**Dataset Used :** Titanic Dataset.csv

Imagine stepping back in time to the tragic voyage of the Titanic. As we delve into the dataset, we uncover a wealth of information that paints a vivid picture of the passengers onboard. Let's embark on a journey filled with intriguing insights and compelling stories that capture the essence of the Titanic dataset. *(Storytelling continued from* ***Insights Gained*** *section along with visualization diagrams).*

Let us consider the Titanic Dataset sourced from Kaggle website and explore the data by visualizing a few key analyses using univariate, bivariate and multivariate analysis. To make it easier to understand, I discussed each visualization definition and story behind visualizing the solution.

**First 10 rows of dataset**



**Data Dictionary**

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Definition** | **Key sets** |
| Passenger Id | Represents unique id of passenger |  |
| Survival | People Survived | 0=No, 1= Yes |
| Pclass | Ticket class | 1 = 1st, 2= 2nd, 3= 3rd |
| Sex | Gender of passenger | Male, female |
| Age | Age of Passenger |  |
| Sibsp | Represents number of siblings/ spouses |  |
| Parch | Represents parents & children count |  |
| Ticket | Ticket Number of passenger |  |
| Fare | Passenger fare |  |
| Cabin | Cabin number of ship |  |
| Embarked | Port to be Embarked by passenger | C= Cherbourg, Q= Queenstown, S= Southampton |

**Design Decisions:**

**Libraries:** I imported necessary libraries such as NumPy, pandas, matplotlib, seaborn, and plotly express to perform data analysis and visualization.

**Data Source:** I assumed that the Titanic dataset is stored in a CSV file named 'Titanic Dataset.csv'. You can replace this with the actual file name and path to match your dataset location.

**Visualization Types:**

* For ***univariate analysis,*** I used various visualization techniques such as pie charts, bar plots, histograms, and scatter plots to explore different aspects of the Titanic dataset.
* For ***Bivariate Analysis,*** I used scatter plots, kernel density estimation (KDE) plots, and count plots to visualize the relationships between variables.
* For ***multivariate analysis,*** I used Tree maps, Scatter plots, 3D Visualization and bubble chart to visualize the relationship between three or more than 3 variables.

**Insights Gained:**

As we set sail, we are immediately drawn to the question of survival. The pie chart depicting the count and percentage of survivors reveals a heartbreaking reality. Only a fraction of the passengers managed to survive the fateful night, highlighting the immense tragedy that unfolded. We begin to empathize with the individuals who faced unimaginable circumstances.

**Count and Percentage of Survived:** So, the bar chart shows the distribution of surviving and deceased passengers, providing a visual representation of the survival rate.

A picture containing text, screenshot, number, diagram

Description automatically generated

When we look at the passenger classes, we see a very interesting distribution. The pie chart showing the number and percentage of passengers in each class reveals a big difference. Most of the passengers were in third class, which represents the working class who were looking for a better future. On the other hand, the wealth of the first-class passengers is clear, as they make up a significant portion of the data. This difference in class creates the foundation for a complex story of social divisions during chaos.

**Count and Percentage of Pclass:** So, the pie chart displays the distribution of passengers across different passenger classes (1st, 2nd, and 3rd). It gives an overview of the class composition within the dataset.

A picture containing text, screenshot, diagram, font

Description automatically generated

As we dig deeper, the age variable comes into focus. The summary statistics of age provide us with a glimpse into the passengers' demographic makeup. From the youngest infants to the elderly, the age range encompasses a spectrum of life stages. We begin to envision the hopes, dreams, and aspirations of these individuals, all tragically extinguished on that ill-fated night.

**Summary Statistics of Age:** So, the summary statistics provide key information about the age variable, including count, mean, standard deviation, minimum, maximum, and quartile values.

A picture containing diagram, plot, line, screenshot

Description automatically generated

Gender, an essential aspect of our exploration, takes center stage. The count and percentage of male and female passengers are displayed through a captivating pie chart. We discover that men dominate the dataset, their presence significantly outnumbering women. Yet, in the face of adversity, gender plays a pivotal role in shaping survival outcomes.

**Count and Percentage of Sex:** So, the pie chart illustrates the distribution of male and female passengers, providing insights into the gender composition of the dataset.

A blue and orange pie chart

Description automatically generated

Our attention is captured by the scatter plot depicting the count of male and female passengers based on the port of embarkation. We observe intriguing patterns emerging from this visualization. The distribution of male and female passengers across different embarkation points reveals a story of diverse origins and destinies. We can't help but wonder how these factors influenced their experiences onboard.

**Scatter Plot of Fare:** So, this scatter plot displays the relationship between the passenger ID and fare. It allows us to identify any patterns or outliers in the fare data.

A picture containing line, plot, text, screenshot

Description automatically generated

**Scatter Plot:** The scatter plot shows the count of male and female passengers based on the port of embarkation. It allows us to observe any patterns or differences in gender distribution across the embarkation points.

A picture containing text, screenshot, display, software

Description automatically generated

Age and gender intertwine in the mesmerizing KDE plot. The density of male and female passengers across age groups reveals nuanced variations. We witness the unique challenges faced by each gender as they navigate the unforgiving circumstances of the sinking ship. These visual cues allow us to empathize with their struggles and reflect on the disparities they encountered.

**KDE Plot:** So, this KDE plot helps in displaying the density of male and female passengers based on their age. It helps in understanding the age distribution of male and female passengers and identifying any variations in density between the two genders.

A picture containing text, screenshot, diagram, plot

Description automatically generated

The count plot, grouped by age intervals, provides a compelling narrative of gender composition. As we explore the count of males and females within each age group, we gain insights into the complexities of human interaction onboard the Titanic. It is within these age-defined segments that we witness the unfolding dynamics of companionship, sacrifice, and resilience.

**Count Plot:** The count plot represents the count of males and females based on age intervals. It provides a visual representation of the gender composition within different age groups.

A picture containing screenshot, text, software, plot

Description automatically generated

Transitioning into more immersive visualizations, we delve into the scatter plot. Here, the intertwining dimensions of age, gender, and survival become apparent. Each data point represents a passenger, with color indicating their survival status. We are confronted with a heartbreaking reality as the plot reveals the stories of those who managed to survive and those who perished. The significance of age and gender in determining survival becomes undeniably clear.

**Scatter Plot:** The scatter plot visualizes the distribution of male and female passengers across different age groups, with the color indicating survival status. It provides insights into age and gender-based survival patterns.

A picture containing text, screenshot, software, display

Description automatically generated

The tree map visualization invites us to explore the hierarchical breakdown of passengers based on embarkation, gender, and passenger class. As we navigate through the interconnected rectangles, we gain a deeper understanding of the diverse tapestry of individuals onboard. We are reminded that each rectangle represents a life and a unique narrative that was tragically altered by the sinking of the Titanic.

**Tree map:** The tree map displays the hierarchical breakdown of passengers based on Embarked, Sex, and Pclass. It helps in understanding the distribution of passengers across different categories and their relative proportions.

A screenshot of a computer

Description automatically generated with medium confidence

The story culminates in the mesmerizing bubble chart. Here, age, fare, gender, and passenger class merge to reveal an intricate web of relationships. The size of the bubbles indicates the passenger class, with color representing gender. We witness the multidimensional nature of their experiences, as age, fare.

**Bubble Chart:** The bubble chart reveals the relationship between age, fare, gender, and passenger class. The size of the bubbles represents the passenger class, allowing for the identification of patterns or outliers.

A picture containing text, plot, software, line

Description automatically generated

In our final exploration, we immerse ourselves in the captivating 3D scatter plot. As we navigate this multidimensional space, we uncover fascinating insights into the intertwined factors of age, gender, and survival. The three axes come alive with age, gender, and survival status, each represented by a data point in this dynamic visualization. The color of each point reveals whether the individual survived or perished aboard the Titanic. We find ourselves amidst a sea of data, each point representing a unique passenger and their harrowing tale.

Age takes center stage, with its position on the x-axis. We witness the range of ages spanning from infants to the elderly, each point representing an individual's place within this spectrum. The y-axis represents gender, with male and female passengers taking their positions in this dimensional space. The z-axis brings the ultimate question of survival to light, with points either representing the triumph of survival or the tragedy of loss.

**3D Scatter Plot:** The 3D scatter plot illustrates the relationship between age, gender, and survival status. It allows for the exploration of any clusters or trends in the data based on these three variables.

A picture containing screenshot, diagram, text, line

Description automatically generated