Name	Rucha Kulkarni
Class	BE Computer Engineering (Batch F)
UID	2021300067
Exp No.	5

Aim: Create advanced charts using Power BI / Tableau / R / Python / D3.js on the dataset - Housing data.

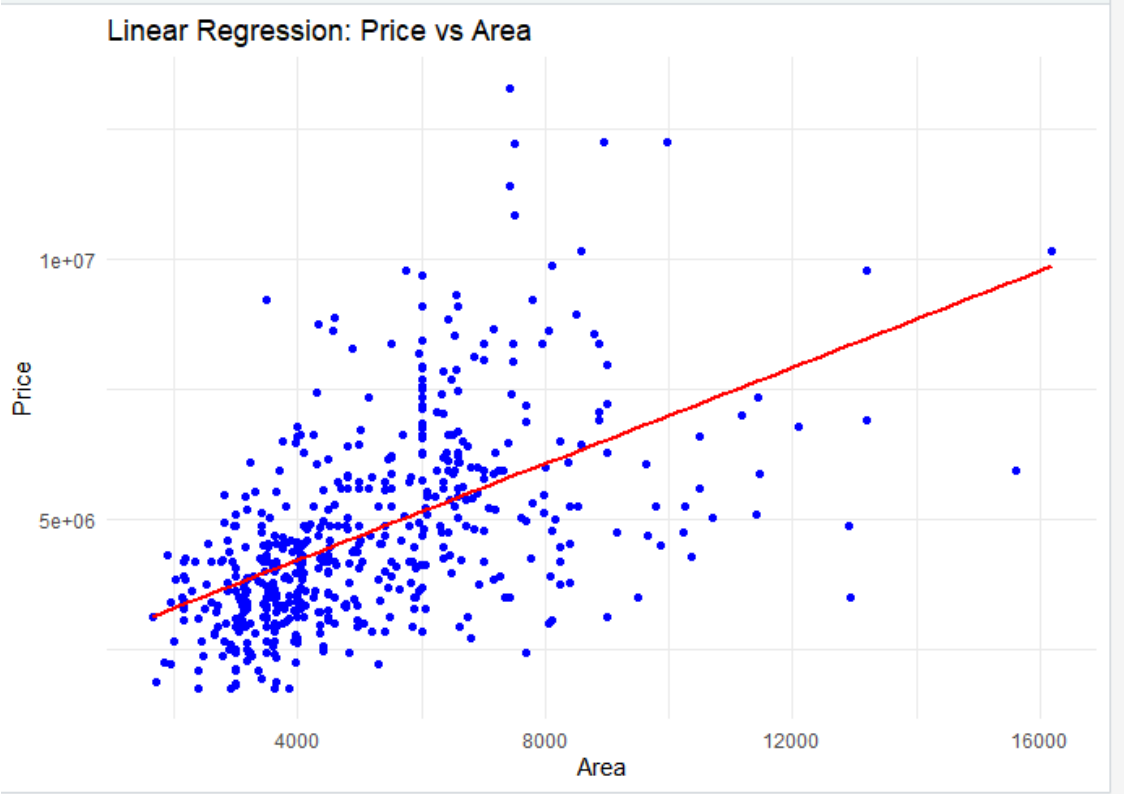
Description:

Dataset used is Housing Dataset available at

https://www.kaggle.com/datasets/ashydv/housing-dataset

Graphs and Observations:

Linear Regression:

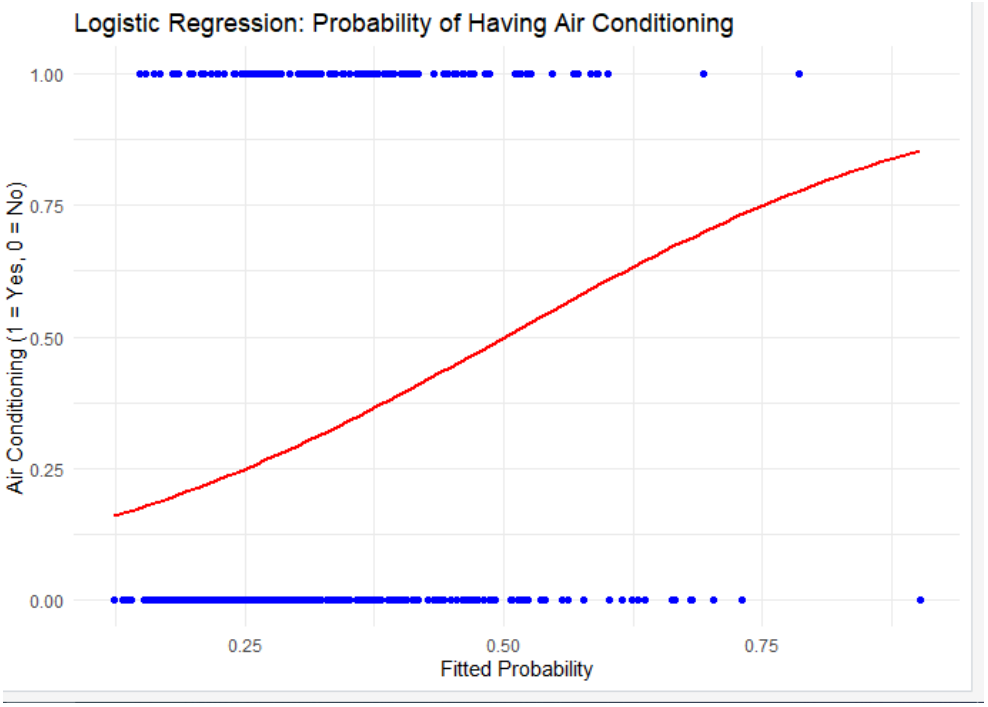


```
> model <- lm(price ~ area, data = housing_data)
>
> summary(model)

Call:
lm(formula = price ~ area, data = housing_data)

Residuals:
     Min       1Q   Median       3Q      Max
-4867112 -1022228  -200135   683027  7484838

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.387e+06  1.745e+05   13.68   <2e-16 ***
area        4.620e+02  3.123e+01   14.79   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1581000 on 543 degrees of freedom
Multiple R-squared:  0.2873,    Adjusted R-squared:  0.286
F-statistic: 218.9 on 1 and 543 DF,  p-value: < 2.2e-16
```

Observation:

The red line (regression line) has a positive slope, indicating that as the area of a house increases, the price also tends to increase. There is some variability around the line, indicating that other factors might also influence the price.

Logistic Regression:




```
> housing_data$fitted_ac <- predict(log_model_ac, type = "response")
> ggplot(housing_data, aes(x = fitted_ac, y = airconditioning_bin)) +
+     geom_point(color = "blue") +
+     geom_smooth(method = "glm", method.args = list(family = "binomial"), se
= FALSE, color = "red") +
+     labs(x = "Fitted Probability", y = "Air Conditioning (1 = Yes, 0 = No)",
+          title = "Logistic Regression: Probability of Having Air Conditionin
g") +
+     theme_minimal()
`geom_smooth()` using formula = 'y ~ x'
> summary(log_model_ac)
```

```
Call:
glm(formula = airconditioning_bin ~ area + bedrooms, family = binomial,
    data = housing_data)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -3.052e+00  4.583e-01  -6.661 2.72e-11 ***
area         2.017e-04  4.431e-05   4.552 5.32e-06 ***
bedrooms     4.019e-01  1.309e-01   3.071  0.00213 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 679.62  on 544  degrees of freedom
Residual deviance: 643.82  on 542  degrees of freedom
AIC: 649.82

Number of Fisher Scoring iterations: 4
```

Observation: The curve shows that both area and bedrooms are significant predictors of having air conditioning. The positive coefficients indicate that as area and bedrooms increase, the probability of having air conditioning also increases.

Conclusion:

From this experiment, we learned about regression by analyzing logistic and linear models. Logistic regression showed that area and number of bedrooms significantly affect the likelihood of having air conditioning. Linear regression taught us to predict continuous outcomes, like house prices.These insights improved our understanding of interpreting relationships between predictors and outcomes in statistics.