Project Report On Titanic Data Set

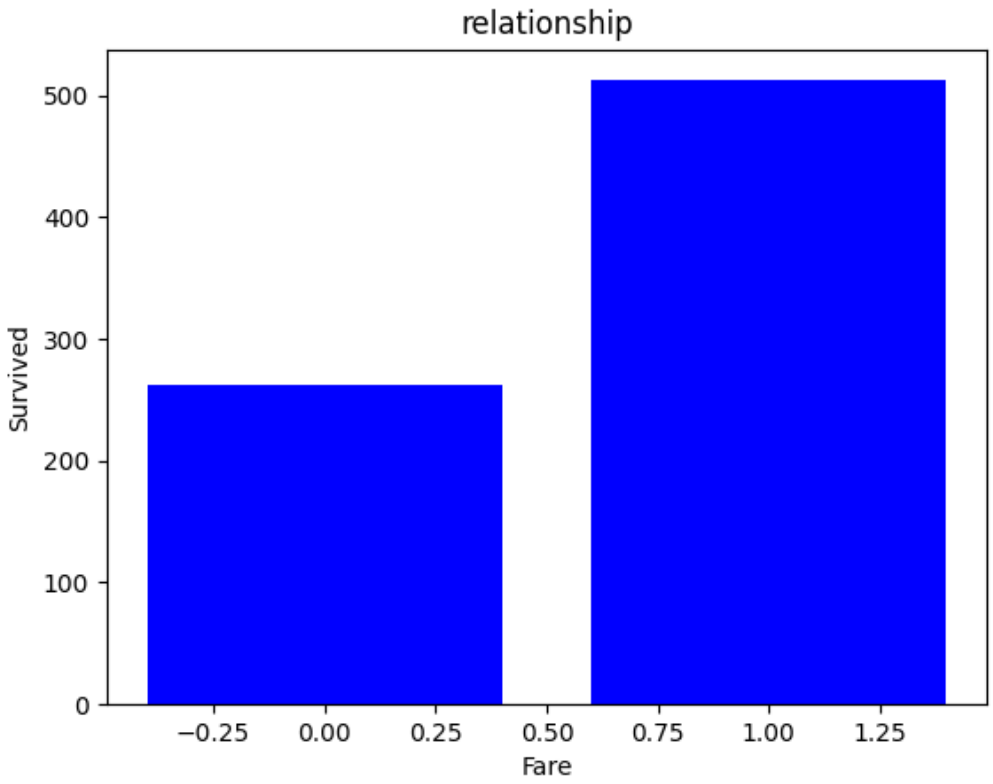
Chapter 0 : Introduction of Titanic Incident

The sinking of the Titanic is one of the most infamous shipwrecks in history.

On April 15, 1912, during her maiden voyage, the widely considered â€œunsinkableâ€• RMS Titanic sank after colliding with an iceberg. Unfortunately, there werenâ€™t enough lifeboats for everyone on board, resulting in the death of 1502 out of 2224 passengers and crew.

While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

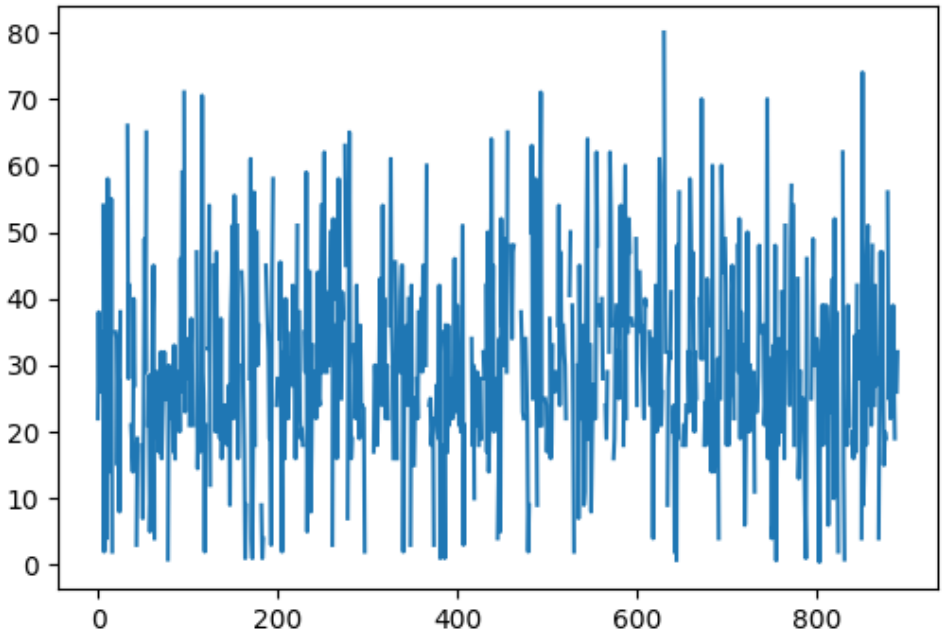
Project Report On Titanic Data Set



Chapter 1 : Variation in the Age of passengers

This line plot represents the variation in the Age of passengers from the Titanic dataset available on Kaggle. Each point on the x-axis corresponds to a unique passenger, while the y-axis indicates their respective age. The plot highlights the wide age distribution among passengers, ranging from infants to elderly individuals, with noticeable fluctuations throughout. Such visualization helps in understanding the demographic spread of the passengers onboard the Titanic, which is crucial for analyzing survival patterns and other correlations based on age. The presence of sharp peaks and drops reflects the randomness and diversity in age without any specific trend, making it an important feature for further analysis.

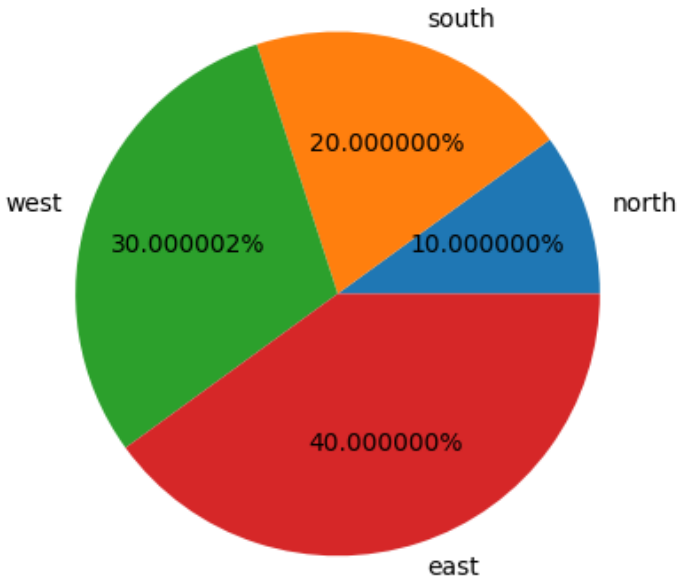
Project Report On Titanic Data Set



Chapter 2 : Relationship between Fare and Survival

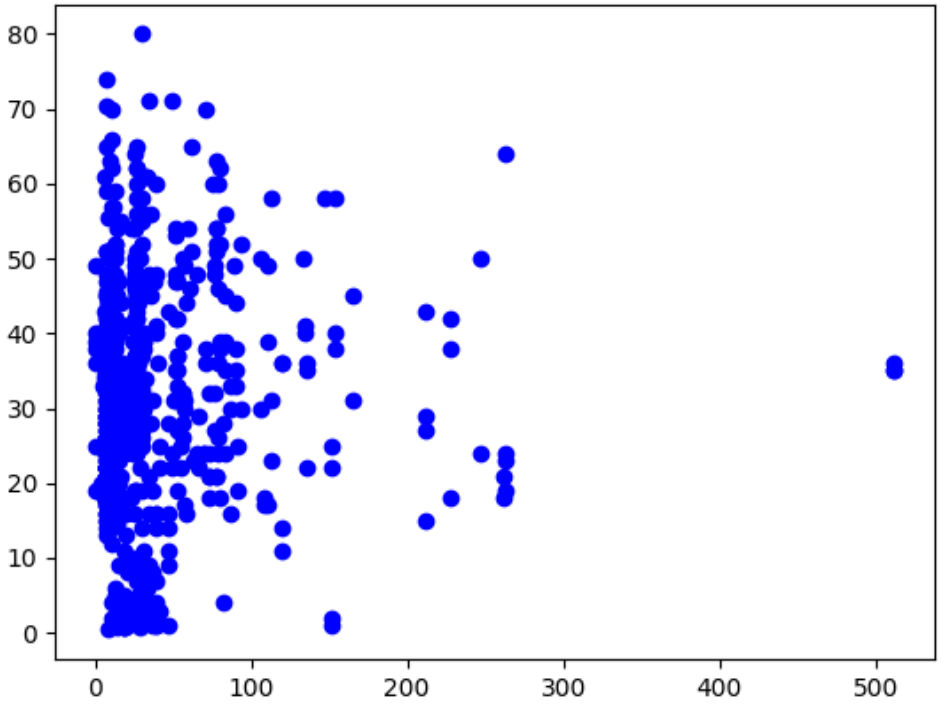
This line plot represents the variation in the Age of passengers from the Titanic dataset available on Kaggle. Each point on the x-axis corresponds to a unique passenger, while the y-axis indicates their respective age. The plot highlights the wide age distribution among passengers, ranging from infants to elderly individuals, with noticeable fluctuations throughout. Such visualization helps in understanding the demographic spread of the passengers onboard the Titanic, which is crucial for analyzing survival patterns and other correlations based on age. The presence of sharp peaks and drops reflects the randomness and diversity in age without any specific trend, making it an important feature for further analysis.

Project Report On Titanic Data Set



hapter 3 : Distribution of Titanic passengers based on their port of embarkatic

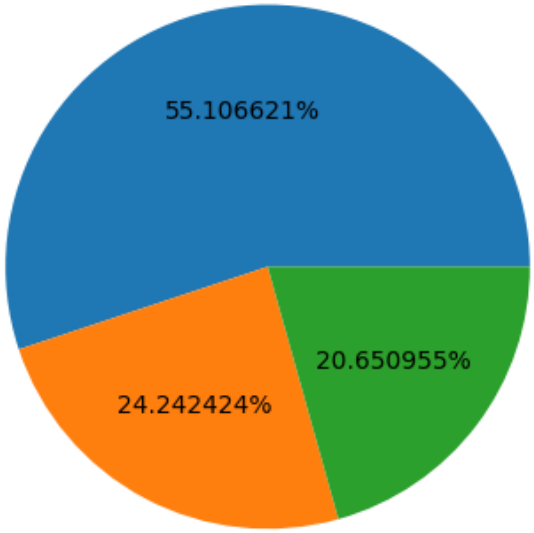
This pie chart illustrates the distribution of Titanic passengers based on their port of embarkation, derived from the Titanic dataset available on Kaggle. Although the chart labels are represented directionally as north, south, east, and west, they symbolically correspond to the actual embarkation points in the datasetâ€”Cherbourg, Queenstown, and Southampton. The chart shows that the majority of passengers, approximately 40%, boarded the Titanic from the 'east' (likely representing Southampton), followed by 30% from the 'west' (possibly Cherbourg), and 20% from the 'south' (representing Queenstown), while the remaining 10% may account for missing or undefined data. This distribution helps in understanding regional demographics and provides context for further analysis of survival trends based on embarkation locations.



Chapter 4 : Scatter Plot of Fare vs. Age

This plot displays individual passengers based on the fare they paid (x-axis) and their age (y-axis). Most passengers cluster in the lower fare range (0â€“100) with ages between 20 and 40. A small number of high-fare outliers (above 300) represent wealthy first-class travelers. The spread indicates that fare was more strongly linked to passenger class and accommodations than to age.

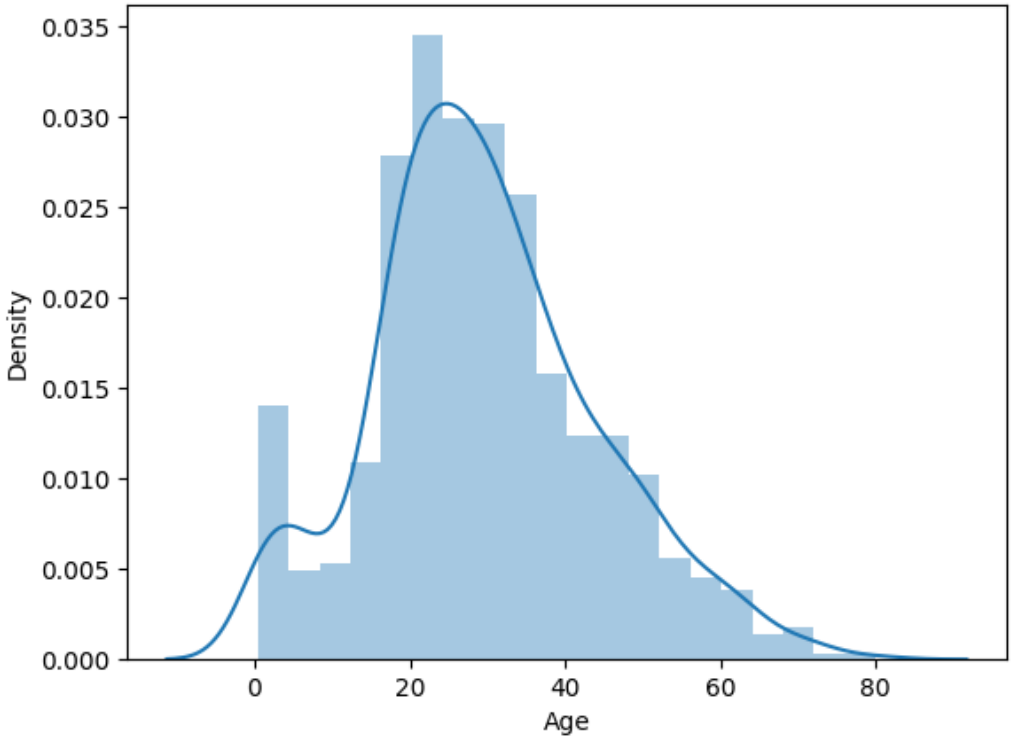
Project Report On Titanic Data Set



Chapter 5 : Proportion of Passengers by Travel Class

The pie chart reveals a heavy class imbalance: third-class passengers made up 55.11% of the shipâ€™s population, first-class accounted for 24.24%, and second-class for 20.65%. This reflects the Titanicâ€™s role as both a luxury liner for the wealthy and a vessel for emigrants and working-class travelers. This distribution is critical for survival analysis because first-class passengers were generally located on higher decks, nearer to lifeboats, while third-class passengers were in lower decks with restricted mobility and delayed access during the evacuation.

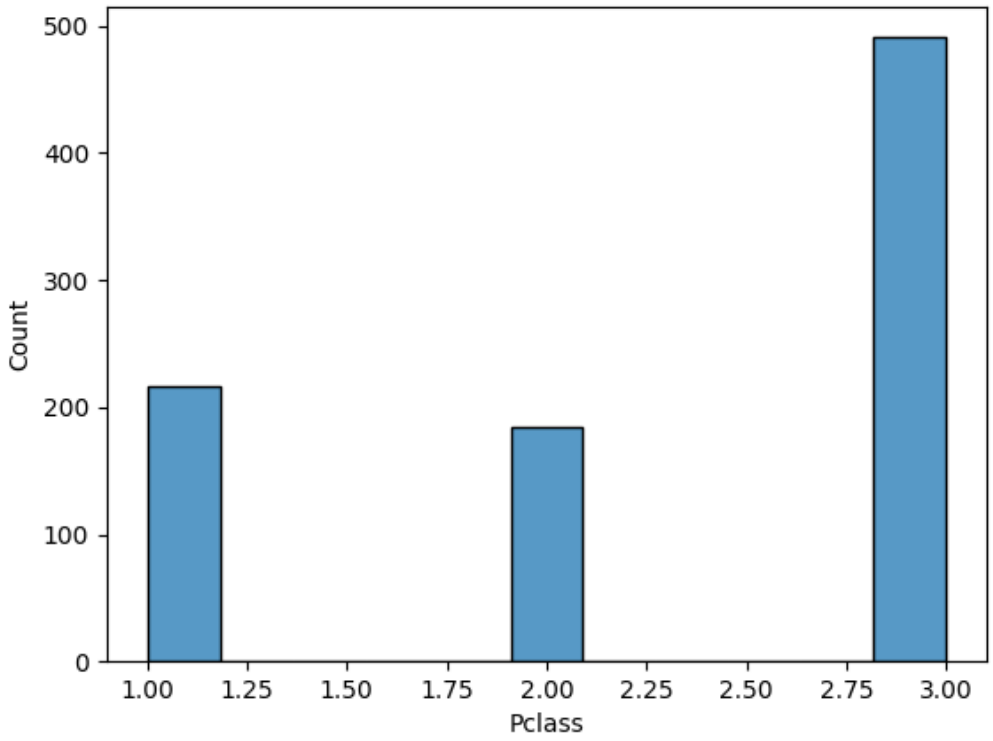
Project Report On Titanic Data Set



Chapter 6 : Age Distribution of Passengers

The histogram with a Kernel Density Estimate (KDE) curve shows that the most common passenger age was around 25 years, with the majority falling between 20 and 40. Children under 10 make up a small portion of the passenger list, while elderly individuals over 60 are rare. The distributionâ€™s right skew indicates that the passenger population leaned toward younger adults. This age structure is historically relevant because the â€œwomen and children firstâ€• evacuation policy meant children had a higher survival priority, while younger adults in lower classes often faced disadvantages due to location on the ship.

Project Report On Titanic Data Set



Chapter 7 : Passenger Count by Class

The bar chart further confirms the class imbalance. The largest group â€” third-class passengers â€” numbered around 490, followed by 220 in first class and 185 in second class. The size of the third-class group, combined with their physical distance from lifeboats and barriers such as locked gates, contributed to their disproportionately low survival rate. In contrast, first-class passengers benefited from proximity to lifeboats, better crew assistance, and quicker evacuation opportunities.