*Employee Salary Analysis*



**Name:**

Amit Kumar Verma

**Roll No:**

202401100400032  
**Course:**

Introduction to AI

**Problem Statement:**

Employee Salary Analysis – Explore correlations in employee salaries and positions with visualizations.

*Introduction*

This project focuses on analyzing employee salaries based on various factors such as position, experience, department, and gender. The goal is to explore correlations, detect patterns, and visualize insights using Python and data visualization libraries like Seaborn and Matplotlib.

Organizations often struggle with understanding salary distributions, disparities, and factors influencing compensation. This project aims to:

* Analyze salary distributions across different job positions and departments.
* Explore the impact of experience on salary.
* Identify potential salary inequalities.
* Provide data-driven insights for better decision-making.

*Methodology*

**Step 1: Data Collection**

* The dataset contains employee records with attributes such as Employee ID, Position, Department, Salary, Experience, and Gender.
* The dataset was loaded into Python using Pandas.

**Step 2: Data Preprocessing**

* Checked for missing values and handled them appropriately.
* Converted categorical data into a format suitable for analysis.

**Step 3: Exploratory Data Analysis (EDA)**

* **Statistical Analysis:** Summary statistics for salary.
* **Visualizations:**
  + Histogram & KDE Plot for salary distribution.
  + Box Plot to compare salaries across positions.
  + Heatmap for correlation analysis.
  + Bar Plot for department-wise salary expenditure.
  + Scatter Plot to analyze salary vs. experience.
  + Count Plot for the number of employees per position.
  + Pairplot for analyzing relationships between numerical features.

**Step 4: Interpretation**

* Identified trends in salary distribution.
* Analyzed how experience impacts salary.
* Detected salary gaps across different job roles.

Code

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

**# Import different libraries for statistical calculations**

**# Load the dataset (Make sure the CSV file is in the same directory)**

data = pd.read\_csv("Employee\_Salary\_Dataset.csv")

**# Basic salary statistics (Gives us an idea of the salary distribution)**

print("\nSalary Statistics:")

print(data['Salary'].describe()) **# Provides summary statistics for salary data**

**# Checking the basic info about the dataset**

print(data.info())

print(data.head())

**#  Visualizing Salary Distribution (Histogram + KDE for smoothness)**

plt.figure(figsize=(8, 5))

sns.histplot(data['Salary'], bins=30, kde=True, color='blue')

plt.title("Salary Distribution")

plt.xlabel("Salary")

plt.ylabel("Frequency")

plt.show()

**# Checking for missing values**

print("Missing Values:\n", data.isnull().sum())

**# Printing all th ecoloumns in the data**

print(data.columns)

plt.figure(figsize=(12, 6))

sns.boxplot(x='Experience\_Years', y='Salary', data=data, palette='coolwarm')

plt.xticks(rotation=45)

plt.title("Salary by Experience\_Years")

plt.xlabel("Experience\_Years")

plt.ylabel("Salary")

plt.show()

**#   Making a correlation Matrix for the salary and the Experience of the employees**

correlation\_matrix = data[['Salary', 'Experience\_Years']].corr()

plt.figure(figsize=(6, 4))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm', fmt=".2f")

plt.title("Correlation Matrix")

plt.show()

**# Comparing the Employees Salaries on the basis of their experience**

plt.figure(figsize=(8, 5))

sns.scatterplot(x='Experience\_Years', y='Salary', data=data, hue='Age', palette='deep')

plt.title("Salary vs. Experience")

plt.xlabel("Years of Experience")

plt.ylabel("Salary")

plt.legend(title="Age", bbox\_to\_anchor=(1, 1))

plt.show()

**# Using PairPlot**

print(data.info())

sns.pairplot(data, hue='Gender')

plt.show()

**# Comparing the salary of the emplyees on the basis of their ages**

plt.figure(figsize=(12, 6))

sns.boxplot(x='Age', y='Salary', data=data, palette='coolwarm')

plt.xticks(rotation=45)

plt.title("Salary by Age")

plt.xlabel("Age")

plt.ylabel("Salary")

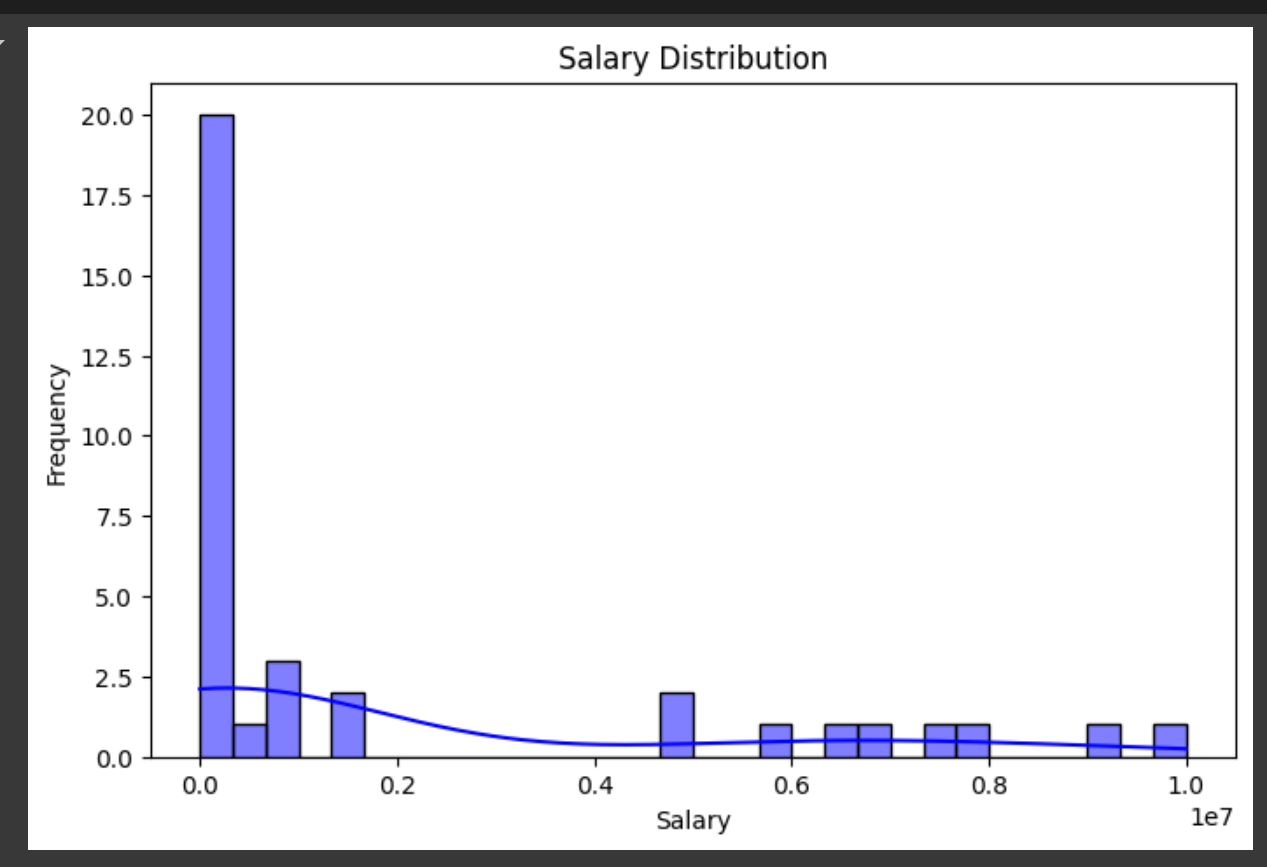
plt.show()

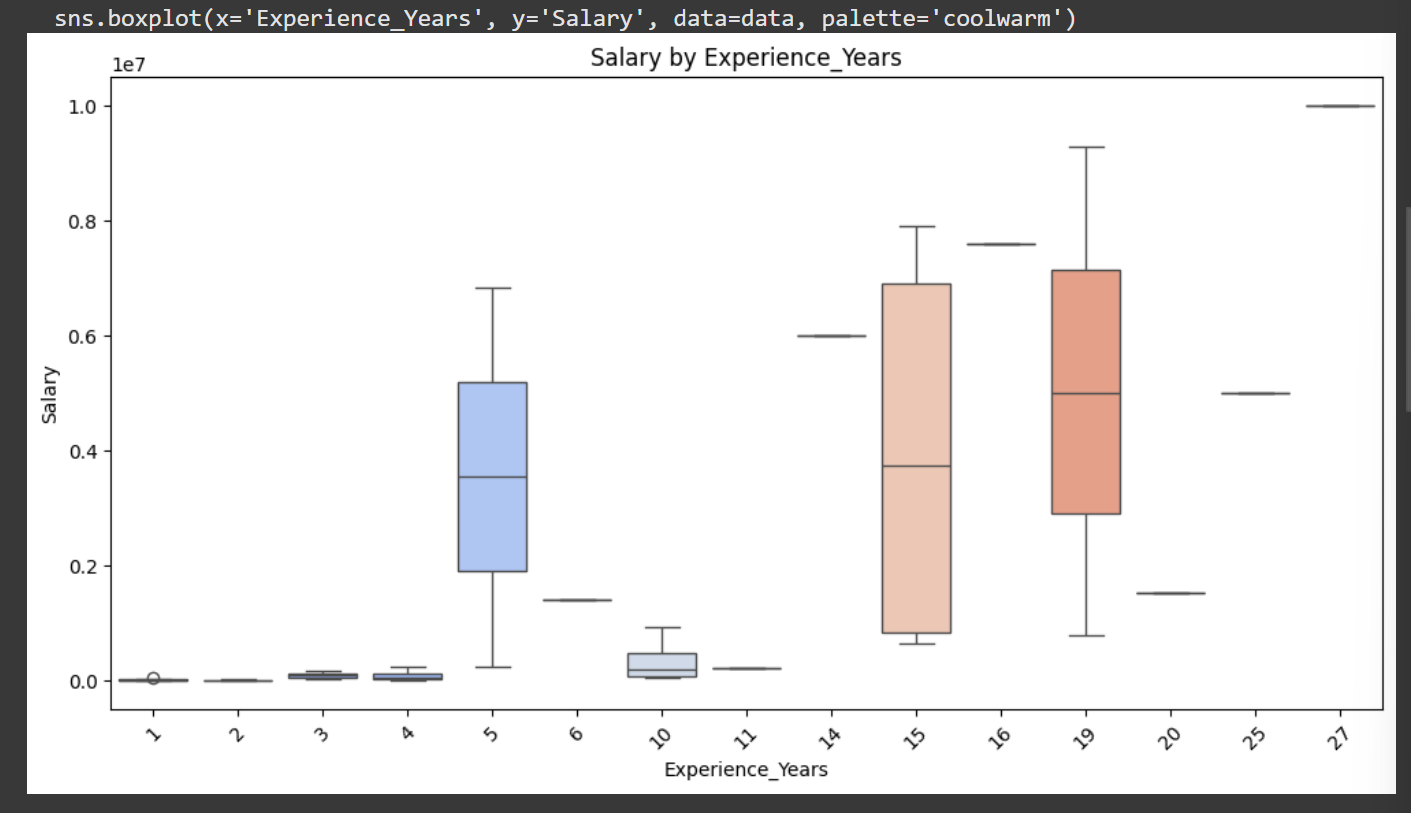
**# title sepal\_length**

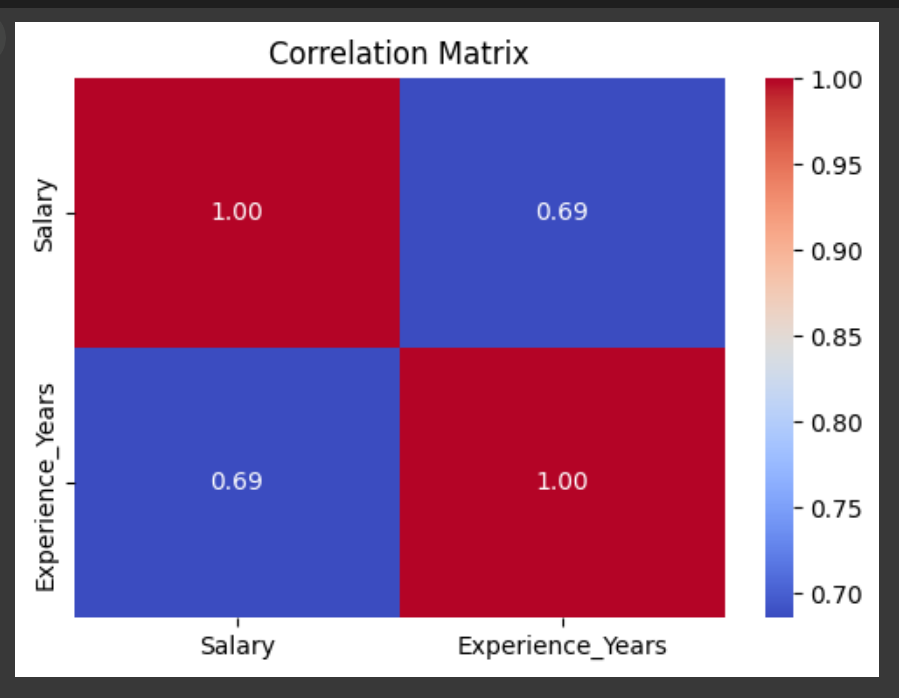
data['Salary'].plot(kind='line', figsize=(8, 4), title='sepal\_length')

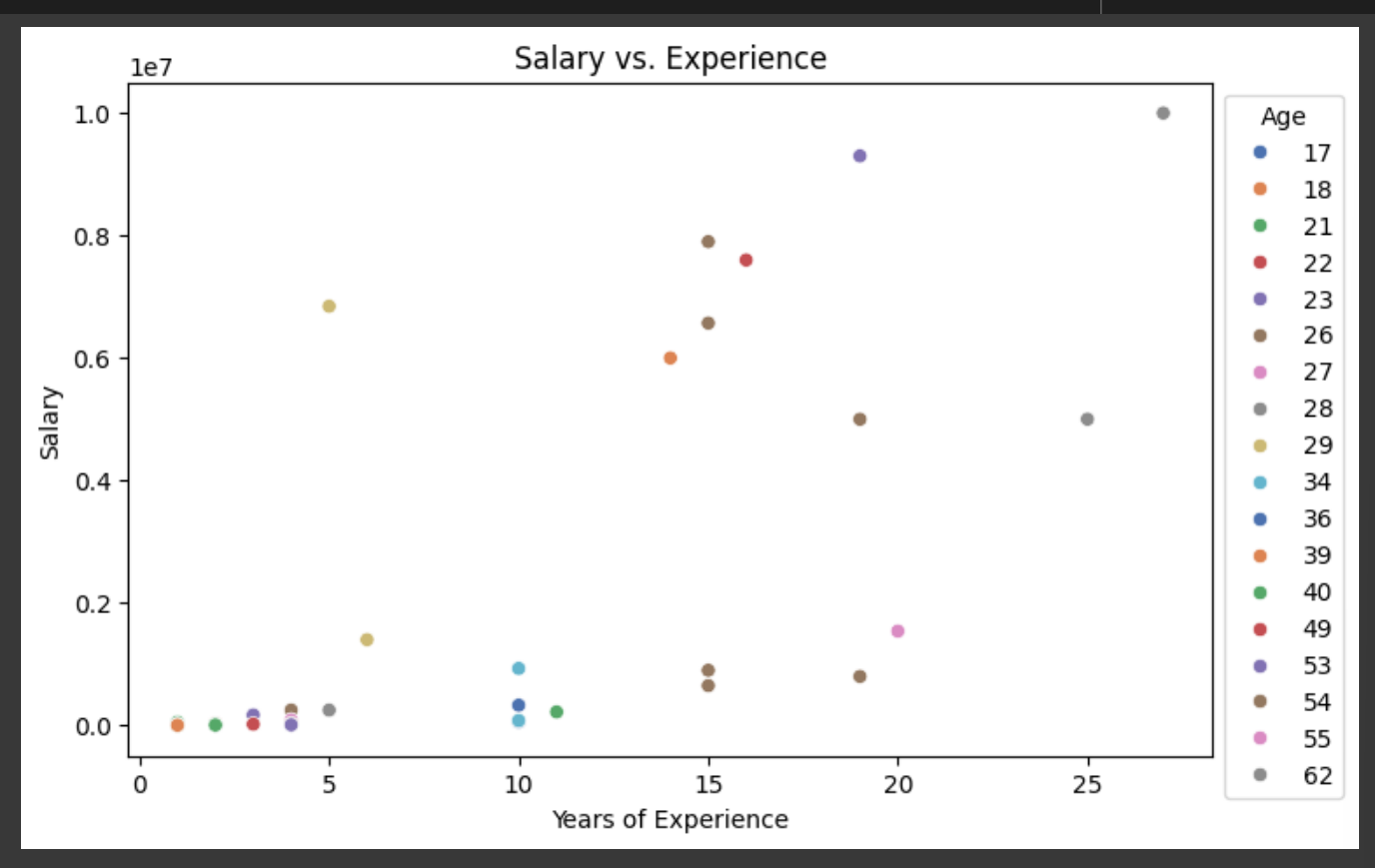
plt.gca().spines[['top', 'right']].set\_visible(False)

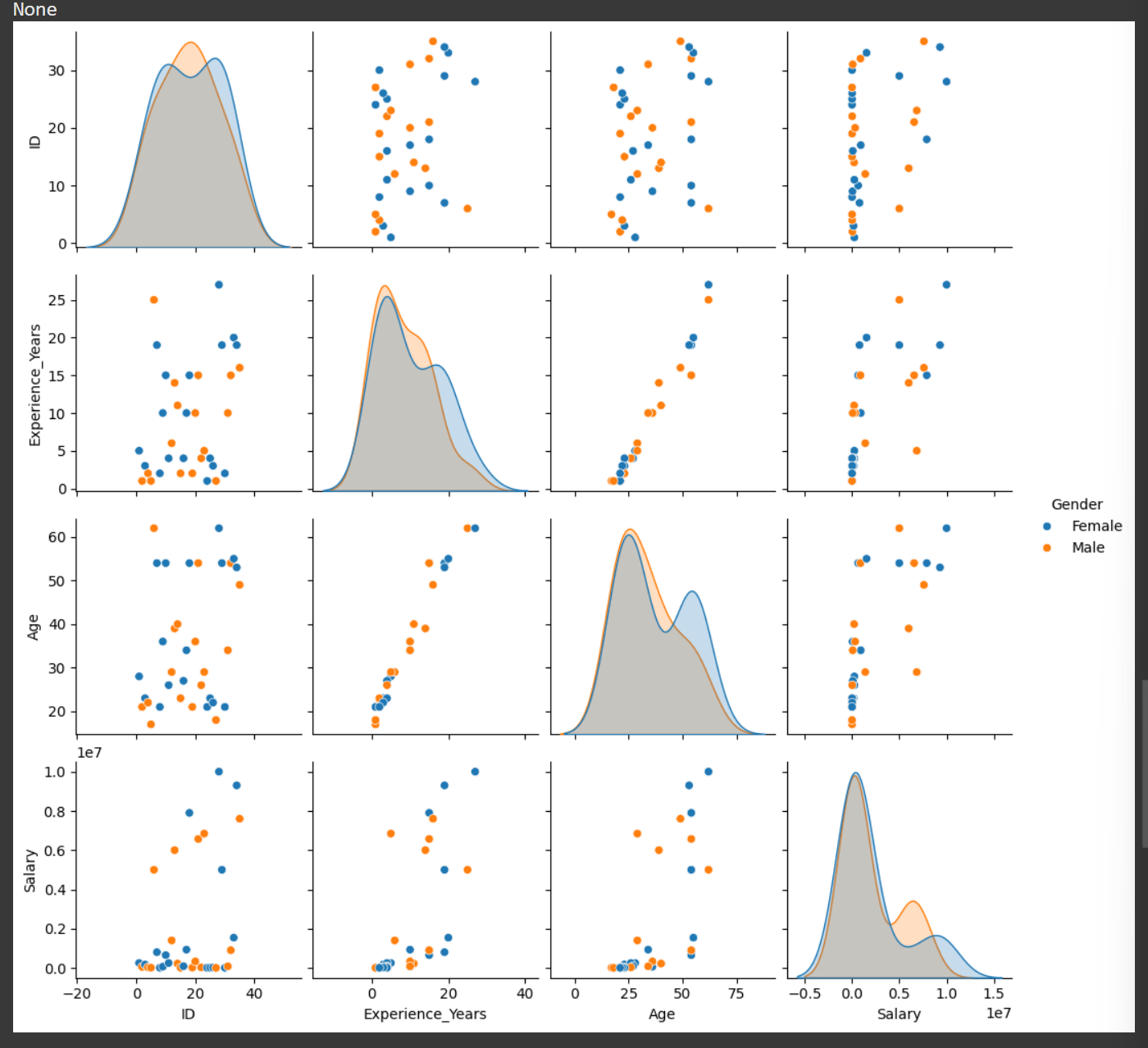
Screen Shots and Graphs

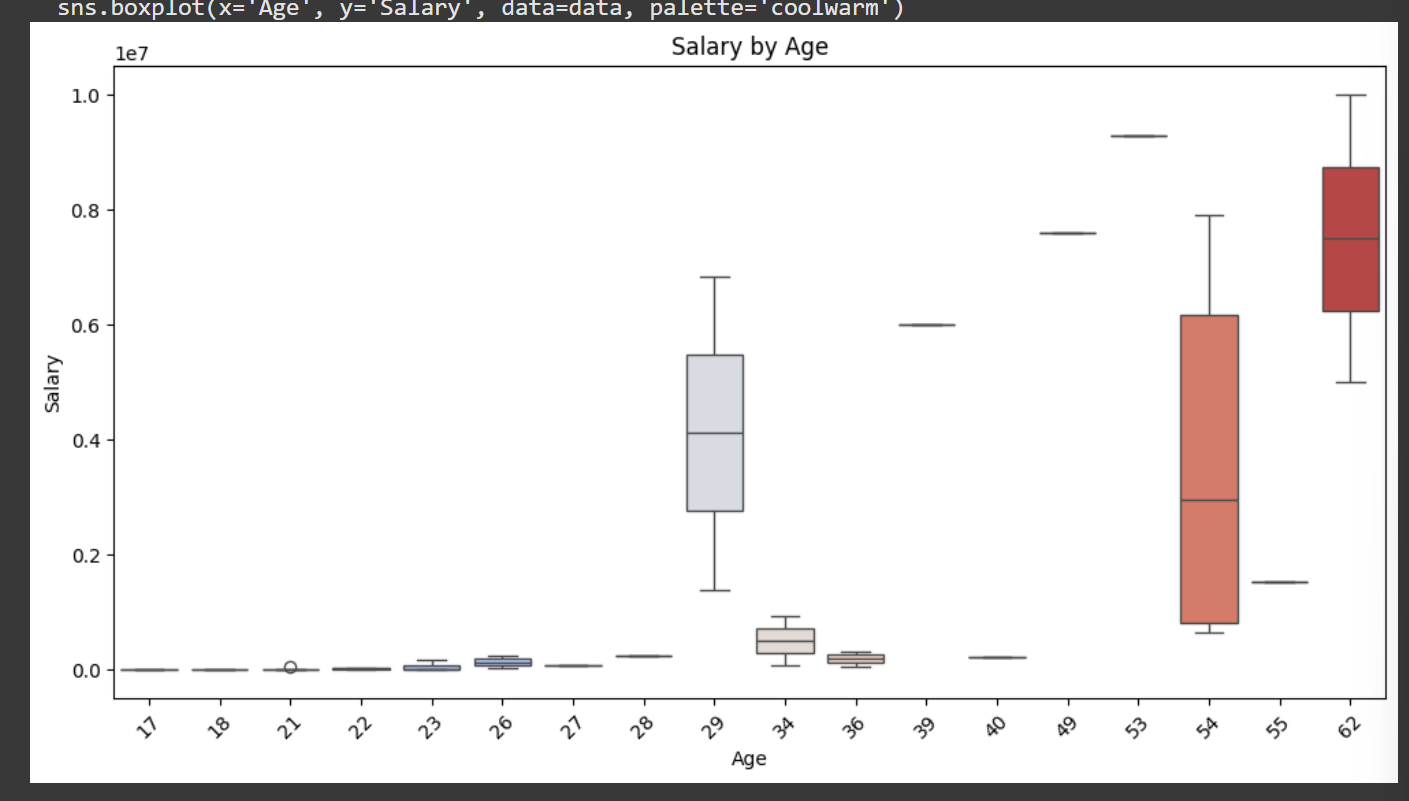


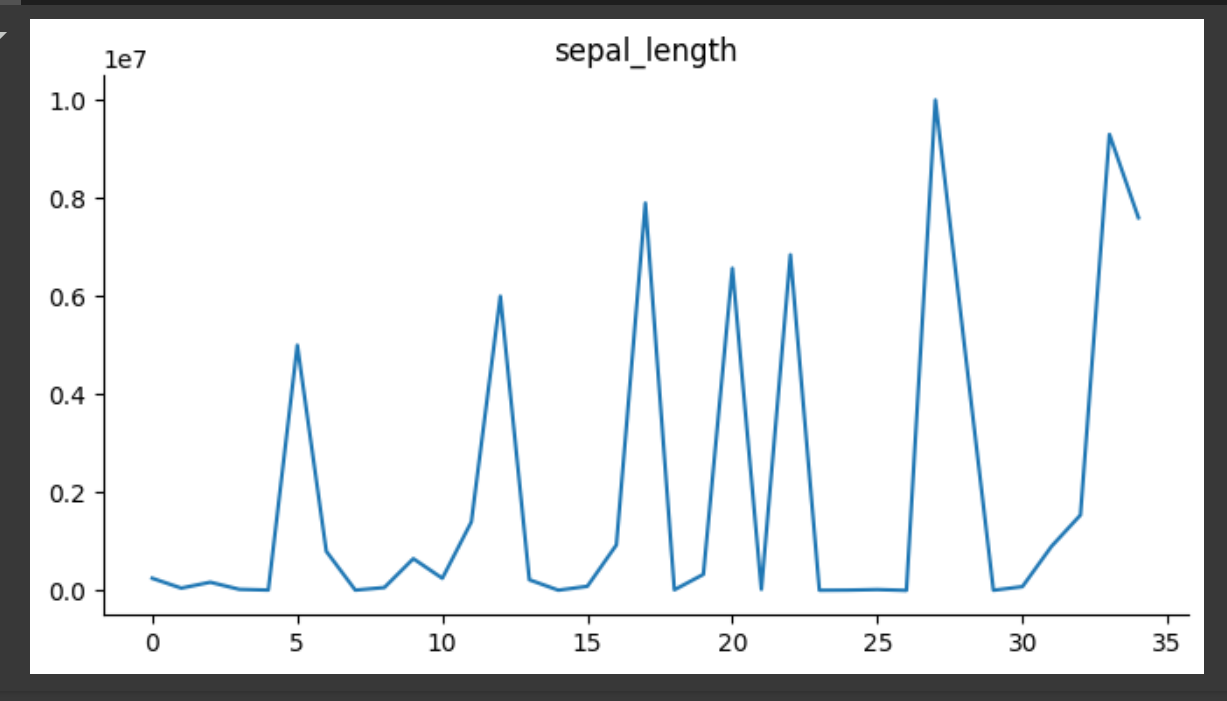












References/Credits

* Dataset: From Kaggle
* Python Libraries: Pandas, Matplotlib, Seaborn, NumPy
* AI MSE Guidelines