**AI-Driven Exploration and Prediction of Company Registration Trends with Registrar of Companies (RoC)**

**TEAM MEMBER**

**Au622021243053 Selladurai S**

**Phase 2 Submission Document**

**Project: Exploration and Prediction of Company Registration Trends**



**Introduction:**

* The Registrar of Companies (RoC) is a government agency that registers and regulates companies in India. It maintains a database of company information, such as name, address, directors, shareholders, financial statements, etc.
* AI-Driven Exploration and Prediction of Company Registration Trends with RoC is a project that aims to use machine learning techniques to analyze the RoC data and discover patterns and insights related to company registration trends in India.
* The project can help various stakeholders, such as entrepreneurs, investors, policymakers, researchers, and regulators, to understand the dynamics of the Indian corporate sector and make informed decisions based on data-driven insights.
* The project can also help to identify potential opportunities and challenges for the Indian economy and society, such as emerging sectors, regional disparities, innovation potential, social impact, etc.
* The project can also provide predictions and forecasts of future company registration trends based on historical data and current conditions, using advanced machine learning models such as neural networks, deep learning, etc.

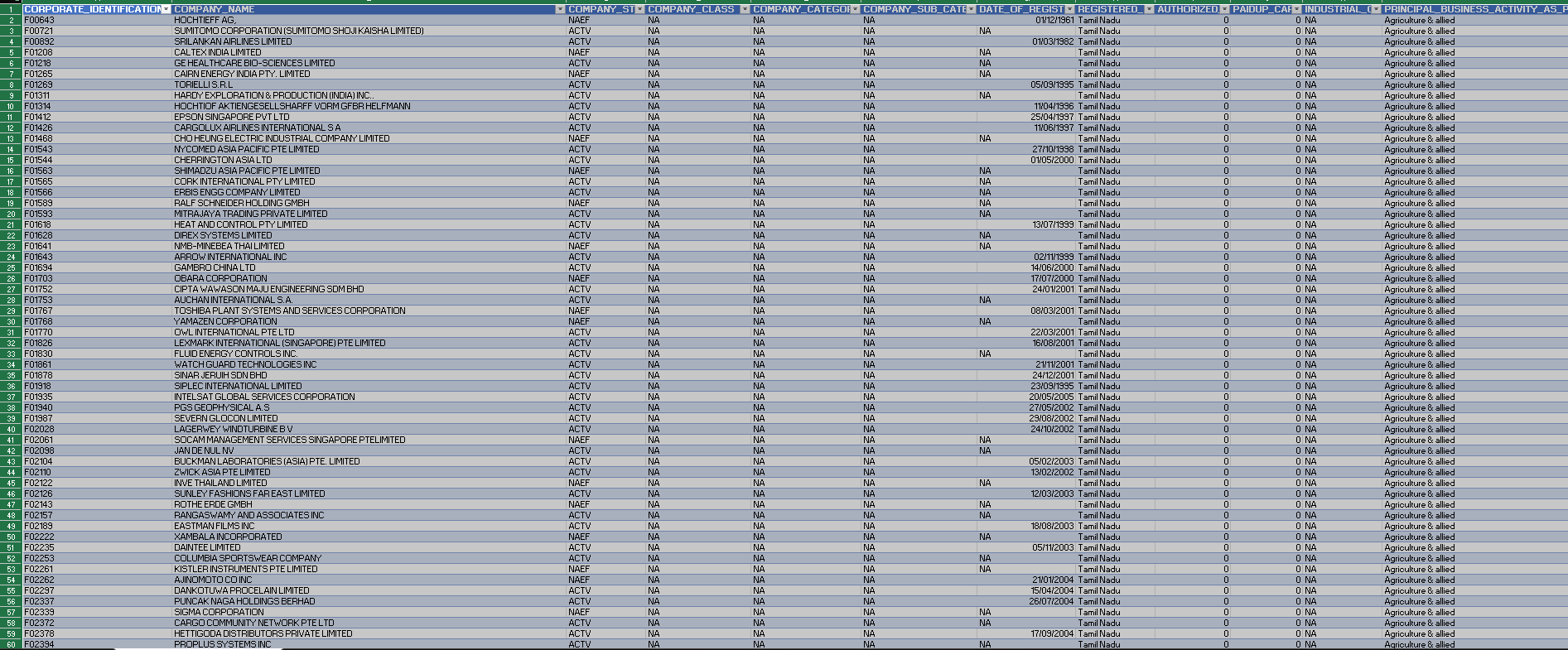
**Content for Project Phase 2 :**

Consider exploring advanced AI algorithms like time series forecasting or ensemble methods for improved predictive accuracy.

**Data Source :**

A good data source Exploration and Prediction of Company Registration Trends should be Accurate, Complete, covering all type of features.

**DatasetLink:**[**https://tn.data.gov.in/resource/company-master-data-tamil-nadu-upto-28th-february-2019**](https://tn.data.gov.in/resource/company-master-data-tamil-nadu-upto-28th-february-2019)

****

**Data Collection and Preprocessing:**

* Data Source: Utilize the dataset containing information about registered companies, including columns like company name, status, class, category, registration date, authorized capital, paid-up capital, and more.
* Data Preprocessing: Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.

**Exploratory Data Analysis (EDA):**

* Perform EDA to understand the distribution, relationships, and unique characteristics of registered companies.
* Visualize and analyze the dataset to gain insights into the relationships between

variables.

* Identify correlations and patterns that can inform feature selection and engineering.
* Present various data visualizations to gain insights into the dataset.
* Explore correlations between features and the target variable (house prices).
* Discuss any significant findings from the EDA phase that inform feature selection.

**Feature Engineering:**

* **Create new features or transform existing ones to capture valuable information.**
* Creating dummy variables for categorical features such as status, class, category, sub-category, and RoC.
* **Explain the process of creating new features or transforming existing ones.**
* **Showcase domain-specific feature engineering, such as proximity scores or composite**
* **indicators.**
* **Emphasize the impact of engineered features on model performance.**
* Applying dimensionality reduction techniques such as PCA or LDA to reduce the number of features and improve computational efficiency.

**Predictive Modelling:**

* Applying AI algorithms to develop predictive models for future company registrations
* There are various types of predictive models, such as linear regression, logistic regression, decision trees, random forests, gradient boosting, etc. The choice depends on the type and complexity of the data, the objective and the performance criteria.
* Build the model: This involves training the model on the data, tuning the hyperparameters and validating the model using appropriate metrics

**Model Evaluation and Selection:**

* Split the dataset into training and testing sets.
* Evaluate models using appropriate metrices to assess their performance.
* Use cross-validation techniques to tune hyperparameters and ensure model stability.
* Compare the results with traditional linear regression models to highlight improvements.
* Select the best-performing model for further analysis.

**Deployment and Prediction:**

* Deploy the chosen regression model to extract trends.
* Develop a dashboard for users to visualize the extracted trends by AI driven tools

**Program:**

**AI-Driven Exploration and Prediction of Company Registration Trends with Registrar of Companies (RoC)**

import pandas as pd

import matplotlib.pyplot as plt

# Load the data from the CSV file

df = pd.read\_csv('Book1.csv')

# Convert the 'Date' column to datetime format

df['Date'] = pd.to\_datetime(df['Date'])

# Sort the data by date

df = df.sort\_values('Date')

import csv

def print\_csv\_data(filename):

    with open(filename, 'r') as file:

        reader = csv.reader(file)

        for row in reader:

            print(row)

# Call the function with the path to your CSV file

print\_csv\_data('Book1.csv')

# Plot the trend

plt.figure(figsize=(12,6))

plt.plot(df['Date'], df['Value'])

plt.title('Trend Over Time')

plt.xlabel('Year')

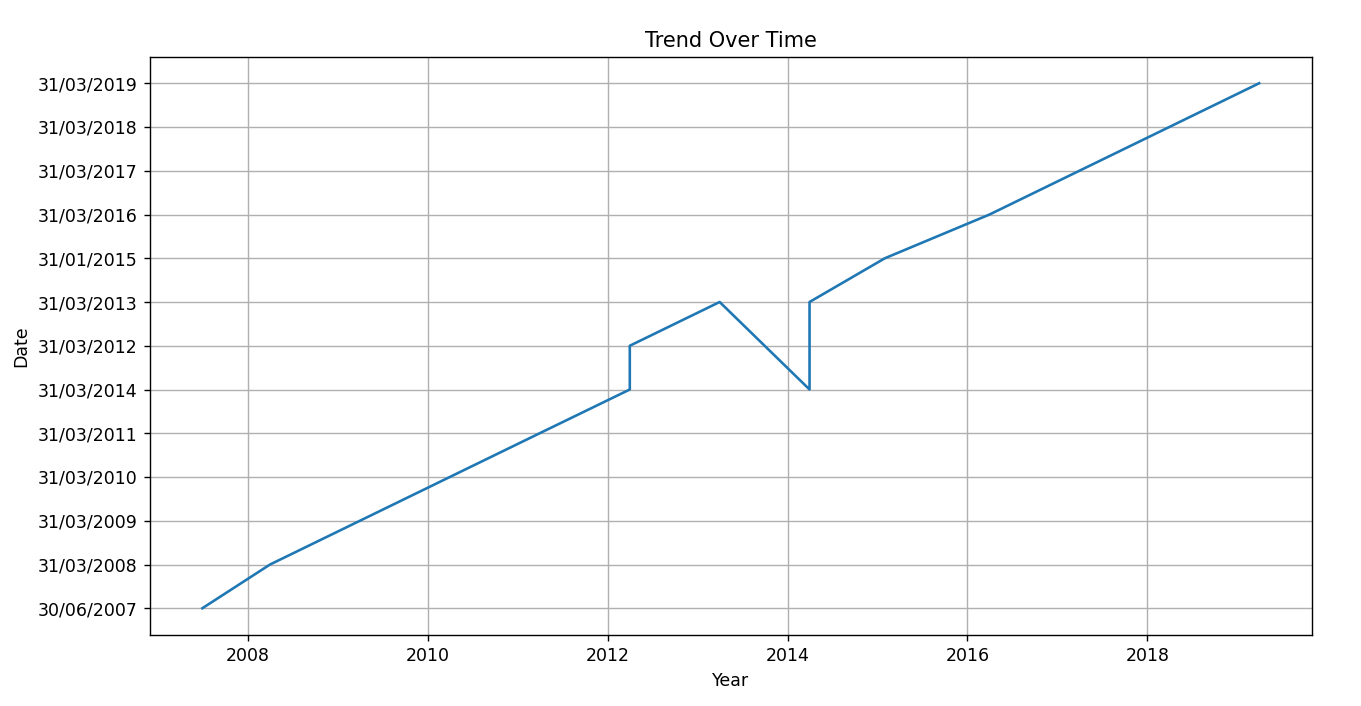
plt.ylabel('Date')

plt.grid(True)

plt.show()



**OUTPUT:**

****

**Conclusion and Future Work (Phase 2):**

**Project Conclusion:**

* In the Phase 2 conclusion, we were summarize the key findings and trends from the

advanced plotting techniques. We were reiterate the impact of these techniques on

improving the accuracy and robustness of extracting trends.