**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in this Minor thesis titled **“AMAZON DELIVERY DATA ANALYSIS”** in fulfilment of the requirement for the degree of Master of Technology in Computer Engineering (specialization in Networking) and submitted to

“**SATYUG DARSHAN INSTITUTE OF ENGINEERING AND TECHNOLOGY**”*,* is an authentic record of my own work carried out under the supervision of **Mr. Ankit Mishra.**

The work contained in this thesis has not been submitted to any other University or Institute for the award of any other degree or diploma by me.

Vishnu Garg Yash

BCA(DS-23/057) BCA(DS-23/058)

1

**ACKNOWLEDGEMENT**

It is with deep sense of gratitude and reverence that I express my sincere thanks to my supervisor **Mr.Ankit Mishra** for her guidance, encouragement, help and useful suggestions throughout. Her untiring and painstaking efforts, methodical approach and individual help made it possible for me to complete this work in time. I consider myself very fortunate for having been associated with the scholar like her. Her affection, guidance and scientific approach served a veritable incentive for completion of this work.

I shall ever remain indebted to the faculty members of **Satyug Darshan Institute of Engineering and Technology, Faridabad, CAREERERA** team and all my classmates for their cooperation, kindness and general help extended to me during the completion of this work.

Although it is not possible to name individually, I cannot forget my well-wishers at Satyug Darshan Institute of Engineering and Technology, Faridabad and outsiders for their persistent support and cooperation.

This acknowledgement will remain incomplete if I fail to express my deep sense of obligation to my parents and God for their consistent blessings and encouragement

(Vishnu Garg)

(Yash)

**2**

**CERTIFICATE**

3

**ABOUT TRAINING**

As each and every sector of the market is growing, data is building up day by day, we need to keep the record of the data which can be helpful for the analytics and evaluation. Now we don’t have data in gigabyte or terabyte but in zetta byte and petabyte and this data cannot be handled with the day By day software such as Excel or MATLAB. Therefore, in this report we will be dealing with large data sets with the high-level programming language ‘Python’.

The main goal of this training is to aggregate and analyse the data collected from the different data sources available on the internet like Kaggle etc., This project mainly focuses on the usage of the python programming language and Data Analysis. This language has not only it’s application in the field of just analysing the data and represent it graphically.

4

**LIST OF ABBRIVATIONS**

*•* pd for pandas

*•* np for NumPy

*•* plt for pyplot

*•* sns for seaborn

*•* df for data frame

*•* int for integer

*•* len for length

*•* str for string

*•* bool for Boolean

*•* arr for array

*•* loc for location

*•* info for information

*•* col for columns

*•* hist for histogram

*•* sqrt for square root

5

**LIST OF FIGURES**

**Figures Page no.**

1.2.1 Arithmetic operator 10

1.2.2 Assignment operator 11

1.2.3 Logical operator 12

1.2.4 Comparison operator 12

1.2.5 Membership operator 13

1.2.6 if else statement 14

1.2.7 Elif statement 16

1.3.1 Matplotlib basic example 18

1.3.2 Pandas basic example 20

1.3.3 NumPy basic example 21

6

**CHAPTER 1: INTRODUCTION TO DATA SCIENCE 1.1 Introduction to topic**

* **Optimizing Operational Efficiency**: This report investigates the performance metrics of Amazon's delivery services, exploring how data analysis can enhance operational efficiency.
* **Addressing Multifaceted Challenges**: With an extensive global network, Amazon's delivery system faces complex challenges. This analysis aims to uncover insights that streamline processes and enhance delivery reliability.
* **Enhancing the Customer Experience**: By analyzing delivery data, this report seeks to improve the delivery experience for customers, ensuring timely arrivals and meeting expectations.
* **Data-Driven Strategies**: Leveraging data-driven approaches, this report examines trends and patterns to inform strategic decisions, aiming for continuous improvement in Amazon's delivery operations.

**1.2 Motivation**

* **Optimizing Delivery Routes**: Analyzing data can reveal patterns that enable the optimization of delivery routes, leading to reduced delivery times and costs while enhancing service reliability.
* **Enhancing Customer Satisfaction**: Insights gained from data analysis help ensure on-time deliveries, accurate tracking, and proactive customer communication, ultimately improving overall customer satisfaction.
* **Operational Efficiency**: By identifying bottlenecks and inefficiencies in the delivery process, Amazon can streamline operations, allocate resources more effectively, and enhance fleet management, thereby improving operational efficiency.
* **Competitive Advantage**: Leveraging data to enhance delivery logistics provides Amazon with a competitive edge by offering faster and more reliable service compared to competitors..

**1.3 Objective of training**

* To give participants the ability to create and refine predictive models that precisely categorize, regress, and cluster data in order to address practical issues.
* To give participants the know-how to prepare data for machine learning algorithms by preprocessing, transforming, and visualizing it.
* To teach participants how to choose, apply, and assess machine learning algorithms to handle challenging issues in computer vision, natural language processing, and recommender systems, among other fields.
* Giving participants the tools they need to implement and incorporate machine learning models into bigger systems while maintaining maintainability, scalability, and dependability.

7

**CHAPTER 2: PYTHON FOR DATA SCIENCE**

**1.1. Introduction to Python**

“Python is an interpreted, object-oriented, high-level programming language with dynamic semantics”. This language consist of mainly data structures which make it very easy for the data scientists to analyse the data very effectively. It does not only help in forecasting and analysis it also helps in connecting the two different languages. Two best features of this programming language is that it does not have any compilation step as compared to the other programming language in which compilation is done before the program is being executed and other one is the reuse of the code, it consist of modules and packages due to which we can use the previously written code anywhere in between the program whenever is required. There are multiple languages for example R, Java, SQL, MATLAB available in market which can be used to analyse and evaluate the data, but due to some outstanding features python is the most famous language used in the field of data science.

Python is mostly used and easy among all other programming languages.

8

**1.2 Operators, Conditional Statements …..**

**OPERATORS -** Operators are the symbols in python that are used to perform Arithmetic or logical operations. Following are the different types of operators in python.

**Arithmetic operators** - Arithmetic operators carry out mathematical operations and they are mostly used with the numeric values.

|  |  |  |
| --- | --- | --- |
| Arithmetic operators | | |
| Operator | Name | Example |
| + | Addition | A+B |
| - | Subtraction | A-B |
| \* | Multiplication | A\*B |
| / | Division | A/B |
| % | Modulus | A%B |
| \*\* | Exponentiation | A\*\*B |
| // | Quotient | A//B |

Fig. 1.2.1: Arithmetic

operators A and B are the numeric value

**Assignment operators** - As the name decides this operators are used for assigning the values to the variables.

|  |  |  |
| --- | --- | --- |
| ASSIGNMENTOPERATORS | | |
| Operator | Example | mayalsobe written |
| = | a=6 | a=6 |
| += | a+=3 | a=a+3 |
| -= | a-=4 | a=a-4 |
| \*= | a\*=5 | a=a\* 5 |
| /= | a /= 6 | a =a / 6 |
| %= | a%=7 | a=a%7 |
| //= | a//=8 | a=a// 8 |
| \*\*= | a\*\*=9 | a=a\*\* 9 |
| &= | a&= 1 | a=a&1 |

Fig. 1.2.2: Assignment Operators

Here a is any value and number of operations are performed on this value.

9

**Logical operators** - These operators are used to join conditional statements

|  |  |  |
| --- | --- | --- |
| Logical Operators | | |
| Operator | Description | Example |
| and | If both statements are true it | x *<*5 **and**x  *<*10  x *<*4 **or**x  *<*8  **not**(x*<*4  **and**x *<*8) |
|  | Returns true |
| or | If any of the two statement |
|  | Is true it returns true |
| not | If the result is true it reverses the result and gives false |

Fig.1.2.3:Logical Operators

Here a is any value provided by us and on which multiple operations can be performed.

**Comparison operators** - These operators are used to compare two different values.

|  |  |  |
| --- | --- | --- |
| Comparison operators | | |
| Operator | Name | Example |
| == | Equal | a==b |
| != | Not equal | a!=b |
| *>* | Greatert han | a*>*b |
| *<* | less than | a*<*b |
| *>*= | Greater than | a*>*=b |
|  | Equal to |  |
| *<*= | Less than equal to | a*<*=b |

Fig. 1.2.4: Comparison operators

Here a and b are two different values and these values are compared.

10

**Membership operators** - These operators are used to check membership of a particular value. It is used to check whether a specific value is present in the object or not.

|  |  |  |
| --- | --- | --- |
| Membershipoperators | | |
| Operator | Description | Example |
| in  not in | It returns a True if the value is present inside the object  It returns a True if the value is not present inside the object | A **in** b  A **not in** b |

Fig.1.2.5:Membershipoperators

# Condition statements

**If elsestatements**

“Themost common type of statement is the if statement. ifstatement consistof a block which is called as clause”, it is the block after if statement, it executed the statement if the condition is true.The statement is omitted if the conditionis False. then the statement in the else part is printed

If statement consist of following-

*•* **If keyword itself**

*•* **Condition which may be True or False**

*•* **Colon**

*•* **If clause or a block of code** Below is the figure shows how If and else statements are used with description inside it.



Figure 1.2.6 : if else statement

**elif statements**

In this statement only one statement is executed, There are many cases in which there is only one possibility to execute. ”The elif statement is an else if statement that always follows an if or another elif statement”[8]. The elif statement provides another condition that is checked only if any of the previous conditions were False. In code, an elif statement always consists of the following:. The only difference between if else and elif statement is that in elif statement we have the condition where as in else statement we do not have any condition.

elIf statement consist of following -

*•* **elIf keyword itself**

*•* **Condition which may be True or False**

*•* **Colon**

*•* **elIf clause or a block of code**

Below is the figure shows how elIf statement is used with description inside it.

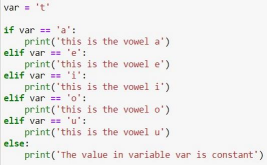


Figure 1.2.7: elif example

**1.3 Understanding Standard Libraries Pandas, Numpy…..**

Libraries in Python

Python library is vast. There are built in functions in the library which are written in C lan- guage. This library provide access to system functionality such as file input output and that is not accessible to Python programmers. This modules and library provide solution to the many problems in programming.

Following are some Python libraries.

Matplotlib

Pandas

Numpy

12

**Matplotlib**

”Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy”[11]. MATLAB provides an application that is used in graphical user interface tool kits. Another such library is pylab which is almost same as MATLAB.

It is a library for 2D graphics, it finds its application in web application servers, graphical user interface toolkit and shell. Below is the example of a basic plot in python.

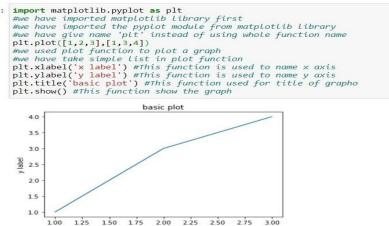


Figure 1.3.1: Matplotlib basic example

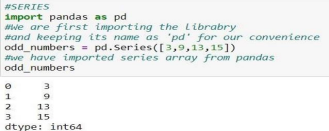
13

**Pandas**

Pandas is also a library or a data analysis tool in python which is written in python program- ming language. It is mostly used for data analysis and data manipulation. It is also used for data structures and time series.

We can see the application of python in many fields such as - Economics, Recommendation Sys- tems - Spotify, Netflix and Amazon, Stock Prediction, Neuro science, Statistics, Advertising, Analytics, Natural Language Processing. Data can be analyzed in pandas in two ways -

**Data frames -** In this data is two dimensional and consist of multiple series. Data is always represented in rectangular table.

**Series -** In this data is one dimensional and consist of single list with index. 

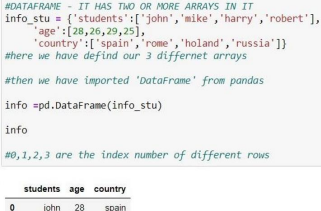


Figure 1.3.2: series and data frame in pandas

14

**NumPy**

”NumPy is a library for the Python programming language, adding support for large, multi- dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays”. The previous similar programming of NumPy is Numeric, and this language was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications. [12] It is an opensource library and free of cost.

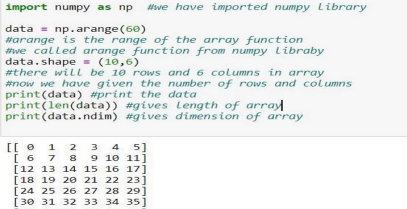


Figure 1.3.3: NumPy basic example

15

**CHAPTER 4: APPROACH USED (REQUIRED TOOLS)**

* **Decision Tree:** A decision tree is a type of supervised machine learning used to categorize or make predictions based on how a previous set of questions were answered.
* **KNN Algorithm:** K-NN algorithm stores all the available data and classifies a new data point based on the similarity.
* **Linear Regression :** A machine learning technique called linear regression uses a linear equation to express the linear relationship between one or more input data and a continuous output variable to predict the latter.
* **Ridge And Lasso:** While Lasso (Least Absolute Shrinkage and Selection Operator) Regression adds a penalty term to the cost function to reduce overfitting by setting some coefficients to zero, thereby performing feature selection, Ridge Regression adds a penalty term to the cost function to reduce overfitting by shrinking coefficients towards zero.
* **RandomForest Regressor:** Random Forest Regressor in Python is a supervised learning algorithm that combines multiple decision trees to predict a continuous output variable. With the aid of methods like fit, predict, and score, it can be applied to regression tasks.
* **Gradient Boosting Regressor** : is a supervised learning algorithm that combines multiple weak models to create a strong predictive model, iteratively training each tree to correct the errors of the previous tree

**REQUIRED TOOLS:**

For application development, the following Software Requirements are: Operating System: Windows 11

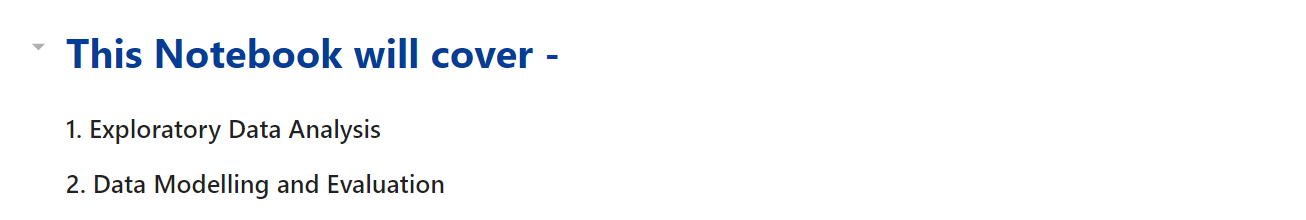
Language: python

Tools: JUPYTER notebook or COLAB, Microsoft Excel (Optional). Technologies used: python.

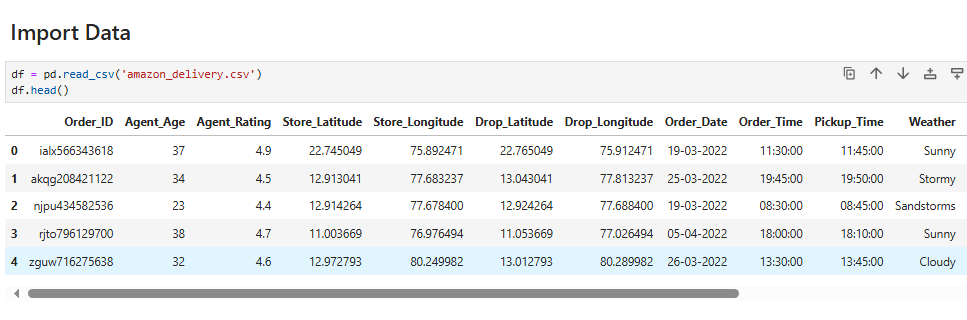
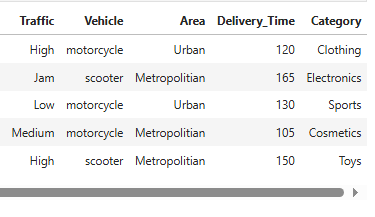
16

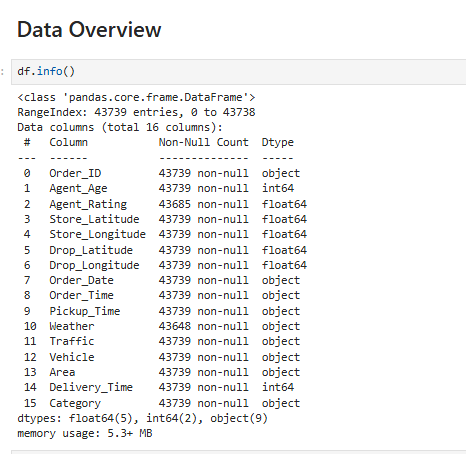
**CHAPTER 5: RESULTS**

****

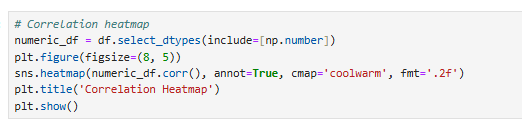
****

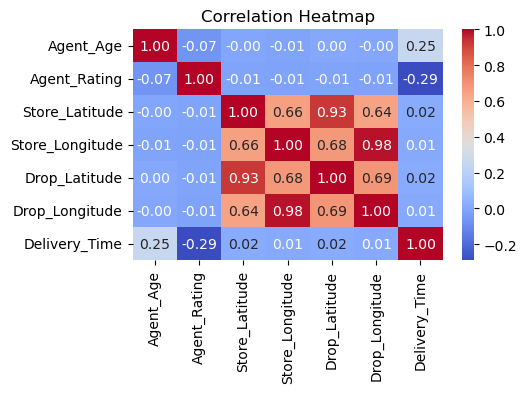




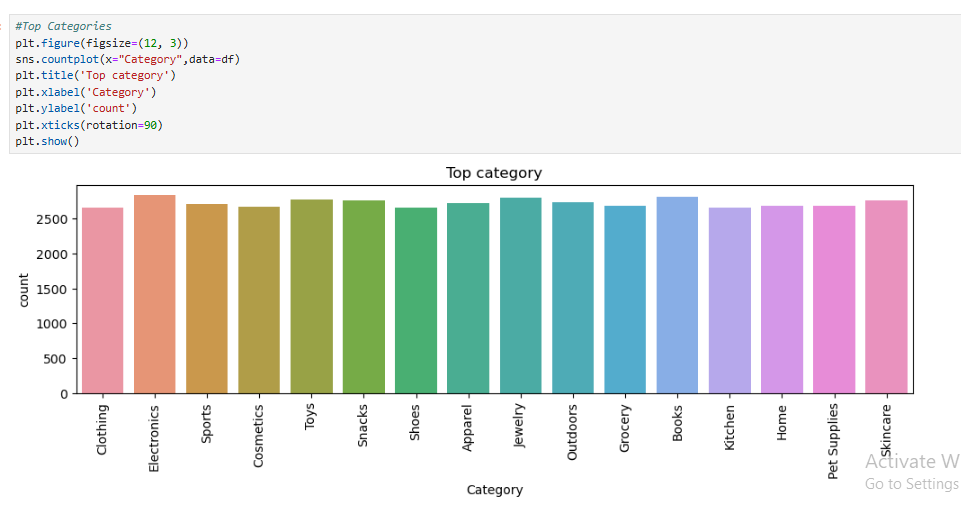


**Correlation Heatmap**

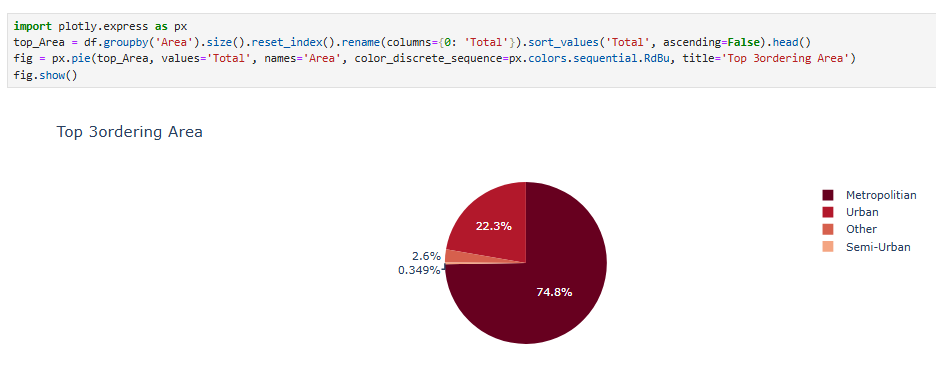
****

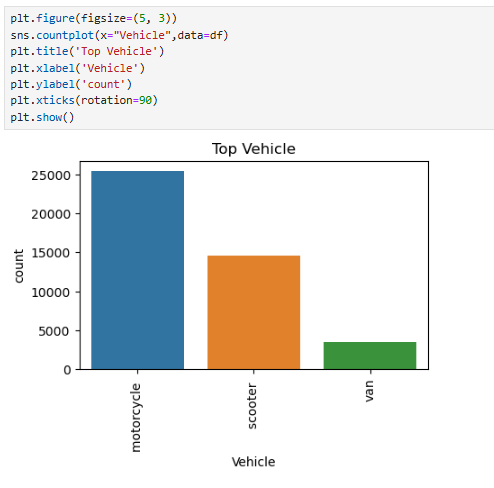
****

**Top Categories**

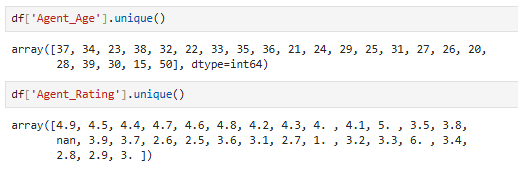
****

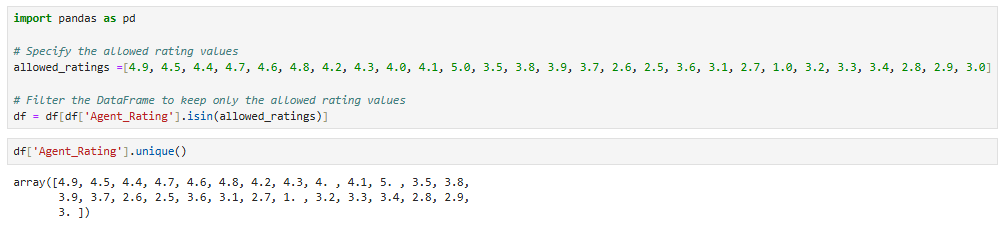
**Top 3 Ordering Area**



**Delivery Time And Vehicle Graph**

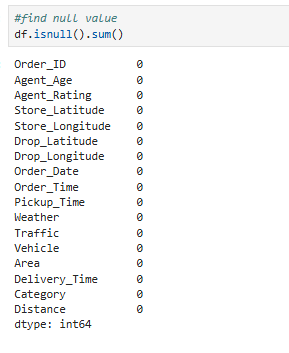


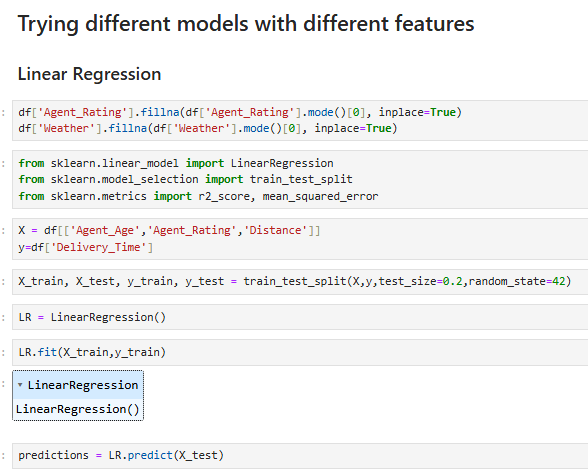


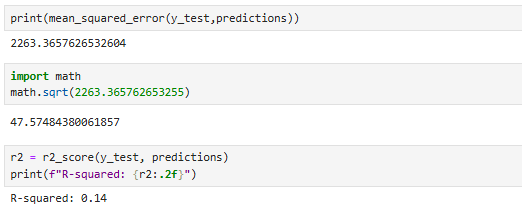


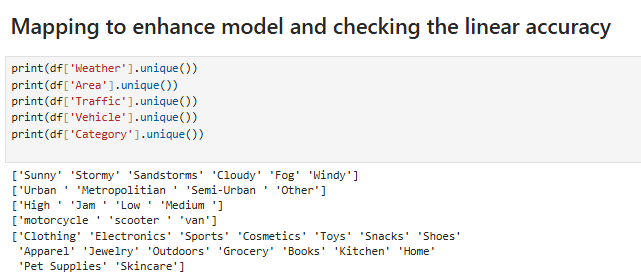








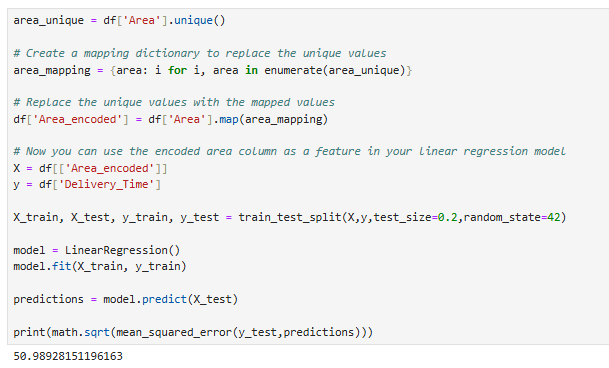




**Mapping Weather column**



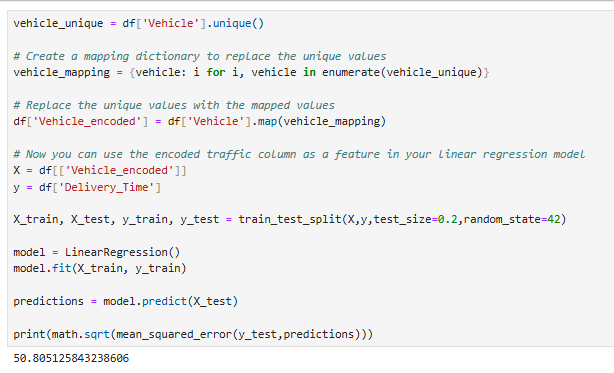
**Mapping Area Column**



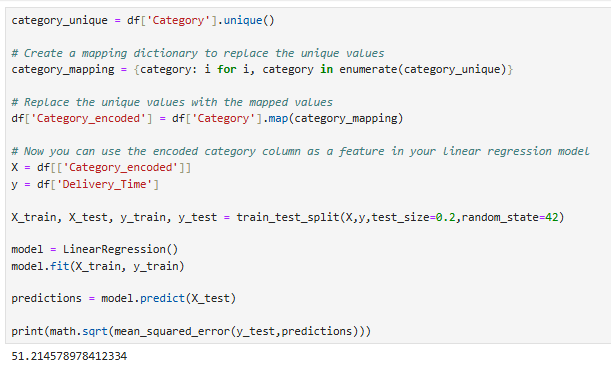
**Mapping Traffic Column**



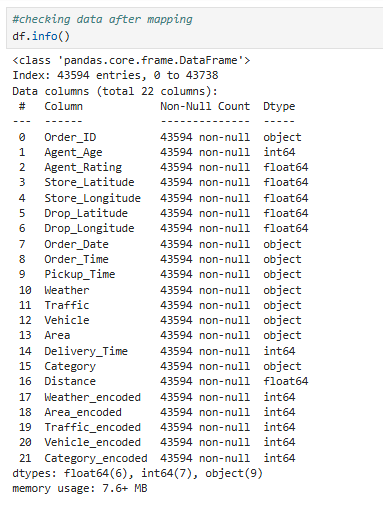
**Mapping Vehicle Column**



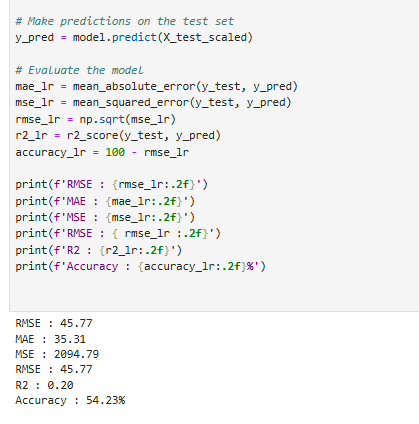
**Mapping Category Column**

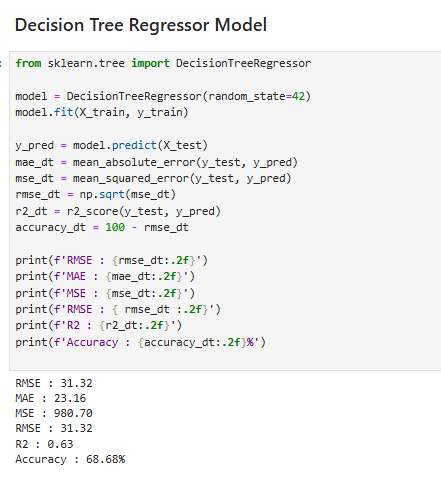


**Check Mapping Column info**

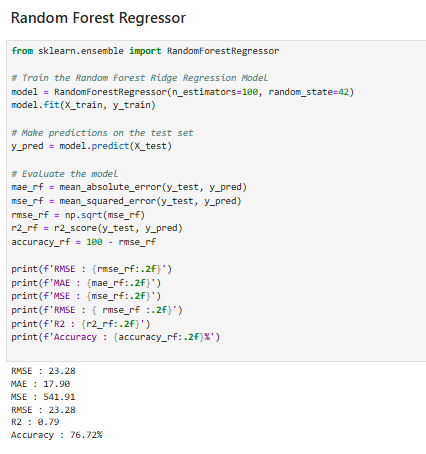


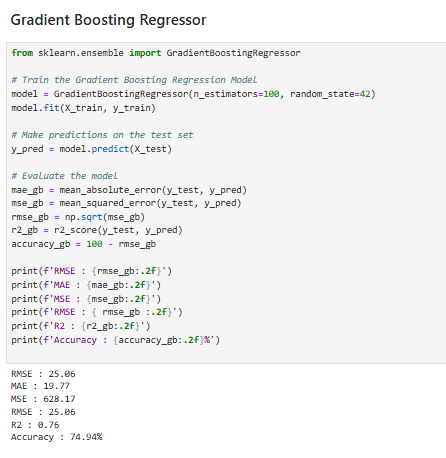


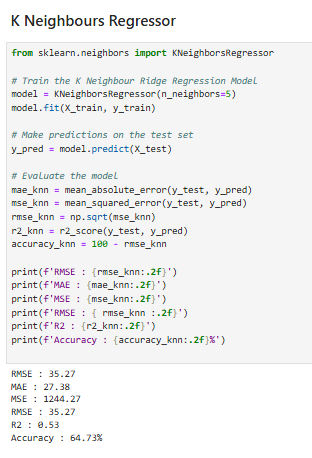


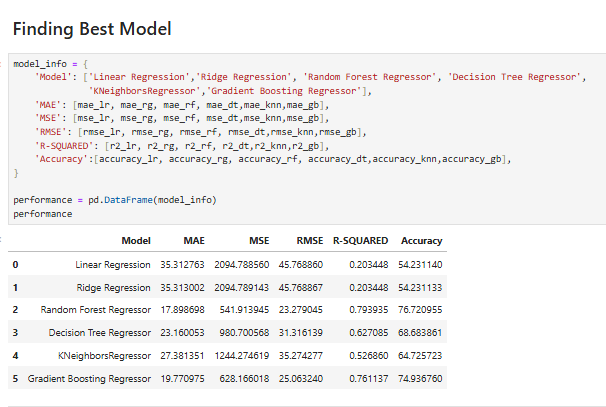


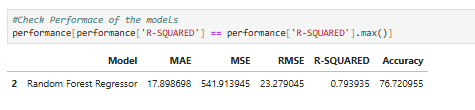


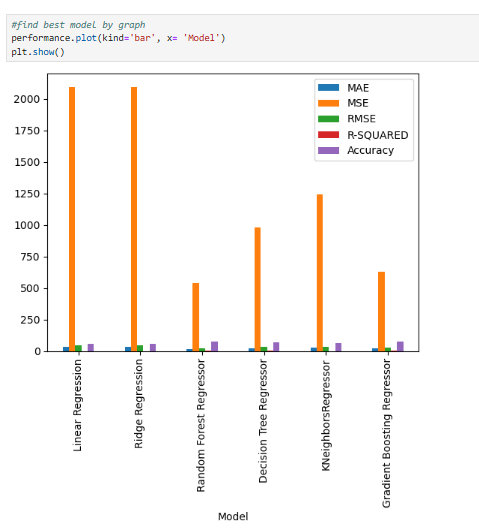




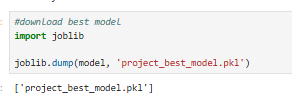














**SUMMARY & CONCLUSIONS**

 **Route Optimization**: The ML model optimizes delivery routes by considering factors such as traffic patterns, weather conditions, and delivery schedules. This minimizes delivery times and reduces operational costs.

 **Real-Time Adaptability**: The model adjusts dynamically to real-time data inputs, ensuring that deliveries are adjusted promptly in response to changing circumstances like traffic congestion or unexpected events.

 **Enhanced Customer Experience**: By predicting delivery times accurately and offering real-time updates, Amazon improves the overall customer experience. Customers benefit from reliable and timely deliveries, leading to higher satisfaction and loyalty.

**REFERENCES & BIBLIOGRAPHY**

**www.simplilearn**

**www.kaggle.com**

**www.androidheadline**

**www.wikipedia.com**

**www.geeksofgeeks.com**

**www.javatpoint.com**

[**www.edubca.com**](http://www.edubca.com)

**GitHub Link for the Project File**

**https://github.com/Vishnugarg897/Projects-Report**

20