

# **CTC Prediction System**

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE DEGREE OF

**BACHELOR OF ENGINEERING**

IN

**INFORMATION TECHNOLOGY**

BY

**Nelson Kolas**

**Divyajothi Raja**

**Khizar Shaikh**

**Samuel Pallikonda**

UNDER THE GUIDANCE OF

**Ms. Sulochana Devi**

(Department of Information Technology)



**INFORMATION TECHNOLOGY DEPARTMENT**

**XAVIER INSTITUTE OF ENGINEERING**

**UNIVERSITY OF MUMBAI**

**2023 – 2024**



**XAVIER INSTITUTE OF ENGINEERING**

Mahim, Mumbai 400016

**Department of Information Technology**

(NBA Accredited)

(Approved by AICTE, Govt. of Maharashtra and Affiliated to University of Mumbai)

**CERTIFICATE**

This to certify that

**Nelson Kolas (2022032007)**

**Divya Jothi Raja (202103011)**

**Khizar Shaikh (2022032002)**

**Samuel Pallikonda (2022032004)**

Have satisfactorily carried out the MINI-PROJECT work titled “**CTC Prediction System**” in partial fulfillment of the degree of Bachelor of Engineering as laid down by the University of Mumbai during the academic year 2022-2023.

**Internal Examiner/Guide**

**External Examiner**

**Date:**

**Place: MAHIM, MUMBAI**

## DECLARATION

I declare that this written submission represents my ideas in my own words and where others' Ideas or words have been included, I have adequately cited and referenced the original sources.

I also declare that I have adhered to all the principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which thus have not been properly cited or from whom proper permission have not been taken when needed.

**Nelson Kolas (2022032007)**

-----

**Divya Jothi Raja (2022032005)**

-----

**Khizar Shaikh (2022032002)**

-----

**Samuel Pallikonda (2022032004)**

-----

Date:

## TABLE OF CONTENTS

<b>SR. NO</b>	<b>TOPIC</b>	<b>PAGE NO.</b>
A.	LIST OF FIGURES	i
B.	ACKNOWLEDGEMENT	ii
C.	COURSE OUTCOMES	iii
	Rubric page for outcomes	
1.	1.1 Introduction. 1.2 Problem Statement. 1.3 Objectives of the project 1.4 Scope of the project 1.5 Features of the project 1.6 Constraints on Project. 1.7 GUI of webpages	1 2 3 4 6 9 11
2.	Implementation Methodology 2.1 Hardware, Software Requirements.  2.2 Code	14  19
3.	REFERENCES	

## LIST OF FIGURES

SR NO.	FIGURE CAPTION	PAGE NO.
1		11
2.		11
3.		12
4.		12
5.		13

## Acknowledgement

We would like to thank Fr. Dr. John Rose S.J (Director of XIE) for providing us with such an environment so as to achieve goals of our project and supporting us constantly.

We express our sincere gratitude to our Honorable Principal Dr. Y.D.Venkatesh for encouragement and facilities provided to us.

We would like to place on record our deep sense of gratitude to Dr. Jaychand Upadhyay, Head of Dept Of Information Technology, Xavier Institute of Engineering, Mahim, Mumbai, for her generous guidance help and useful suggestions.

With deep sense of gratitude, we acknowledge the guidance of our project guide **Prof. Sulochana Devi**.

The time-to-time assistance and encouragement by her has played an important role in the development of our project.

We would also like to thank our entire Information Technology staff who have willingly cooperated with us in resolving our queries and providing us all the required facilities on time.

**Nelson Kolas**

-----

**Divya Jothi Raja**

-----

**Khizar Shaikh**

-----

**Samuel Pallikonda**

-----



# XAVIER INSTITUTE OF ENGINEERING

Mahim, Mumbai 400016

## Department of Information Technology

(NBA Accredited)

(Approved by AICTE, Govt. of Maharashtra and Affiliated to University of Mumbai)

CLASS: TE

SEM: V

COURSE CODE: ITM 501

COURSE NAME: Miniproject- 2A- Machine Learning & Artificial Intelligence

AY: 2023-24

SUB IN-CHARGE: Ms. Sulochana Devi

**Course Objectives:** The course aims:

1. To acquaint with the process of identifying the needs and converting it into the problem.
2. To familiarize the process of solving the problem in a group.
3. To acquaint with the process of applying basic engineering fundamentals to attempt solutions to the problems.
4. To inculcate the process of self-learning and research.

**Bloom's Taxonomy Levels:**

**1 = Remembering, 2= Understanding, 3 = Applying, 4 = Analyzing, 5 = Evaluating, 6 = Creating**

COURSE OUTCOMES:

Sr. No	Course Outcome	Cognitive levels of attainment as per Bloom's Taxonomy
On successful completion of course, learner/student will be able to:		
ITM 501.1	<b>Identify</b> problems based on societal /research needs and <b>apply</b> knowledge and skill to solve these problems in a group.	L1, L2, L3
ITM 501.2	<b>Use</b> standard norms of engineering practices to <b>analyse</b> the impact of solutions in societal and environmental context for sustainable development.	L3, L4
ITM 501.3	<b>Examine</b> interpersonal skills and <b>ethical</b> awareness to work as member of a group or leader and <b>demonstrate</b> capabilities of self-learning in a group, which leads to lifelong learning.	L2, L3, L4
ITM 501.4	<b>Build</b> project management principles during project work and <b>excel</b> in written and oral communication.	L5, L6

Project Guide Sign



# XAVIER INSTITUTE OF ENGINEERING

Mahim, Mumbai 400016

## Department of Information Technology

(NBA Accredited)

(Approved by AICTE, Govt. of Maharashtra and Affiliated to University of Mumbai)

**CLASS/ SEM: TEIT / V**

**AY: 2023-24**

**Subject: Mini-Project-2A- Machine Learning & Artificial Intelligence**

### **Guidelines for Mini-Project Report**

- A. It should contain Title Sheet, Internal Approval Sheet (Certificate), Approval Sheet, Declaration, LO List, Abstract, Tables of Contents, List of Figures, List of Tables, Abbreviation, Notation and Nomenclature in the same order and followed by following chapters:

#### **Rubrics page**

#### **1. Introduction**

- 1.1 Background** (Need for developing the website).
- 1.2 Problem Definition** (Give the problem definition presenting the current state of art.)
- 1.3 Objectives of the project**
- 1.4 Scope of the project** (Define the scope of the project.)
- 1.5 Features of the project**
- 1.6 Constraints on Project.**
- 1.7 GUI of webpages**
- 1.8 Test Cases/ Validation** (Screenshots of errors / validation checking)

#### **2. Implementation Methodology** This is the main body of the report and should contain

- 2.1 Design** (Block diagram, Flowchart of proposed system,)
- 2.2 Hardware, Software Requirements.**
- 2.3 ER Diagram for the website (if project has database)**
- 2.4 Module Implementation** (Explain how you have implemented the system step by step with a brief explanation of each step)
- 2.5 Database Connectivity**
- 2.6 Code**  
Insert Code for Homepage, Registration page and Database Connectivity
- 2.7 Steps to Launch the website on Internet.**





# **XAVIER INSTITUTE OF ENGINEERING**

**Mahim, Mumbai 400016**

## **Department of Information Technology**

(NBA Accredited)

(Approved by AICTE, Govt. of Maharashtra and Affiliated to University of Mumbai)

### **Appendix (If any)**

### **References**

(Give a list of references that were helpful in completing this prototype. References may include research papers, books, blogs, and YouTube videos etc. Do proper citation in the report.)

**Publications** Certificates and papers published need to be attached.

### **Acknowledgement**

Acknowledge the guide / department / college

**Name of Project Guide & Sign**

**Ms. Sulochana Devi**  
**Mini Project Co-ordinator**

# **1. INTRODUCTION**

In today's dynamic job market, where transparency and fairness in salary negotiations are paramount, a salary prediction system employing machine learning techniques emerges as a pivotal tool. By harnessing vast datasets encompassing diverse factors such as educational background, skill sets, industry standards, and regional variations, this system offers a sophisticated approach to estimating compensation for entry-level positions. Through meticulous data preprocessing and feature engineering, the system transforms raw information into meaningful predictors, facilitating accurate salary predictions. Machine learning algorithms, trained on historical salary data and validated through rigorous evaluation processes, learn intricate patterns and relationships, enabling them to make informed predictions. Upon deployment, this system empowers both job seekers and employers, providing clarity and insight into fair compensation practices. Moreover, its iterative nature allows for continuous refinement, ensuring that predictions remain aligned with evolving market trends. Ultimately, by fostering transparency and data-driven decision-making, this innovative system contributes to a more equitable and efficient job market for all stakeholders involved. By leveraging data analytics, this system evaluates diverse factors including skills, knowledge, academic achievements, and industry trends to estimate an individual's worth to an organization accurately. Such a data-driven approach not only enhances transparency but also ensures fair and competitive compensation for entry-level positions, benefitting both employers and prospective employees in navigating the intricacies of salary negotiations.

## **2. Problem Statement**

1. **Skill Set Evaluation:** Existing systems often fail to thoroughly evaluate an individual's skill set, including both technical and soft skills, which are crucial determinants of job performance and value to the organization.
2. **Logical Thinking Assessment:** The ability to think critically and solve problems logically is a fundamental aspect of many roles, yet current methods may overlook or inadequately measure this competency.
3. **Technical Knowledge Consideration:** In today's technology-driven world, technical expertise is highly valued. However, traditional systems may not sufficiently account for variations in technical skills and their impact on salary determination.
4. **Domain Expertise Recognition:** Different domains require specific knowledge and expertise. Existing models may lack the granularity to accurately assess the value of domain expertise, resulting in disparities in salary allocations.
5. **Background and Experience:** A person's background, including prior work experience and industry exposure, significantly influences their suitability for a role and, consequently, their salary expectations. However, this aspect is often oversimplified or overlooked in current systems.
6. **Academic Achievement Integration:** While academic qualifications are considered, the depth of their impact on salary determination may vary. Some systems may not appropriately weigh academic achievements against other factors, leading to skewed outcomes.
7. **Inadequate Coverage of Essential Dimensions:** Existing methods may fail to encompass all essential dimensions necessary for a comprehensive salary assessment, resulting in incomplete or inaccurate evaluations.
8. **Risk of Inaccuracy and Dissatisfaction:** When critical aspects such as skills, knowledge, and background are not adequately accounted for, there's a heightened risk of inaccuracies in salary determinations, leading to dissatisfaction among employees and potential talent attrition.
9. **Need for a Comprehensive Approach:** Given the complexity of today's job market, there's a clear need for a more comprehensive and nuanced approach to salary determination that considers all relevant factors to ensure fairness, transparency, and accuracy in compensation allocations.

## **2.1 Objectives of the project**

- 1. Finding Relationship between skills and salary**
- 2. Predicting salary based on skillset.**
- 3. Filling gaps, by increasing accuracy of existing models.**
- 4. Increasing confidence and self-esteem of students and employers**  
**give me objectives in points.**
  - 1. Enhance Predictive Accuracy:**
    - Refine existing models by incorporating more granular skill data and utilizing advanced machine learning techniques.
    - Explore ensemble methods to combine multiple models for improved accuracy.
    - Regularly update the models with new data to capture evolving skill trends and market dynamics.
  - 2. Education and Training:**
    - Develop educational resources and workshops aimed at students, job seekers, and employers to improve understanding of the factors influencing salary determination.
    - Provide guidance on acquiring and developing in-demand skills that correlate with higher salaries, thereby empowering individuals to make informed career choices.
  - 3. Feature Engineering:**
    - Conduct in-depth analysis to identify additional relevant features beyond skillset, such as experience level, education, industry, and geographic location.
    - Utilize natural language processing (NLP) techniques to extract insights from job descriptions and resumes, potentially uncovering latent features that impact salary

## 2.2 Scope of the Project

The scope of a CTC (Cost to Company) prediction system project typically encompasses several key aspects:

1. **Data Collection:** Gather comprehensive data on employees' profiles, including their skills, experience, education, role, and current salary. Additionally, collect data on external factors such as industry benchmarks, market trends, and economic indicators.
2. **Feature Engineering:** Identify relevant features that may influence CTC, such as job role, experience level, location, educational qualifications, performance metrics, and specialized skills.
3. **Data Preprocessing:** Cleanse and preprocess the collected data, handling missing values, outliers, and inconsistencies. Perform data transformations and encoding categorical variables as necessary for modeling.
4. **Model Selection:** Choose appropriate machine learning or statistical modeling techniques for CTC prediction, considering factors like interpretability, scalability, and prediction accuracy. Common models include regression analysis, decision trees, random forests, and neural networks.
5. **Model Training:** Train the selected model using historical data, where the features represent employee profiles and the target variable is the corresponding CTC. Utilize techniques like cross-validation to assess model performance and avoid overfitting.
6. **Model Evaluation:** Evaluate the trained model's performance using relevant evaluation metrics such as mean absolute error (MAE), root mean square error (RMSE), or coefficient of determination (R-squared). Compare the model's predictions against actual CTC values to assess accuracy.
7. **Deployment and Integration:** Integrate the trained model into a user-friendly interface or application, allowing stakeholders such as HR professionals, managers, and executives to input employee profiles and receive predicted CTC estimates. Ensure scalability, reliability, and security of the deployed system.
8. **Monitoring and Maintenance:** Implement mechanisms for monitoring the model's performance in production, including tracking prediction errors and recalibrating the model periodically with new data. Address any drift in data distributions or model degradation over time to maintain prediction accuracy.

**9. Documentation and Training:** Provide comprehensive documentation of the CTC prediction system, including model architecture, data sources, preprocessing steps, and deployment instructions. Offer training sessions or materials to users on how to effectively utilize the system for decision-making.

**10. Ethical Considerations:** Consider ethical implications such as fairness, transparency, and privacy in developing and deploying the CTC prediction system. Ensure that the system does not perpetuate biases or discrimination based on protected characteristics such as gender, race, or age.

By defining the scope of the project across these dimensions, organizations can develop a robust CTC prediction system that provides valuable insights for workforce planning, compensation management, and talent acquisition strategies.

## 1.4 Features of the Project

### 1. Employee Profile Features:

- **Experience Level:** Years of relevant work experience.
- **Education:** Highest level of education attained (e.g., high school, bachelor's degree, master's degree, PhD).
- **Skills and Qualifications:** Specific skills and qualifications relevant to the job role.
- **Certifications:** Any professional certifications or licenses held by the employee.
- **Performance Metrics:** Past performance evaluations or ratings, if available.
- **Job Role/Title:** Specific job title or role within the organization.
- **Department/Division:** The department or division where the employee works.
- **Location:** Geographic location of employment, which may influence salary due to cost-of-living differences.
- **Industry:** The industry in which the organization operates, as salary levels can vary across industries.

### 2. External Factors:

- **Market Trends:** Trends in the job market and industry-specific salary benchmarks.
- **Economic Indicators:** Macroeconomic factors such as inflation rates, unemployment rates, and GDP growth that may impact salary trends.
- **Regional Factors:** Local labor market conditions and regulatory environment.

### 3. Historical Salary Data:

- **Previous Salary:** Employee's historical salary or compensation package.
- **Salary Changes Over Time:** Any changes or adjustments in salary over the employee's tenure.
- **Promotions or Bonuses:** Occurrence of promotions, bonuses, or other compensation-related events.

**4. Feedback and Reviews:**

- **Employee Feedback:** Feedback from performance reviews, 360-degree assessments, or employee satisfaction surveys.
- **Managerial Feedback:** Ratings or feedback provided by managers or supervisors.

**5. Market Demand for Skills:**

- **Trending Skills:** Identification of in-demand skills or emerging technologies that may impact salary trends.
- **Skill Scarcity:** Assessment of skill shortages or areas where specific skills command higher salaries.



## 1.5 Constraint of Project

- 1. Data Availability and Quality:** Limited availability of comprehensive and accurate data on employee profiles, historical salaries, and relevant external factors can constrain the development and accuracy of the prediction model.
- 2. Privacy and Confidentiality:** Adherence to privacy regulations and organizational policies regarding the handling of employee data may limit access to certain types of data or require anonymization techniques to protect sensitive information.
- 3. Bias and Fairness:** Addressing potential biases in the data or model that could lead to unfair or discriminatory outcomes, particularly related to factors such as gender, race, age, or other protected characteristics.
- 4. Model Complexity vs. Interpretability:** Balancing the complexity of the prediction model with the need for interpretability and transparency, as overly complex models may be difficult to understand and explain to stakeholders.
- 5. Scalability and Performance:** Ensuring that the prediction system is scalable to handle large volumes of data and can provide timely predictions without compromising performance.
- 6. Regulatory Compliance:** Compliance with legal and regulatory requirements, such as labor laws, equal pay regulations, and data protection laws, which may impose constraints on how salary prediction models are developed and deployed.
- 7. Resource Constraints:** Limited resources in terms of budget, time, and expertise for developing, training, and maintaining the prediction system, which may impact the scope and complexity of the project.
- 8. Feedback and Iteration:** Incorporating feedback from stakeholders and iterating on the model based on new data or changing requirements while operating within project constraints.
- 9. Model Performance Evaluation:** Challenges in accurately evaluating the performance of the prediction model, particularly in real-world settings where multiple factors can influence salary outcomes.

By identifying and addressing these constraints early in the project lifecycle, organizations can better manage risks and optimize the development and deployment of the salary prediction system.

## **1.6 GUI of the Project**

- **Home Page:**

### **1.6.1 Home Page**

## 2.1 Hardware and Software Requirements

- **Hardware Requirements:**

1. RAM : 2gb minimum
2. Processor: Intel Core i3 onwards(Windows), Android 4 minimum(Mobile)
3. HDD space: 2gb minimum

- **Software Requirements:**

1. Windows Operating System, Mobile Operating System
2. Google Chrome, Microsoft Edge, Mozilla Firefox
3. Python- For Development of Model.
4. Libraries & Packages- NumPy, Pandas, Seaborn, Matplotlib, Sklearn, Scikit-learn.
5. Google Colab & VS Code-Code Editors.
6. Front-End-HTML, CSS, JavaScript.
7. Back-End-Node JS/Django
8. Algorithms-Decision Tree Classifier, Naïve Bayes Classifier, Linear Regression, Polynomial Regression.

## **2.2 Activity Diagram**

## **2.3 Module Implementation**

## **2.4 Database Connectivity**

## 2.5 Code

```
import pandas as pd

# Load the dataframe from CSV file, explicitly specifying data types
data3 = pd.read_csv("/content/drive/MyDrive/Mini Project Data Science/Main Mini Project/CTC
DataSets/Engineering_graduate_salary.csv")

# Check the data types of all columns
# print(data3.dtypes)
print(data2.dtypes)

testingData = data2.head(10)
testingData=testingData.drop(columns=['Salary'])
testingData.head(10)
data2.head(10)

import xgboost as xgb
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt

# Prepare the data
X2 = data2.drop(columns=['Salary'])
y2 = data2['Salary']

# Split the data into training and testing sets
X2_train, X2_test, y2_train, y2_test = train_test_split(X2, y2, test_size=0.2, random_state=42)

# Define XGBoost regressor
xgb_reg =

xgb.XGBRegressor(objective ='reg:squarederror', random_state=42)

# Train the model
xgb_reg.fit(X2_train, y2_train)

# Make predictions on the test set
y2_pred = xgb_reg.predict(X2_test)

# Evaluate the model
mse = mean_squared_error(y2_test, y2_pred)
print("Mean Squared Error:", mse)

xgb_predict = xgb_reg.predict(testingData)
for salary in xgb_predict:
```

```
print(salary)
```

```
import pickle
```

```
# Save the model to a file
```

```
with open('xgb_model.pkl', 'wb') as f:
```

```
    pickle.dump(xgb_reg, f)
```



## **2.6 Steps to Launch Website**

### **1. Planning and Strategy:**

Define your target audience: Identify the farmers and consumers you want to target with your platform.

Set clear goals and objectives: Determine what you want to achieve with the website.

Develop a business plan and budget: Outline your revenue model, funding requirements, and financial projections.

Choose the technology stack: Select the programming languages, frameworks, and tools for website development.

### **2. Domain Name and Hosting:**

Register a domain name that reflects your brand (e.g., ecogreenmarketplace.com).

Choose a reliable web hosting provider to host your website. Consider factors like performance, scalability, and security.

### **3. Website Development:**

Design the user interface (UI) and user experience (UX) for the website.

Develop the website using web development technologies. You can use web development frameworks like Django, Ruby on Rails, or build a custom solution.

Implement user registration and authentication systems for farmers and consumers.

Create product listing and management features for farmers.

Build a shopping interface for consumers to browse and purchase products.

Develop a review and feedback system.

Ensure the website is mobile-responsive for a seamless user experience on different devices.

Implement security measures to protect user data and transactions.

Set up a content management system (CMS) to manage educational resources and articles.

### **4. Database Setup:**

Create a relational database to store data related to farmers, products, consumers, reviews, farms, certifications, and other relevant information.

Define tables and establish relationships between them.

Implement CRUD operations for data management.

### **5. Quality Assurance and Testing:**

Thoroughly test the website for functionality, usability, and security.

Identify and fix any bugs or issues.

Conduct user testing to gather feedback and make improvements.

### **6. Content Creation:**

Create high-quality content for product listings, farmer profiles, educational resources, and blog articles.

Ensure that content is accurate, informative, and engaging.

### **7. SEO and Optimization:**

Optimize the website for search engines (SEO) to improve its visibility in search results.

Implement on-page SEO techniques, including keyword optimization and meta tags.

**8. Payment Gateway Integration:**

Integrate a secure payment gateway to facilitate transactions between consumers and farmers.

**9. Launch and Marketing:**

Set a launch date for your website.

Develop a marketing strategy to promote the launch and attract users.

Consider social media marketing, email marketing, and partnerships with local communities.

**10. User Support and Feedback Mechanism:**

Offer customer support channels for user inquiries and assistance.

Encourage user feedback and reviews to build a community-driven ecosystem.

**11. Legal and Compliance:**

Ensure compliance with agricultural and e-commerce regulations in the regions where EcoGreen operates.

**12. Monitor and Iterate:**

Continuously monitor website performance, user behavior, and sales metrics.

Gather user feedback to make improvements and enhancements to the platform.

**13. Scaling and Expansion:**

Plan for scalability as your user base grows.

Explore opportunities for expanding your platform to serve more regions or offer additional services.

**14. Analytics and Reporting:**

Implement analytics tools to track website performance and user engagement.

Provide reports and insights to farmers to help them optimize their listings.

Launching a website is an ongoing process, and it's important to regularly update and improve it based on user feedback and changing market dynamics.

## ● Data Preprocessing and Modelling

```

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
data = pd.read_csv("../content/drive/MyDrive/Mini Project Data Science/Main Mini Project/CTC DataSets/Engineering_graduate_salary.csv")
data.head()

```

	ID	Gender	DOB	10percentage	10board	12graduation	12percentage	12board	CollegeID	CollegeTier	...	MechanicalEngg	ElectricalEngg	TelecomEngg	CivilEngg	conscientiousness	agreeableness	extraversion	neroticism
0	604399	f	1990-10-22	87.80	cbse	2009	84.00	cbse	6920	1	...	-1	-1	-1	-1	-0.1590	0.3789	1.2396	0.14590
1	988334	m	1990-05-15	57.00	cbse	2010	64.50	cbse	6624	2	...	-1	-1	-1	-1	1.1336	0.0459	1.2396	0.52620
2	301647	m	1989-06-21	77.33	maharashtra state board.pune	2007	85.17	amravati divisional board	9084	2	...	-1	-1	260	-1	0.5100	-0.1232	1.5428	-0.29020
3	582313	m	1991-05-04	84.30	cbse	2009	86.00	cbse	8195	1	...	-1	-1	-1	-1	-0.4463	0.2124	0.3174	0.27270
4	339001	f	1990-10-30	82.00	cbse	2008	75.00	cbse	4889	2	...	-1	-1	-1	-1	-1.4992	-0.7473	-1.0697	0.06223

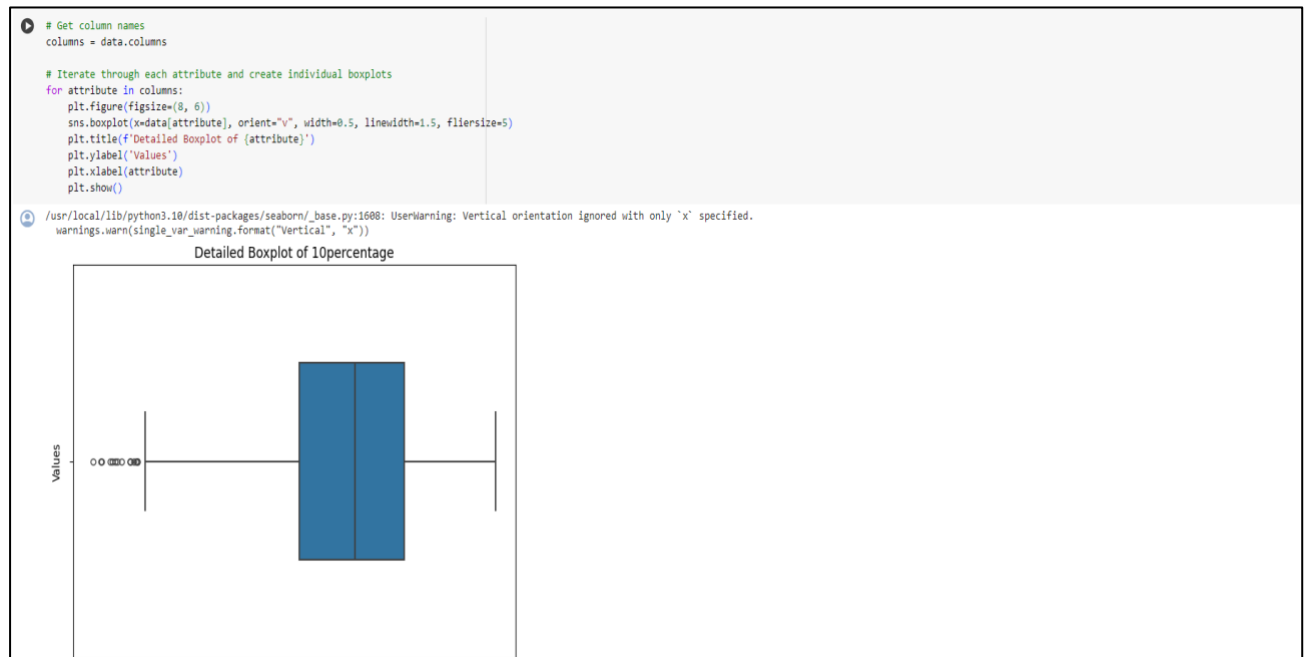
5 rows x 34 columns

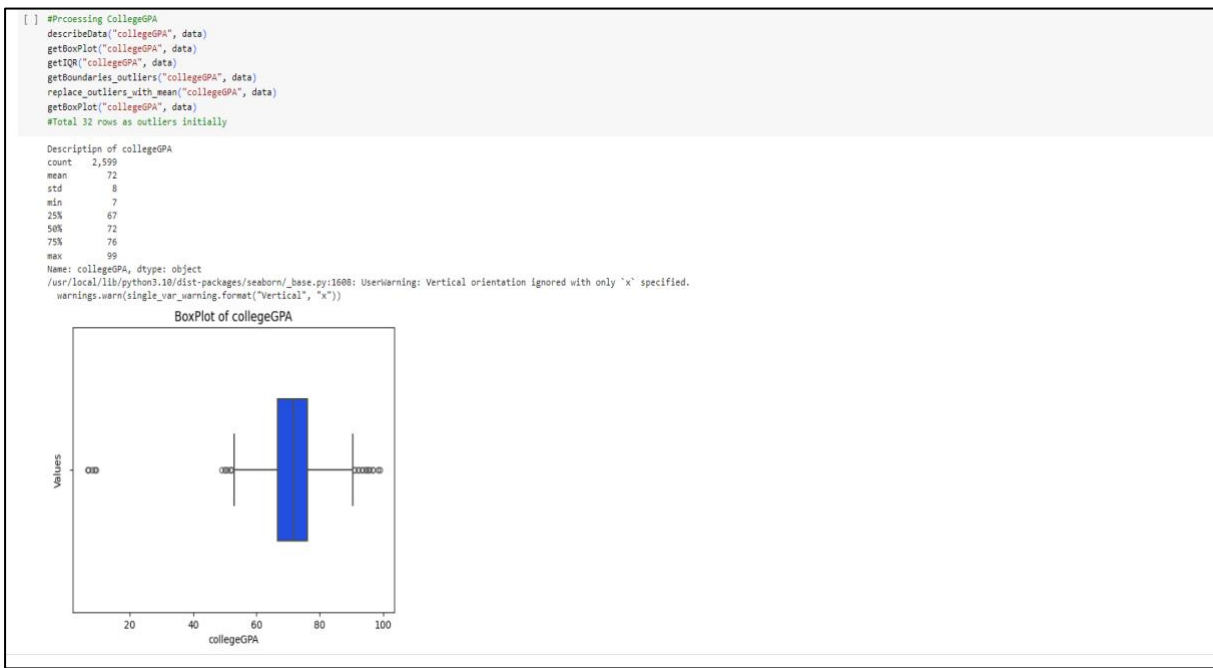
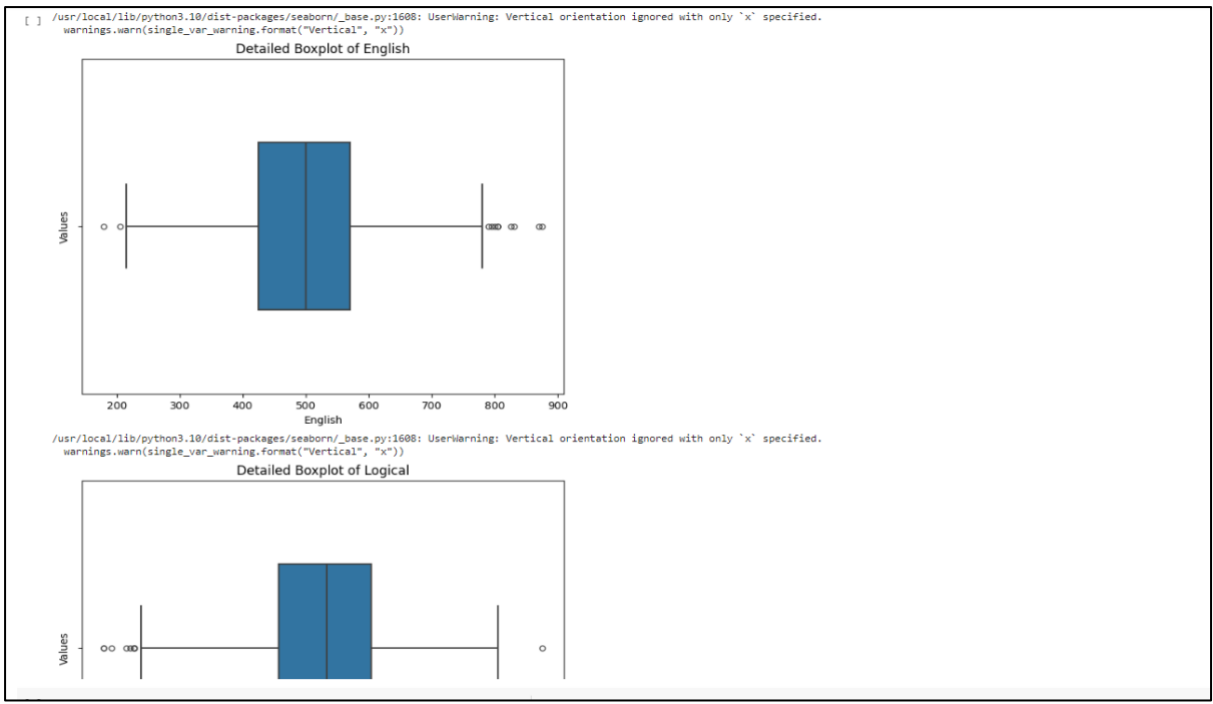
```

[ ] #removing unwanted attributes from the data
data = data.drop(columns=["ID", "DOB", "10board", "12graduation", "12board", "CollegeID", "CollegeTier", "CollegeCityID", "CollegeCityTier", "CollegeState", "GraduationYear", "Quant", "Domain", "conscientiousness", "agreeableness", "extraversion", "neroticism"])
data.head()

```

	Gender	10percentage	12percentage	Degree	Specialization	collegeGPA	English	Logical	ComputerProgramming	ElectronicsAndSemicon	ComputerScience	MechanicalEngg	ElectricalEngg	TelecomEngg	CivilEngg	Salary
0	f	87.80	84.00	B.Tech/B.E.	instrumentation and control engineering	73.82	650	665	485	366	-1	-1	-1	-1	-1	445000
1	m	57.00	64.50	B.Tech/B.E.	computer science & engineering	65.00	440	435	365	-1	-1	-1	-1	-1	-1	110000
2	m	77.33	85.17	B.Tech/B.E.	electronics & telecommunications	61.94	485	475	-1	400	-1	-1	-1	260	-1	255000
3	m	84.30	86.00	B.Tech/B.E.	computer science & engineering	80.40	675	620	655	-1	-1	-1	-1	-1	-1	420000





```
[ ] import pandas as pd

# Load the dataframe from CSV file, explicitly specifying data types
data3 = pd.read_csv("/content/drive/MyDrive/Mini Project Data Science/Main Mini Project/CTC DataSets/Engineering_graduate_salary.csv")

# Check the data types of all columns
# print(data3.dtypes)
print(data2.dtypes)
```

Gender int64  
10percentage float64  
12percentage float64  
Degree int64  
Specialization int64  
collegeGPA float64  
English float64  
Logical float64  
ComputerProgramming float64  
ElectronicsAndSemicon float64  
Salary float64  
dtype: object

```
[ ] testingData = data2.head(10)
testingData=testingData.drop(columns=['Salary'])
testingData.head(10)
data2.head(10)
```

	Gender	10percentage	12percentage	Degree	Specialization	collegeGPA	English	Logical	ComputerProgramming	ElectronicsAndSemicon	Salary
1	1	57.00	64.50	0	1	65.00	440.0	435.0	365.000000	336.41039	110000.0
2	1	77.33	85.17	0	0	61.94	485.0	475.0	454.278545	400.00000	255000.0
3	1	84.30	86.00	0	1	80.40	675.0	620.0	655.000000	336.41039	420000.0
7	0	77.00	88.00	0	1	71.00	370.0	470.0	465.000000	336.41039	105000.0
8	1	76.80	87.70	0	1	73.15	510.0	555.0	525.000000	336.41039	195000.0
9	0	57.00	73.00	0	1	70.08	500.0	410.0	385.000000	336.41039	200000.0
10	1	77.00	75.00	0	0	62.00	675.0	630.0	405.000000	260.00000	335000.0
12	0	85.00	88.00	0	1	85.00	495.0	445.0	485.000000	336.41039	480000.0
13	0	90.00	82.10	0	1	85.00	640.0	530.0	615.000000	336.41039	550000.0
14	1	86.40	86.20	0	0	81.40	720.0	630.0	485.000000	292.00000	325000.0

```
[ ] import xgboost as xgb
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
```

```
[ ] import xgboost as xgb
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt

# Prepare the data
X2 = data2.drop(columns=['Salary'])
y2 = data2['Salary']

# Split the data into training and testing sets
X2_train, X2_test, y2_train, y2_test = train_test_split(X2, y2, test_size=0.2, random_state=42)

# Define XGBoost regressor
xgb_reg = xgb.XGBRegressor(objective='reg:squarederror', random_state=42)

# Train the model
xgb_reg.fit(X2_train, y2_train)

# Make predictions on the test set
y2_pred = xgb_reg.predict(X2_test)

# Evaluate the model
mse = mean_squared_error(y2_test, y2_pred)
print("Mean Squared Error:", mse)

xgb_predict = xgb_reg.predict(testingData)
for salary in xgb_predict:
    print(salary)
```

Mean Squared Error: 16029888886.800943  
130766.78  
225334.48  
408507.5  
130849.16  
215095.25  
203637.2  
318318.66  
482767.88  
515273.06  
330181.97

