**THE TITANIC DATASET:**

General overview of data and who were the passengers:

* + - The titanic data set is imported from kaggle website
    - The dataset is initially explored, it has the following attributes as its columns:
      * 'PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp', 'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'
    - A factor plot on the sex of the passengers is plotted and reveals the disproportionately high male population in the ship.
      * In the factorplot itself, the hue is mad as the ‘Pclass’ and reveals the distribution of the male and female population over the 3 different passenger classes.
    - A function is explicitly defined to differentiate between males, females and children
      * This function is then applied over the dataframe and the resulting column is named ‘person’.
      * The same distribution is then plotted over a factor plot same as before, this time plotting the ‘person’ column instead of the ‘Sex’ column.
    - A histogram is also plotted over the various bins of ages.
    - The counts of each of the person types were also found using the ‘.value\_counts()’ member function .
    - The kde plot of the distribution of the different ages of the passengers is made. Multiple kde plots can be constructed in the same plot using the function ‘FactGrid()’ wherein multiple kde plots are crated which are distinguished by the sex of the passengers.
      * The resulting graph interpretation was that the male population was more or less more dominant than the female population, but the female population was more in the age group of 0-20 years. The male population was dominant in the age group of 20-40 years and also marginally in the elder age group.

Distribution of the classes, decks and where did the passengers come from:

* The cabin column from the dataframe is cleansed of NaN values and stored in another Series. The first letter of this series is then extracted from the series and stored in a list using the for loop. Then it is converted to a dataframe.
* The information about the decks are then plotted using the factor plot same as before.
  + The resulting plot showed an increased count in the cabin ’C’ and last in cabin ‘G’.
* The same is done after eliminated the defaulting attribute ‘T’.
* A factorplot is plotted over the embarked destination, with the hues being the ‘Pclass’
  + The resulting interpretation is that irrespective of the embarked destination, the population of the people in the lower economic strata namely the ones in the 3rd class dominate. Especially Queenstown, where almost all the passengers embarked were third class.

Who was alone and who was with family:

* A new attribute alone is created as the sum of the ‘SibSp’ and ‘Parch’ attributes, and then in all the locations of alone where it equals 0, we changed it to ‘alone’ else gave the value ‘with family’.
* Upon constructing a facet plot on ‘alone ‘,with varying hues:
  + Most of the passengers travelled in titanic alone
  + Most of the people who travelled alone were males while the sex ratio of the ones travelled with family were almost the comparable
  + A large number of people who travelled alone comprised of the 3rd class
  + The ratios were comparable on the ones with family

Who all survived ?:

* ‘survivor’ column is created by mapping with the ‘Survived’ column. The zeroes correspond to ‘no’ and the ones correspond to ‘yes’.
* On plotting the ‘survivor’ column along with other dependencies using factorplot and lmplot :
  + The number of people who died is significantly more than the ones survived.
  + The 3rd class people were the ones that suffered maximum percentage of casualty, followed by 2nd class, followed by the 1st class.
  + The males have a higher percentage casualty irrespective of the class, with the second and 3rd being the worst hit.
  + The females have a better survival, but among them, the lower the class, the lesser is their survival.
  + The lower aged ones have better survival as compared to the middle and older aged ones for the second and 3rd classes.
  + The middle aged people have the lowest survival for the 2nd and 3rd class, but interestingly for the 1st class, the middle aged people had better survival.
  + Overall, Irrespective of the classes, the older the age the lesser is the chance of survival.
  + On interpreting the dependencies of the survival , age and sex it was found that overall males had a lesser percentage of survival than females.
  + But, as the age increases, the survival for males decreased and the survival for females increased.
  + Ie, an old aged male had the least survival probability while an old aged female had a relatively high survival probability.
  + The chances of survival in deck ‘B’ was the highest while it was more probable for a person of deck ‘A’ to die.
  + All the other decks have comparable survivors and casualties.
  + A person travelling alone is far more likely to die as compared to with family.
  + Families have equal percentages of survival and death.