**HACKATHON**

**DATA SCIENCE USING PYTHON**

**.**

**INDUSTRY-CONSTRUCTION**

**DEPARTMENT-FINANCE PAIN POINTS**

**ABSTRACT:**

**Construction :** Construction industry refers to the industrial branch of manufacturing and trade related to building, repairing, renovating, and maintaining infrastructures. It is a determinant of the country's technological and technical advancement, often regulating the growth of the country's infrastructural development that often directs to the country's advancement in terms of sustainability assurance.

**Finance :** Finance plays an important role in construction.It would be impossible for any construction company to survive without proper financial management and since the competition in the sector has grown rapidly and the profit margins are often slim, it plays a pivotal role.

**Links:**

**1.**[**https://www.sciencedirect.com/topics/earth-and-planetary-sciences/construction-industry**](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/construction-industry)

**2.**[**https://indieseducation.com/construction-financial-management/#:~:text=Financial%20Planning%20in%20construction%20management,-It%20is%20an&text=It%20is%20the%20process%20of,capital%2C%20capital%20expenditure%2C%20etc**](https://indieseducation.com/construction-financial-management/#:~:text=Financial%20Planning%20in%20construction%20management,-It%20is%20an&text=It%20is%20the%20process%20of,capital%2C%20capital%20expenditure%2C%20etc)**.**

1. **IDEATION**

**PROBLEM STATEMENT:**

**BUDGET PREDICTION OF CONSTRUCTION -** Budget management is a critical function for any organization, and involves the planning, organizing, directing, and controlling of financial resources in order to achieve the organization's goals. It includes activities such as budgeting, forecasting, financial reporting, and risk management.

➤This Model predicts the overall budget for the Construction of the project.

**PAIN POINT :**

As the client/customer cannot predict overall budget for the project.So this model takes the clear budget for each factor and predicts the overall budget of the project.so it is easy for the customer/client to manage his finance.

1.Budgeting and cost control.

**Link:** [**https://chat.openai.com/chat**](https://chat.openai.com/chat)

**OUTCOME OF THE PROJECT:**

To predict the overall budget for the project.Based on the observations from the features this model also predicts the factors which mostly influence the outcome.

1. **FEATURE ENGINEERING**

**1.FEATURE CREATION:** Factors that influence the outcome of a project.

Different features gathered by the team includes:

**Link**:<https://www.ferkeybuilders.com/3-big-financial-pain-points-for-construction-companies/>

1.project delays

2.material waste

3.labor shortage

**Link**: <https://preferredcfo.com/7-most-common-financial-mistakes-construction-companies-make/>

1.Invoicing Late & Missing Bank Draws

2.Fixed Material Cost in Bid Contracts

**Link:** [**https://chat.openai.com/chat**](https://chat.openai.com/chat)

**Searched for**: Find factors that can impact finance in the construction industry.

1.Economic conditions

2.Cost of materials

3.Labour costs

4.Government regulations

5.Competition

6.Availability of financing

7.Natural disasters and weather events

8.Technological advancements

9.cost overruns

10.Delays and Disruptions

11.Lack of transparency

12.Risk Management

13.Payment Disputes

14.Competition and pricing pressure.

15.project management

16.Data Management

17.Budget cost

18.Time

**2.FEATURE SELECTION**:Based on the feature creation the main features selected are:

1.Economic conditions

2.Cost of materials

3.Labour costs

4.Government regulations

5.Competition

6.Competition and pricing pressure.

7.project management

8.Data Management

9.Budget cost

10.Time

11.Technology Advancement

12.Payment Disputes.

13.Natural disasters and weather events.

14.Availability of financing

15.cost overruns

16.Lack of transparency

17.Risk Management.

**3.FEATURE EXTRACTION:**

After the selection the fields which come under the same category are grouped.

1.Interest rates.

2.Market Demand

Competition

3.Project Management

Regulation

Data Management

4.Material Cost

5.Labour cost.

6.Cost Overruns

7.Access to funding

8.Complexity

9.Time

10.Technology

11.Budget Prediction

**4.DIMENSION REDUCTION:**The following features are discarded:

1.Natural disasters and weather events- This feature is discarded as natural calamities are unpredictable. It is the value that is unpredictable by humans.

If the assumption that there is a natural calamity and consider the budget .Then if any natural calamity doesn’t happen the interest for the money from the bank is lost.

If assumption is that there will not be a natural calamity if that happens then there is huge loss of finance .

So due to these aspects are not at all predicted by humans so the factor is discarded.

Market Demand :

Competition: Market Demand and competition has similar meaning like Market Demand for construction increases competition for the construction of project increases. Competition for construction contracts can drive down margins and reduce profitability, especially in markets with an oversupply of contractors.

Project Management:

Regulation :Obtaining the necessary permits and complying with regulations can also impact the cost of a construction project. Delays in obtaining permits or changes in regulations can cause cost overruns, which can impact the budget for a project.

Data Management: Data Management includes storing the payment details and rules of the project.

As the three factors gives the same meaning these are grouped as one.

**Reasons for considering the above features:**

1.**Interest Rates:** The cost of raw materials such as steel, concrete, and lumber can greatly impact the financial viability of construction projects.

2.**Market Demand:** Land Demand and cost of the Land at a particular time.

3.**Project Management**:Perfect project Management with clearly defined rules can be completed on time. Inorder to avoid disputes about the finance among the clients and manager it is necessary to have clear project Management.

4.**Material Cost**: The cost of raw materials such as steel, concrete, and lumber can greatly impact the financial viability of construction projects.

5.**Technological advancements**: Technological advancements in construction can increase efficiency and reduce costs, but can also require significant investment to adopt and integrate into construction processes.

6.**Labour Costs**: The cost of labor, including wages, benefits, and other employment costs, is a major factor in the cost of construction projects.

7.**Cost Overruns**: Cost overrun is the increased cost compared to the actual cost of construction as per the cost estimation in the initial project planning.

8.**Access to funding**: It also depends whether the money for the project is easily sanctioned or not .If the money for the project is not easily sanctioned then it increases the time .As the time increases if the cost of materials increases finance increases.

9.**Complexity**: Construction projects are complex and can involve many different stakeholders, including architects, engineers, contractors, and suppliers. This complexity can make it difficult to secure financing and can increase the risk of miscommunication and misunderstandings.

10.**Time:** As the time increases , material demand and labor demand increases so the finance increases.

11**.Budget Prediction**: Based on the above features we calculate the overall cost for the project.

**Link :** [**https://chat.openai.com/chat/cf4c134d-e8b7-4ebb-b101-9025b8acbfd3**](https://chat.openai.com/chat/cf4c134d-e8b7-4ebb-b101-9025b8acbfd3)

**5.FEATURE SCALING :**

1.**INTEREST\_RATES**:

0 - determines no Loan is taken from the bank.

1.1% Interest for the finance

2. 2% Interest.

5. 5% Interest for the loan

**2.MARKET\_DEMAND:**

**1->**Low demand for the land in the particular area.

2->Moderate demand for the land in the particular area.

3->High demand for the land in the particular area.

**3.MATERIAL\_COST:**

0->No Material cost

30 -> determines high material cost

**4.LABOR COST:**

0-> Labor cost

20-> determines high labor cost

**5.COST OVERRUNS:**

1->determines least wastage and extra finance.

10->determines high wastage of materials.

**6.ACCESS\_TO\_FUNDING:**

1->determines fast sanction of money.

2->determine the sanction of money slowly.

**7.COMPLEXITY:**

1->determine low infrastructure.

3->determines moderate level infrastructure.

5->determine high level infrastructure.

**8.TIME:**

1->determine one month.

24-> denotes maximum completion of a project.

**9.TECHNOLOGY:**

1->Modern Technology.

2->old Technology

**10.PROJECT BUDGET**:

Numerical value. That denotes the overall budget of the project.

**11.CLASSIFICATION:**

2->High Budget

1->Low Budget

**3.DATASET GENERATION**

Values are based on the real data.Data is taken from the survey .

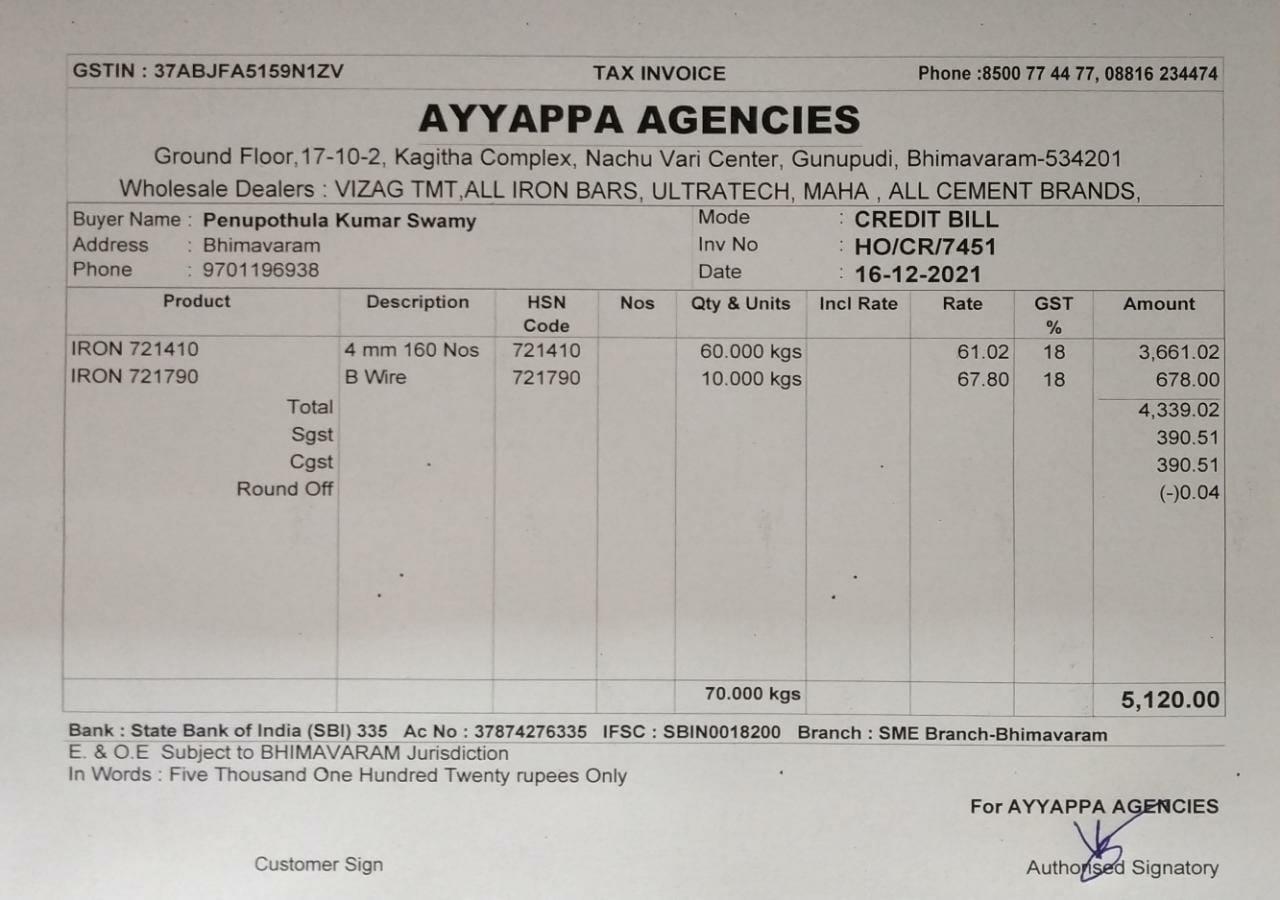
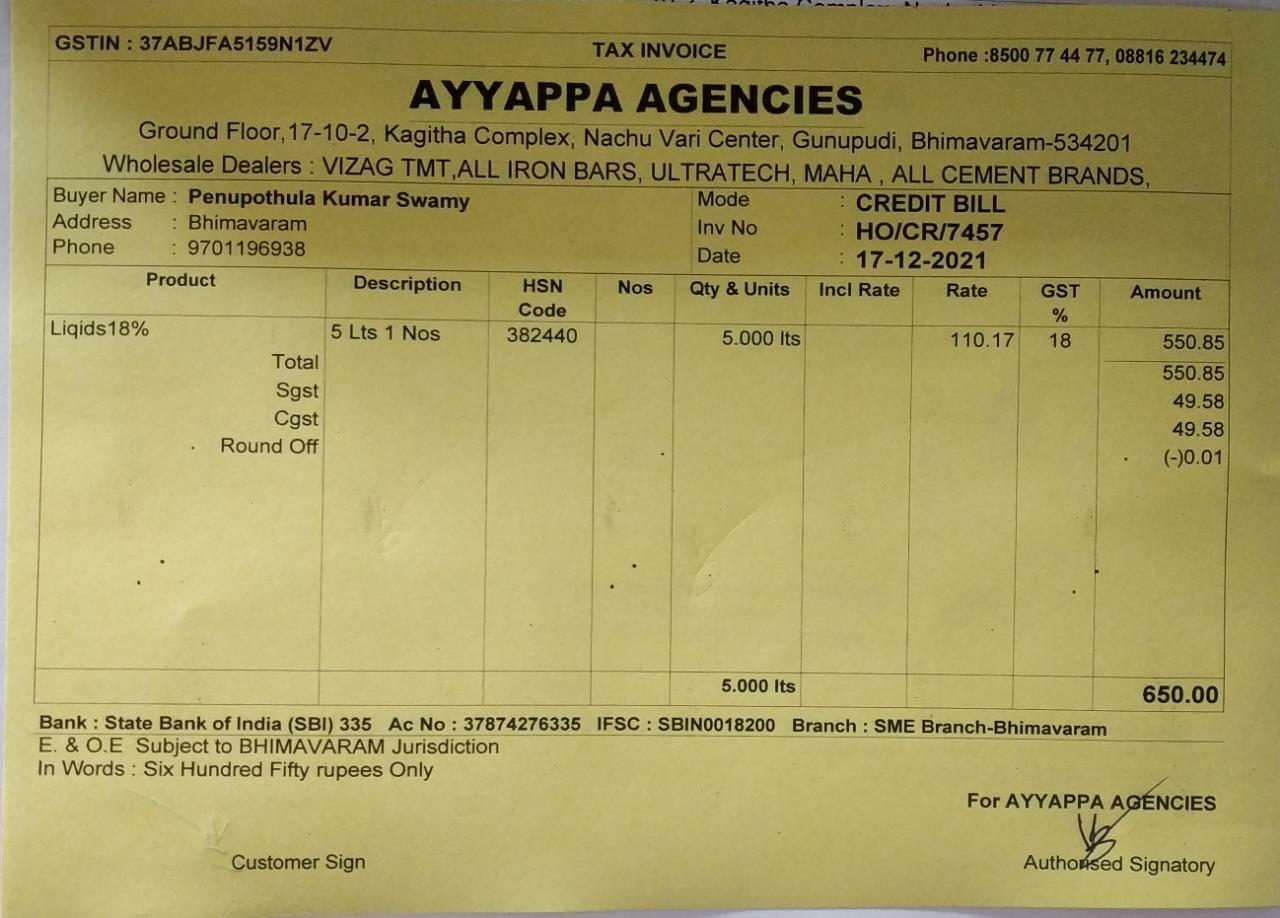
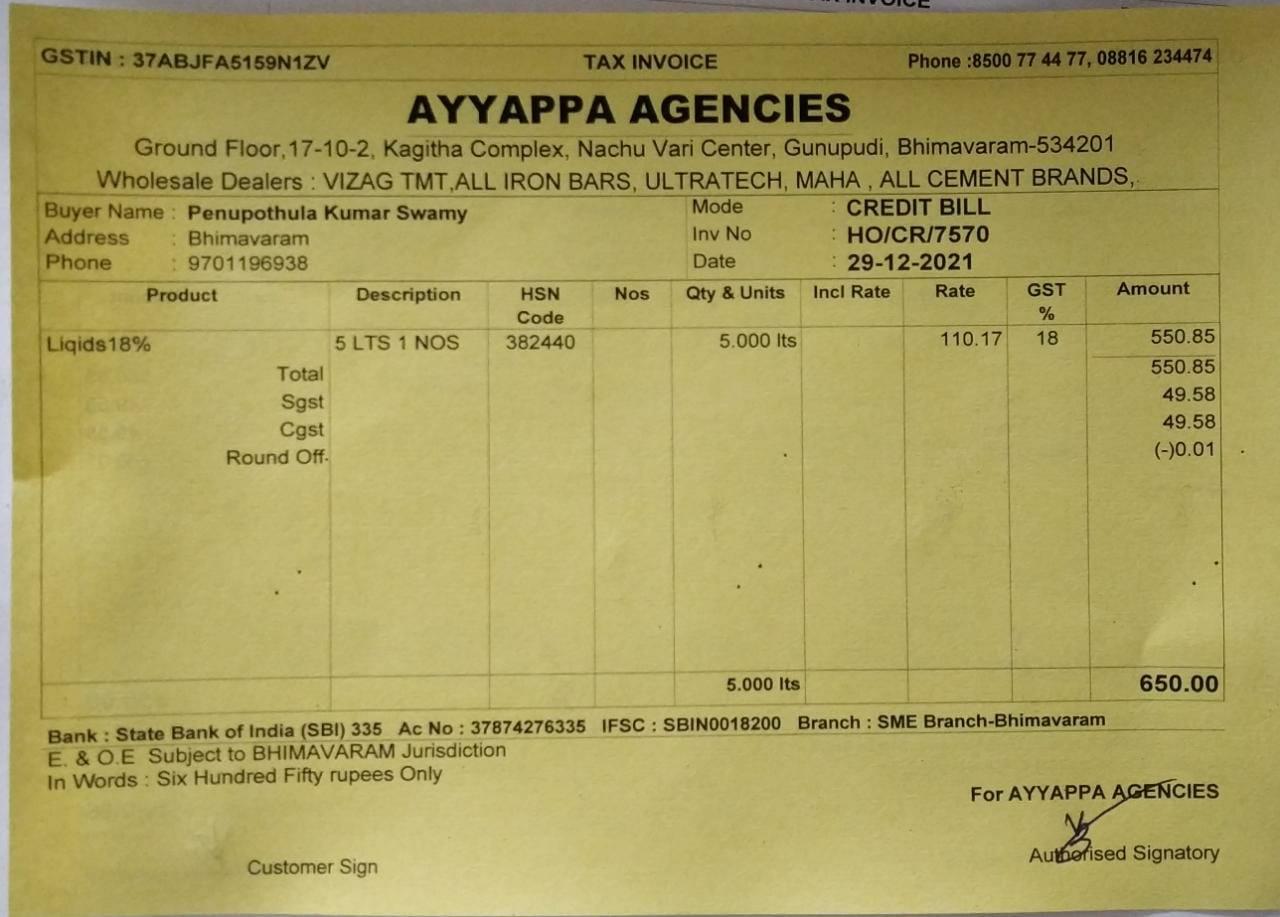
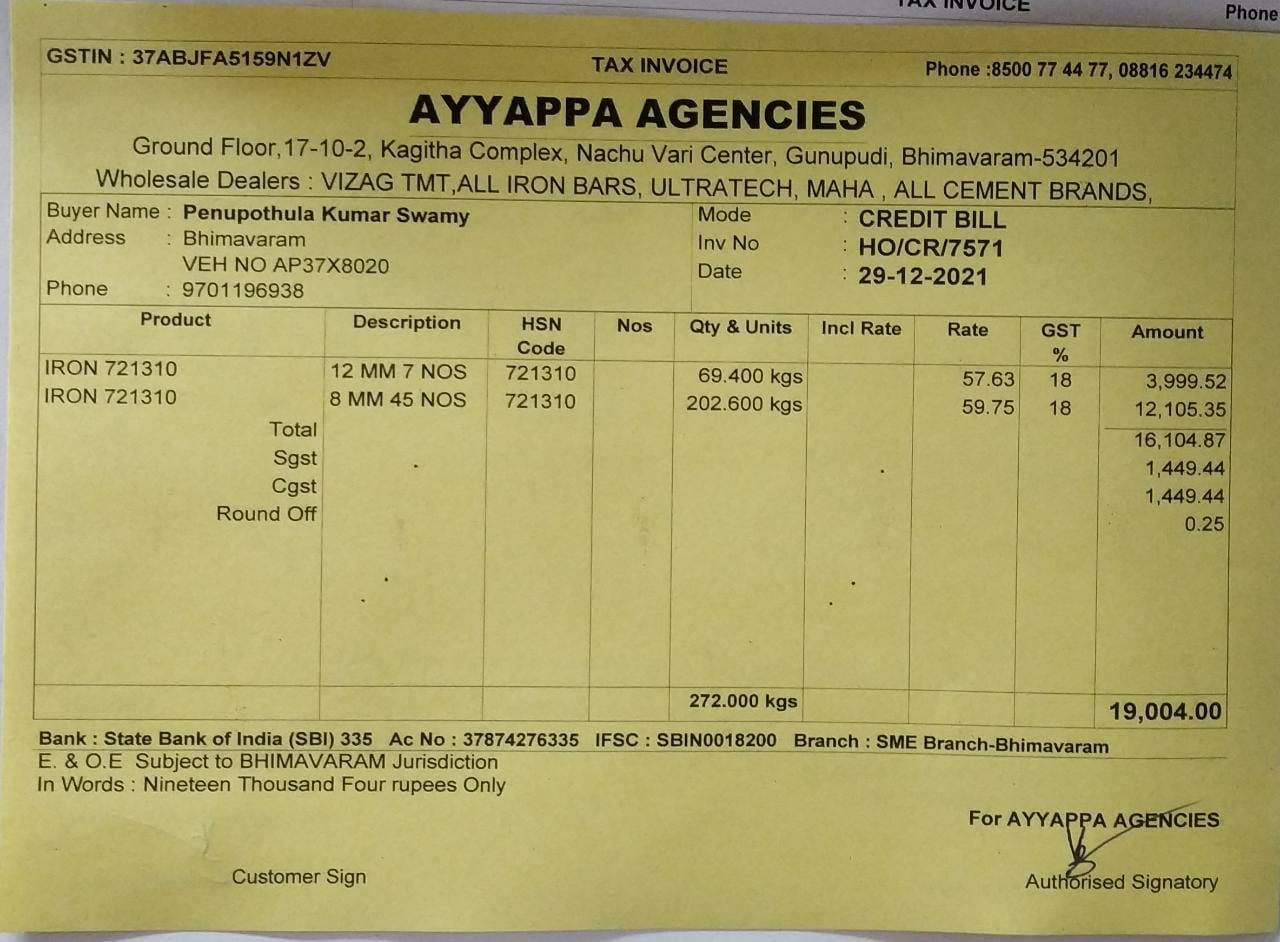
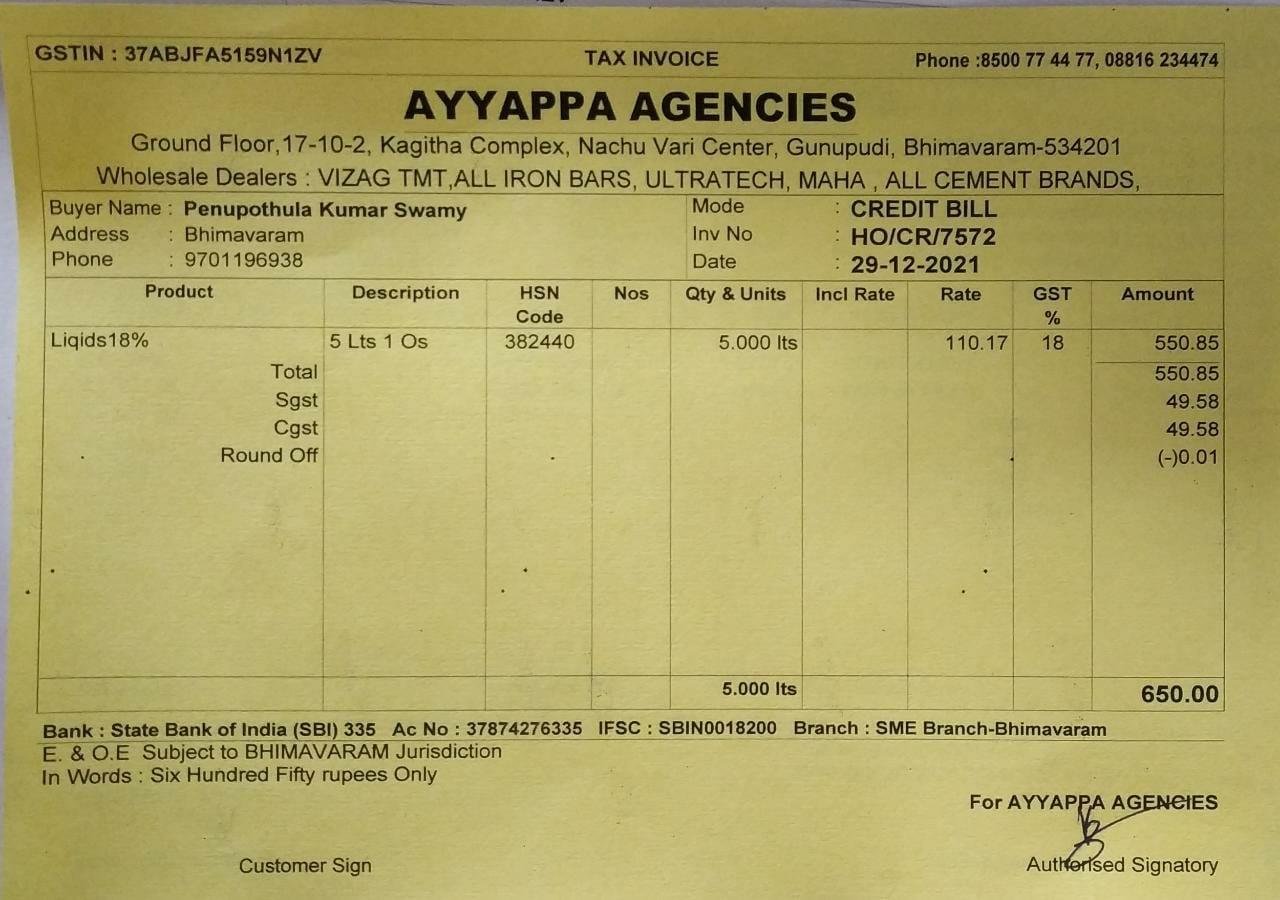
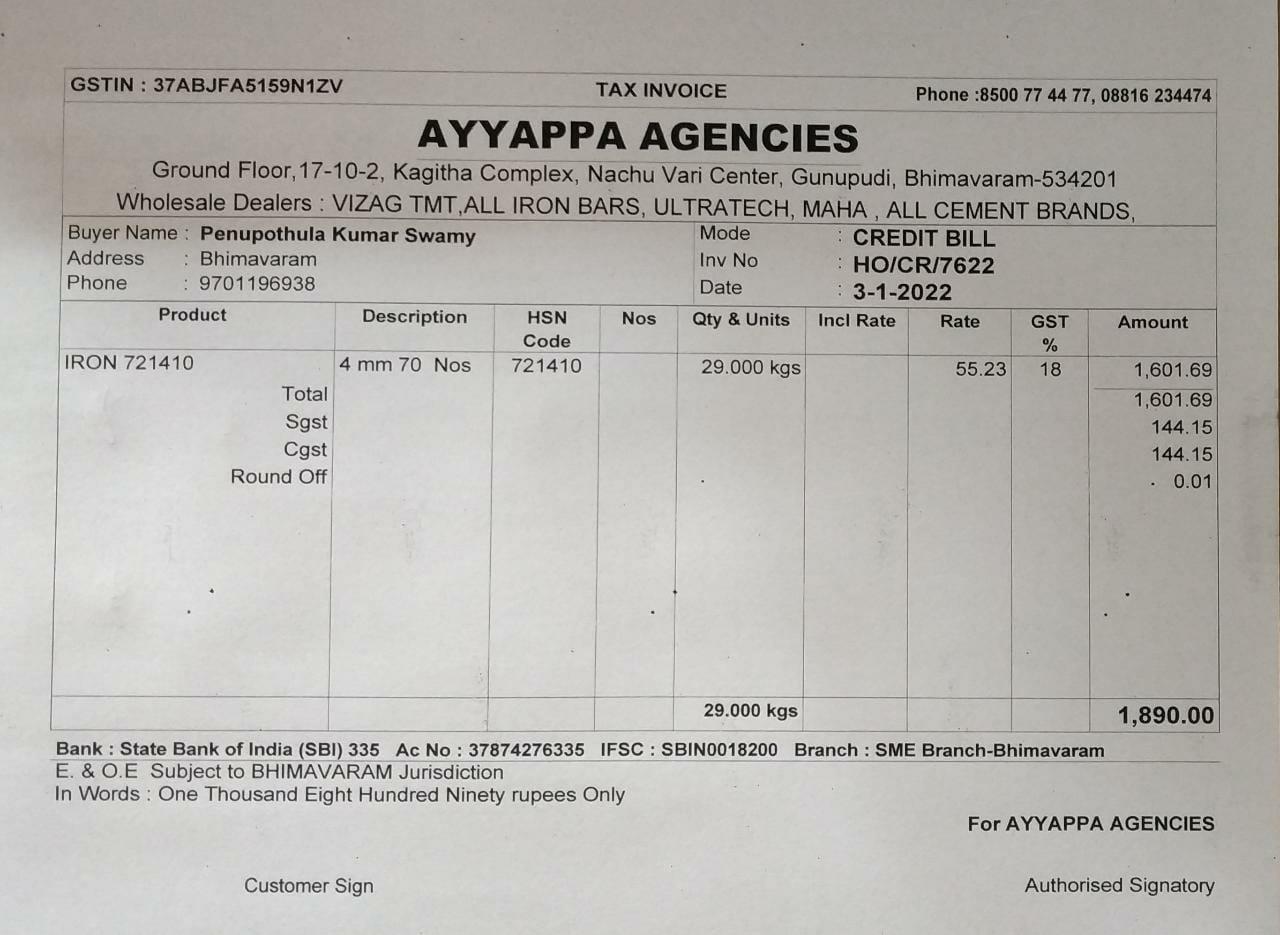
**The Real data includes :**

<https://drive.google.com/file/d/1Pi6ZrJF6fT-bBH80CmQnDSwGwwvsHX-l/view?usp=share_link>

<https://drive.google.com/file/d/1Pi6ZrJF6fT-bBH80CmQnDSwGwwvsHX-l/view?usp=share_link>

<https://drive.google.com/file/d/1i1_h2OSQ-Ecu90FwnUzQrwDatD2J4b1E/view?usp=share_link>

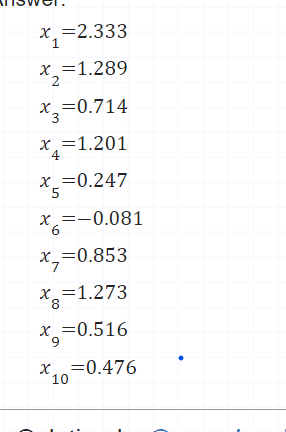
**Photo proofs of the data :**

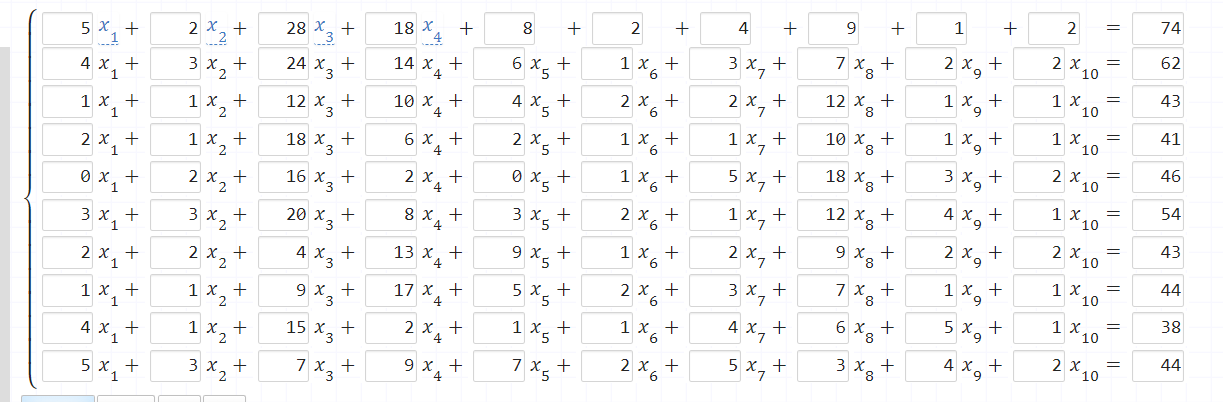


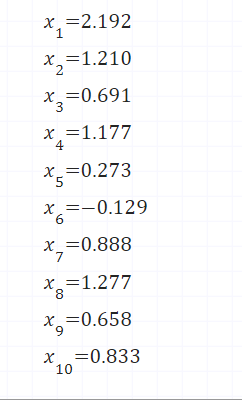
The values after the multiple iteration are :

**Link:**matrixcalc.org\slu.html

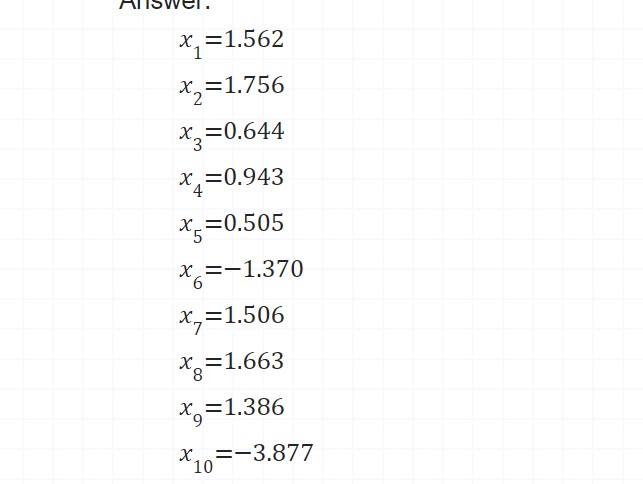
**1st Iteration :**



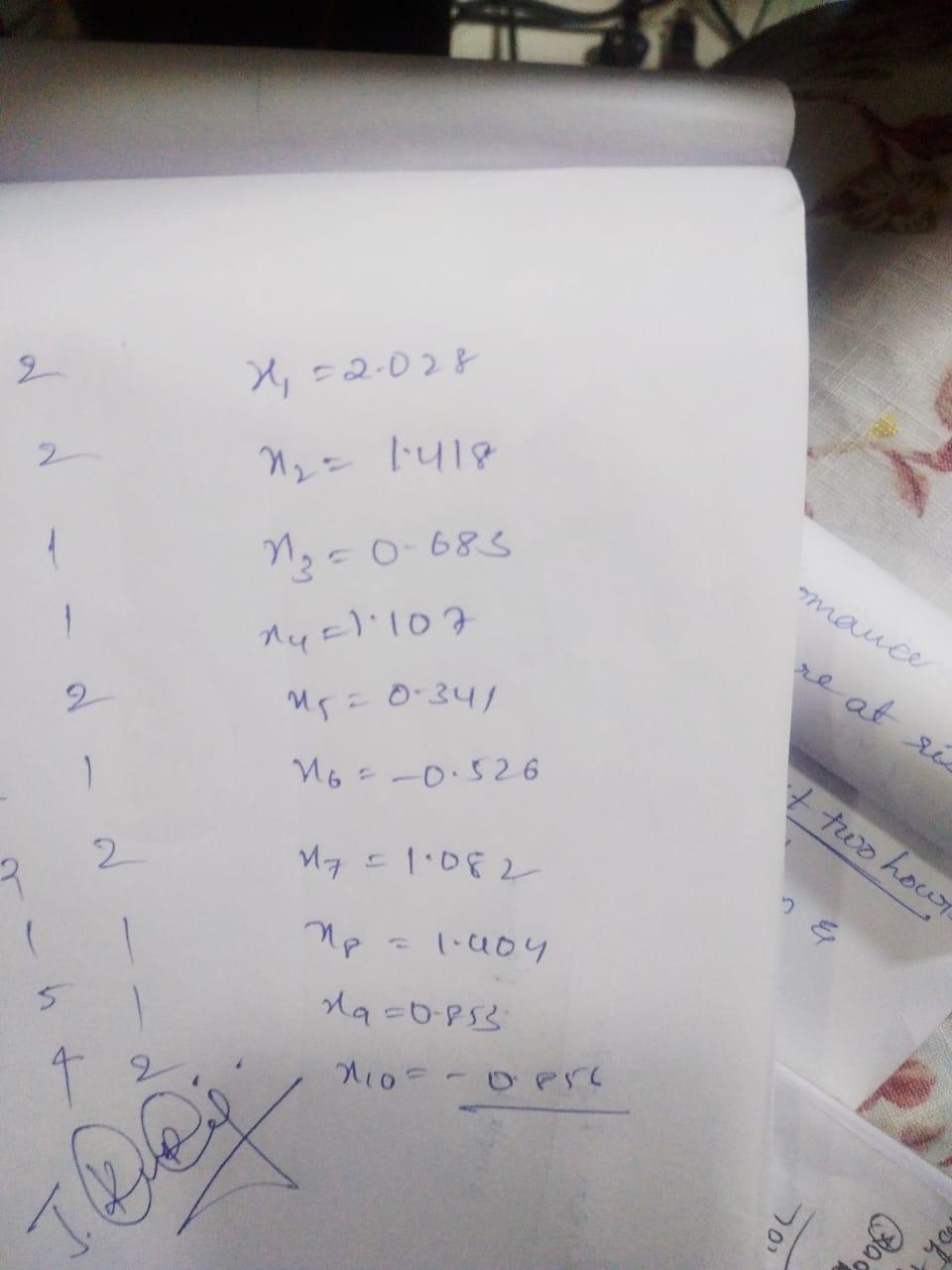
**2nd Iteration:** 



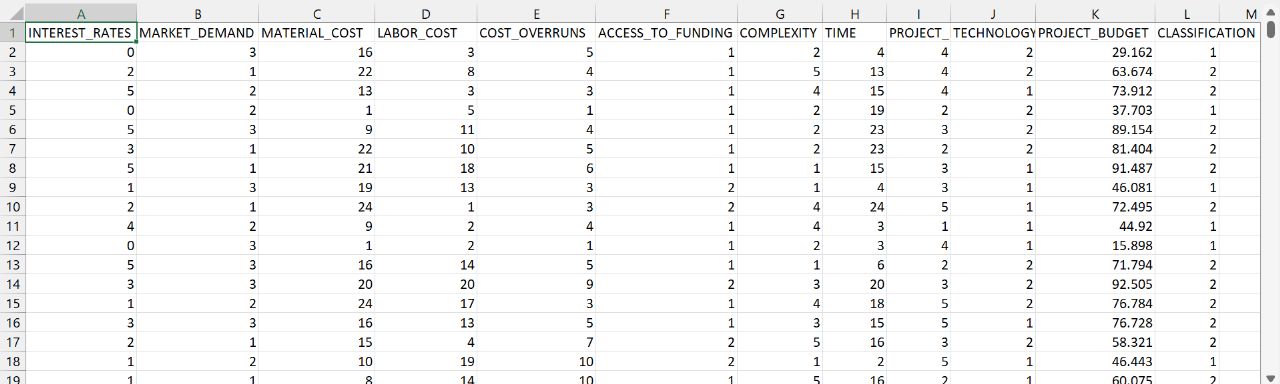
**3rd Iteration :**



Final Values after Mean :



**Code for Generation of dataset:**

****

**4.MODEL SELECTION**

Target variable which the model is predicting is -**Random Forest**.

Why Random Forest ?

As the Linear Regression is concerned there is a more chance of overfitting. So in order to avoid overfitting linear regression is avoided.

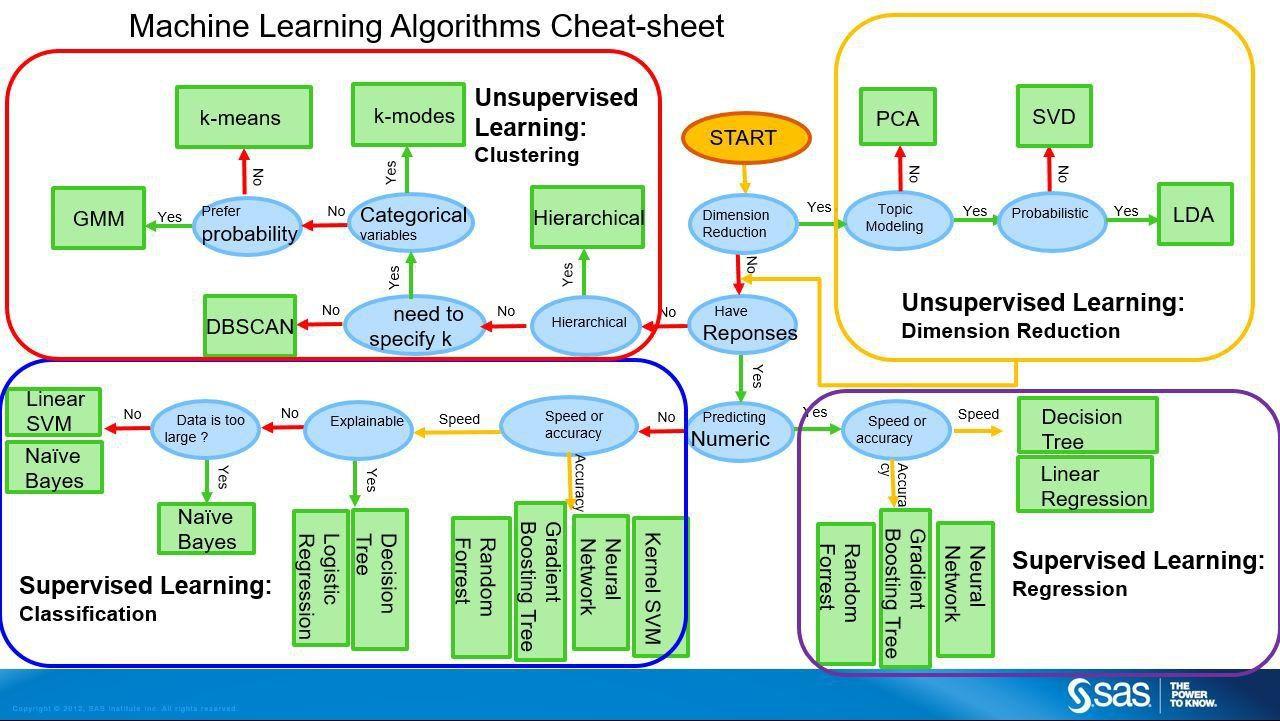
Linear regression makes predictions for continuous/real or numeric variables.

In linear regression overfitting occurs when the model is "too complex". This usually happens when there are a large number of parameters compared to the number of observations. Such a model will not generalize well to new data.

As the Budget Prediction is continuous **Logistic Regression** is not considered .It is only useful binary valued data.

**KMeans:** As it makes the clusters of the data .This model predicts the continuous variable so the K Means model is discarded.

**Naive Bayes :** It predicts the wrong output in some cases. It is most suitable for categorical data. So this model is not taken into consideration.

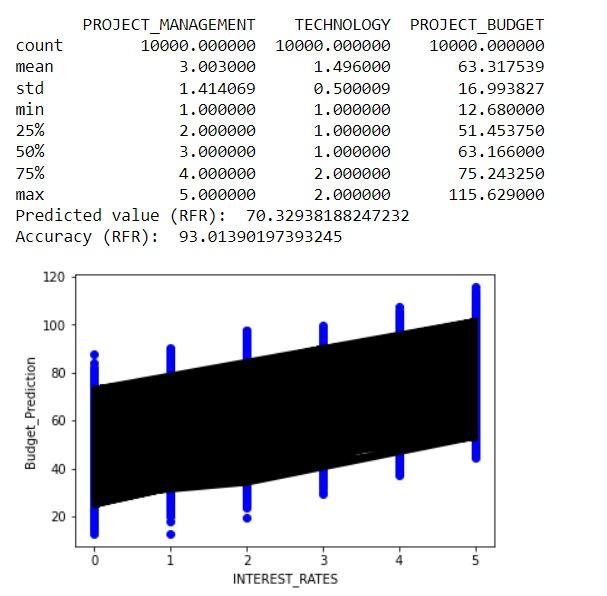


**5.CODE GENERATION AND OBSERVATIONS**

**CODE:** Random Forest



**OUTPUT :**

****

Describe function gives the summary of the code .

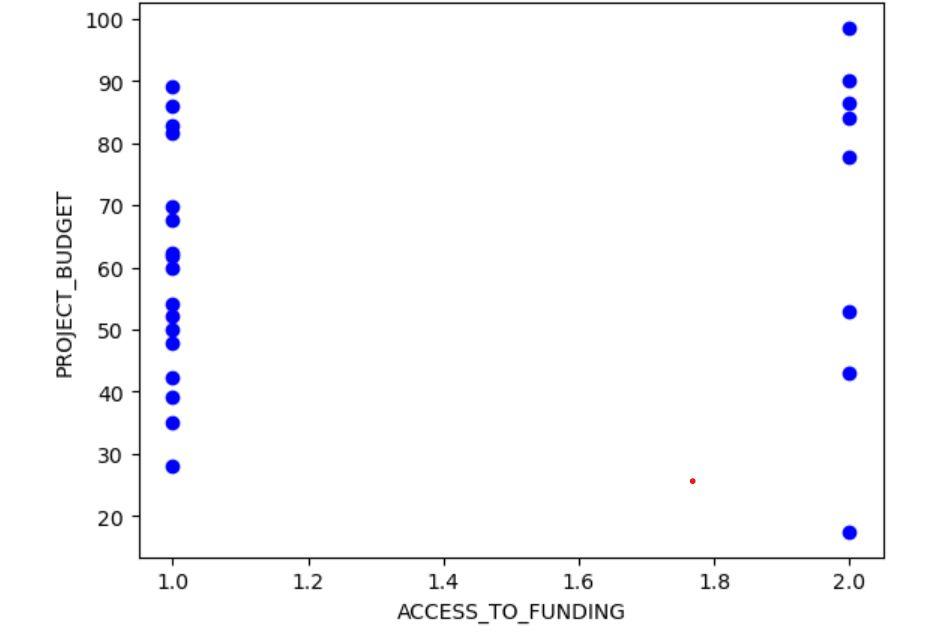
X indicates the data without the prediction variable.

Y indicates the target variable.

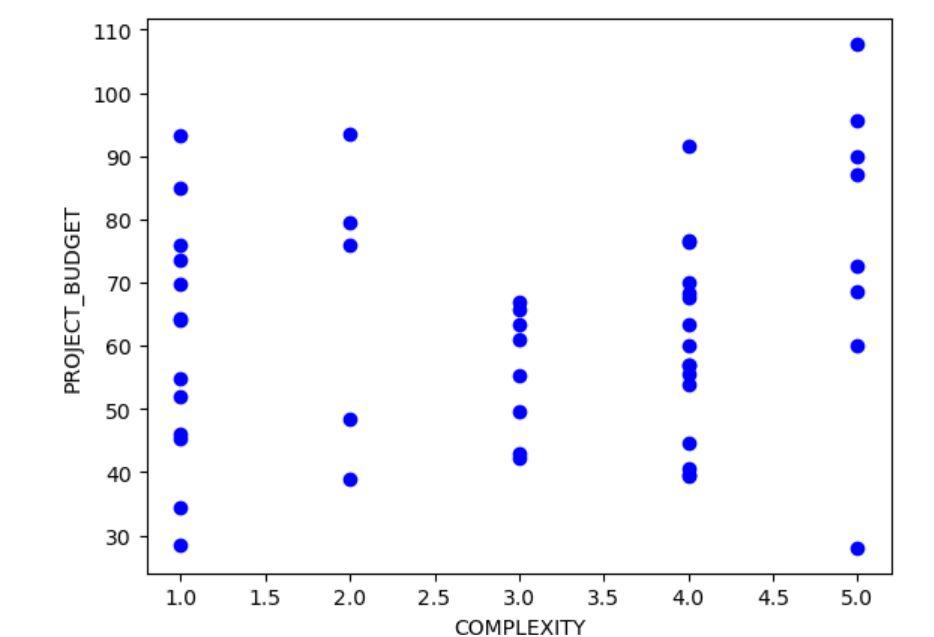
Importing the libraries of Random Forest Regression.

Scatter plot gives the graphical representation of data.

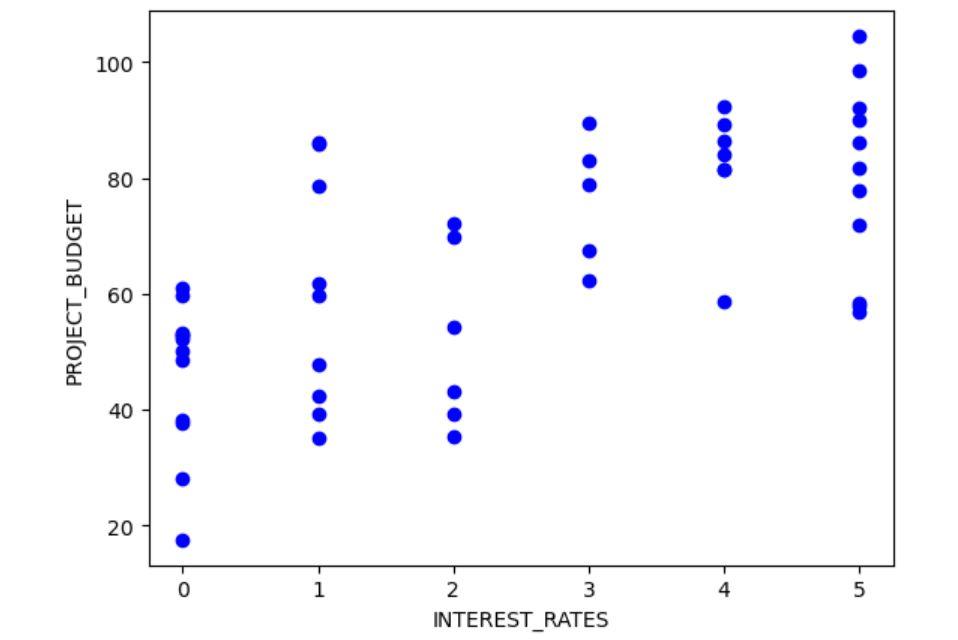
**OBSERVATIONS :**

**A**

**ACCESS TO FUNDING VS PROJECT\_BUDGET :** If the sanctioning of money becomes late then it affects the budget like increase in cost of labor and material.

****

**COMPLEXITY VS PROJECT\_BUDGET :** As the complexity like interior design , area ,infrastructure is more complex budget cost increases. In order to construct the building.

****

**INTEREST\_RATES VS PROJECT\_BUDGET:** As the interest rate increases, the budget increases in order to pay the loans.