Phase-5

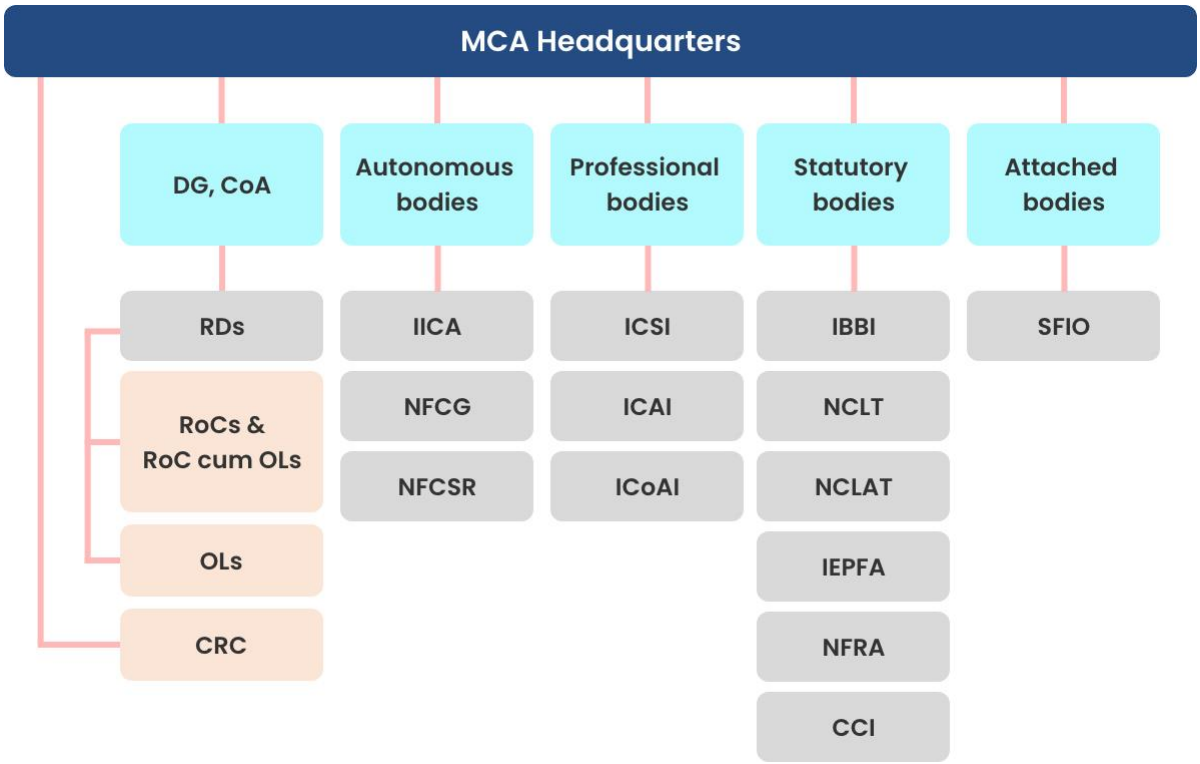
[AI-DRIVEN EXPLORATION AND PREDICTION OF COMPANY REGISTRATION TRENDS WITH REGISTRAR OF COMPANIES (ROC) IN DATA PRE PROCESSING]

COMPANY MASTER DATA

Get data on master details of any company registered with Registrar of Companies (RoC). Data contains various information like Corporate Identification Number(CIN), Company Name, Company Status, Company Class, Company Category, Authorized Capital in INR, Paid-up Capital in INR, Date of Registration, Registered State, Registrar of Companies, Principal Business Activity, Registered Office Address and Sub Category.

ROC-WISE STATISTICS OF PROSECUTIONS

Get the data regarding ROC-wise statistics of prosecutions. The data has been published by Ministry of Corporate Affairs.Economic- Activity wise Active Non-Government CompaniesThe data describes details on Economic- Activity wise Active Non-Government Companies. Data has been published by The Ministry of Corporate Affairs, Government of India.

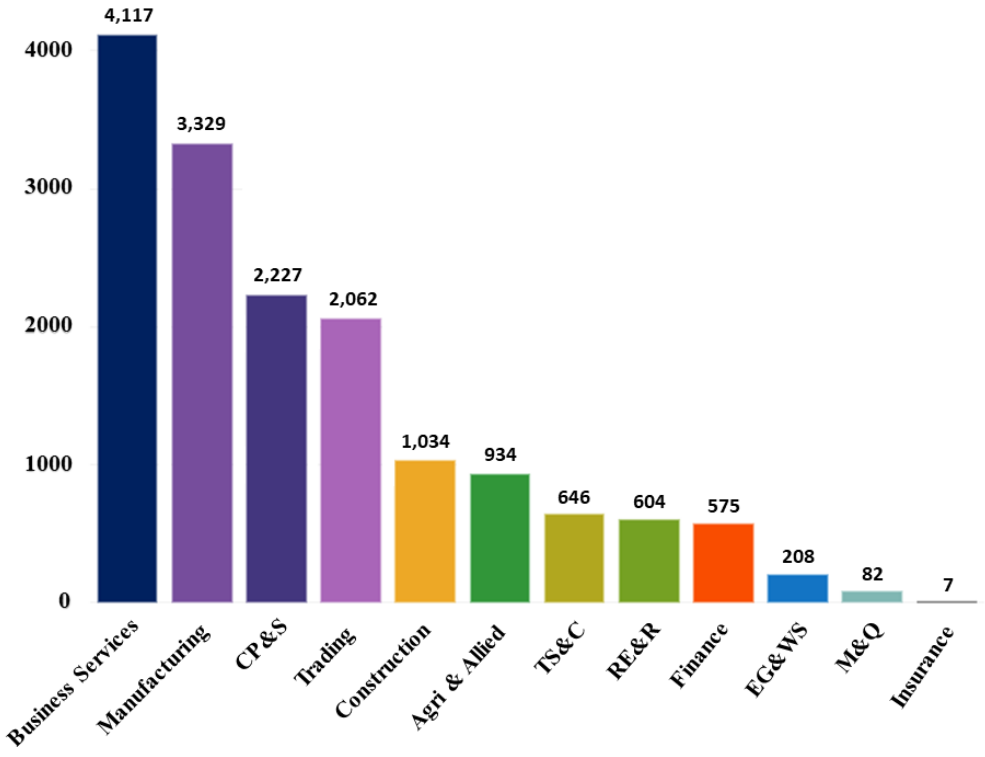


DATA PREPROCESSING:

Clean the data by handling missing values, duplicates, and outliers convert text-based information into numerical features using techniques like one-hot encoding or word embedding.

Exploratory Data Analysis (EDA):

Perform data visualization and statistical analysis to understand trends, correlations, and patterns in the data.Identify factors that might influence company registrations, such as economic indicators or regional demographics



PREDICTIVE MODELLING:

Predictive modeling is a mathematical process used to predict future events or outcomes by analyzing patterns in a given set of input data. It is a crucial component of predictive analytics, a type of data analytics which uses current and historical data to forecast activity, behavior and trends.

Examples of predictive modeling include estimating the quality of a sales lead, the likelihood of spam or the probability someone will click a link or buy a product. These capabilities are often baked into various business applications, so

it is worth understanding the mechanics of predictive modeling to troubleshoot and improve performance.

Although predictive modeling implies a focus on forecasting the future, it can also predict outcomes (e.g., the probability a transaction is fraudulent). In this case, the event has already happened (fraud committed). The goal here is to predict whether future analysis will find the transaction *is* fraudulent. Predictive modeling can also forecast future requirements or facilitate what-if analysis.

"Predictive modeling is a form of data mining that analyzes historical data with the goal of identifying trends or patterns and then using those insights to predict future outcomes," explained Donncha Carroll a partner in the revenue growth practice of Axiom Consulting Partners. "Essentially, it asks the question, 'have I seen this before' followed by, 'what typically comes after this pattern.'"

TOP TYPES OF PREDICTIVE MODELS

There are many ways of classifying predictive models and in practice multiple types of models may be combined for best results. The most salient distinction is between unsupervised versus supervised models.

> Unsupervised models use traditional statistics to classify the data directly, using techniques like logistic regression, time series analysis and decision trees.Supervised models use newer machine learning techniques such as neural networks to identify patterns buried in data that has already been labeled.

FEATURE ENGINEERING :

Private Limited is an unlisted private company incorporated on 31 March, 2023. It is classified as a private limited company and is located in , Karnataka. It's authorized share capital is INR 10.00 lac and the total paid-up capital is INR 1.00 lac.

The current status of Feature Engineering Private Limited is - Active.

Details of the last annual general meeting of Feature Engineering Private Limited are not available. The company is yet to submit its first full-year financial statements to the registrar.

Feature Engineering Private Limited has two directors - Shubhangi Praveen Passi and Sudhanshu Praveen Passi.

The Corporate Identification Number (CIN) of Feature Engineering Private Limited is U62099KA2023PTC171793. The registered office of Feature Engineering Private Limited is at Prestige Atlanta, 80 Feet Rd, Koramangala 1A Block,Koramangala 3 Block, Koramangala Bangalore South , Karnataka.

NAME	INCORPORATION YEAR	STATE	PAID UP CAPITAL	
SRI MUNDRIKA INFOTECH PRIVATE LIMITED	2023	Bihar	1.00 lac	Buy financial reports
BLUE ARROW TECHNOLOGY CONSULTING SERVICES PRIVATE LIMITED	2023	Karnataka	1.00 lac	Buy financial reports
SADGURU CONTROLS AND AUTOMATION PRIVATE LIMITED	2023	Maharashtra	1.00 lac	Buy financial reports
TECREC TECHNOLOGIES PRIVATE LIMITED	2023	Andhra Pradesh	1.00 lac	Buy financial reports
TECHFORTH ITSEVA PRIVATE LIMITED	2023	West Bengal	1.00 lac	Buy financial reports
BEAMIO ENGINEERING SOLUTIONS PRIVATE LIMITED	2023	Andhra Pradesh	1.00 lac	Buy financial reports

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis or EDA is used to take insights from the data. Data Scientists and Analysts try to find different patterns,

relations, and anomalies in the data using some statistical graphs and other visualization techniques. Following things are part of EDA :

1. Get maximum insights from a data set
2. Uncover underlying structure
3. Extract important variables from the dataset
4. Detect outliers and anomalies(if any)
5. Test underlying assumptions
6. Determine the optimal factor settings

EDA IS IMPORTANT

The main purpose of EDA is to detect any errors, outliers as well as to understand different patterns in the data. It allows Analysts to understand the data better before making any assumptions. The outcomes of EDA helps businesses to know their customers, expand their business and take decisions accordingly.

COMPANY_NAME
HOCHTIEFF AG,
SUMITOMO CORPORATION (SUMITOMO SHOJI KAISHA LIMITED)
SRILANKAN AIRLINES LIMITED
CALTEX INDIA LIMITED
GE HEALTHCARE BIO-SCIENCES LIMITED
CAIRN ENERGY INDIA PTY. LIMITED
TORIELLI S.R.L
HARDY EXPLORATION & PRODUCTION (INDIA) INC..
HOCHTIOF AKTIENGESELLSHARFF VORM GFBR HELFMANN
EPSON SINGAPORE PVT LTD

Table 1: Summary Statement on Companies in India as on 30th June, 2022		
	Number of Registered Companies	**23,66,549**
1	**Number of Active companies**	**14,81,069**
	Of which	
i	*Companies Limited by Shares*	14,71,127
ii	*Companies Limited by Guarantee*	9,643
iii	*Unlimited Companies*	299
2	**Number of Closed Companies**	**8,47,513**
i	*No. of Companies Liquidated/Dissolved*	11,334
ii	*No. of Companies Defunct/ Struck-off (Section 248 of CA, 2013)*	7,84,841
iii	*No. of Companies Amalgamated/merged*	29,722
iv	*No. of Companies Converted to LLP*	16,740
v	*No. of Companies Converted to LLP and dissolved*	4,876
3	**Number of Companies lying dormant u/s 455 of Companies Act, 2013**	**2,441**
4	**Number of Companies under Liquidation**	**7,041**
5	**Number of Companies which are under the process of Striking-Off (Section 248 of the Companies Act, 2013 and Section 560 of Companies Act 1956)**	**28,485**

Table 2: Active Companies as on 30th June, 2022				
S. No.	**Particulars**	**Government**	**Non-Government**	**Total**
1	**Companies Limited by Shares**	**2,198**	**14,68,929**	**14,71,127**
a	Public Limited	1,551	66,536	68,087
i	*Listed*	79	6,662	6,741
ii	*Unlisted*	1,472	59,874	61,346
b	Private Limited	647	14,02,393	14,03,040
	One Person Company	0	45,447	45,447
2	**Companies Limited by Guarantee**	**30**	**9,613**	**9,643**
a	Public Limited	10	2,298	2,308
i	*Listed*	1	14	15
ii	*Unlisted*	9	2,284	2,293
b	Private Limited	20	7,315	7,335
3	**Companies with Unlimited Liability**	**4**	**295**	**299**
	Total	**2,232**	**14,78,837**	**14,81,069**

```
# This Python 3 environment comes with many helpful analytics libraries installed

# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python

# For example, here's several helpful packages to load

import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory

# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os

for dirname, _, filenames in os.walk('/kaggle/input'):

for filename in filenames:

print(os.path.join(dirname, filename))

# You can write up to 20GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version using "Save & Run All"

# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session
```

/kaggle/input/monthly-list-of-legal-entities-registered-in-india/2021_april_registered_companies.csv

In [2]:

```
import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt
```

In [3]:

```
df=pd.read_csv("../input/monthly-list-of-legal-entities-registered-in-india/2021_april_registered_companies.csv")
```

In [4]: linkcode

```
df.head()

df.shape
```

```
df.shape
Out[5]:
(12554, 10)
In [6]:
df.info()
<class 'pandas.core.frame.DataFrame'>
```

RangeIndex: 12554 entries, 0 to 12553
Data columns (total 10 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 company_name 12554 non-null object
1 date_of_registration 12554 non-null object
2 month_name 12554 non-null object
3 state 12554 non-null object
4 roc 12554 non-null object
5 category 12554 non-null object
6 class 12554 non-null object
7 company_type 12554 non-null object
8 activity_code 12554 non-null int64
9 activity_description 12554 non-null object
dtypes: int64(1), object(9)
memory usage: 980.9+ KB
In [7]:
sns.heatmap(df.corr())

```
fig,ax=plt.subplots(1,2,figsize=(20,5))
df['class'].value_counts().plot.pie(explode=None,ax=ax[0],autopct='%1
.1f%%',shadow=True)
ax[0].set_title("Count of Class of Companies")
ax[0].set_ylabel("Count")
sns.countplot("category",data=df,order=df['category'].value_counts().i
ndex)
ax[1].set_title("Count of types of Categories Companies are Listed")

plt.show()
```

Out[5]:

(12554, 10)

In [6]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>

RangeIndex: 12554 entries, 0 to 12553

Data columns (total 10 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 company_name 12554 non-null object

1 date_of_registration 12554 non-null object

2 month_name 12554 non-null object

3 state 12554 non-null object
```

4 roc 12554 non-null object

5 category 12554 non-null object

6 class 12554 non-null object

7 company_type 12554 non-null object

8 activity_code 12554 non-null int64

9 activity_description 12554 non-null object

dtypes: int64(1), object(9)

memory usage: 980.9+ KB

In [7]:

sns.heatmap(df.corr())

fig,ax=plt.subplots(1,2,figsize=(20,5))
df['class'].value_counts().plot.pie(explode=None,ax=ax[0],autopct='%1.1f%%',shadow=True) ax[0].set_title("Count of Class of Companies") ax[0].set_ylabel("Count")
sns.countplot("category",data=df,order=df['category'].value_counts().index) ax[1].set_title("Count of types of Categories Companies are Listed")
plt.show()