TITANIC DATASET FINDING REPORT

INSIGHTS

AROUND 62% PASSENGER DIED.

CLASS 3RD WAS THE MOST POPULATED FOLLOWED BY 1ST AND THEN 2ND.

FARE IS VERY MUCH RIGHT SKEWED.

MOST PEOPLE BUYED THE CHEAPEST TICKET

SHOCKING INSIGHT OUT HERE ALMOST MORE THAN HALF OF THE PEOPLE HEADING FOR CHERBOURG SURVIVED THE DISASTER. NEED TO DO FURTHER ANALYSIS ON THIS POINT WHETHER THIS IS BECAUSE THEY ARE MOSTLY FROM 1ST CLASS HIGH PROFILE PEOPLE. NEED TO PLOT A MAP TO UNDERSTAND THE GEOGRAPHIC POINT OF VIEW OR ECONOMIC POINT OF VIEW.

MOST OF THEM ARE HEADING TOWARDS SOUTHAMPTON, FOLLOWED BY CHERBOURG AND QUEENSTOWN.

THIS GRAPH IS VERY CLOSE NORMALLY DISTRIBUTED WHICH MEANS LESS CHANCE OF HAVING OUTLIER ALSO THE SKEW AND KURT IS ALMOST NORMAL

MOST OF THE PEOPLE WERE IN THE RANGE OF 20 TO 40 WITH THE HIGHEST SPIKE IN 24

MOST OF THEM WERE SOLO TRAVELLING

MORE MALES WERE TRAVELLING.

PCLASS 3 WAS THE DEADLIEST IMPACTED CLASS WITH THE MOST MORTALITY RATE WHILE IN THIS DISASTER SAFEST OPTION WAS PCLASS 1.

MORE MEN DIED THAN FEMALE IN THIS DISASTER ALMOST 3/4 FEMALE SURVIVED WHILE ONLY 1/5 OF MEN SURVIVED.

INFANTS AND LOWER AGE GROUP BELOW 15 SURVIVED MORE AS TO SAVE THEIR LIFE WAS THE PRIORITY BUT AFTER 15 TILL 30 THE PROBABILITY OF MORE LIKELY TO HAVE NOT SURVIVED INCREASES, THIS TREND FOLLOWS THE PATH BY AGE OF 33-34 BUT SURPRISINGLY AT 35-36 THE SURVIVING PROBABILITY AGAIN INCREASED TILL 43-44 AGAIN DROPPED BUT AGAIN RISED WHEN THE AGE GROUP IS 50 TO 60 BUT AFTER THAT NO TO NEGLIGIBLE CHANCE FOR SURVIVING.

CLEARLY VISIBLE HIGHER CLASS PEOPLE WHO PAID MORE FOR THE TICKET HAVE THE MOST CHANCE OF SURVIVING THAN POPULATION WHO BOUGHT CHEAP TICKETS.

SibSp AND Parch HAVE HIGH CORRELATION ALSO Parch AND Fare HAVE HIGH CORRELATION.

IF SOMEONE TRAVELLING ALONE THERE IS ALMOST 70% CHANCE OF NOT SURVIVING ALSO IF U HAVE LARGE FAMILY SIZE IT IS LESS LIKELY TO SURVIVE WHEREAS IF THE FAMILY SIZE IS MEDIUM (1 TO 4 MEMBERS) CHANCES OF SURVIVING IS MORE THAN 56%.

CONCLUSION

The EDA shows a dominant gender effect: women survived at roughly three to four times the rate of men (about 74% vs 19%), making sex the single strongest observed determinant in the dataset. Class was highly stratified, with 1st-class passengers surviving at about 63%, 2nd-class near 47%, and 3rd-class around 24%, reinforcing a consistent socioeconomic gradient. Embarkation patterns aligned with this gradient: Cherbourg passengers survived at about 55% compared with 39% at Queenstown and ~34% at Southampton, and the notebook flags this as a notable insight likely tied to 1st-class mix and affluence. Family structure mattered: medium families (56% survival) faced best, while those traveling alone (30%) or in large families (15%) faced worst, suggesting benefits from limited social support without the coordination costs of large groups. Age patterns indicate children and early teens had higher survival, rates dipped through young adulthood, rose again in the mid-30s to early-40s, showed a smaller lift between 50–60, and fell sharply after 60, consistent with evacuation priorities and vulnerability at older ages. Fare distributions and visuals indicate survivors were concentrated at higher fares, further corroborating the class–wealth survival advantage captured by Pclass, Embarked, and Fare together. Given substantial missing Cabin data (77%) and moderate Age missing (20%), the most reliable, high-signal features for modeling and interpretation from this EDA are Sex, Pclass, Fare, Age, Embarked, and engineered family size/type, which collectively encapsulate the study's core patterns of advantage and risk.

--THANK YOU--