

FOUNDATION OF DATA SCIENCE

Dataset: MELBOURNE HOUSING PRICE



ANJALI CHAUHAN
Roll no: 3154
FOUNDATION OF DATA SCIENCE
2022 - 2023
Prof: ANKIT GANESH

MELBOURNE HOUSING PRICE

Melbourne, city, capital of the state of Victoria, Australia. The central city is home to about 136,000 people and is the core of an extensive metropolitan area—the world's most southerly with a population of more than 1,000,000.

The dataset has data on Melbourne housing prices from the year 2016 - 2018.

Dataset contains columns: **Suburb, Address, Rooms, Type, Price, Sales_Method, Agent, Date_Sold, Distance, Postcode, No_of_Bedroom, Bathroom, Council_Area, Region, Property_Count**.

Feature Description:

- **Suburb:** Suburb
- **Address:** Address
- **Rooms:** Number of rooms
- **Type:** type of real estate including : br - bedroom(s); h - house,cottage,villa, semi,terrace; u - unit, duplex; t - townhouse; dev site - development site; o res - other residential.
- **Price:** Price in Australian dollars
- **Sales_Method:** a method for selling real estate including S - property sold; SP - property sold prior; PI - property passed in; PN - sold prior not disclosed; SN - sold not disclosed; NB - no bid; VB - vendor bid; W - withdrawn prior to auction; SA - sold after auction; SS - sold after auction price not disclosed. N/A - price or highest bid not available.
- **Agent:** Real Estate Agent:
- **Date_Sold:** Date Sold
- **Postcode:** Postcode
- **Region:** General Region (West, North West, North, Northeast ...etc)
- **Property_Count:** Number of properties that exist in the suburb
- **Distance:** Distance from Central Business Development (in km)
- **Council_Area** Governing council for the area

IMPORTING DATASET

#Import Dataset

```
getwd()
setwd("C:/Users/ADMIN/Downloads")
melbourne<-read.csv("C:/Users/ADMIN/Downloads/Melbourne_housing.csv")
```

#Packages & libraries

```
install.packages('tidyverse')
install.packages('janitor')
install.packages("dplyr")
install.packages("ggalt")
install.packages('GGally')
install.packages('ggridges')
install.packages('ggcorrplot')
install.packages('corrplot')
install.packages('treemapify')
install.packages("mosaic")
install.packages('visreg')
```

#libraries Used

```
library("RColorBrewer")
library(tidyverse)
library(treemapify)
library(dplyr)
library(plotly)
library(ggplot2)
library(ggalt)
library(GGally)
library(ggridges)
library(ggcorrplot)
library(corrplot)
library(lubridate)
library(visreg)
library(treemapify)
```

```
mycolors <- c( "#AEBDCA", "#DFF6FF", "#B4CDE6", "#98A8F8", "#5F9DF7",
              "#CDFCF6", "#2146C7", "#7DE5ED")
```

DATA CLEANING

#checking for null values

```
colSums(is.na(melbourne))
```

```
str(melbourne)
```

```
> #checking for null values
```

```
> colSums(is.na(melbourne))
```

Suburb	Address	Rooms	Type	Price	Method
0	0	0	0	7610	0
sellerG	Date	Distance	Postcode	Bedroom2	Bathroom
0	0	0	0	8217	8226
Car	Landsize	BuildingArea	YearBuilt	CouncilArea	Latitude
8728	11810	21115	19306	0	7976
Longitude	Regionname	Propertycount			
7976	0	0			

```
> str(melbourne)
```

```
'data.frame': 34857 obs. of 21 variables:
 $ Suburb      : chr "Abbotsford" "Abbotsford" "Abbotsford" "Abbotsford" ...
 $ Address     : chr "68 studley st" "85 Turner St" "25 Bloomburg St" "18/659 victori
a St" ...
 $ Rooms       : int 2 2 2 3 3 3 4 4 2 2 ...
 $ Type        : chr "h" "h" "h" "u" ...
 $ Price       : int NA 1480000 1035000 NA 1465000 850000 1600000 NA NA NA ...
 $ Method      : chr "SS" "S" "S" "VB" ...
 $ sellerG     : chr "Jellis" "Biggin" "Biggin" "Rounds" ...
 $ Date        : chr "3/09/2016" "3/12/2016" "4/02/2016" "4/02/2016" ...
 $ Distance    : chr "2.5" "2.5" "2.5" "2.5" ...
 $ Postcode    : chr "3067" "3067" "3067" "3067" ...
 $ Bedroom2    : int 2 2 2 3 3 3 3 4 3 ...
 $ Bathroom    : int 1 1 1 2 2 2 1 2 1 2 ...
 $ Car         : int 1 1 0 1 0 1 2 2 2 1 ...
 $ Landsize    : int 126 202 156 0 134 94 120 400 201 202 ...
 $ BuildingArea : num NA NA 79 NA 150 NA 142 220 NA NA ...
 $ YearBuilt   : int NA NA 1900 NA 1900 NA 2014 2006 1900 1900 ...
 $ CouncilArea : chr "Yarra City Council" "Yarra City Council" "Yarra City Council"
"Yarra City Council" ...
 $ Latitude    : num -37.8 -37.8 -37.8 -37.8 -37.8 ...
 $ Longitude   : num 145 145 145 145 145 ...
 $ Regionname  : chr "Northern Metropolitan" "Northern Metropolitan" "Northern Metropol
olitan" "Northern Metropolitan" ...
 $ Propertycount: chr "4019" "4019" "4019" "4019" ...
```

#removing unwanted columns

```
house <- subset( melbourne, select = -c(Car,Latitude, YearBuilt, BuildingArea,Longitude , Landsize,
Longitude) )
```

```
dim(house)
```

```
house2 <- subset( melbourne, select = c(Rooms, Distance, Price, Bedroom2, Bathroom, Car, Landsize,
Latitude, Longitude, YearBuilt ) )
```

```
dim(house2)
```

#Replace missing value Price with NA

```
house$Price[is.na(house$Price)] <- NA
```

#replacing empty values with NA

```
house[house==""]<-NA
```

#removing rows with NA values

```
data <- na.omit(house)
colSums(is.na(data))
dim(data) #checking if removed
> colSums(is.na(data))
      Suburb      Address      Rooms      Type      Price      Method
      0         0         0         0         0         0
  SellerG      Date      Distance      Postcode      Bedroom2      Bathroom
      0         0         0         0         0         0
  CouncilArea      Regionname      Propertycount
      0         0         0
> dim(data) #checking if removed
[1] 20800 15
```

#renaming columns name

```
names(data)
names(data) <- c("Suburb","Address","Rooms","Type","Price","Sales_Method",
  "Agent","Date_Sold","Distance","Postcode","No_of_Bedroom",
  "Bathroom","Council_Area","Region","Property_Count")
names(data) #Checking for changes made
```

```
> names(data)
 [1] "Suburb"      "Address"      "Rooms"      "Type"
 [5] "Price"       "Sales_Method" "Agent"      "Date_sold"
 [9] "Distance"    "Postcode"    "No_of_Bedroom" "Bathroom"
[13] "Council_Area" "Region"      "Property_Count"
```

#Changing datatypes of columns

```
data$Rooms <- as.factor(data$Rooms)
data$Type <- as.factor(data$Type)
data$Sales_Method <- as.factor(data$Sales_Method)
data$Region <- as.factor(data$Region)
data$Property_Count <- as.integer(data$Property_Count)
data$Postcode <- as.integer(data$Postcode)
data$Distance <- as.numeric(data$Distance)
data$Date_Sold <- as.Date(data$Date_Sold, format = "%d/%m/%Y")
```

#renaming level of factor

```
levels(data$Type) <- list(House = "h", Townhouse = "t", Unit = "u")
levels(data$Sales_Method) <- list(`property sold` = "S",
  `property priorly sold` = "SP",
  `passed in` = "PI",
  `sold prior not disclosed` = "PN",
  `sold not disclosed` = "SN",
  `no bid` = "NB",
  `vendor bid` = "VB",
  `withdrawn prior to auction` = "W",
  `sold after auction` = "SA",
  `sold after auction price not disclosed` = "SS")
```

str(data) #Checking for changes

```
> str(data) #Checking for changes
'data.frame': 20800 obs. of 15 variables:
 $ Suburb : chr "Abbotsford" "Abbotsford" "Abbotsford" "Abbotsford" ...
 $ Address : chr "85 Turner St" "25 Bloomberg St" "5 Charles St" "40 Federation La" ...
 $ Rooms : Factor w/ 11 levels "1","2","3","4",...: 2 2 3 3 4 2 3 2 1 2 ...
 $ Type : Factor w/ 3 levels "House","Townhouse",...: 1 1 1 1 1 1 1 3 1 ...
 $ Price : int 1480000 1035000 1465000 850000 1600000 941000 1876000 1636000 300000 1097000 ...
 $ Sales_Method : Factor w/ 10 levels "property sold",...: 1 1 2 3 7 1 1 1 1 1 ...
 $ Agent : chr "Biggin" "Biggin" "Biggin" "Biggin" ...
 $ Date_Sold : Date, format: "2016-12-03" "2016-02-04" ...
 $ Distance : num 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 ...
 $ Postcode : int 3067 3067 3067 3067 3067 3067 3067 3067 3067 ...
 $ No_of_Bedroom : int 2 2 3 3 3 2 4 2 1 3 ...
 $ Bathroom : int 1 1 2 2 1 1 2 1 1 1 ...
 $ Council_Area : chr "Yarra City Council" "Yarra City Council" "Yarra City Council" ...
 $ Region : Factor w/ 8 levels "Eastern Metropolitan",...: 3 3 3 3 3 3 3 3 3 ...
 $ Property_Count: int 4019 4019 4019 4019 4019 4019 4019 4019 4019 ...
 - attr(*, "na.action")= 'omit' Named int [1:14057] 1 4 8 9 10 13 14 16 17 20 ...
 .. attr(*, "names")= chr [1:14057] "1" "4" "8" "9" ...
```

summary(data)

```
> summary(data)
Suburb      Address      Rooms      Type      Price
Length:20800 Length:20800 3 :9392 House :15728 Min. : 85000
Class :character Class :character 4 :4857 Townhouse: 1579 1st Qu.: 660000
Mode :character 2 :4608 Unit : 3493 Median : 910000
                    5 :1021 Mean : 1089971
                    1 : 770 3rd Qu.: 1335000
                    6 : 117 Max. :11200000
                    (other): 35

Sales_Method Agent      Date_Sold      Distance      Postcode
property sold :13578 Length:20800 Min. :2016-01-28 Min. : 0.00 Min. :3000
property priorly sold : 2650 Class :character 1st Qu.:2016-11-19 1st Qu.: 6.40 1st Qu.:3046
passed in : 2408 Mode :character Median :2017-07-01 Median :10.40 Median :3087
vendor bid : 2017 Mean :2017-05-15 Mean :11.38 Mean :3114
sold after auction : 147 3rd Qu.:2017-10-14 3rd Qu.:14.20 3rd Qu.:3152
sold prior not disclosed: 0 Max. :2018-03-17 Max. :48.10 Max. :3978
(other) : 0

No_of_Bedroom Bathroom Council_Area Region Property_Count
Min. : 0.000 Min. :0.000 Length:20800 Southern Metropolitan :6319 Min. : 83
1st Qu.: 2.000 1st Qu.:1.000 Class :character Northern Metropolitan :5949 1st Qu.: 4380
Median : 3.000 Median :1.000 Mode :character Western Metropolitan :4599 Median : 6567
Mean : 3.046 Mean :1.592 Eastern Metropolitan :2543 Mean : 7502
3rd Qu.: 4.000 3rd Qu.:2.000 South-Eastern Metropolitan:1023 3rd Qu.:10331
Max. :20.000 Max. :9.000 Eastern Victoria : 140 Max. :21650
                    (other) : 227
```

1) The Maximum number of properties has 3 rooms and very few properties have 6 rooms.

2) Most type of property in Melbourne is of type of house, the least are townhouses.

3) Least expensive property is \$85000 and the most expensive property is \$11200000.

4) Most selling methods were directly agent to buyers.

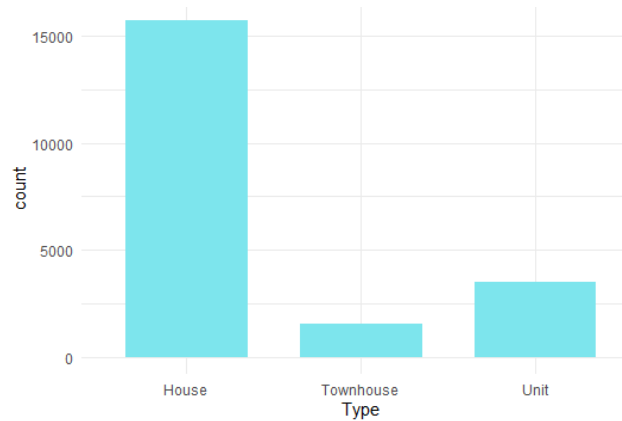
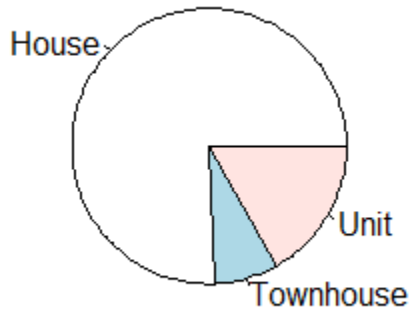
5) The first property was sold on 28/1/2016.

6) Most properties are in Southern Metropolitan Region.

UNIVARIATE ANALYSIS

#Proportions type of property

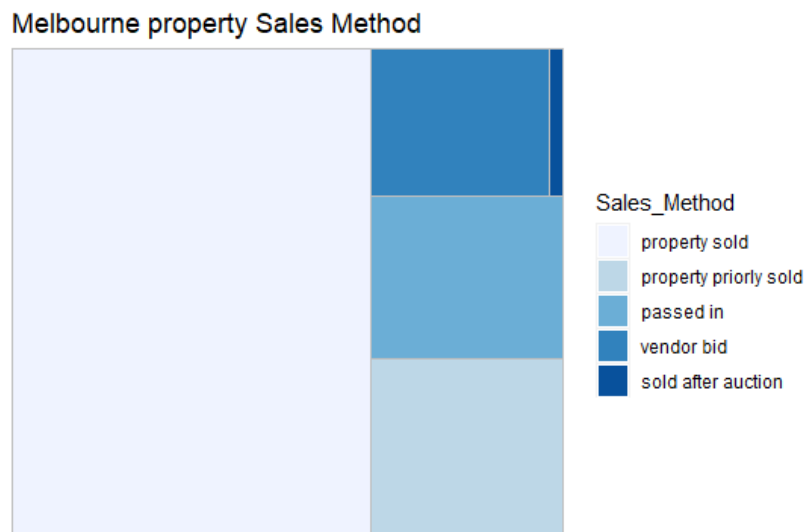
```
pie(xtabs(~Type, data))
```



In Melbourne type of property most of them are houses followed by units/ apartments then Townhouses.

#treemap spread of Sales Method

```
plotdata <- data %>%
  count(Sales_Method)
ggplot(plotdata,
  aes(fill = Sales_Method,
    area = n)) +
  geom_treemap() +
  scale_fill_brewer(palette = "Blues")+
  labs(title = "Melbourne property Sales Method")
```



1) The majority of property are sold directly to buyers by agent (direct deal)

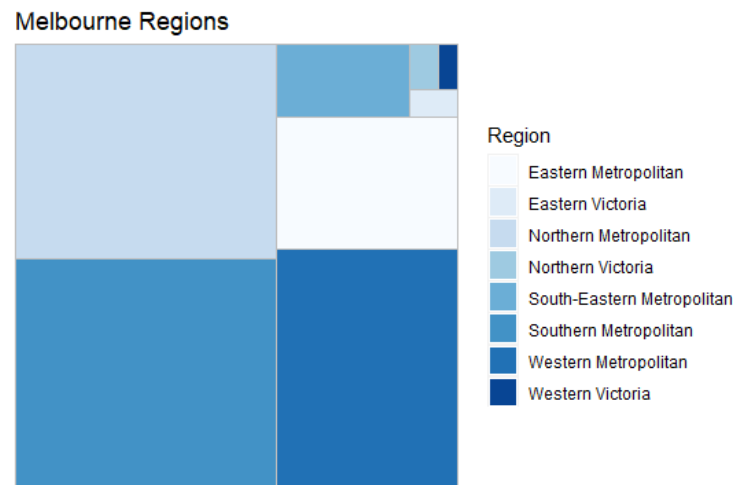
2) Least selling method to buyers is after auction.

#treemap of Regions

```

plotdata <- data %>%
  count(Region)
ggplot(plotdata,
  aes(fill = Region,
    area = n)) +
  geom_treemap() +
  scale_fill_brewer(palette = "Blues")+
  labs(title = "Melbourne Regions")

```



1) Most properties are in southern and Northern metropolitan regions.

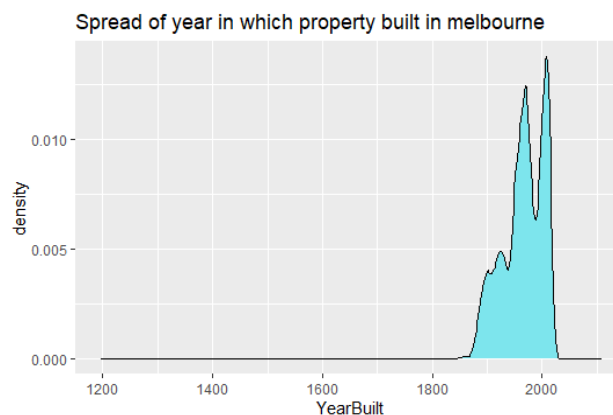
2) Western Victoria region is maybe at a longer distance hence less property out there.

#kernel density plot of year built

```

ggplot(house2, aes(x = YearBuilt)) +
  geom_density(fill = "#7DE5ED") +
  labs(title = "Spread of year in which property built in melbourne")

```

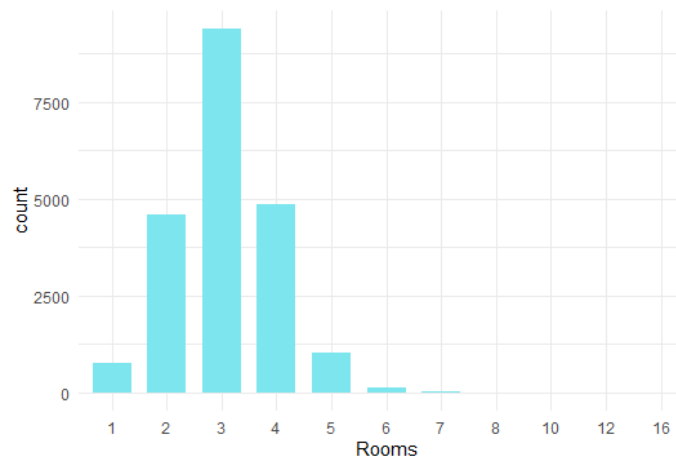


1) Properties listed were started to build in early 1900's.

2) There's been a spike in the construction of properties from the late 1900's to the early 2000's. Looks like people in Australia has some financial gain at that period of time.

#Spread of Rooms


```
ggplot(data, aes(x=Rooms))+  
  geom_bar(stat="count", width=0.7, fill="#7DE5ED")+  
  theme_minimal()
```

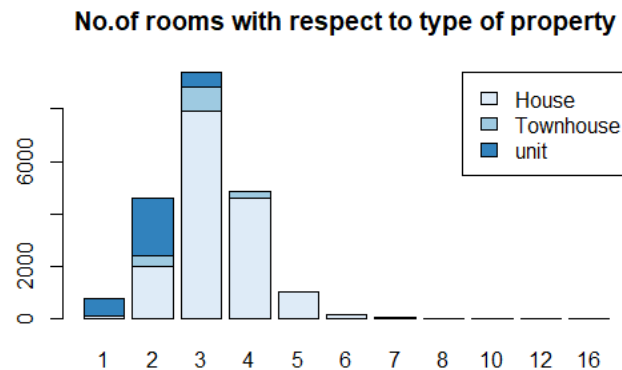


- 1) The bulk properties in Melbourne commonly have 3 rooms.
- 2) Very few properties have 7 bedrooms.

BIVARIATE ANALYSIS

#Spread of no.of Rooms with respect to type of property

```
table(data$Type, data$Rooms)
barplot(table(data$Type, data$Rooms),
        main = "No.of rooms with respect to type of property",
        col=brewer.pal(n = 3, name = 'Blues'))
legend("topright",
       c("House", "Townhouse", "unit"),
       fill=brewer.pal(n = 3, name = 'Blues'))
```



1) Property with 3 rooms are mostly houses.

2) Melbourne townhouses are mainly with 3 rooms.

3) Apartments in Melbourne are ordinarily with 2 rooms.

#Which type of real estate has highest Price based on the numbers of

```
Price_high <- aggregate(Price~Type+Rooms, data, FUN = max)
```

```
Price_high[order(Price_high$Price, decreasing = T),][1,]
```

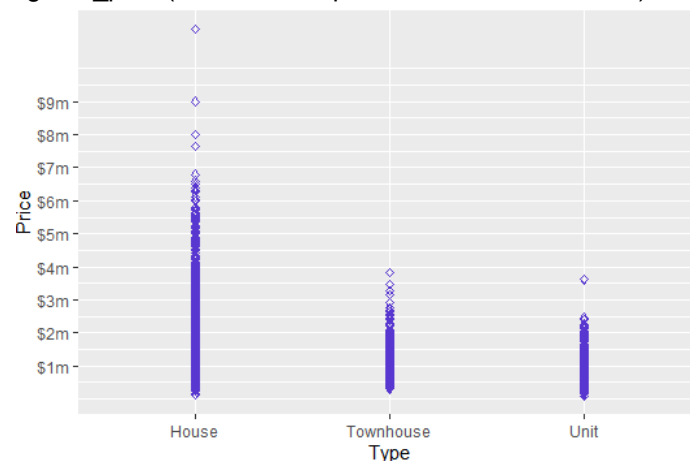
```
ggplot(data, aes(x=Type, y=Price)) + geom_point()
```

```
ggplot(data, aes(x=Type, y=Price)) +
```

```
  scale_y_continuous(breaks =
c(1000000,2000000,3000000,4000000,5000000,6000000,7000000,8000000,9000000),
```

```
  labels = c("$1m", "$2m", "$3m", "$4m", "$5m", "$6m", "$7m", "$8m", "$9m"))+
```

```
  geom_point(size=1.5, shape=23, colour="#5837D0")
```



Type	Rooms	Price
House	4	11200000

	House	Townhouse	Unit
Eastern Metropolitan	2161	158	224
Eastern Victoria	135	2	3
Northern Metropolitan	4549	449	951
Northern Victoria	139	0	1
South-Eastern Metropolitan	896	53	74
Southern Metropolitan	3976	595	1748
Western Metropolitan	3785	322	492

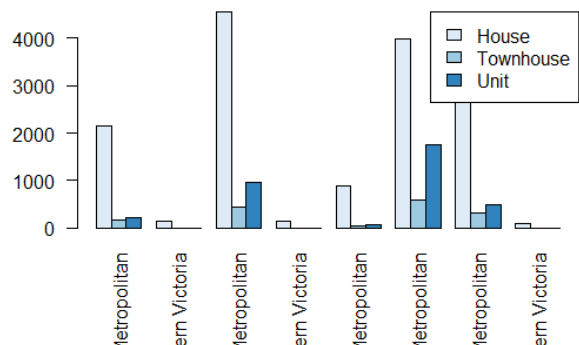
- 1) Melbourne houses are overly and mighty pricey than townhouses and apartments.
- 2) House cost around \$7 million at the most with some over \$7 million. If you need to own a house type of property in Melbourne you'll need a hefty amount of money
- 3) Apartments are less expensive to buy if looking for property in Melbourne.
- 4) Northern Metropolitan has the most house type of properties.
- 5) You can find more townhouses and apartments in Southern Metropolitan.

#Property Price of region wrt to its price

```
table(data$Region, data$Type) %>%head(7)
```

```
barplot(table(data$Type, data$Region),
  main = "Region and it's property price",
  las=2,srt=45,xpd=TRUE,
  beside = TRUE,
  col=brewer.pal(n = 3, name = 'Blues'))
legend("topright",
  c("House","Townhouse","Unit"),
  fill=brewer.pal(n = 3, name = 'Blues'))
```

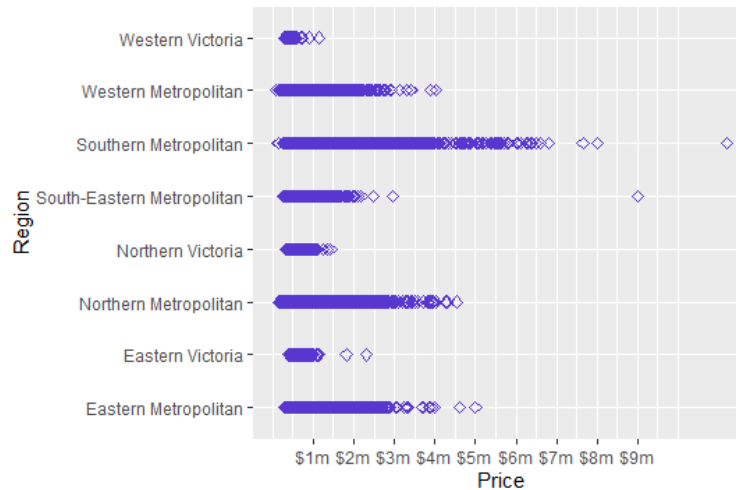
Region and it's property price



- 1) Northern Metropolitan has the most house type of properties.
- 2) You can find more townhouses and apartments in Southern Metropolitan.

#scatter plot region vs Price

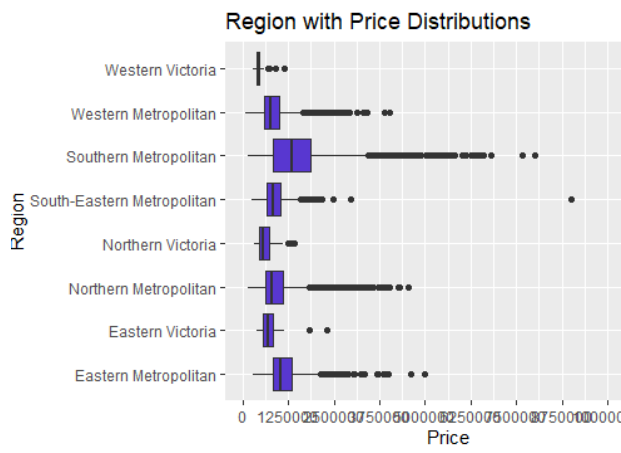
```
ggplot(data, aes(x=Price, y=Region)) + geom_point()
ggplot(data, aes(x=Price, y=Region)) +
  scale_x_continuous(breaks =
c(1000000,2000000,3000000,4000000,5000000,6000000,7000000,8000000,9000000),
  labels = c("$1m","$2m","$3m","$4m","$5m","$6m","$7m","$8m","$9m"))+
  geom_point(size=2, shape=23, colour="#5837D0")
```



- 1) Southern Metropolitan is an expensive neighborhood. That's where the riches are at,
- 2) It has properties from \$1 million to \$11 million.
- 3) Northern Metropolitan Region is second most expensive after Southern Metropolitan.
- 4) If looking for properties at low prices head towards Eastern and Western Victoria.

#Region with Price Distributions

```
ggplot(data, aes(Region, Price)) +
  geom_boxplot(fill="#5837D0") +
  scale_y_continuous(breaks =
c(1000000,2000000,3000000,4000000,5000000,6000000,7000000,8000000,9000000),
  labels = c("$1m", "$2m", "$3m", "$4m", "$5m", "$6m", "$7m", "$8m", "$9m"))+
  theme(legend.position = "none") +
  xlab("Region") +
  ylab("Price") +
  ggtitle("Region with Price Distributions") +
  coord_flip()
```



Co-relation of data

```
df <- dplyr::select_if(data, is.numeric)
corG <- cor(df, use="complete.obs")
ggcorrplot(corG)
```



- 1) There's high positive correlation between No. of bedrooms and price also no. of Bathroom and price; have they increase in numbers the fancy the price gets.
- 2) Bedrooms and bathrooms also have a positive correlation, obviously.
- 3) We see here some negative correlation between distance and price, which also makes sense, as the distance increases the price of property drops. The further it gets from the city and rate drops.

#Rate price of townhouse every suburb

```
townhouse <- data[data$Type == "Townhouse",]
townhouse_2 <- aggregate(Price~Type+Suburb, townhouse, mean)
townhouse_2[order(townhouse_2$Price, decreasing = T), ][1:3,]
#Rate price of house every suburb
house1 <- data[data$Type == "House",]
house_2 <- aggregate(Price~Type+Suburb, house1, mean)
house_2[order(house_2$Price, decreasing = T), ][1:3,]
#Rate price of unit every suburb
unit <- data[data$Type == "Unit",]
unit_2 <- aggregate(Price~Type+Suburb, unit, mean)
unit_2[order(unit_2$Price, decreasing = T), ][1:3,]
```

```
> townhouse_2[order(townhouse_2$Price, decreasing = T), ][1:3,]
  Type Suburb Price
188 Townhouse Toorak 1914167
32 Townhouse Brighton 1834338
27 Townhouse Bonbeach 1830000
> house_2[order(house_2$Price, decreasing = T), ][1:3,]
  Type Suburb Price
114 House East Melbourne 3295714
302 House Toorak 3068171
199 House Malvern 2786100
> unit_2[order(unit_2$Price, decreasing = T), ][1:3,]
  Type Suburb Price
33 Unit Brighton 943340.9
101 Unit Glen Waverley 941916.7
18 Unit Balwyn North 912192.3
```

- 1) Toorak Suburb has the most expensive townhouses in Melbourne.
- 2) East Melbourne has the most expensive houses in Melbourne.
- 3) Brighton has the most expensive Apartments/ Units in Melbourne.

#expensive Suburb with it's region wrt price

```
table(data$Price, data$Region)
region_high <- aggregate(Price~Region+Suburb, data, FUN = max)
region_high[order(region_high$Price, decreasing = T), ][1:5,]
```

	Region	Suburb	Price
43	Southern Metropolitan	Brighton	11200000
225	South-Eastern Metropolitan	Mulgrave	9000000
63	Southern Metropolitan	Canterbury	8000000
157	Southern Metropolitan	Hawthorn	7650000
12	Southern Metropolitan	Armada	6800000

- 1) In Southern Metropolitan Region Brighton Suburb have properties at high rate.
- 2) Mulgrave Suburb in southern Eastern Metropolitan is a little less expensive than Brighton.
- 3) Brighton, Mulgrave, Canterbury, Hawthorn, and Armadale are luxury Suburbs.

#Which suburb region has the most property count?

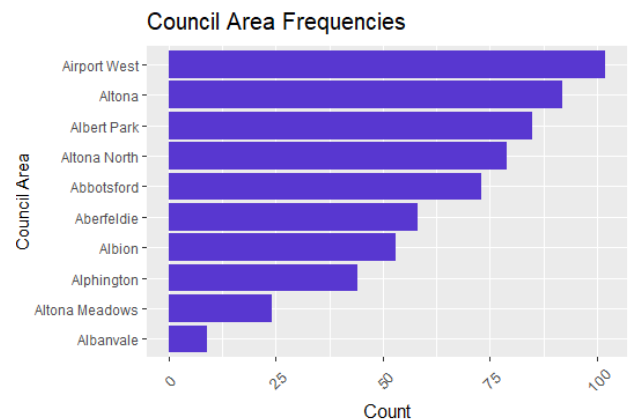
```
table(data$Property_Count, data$Region)
table(data$Suburb, data$Price) %>% head(5)
sub_high <- aggregate(Property_Count~Suburb+Region, data, sum)
sub_high[order(sub_high$Property_Count, decreasing = T),][1:5,]
```

	Suburb	Region	Property_Count
126	Reservoir	Northern Metropolitan	10630150
127	Richmond	Northern Metropolitan	4903272
124	Preston	Northern Metropolitan	4708371
208	Bentleigh East	Southern Metropolitan	3828181
254	South Yarra	Southern Metropolitan	3677089

- 1) Northern Metropolitan Region has the most properties in Melbourne.
- 2) In Melbourne property count more properties are in Reservoir Suburb in the Northern Metropolitan region, followed by Richmond and Preston Suburb.
- 3) Also Bentleigh East and South Yarra in Southern and Metropolitan after Northern Metropolitan.

#Council area frequency with suburb

```
data %>%
  group_by(Suburb) %>%
  summarise(Count = n()) %>%
  head(10) %>%
  ggplot(aes(reorder(Suburb, Count), Count)) +
  geom_bar(stat = 'identity', fill="#5837D0") +
  mytheme+
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5)) +
  coord_flip() +
  xlab("Council Area") +
  ylab("Count") +
  ggtitle("Council Area Frequencies")
```



#The highest of total price based suburb and type of real estate ?

```
total_price <- aggregate(Price~Suburb+Type, data, sum)
total_price[order(total_price$Price, decreasing = T),][1:3,]
```

	Suburb	Type	Price
23	Balwyn North	House	394842300
43	Brighton	House	389115888
181	Kew	House	384190000

- 1) We got that houses are the most expensive property type in Melbourne.
- 2) House in the Balwyn North suburb has a high price of almost \$400 Million.
- 3) Suburb like Brighton and Kew follows next.

#Which of council area is closest to Central Business Development?

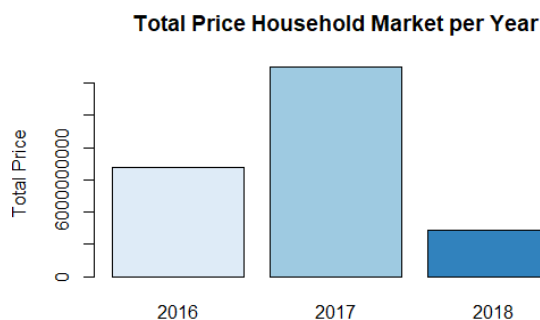
```
council <- aggregate(Distance~Council_Area, data, min)
council[order(council$Distance, decreasing = F),][1:10,]
```

	Council_Area	Distance
19	Melbourne City Council	0.0
32	Yarra City Council	1.6
27	Port Phillip City Council	1.9
25	Moreland City Council	3.5
28	Stonnington City Council	4.1
17	Maribyrnong City Council	4.3
23	Moonee Valley City Council	4.3
3	Boroondara City Council	4.6
7	Darebin City Council	5.3
11	Hobsons Bay City Council	6.2

1) Properties in Melbourne city Council is at shortest distance from the central Business Development are.

#Total price every year?

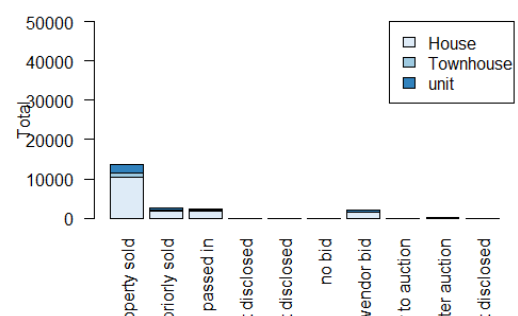
```
year_total <- xtabs(Price~year, data)
options(scipen = 999)
barplot(year_total,
        col=brewer.pal(n = 3, name = 'Blues'))
title(main = "Total Price Household Market per Year",
      ylab = "Total Price")
```



1) In 2017 property in Melbourne was at it's highest market rate in the dataset of data from year 2016- 2018. People who bought property in 2017 has to pay lot compared to the year previous. Of course, day by day property price keeps spiking up.

#Method of Sales every Type of real estate?

```
table(droplevels(data$Sales_Method), data$Type)
barplot(xtabs(~Type + Sales_Method, data), las=2,
        ylim = c(0,50000), ylab = "Total",
        col=brewer.pal(n = 3, name = 'Blues'))
legend("topright",
      c("House", "Townhouse", "unit"),
      fill=brewer.pal(n = 3, name = 'Blues'))
```



#Which property agent has the highest rate price for house?

```
mean_house <- data[(data$Type == "House"),]
mean_high <- aggregate(Price~Agent+Type, mean_house, mean)
mean_high[order(mean_high$Price, decreasing = T),][1,]
#Which property agent has the highest rate price for unit?
mean_unit <- data[(data$Type == "Unit"),]
mean_unit <- aggregate(Price~Agent+Type, mean_unit, mean)
mean_unit[order(mean_unit$Price, decreasing = T),][1,]
#Which property agent has the highest rate price for townhouse?
mean_Thouse <- data[(data$Type == "Townhouse"),]
mean_Thouse <- aggregate(Price~Agent+Type, mean_Thouse, mean)
mean_Thouse[order(mean_Thouse$Price, decreasing = T),][1,]
```

```
> mean_high[order(mean_high$P
      Agent Type Price
247 Sotheby's House 3531250
> mean_unit[order(mean_unit$P
      Agent Type Price
140 Scott Unit 1860000
> mean_Thouse[order(mean_Thou
      Agent Type Price
23 Castran Townhouse 2175000
> |
```

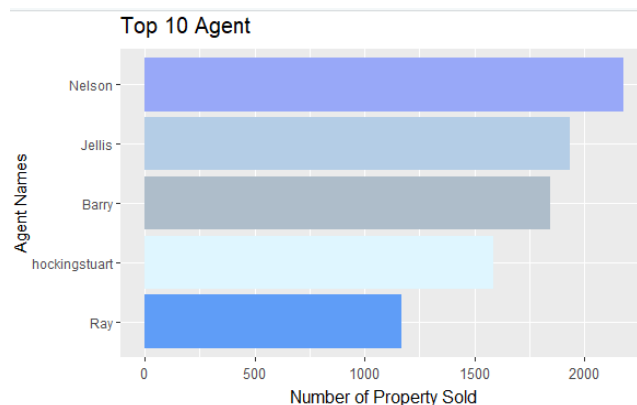
1) House sold in Melbourne at highest price was by Sotheby's.

2) Unit/Apartment sold in Melbourne at the highest price was by Scott.

3) Townhouse sold in Melbourne at the highest price was by Castran.

#Top10 Agent wrt property sold

```
top10_agent <- data %>% group_by(Agent) %>%
  summarise(Number = n()) %>% arrange(desc(Number)) %>%
  head(5)
ggplot(top10_agent, aes(reorder(Agent, Number), Number, fill = Agent))+
  geom_bar(stat = "identity")+
  mytheme2+
  theme(legend.position = "none")+
  labs(x = "Agent Names", y = "Number of Property Sold",
       title = "Top 10 Agent")+
  scale_fill_manual(values = mycolors)+
  coord_flip()
```



1) The graph shows top 10 agent that has sold most properties. Agent Nelson selling out more than 2000 properties.

#Price Distribution of property Type

```
ggplot(data, aes(Type, Price)) +
  geom_boxplot(outlier.colour = "black", fill="#5837D0") +
  scale_x_discrete(labels = c('Houses', 'Townhouses', 'Units')) +
```



```
scale_y_continuous(breaks=seq(0,10000000,1250000)) +
xlab("Type of Property") +
ylab("Price") +
ggtitle("Price Distribution of Home Type")
```



1) We got that houses are the most expensive property type in Melbourne than townhouses and apartments.

2) If you need to own a house-type of property in Melbourne you'll need a hefty amount of money.

3) You can go for townhouses if you are looking for something like a house but don't have enough budget.

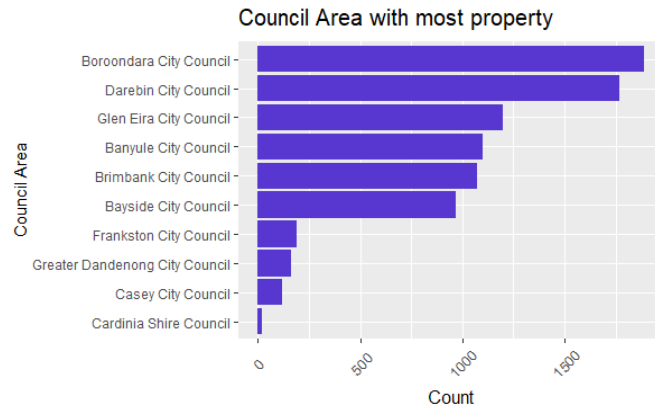
3) Apartments are less expensive to buy if looking for property in Melbourne, Apartments rock!

#Council Area that has most property

```
data %>%
  group_by(Council_Area) %>%
  summarise(Count = n()) %>%
  head(10) %>%
  ggplot(aes(reorder(Council_Area, Count), Count)) +
  geom_bar(stat = 'identity', fill="#5837D0") +
  mytheme+
  theme(axis.text.x = element_text(angle = 45, vjust = 0.5)) +
  coord_flip() +
  xlab("Council Area") +
  ylab("Count") +
  ggtitle("Council Area Frequencies")
```

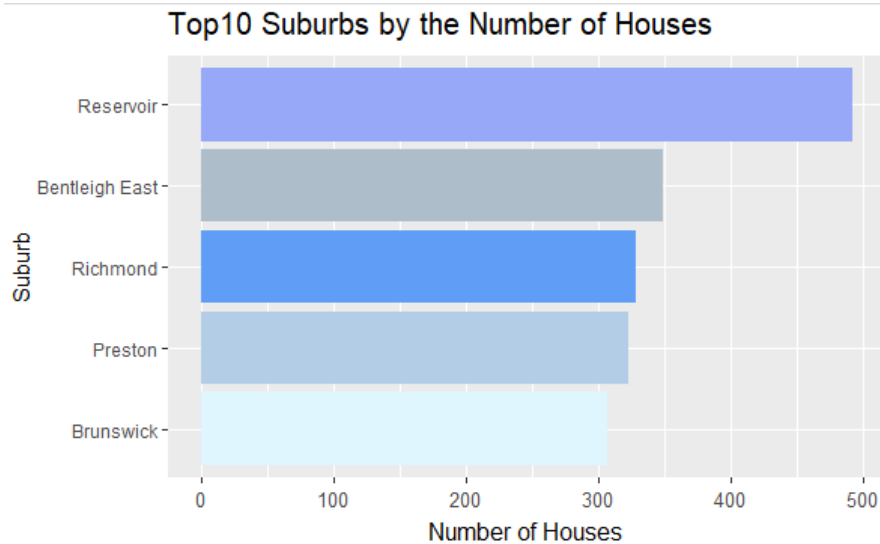
1) Boroondara City council as most property count, probably in southern Metropolitan Region.

2) Darebin and Glen Eira city council comes next after Boroondara city council.



#Top10 Suburbs by the Number of Houses

```
top10sub_by_houses <- data %>% group_by(Suburb) %>%
  summarise(Number = n()) %>% arrange(desc(Number)) %>%
  head(5)
ggplot(top10sub_by_houses, aes(reorder(Suburb, Number), Number, fill = Suburb))+
  geom_bar(stat = "identity")+
  mytheme2+
  theme(legend.position = "none")+
  labs(x = "Suburb", y = "Number of Houses",
       title = "Top10 Suburbs by the Number of Houses")+
  scale_fill_manual(values = mycolors)+
  coord_flip()
```

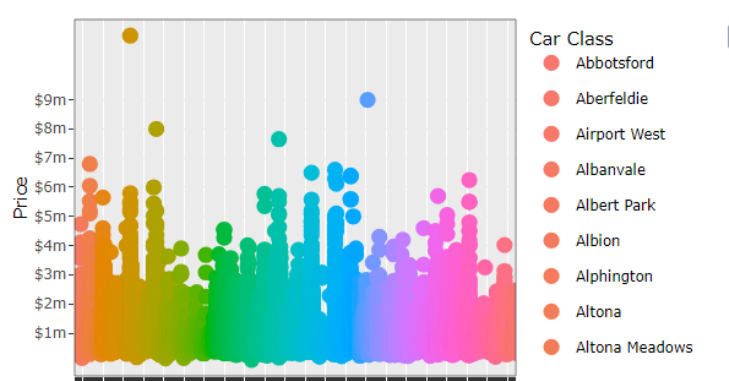


1) In Melbourne property count more properties are in Reservoir Suburb in the Northern Metropolitan region, followed by Richmond and Preston Suburb.

2) Also Bentleigh East and South Yarra in Southern and Metropolitan after Northern Metropolitan.

#Suburb vs Price

```
p <- ggplot(data, aes(x=Suburb,
                      y=Price,
                      las=2, srt=45, xpd=TRUE,
                      color=Suburb)) +
  geom_point(size=3) +
  scale_y_continuous(breaks =
c(1000000,2000000,3000000,4000000,5000000,6000000,7000000,8000000,9000000),
  labels = c("$1m", "$2m", "$3m", "$4m", "$5m", "$6m", "$7m", "$8m", "$9m"))+
  labs(color = "Car Class") +
  theme_bw()
ggplotly(p)
```

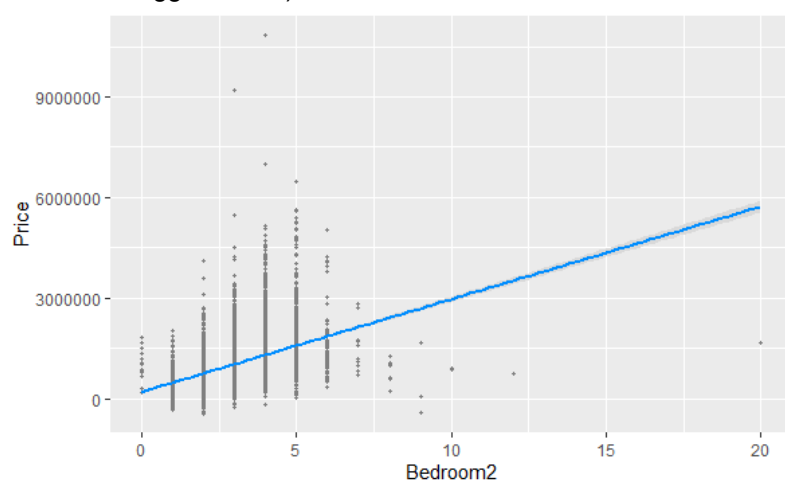


1) This is plotly graph an interactive graph displaying it's price and suburb when hovered over the points.

2) It's a pretty graph the image doesn't justify it.

#Linear Regression

```
houseDist <- lm(Price ~ Distance + Bathroom + Bedroom2 + Landsize +
  BuildingArea,
  data = house2)
visreg(houseDist, "Bedroom2", gg = TRUE)
```



1) No. of bedrooms increases the fancy the price gets.

CONCLUSION:

- 1) *By this analysis we can conclude that Melbourne House rates are no joke.*
- 2) *We got the priciest region and suburb in Melbourne.*
- 3) *We got the most expensive type of property i.e house.*
- 4) *The most selling method of property in Melbourne.*
- 5) *Council city area with the most property count.*
- 6) *How many rooms do properties in Melbourne have, commonly found no. of rooms that properties have all over Melbourne.*
- 7) *Which Agent sold the most properties from the year 2016- 2018, and the Highest price at which the agent sold what kind of property.*
- 8) *We have the day the first and last property sold in the dataset.*
- 9) *The area that is at the closest distance from the central business development area.*
- 10) *The region and Suburb that has the most property.*