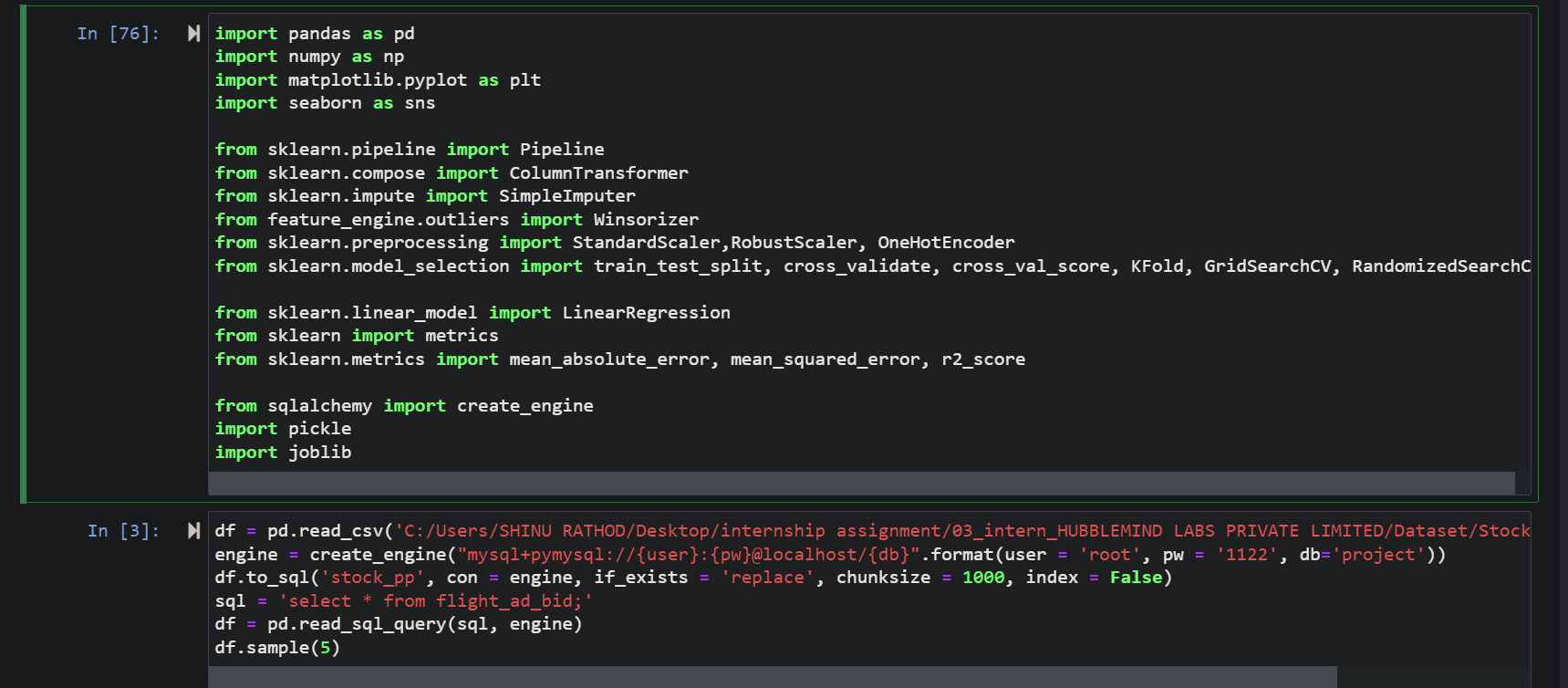
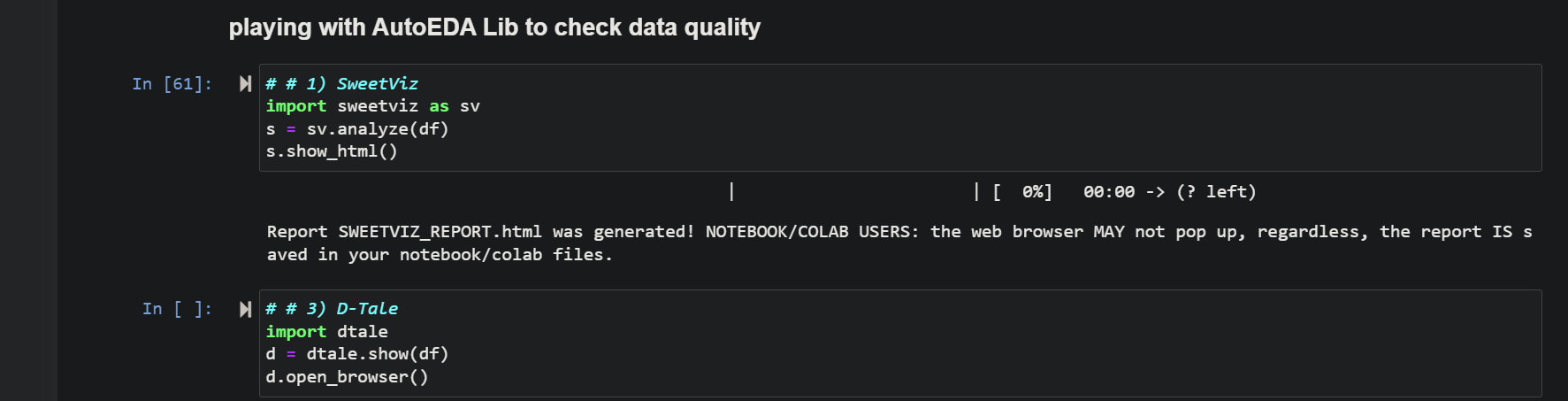
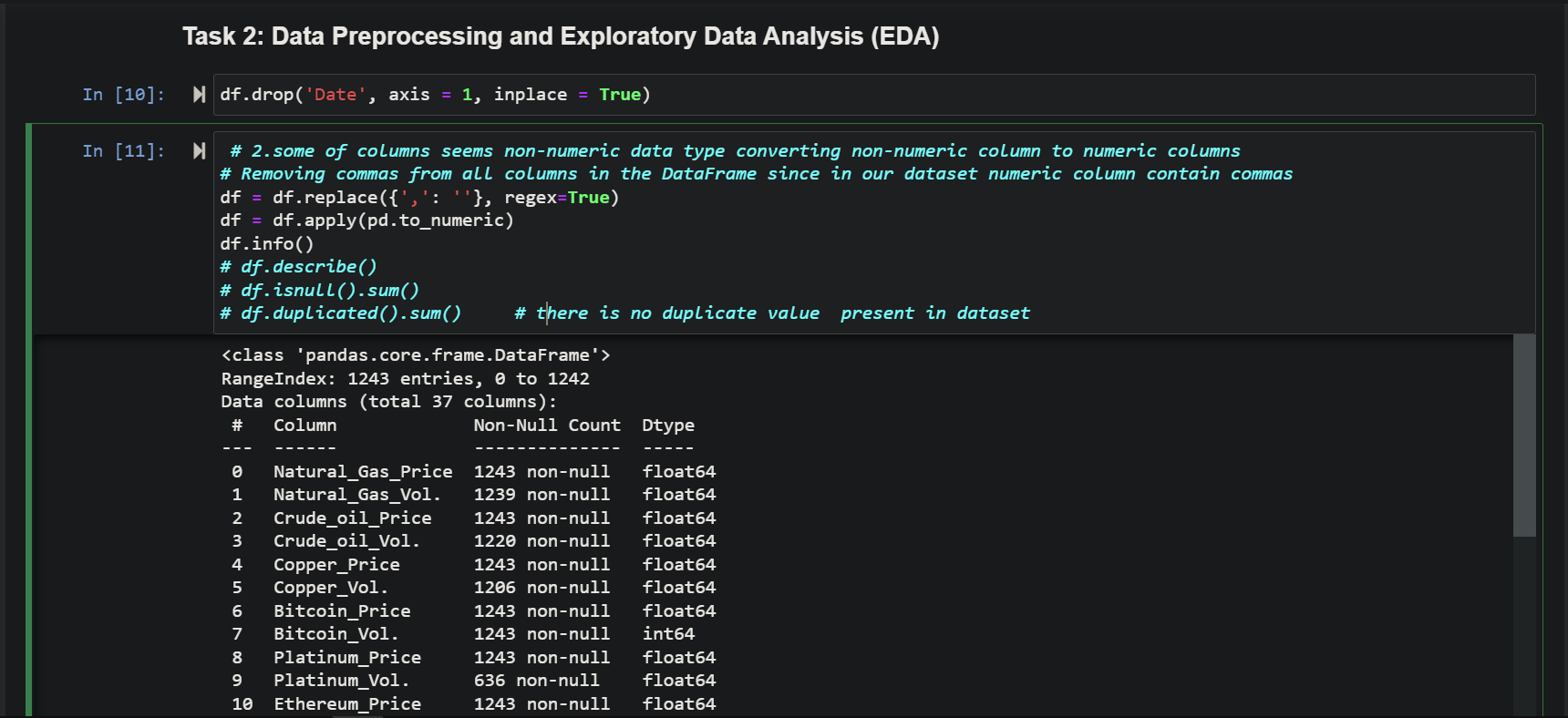
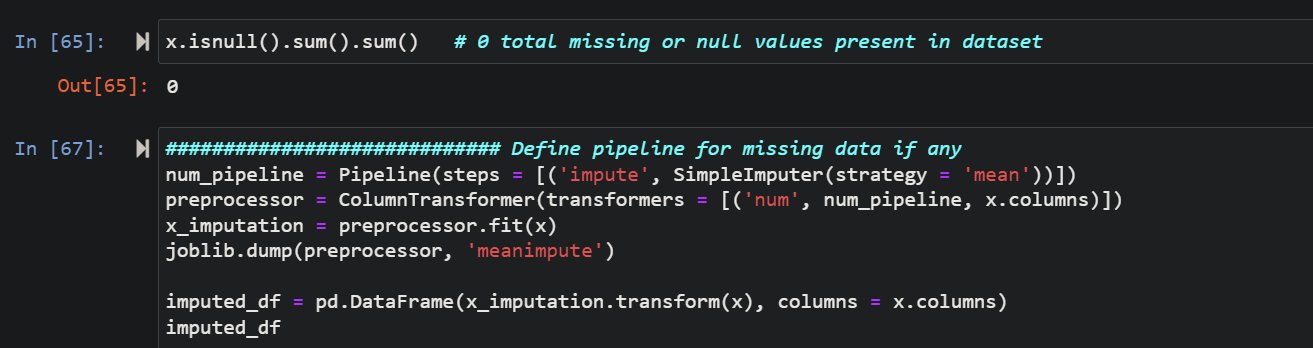
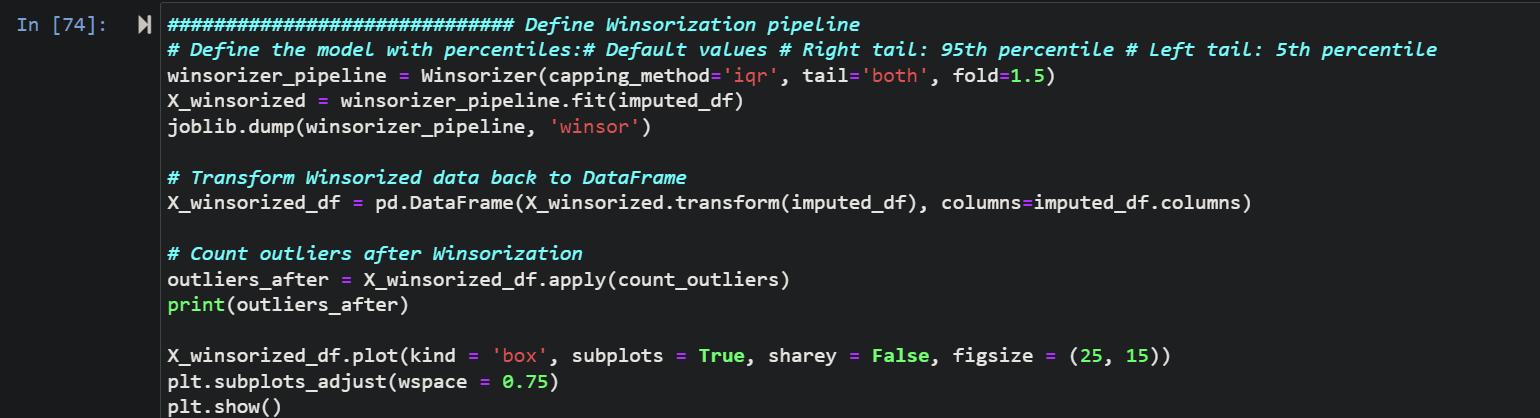
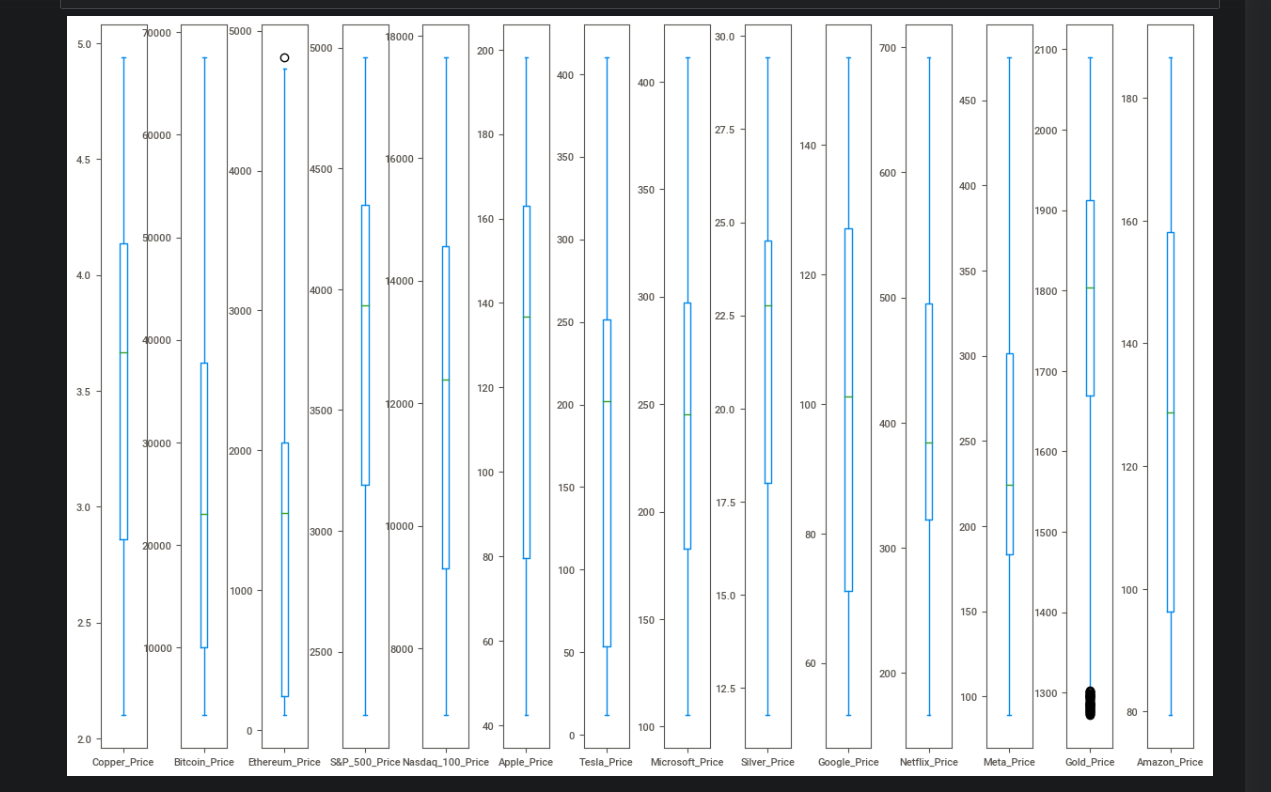
***Task 1: Document the Entire Process***

**1. Importing all Required lib and Dataset** :



Here importing all required lib and dataset then pushing dataset to mysql database after pushing to mysql database retrieving all data again for further analysis on that data using the sql squery

**1. Data Preprocessing**

* **Loading the Dataset**: The dataset was loaded into a DataFrame from a CSV file. An initial inspection was done using **df.head(), df.info(),** **and df.describe()** to understand the data types, missing values, and summary statistics.
* **AutoEDA** : Significant time is required in the initial steps to analyze the data. The univariate analysis will reveal a lot of information about the business conditions and the Data. To conserve time in Exploratory Data Analysis (EDA), automation has been introduced with the help of Python libraries.
* **Data Cleaning**:
* 
  + **Date Column**: The Date column was dropped as it was not needed for the analysis.
  + **Comma Removal**: Non-numeric columns containing commas were cleaned by replacing commas with empty strings to convert them to numeric.
  + **Numeric Conversion**: The cleaned columns were converted to numeric types using ***pd.to\_numeric().***
  + **Missing Values**: Missing values were imputed using the mean strategy through a pipeline with ***SimpleImputer****.*
* **Outlier Treatment**:
  + **Winsorization**: A Winsorizer was applied to limit extreme values to reduce the effect of outliers. Before applying winsorization tech plotted boxplot we can see the there is outliers found
* **Feature Scaling**:
  + **Robust Scaling**: The features were scaled using ***RobustScaler*** to handle the presence of outliers more effectively.