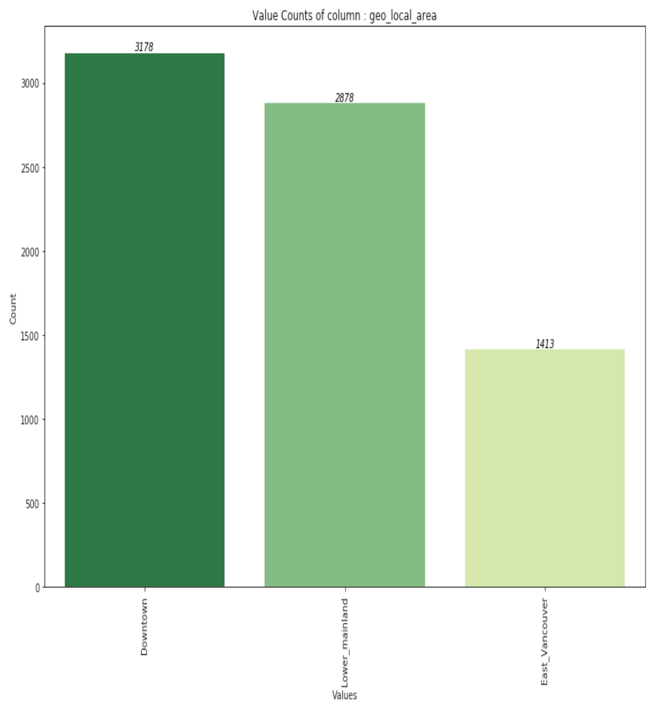
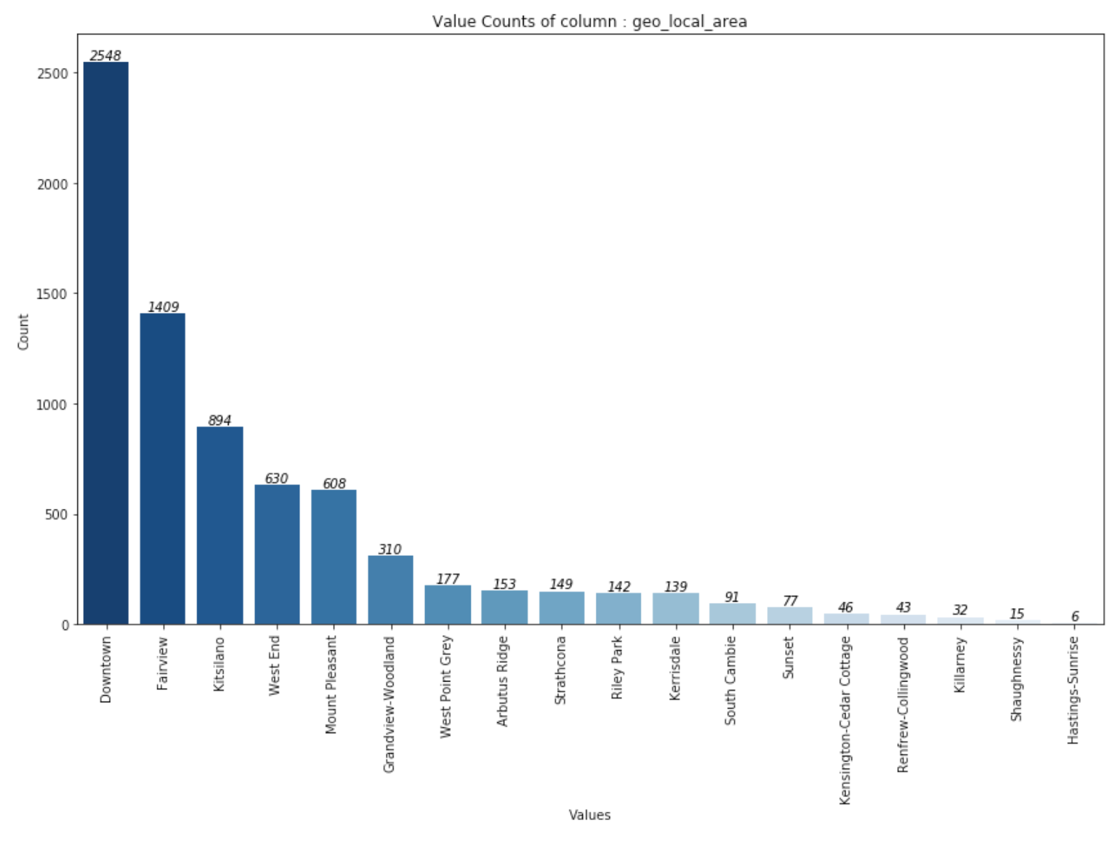


The dataset contains information about various types of parking meter heads, the rate per hour for different time slots of the weekdays, Saturday and Sunday and the time limits during the same, if the meter accepts credit card and where it is located. The dataset was obtained from open data of the Vancouver corporation.

All the non-numerical columns were cleaned to be numerical

(i.e. dollar signs were removed from currency column, etc.,). New columns, representing aggregated values for weekdays, Saturday and Sunday with regards to pay rate per hour and time limit was created.

The locality column was created from the ‘geo\_local\_area’ column, based on the map provided below.



(The above plots were taken from the jupyter notebook)

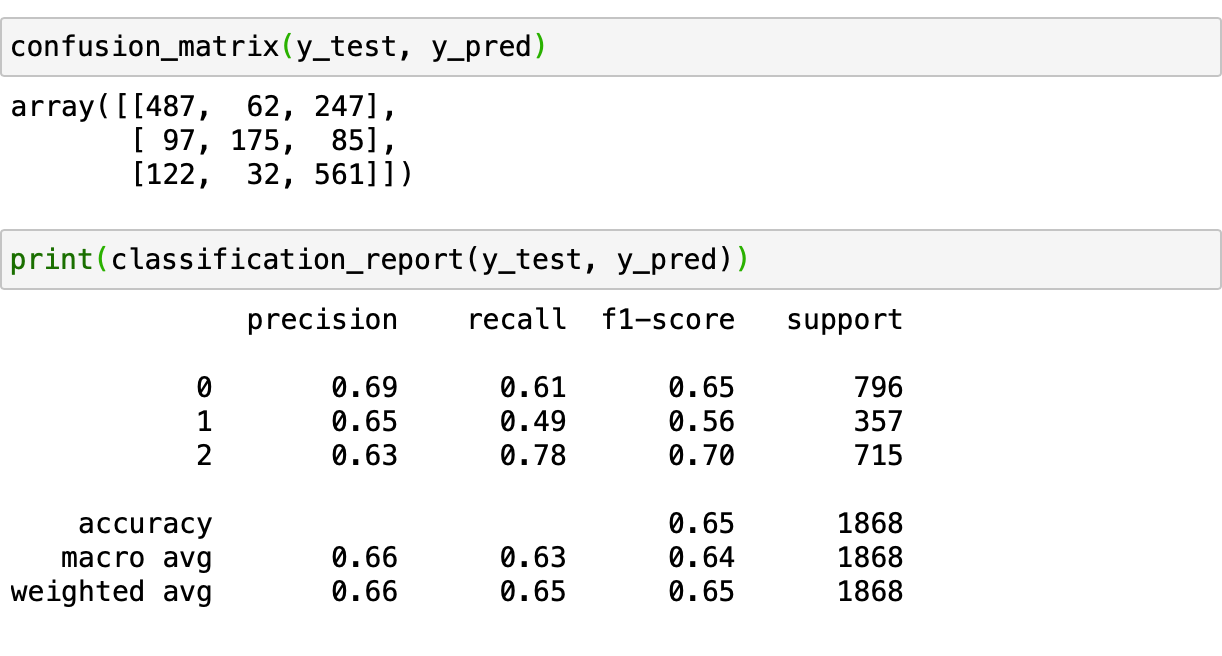
**3) Model Implementation**

\*CLASSIFICATION : A variety of classification models are tested, with ‘geo\_local\_area’ target, that has 3 classified localities, as shown in the green bar plot above. A multi-class classification model suits best for this situation and iterating through different models, helps choose the best fitting one for the dataset. The classification model, helps to understand if the parking meters are actually classifiable based on the given factors.

\*REGRESSION : Numerous regression algorithms are tested, with ‘avg\_amount\_spent\_week’ as the target. Multiple-regression algorithms helps to predict the average amount that will be spent in a week by someone. The model helps to understand how high or how low the expenditure on parking can be.

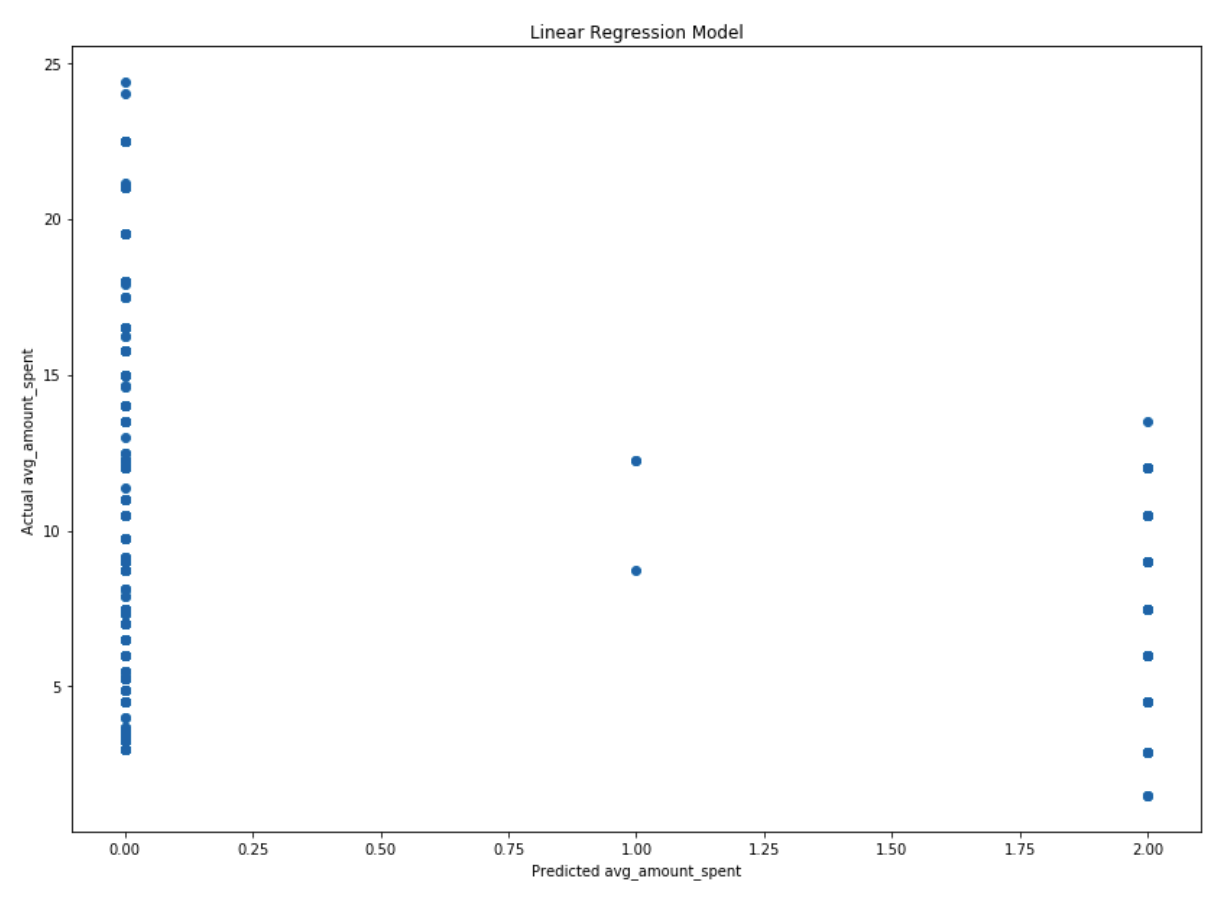
**4) Results Interpretation and Implications**

*CLASSIFICATION MODELS*



The classification model performed with an accuracy of 65%, which even in regards to real world data is a poorly performing model. This might be because of too much aggregated data which is almost redundant, from the data engineering performed. Although I am unable to assess mistakes in the steps carried out during the implementation of the model, the model is not of any significance. The analysis needs proper interpretation and use of geographic data, which might divert the analysis in a meaningful direction. The results in reality make no sense.

*REGRESSION MODELS*



The regression model performed with a 100% accuracy for prediction. This is also because of the aggregated nature of the target and also multiple features, where the target has very minimal variance to even make sense. The model however was carried out with efficiency.

The Key findings would be that majority of parking meters are centered around downtown and West Vancouver. A person looking to reduce parking costs should look to use public transit when travelling downtown as the transit system is extremely efficient and quick and hassle-free in and around downtown area.

**5) Concluding Remarks**

The dataset was engineered to be run through classification and regression algorithms. But the choice of dataset was very poor. The data might serve well for data visualization, involving geographic visuals of locality, parking meter concentrations etc.

**6) Video Presentation Link**

<https://youtu.be/8_kZQNVUIfo>