Data analytics with Cognos

COVID VACCINE ANALYSIS

Phase 4: Development Part 2

Continue conducting the Covid-19 vaccines analysis by performing:

* Exploratory data analysis
* Statistical analysis
* Visualization

Loading the dataset

import pandas as pd

# Load the dataset

data = pd.read\_csv("/content/covid\_vaccines\_analysis.csv")

* Exploratory data analysis

Hypothesis Testing

Hypothesis:

# Hypothesis Testing (T-test for two independent samples)

from scipy.stats import ttest\_ind

Albania\_a\_vaccination = data[data['country'] == 'Albania']['total\_vaccinations']

Argentina\_b\_vaccination = data[data['country'] == 'Argentina ']['total\_vaccinations']

t\_stat, p\_value = ttest\_ind(Albania\_a\_vaccination,Argentina\_b\_vaccination)

# Output the results

print("T-statistic:", t\_stat)

print("P-value:", p\_value)

# Interpretation: If p-value is less than the chosen alpha level (e.g., 0.05), reject the null hypothesis.

Output

T-statistic: nan

P-value: nan

Correlation Analysis

# Correlation Analysis (Pearson correlation coefficient)

correlation\_coefficient = data['people\_vaccinated\_per\_hundred'].corr(data['daily\_vaccinations\_per\_million'])

# Output the correlation coefficient

print("Correlation Coefficient between people\_vaccinated\_per\_hundred and daily\_vaccinations\_per\_million:", correlation\_coefficient)

# Interpretation: Positive value indicates a positive correlation; negative value indicates a negative correlation.

Output

Correlation Coefficient between people\_vaccinated\_per\_hundred and daily\_vaccinations\_per\_million: 0.23850717638405164

Visualization Code (Output)

Visualization 1: Bar Chart for Vaccination Rates by country

import matplotlib.pyplot as plot

# Bar chart for Vaccination Rates by Region

plt.figure(figsize=(26, 6))

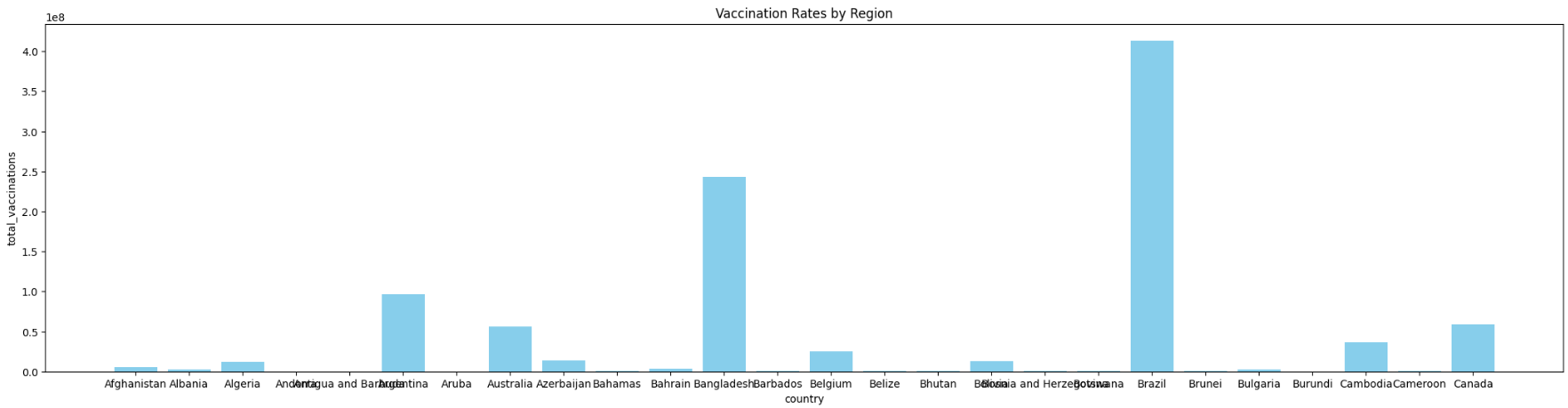
plt.bar(data['country'], data['total\_vaccinations'], color='skyblue')

plt.xlabel('country')

plt.ylabel('total\_vaccinations')

plt.title('Vaccination Rates by country')

plt.show()



Visualization 2: Scatter Plot for country vs. people\_vaccinated\_per\_hundred'

# Scatter plot for country vs. people\_vaccinated\_per\_hundred

plt.figure(figsize=(20, 6))

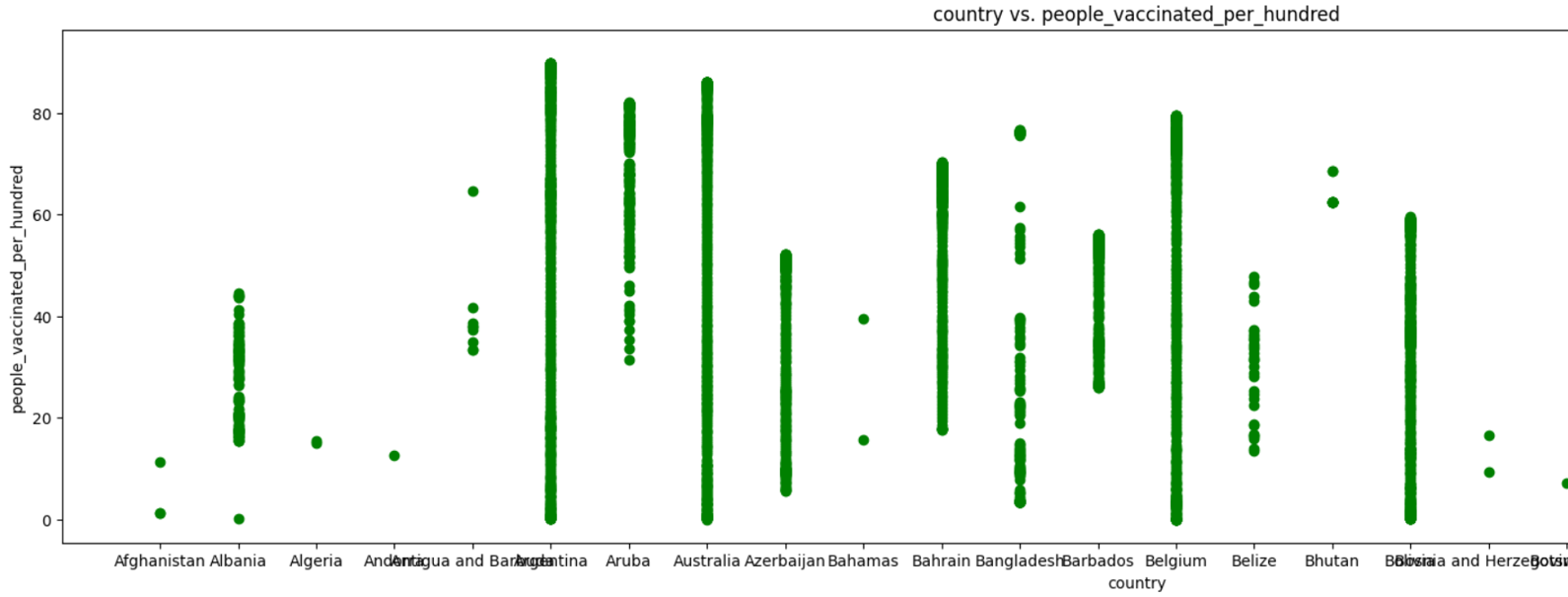
plt.scatter(data['country'], data['people\_vaccinated\_per\_hundred'], color='green')

plt.xlabel('country')

plt.ylabel('people\_vaccinated\_per\_hundred')

plt.title('country vs. people\_vaccinated\_per\_hundred')

plt.show()



* Statistical analysis

Mean and Standard Deviation

Mean Calculation:

# Calculate mean of 'total\_vaccinations' column

mean\_total\_vaccinations = data['total\_vaccinations'].mean()

# Output the mean

print("Mean of Total Vaccinations:", mean\_total\_vaccinations)

output

Mean of Total Vaccinations: 30959612.719471946

Standard Deviation Calculation:

# Calculate standard deviation of 'total\_vaccinations' column

std\_dev\_total\_vaccinations = data['total\_vaccinations'].std()

# Output the standard deviation

print("Standard Deviation of Total Vaccinations:", std\_dev\_total\_vaccinations)

output

Standard Deviation of Total Vaccinations: 68293567.07793456

Linear Regression Analysis

from sklearn.linear\_model import LinearRegression

from sklearn.model\_selection import train\_test\_split

# Selecting features and target variable

X = data[['people\_vaccinated']]

y = data['total\_vaccinations']

# Split the data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Initialize the linear regression model

model = LinearRegression()

# Fit the model with the training data

model.fit(X\_train, y\_train)

# Make predictions on the test data

predictions = model.predict(X\_test)

# Output the coefficients and intercept

print("Coefficients:", model.coef\_)

print("Intercept:", model.intercept\_)

output

Coefficients: [1.86404017]

Intercept: -1013885.9070344828

* Visualization

Visualization with Matplotlib and Seaborn

1. Bar Chart: Vaccination Rates by Country

import matplotlib.pyplot as plt

# Bar chart for Vaccination Rates by Country

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

plt.bar(data['country'], data['total\_vaccinations'], color='skyblue')

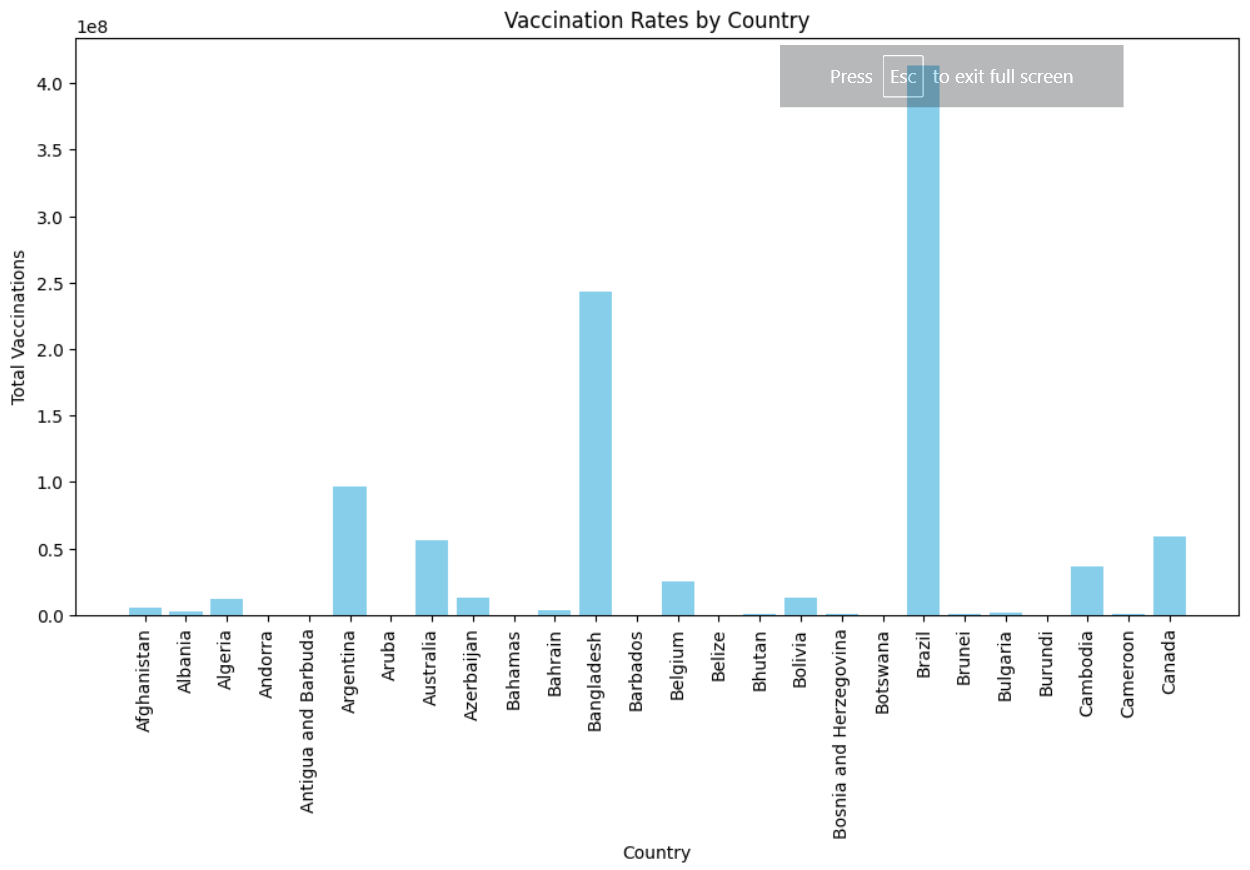
plt.xlabel('Country')

plt.ylabel('Total Vaccinations')

plt.title('Vaccination Rates by Country')

plt.xticks(rotation=90)

plt.show()



1. Scatter Plot: Total Vaccinations vs. People Vaccinated

import seaborn as sns

# Scatter plot for Total Vaccinations vs. People Vaccinated

plt.figure(figsize=(8, 6))

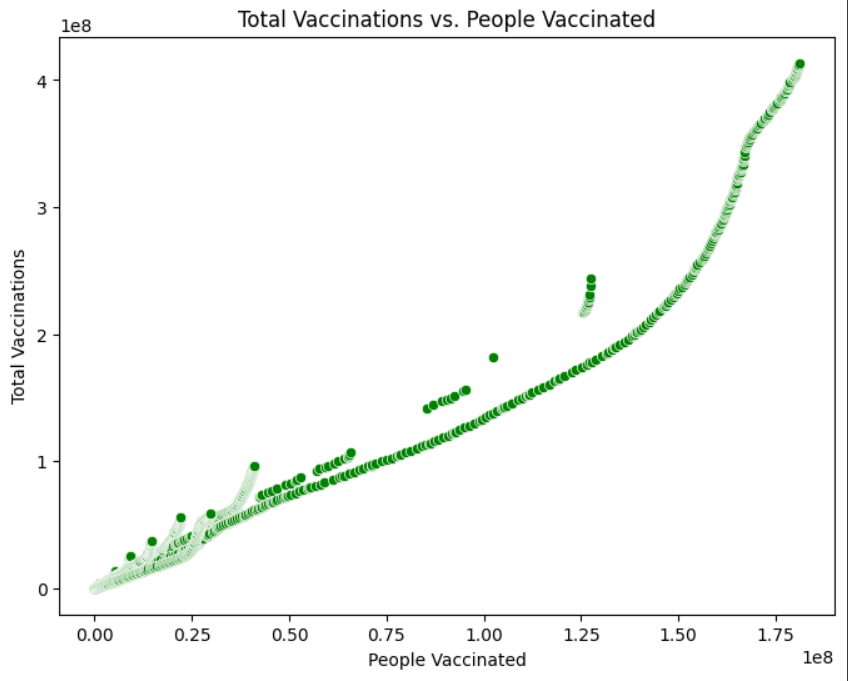
sns.scatterplot(data=data, x='people\_vaccinated', y='total\_vaccinations', color='green')

plt.xlabel('People Vaccinated')

plt.ylabel('Total Vaccinations')

plt.title('Total Vaccinations vs. People Vaccinated')

plt.show()



1. Box Plot: Vaccination Rates by country

import seaborn as sns

# Box plot for Vaccination Rates by country

plt.figure(figsize=(26, 6))

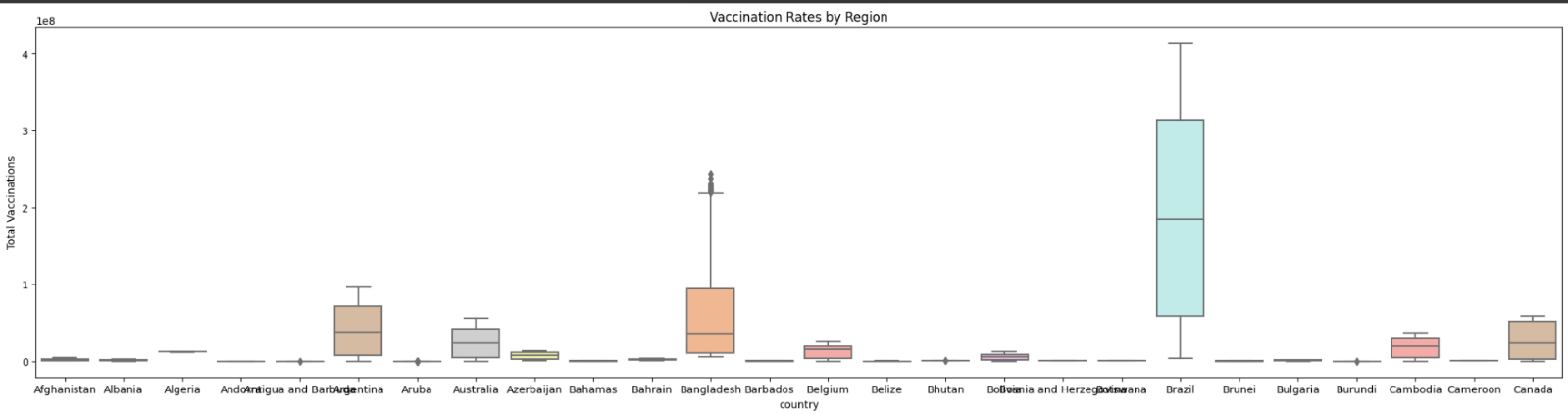
sns.boxplot(data=data, x='country', y='total\_vaccinations', palette='pastel')

plt.xlabel('country')

plt.ylabel('Total Vaccinations')

plt.title('Vaccination Rates by country')

plt.show()



TEAMMATES

K.S.SRINITHI

S.LEENA

C.ELAKKIYA

N.LOGESWARI

P.ANUCIYA