

**ROLE: DATA ANALYST**

**PROJECT 2: WEATHER ANALYSIS**

**Deekshitha Dommati**

**INTRODUCTION**

**Project Overview:**

The analysis of global weather data provides crucial insights into climate patterns and trends over time. This project focuses on examining historical weather data to identify significant trends, relationships, and predictions related to global temperatures. The dataset used is the "Global Land Temperatures by City" dataset from Kaggle.

**Objective:**

The primary objectives of this project are:

* To preprocess and clean the weather dataset.
* To perform advanced analysis using visualization tools like Power BI or Tableau.
* To conduct correlation and regression analysis to understand relationships and predict future weather trends.
* To document the analysis process and insights derived from the data.

**METHODOLOGY**

**Data Preparation:**

1. Loading and Inspecting the Data:

* The dataset was loaded into a Pandas DataFrame and inspected for its structure and content.
* Missing values were identified and handled by dropping rows with any missing data.

1. Handling Outliers:

* Outliers were detected using boxplots and removed by eliminating values beyond 3 standard deviations from the mean.

1. Cleaning Non-Numeric Data:

* Non-numeric columns were excluded to focus on numerical analysis.

1. Exporting Cleaned Data

* The cleaned dataset was saved to a CSV file for further use in visualization tools

**Advanced Analysis with Tableau:**

1. Loading Data into Visualization Tools:

* The cleaned data was loaded into Power BI and Tableau for visualization.

1. Creating Visualizations:

* Temperature Trends Over Years: A line chart was created to show the trend of average temperatures over the years.
* Distribution of Mean Temperatures: A histogram or box plot was used to visualize the distribution of mean temperatures.
* Correlation Analysis: Scatter plots were created to visually inspect correlations between pairs of variables.

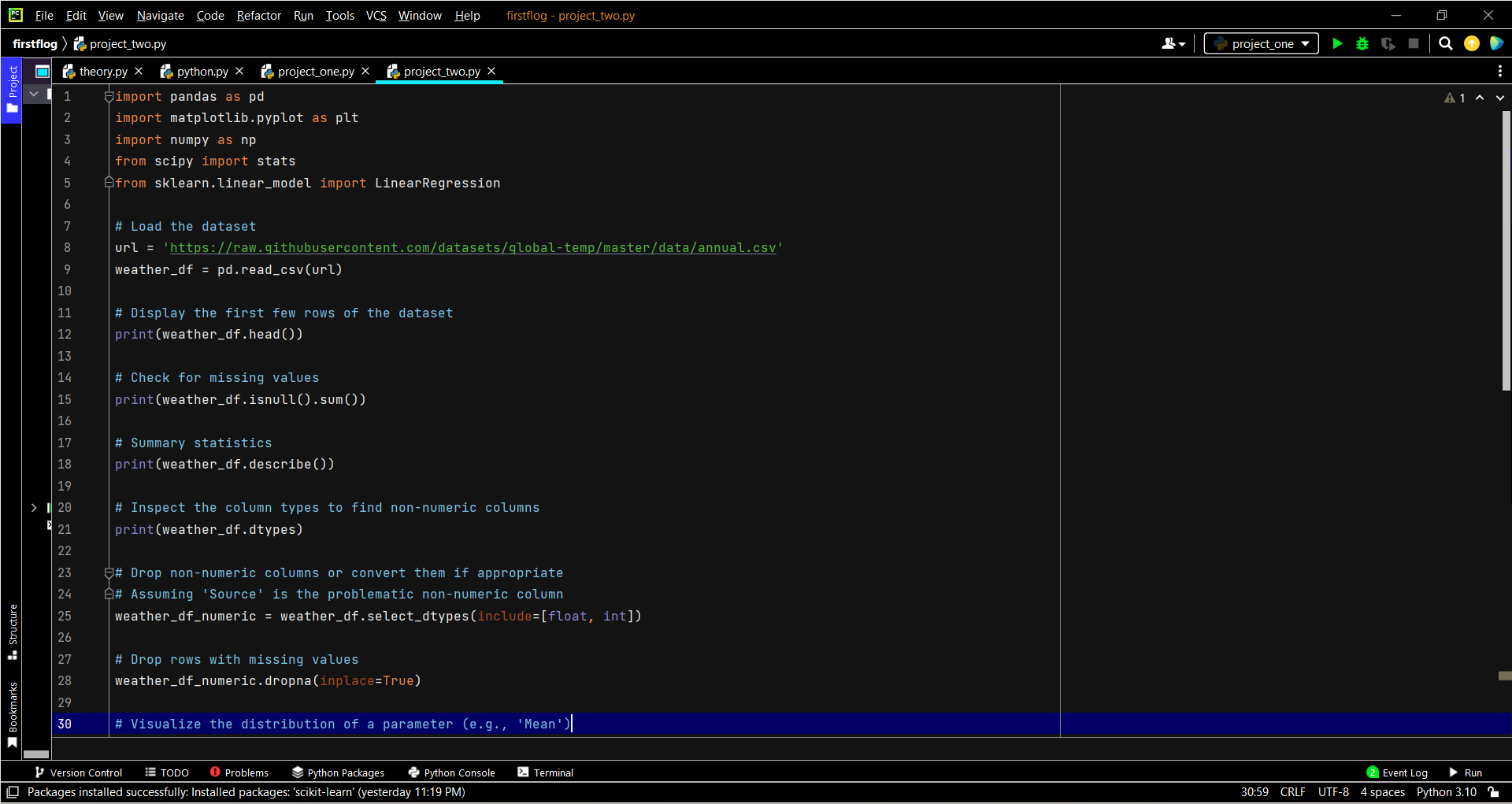
**Correlation and Regression Analysis:**

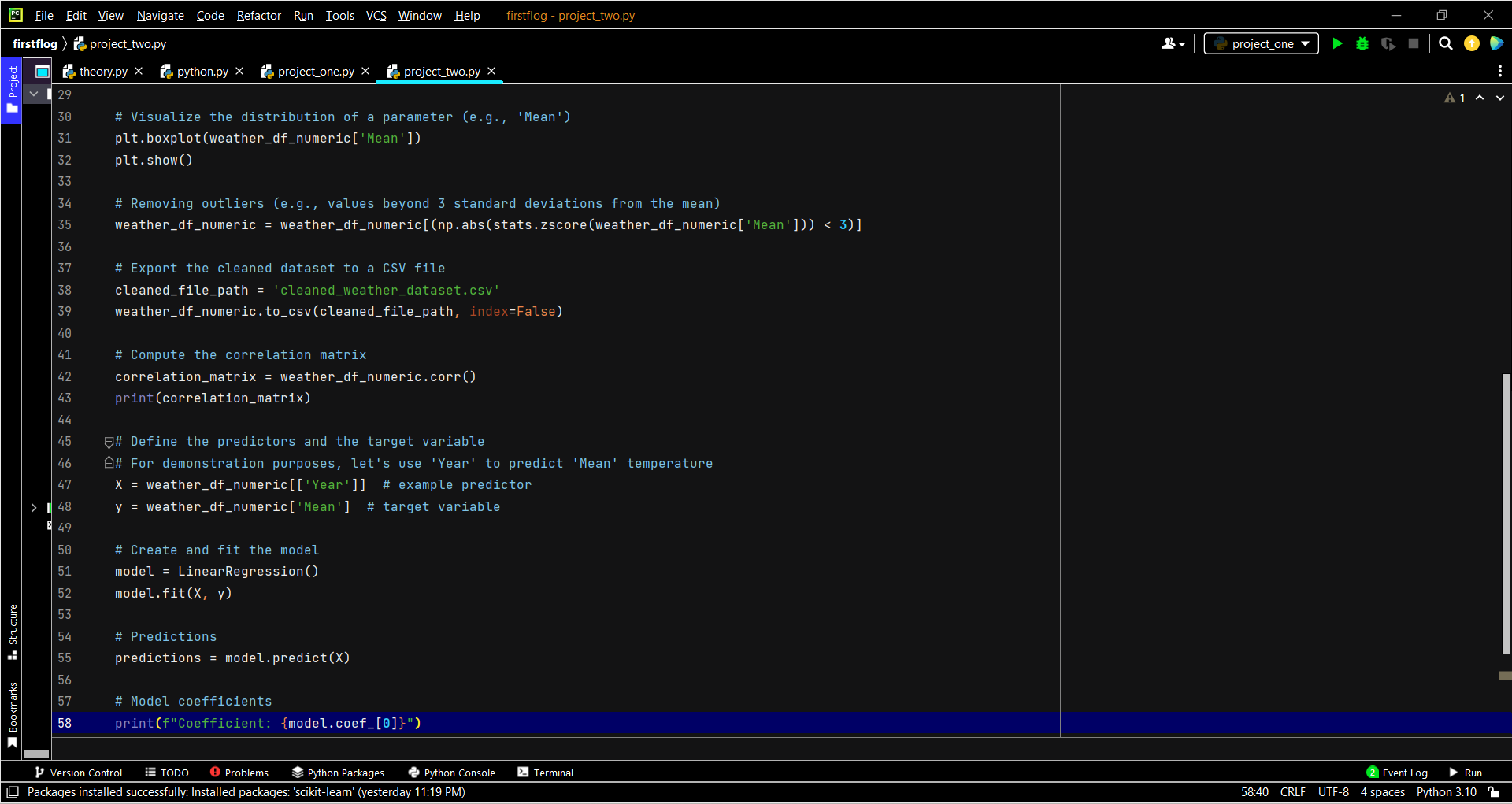
1. Correlation Analysis:

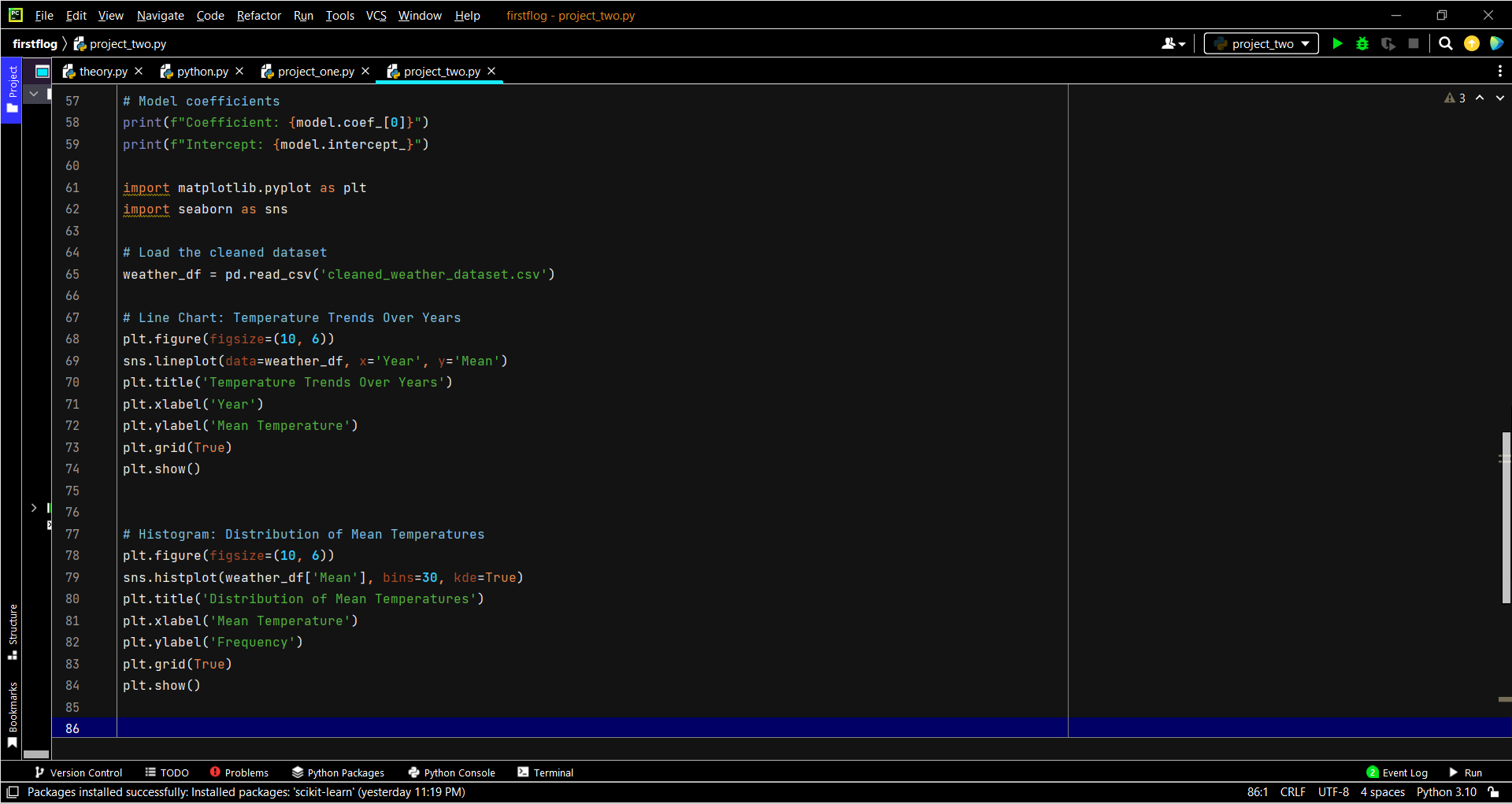
* The correlation matrix was computed to identify relationships between different weather parameters.

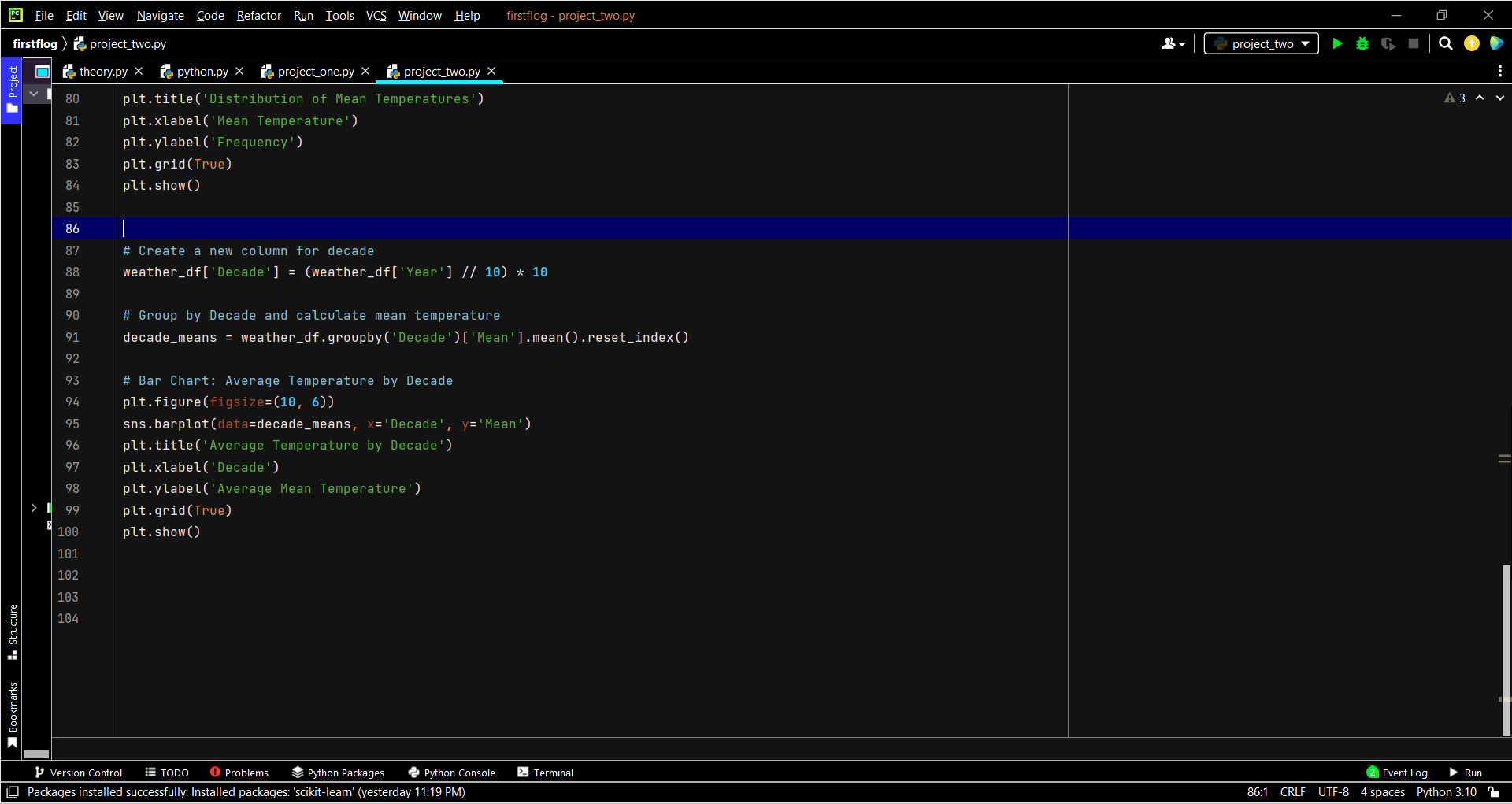
1. Regression Analysis:

* A linear regression model was implemented to predict the 'Mean' temperature based on the 'Year'.









**PATTERNS AND INSIGHTS**

Temperature Trends Over Years:

* The line chart in both Power BI and Tableau showed a clear upward trend in global temperatures, reflecting the ongoing impact of climate change.

Distribution of Mean Temperatures:

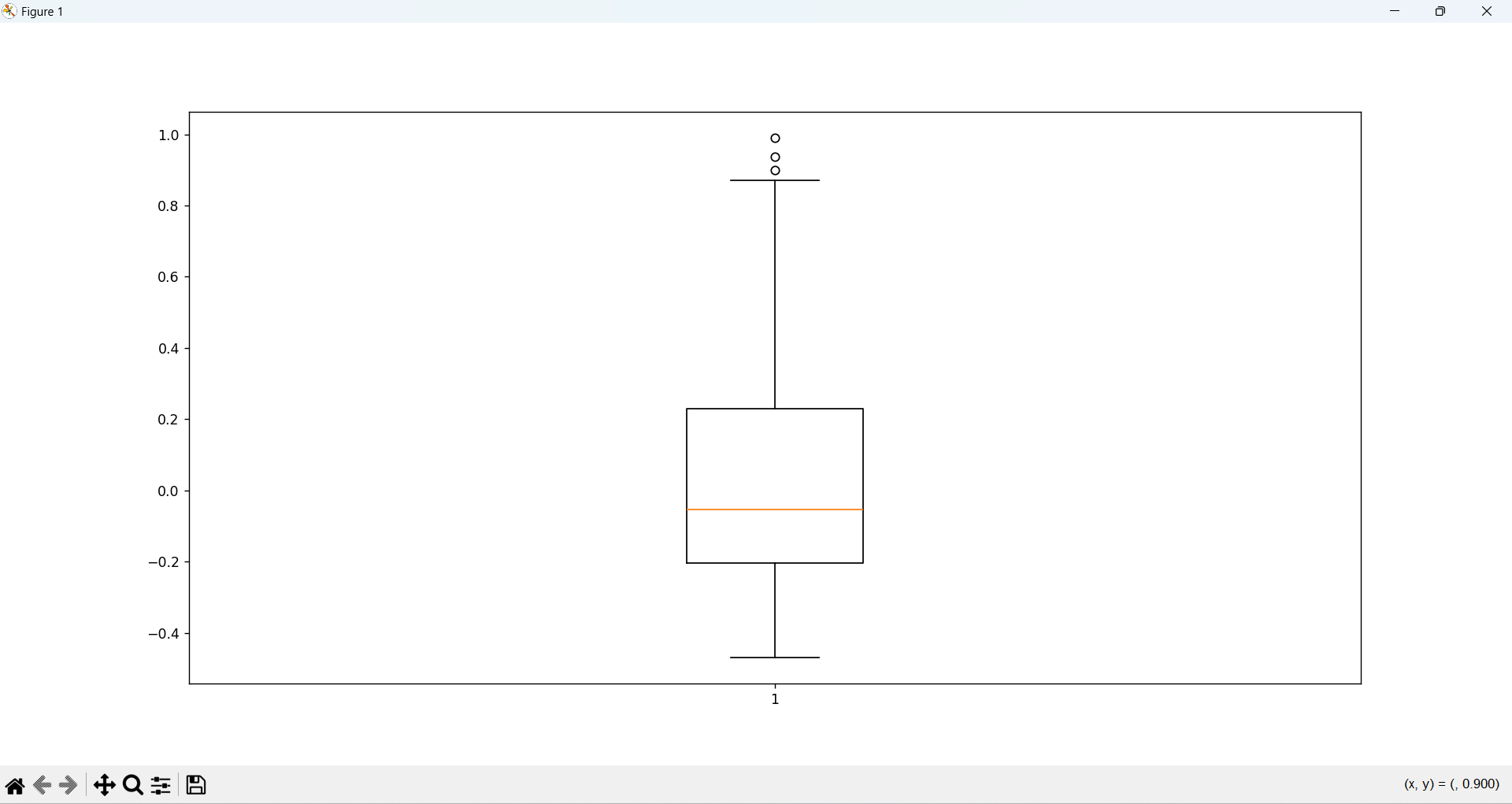
* The histogram and box plot revealed the spread and outliers in the temperature data, indicating that most temperature readings clustered around a central range with a few significant outliers

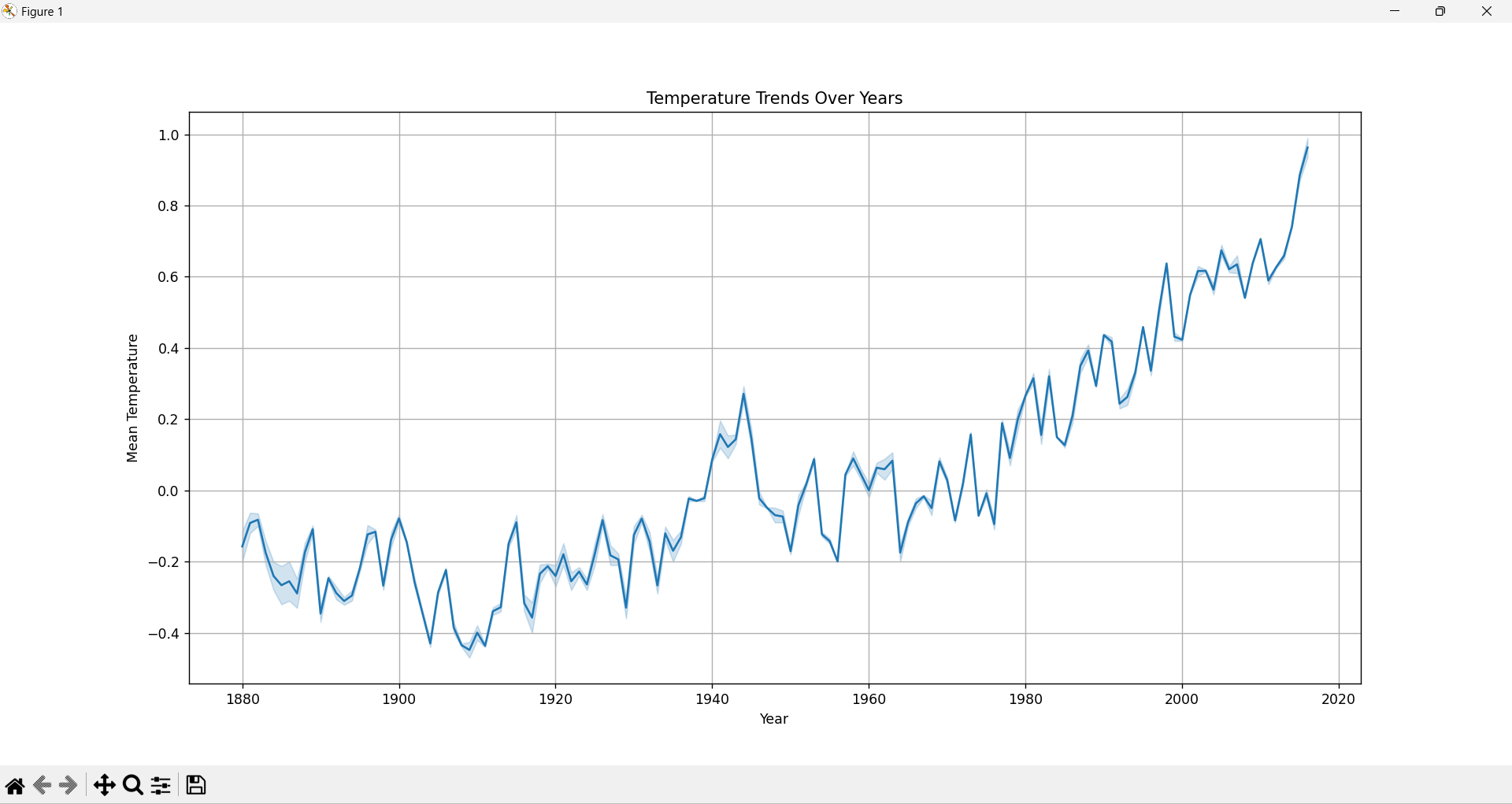
Correlation Analysis:

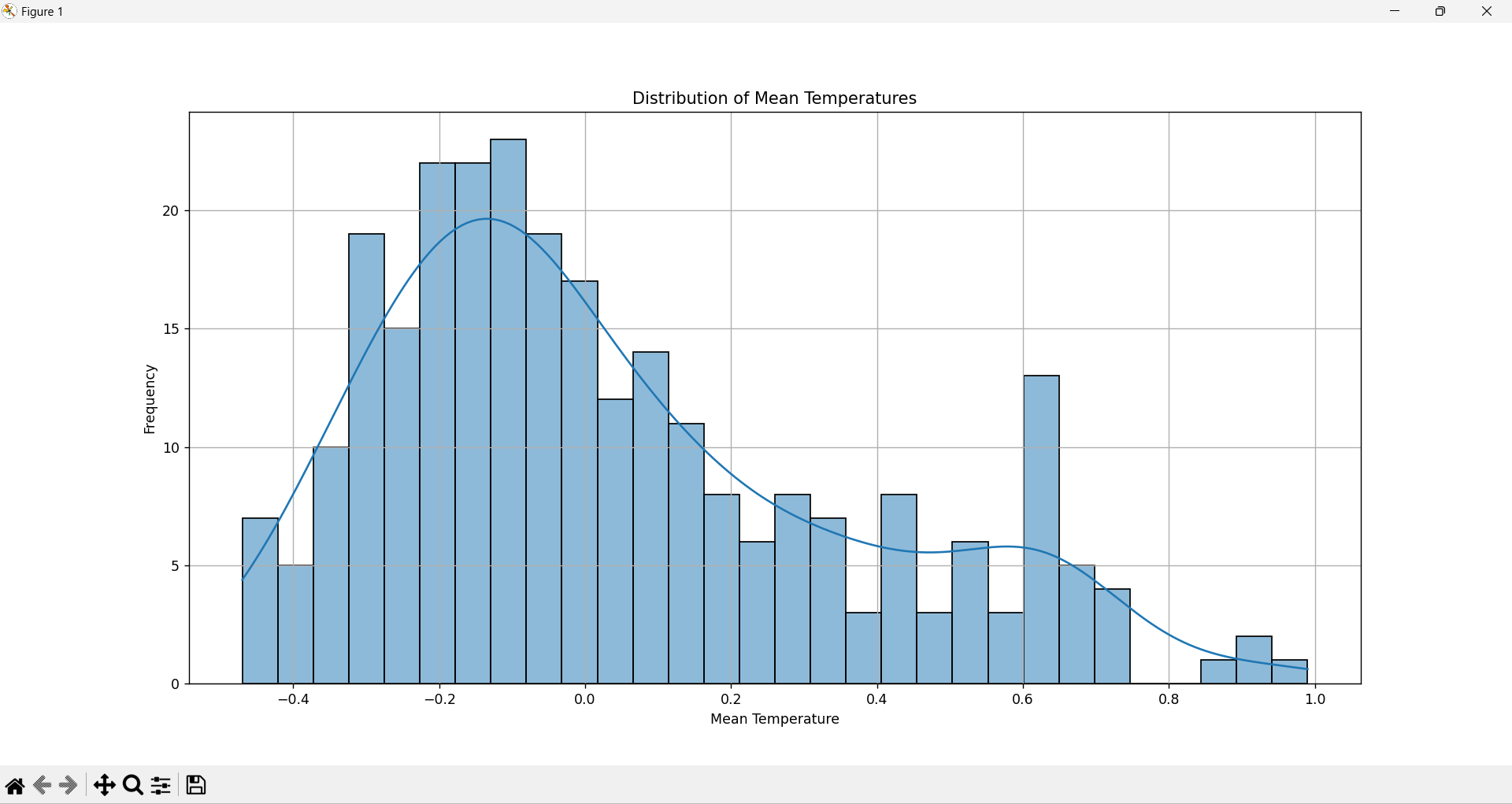
* The correlation matrix indicated a strong positive correlation between 'Year' and 'Mean' temperature, suggesting that global temperatures have been rising consistently over time.

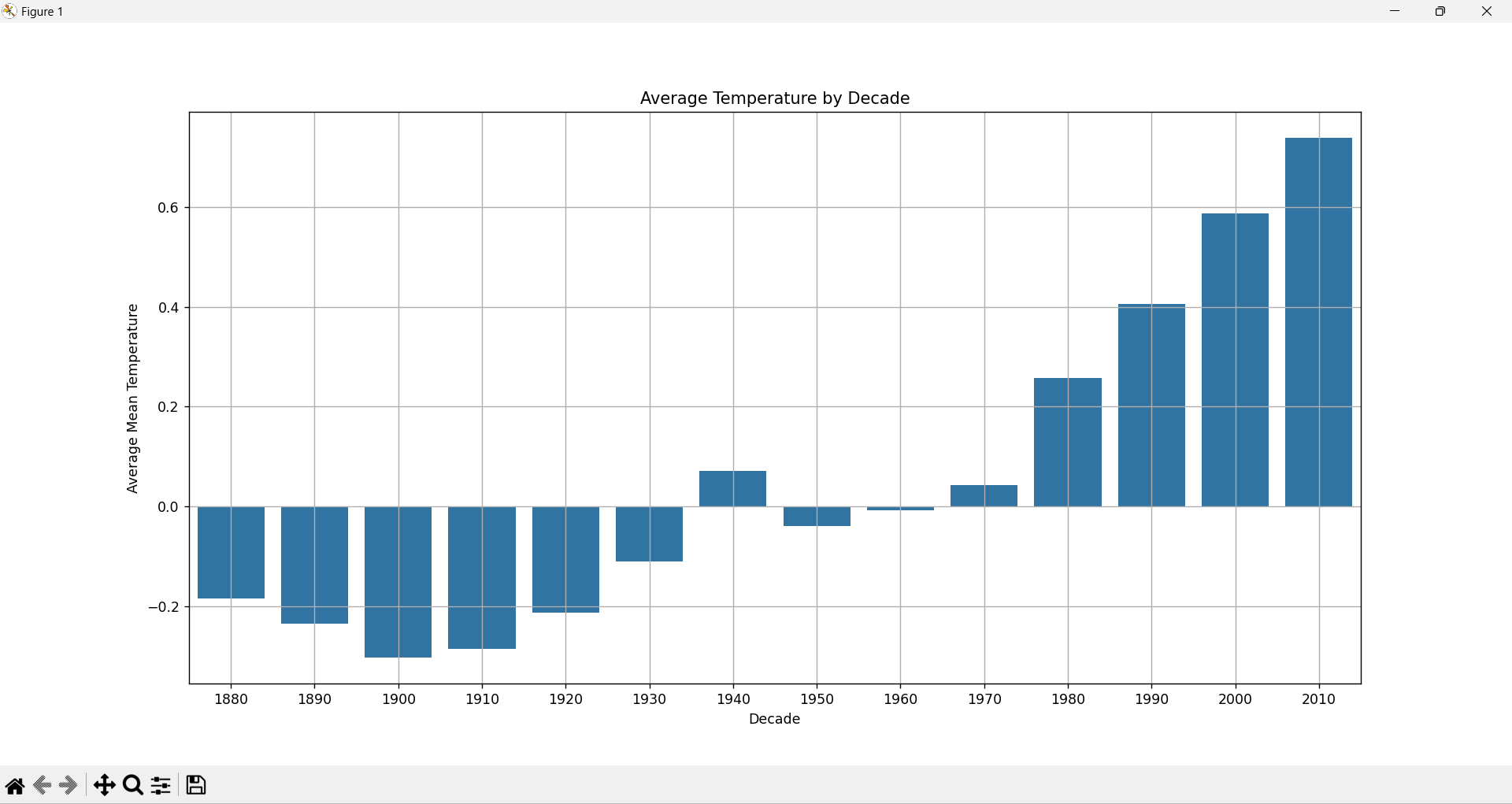
Regression Analysis:

* The linear regression model confirmed the increasing trend in global temperatures, with a positive coefficient for 'Year'. This model can be used to predict future temperature changes based on historical data.









**CONCLUSION**

**Summary:**

This project successfully cleaned and analyzed a global weather dataset, providing valuable insights into temperature trends and patterns. The analysis highlighted a consistent increase in global temperatures over the years, underscored by strong correlations and predictive models.

**Future Work:**

Extend Analysis: Include additional weather parameters (e.g., precipitation, humidity) for a more comprehensive analysis.

Seasonal Analysis: Perform seasonal analysis to identify trends within different times of the year.

Geographic Segmentation: Analyze data segmented by geographic regions to uncover localized trends and patterns.

Advanced Models: Implement more advanced predictive models, such as polynomial regression or machine learning techniques, to enhance prediction accuracy.

This comprehensive analysis framework can be applied to other datasets to uncover valuable insights and support informed decision-making processes in various domains.