

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

NAME: KARTHIKA A

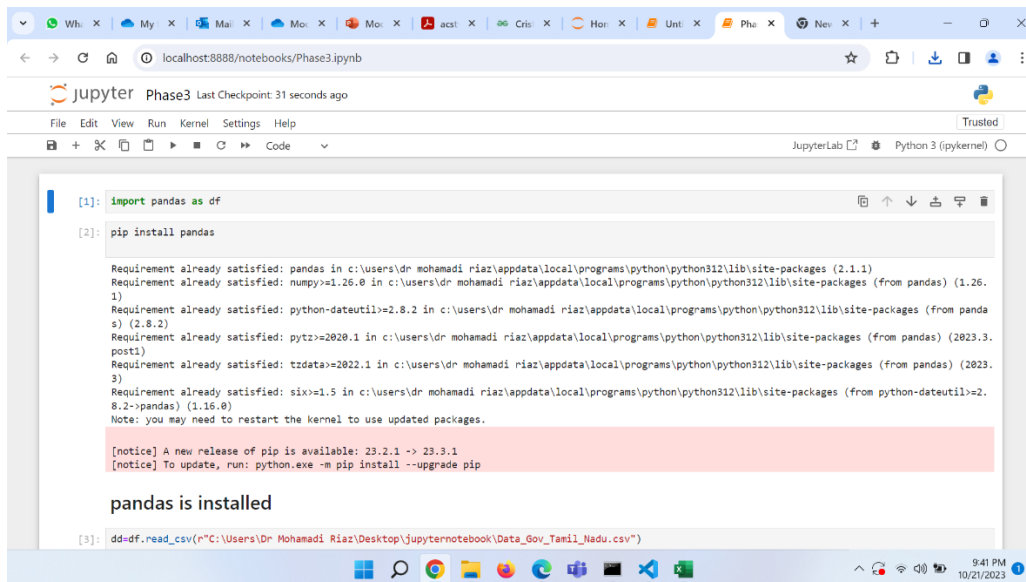
REGISTER NUMBER: 61772131018

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
```



The screenshot shows a Jupyter Notebook window titled "Phase3" with a last checkpoint 31 seconds ago. The notebook is running on Python 3 (ipykernel). The code cell contains the following commands:

```
[1]: import pandas as df
[2]: pip install pandas
```

The output of the second command shows that pandas is already installed (version 2.1.1) and its dependencies are satisfied. A notice indicates that a new release of pip is available (23.2.1 -> 23.3.1) and suggests running `python.exe -m pip install --upgrade pip` to update it.

pandas is installed

```
[3]: dd=df.read_csv(r"C:\Users\Dr Mohamadi Riaz\Desktop\jupyternotebook\Data_Gov_Tamil_Nadu.csv")
```

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:

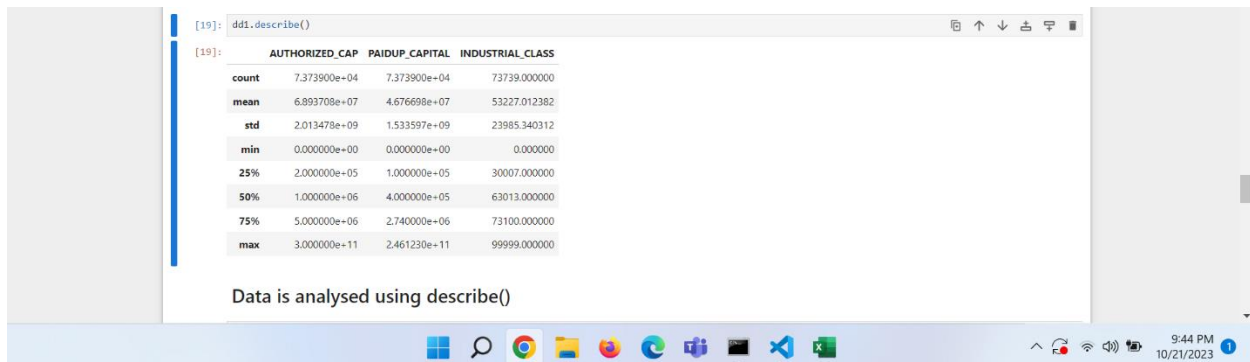
- Handling Missing Data
- Handling Duplicates
- Data Transformation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
73723	U74996TNj.s.exim pr	ACTV	Private	Company Non-govt	7/9/2006	Tamil Nad	50000000	100000	74996 Real estat	No. 55, Nair	ROCA	CHEjsexim	pltd	31-03-201	31-03-2019				
73724	U74996TNyouth spo	ACTV	Private	Company Non-govt	4/5/2016	Tamil Nad	100000	100000	74996 Real estat	No.45, Ter	ROCA	CHEaskask	199	31-03-201	31-03-2019				
73725	U74996TNlatentech	ACTV	Private	Company Non-govt	3/8/2016	Tamil Nad	1000000	100000	74996 Real estat	330, UG C	ROCA	CHEsales	@lati	31-03-201	31-03-2019				
73726	U74996TNonena ecc	ACTV	Private	Company Non-govt	18-08-201	Tamil Nad	100000	100000	74996 Real estat	No Y-44, K	ROCA	CHEsangeeth	..	31-03-201	31-03-2019				
73727	U74996TNvlightup in	ACTV	Private(Or	Company Non-govt	17-08-201	Tamil Nad	100000	100000	74996 Real estat	E5(EBLOC	ROCA	CHEvivek	@sai	31-03-201	31-03-2019				
73728	U74996TNhyrespot r	ACTV	Private(Or	Company Non-govt	#####	Tamil Nad	100000	100000	74996 Real estat	NO.2, VENI	ROCA	CHEPRIYAKUM		31-03-201	31-03-2019				
73729	U74996TNagr logistic	ACTV	Private	Company Non-govt	9/3/2018	Tamil Nad	1000000	200000	74996 Real estat	D-2, Third	ROCA	CHEagr	@agric	31-03-201	31-03-2019				
73730	U74996TNshanpros	ACTV	Private(Or	Company Non-govt	20-03-201	Tamil Nad	1000000	100000	74996 Real estat	438, FIRST	ROCA	CHEkrjb	7777	31-03-201	31-03-2019				
73731	U74996TNaddthone	ACTV	Private	Company Non-govt	26-03-201	Tamil Nad	100000	100000	74996 Real estat	NO.9/8, M	ROCA	CHEcomplanc		31-03-201	31-03-2019				
73732	U74996TZbellie & hi	ACTV	Private	Company Non-govt	24-08-201	Tamil Nad	1000000	100000	74996 Real estat	NO.9, HUD	ROCA	COI sashimani		31-03-201	31-03-2019				
73733	U74996TNadvaith p	STOF	Private	Company Non-govt	19-11-201	Tamil Nad	100000	100000	74997 Real estat	193, Peter	ROCA	CHEjbadrin	@g	31-01-201	31-01-2015				
73734	U74996TNglassic eye	ACTV	Private	Company Non-govt	16-04-201	Tamil Nad	12200000	10781630	74997 Real estat	Fat No.6-I	ROCA	CHEkailash	@g	31-03-201	31-03-2019				
73735	U74996TNauroflux t	ACTV	Private	Company Non-govt	23-04-201	Tamil Nad	500000	500000	74997 Real estat	Plot No. 3,	ROCA	CHEmcmadha		31-03-201	31-03-2019				
73736	U74996TNmrkr comi	ACTV	Private	Company Non-govt	22-08-201	Tamil Nad	100000	100000	74997 Real estat	11C, AASA	ROCA	CHEKRPR	PRIVAI	31-03-201	31-03-2019				
73737	U74996TNethnicindi	ACTV	Private	Company Non-govt	30-08-201	Tamil Nad	1000000	1000000	74997 Real estat	VNR Milfo	ROCA	CHEALEX	CHAI	31-03-201	31-03-2019				
73738	U74996TNSavidya ec	ACTV	Private	Company Non-govt	1/9/2016	Tamil Nad	1000000	100000	74997 Real estat	Mahalaksl	ROCA	CHEsridevis	77	31-03-201	31-03-2019				
73739	U74996TNquad42 m	ACTV	Private	Company Non-govt	19-09-201	Tamil Nad	1000000	100000	74997 Real estat	Old No.37	ROCA	CHEezhil	@qu	31-03-201	31-03-2019				
73740	U74996TZpandiya a	ACTV	Private	Company Non-govt	16-03-201	Tamil Nad	2500000	1500000	74997 Real estat	10/10 C3,	ROCA	COI sathishpai		31-03-201	31-03-2019				

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.

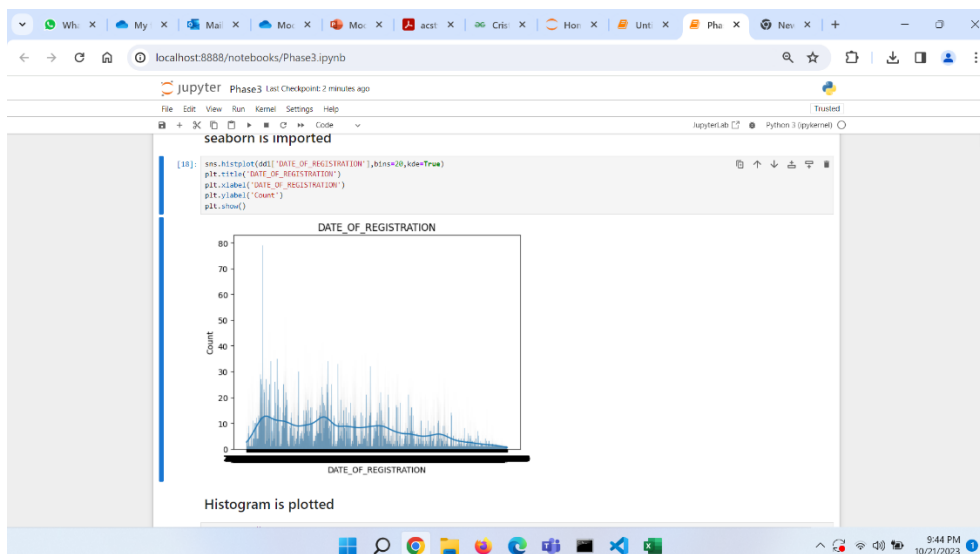


DATA VISUALIZATION:

Data visualization is the representation of data through use of common graphics, such as charts, plots, infographics, and even animations.

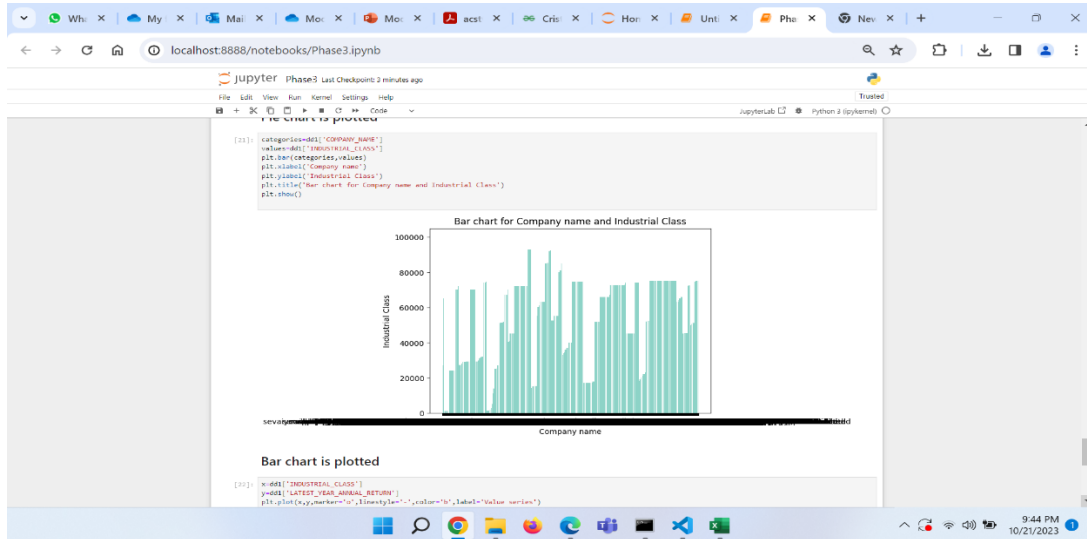
In python, it is done using matplotlib library that is imported to visualize different charts.

DATA VISUALIZATION USING HISTOGRAM:



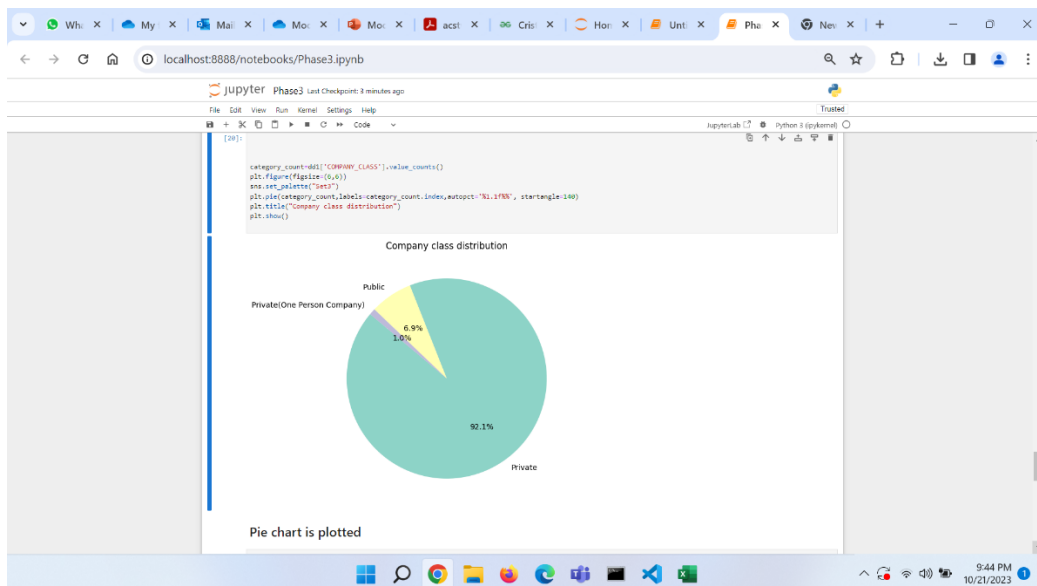
Histogram is plotted for the “DATE_OF_REGISTRATION” and “COUNT” in x and y axes respectively.

DATA VISUALIZATION USING BARCHART:



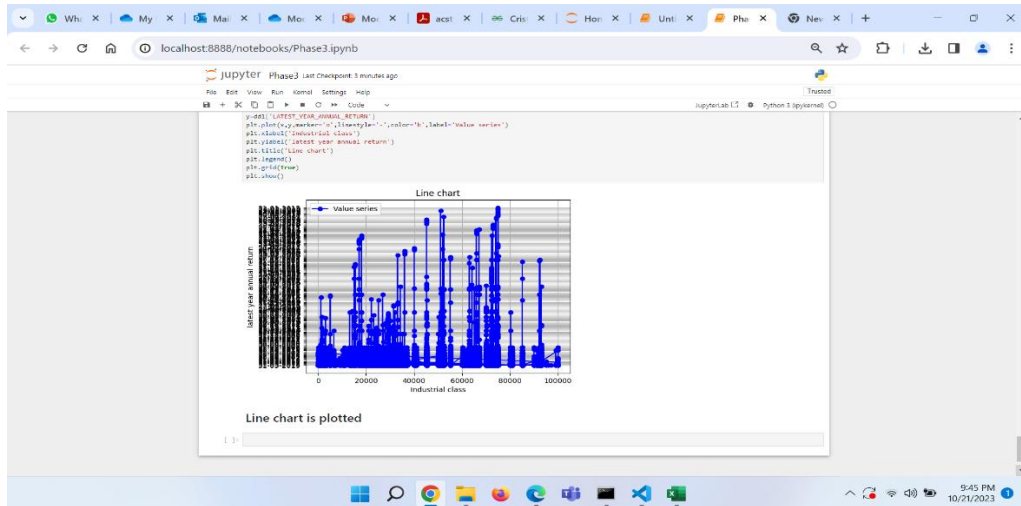
Bar chart is plotted for “COMPANY_NAME” and “INDUSTRIAL_CLASS” in x and y axes respectively.

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”.