Ishdeep Kaur
EPPS 6356
Project Presentation

REAL ESTATE

" Analysis of Real Estate Prices in Melbourne and the factors affecting thereof "

OBJECTIVE

In this research paper, the objective is to study the various factors on which the real estate housing prices depend; the importance of location, the number of bedrooms and bathrooms in the property, distance from the city and number of car parking.

METHODOLOGY

- The study is based on analyzing the Melbourne Housing Market for the period 2016 to 2019.
- Insights on the real estate market are developed from the data using Python.
- Graphical representation of the data has been done to visualize the outcomes from the study.

PACKAGES USED

- seaborn as sns
- matplotlib.pyplot as plt
- statsmodels.formula.api as smf
- sklearn.linear_model import LinearRegression

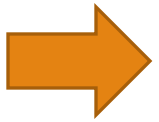
OUTLINE

- Explore the contents of the data and do data cleaning and modification.
- Check and replace all the missing values in the data for better results. It includes treating the outliers.
- Data presentation and Melbourne price trend over the years.
- Predict house prices in Melbourne region for 2019, 2020 and 2021.
- Seasonal performances of housing market.
- Relationship of housing factors affecting the real estate prices over the years.

DATA CLEANING

- Change all objects to category
- Change date to datetime
- Change postcode to category

#	Column	Non-Null Count	Dtype
0	Suburb	34857 non-null	object
1	Address	34857 non-null	object
2	Rooms	34857 non-null	int64
3	Type	34857 non-null	object
4	Price	27247 non-null	float64
5	Method	34857 non-null	object
6	SellerG	34857 non-null	object
7	Date	34857 non-null	object
8	Distance	34856 non-null	float64
9	Postcode	34856 non-null	float64
10	Bedroom2	26640 non-null	float64
11	Bathroom	26631 non-null	float64
12	Car	26129 non-null	float64
13	Landsize	23047 non-null	float64
14	BuildingArea	13742 non-null	float64
15	YearBuilt	15551 non-null	float64
16	CouncilArea	34854 non-null	object
17	Lattitude	26881 non-null	float64
18	Longtitude	26881 non-null	float64
19	Regionname	34854 non-null	object
20	Propertycount	34854 non-null	float64



Dtype
category
category
int64
category
float64
category
category
category
float64
category
float64
float64
float64
float64
float64
float64
category
float64
float64
category
float64

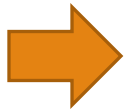
Table 1 : Original Data Type

Table 2 : Cleaned Data Type

DATA MODIFICATION

- Remove 'Bedroom2' because 'Rooms' and 'Bedroom2' are the same.
- Since we will not be using landsize, building area and year built we drop those variables

	count
Rooms	34857.0
Price	27247.0
Distance	34856.0
Bedroom2	26640.0
Bathroom	26631.0
Car	26129.0
Landsize	23047.0
BuildingArea	13742.0
YearBuilt	15551.0
Lattitude	26881.0
Longtitude	26881.0
Propertycount	34854.0



	count
Rooms	34857.0
Price	27247.0
Distance	34856.0
Bathroom	26631.0
Car	26129.0
Lattitude	26881.0
Longtitude	26881.0
Propertycount	34854.0

Table 3 : Original Variables

Table 4: Modified Variables

MISSING VALUES

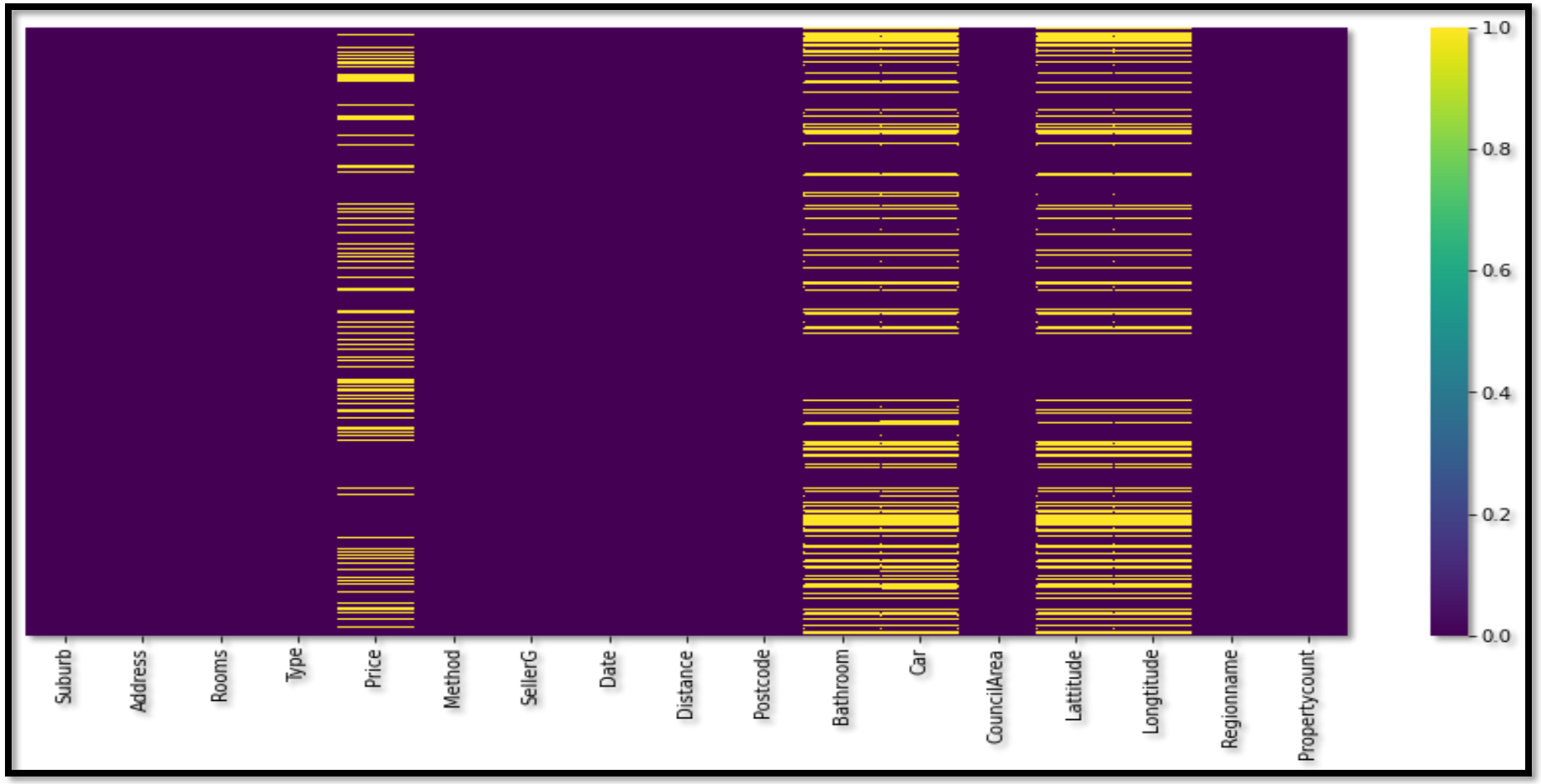


Figure 1 : Heatmap for visualising missing values

MISSING VALUES

Suburb	0.000000
Address	0.000000
Rooms	0.000000
Type	0.000000
Price	21.832057
Method	0.000000
SellerG	0.000000
Date	0.000000
Distance	0.002869
Postcode	0.002869
Bathroom	23.599277
Car	25.039447
CouncilArea	0.008607
Lattitude	22.882061
Longtitude	22.882061
Regionname	0.008607
Propertycount	0.008607

dtype: float64

Table 5 : Percentage of Missing Values

- Price has 21% missing values so we drop them.
- Car and bathroom have 23% and 25% missing values.
- Latitude and Longitude have 22% missing values.
- Replace them by the mode of respective variables

Suburb	0.00000
Address	0.00000
Rooms	0.00000
Type	0.00000
Price	0.00000
Method	0.00000
SellerG	0.00000
Date	0.00000
Distance	0.00367
Postcode	0.00367
Bathroom	0.00000
Car	0.00000
CouncilArea	0.01101
Lattitude	0.00000
Longtitude	0.00000
Regionname	0.01101
Propertycount	0.01101

dtype: float64

```
melbourne_data.shape
```

```
(27247, 17)
```

Table 6 : Cleaned data without missing values

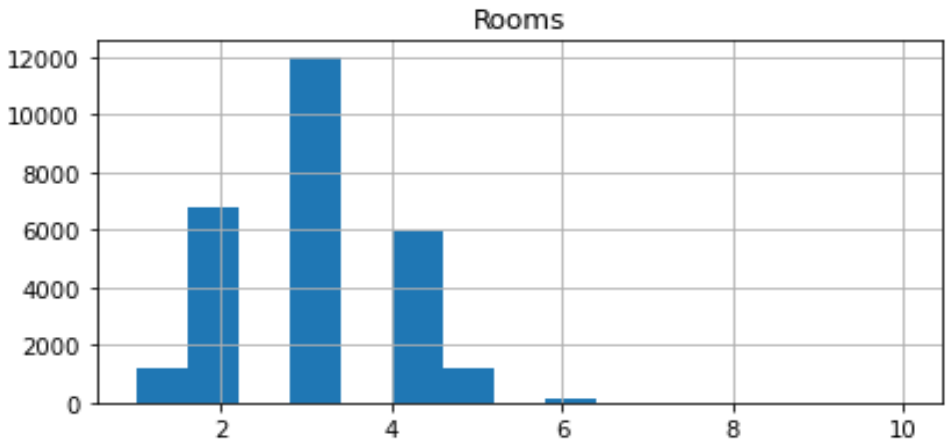
OUTLIERS

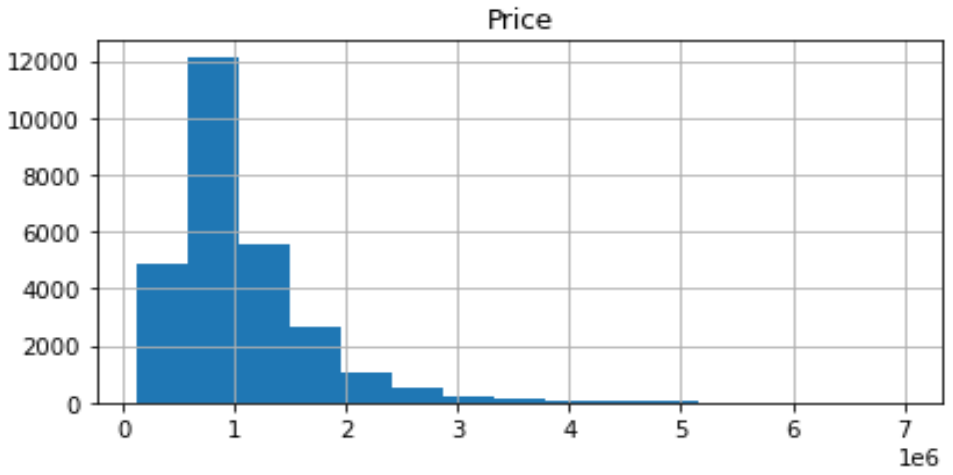
- Find outliers in our data to make sure that the final data is not skeewed or not symmetric.
- Divide prices into groups to find outliers and then drop the extreme values.

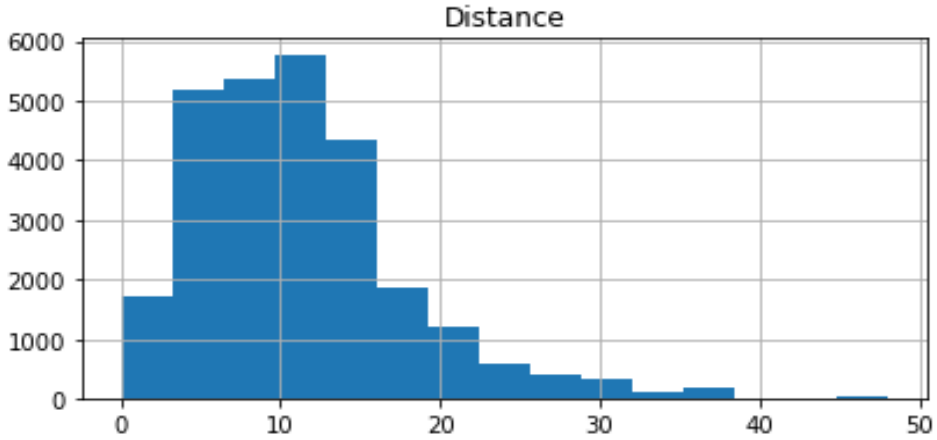
PriceRange	count
0-100,000	1
100,001 - 1M	16496
11M-12M	1
1M - 3M	10304
3M - 5M	388
5M - 6M	40
6M - 7M	14
7M-8M	2
8M-9M	1

Table 7: Outliers for Price Range

OUTLIERS







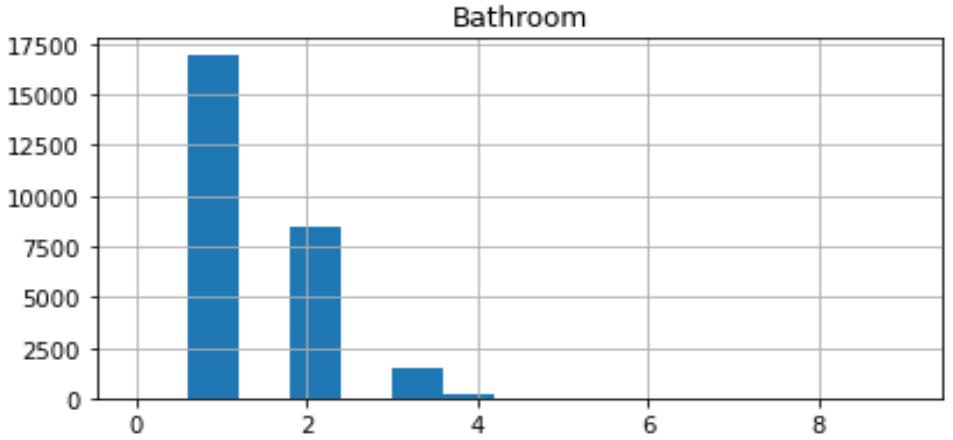


Figure 2: Skewness of Rooms, Price, Distance and Bathroom

OUTLIERS

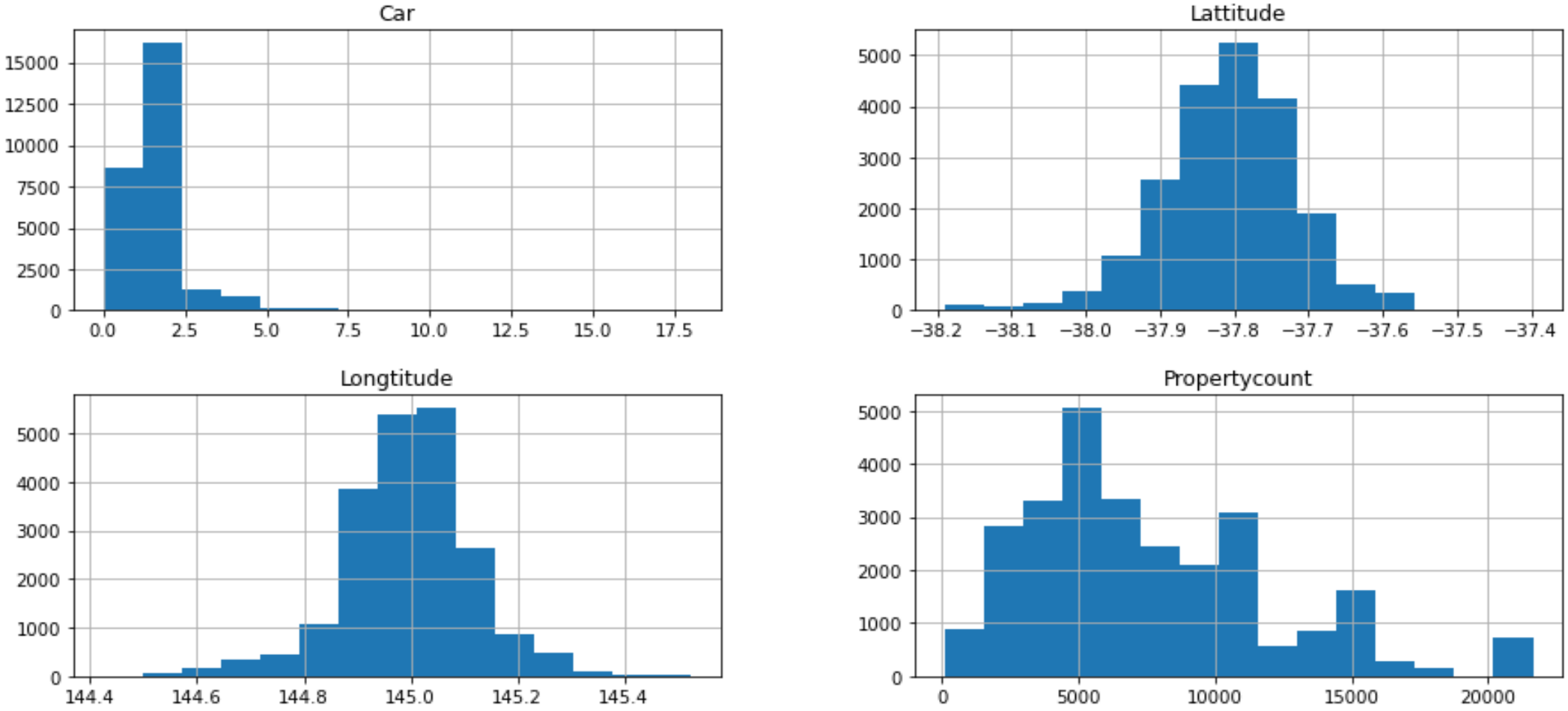


Figure 3: Skewness of Car, Latitude, Longitude and Property Count

DATA PRESENTATION

PRICE TREND VS YEAR

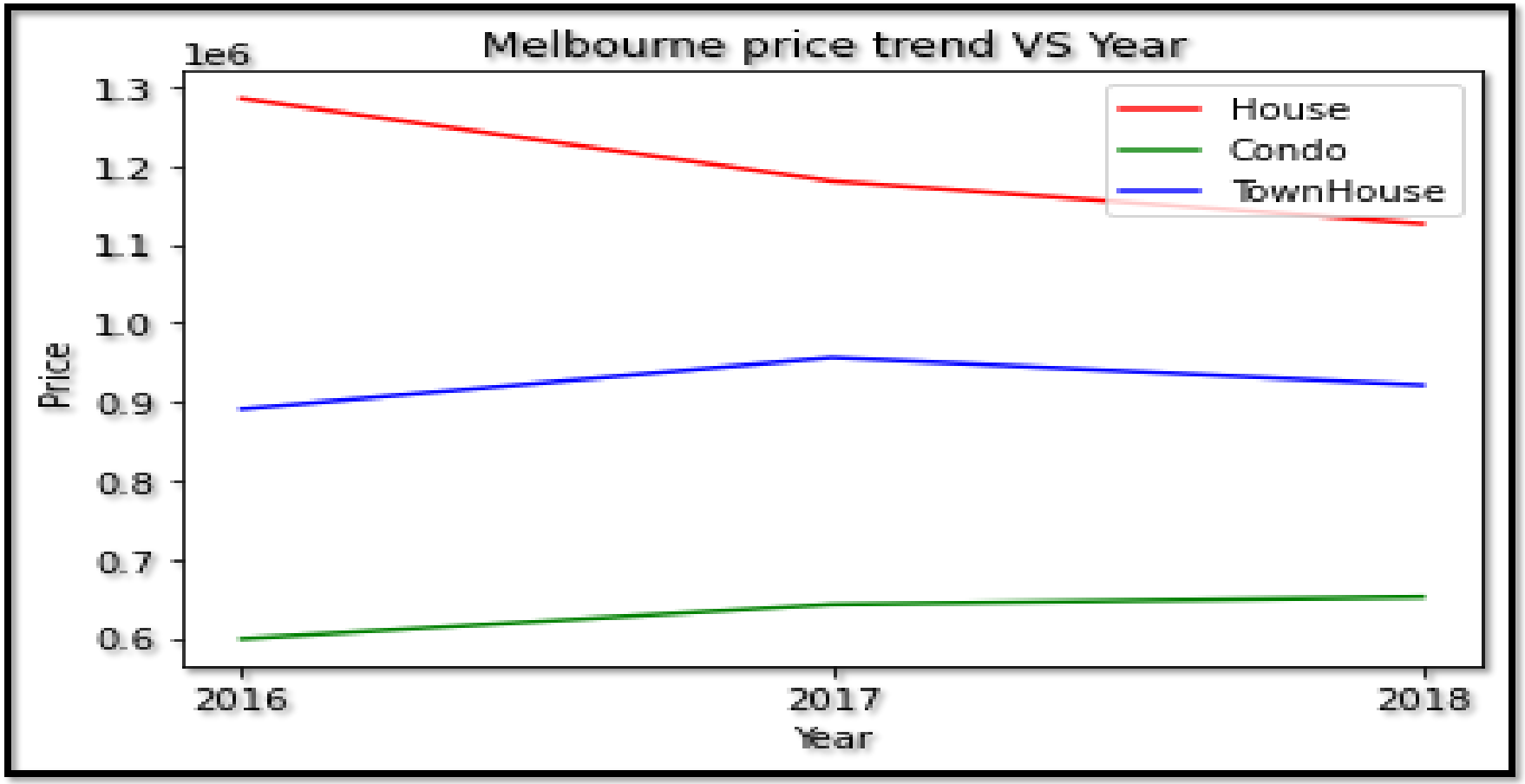


Figure 4: Price trend of houses in Melbourne over the years

PRICE PREDICTION FOR CONDOS

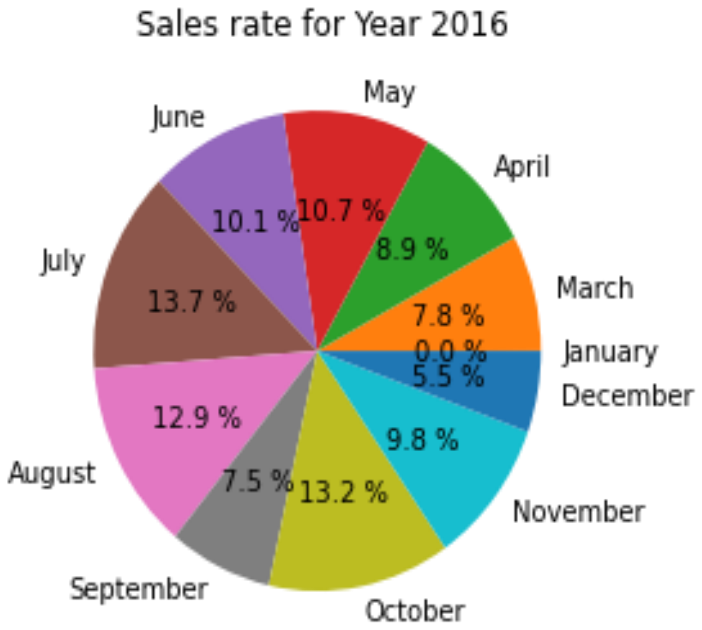
Year	Price Range
2016	659000.900688
2017	704424.661081
2018	749848.421474
2019	795272.181867
2020	840695.942260
2021	886119.702653

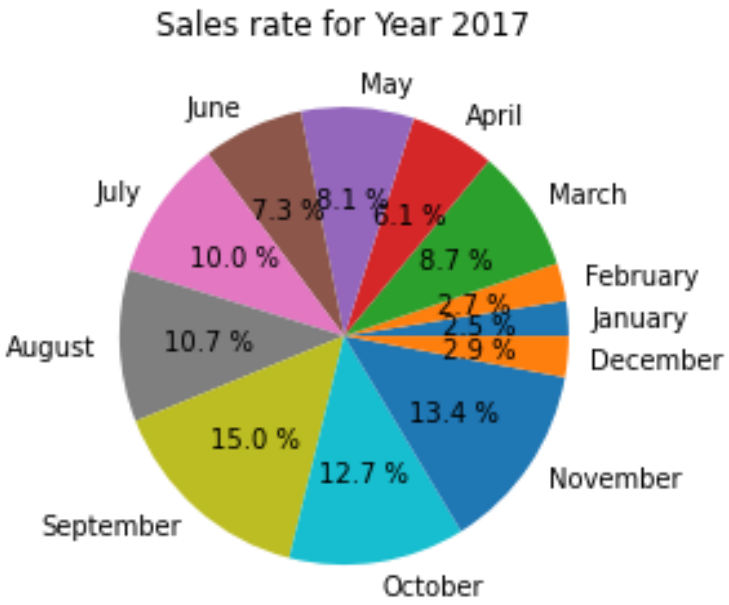
Table 8: South Metro Price Range for Condos

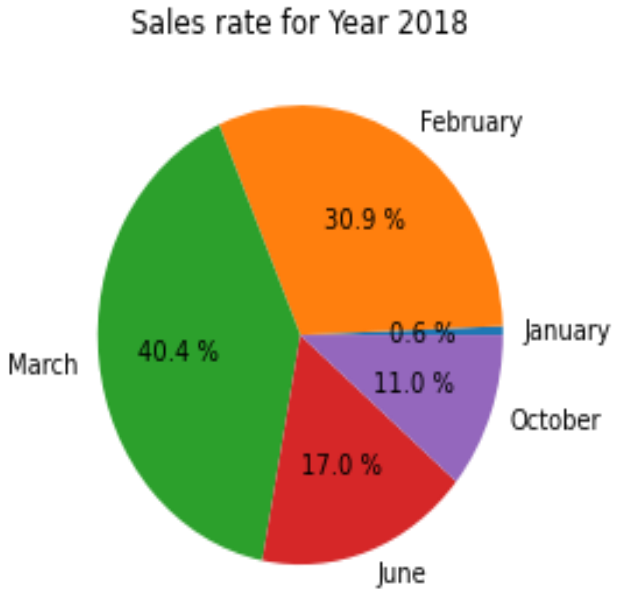
Year	Price Range
2016	644313.326193
2017	709498.532146
2018	774683.738098
2019	839868.944051
2020	905054.150004
2021	970239.355957

Table 9: East Metro Price Range for Condos

SEASONAL PERFORMANCE







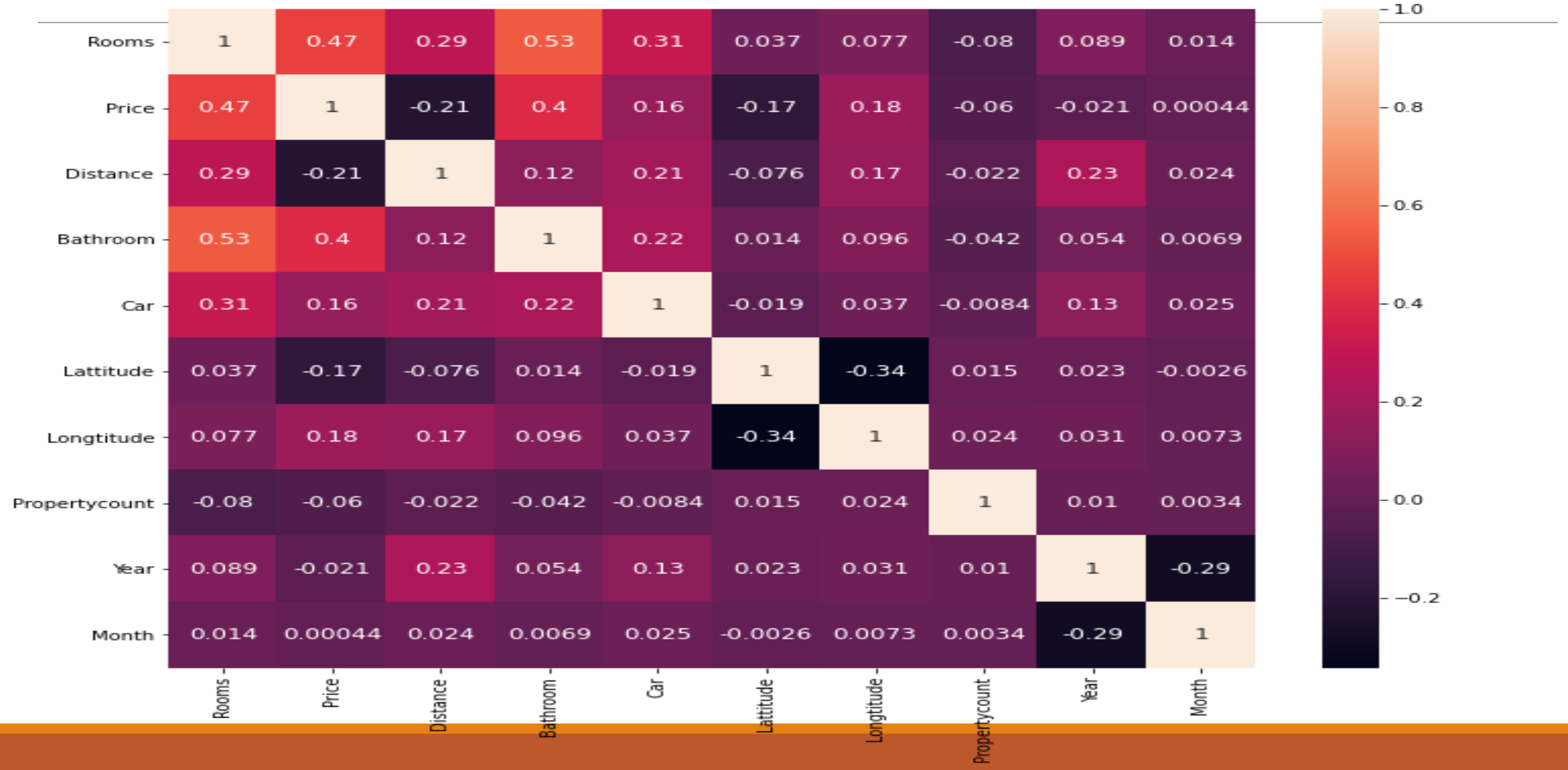
SALES COUNT FOR REGIONS

Regionname	Type	
Northern Metropolitan	h	5309
Southern Metropolitan	h	4718
Western Metropolitan	h	4435
Southern Metropolitan	u	2782
Eastern Metropolitan	h	2551
Northern Metropolitan	u	1689
South-Eastern Metropolitan	h	1036
Southern Metropolitan	t	1020
Northern Metropolitan	t	866
Western Metropolitan	u	810

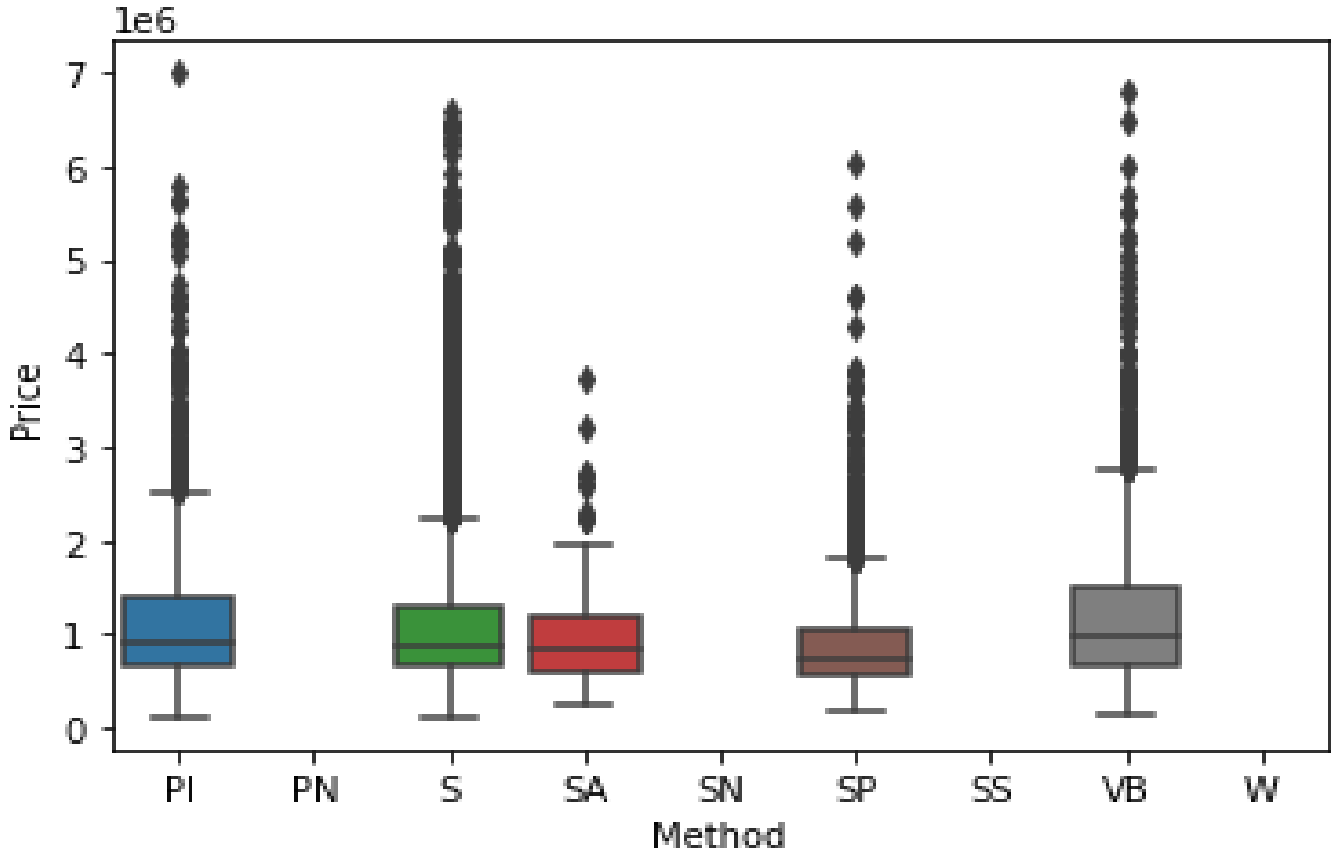
Name: Price, dtype: int64

Table 10: Top 10 regions with the most sold houses over the years

HEATMAP FOR CORELATION



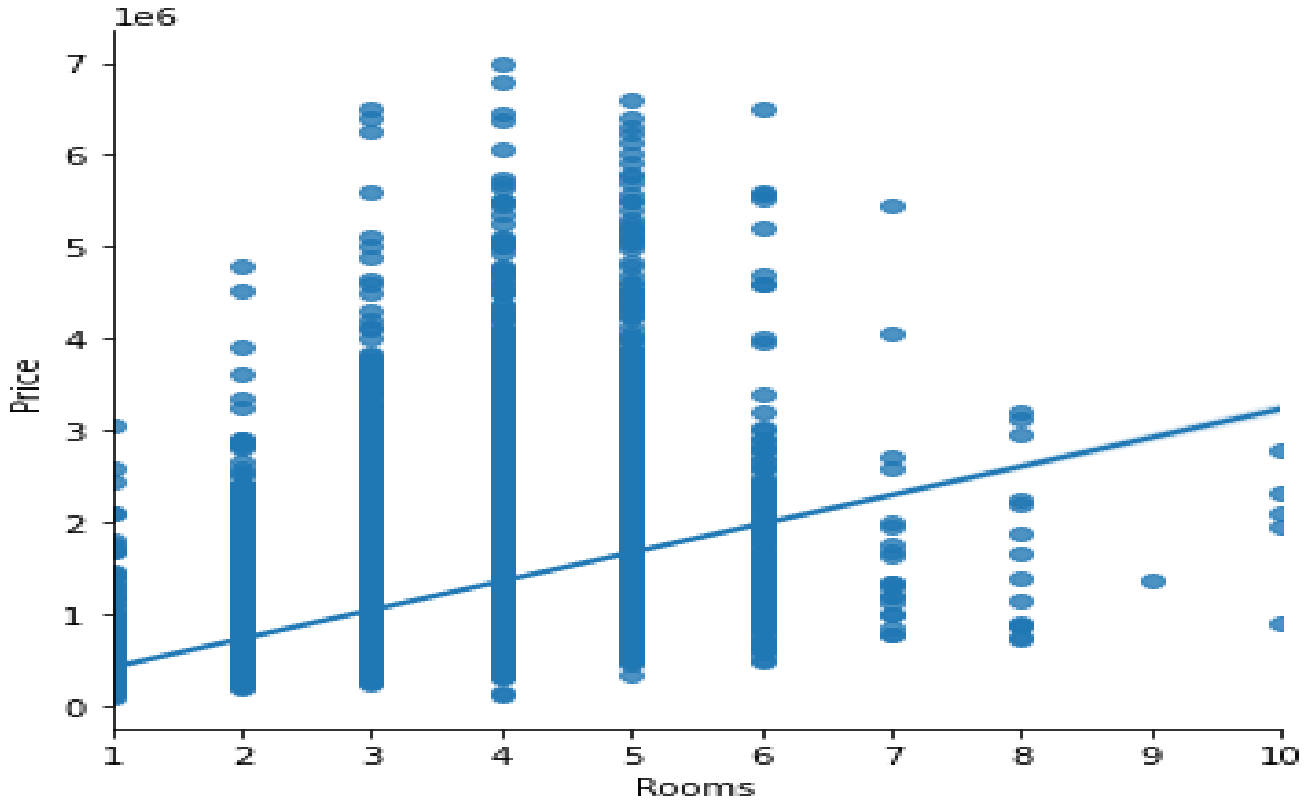
METHOD VS PRICE



- Relationship between sales method and prices of houses.
- Graph shows that all methods have the same price range.
- It shows that Method of Sales does not determine Price Range

Figure 5: Relationship of Sales method with Prices

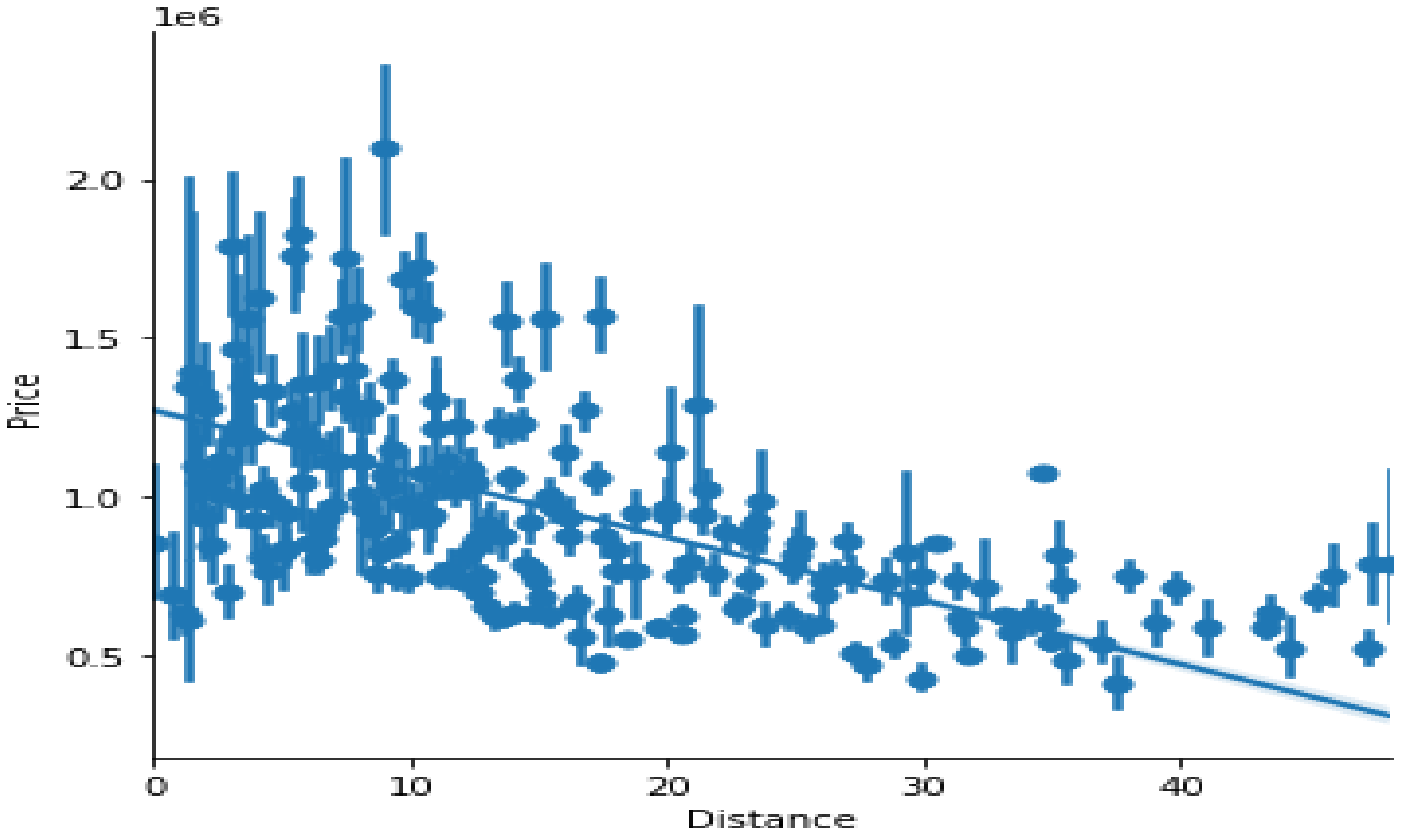
NUMBER OF ROOMS VS PRICE



- Relationship between number of rooms and house price
- Positive relationship between rooms and prices
- As number of rooms increases the house prices also increases
- With 10 rooms, prices are ower because of less demand

Figure 6: Relationship of Rooms with Prices

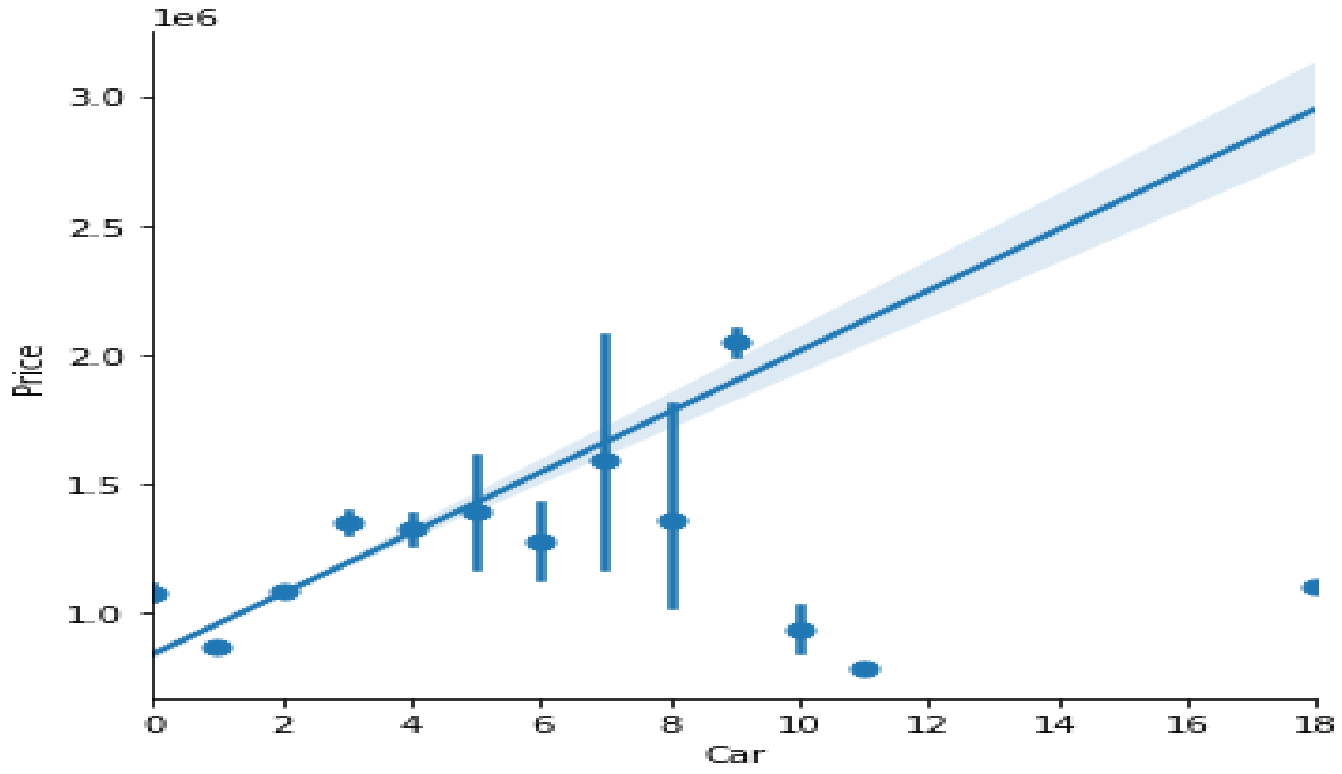
DISTANCE VS PRICE



- Relationship of distance of the house from the city with the price of the house
- Negative relationship
- As distance increases the price of the house decreases significantly.

Figure 7: Relationship of distance from the city to the price of the house

CAR PARK VS PRICE



- Relationship of number of cars in a household with the price of the house
- Highly positive corelation
- As the number of car parkings increase the prices increase

Figure 8: Relationship between car parking and prices

IDEAL HOUSE TYPE

- Ideal house includes major details of type of house, number of rooms and bathrooms and price.
- Region wise distribution of type of houses with rooms and bathrooms
- Houses with 3 Rooms and 1 Bathroom in North metropolitan seems to be an ideal purchase.

Regionname	Type	Rooms	Bathroom	
Northern Metropolitan	h	3	1.0	2034
Western Metropolitan	h	3	1.0	1659
Southern Metropolitan	u	2	1.0	1608
	h	3	1.0	1153
Northern Metropolitan	u	2	1.0	931
	h	2	1.0	912
Southern Metropolitan	h	3	2.0	817
		4	2.0	785
Northern Metropolitan	h	3	2.0	779
Western Metropolitan	h	3	2.0	769

Name: Price, dtype: int64


THANK YOU !