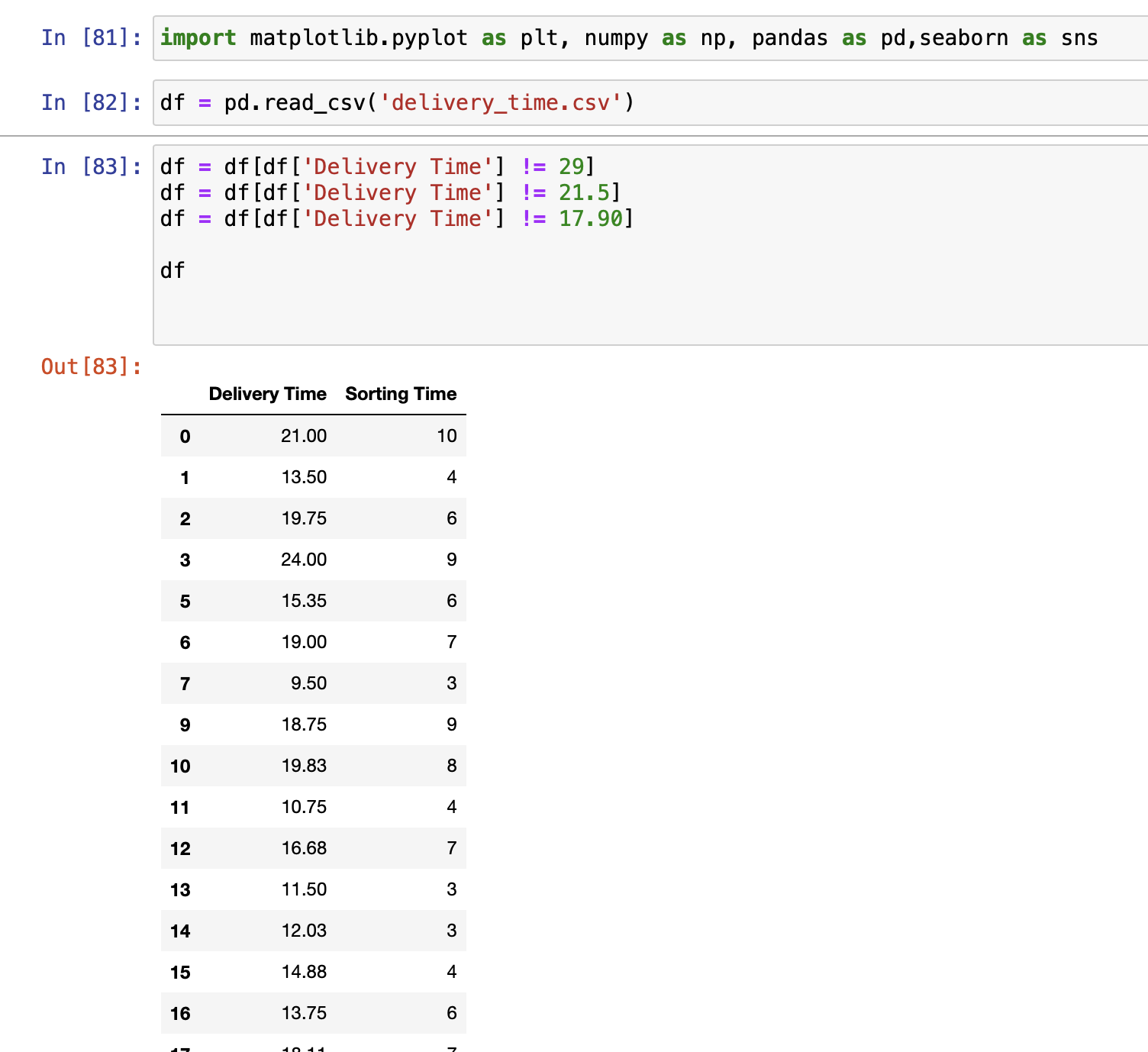
Simple Linear Regression

1) Delivery\_time -> Predict delivery time using sorting time

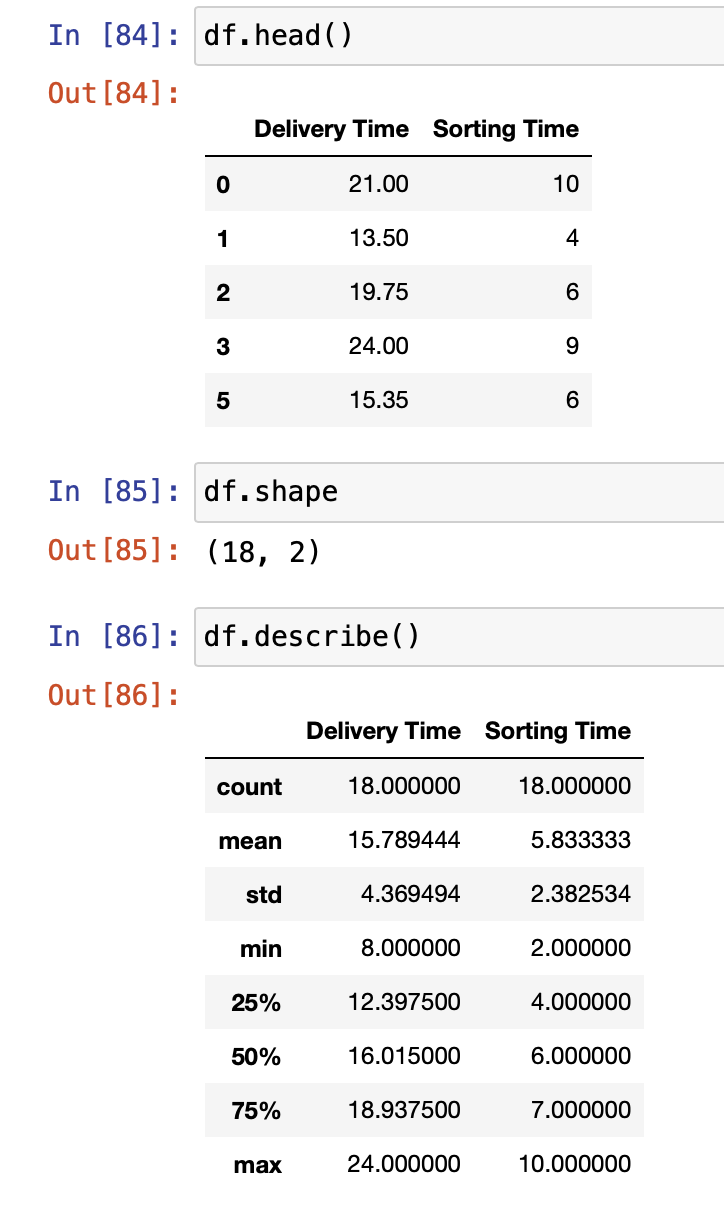
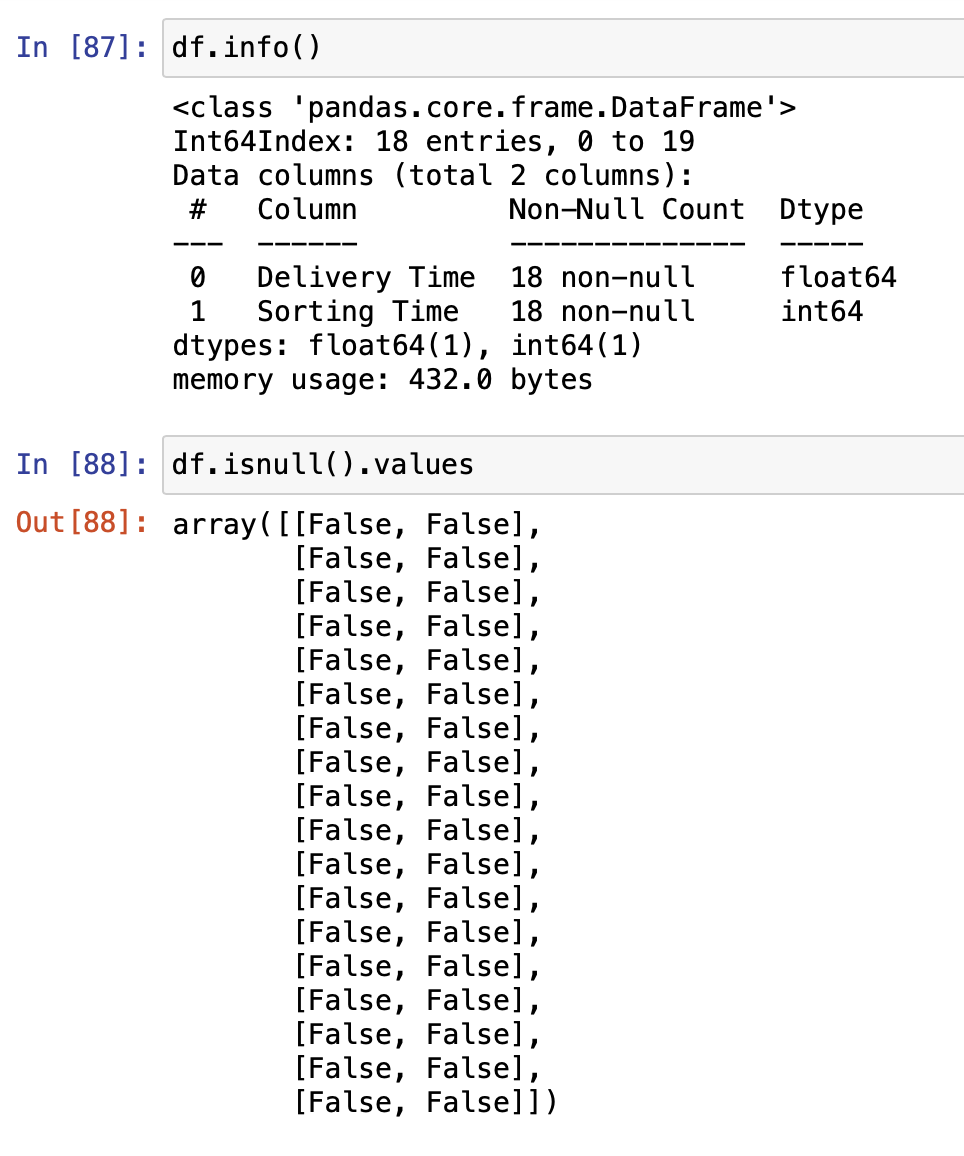
### Data Import and Cleaning:

* We start by importing the necessary Python libraries: **matplotlib.pyplot**, **numpy**, **pandas**, and **seaborn**.
* The delivery time data is read from the 'delivery\_time.csv' file into a Pandas DataFrame called **df**.
* To improve data quality, we remove outliers from the 'Delivery Time' column, specifically values of 29, 21.5, and 17.90.



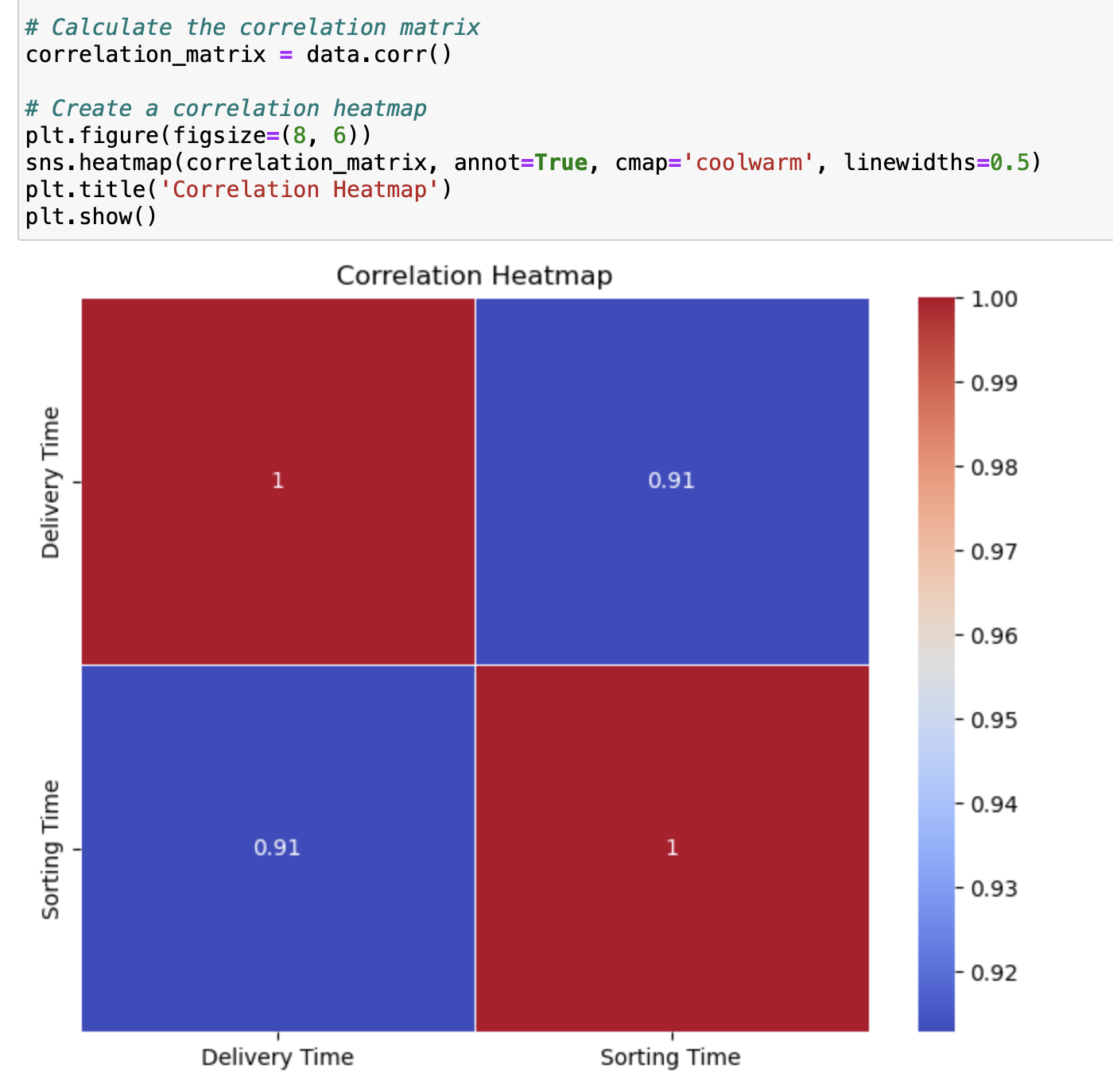
### Data Exploration:

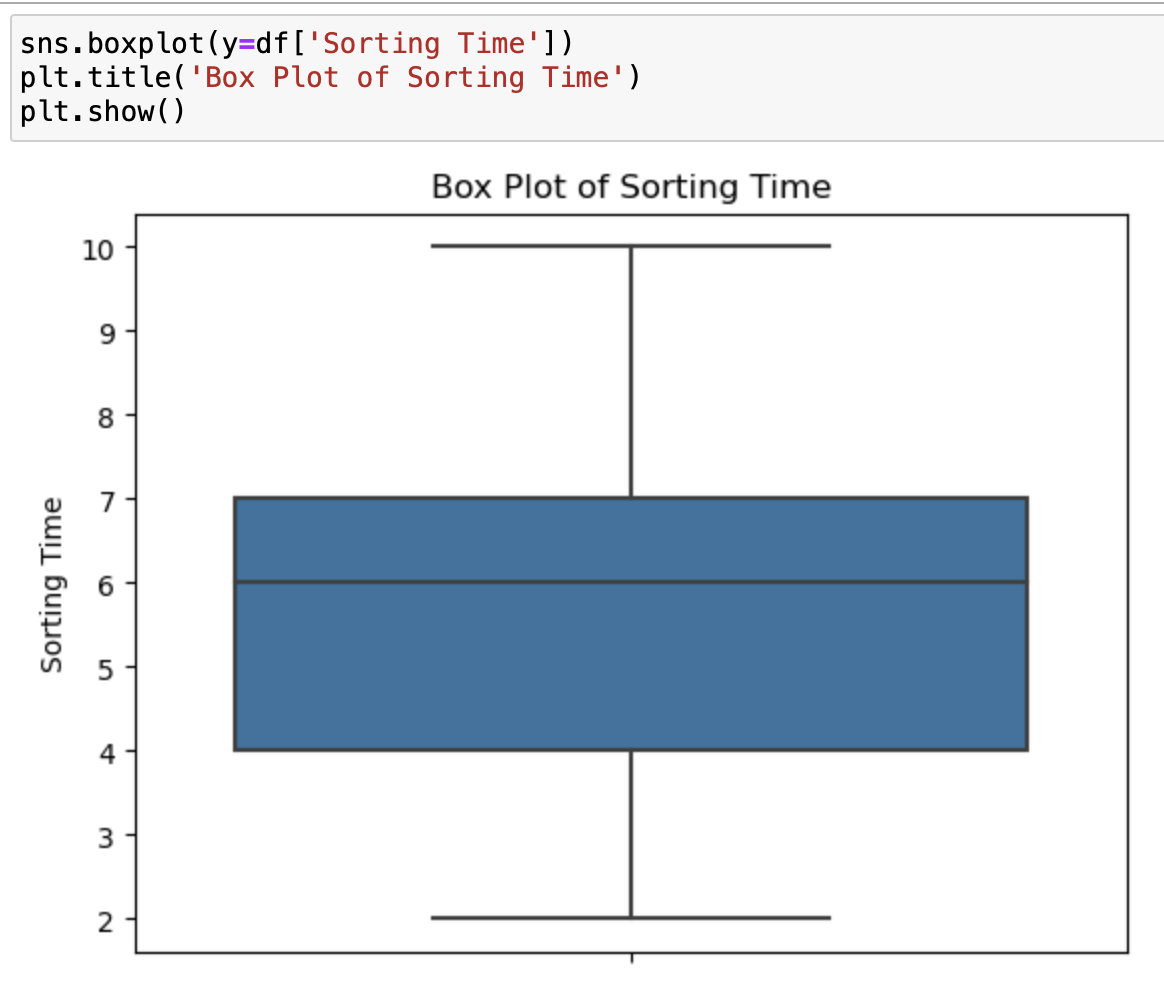
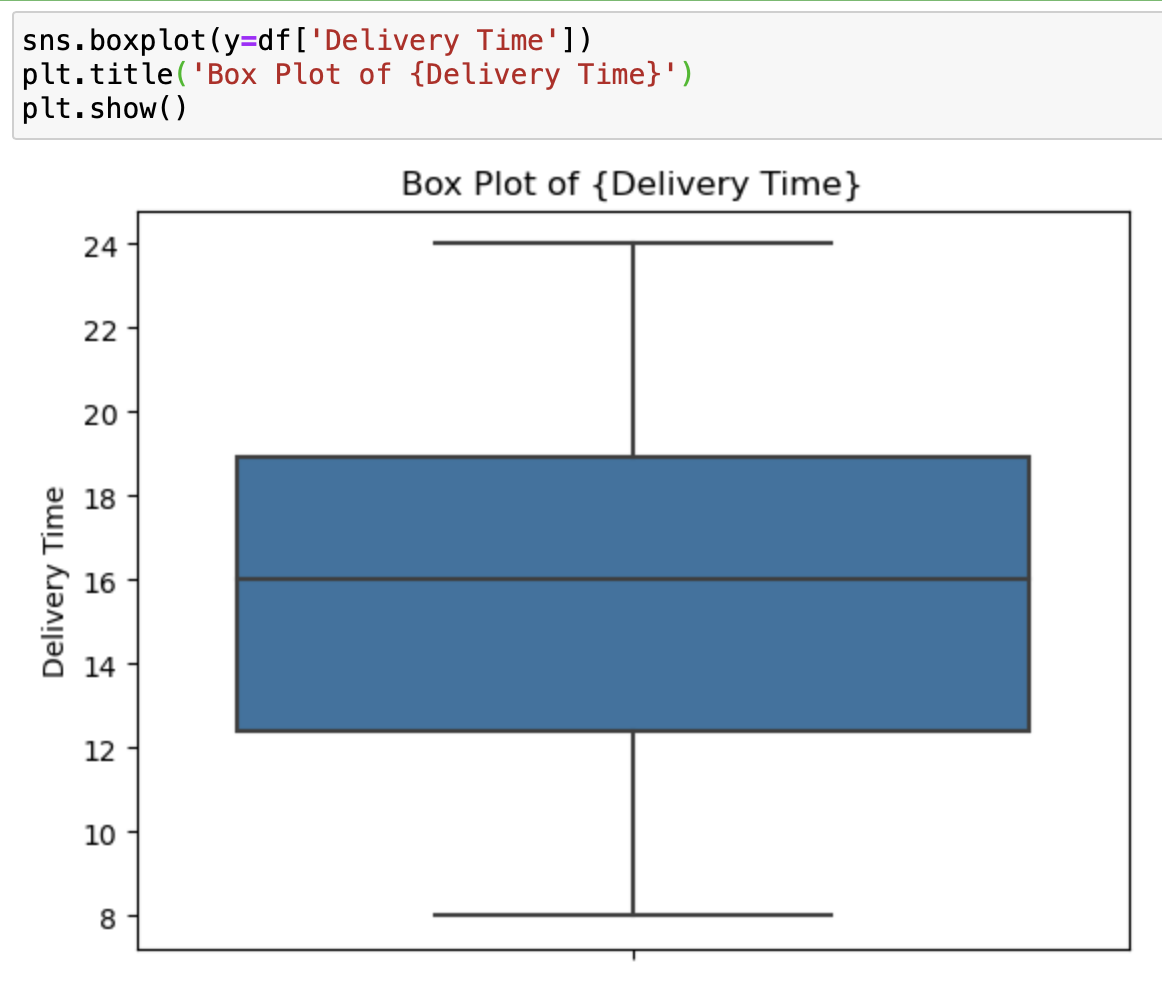
* **df.head()** displays the first few rows of the dataset.
* **df.shape** provides information about the number of rows and columns in the DataFrame.
* **df.describe()** offers summary statistics for the numerical columns.
* **df.info()** reveals details about data types and potential missing values.
* **df.isnull().values** checks for any missing values within the DataFrame.



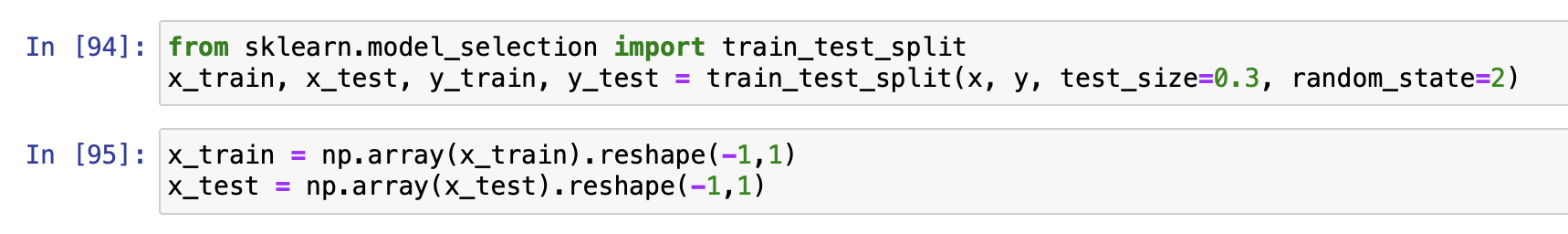
### Data Visualization:

* We utilize Seaborn to create a box plot of the 'Delivery Time' column, visualizing the distribution of delivery times.
* A correlation heatmap is generated using Seaborn to visualize relationships between the numerical variables within a separate dataset called **data**.



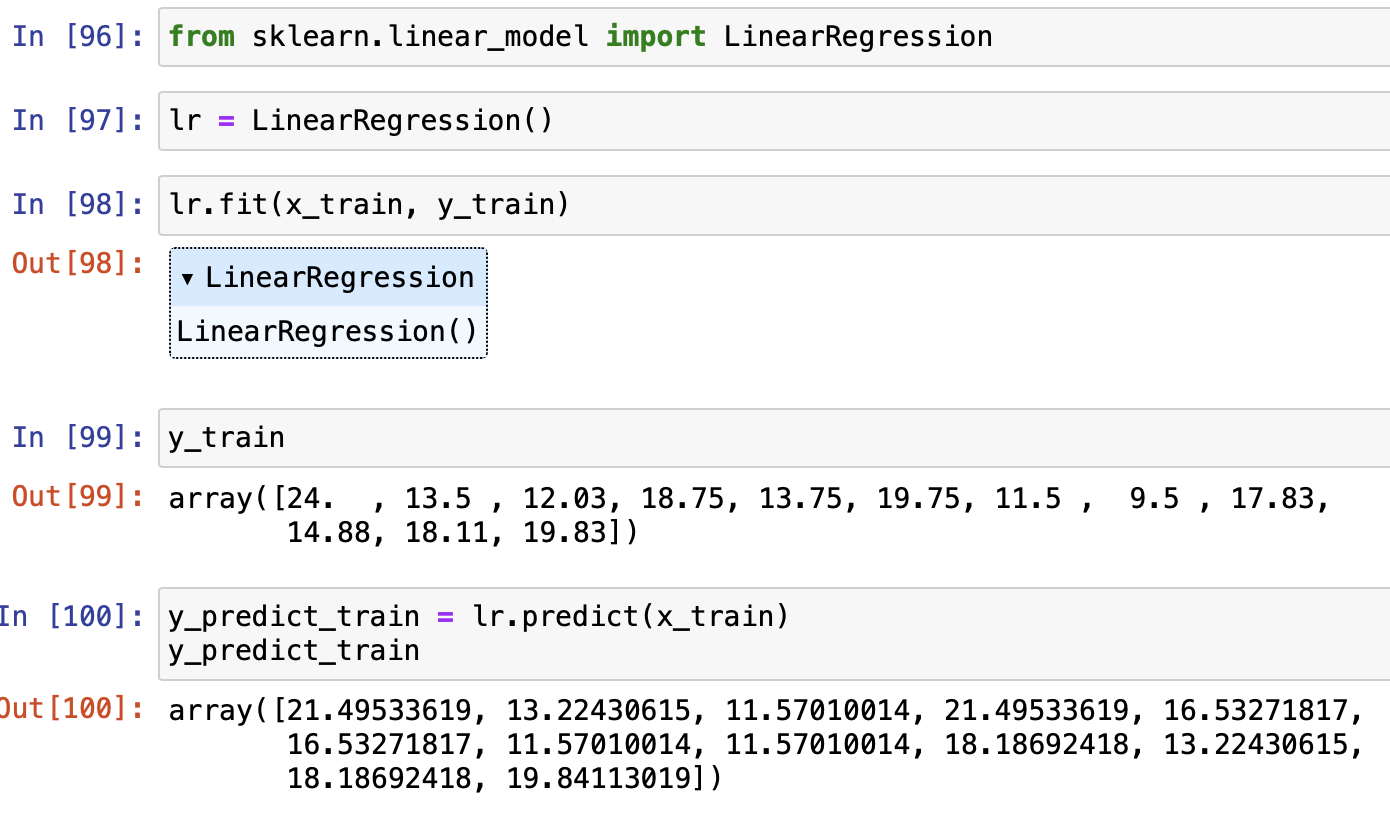


### Screenshot 2023-09-16 at 1.56.01 PM.pngLinear Regression Model:

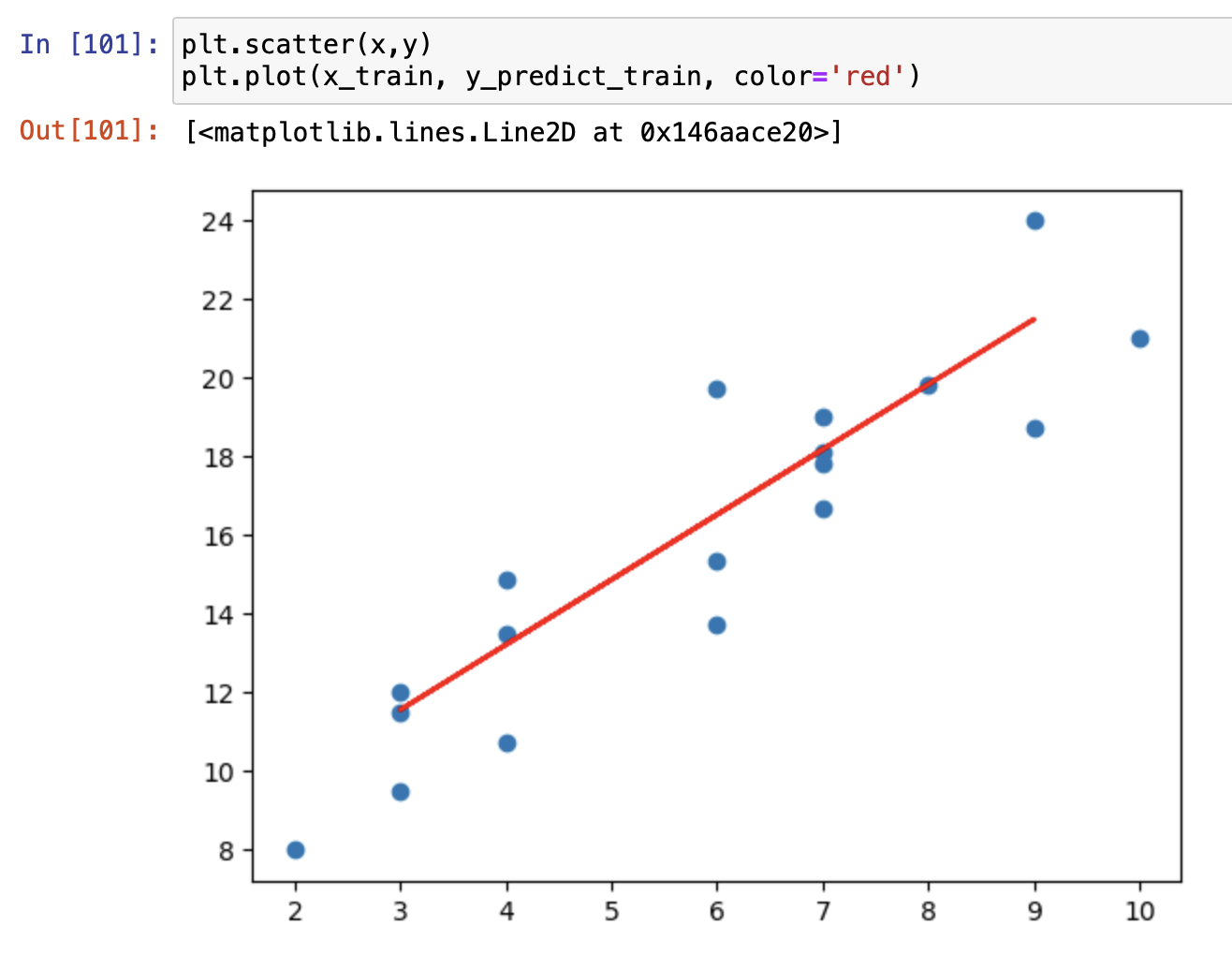
* We prepare the data for building a linear regression model.
* The 'Sorting Time' column is designated as the independent variable (**x**), and the 'Delivery Time' column is designated as the dependent variable (**y**).
* Data is divided into training and testing sets using **train\_test\_split** from **sklearn.model\_selection**.
* We reshape the independent variables in both the training and testing sets using NumPy to accommodate the model's requirements.

### Linear Regression Modeling:

* We import the **LinearRegression** model from **sklearn.linear\_model**.
* An instance of the linear regression model (**lr**) is created and fitted with the training data.

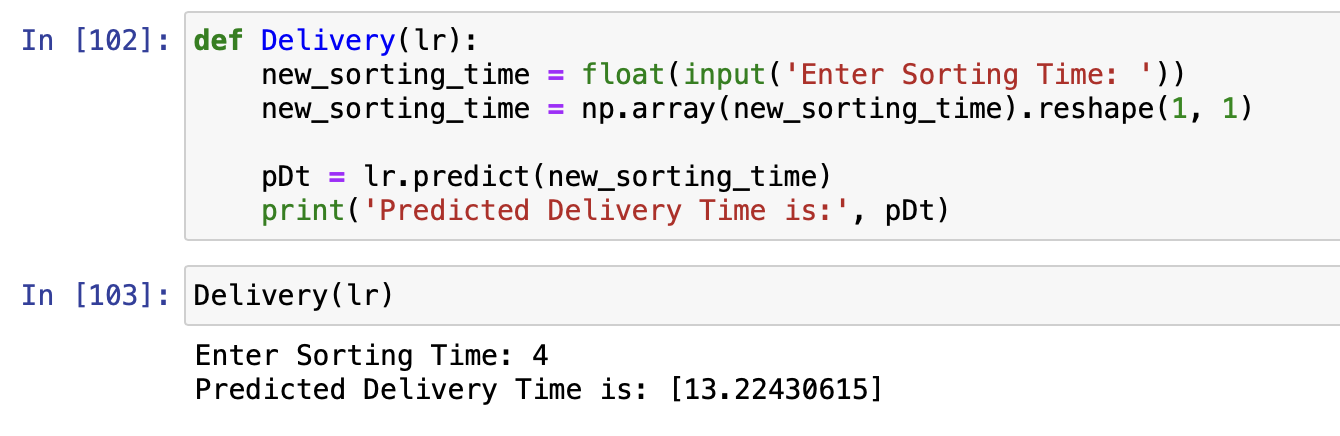


### Model Visualization:

* A scatter plot is generated to visualize the relationship between 'Sorting Time' and 'Delivery Time.' Additionally, the linear regression line is overlaid on the training data for further insight.

### Delivery Time Prediction:

* We define a function, **Delivery(lr)**, to make predictions for delivery times based on sorting times.
* Users are prompted to input a sorting time, and the function utilizes the trained linear regression model to predict the corresponding delivery time.
* The predicted delivery time is then displayed as 'Predicted Delivery Time.



### Conclusion:

* The assignment is summarized by highlighting the key steps involved, including data import, cleansing, exploration, visualization, model construction, and delivery time prediction.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_