1   Tools and Technologies

This visualization project was built using python and the following libraries:

1. Dash: which was used for creating the interactive dashboard on the web.
2. Plotly: It was used to generate interactive visualization such as bar graphs choropleths and bubble charts.
3. Pandas: Used of manipulation of csv data sets.
4. Dash Bootstrap: It was used for styling of the web application.

2   Data Preprocessing Steps:

- Normalized the country name across the datasets for example replace "USA" with "United States" to ensure consistent merging.

- Converted data columns into datetime objects and also aggregated the covid-19 cases monthly for better viewing.

- derived additional attributes such as the employment_impact which was based on the percentage of working hours lost.

- Merged multiple datasets to get the total cases population total deaths employment metrics and monthly cases data based on country.

3   Dataset Description

- General Description of the dataset:
  - The dataset contains a combination of covid-19 data from OWID dataset and worldometer data which consists of the entire world's data dn employment metrics. It includes information on global cases, deaths populations and the economic impact.
- Detailed Analysis:
  - Attribute types: It includes numeric data like total cases, population and categorical data like the country/region, continent and datetime fields like the month_start_date.
  - Derived Attributes:
    - Employment_impact: calculated as the product of **percentage_of_working_hrs_lost** and population.
    - **population_scaled:** population scaled for visual clarity.
- Complexity:
  - The dataset has a combination of multi-source data with varying scales and requires visual aggregation to have meaningful insights. It's temporal nature and multi-dimensional attribute with interactive visualizations to convey relationships.

4   Tasks Supported by Visualization

- **Task 1:** Analyzing the monthly rise and fall of covid-19 cases
- **Task 2:** Identifying the correlations between the totalCases Vs TotalDeaths using scatter plots.
- **Task 3:** Assessing the geographic spread of COVID-19 and its relation to population and employment metrics
- **Task4:** Explore country-level employment impacts.

5   Encoding Channels and Idioms

- **Encoding Channels:**
  - **Position:** x-axis for population and y-axis for total cases or employment_impact.
  - **Color:** Encodes categorical attributes like Country/Region and continent.
  - **Size:** Bubbles in the geo chart represent the population_scaled.
  - **Animation Frame:** Temporal progression visualized through month_start_date.
- **Idioms:**
  - **Bar Charts:** Highlight the top 10 countries based on deaths in each frame.
  - **Bubble Charts:** Shows the trends in employment impact and population.
  - **Scatter Geo Maps:** shows the distribution of cases.
  - **choropleths:** shows the global employment impact.
- **Interactive Operations:**
  - Dropdown menus for selecting countries or metrics.
  - Hover interactions to reveal detailed data points.
  - Animation for observing monthly trends.

Justification: These encoding helps users to explore relationships between metrics, identify trends and gain insights into the pandemic's impact.

6   Novelty and Complexity

- **Novelty:**
  - The dashboard integrates COVID-19 data and employment metrics which will provide a comprehensive view of the economic activities.
  - It uses dual Y-axes for visualizing totalCases alongside employment_impact provides a unique perspective.
- **Implementation Complexity:**
  - Preprocessing involved merging different datasets with varying formats and scales
  - The dynamic visualization with temporal animation which require precise data structuring and plolty/Dash expertise.

7   Critical Analysis:

- **Strengths:**
  - Highly interactive and visually engaging.
  - Combines geographic, temporal and quantitative data in a cohesive manner.
  - Enables multi-dimensional analysis.

- **Weaknesses:**
  - Dual-axis bubble charts may confuse users.
  - The density of information in geo map can overwhelm users.
  - and derived metrics like employment_impact rely on assumptions and may not represent ground reality.

References

[1] https://www.kaggle.com/datasets/vineethakkinapalli/impact-of-covid19-on-employment-ilostat
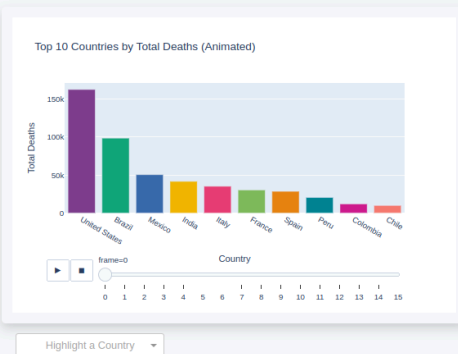
[2] https://plotly.com/

[3] https://www.kaggle.com/datasets/sandhyakrishnan02/latest-covid-19-dataset-worldwide

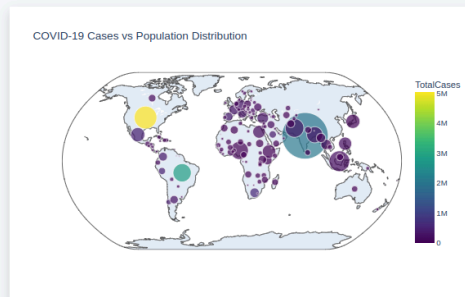Video Google Drive Link: 📹 video1657927139.mp4

Github Repo link for code: https://github.com/itzVort3x1/dv_assignemnt-3

NOTE: I have uploaded the video to github as well as provided the link for g-drive.