Data Analysis - EDA - Covid_19

Abstract:
The goal of this project was to do EDA(Exploratory Data Analysis) to determine the effects of the corona virus in the world. The was acquired from OurWorldInData.org which is updated weekly. The data included various features of which a few are picked to carry out the analysis and answer the questions. The data included many countries but only the two extreme groups were picked: the top(x) and bottom(x), because the rest would've followed the generalization making the visualization too compact for the questions and analysis previously decided. No model is required for this project.

Design:
The project shows a few of the many questions that can be answered using the same data. The data shows which countries took the better decisions when dealing with the pandemic and how geographical location can be of benefit. Furthermore it also shows that even though some countries are wealthy they took the poorest decisions.

Data:
The dataset contains roughly 140k data points with 65 features/columns 5 is categorical and the rest is numerical(float). One of the categoricals was turned to a numeric type(Monday=1,Sunday=2..etc) for ease of use and to perform SQL queries more efficiently. Originally most of the data was numbers in string type and was converted to float.

Data cleaning - Major:
The data once converted to SQL had issues of types, nulls and more. The most troublesome was the types of the database columns. To solve that issue the notebook 'Fixing database' was built to deal with that issue by getting the data to a dataframe from the database, then fixing the issue with the columns of the sql, while doing that fixing the Null data presented itself, which was a lot, as it also was proving to be causing issues with the SQL queries so the nulls were replaced with a zero which made sense since the columns types were changed float.

Algorithms:
None were required

Tools:
Numpy,SQL and Pandas for data manipulation | Pyodbc for connecting to database server | SQLalchemy for data queries | Matplotlib and Seaborn for plotting | SQL Server Microsoft

Communication:

Data Analysis - EDA - Covid_19

In addition to the slides and visuals presented, the code is on github bushnag-ai and data is https://ourworldindata.org/covid-deaths.