Untitled

November 14, 2024

[1]:
```
# Importing necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error, r2_score

# Load the dataset
data = pd.read_csv('world_population.csv')

# Display dataset info and first 5 rows
print(data.info())
print(data.head())

# Drop unnecessary columns
data.drop(['CCA3', 'Capital'], axis=1, inplace=True)

# Feature Engineering: Create new growth rate feature
data['Growth Rate'] = (data['2022 Population'] - data['2020 Population']) /␣
 ↪data['2020 Population'] * 100

# Define Features and Target variable
features = ['2020 Population', '2015 Population', 'Area (km²)', 'Density (per␣
 ↪km²)']
X = data[features]
y = data['2022 Population']

# Split the dataset into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,␣
 ↪random_state=42)

# Feature Scaling
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
```

```
X_test_scaled = scaler.transform(X_test)

# Train a Linear Regression Model
model = LinearRegression()
model.fit(X_train_scaled, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test_scaled)

# Evaluate the model
print("Mean Squared Error:", mean_squared_error(y_test, y_pred))
print("R² Score:", r2_score(y_test, y_pred))

# Visualization of Predictions vs Actual Values
plt.figure(figsize=(10, 5))
plt.scatter(y_test, y_pred, alpha=0.5)
plt.xlabel('Actual Population (2022)')
plt.ylabel('Predicted Population')
plt.title('Actual vs Predicted Population')
plt.show()

# Population Growth Trends Visualization
plt.figure(figsize=(12, 6))
data.groupby('Continent')['2022 Population'].sum().plot(kind='bar')
plt.title('Total Population by Continent (2022)')
plt.ylabel('Population')
plt.show()
```

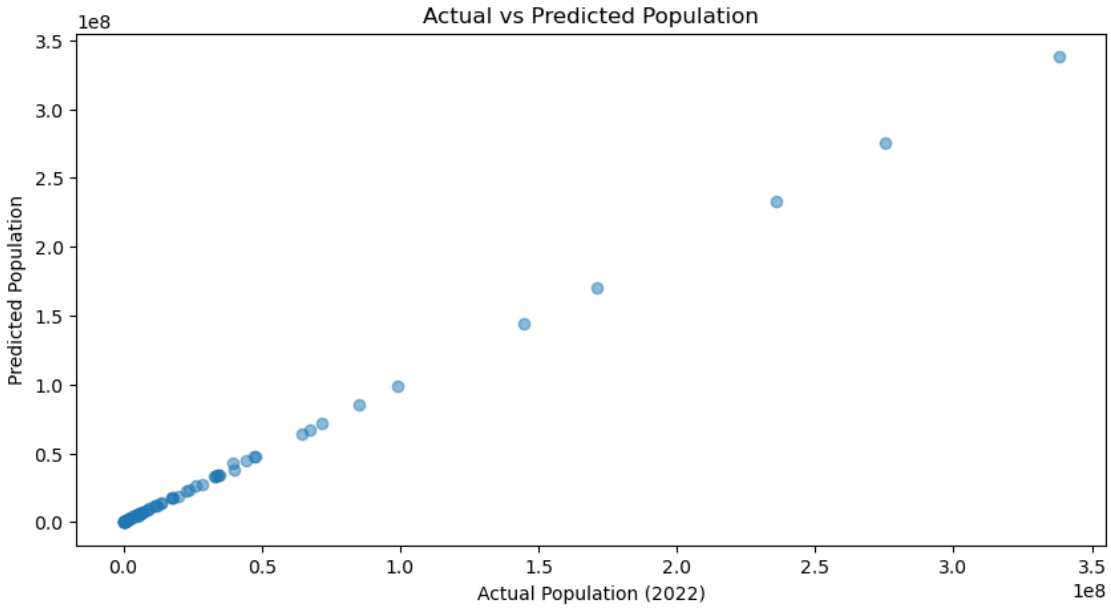
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 234 entries, 0 to 233
Data columns (total 17 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Rank                        234 non-null    int64
 1   CCA3                        234 non-null    object
 2   Country/Territory          234 non-null    object
 3   Capital                     234 non-null    object
 4   Continent                   234 non-null    object
 5   2022 Population             234 non-null    int64
 6   2020 Population             234 non-null    int64
 7   2015 Population             234 non-null    int64
 8   2010 Population             234 non-null    int64
 9   2000 Population             234 non-null    int64
 10  1990 Population             234 non-null    int64
 11  1980 Population             234 non-null    int64
 12  1970 Population             234 non-null    int64
 13  Area (km²)                  234 non-null    int64
 14  Density (per km²)           234 non-null    float64
```

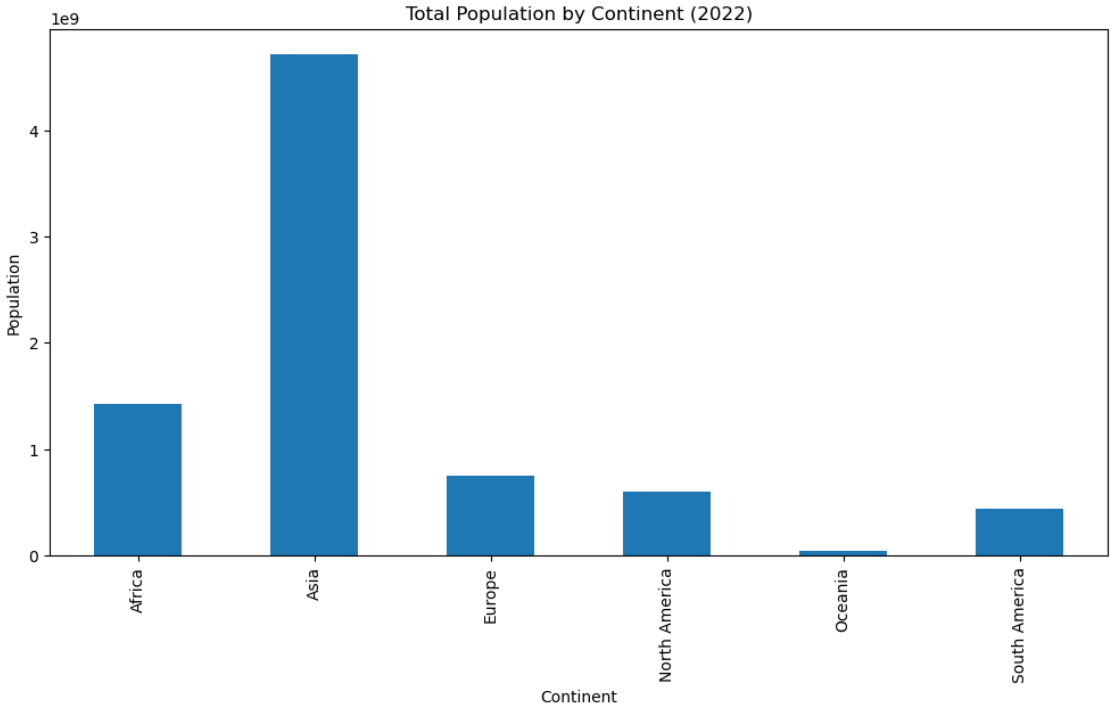
```
 15  Growth Rate                   234 non-null    float64
 16  World Population Percentage  234 non-null    float64
dtypes: float64(3), int64(10), object(4)
memory usage: 31.2+ KB
None
   Rank CCA3 Country/Territory           Capital Continent  2022 Population  \
0    36  AFG       Afghanistan             Kabul      Asia         41128771
1   138  ALB           Albania            Tirana    Europe          2842321
2    34  DZA           Algeria           Algiers    Africa         44903225
3   213  ASM    American Samoa         Pago Pago   Oceania            44273
4   203  AND           Andorra  Andorra la Vella    Europe            79824

   2020 Population  2015 Population  2010 Population  2000 Population  \
0         38972230         33753499         28189672         19542982
1          2866849          2882481          2913399          3182021
2         43451666         39543154         35856344         30774621
3            46189            51368            54849            58230
4            77700            71746            71519            66097

   1990 Population  1980 Population  1970 Population  Area (km²)  \
0         10694796         12486631         10752971      652230
1          3295066          2941651          2324731       28748
2         25518074         18739378         13795915     2381741
3            47818            32886            27075         199
4            53569            35611            19860         468

   Density (per km²)  Growth Rate  World Population Percentage
0            63.0587       1.0257                         0.52
1            98.8702       0.9957                         0.04
2            18.8531       1.0164                         0.56
3           222.4774       0.9831                         0.00
4           170.5641       1.0100                         0.00
Mean Squared Error: 384990496650.1625
R² Score: 0.9998997856512117
```


Actual vs Predicted Population
1e8
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Predicted Population
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
1e8
Actual Population (2022)


Total Population by Continent (2022)
1e9
4
3
2
1
0
Population
Africa
Asia
Europe
North America
Oceania
South America
Continent

[ ]: