AI-Driven Exploration and Prediction of Company Registration Trends with the Registrar of Companies (ROC)

NAME: KARTHIKA A
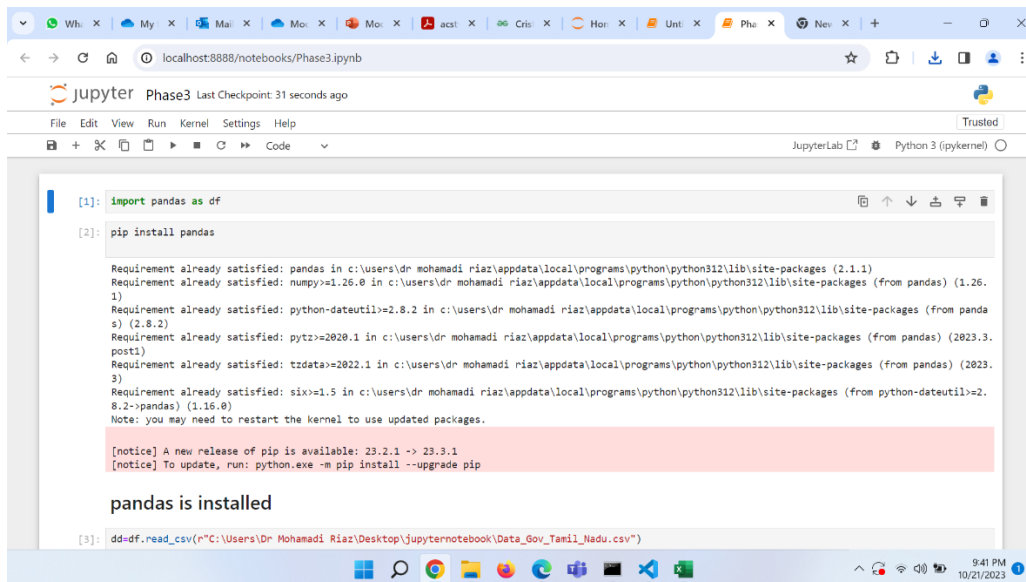
REGISTER NUMBER: 61772131018

DATA IMPORT:

- The given dataset is imported into jupyter notebook. The required modules are imported to perform the cleaning operation.
- The necessary libraries are imported by the following commands:

  pip install pandas

  import pandas as df

DATA CLEANING:

Data cleaning means fixing bad data in the dataset. The bad data could be empty cells, data in wrong format and wrong data.

Data cleaning is a critical step in preparing data for the prediction of company registration using AI-driven exploration. Clean and well-structured data is essential for building accurate and reliable predictive models. Here are the steps to clean the data:

- Handling Missing Data

- Handling Duplicates

- Data Transformation

After the data cleaning is done, the number of cells in the Excel sheet in the final dataset is reduced to 73740 by removing the duplicate data, wrong data.

DATA ANALYSIS:

Data analysis is a systematic approach which follows the process of inspecting, cleaning, transforming and interpreting data to extract valuable insights.
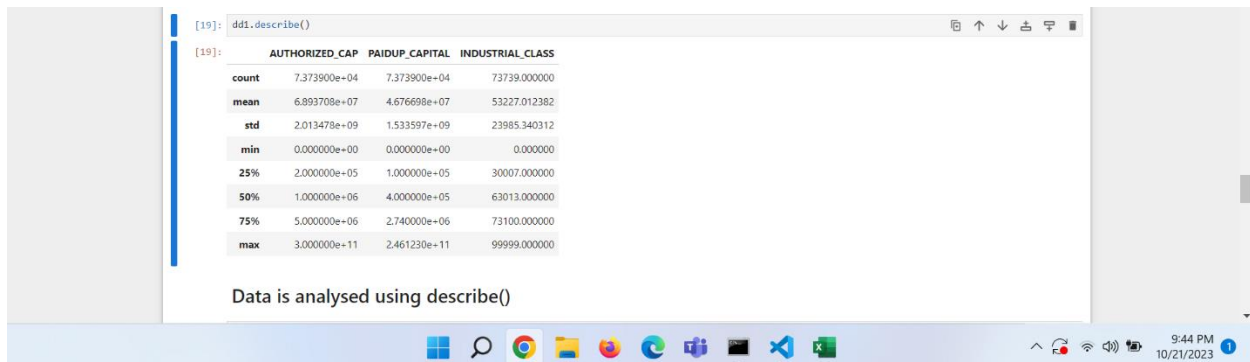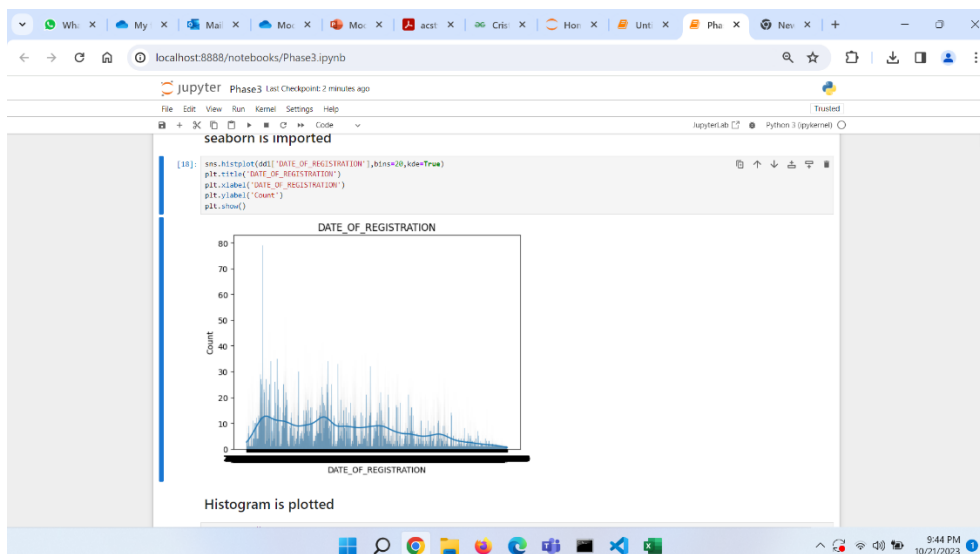

DATA VISUALIZATION:

Data visualization is the representation of data through use of common graphics, such as charts, plots, infographics, and even animations.
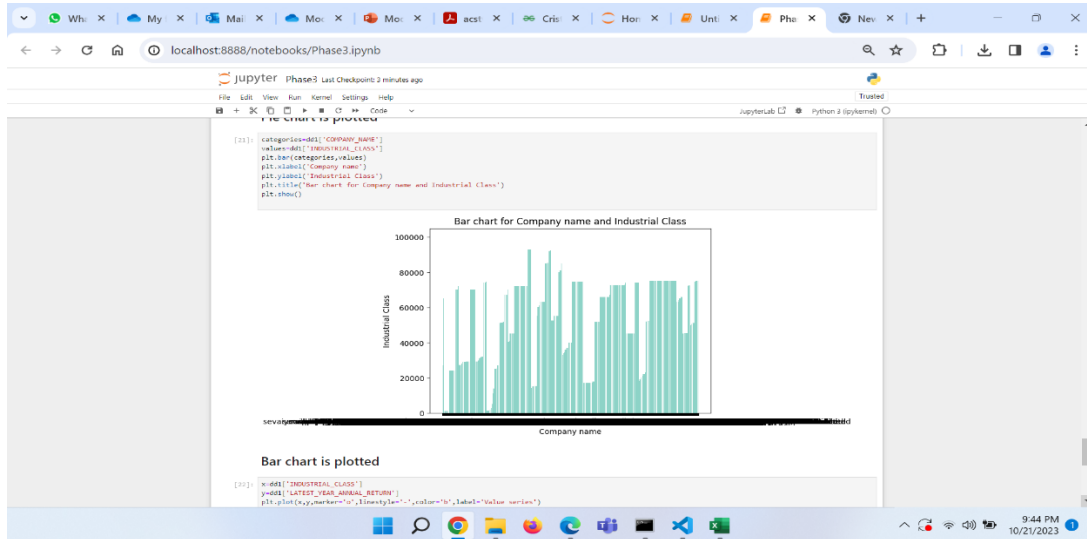
In python, it is done using matplotlib library that is imported to visualize different charts.

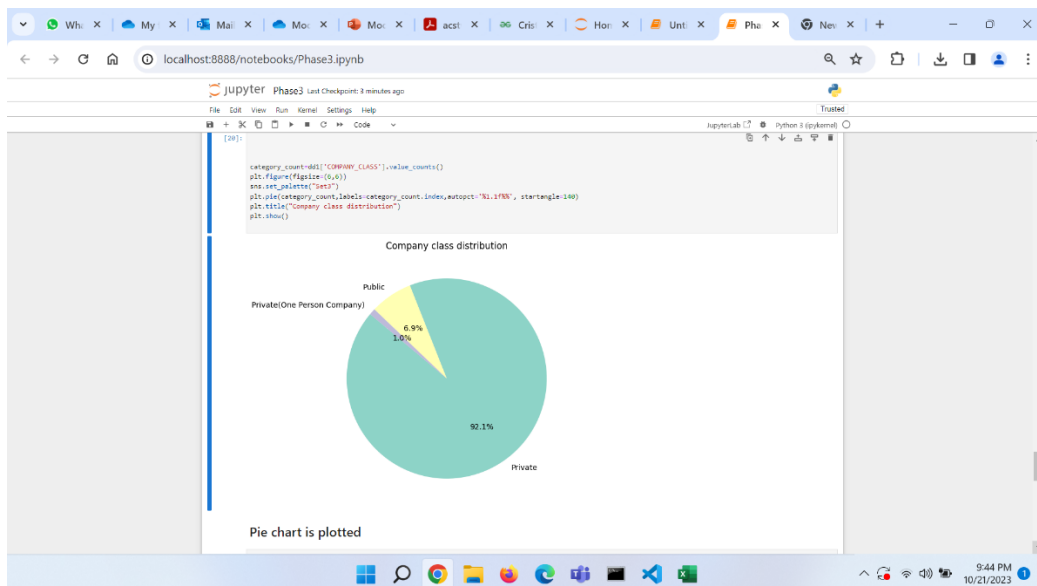## DATA VISUALIZATION USING HISTOGRAM:

Histogram is plotted for the "DATE_OF_REGISTRATION" and "COUNT" in x and y axes respectively.
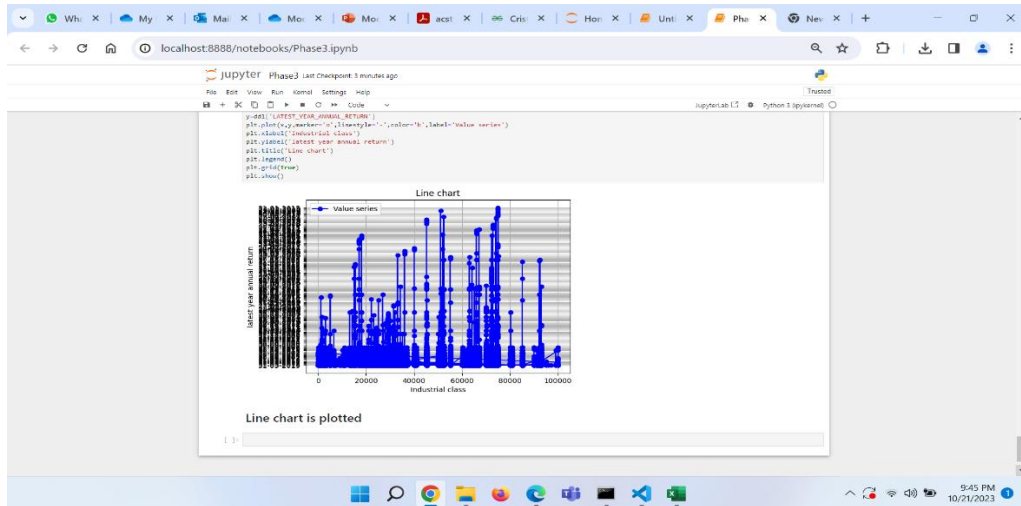
DATA VISUALIZATION USING BARCHART:

Bar chart is plotted for "COMPANY_NAME" and "INDUSTRIAL_CLASS" in x and y axes respectively.

DATA VISUALIZATION USING PIE CHART:

Data visualization is done using pie chart for "COMPANY_CLASS_DISTRIBUTION".

DATA VISUALIZATION USING LINECHART:

Data visualization is done using line chart for "INDUSTRIAL_CLASS" and "LATEST_YEAR_ANNUAL_RETURN".