AI DRIVEN EXPLORATION AND PREDICTION OF COMPANY REGISTRATION TRENDS WITH REGISTRAR OF COMPANIES(ROC)

PHASE 4: DEVELOPMENT PART 2

EXPLORATORY DATA ANALYSIS:

PROGRAM:

```
#import necessary libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt

# Load the CSV data
data = pd.read_csv('/content/drive/MyDrive/DataGov_TamilNadu.csv',
encoding = "ISO-8859-1")

# Data Preprocessing and Feature Engineering (customize this part based on
your data)
data['DATE_OF_REGISTRATION'] =
pd.to_datetime(data['DATE_OF_REGISTRATION'])
data['year'] = data['DATE_OF_REGISTRATION'].dt.year

#exploratory data analysis
registrations_by_year = data.groupby('year').size()

# Plot the number of registrations by year
plt.figure(figsize=(10, 6))
plt.plot(registrations_by_year.index, registrations_by_year.values,
marker='o')
plt.xlabel('Year')
plt.ylabel('Number of Registrations')
plt.title('Company Registration Trends')
plt.grid(True)
plt.show()


# Machine Learning Model (Random Forest Regression)
X = data[['year']]
y = data['INDUSTRIAL_CLASS']
```

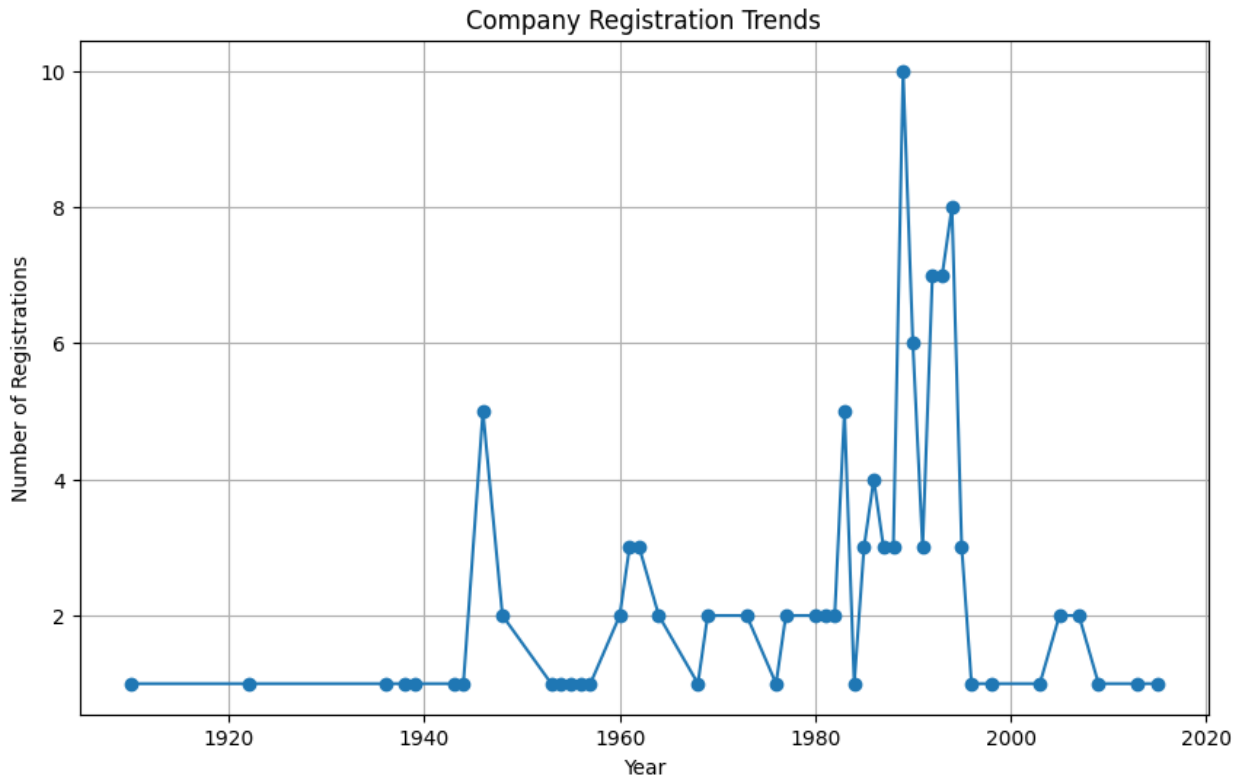
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

# Model Evaluation
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print(f'Mean Squared Error: {mse}')

# Prediction (Predict registrations for future years)
future_years = pd.DataFrame({'year': [2023, 2024, 2025]})
future_registrations = model.predict(future_years)
print(f'Predicted Registrations for 2023: {future_registrations[0]}')
print(f'Predicted Registrations for 2024: {future_registrations[1]}')
print(f'Predicted Registrations for 2025: {future_registrations[2]}')
```

OUTPUT:



```
Mean Squared Error: 38259115.42978392
Predicted Registrations for 2023: 16442.636666666665
Predicted Registrations for 2024: 16442.636666666665
Predicted Registrations for 2025: 16442.636666666665
```