**CSCE 5320 SCIENTIFIC DATA VISUALIZATION**

**DATASETS FOR PROJECT**

**Students must choose only the below given10 datasets for your project:**

**You can choose any one of them:**

1. [**https://www.kaggle.com/datasets/rajkumarpandey02/2023-world-population-by-country**](https://www.kaggle.com/datasets/rajkumarpandey02/2023-world-population-by-country)
2. [**https://www.kaggle.com/datasets/georgesaavedra/covid19-dataset**](https://www.kaggle.com/datasets/georgesaavedra/covid19-dataset)
3. [**https://www.kaggle.com/datasets/sazidthe1/data-science-salaries**](https://www.kaggle.com/datasets/sazidthe1/data-science-salaries)
4. [**https://www.kaggle.com/datasets/nextmillionaire/car-accident-dataset**](https://www.kaggle.com/datasets/nextmillionaire/car-accident-dataset)
5. [**https://www.kaggle.com/datasets/vineetkukreti/indian-agriculture-dataset**](https://www.kaggle.com/datasets/vineetkukreti/indian-agriculture-dataset)
6. [**https://www.kaggle.com/datasets/nayansubedi1/airplane-crashes-and-fatalities-upto-2023**](https://www.kaggle.com/datasets/nayansubedi1/airplane-crashes-and-fatalities-upto-2023)
7. [**https://www.kaggle.com/datasets/thedevastator/unlock-profits-with-e-commerce-sales-data**](https://www.kaggle.com/datasets/thedevastator/unlock-profits-with-e-commerce-sales-data)
8. [**https://www.kaggle.com/datasets/nelgiriyewithana/countries-of-the-world-2023**](https://www.kaggle.com/datasets/nelgiriyewithana/countries-of-the-world-2023)
9. [**https://www.kaggle.com/datasets/nelgiriyewithana/australian-vehicle-prices**](https://www.kaggle.com/datasets/nelgiriyewithana/australian-vehicle-prices)
10. [**https://www.kaggle.com/datasets/neharana404/maternal-indicators-in-us-states2016-2021**](https://www.kaggle.com/datasets/neharana404/maternal-indicators-in-us-states2016-2021)

**Student must not use a single data instead you should use multiple data from different areas such as health, finance, education data and show correlation in the data by analyzing the data with the help of visualization tools (d3.js, python/tableau/Power BI).**

**In your project you should use at least 3 tools and it is mandatory to use D3.Js.**

**Example:**

Let’s examine two datasets to derive insights on a specific topic. For example, we can analyze a dataset providing COVID-19 information which is related to health alongside another dataset focusing on the world economy, which is related to Economy. Our goal is to illustrate how COVID-19 has affected the global economy and present relevant statistics.

**Examples of Data Visualization Projects with multiple datasets:**

**A screenshot of a computer

Description automatically generated**

In above screenshot, you can see they have combined both data sources all\_sales.xlsx and Superstore.xlsx. So, both the data sources available under the Dashboard. By using these two data sources we can generate visualizations and relations between two datasets. The attributes and metrics of both sources are available under their respective names.

**A screenshot of a computer

Description automatically generated**

Next, drag the attribute “Business Line” from All\_sales.xlsx to the rows box. Drag the attributes “customer segment” and “Product Category” from the second data set to the rows box.

The grid visualization appears showing data from both the data sets.

In above grid visualization we can see the different types of business lines are related to different types of customer segment and Product category.

Examples of Visualizations:

A graph of a number of people

Description automatically generated

In above screenshot, we can see the population of different states which is represented by the horizontal bar chart.

We can create more visualizations by adding some information like literacy rate, economic rate of the different countries.

So, here we can combine the population with different areas like education, economy etc.

**Interactive Visualizations:**

**A screenshot of a computer screen

Description automatically generated**

In above visualization, we can seethe interactive, colorful map details exactly where the United States electricity comes from and how much energy is generated. Each circle represents an individual power source, which are color-coded by type, with the source key demonstrating total capacity nationwide. The circle size also indicates the power output generated by the source.