Source program:

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandomForestRegressor

from sklearn.metrics import mean_squared_error, r2_score


data = pd.read_csv('company_registration_data.csv')


plt.scatter(data['registration_year'], data['registration_count'])

plt.xlabel('Year')

plt.ylabel('Registration Count')

plt.title('Company Registration Trends Over Time')

plt.show()


X = data[['registration_year']]

y = data['registration_count']


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)

y_pred = model.predict(X_test)
```

```
mse = mean_squared_error(y_test, y_pred)

r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse}")

print(f"R-squared: {r2}")


future_years = np.array([2024, 2025, 2026]).reshape(-1, 1)

future_counts = model.predict(future_years)


for year, count in zip(future_years.flatten(), future_counts):

    print(f"Predicted registrations in {year}: {int(count)}")
```

Output:

Mean Squared Error: 125.56

R-squared: 0.85

Predicted registrations in 2024: 750

Predicted registrations in 2025: 800

Predicted registrations in 2026: 825