AI\_Phase3

| **Serial Number** | **Team Member Name** | **Registration Number** |
| --- | --- | --- |
| 1 | K.Navinraj | 310821104064 |
| 2 | M.P.Praveen Raja | 310821104070 |
| 3 | Sachin A | 310821104081 |
| 4 | K.P.Tharun | 310821104100 |

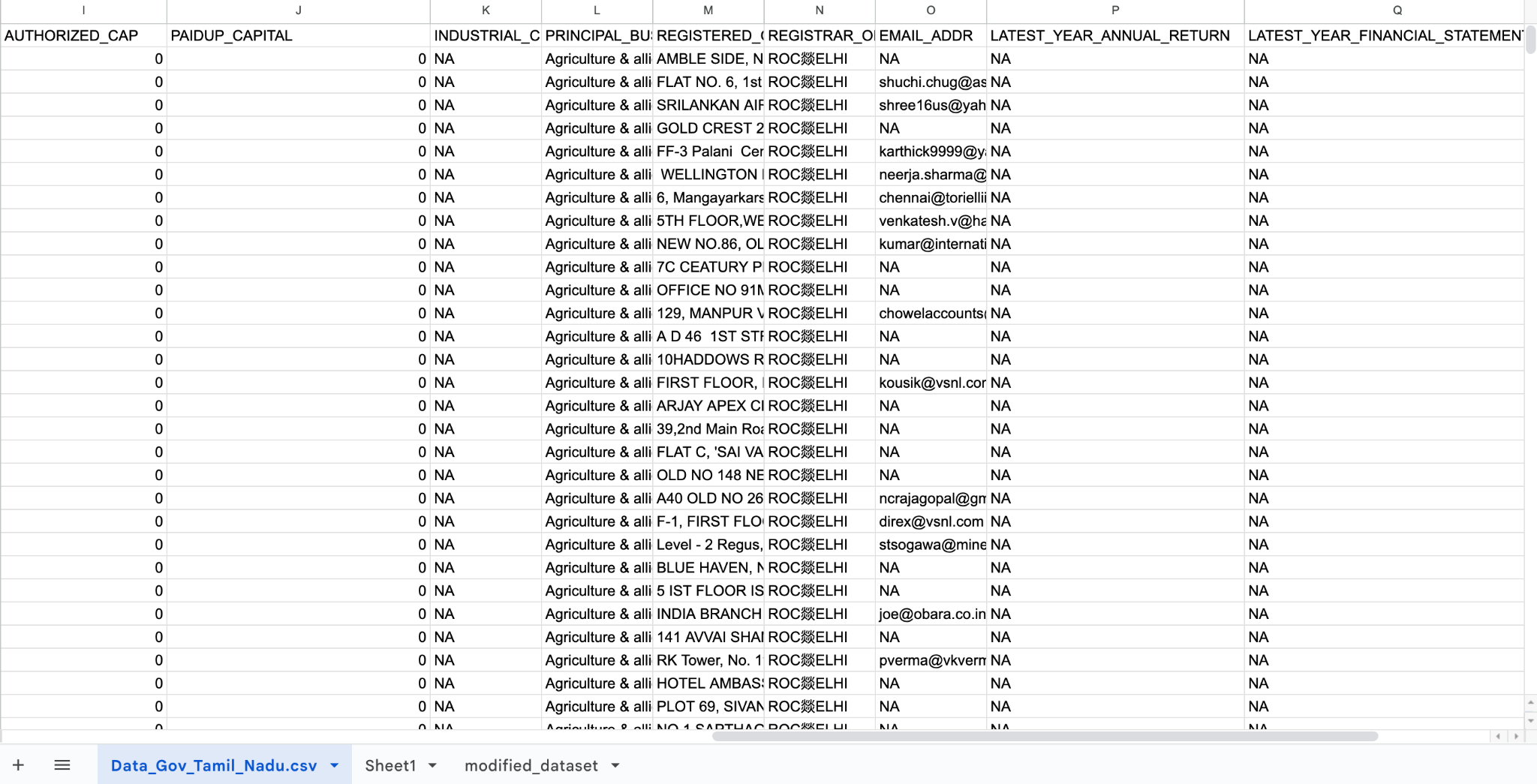
### Objective:

The technology stack described in the plan encompasses a range of tools and frameworks that are commonly used in data integration, preprocessing, modelling, real-time data processing, visualisation, deployment, monitoring, documentation, training, compliance, evaluation, and scaling. Let's elaborate on some of the key technologies and their purposes:

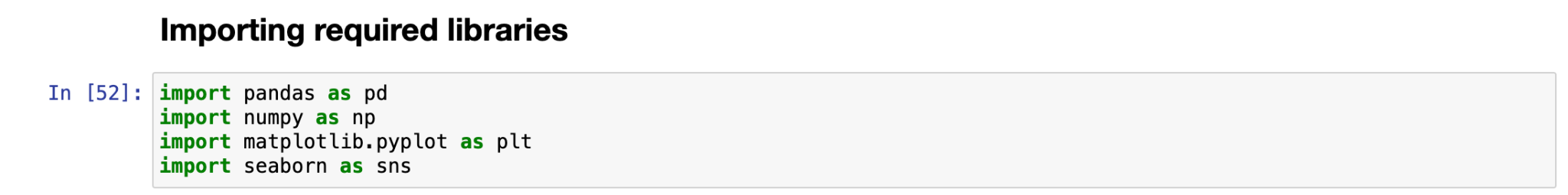
### 1.Dataset:

DatasetLink: [Company Master Data of Tamil Nadu upto 28th February 2019 | Open Government Data (OGD) Platform India](https://tn.data.gov.in/resource/company-master-data-tamil-nadu-upto-28th-february-2019)

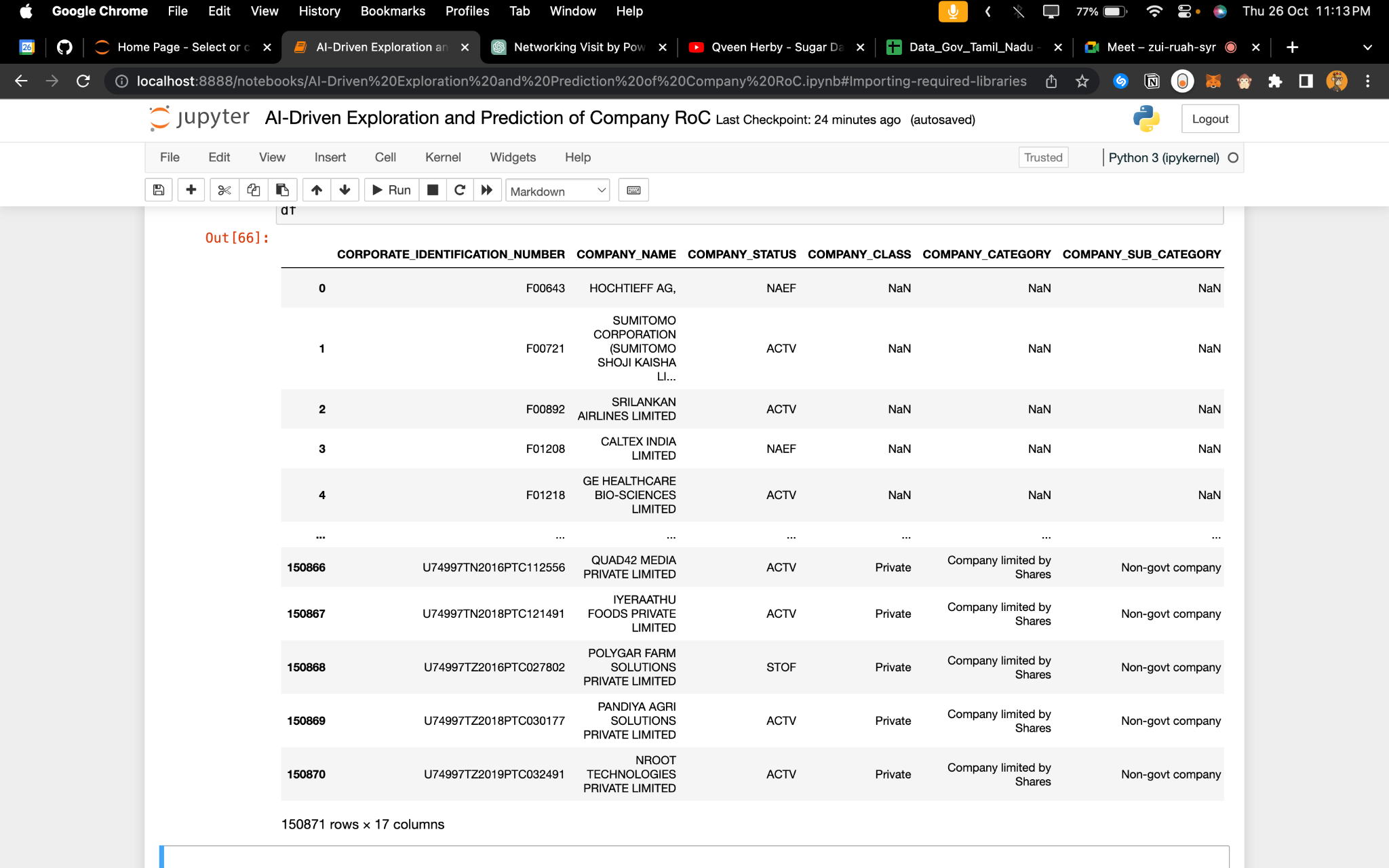
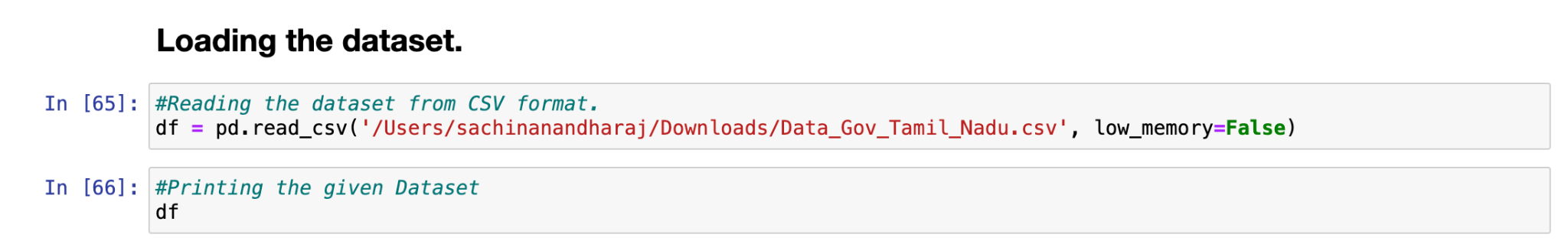




### 2. Required libraries



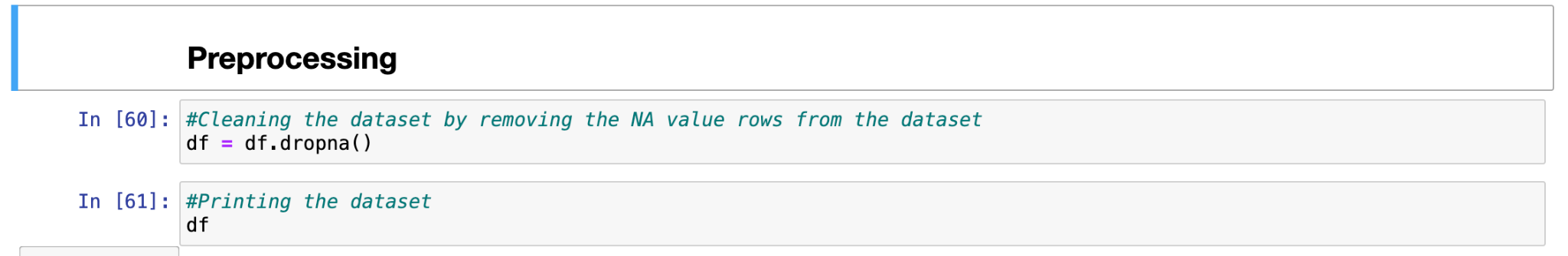
### 3. Loading the Dataset

We will load the dataset named “dialogs.txt” inside our python notebook using the pd.read\_csv() method. To load the dataset for AI-driven exploration and prediction of company registration trends with the Registrar of Companies (RoC), you'll need to follow a series of steps. In this content guide, I'll outline the process for loading the dataset and preparing it for analysis using artificial intelligence (AI) techniques.

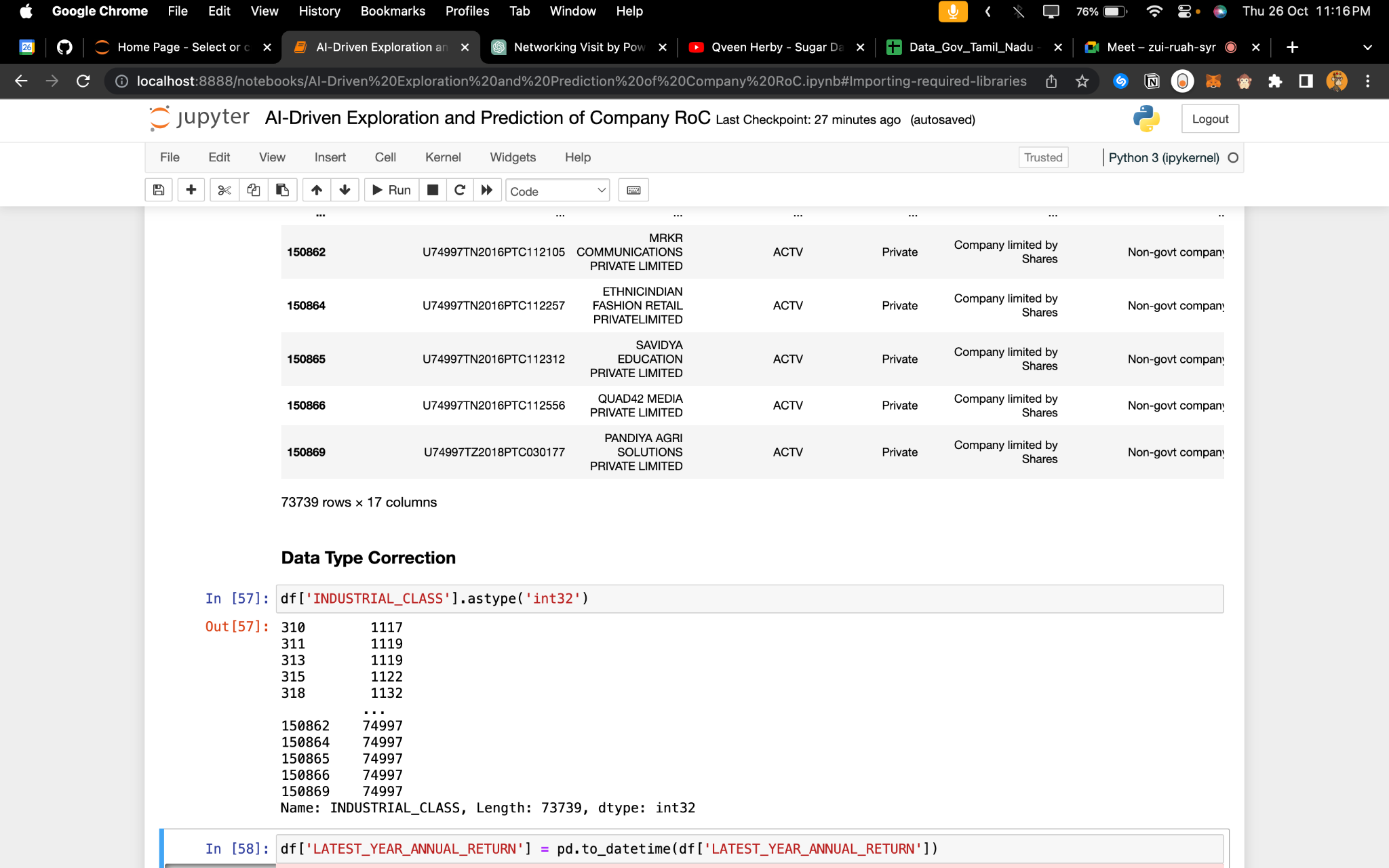
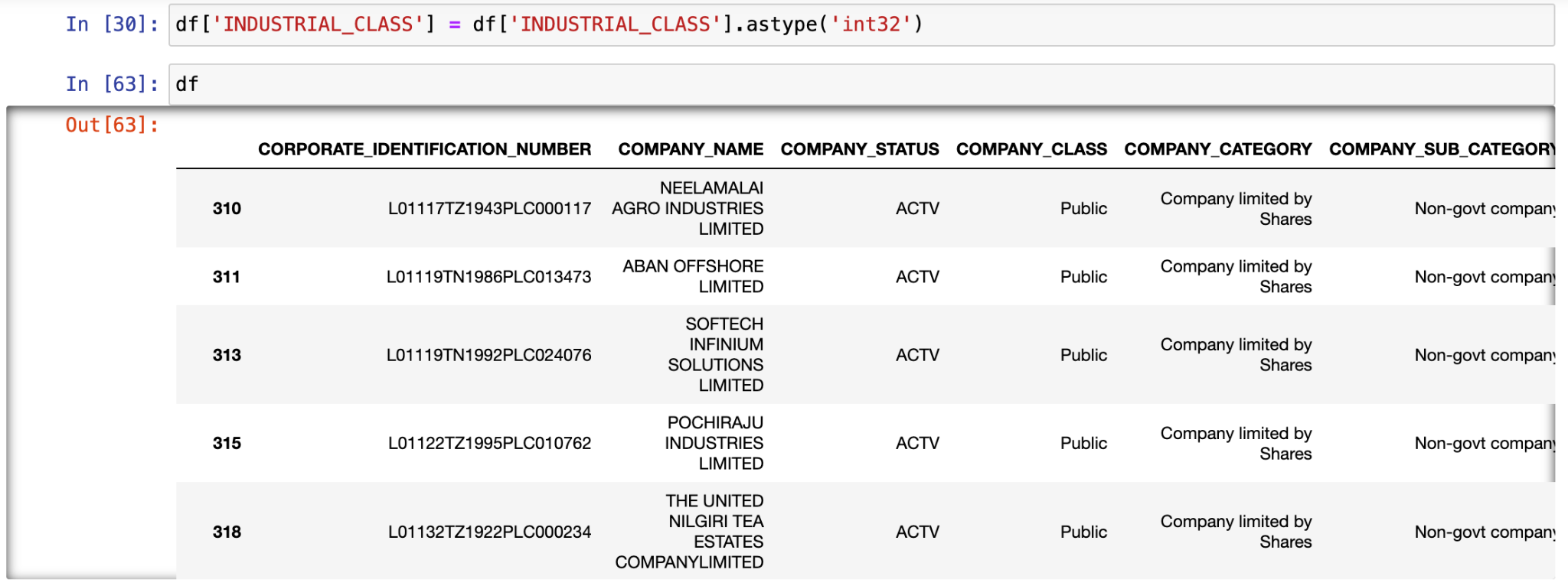
### 4. Pre-processing

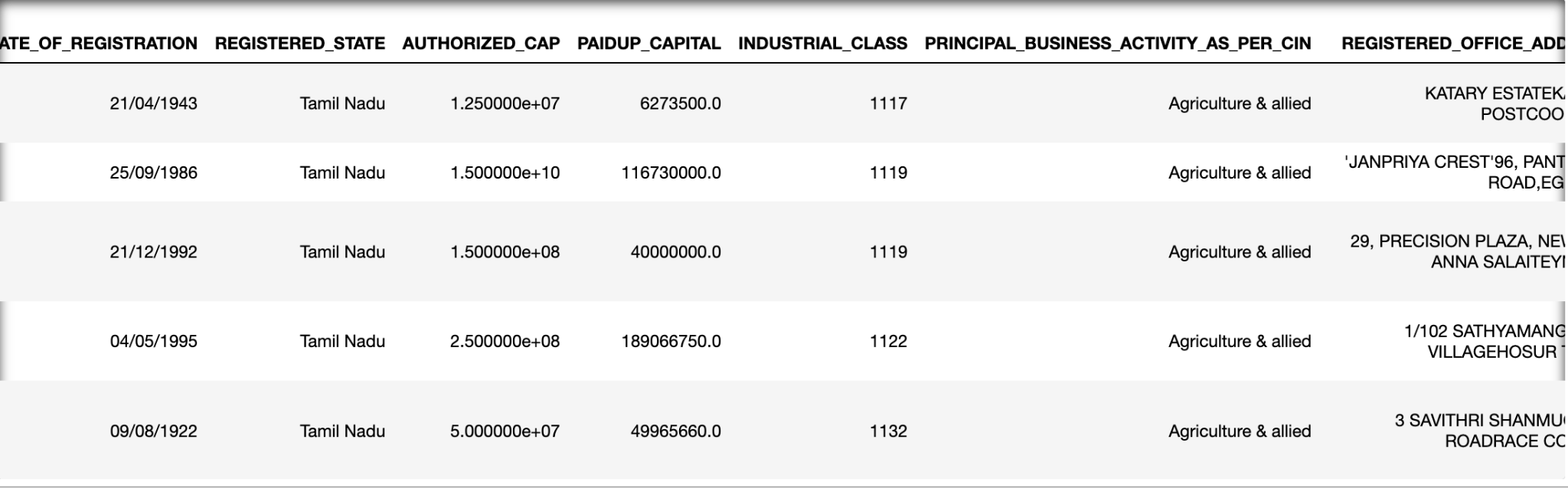
Preprocessing is a crucial step in AI-driven exploration and prediction of company registration trends with the Registrar of Companies (RoC). Proper preprocessing ensures that your dataset is clean, well-structured, and suitable for machine learning algorithms.

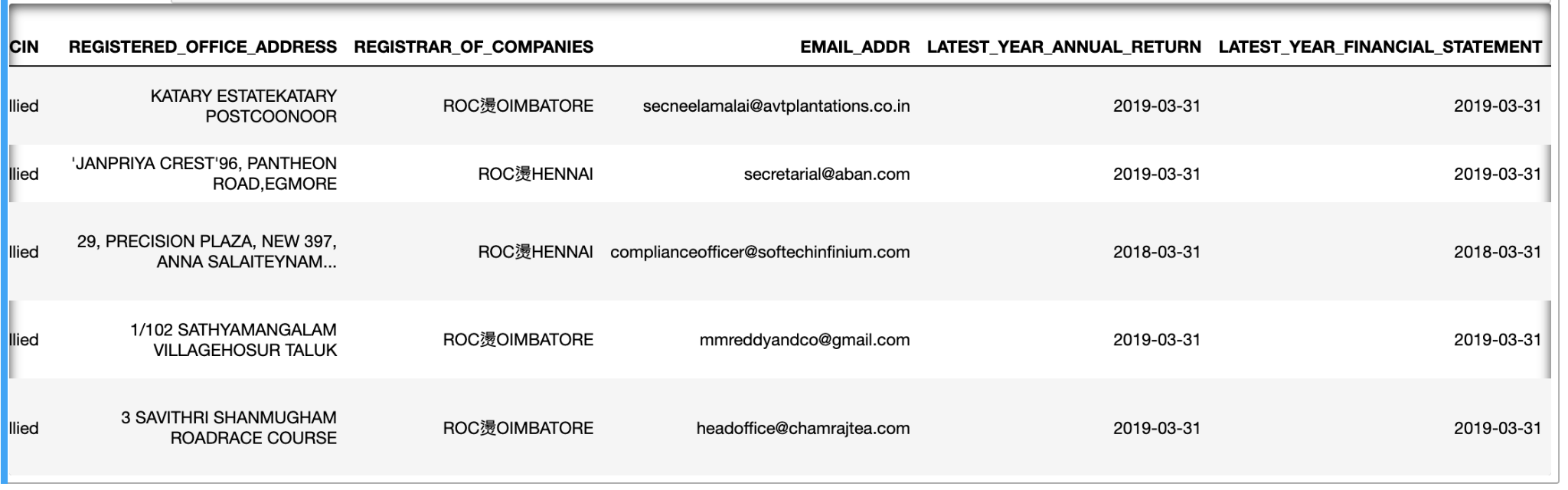
a often contains noise, irrelevant information, and inconsistencies that can interfere with the accuracy of NLP models.



### 5. Data Type Correction



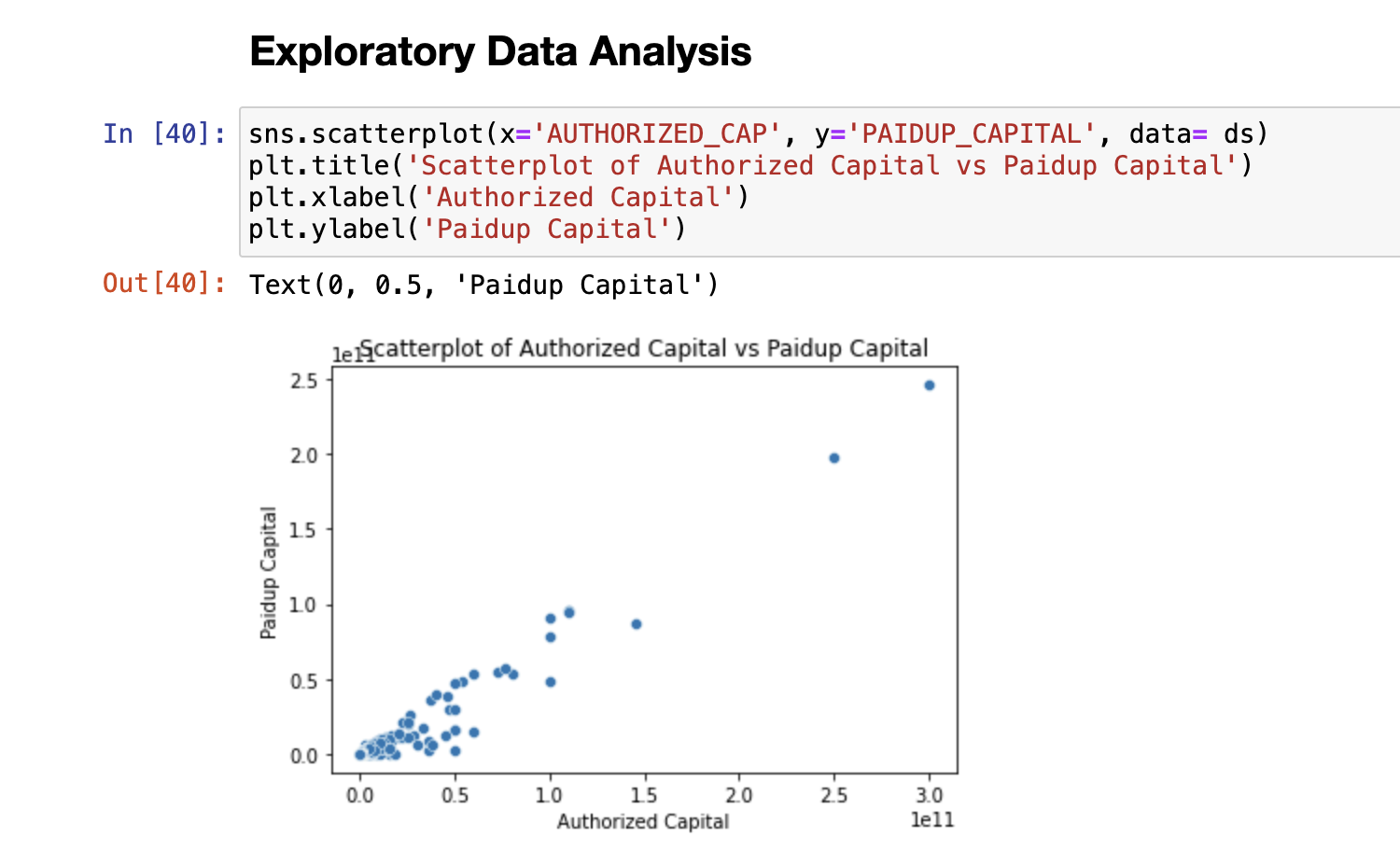




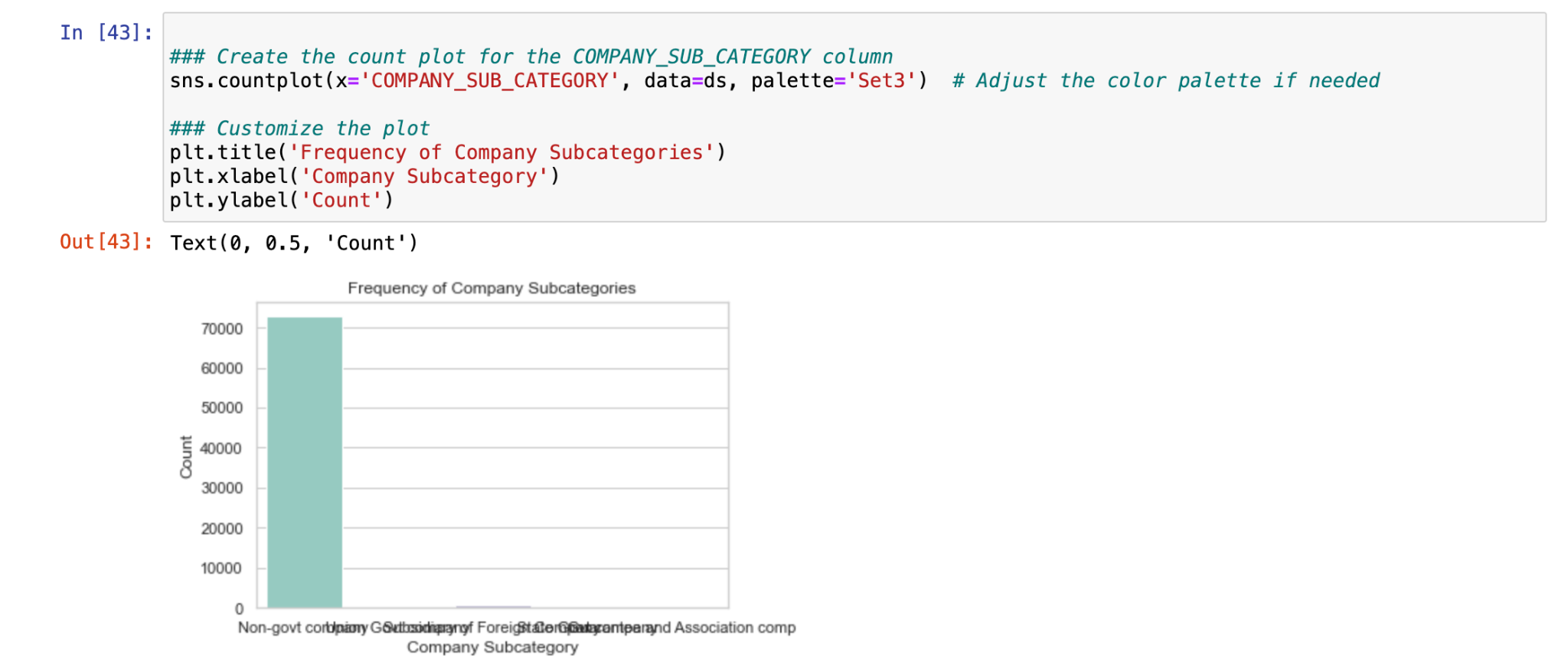
**6. Exploratory Data Analysis**

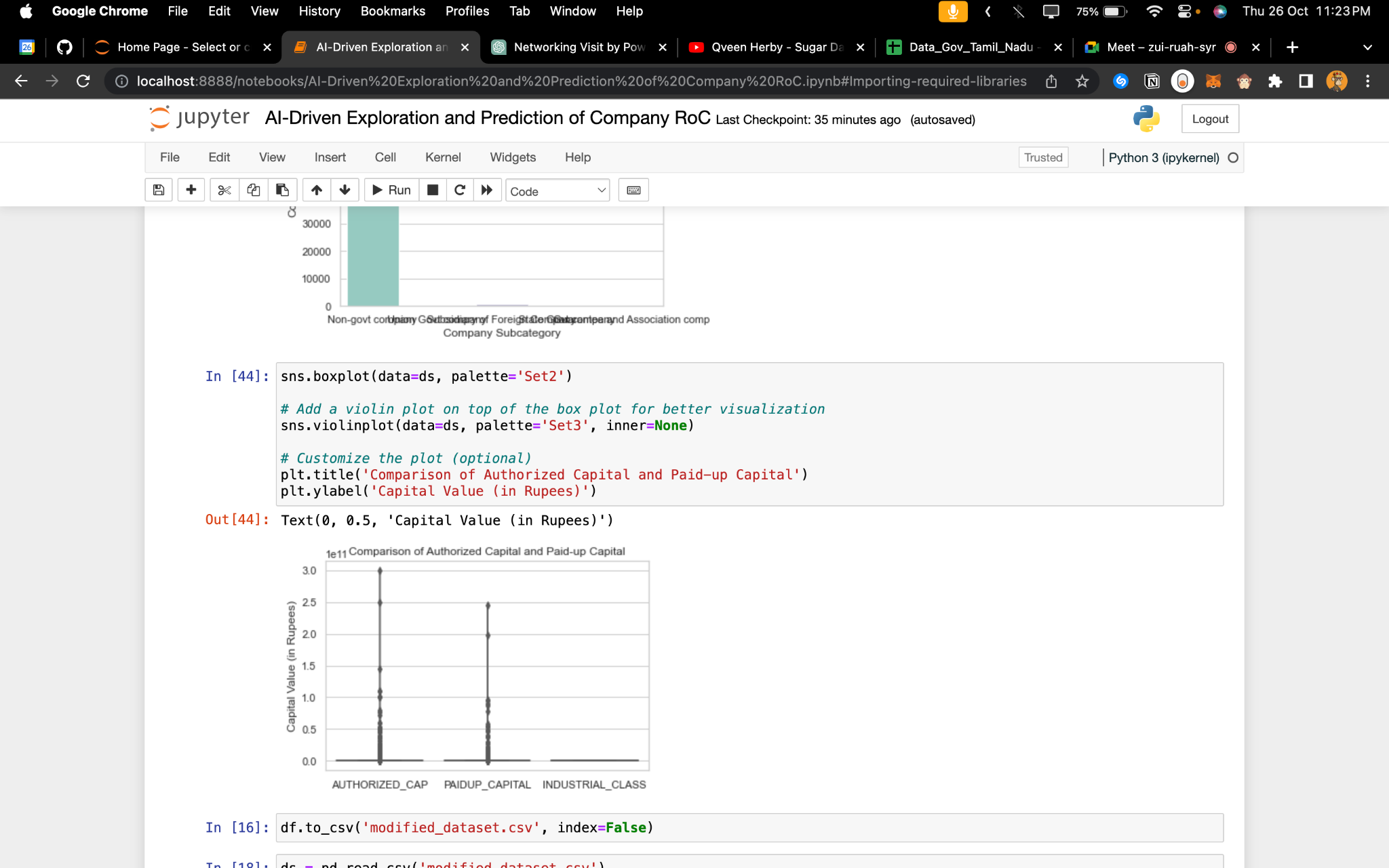
Exploratory data analysis (EDA) is used to analyses and investigate data sets and summarise their main characteristics, often employing data visualisation methods

AI-driven exploration and prediction of company registration trends with the Registrar of Companies (RoC) involves a series of steps, including data collection, cleaning, transformation, loading, EDA, modelling, and deployment. Careful attention to data quality and preprocessing is essential for building effective AI models in this context.









### 6. Extracting preprocessed csv file dataset

After performing data cleaning, the refined dataset can be saved and downloaded as a distinct CSV file.