**MACHINE LEARNING**

**Using Regression Model**

**Project submitted**

**IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR AWARD OF DEGREE OF**

**Master of Technology**

**In**

**Industrial Engineering & Management**

**Submitted by**

**ADARH KUMAR**

**Admission No- 19MT0016**



**Department of Management Studies**

**Indian Institute Of Technology (Indian School Of Mines)**

**Dhanbad- 826004(Jharkhand)**

**ACKNOWLEDGEMENT**

**First of all I am very thankful to my institute IIT (ISM) Dhanbad, my department as well as our respected HOD Prof. SAUMYA SINGH for providing me the opportunity for getting the research oriented knowledge.**

**I would like to express my sincere gratitude to Dr.Rashmi Singh, Assistant Professor, Department of Management Studies, for his noble involvement and valuable guidance during completion of my term paper. His forth coming motivation, untiring zeal for work were a great source of inspiration for me. I have also got full support from all professors of my department. I am also very thankful to whole Management Studies Department for providing me the best facility & good environment to complete my term paper successfully. I express my gratitude to each & every person of the Management Studies Department for their co-operation & for giving their most valuable time. I wish to extend my thanks to all the staff members of Department of Management Studies who helped me whenever I needed help.**

**CONTENTS**

|  |  |
| --- | --- |
| **S.NO** | **TOPIC** |
| **1** | Problem Description |
| **2** | Python Code |
| **3** | Mathematical Representation of Model |
| **4** | Conclusion |

**Problem Description**

**Delivery Time Prediction**

**Delivery time of Pizza order is affected when Coke, Crust and Bread are present in an order.**

**Data Description**

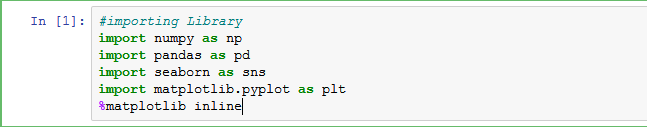
There are four variables in the dataset which are **Coke, Crust , Bread and Delivery (Time).**

**Coke, Crust and Bread are categorical data and Delivery (time) is numerical data.**

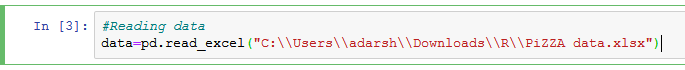
Coke, Crust, Bread are coded as 0 or 1. 1 is coded as presence of item in order and 0 is coded as absence of item in order

**Python Code**

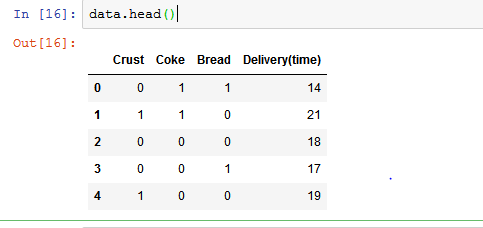
Importing useful library functions



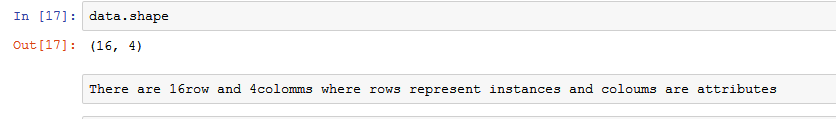
Loading data set in Jupyter Notebook



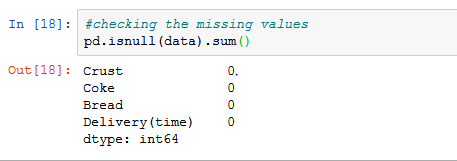
Checking how data is in the file by looking to some rows and columns.



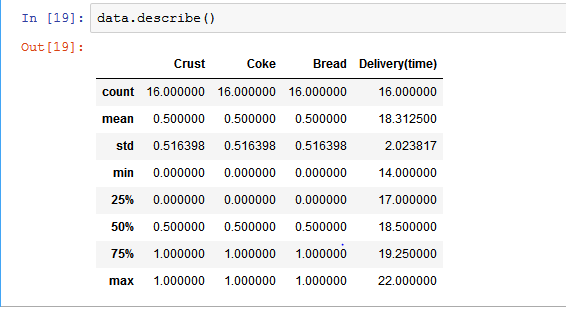
Checking how many rows and columns are there in dataset. Rows represent instances and column represents attributes or features.



Checking if there is any missing values in data set

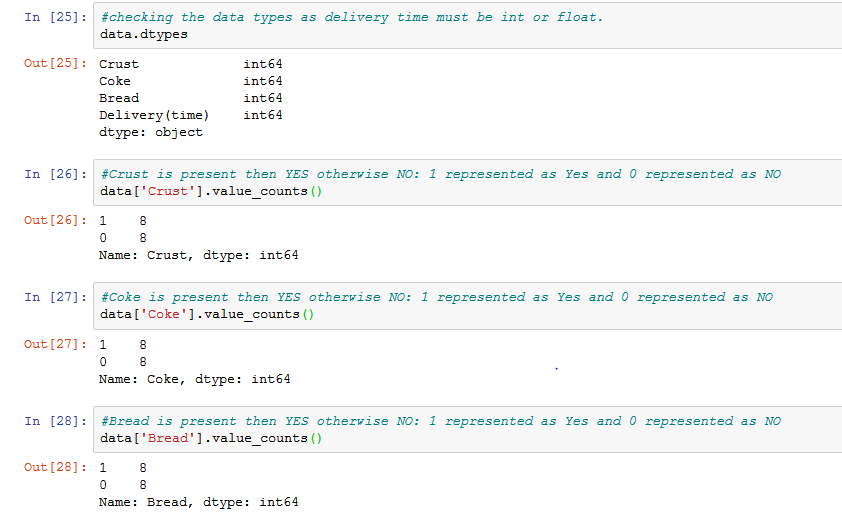


There is no missing values in data set.



This help in getting general idea about numeral data present in data set as only Delivery time is numerical other are categorical so five point summary is for Delivery time.

Checking the data types of data whether there is integer, float or string type.

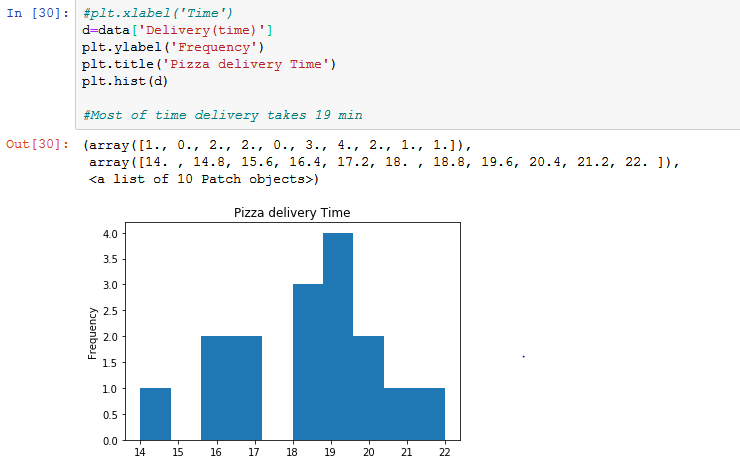


All data sets are integer type as Coke, Crust, Bread are coded as 0,1.

1 is coded as presence of in order and 0 is coded as absence in order.

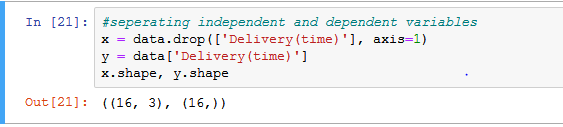
Coke, Crust, Bread all are having 8 presence and 8 absence in whole data set.

Using histogram to analyze Pizza delivery time.



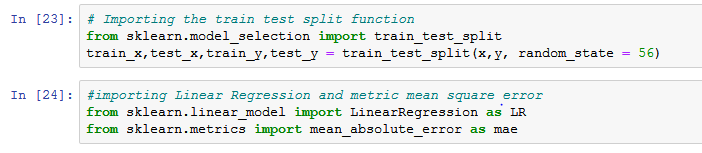
The is delivered in 19 min most of the time.

Separating independent and dependent varibles

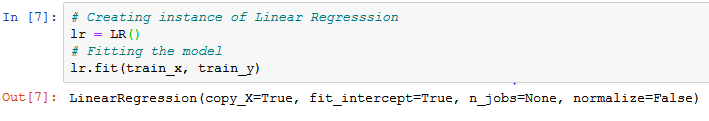


Considering delivery time as dependent variable and rest all as independent variables.

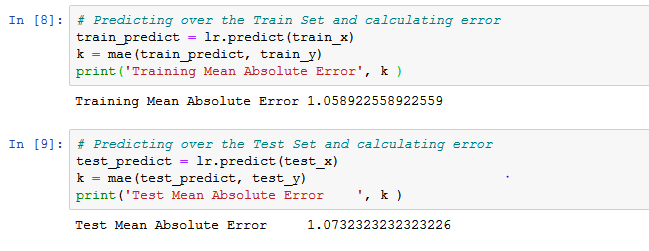
Splitting the dataset into training and test datasets.



Fitting the training data into linear regression model.

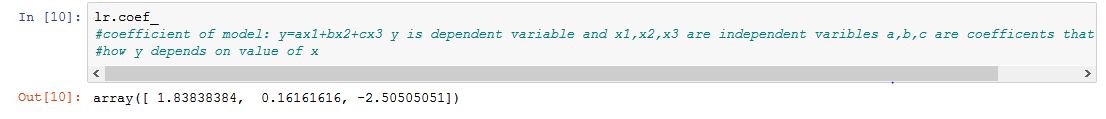


Calculating the errors in training data and test data.



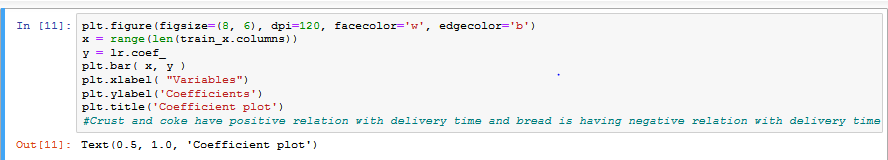
Using MAD(Mean Absolute Error) for error prediction.

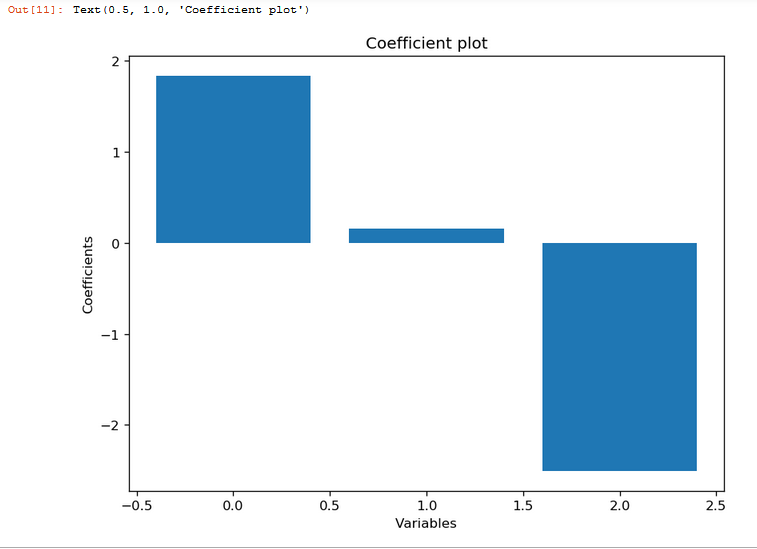
Checking how the independent variables are related to dependent variable.



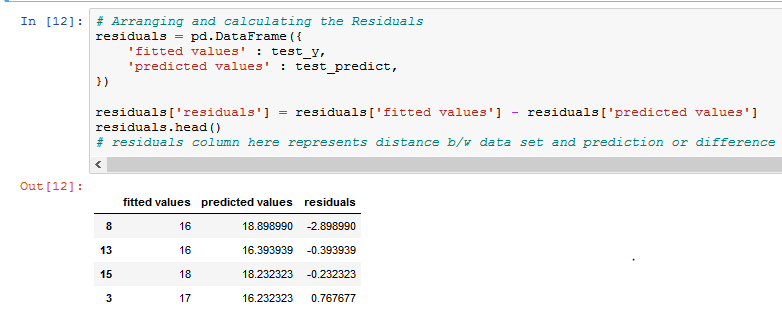
Crust and coke are positively related to delivery time and Bread is negatively related.

Same relation is represent as graph.



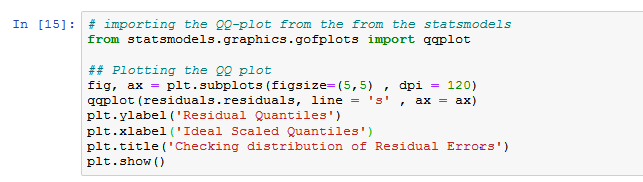


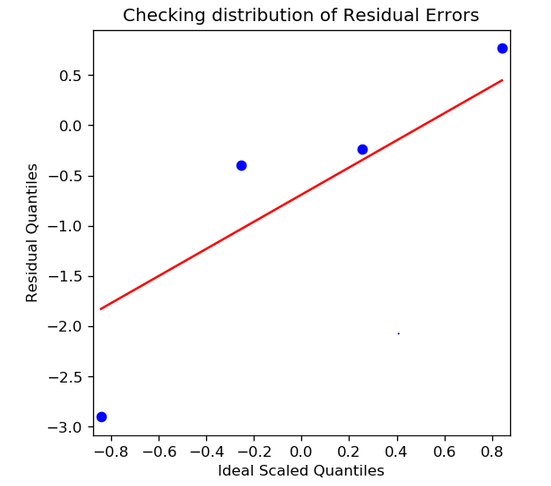
Checking how far is our prediction by calculating the distance of regression line from data points.



Residuals column represents the distance.

Same is plotted by QQ-plot.



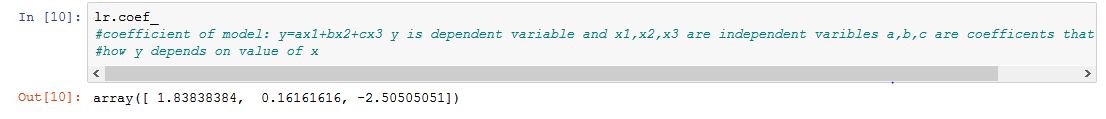


**Mathematical Representation of Model**

Here, we are using Linear Regression model

Y=ax1+bx2+cx3

[Y represents Delivery(time), x1 represents Crust , x2 represents coke , x3 represents Bread ]



Y=1.838384(x1)+0.16161616(x2)-2.50505051(x3)

**Conclusion**

Crust and coke are positively related to delivery time and Bread is negatively related.

On ordering crust and coke in an order delivery time will increase.