Assignment 1 – Fundamentals of Machine Learning

Submitted By - Mohana Krishna Nitesh Varma

Kent ID – 811292363

1. Dataset – House Prices Dataset

(Source: Kaggle, Link to dataset: https://www.kaggle.com/datasets/lespin/house-prices-dataset?resource=download)

This dataset is a combination of quantitative and qualitative (categorical) variables.

Quantitative variables – E.g., sale price, lot area

Qualitative variables – E.g., neighborhood, sale condition, lot config

2. Import the dataset into R

```
library(readr)

data <- read_csv("house_prices.csv")

str(data)
```

3. Print out descriptive statistics for a selection of quantitative and categorical variables

```
# Summary statistics for quantitative variables

summary(data$SalePrice)

summary(data$LotArea)


# Frequency table for a categorical variable ('Neighborhood')

table(data$Neighborhood)
```

4. Transform at least one variable. It doesn't matter what the transformation is.

```
# Transforming a quantitative variable (logarithm of 'LotArea')

data$log_LotArea <- log(data$LotArea)
```

5. Plot at least one quantitative variable, and one scatterplot

```
# Ploting a histogram for a quantitative variable (SalePrice)

hist(data$SalePrice, main = "Histogram of Sale Price", xlab = "Sale Price")
```

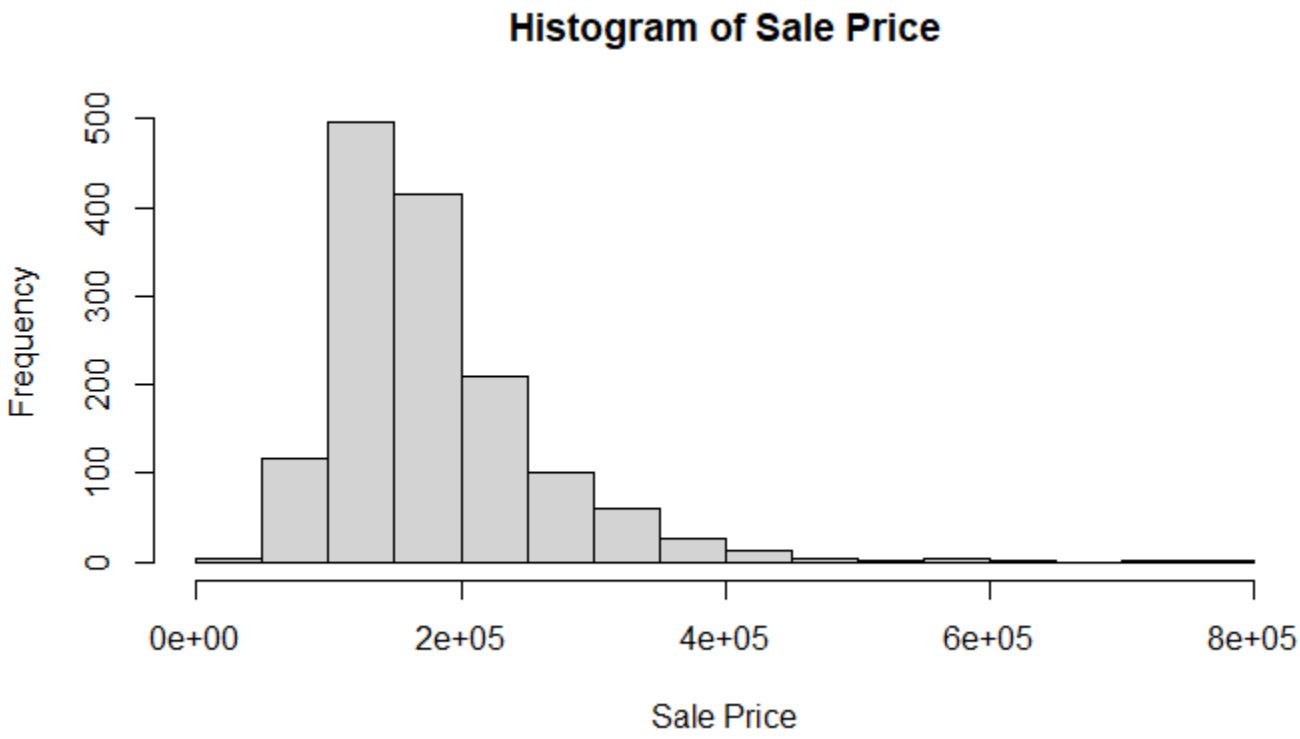
Creating a scatterplot (LotArea vs. SalePrice)

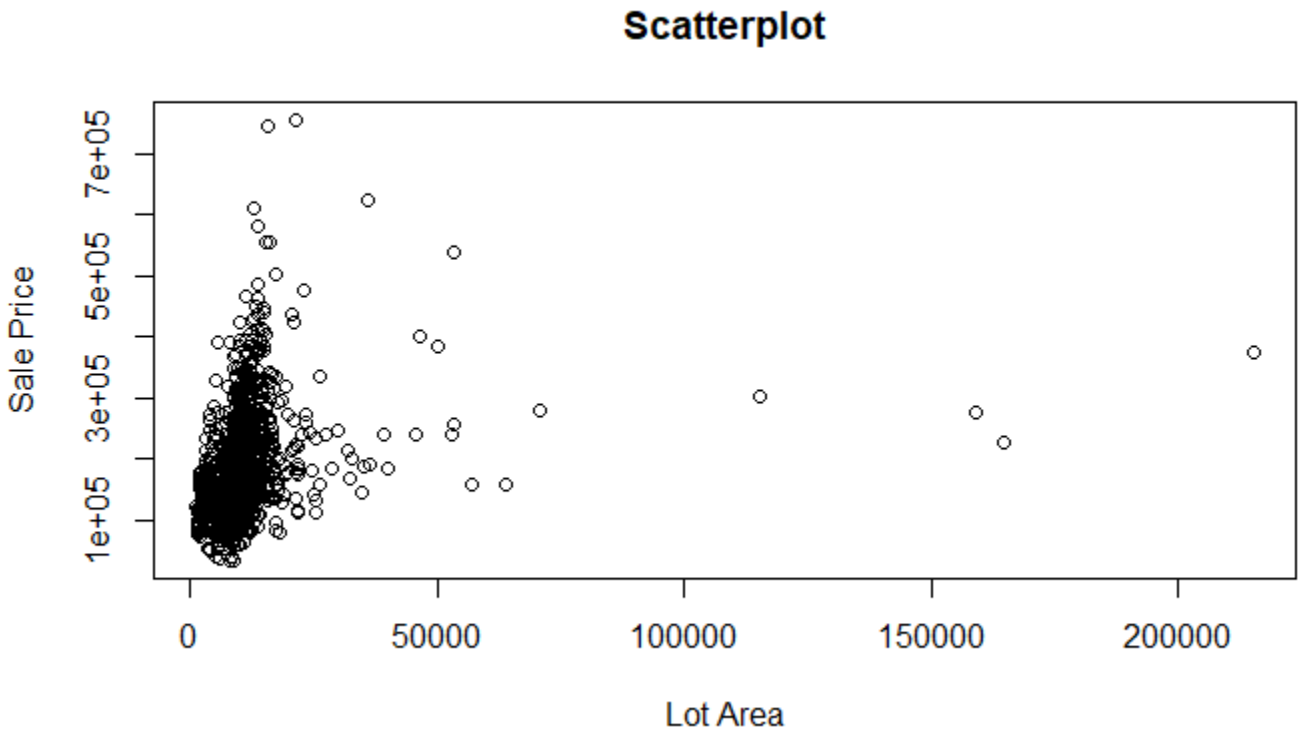
plot(data$LotArea, data$SalePrice, main = "Scatterplot", xlab = "Lot Area", ylab = "Sale Price")

Outputs:

```
> summary(data$SalePrice)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  34900  129975  163000  180921  214000  755000
> summary(data$LotArea)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1300    7554    9478   10517   11602  215245
> table(data$Neighborhood)

Blmngtn Blueste  BrDale BrkSide ClearCr CollgCr Crawfor Edwards Gilbert  IDOTRR MeadowV
     17       2      16      58      28     150      51     100      79      37      17
Mitchel   NAmes NoRidge NPkVill NridgHt  NWAmes OldTown  Sawyer SawyerW Somerst StoneBr
     49     225      41       9      77      73     113      74      59      86      25
  SWISU  Timber Veenker
     25      38      11
>
```




Scatterplot
Sale Price
1e+05
3e+05
5e+05
7e+05
0
50000
100000
150000
200000
Lot Area