Story 1: Exploring the Trends: Visualizing Titanic Passenger Data
Leveraging the Titanic dataset, you will analyze the data and create visualizations that align with a specific story line.

1. Download the dataset provided (Titanic Dataset).
2. Understand the story line and identify key data points for your analysis.
3. Generate visualizations as requested in each task.

The Titanic disaster is one of the most infamous shipwrecks in history. Analyzing data from the Titanic provides insights into survival patterns and socio-economic factors. In this assignment, you will use the Titanic dataset to uncover these patterns and gain an understanding of the factors affecting passenger survival rates.

The dataset includes the following columns:

- PassengerId: Unique identifier for each passenger
- Survived: Survival status (0 = No, 1 = Yes)
- Pclass: Ticket class (1st, 2nd, 3rd)
- Sex: Gender of the passenger
- Age: Age of the passenger
- SibSp: Number of siblings/spouses aboard
- Parch: Number of parents/children aboard
- Ticket: Ticket number
- Fare: Fare paid by the passenger
- Cabin: Cabin number
- Embarked: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)

Task 1: Survival Rate by Passenger Class

- Plot a bar chart to show survival rates for each passenger class. (A bar chart is ideal for comparing categorical data, helping us understand the relationship between class and survival)
- Create a pie chart to visualize the proportion of passengers in each class. (A pie chart effectively shows the distribution of passengers across classes)
- Is there any other visualization for these tasks. Please add the same and add one line why you choose this.

Task 2: Gender and Survival Analysis

- Generate a stacked bar chart comparing survival rates by gender within each class. (Stacked bar charts allow for the simultaneous visualization of two categorical variables, providing clarity on gender-based survival trends across classes)

- Create a heatmap to analyze the correlation between gender, age, and survival. (A heatmap is suitable for identifying patterns and correlations

in numeric and categorical data)
- Is there any other visualization for these tasks. Please add the same and add one line why you choose this.

Task 3: Port of Embarkation Analysis

- Visualize the number of survivors from each embarkation port using a horizontal bar chart. (Horizontal bar charts provide a clear comparison when categories (ports) have long labels)
- Use a clustered bar chart to compare survival rates across ports for different classes. (Clustered bar charts are useful for comparing multiple groups across categories)
- Is there any other visualization for these tasks. Please add the same and add one line why you choose this.

Task 4: Age Distribution Analysis

- Create a histogram to show the age distribution of passengers. (Histograms are ideal for understanding the frequency distribution of a continuous variable like age)

- Add a line chart over-laying the survival rates for each age group. (A line chart allows for tracking survival trends across age groups, making it easier to interpret patterns)

- Is there any other visualization for these tasks. Please add the same and add one line why you choose this.

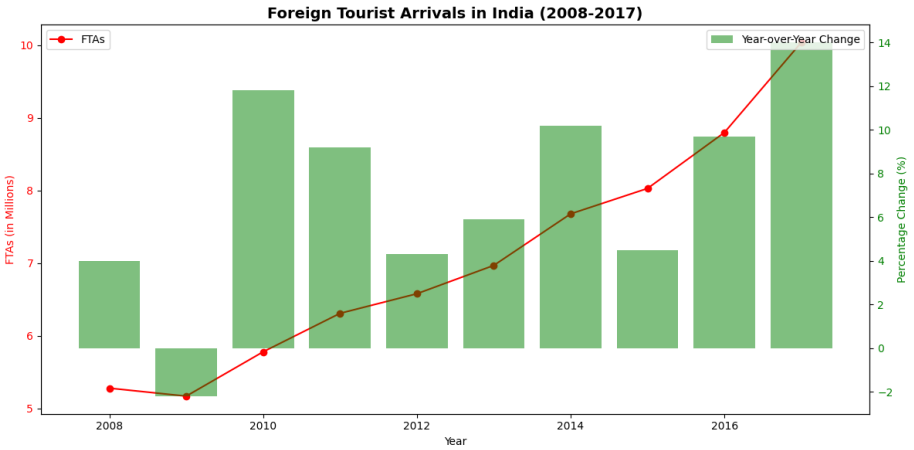
Task 5: Fare and Survival Patterns

- Create a box plot to compare the distribution of fares between survivors and non-survivors (Box plots are ideal for comparing distributions and identifying outliers between two groups)
- Is there any other visualization for these tasks. Please add the same and add one line why you choose this.

Story 2: A Decade of Growth: Tracking Foreign Tourist Arrivals in India India's diverse culture, historical landmarks, and natural beauty have always been a magnet for international tourists. Over the years, foreign tourist arrivals (FTAs) have been an important indicator of India's growing prominence as a global travel destination. But what do the numbers tell us about this journey? Let's explore how FTAs in India have evolved over a decade and uncover trends behind the growth story.

1. Download Data from Data.Gov.in

Task 01: Growth of Foreign Tourist Arrival in India. The visualization that combines the Foreign Tourist Arrivals (FTAs) in India over the years with the percentage change year-over-year. The line chart represents FTAs in millions, while the bar chart indicates the growth percentage for each year. This is a 3-dimensional chart. y-axis is labeled differently for the line and bar chart.



- Generate the similar graph for the data available from 2000 until 2017.
- Is there any other diagram you suggest? Please add the same

Task 02: From where they are coming

- Download appropriate data to visualize country-wise arrival of FTAs (Use lollipop chart)
- Download appropriate data to visualize region-wise arrival of FTAs (Use Sankey diagram)
- Is there any other diagram you suggest? Please add the same.

Task 03: Compositions of FTAs

- Download appropriate data to visualize age-wise arrival of FTAs (Use DOUGHNUT chart)
- Download appropriate data to visualize the distribution of mode of travel and purpose of travel of FTAs in a single diagram year-wise (Use Violin Plot)
- Download appropriate data to visualize year-wise age and gender distributions of the FTAs. (Use multiple box plot or violin plot)
- Download appropriate data to visualize nationality wise age distribution with Treemap diagram.
- Is there any other diagram you suggest? Please add the same.

Story 03: Analysis of COVID-19 India India faced significant challenges during the COVID-19 pandemic, and Maharashtra was one of the hardest-

hit states. Tracking various metrics during the pandemic allows for a better understanding of infection trends, recovery efforts, and the effectiveness of public health measures. In this visualization assignment, we explore different aspects of the pandemic through interactive and informative visualizations.

You can use the COVID-19 Maharashtra, India dataset from IBM COVID-19 India Data.

The following are the dataset variables information:

- date – Date of record.
- cases_new – Daily new COVID-19 cases.
- discharged_today – Daily recovered patients.
- recovery_rate – Percentage of total cases recovered.
- deaths_new – Daily reported deaths.
- active_cases – Current cases under treatment.
- discharged_total – Total recovered patients.
- cfr – Case Fatality Rate (is the proportion of people who have been diagnosed with a disease and end up dying of it.).
- tests_cumulative – Total tests conducted.
- tests_positive_cumulative – Total positive cases from tests.
- current_home_quarantine – Cases under home quarantine.
- current_institutional_quarantine – Cases in institutional quarantine.

Let's begin the analysis by posing questions to the data. Along with each question, we have outlined the approach, specified the relevant variables, and explained the necessity of the chosen visualization.

Task 01 How did Maharashtra's daily recoveries compare to the number of new infections during the peak and decline of the pandemic?

- Compare trends using a dual-color line plot, with red for new cases and green for recoveries.
- Is there any other diagram you suggest? Please add the same.

Task 02 What proportion of patients were quarantined at home versus institutional facilities over time?

- Visualize the share of home and institutional quarantines as stacked proportions.
- Is there any other diagram you suggest? Please add the same.

Task 03 How do new deaths relate to new infections over time?

- Correlate new cases, deaths, recoveries, and active cases using a heatmap.
- Is there any other diagram you suggest? Please add the same.

Task 04 How do daily new cases correlate with deaths and recoveries during the pandemic?

- Use a bubble plot with recoveries represented by bubble size and color.
- Is there any other diagram you suggest? Please add the same.

Task 05 What is the density correlation between daily new cases and deaths during the pandemic?

- Visualize the density of new cases versus deaths using a kernel density estimate (2D Density Plot)
- Is there any other diagram you suggest? Please add the same.

Story 04: IPL Tracker Imagine you are a data analyst working with the IPL dataset covering the seasons from 2008 to 2022. Your task is to create a comprehensive data visualization story that provides insights for various aspects.

Download the dataset provided (IPL Dataset)

The IPL_Matches_2008_2022.csv dataset includes the following columns:

1. ID - Unique identifier for each match.
2. City - The city where the match was played.
3. Date - The date on which the match took place.
4. Season - The IPL season year.
5. MatchNumber - The designation of the match (e.g., Final, Qualifier 1, etc.).
6. Team1 - The name of the first team.
7. Team2 - The name of the second team.
8. Venue - The stadium where the match was held.
9. TossWinner - The team that won the toss.
10. TossDecision - The decision made by the toss-winning team (bat or field).
11. SuperOver - Indicates whether the match went into a Super Over (Yes/No).
12. WinningTeam - The team that won the match.
13. WonBy - The method of victory (Runs/Wickets).
14. Margin - The margin of victory (in terms of runs or wickets).
15. Method - If applicable, the method used to determine the winner (e.g., DLS method).
16. Player_of_Match - The player awarded as the Player of the Match.

17. Team1Players - List of players from Team 1.
18. Team2Players - List of players from Team 2.
19. Umpire1 - The name of the first umpire officiating the match.
20. Umpire2 - The name of the second umpire officiating the match.**

Task 01: Match Outcomes by Team

- Visualize the number of wins for each team across different seasons. Highlight the teams with the most consistent performance and those with significant fluctuations in their win rates. Choose the best visualization you think appropriate from the catalog provided in Data-to-Viz site.

Task 02: Venue Analysis:

- Develop a heatmap showing the distribution of matches across various venues. Identify which venues hosted the most matches and analyze any patterns in home-ground advantages.
- Is there any other diagram you suggest? Please add the same.

Task 03: Toss Decisions Impact:

- Analyze the impact of toss decisions (batting or fielding first) on match outcomes. Use a stacked bar chart to compare the success rates of teams that chose to bat first versus field first.
- Is there any other diagram you suggest? Please add the same.

Task 04: Player Performance:

- Highlight the top-performing players across seasons using a radar chart. Focus on players who won the "Player of the Match" award frequently
- Is there any other diagram you suggest? Please add the same.

Task 05: Head-to-Head Team Analysis:

- Develop a visualization comparing the head-to-head win/loss records of major rival teams. Use a grouped bar chart to showcase the competitive dynamics between these teams over the years.
- Is there any other diagram you suggest? Please add the same.