

Covid-19 cases analysis using cognos

INTRODUCTION :

Covid-19 is an infectious illness caused by a newly identified form of coronavirus. This is a new virus and illness that was previously unknown before the December 2019 outbreak in Wuhan, China. The number of confirmed cases of Covid-19 and the number of deaths due to this virus in World are increasing and quite alarming. Therefore this study will discuss the grouping of Cases and Deaths of COVID-19 in EU/EEA Countries. In this part you will begin building your project by loading and preprocessing the dataset. Building the COVID-19 cases analysis using IBM Cognos for visualization.

DATASET LINK :


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

PREPROCESSING :

Preprocessing of data is a crucial step in data analysis and machine learning. Here are the common steps involved:

1.DATA CLEANING:

This involves handling missing values, dealing with duplicates, and correcting errors.

A screenshot of a Jupyter Notebook interface. The top bar shows two tabs: 'accuracy.ipynb' and 'Untitled2.ipynb'. Below the tabs is a toolbar with icons for saving, adding, deleting, and running code. The main area displays a code cell with the following Python code:

```
[3]: miss_mean_imputer = Imputer(missing_values='NaN', strategy='mean', axis=0)
      miss_mean_imputer = miss_mean_imputer.fit(df)
      imputed_df = miss_mean_imputer.transform(df.values)
      print(imputed_df)
```

2.DATA INTEGRATION:

If data comes from different sources, integration ensures consistency in format and resolving any inconsistencies.

```
df.drop([dateRep] , axis=0)
```

3.DATA TRANSFORMATION:

Covid-19 cases analysis using cognos

This step includes normalization, variables to make data suitable for analysis.

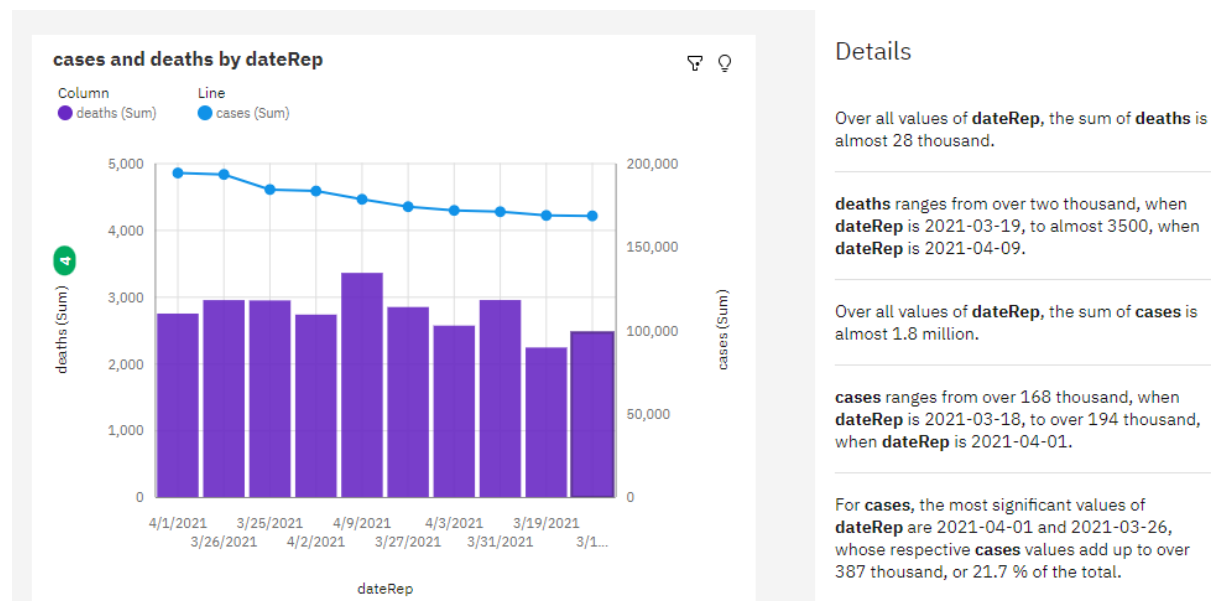
4.DATA REDUCTION:

Reducing the dimensionality of the data through techniques like feature selection or efficiency of the analysis.

5.Data Discretization:

It continuous variables can be transformed into discrete ones for ease of analysis.

VISUALIZATION USING IBM COGNOS:



TRENDS OVER TIME:

Analyzing the trend of cases and deaths over time can help you understand the progression of the pandemic. Are cases and deaths increasing, decreasing, or remaining stable?

REGIONAL VARIATIONS:

Analyzing data by regions or countries can reveal variations in the impact of the pandemic. Some areas may have been more heavily affected than others.

Covid-19 cases analysis using cognos

VACCINATION IMPACT:

You can analyze data before and after vaccination campaigns to see how they impact the number of cases and deaths.

CASE-FATALITY RATE (CFR):

Calculating the CFR, which is the ratio of deaths to cases, can help assess the severity of the virus in a given area or population.

ACCURACY :

The Jupyter Notebook is an open-source web application that allows you to create and share documents that live code. In Jupyter we use python programming language to find the accuracy of the data with the help of total number of samples in the dataset.

```
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression
nb_samples=2731
x,y=make_classification(n_samples=nb_samples,n_features=2,n_informative=2,n_redundant=0,n_clusters_per_class=1)
xtrain,xtest,ytrain,ytest=train_test_split(x,y,test_size=0.2,random_state=42)
model=LogisticRegression()
model.fit(xtrain,ytrain)
print(accuracy_score(ytest,model.predict(xtest)))

0.9597806215722121
```

Here , the accuracy of the given dataset is 0.9597806215722121. The best possible value is 1 (if a model got all the predictions right), and the worst is 0 (if a model did not make a single correct prediction). From our experience, you should consider Accuracy > 0.9 as an excellent score, Accuracy > 0.7 as a good one, and any other score as the poor one. So, this COVID-19 dataset has excellent score in accuracy.

CONCLUSION :

Thus, we conclude that the dataset has been processed by the preprocessing including the missing values which is handled by calculate the mean values and removing of irrelevant column (datarep). Visualization of dataset is done by using IBM Cognos, COVID-19 cases and deaths of 30 EU/EEA countries with the help of the graph. The accuracy can be done by the Jupiter notebook with the python programming language (py.code). So , the accuracy level of COVID-19 is 0.959780621572.