Presentation Agenda : **House pricing High Level Prediction**

Data Scientists : Nasim,Satheesh,Sandeep

- Introduction
- Data Cleaning and Visualization
- Poster Presentation

Business Analytics - House pricing High Level Prediction

Prof. Dr. Itauma,Itauma

Data Scientists : Nasim,Satheesh,Sandeep

Introduction: To predict housing price for maximum of about three bedrooms apartments in Melbourne,Australia by regionwise.Here cosidered the hypothesis can discussed about distance,year of built,no. of bedrooms with integrated washrooms as well as overall occupancy size.

Refining Method:

Imputation for data completeness and here records having 25% missing data removed from the raw dataset.

```
Imputation.py > ...
1   import pandas as pd
2   df = pd.read_csv("Melbourne_housing_FULL.csv")
3   df.head(2)
4   housing = df[['Price','Bedroom2','Bathroom','Landsize','BuildingArea','YearBuilt','Distance','Regionname']].copy()
5   print(housing.isnull().sum())
6   print((housing.isnull().sum()/len(housing))*100)
7
8   #To find the mean and update the missing values in Bedroom mean

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  JUPYTER

PS D:\WISA\USA\NU_MSBA\CLASSES\Subject3and4\MIS653ProgrammingandComputation\PythonFiles> & C:/Users/mnasi/AppData/Local/Microsoft/WindowsApps/python3.10.exe d:/WISA/USA/NU_MSBA/CLASSES/Subject3and4/MIS653ProgrammingandComputation/PythonFiles/Imputation.py
Price          7610
Bedroom2       8217
Bathroom       8226
Landsize      11810
BuildingArea  21115
YearBuilt     19306
Distance          1
Regionname        3
dtype: int64
Price         21.832057
Bedroom2      23.573457
Bathroom      23.599277
Landsize      33.881286
BuildingArea  60.570068
YearBuilt     55.386293
Distance       0.002869
Regionname     0.008607
dtype: float64
PS D:\WISA\USA\NU_MSBA\CLASSES\Subject3and4\MIS653ProgrammingandComputation\PythonFiles>
```

Data Analysis

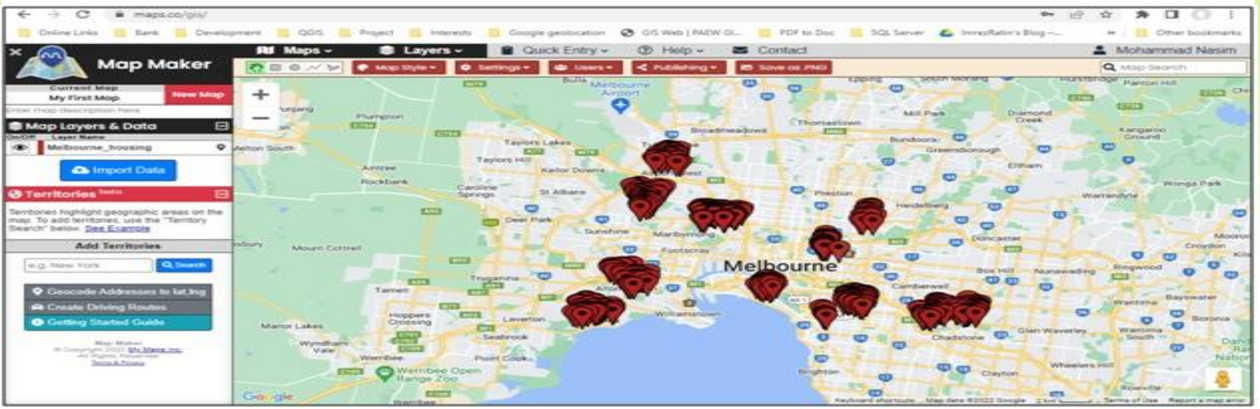
Statistics

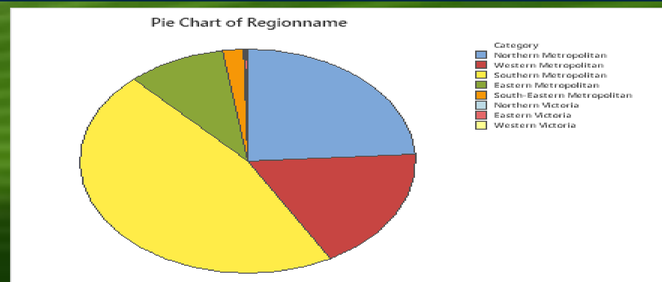
Variable	Total Count	Mean	SE Mean	StDev	Sum	Minimum	Q1	Median
Price	10952	1068580	6225	651431	11703089833	85000	635000	895000
Distance	10952	9.6978	0.0522	5.4594	106210.1000	0.0000	5.9000	9.2000
Bedroom2	10952	1.0000	0.000000	0.000000	10952.0000	1.0000	1.0000	1.0000
Bathroom	10952	1.5415	0.00662	0.6927	16882.0000	1.0000	1.0000	1.0000
YearBuilt	10952	1979.1	0.389	40.7	21674668.0	1196.0	1952.0	1985.0

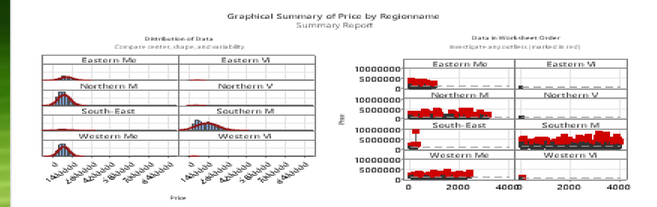
Variable	Q3	Maximum	IQR	Skewness
Price	1322000	9000000	687000	2.31
Distance	12.4000	47.4000	6.5000	1.71
Bedroom2	1.0000	1.0000	0.000000	*
Bathroom	2.0000	8.0000	1.0000	1.42
YearBuilt	2022.0	2022.0	70.0	-1.29

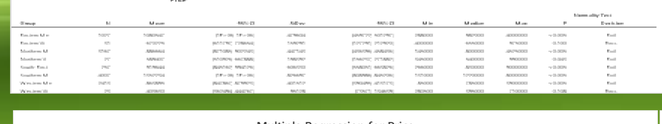
Model Summary

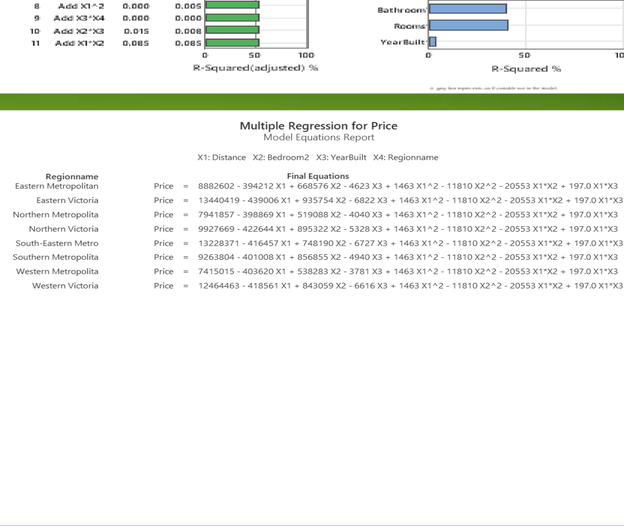
S	R-sq	R-sq(adj)	R-sq(pred)
456901	50.86%	50.81%	50.68%











Conclusion: Based on statistical evaluation ,southern Metropolitan region have highest pricing in terms of number of rooms with integrated washroom as well as distance and age of the building have less significant in Melbourne, Australia.

Reference :

https://www.hindawi.com/journals/acisc/2022/1562942/
https://www.researchgate.net/publication/348604535_Empirical_analysis_of_regression_techniques_by_house_price_and_salary_prediction
https://ieeexplore.ieee.org/document/8473231