

APPLIED DATA SCIENCE

COVID-19 VACCINE ANALYSIS

PHASE-2

PROJECT DEFINITION:

The project we aim to address is the need for timely and insightful reporting on the COVID-19 pandemic's impact on European union(EU) countries. This project involves analyzing COVID-19 cases and deaths data using DATA SCIENCE. The objective is to compare and contrast the mean values and standard deviations of cases and associated deaths per day and by country in the EU/EEA. This project encompasses defining analysis objectives, collecting COVID-19 data, designing relevant visualizations in DATA SCIENCE, and deriving insights from the data.

Dataset Link: <https://www.kaggle.com/datasets/chakradharmattapalli/covid-19-cases>

DATA PREPARATION:

Start by cleaning and preprocessing the COVID-19 data. This involves handling missing values, checking for data consistency, and ensuring that the data is in a format suitable for analysis.

Example:

- Checking for missing values using `isnull()` and `notnull()`
- Filling for missing values using `fillna()` and `replace()`

(Sample code here)

```
import os
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
os.chdir("C:\Users\ELCOT\Downloads")
```

APPLIED DATA SCIENCE

```
df=pd.read_csv("Covid_19_cases4.csv")  
print(df.isnull())
```

output:

```
0 False  
1 False  
2 False  
.  
.  
.
```

```
2731 False
```

EXPLORATORY DATA ANALYSIS (EDA):

Dive deeper into the dataset to uncover patterns and insights. Use statistical tools, such as histograms, scatter plots, and time-series analysis, to visualize and understand the data.

HYPOTHESIS FORMULATION:

Based on your initial observations from EDA, formulate hypotheses about the factors that may influence COVID-19 cases and deaths in the EU/EEA.

For example, you might hypothesize that certain countries have higher mean values due to population density or healthcare infrastructure.

APPLIED DATA SCIENCE

STATISTICAL ANALYSIS:

Test your hypotheses using appropriate statistical methods. This may involve regression analysis, correlation testing, or other statistical tests relevant to your research questions.

VISUALIZATION ENHANCEMENT:

Refine your visualization strategy in DATA SCIENCE. Create detailed charts, graphs, and dashboards that illustrate the relationships and trends you've identified through statistical analysis.

(Sample model chart for few set of covid-19 cases)

(Sample code here) import os import pandas

as pd import matplotlib.pyplot as plt

os.chdir("C:\Users\ELCOT\Downloads")

df=pd.read_csv("Covid_19_cases4.csv")

cases=df["cases"] death=df["deaths"]

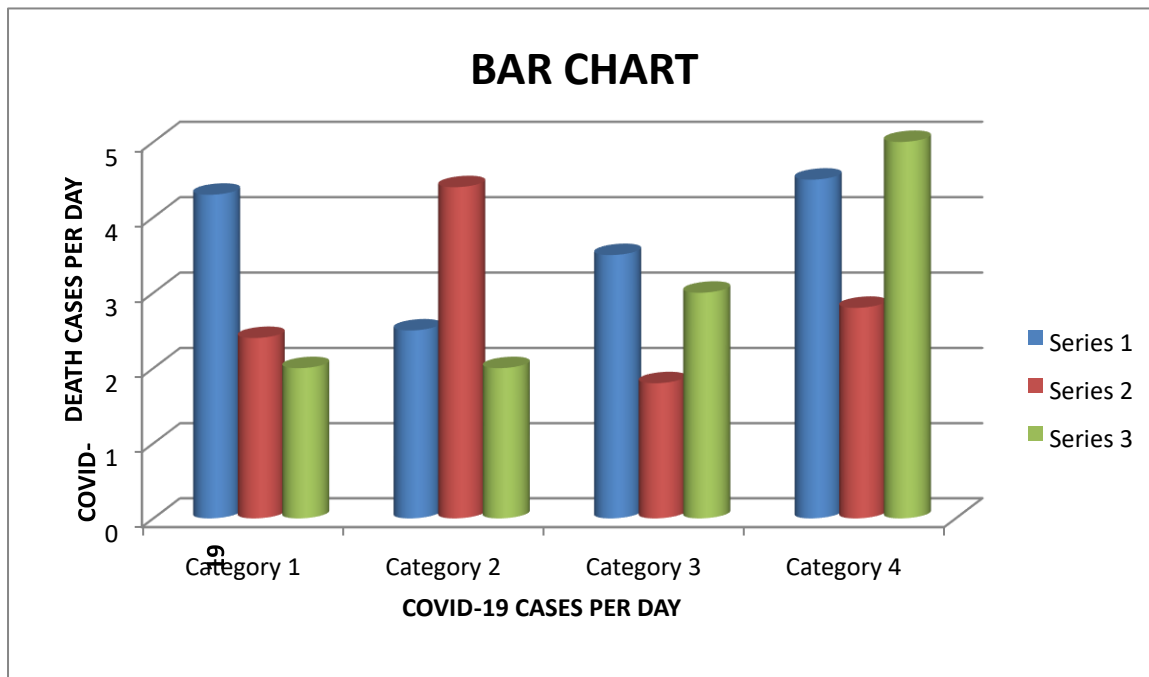
plt.bar(cases,death) plt.xlabel("COVID-19

CASES PER DAY") plt.ylabel("COVID-19

DEATH CASES PER DAY") plt.title("BAR

CHART") plt.show()

APPLIED DATA SCIENCE



RECOMMENDATIONS:

Based on your insights, propose data-driven recommendations or actions. For example, if you find that certain countries have consistently high standard deviations, consider investigating the factors contributing to this variability.

ITERATION:

Keep in mind that data analysis is often an iterative process. If you uncover new questions or areas of interest during this phase, be prepared to revisit earlier stages of the project for further exploration.

APPLIED DATA SCIENCE

TIME-SERIES ANALYSIS:

Analyze how mean values and standard deviations have evolved over time. Are there noticeable trends, seasonality, or sudden changes in the data? Visualize these patterns effectively.

Geospatial Analysis:

If relevant, consider geospatial visualizations to display variations in mean values and standard deviations on a map of the EU/EEA. This can help identify regional disparities.

INSIGHTS GENERATION:

Interpret your analysis results. What insights have you gained from comparing mean values and standard deviations of cases and deaths? Are there countries or time periods that stand out? Are there factors contributing to the observed variations?

ADVANCED VISUALIZATIONS:

Enhance your visualization strategy to include more advanced charts and graphs. For instance, use box plots, heatmaps, or interactive dashboards to provide a comprehensive view of the data.

APPLIED DATA SCIENCE

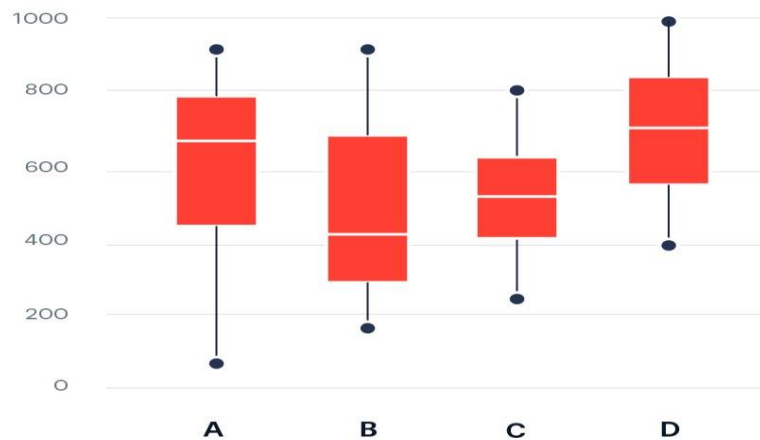
(Sample box plot) (Sample code here)

```
import os
import pandas as pd
import matplotlib.pyplot as plt

os.chdir("C:\Users\ELCOT\Downloads")

df=pd.read_csv("Covid_19_cases4.csv")
cases=df["cases"]
plt.boxplot(cases)

plt.show()
```



APPLIED DATA SCIENCE

DOCUMENTATION:

Document your entire analysis process, including data sources, methods, and assumptions. This documentation is crucial for transparency and reproducibility.

COMMUNICATION:

Communicate your findings and recommendations effectively to stakeholders or your team. Use clear visualizations and concise explanations to convey your insights.