Findings

- **Only around 40% of people survived.**
- **The PassengerId column is just another index column on top of the original index starting from 0, so it cannot influence a passenger's survival.**
- **Observe the Pclass column. There are three categories in this column: 1st Class, 2nd Class and 3rd Class passengers represented by the numbers 1, 2 and 3, respectively. We would expect 1st Class passengers to be given priority in boarding the ship and perhaps priority in being saved in the lifeboats. Hence, we could expect a higher survival rate of 1st Class passengers compared to 2nd or 3rd Class.**
- **We use count_n_plot to explore the number of passengers that survived in each class (by splitting the Pclass column based on the Survived column):**
- **The majority of passengers were 3rd Class, making up 55% of the passengers. Interestingly, there were more 1st Classes (216 passengers) than 2nd Classes (184 passengers) although the difference is not too much. This would make sense because the**

Titanic was designed to be a luxurious cruise ship for to accomodate the 1st class passengers.

- **The barplot indicates above 60% survival rate for 1st Class passengers, below 50% for 2nd Classes and only about 25% for 3rd Classes, even though they were the majority. This is a non-rigorous, informal study using data visualization tools to confirm that survival rate is linked to passenger class.**
- **Although 65% of passengers were male (seen from the leftmost countplot), the majority of females survived, at a survival rate of 75% (from the barplot on the right) compared to the male survival rate at 20%. This confirms another link to the target column Survived: females were more likely to survive than males. To go even further, we can split the Sex columns based on Pclass.**
- **The last categorical feature we can look at is the port of embarkment in the Embarked column. Passengers embarked from three different ports named Cherbourg, Queenstown and Southampton, abbreviated with the letters C, Q and S, respectively. We have already seen that Sex and Pclass influence the survival rate of passengers.**

- **From the countplot (and table), over 72% of passengers embarked from S(outhampton), 19% from C(herbourg) and 9% from Q(ueenstown). However, the survival rate was highest at 55% for Cherbourg, around 40% for Queenstown and 34% for Southampton.**
- **The countplot (1, 2) split by Sex has a very similar shape to the countplot split by Survived (from the previous count_n_plot code). In both cases, the ratio of Survived to Not Survived and Female to Male is roughly similar for the Cherbourg and Queenstown ports but markedly different for Southampton. There were more than twice as many males compared to females from Southampton, which explains why the overall survival rate for Southampton is much lower, at around 34%.**
- **In the barplot (2, 2), survival rates for females are around the expected 75% except for Cherbourg, where female survival rate is around 85%. For males, the unexpected survival rates are at 30% for Cherbourg and about 5% for Queenstown, compared to the expected 20%.**

- **50% of passengers who embarked from Cherbourg were in first class, whereas 72/77 passengers from Queenstown were in third class, thus explaining the slight deviation from the expected 20% survival rate.**

- **The first plot (1,1) indicates that nearly 70% of passengers had no siblings/spouses with them, and the rest had one or more. The second plot (1,2) counts the survived number of passengers, and the third plot (2,1) puts those counts as a fraction. Passengers having one or two sibling/spouse had a higher survival rate (around 55%) than those with no sibling/spouse (35%). The last plot (2,2) is the most interesting, since we can make some conjectures based on the Sex of the passenger and whether they had a spouse or not.**

- **In the last plot, using the expectation of 75% female survival and 20% male survival rates, we can see that 0 and 2 SibSp categories roughly align with this benchmark. The sample size for 3 or more SibSp is too small, so these categories need not adhere to the expection. The interesting bit is the higher than expected 30% survival rate for male passengers with 1 SibSp. This could occur,**

for instance, when both the man and his wife are 1st class passengers on the trip and they board the lifeboats together.

- **similar to the SibSp column, passengers with no parents/children were less likely to survive (around 35% as seen from the third plot (2,1)) than those with one or two parents/children (around 55%). From the last plot (2,2), males with one or two parents/children also had a higher survival rate (30%) compared to those with no parents/children.**