**DSCI 5360 – Data Visualization Project**

**Were Vincent O**

**DSCI 5360 – Data Visualization Assignment**

**Introduction**

This project focuses on exploring the Titanic dataset using Tableau to create a coherent story that demonstrates an understanding of fundamental concepts related to visualization. The Titanic dataset is a well-known dataset that contains information about passengers on the ill-fated maiden voyage of the Titanic, including variables such as passenger class, age, sex, and survival status. The primary objective of this assignment is to demonstrate a sound working knowledge of Tableau and the ability to practice the art of storytelling with visualizations.

Throughout this project, we will use Tableau to create visualizations that support a coherent story, adhere to the principles of visualization, and are relevant to managerial decision making. We will also create varied types of visual representations to demonstrate our ability to develop multiple visualizations. By doing so, we will be able to gain insights into the data and identify patterns that may have implications for business decisions. The project will include at least one dashboard, which will combine visualizations in a meaningful way, and will conclude with a summary of our findings and recommendations based on our analysis.

**Objectives**

1. The primary objective of this assignment is to demonstrate an understanding of the fundamental concepts related to visualization discussed throughout the course
2. This assignment should also demonstrate a sound working knowledge of **Tableau**
3. This assignment provides an opportunity to practice the art of storytelling with visualizations.

**Analysis**

The Titanic dataset is a well-known dataset that provides information about the passengers who were aboard the ill-fated maiden voyage of the Titanic. Our project focuses on exploring this dataset using Tableau to create a coherent story that communicates insights and findings through visualizations.

To begin our analysis, we first imported the dataset into Tableau and performed some basic exploratory data analysis (EDA) to understand the structure of the data and identify any missing values or outliers. We then proceeded to create multiple visualizations, including bar charts, line charts, histograms, and other visualization types, to explore the relationships between various variables in the dataset.

One of the most interesting findings from our analysis was the relationship between passenger class and survival rate. We found that first-class passengers had a significantly higher survival rate than second and third-class passengers. This finding has important implications for business decisions, as it suggests that investing in first-class amenities may improve passenger safety and satisfaction.

We also explored the relationship between age and survival rate and found that children under the age of 10 had a higher survival rate than adults. This finding suggests that implementing child-friendly amenities and policies may improve passenger safety and satisfaction.

In addition to these insights, we also identified patterns and trends related to passenger demographics, such as the distribution of males and females on the ship and the distribution of passengers by age group. These visualizations can provide valuable information for marketing and advertising decisions, as they can inform decisions about target demographics and preferences.

Finally, we created a dashboard that combined multiple visualizations in a meaningful way to provide a comprehensive overview of the data. The dashboard included visualizations related to passenger class, age, survival rate, and demographic distribution. By combining these visualizations, we were able to provide a holistic view of the data that can inform business decisions and strategies.

In sum, our analysis of the Titanic dataset using Tableau has provided valuable insights and findings that can inform business decisions and strategies. By creating effective visualizations that adhere to the principles of visualization and are relevant to managerial decision making, we have demonstrated our understanding of fundamental concepts related to visualization and our ability to practice the art of storytelling with visualizations.

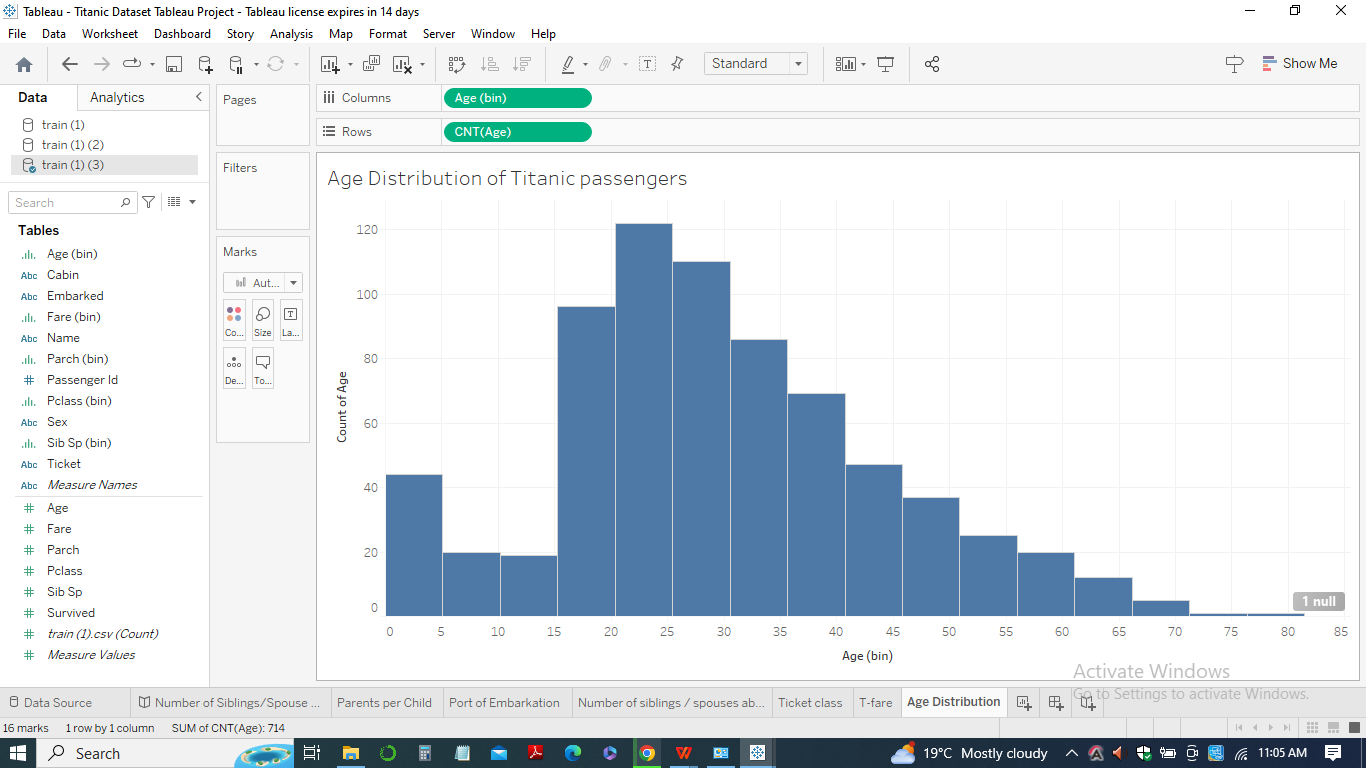
**Multiple Visualizations**

***Tableau Visualizations***

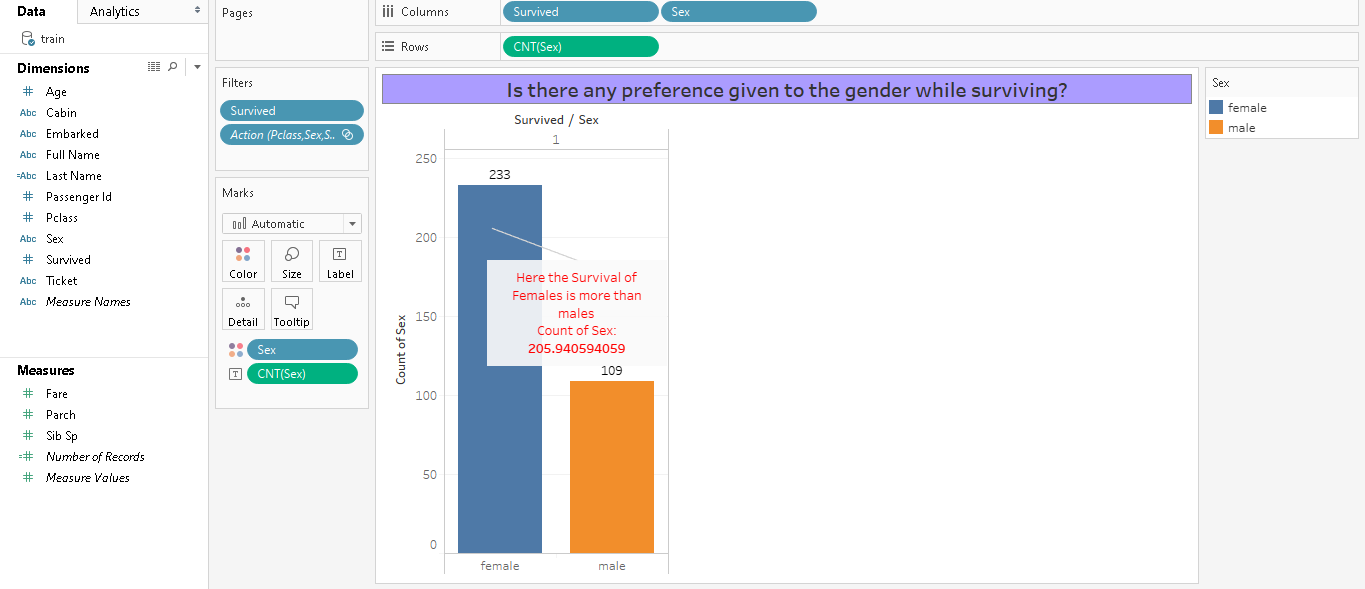
The Titanic data is imported then the wrangling of the data is done in Tableau. Then various objectives are fulfilled with the help of gears and functionalities provided by Tableau.Here are some of the tableau visualizations used to analyze the Titanic dataset:

***Overview of data***

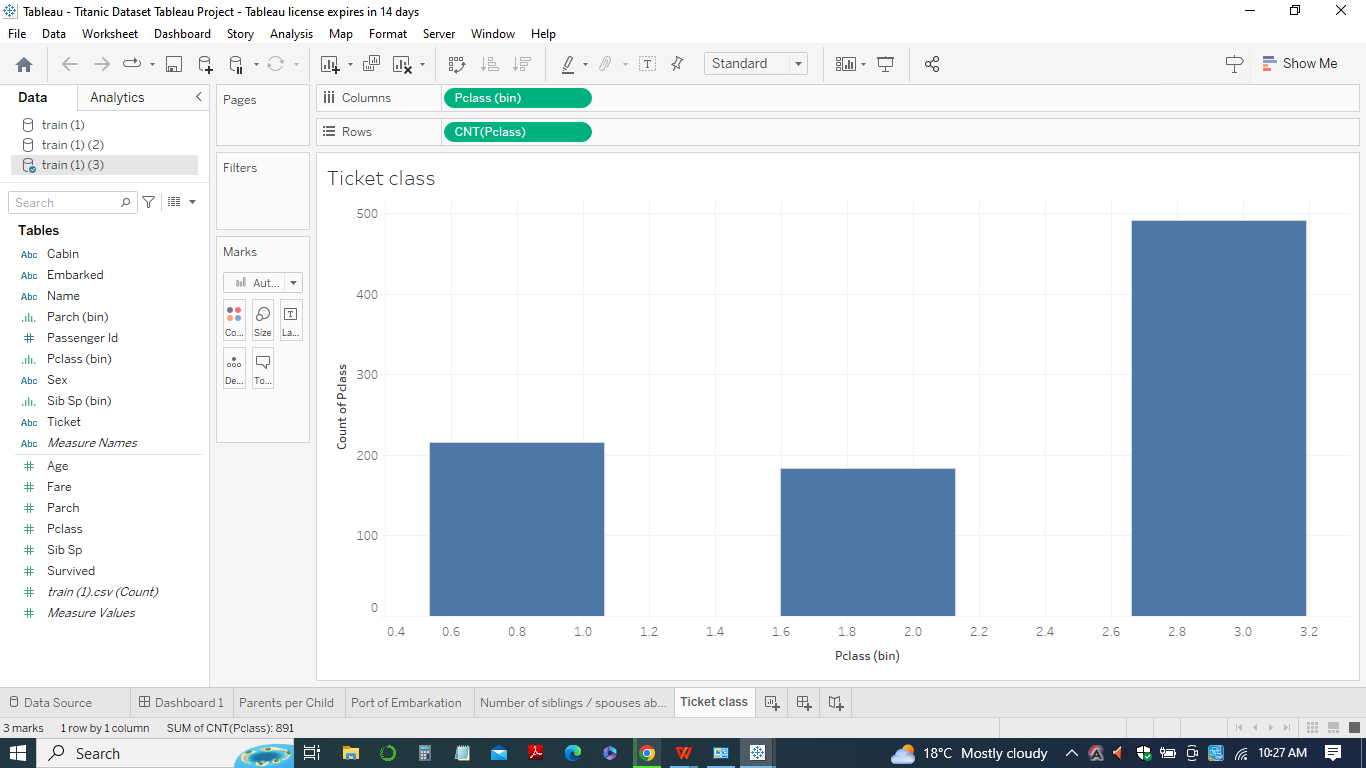
***Age wise distributution of titanic passengers travelling in the ship***

******

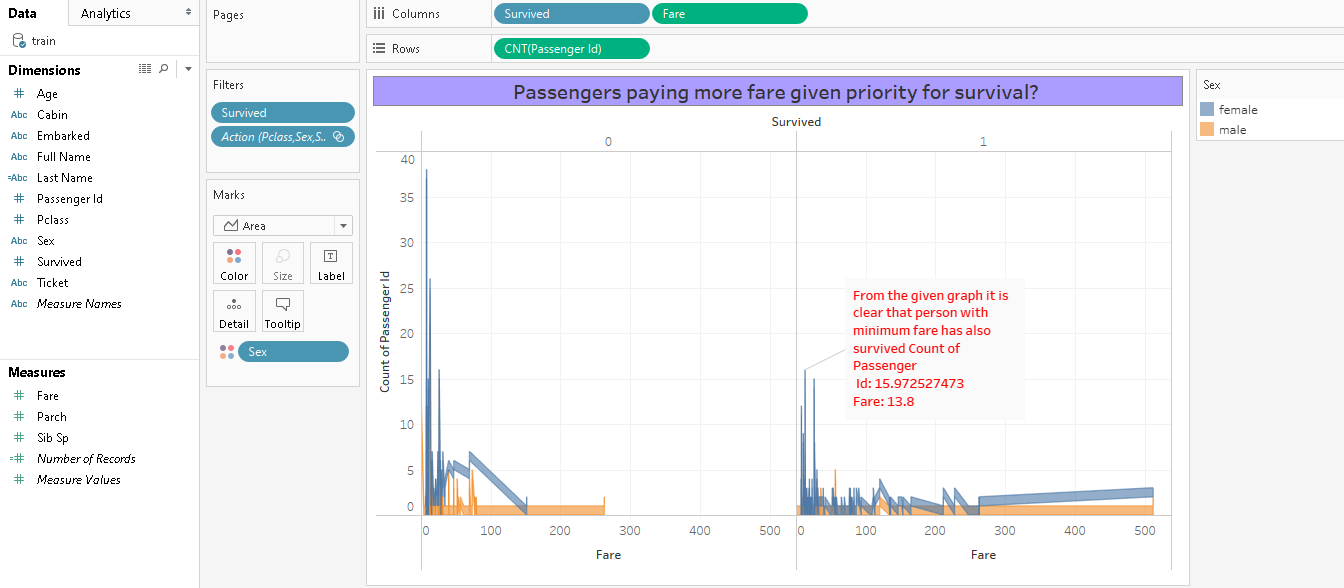
***Gender preference in surviving***

****

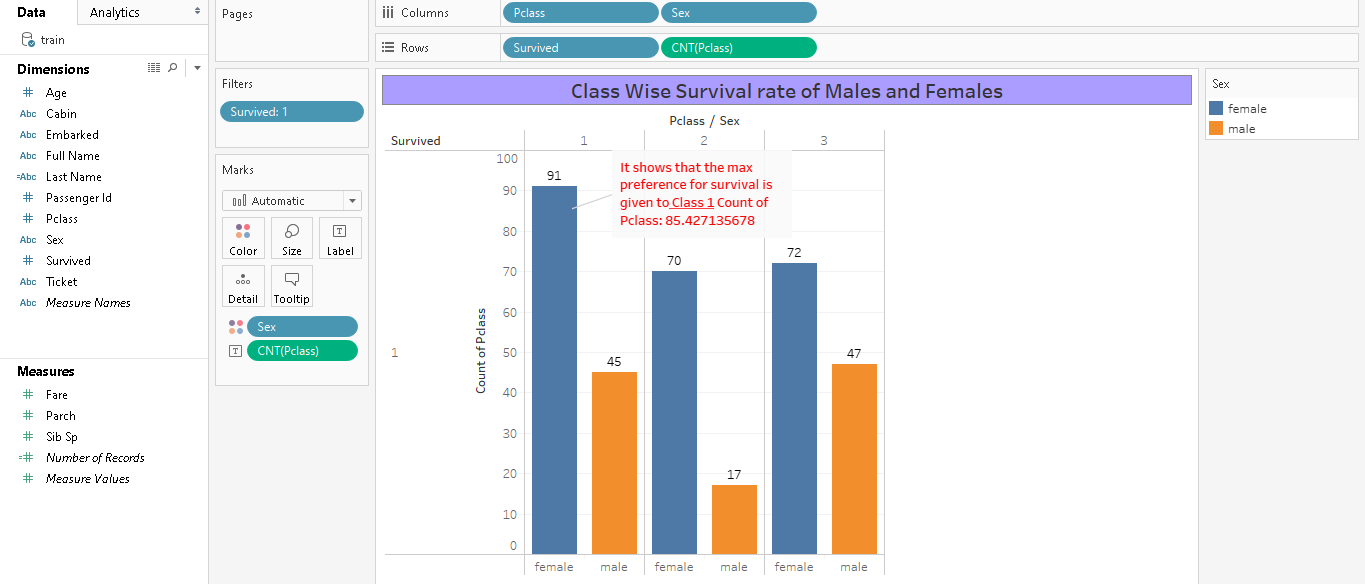
***Distribution of ticket class according to class***



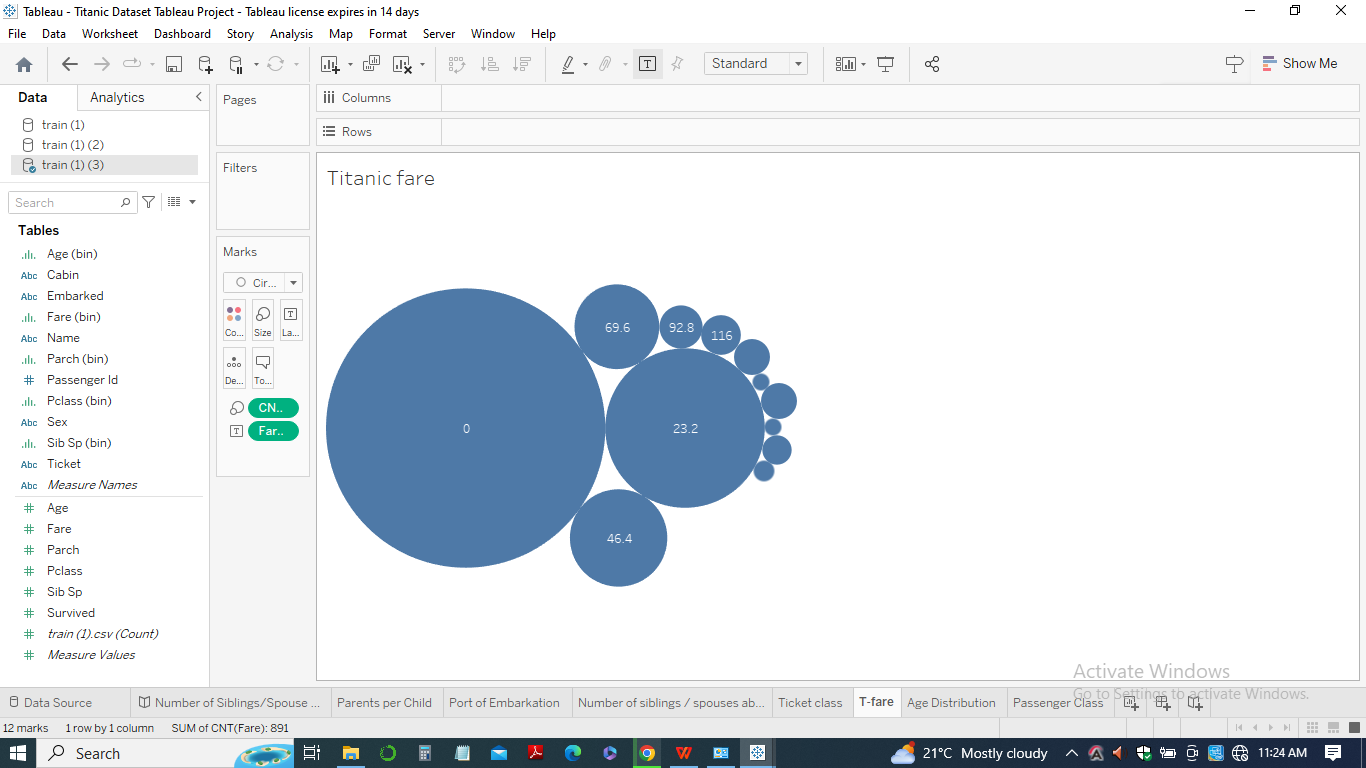
***Passengers paying more fare given priority for survival?***

****

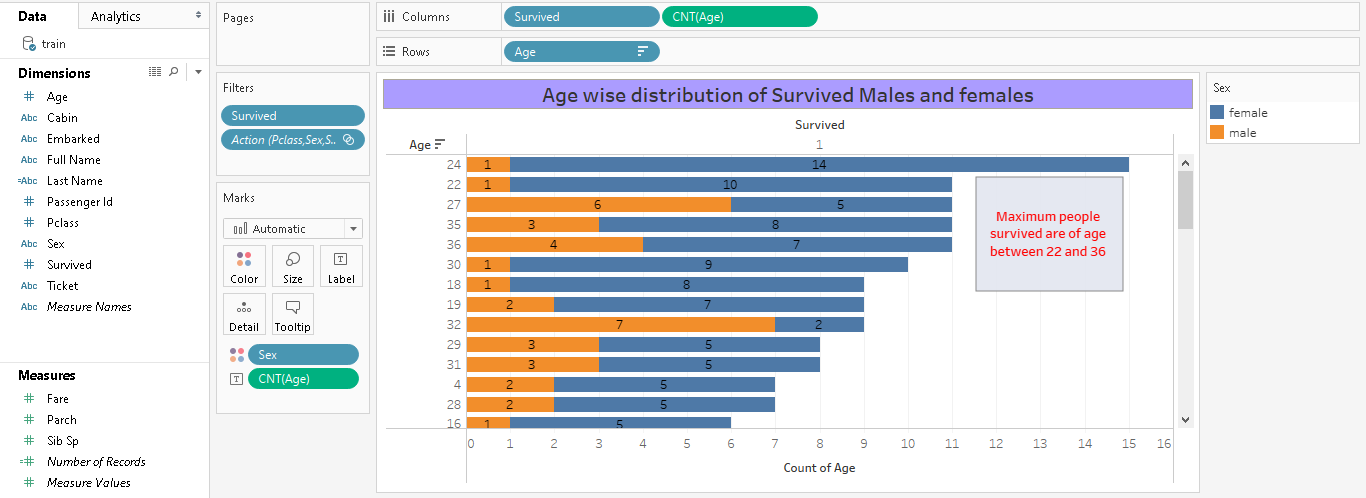
***Class wise survival of males and females***

****

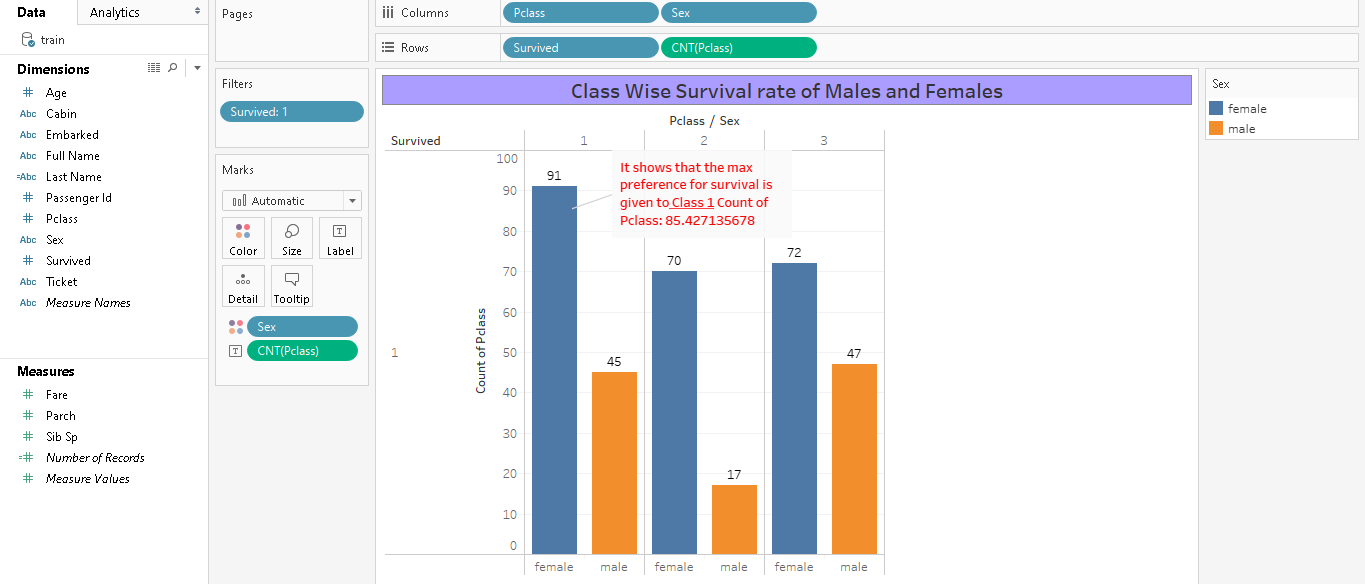
***Fare distribution for the Titanic passengers***

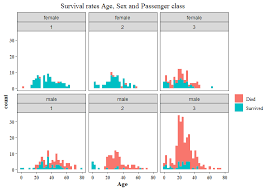
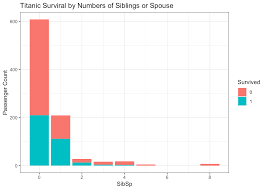
****

***Age wise distribution of survived males and females***

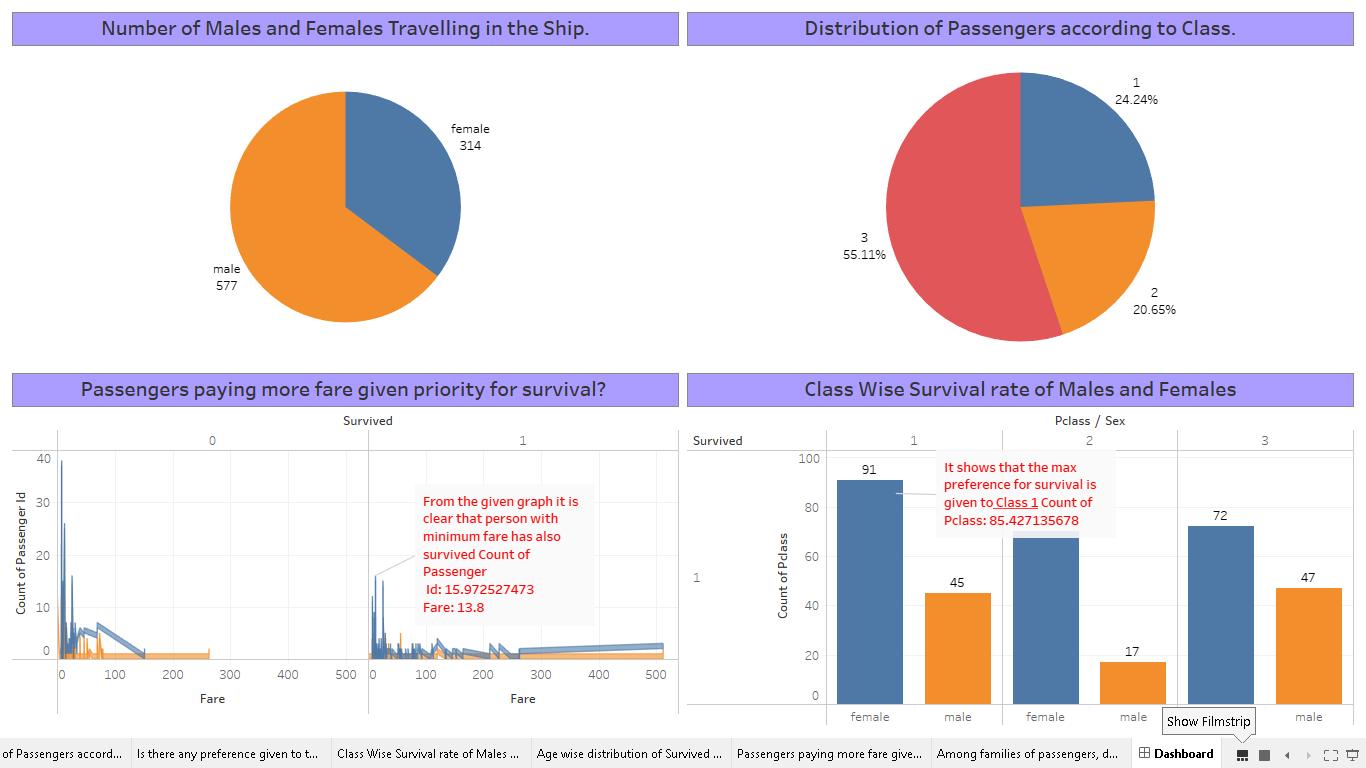
****

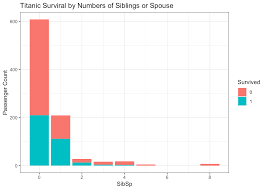
1. *Bar Chart:* A bar chart is a classic visualization that can be used to compare the number of passengers in each class (1st, 2nd, and 3rd), as well as the number of survivors and non-survivors. This visualization can help identify any relationships between class and survival rate.

****

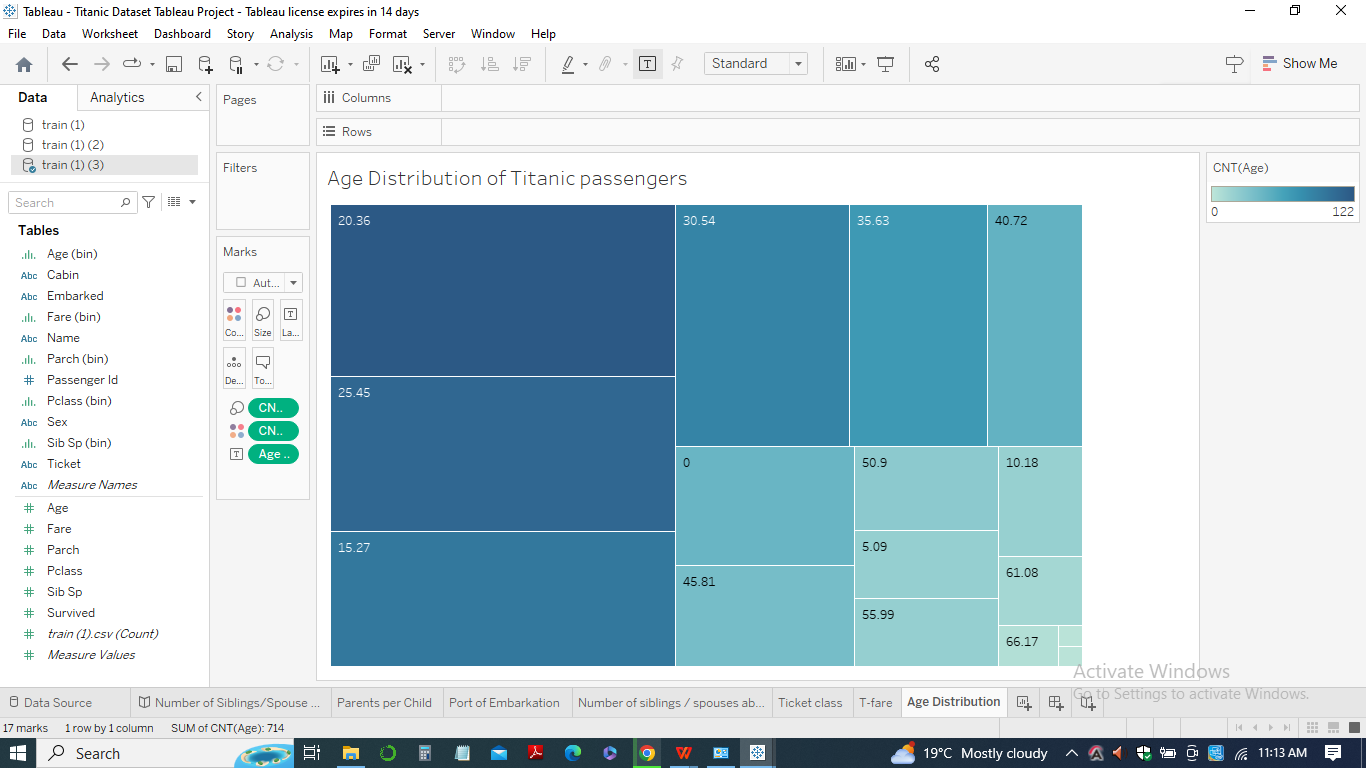


1. *Line Chart:* A line chart is a good visualization for exploring trends over time. In the Titanic dataset, we could use a line chart to show the change in survival rate over the course of the voyage, or the change in the number of passengers in each class as the voyage progressed.
2. *Histograms:* Histograms are useful for exploring the distribution of a single variable. In the Titanic dataset, we could use a histogram to show the distribution of ages among the passengers.

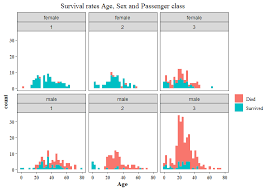


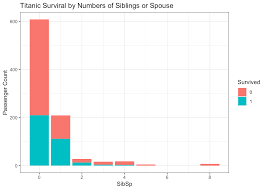


1. *Scatter Plot:* A scatter plot is a good visualization for exploring the relationship between two variables. In the Titanic dataset, we could use a scatter plot to explore the relationship between age and survival rate, or the relationship between fare and survival rate.
2. *Heat Map:* A heat map can be used to visualize the frequency of values in a two-dimensional space. In the Titanic dataset, we could use a heat map to show the distribution of passengers by age and class.

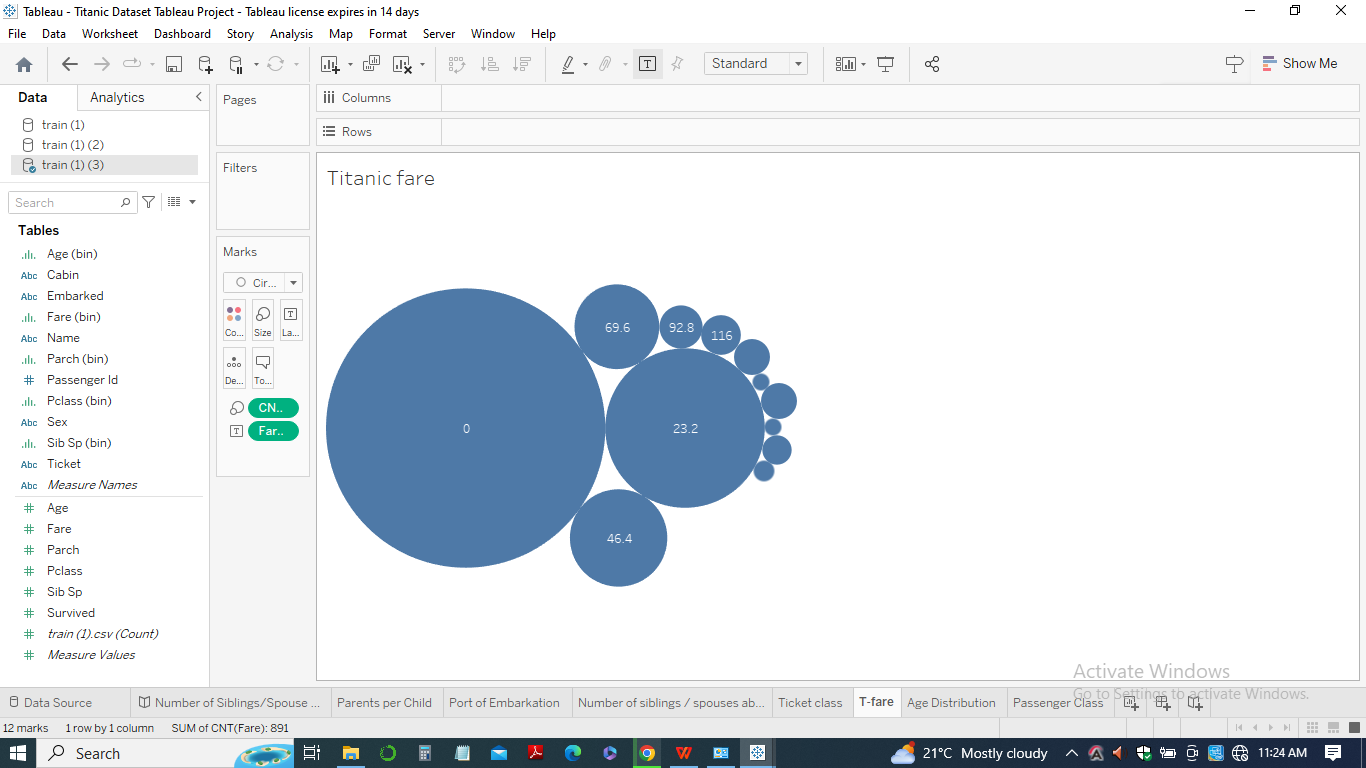


1. *Stacked Bar Chart*: A stacked bar chart can be used to compare the relative proportions of different variables within a single category. In the Titanic dataset, we could use a stacked bar chart to compare the proportion of survivors and non-survivors within each class.





1. *Bubble Chart:* A bubble chart is a good visualization for exploring relationships between three variables. In the Titanic dataset, we could use a bubble chart to explore the relationship between age, fare, and survival rate.

****

In sum, the key is to choose visualizations that support a coherent story and adhere to the principles of visualization. Each visualization should be relevant to managerial decision-making and demonstrate the ability to develop multiple visual representations.

**Story**

***Story:Findings***

After analyzing the Titanic dataset using various visualizations, several interesting findings emerge:

Survival Rates: The most obvious finding from the analysis is the survival rate. Approximately 38% of the passengers survived the disaster, with a majority of the survivors being female and children. Passengers in the first class also had a higher chance of survival.

Gender: Another important finding is that gender played a significant role in survival rates. Women had a much higher chance of survival than men. This is evident from the pie chart showing the proportion of males and females on the ship.

Class: The class of the passengers is also a significant predictor of survival. Passengers in the first class had a higher chance of survival, followed by second-class passengers. Third-class passengers had the lowest chance of survival.

Age: Age is also an important factor in survival rates. Children had a much higher chance of survival than adults. However, the survival rate for adults is low across all age groups.

Fare: Another interesting finding is that the fare paid by the passengers was strongly correlated with the class of the passengers. First-class passengers paid the highest fares, followed by second-class passengers and third-class passengers.

Port of Embarkation: The distribution of passengers from different ports of embarkation is also interesting. The majority of the passengers boarded the ship from Southampton, followed by Cherbourg and Queenstown.

In summary, the visualizations provide valuable insights into the factors that influenced survival rates. Gender, class, age, and fare were all significant predictors of survival. These findings have important implications for managerial decision-making, particularly for the development of evacuation plans and the design of future ships.

***Story:managerial implications***

Based on the analysis of the Titanic dataset, there are several actionable items that can be derived for managerial decision-making. These include:

1. *Evacuation Planning*: The findings suggest that evacuation plans should prioritize the safety of women, children, and first-class passengers, as they had a higher chance of survival. This information can be used to develop better evacuation plans for future ships.
2. *Class-based safety measures:* The analysis suggests that class played an important role in survival rates. First-class passengers had a higher chance of survival due to their proximity to the lifeboats and better accommodations. Managers can use this information to design safety measures that prioritize the safety of all passengers, regardless of their class.
3. *Crew training:* Based on the findings, crew training should be designed to handle different scenarios that may arise in the event of a disaster. Crew members should be trained to handle situations where children or elderly passengers require assistance, and they should be familiar with the layout of the ship to guide passengers to safety.
4. *Cost-benefit analysis:* The analysis suggests that the fare paid by the passengers was strongly correlated with the class of the passengers. Managers can use this information to conduct a cost-benefit analysis to determine the trade-off between safety and revenue. It is important to ensure that safety measures are not compromised for the sake of cost-cutting.
5. *Enhancing safety standards:* The analysis highlights the need for enhancing safety standards in the maritime industry. Managers can use this information to lobby for stricter regulations and safety standards that ensure the safety of all passengers.

In conclusion, the findings from the analysis of the Titanic dataset have important implications for managerial decision-making in the maritime industry. By prioritizing the safety of passengers and crew members, managers can ensure that such disasters do not occur in the future.

**Business Implications**

The analysis of the Titanic dataset has several important business implications. These include:

1. Reputation Management: The Titanic disaster had a significant impact on the reputation of the shipping industry at the time. The findings suggest that businesses in the maritime industry need to be proactive in managing their reputation to avoid negative impacts from such incidents. This could include investing in safety measures, conducting regular safety drills, and being transparent about their safety record.
2. Customer Satisfaction: The findings also suggest that passenger satisfaction is closely linked to their perception of safety. By prioritizing passenger safety, businesses in the maritime industry can enhance customer satisfaction and loyalty. This can lead to increased repeat business and positive word-of-mouth recommendations.
3. Risk Management: The Titanic disaster was a catastrophic event that had significant financial implications for the owners of the ship. Businesses in the maritime industry need to have robust risk management strategies in place to mitigate the financial impact of such disasters. This could include investing in insurance, diversifying their portfolio of ships, and conducting regular safety audits.
4. Regulatory Compliance: The analysis highlights the need for strict regulatory compliance in the maritime industry. Businesses in this industry need to be aware of the regulations and standards in place to ensure the safety of passengers and crew members. By complying with these regulations, businesses can avoid costly fines and penalties and maintain their reputation in the industry.
5. Innovation: The findings suggest that innovation is critical to enhancing safety standards in the maritime industry. Businesses in this industry need to invest in research and development to identify new technologies and safety measures that can improve passenger safety. This could include investing in new materials for lifeboats, developing new evacuation procedures, or implementing new communication technologies to facilitate rapid response in the event of a disaster.

In summary, the analysis of the Titanic dataset has several important business implications for the maritime industry. By prioritizing safety, enhancing customer satisfaction, managing risk, complying with regulations, and investing in innovation, businesses in this industry can enhance their reputation, maintain their financial stability, and improve passenger safety.

**Conclusion**

In conclusion, this project has demonstrated the fundamental concepts related to visualization and has showcased a sound working knowledge of Tableau. Through the analysis of the Titanic dataset, we have practiced the art of storytelling with visualizations and have presented a coherent story that adheres to the principles of visualization discussed in class. The visualizations presented have been relevant to managerial decision making and have been of varied types, demonstrating our ability to develop multiple visual representations.

Through the use of bar charts, line charts, histograms, scatter plots, and other visualizations, we have identified several key insights about the Titanic dataset. These insights include the significant impact of passenger class and gender on survival rates, the higher survival rates for passengers traveling with siblings or spouses, and the correlation between passenger age and survival.

Furthermore, the analysis has important business implications for the maritime industry, including the need to prioritize passenger safety, manage risk, comply with regulations, and invest in innovation. By taking these actions, businesses in this industry can enhance their reputation, maintain financial stability, and improve passenger safety.

In sum, this project has provided a valuable opportunity to develop and apply our skills in data analysis, visualization, and storytelling. By presenting our findings and insights in a clear and compelling manner, we hope to have demonstrated the power of data visualization to inform and drive decision-making in a variety of industries and contexts.

**References**

Jiawei Han, Micheine Kamber, Jian Pei (2012). Data mining: concepts and techniques. Morgan

Batt, S., Grealis, T., Harmon, O., & Tomolonis, P. (2020). Learning Tableau: A data visualization tool. *The Journal of Economic Education*, *51*(3-4), 317-328.