Data Visualization in python

Here in python data visualization we will work on the “Titanic” dataset available in the seaborn module in python package.

Introduction

Python is a great language for doing data analysis, primarily because of the fantastic ecosystem of data-centric Python packages. **Pandas** is one of those packages, and makes importing and analyzing data much easier.

A Data Analyst uses data visualization and manipulation techniques to uncover insights and help organizations make better decisions.

Basic details of “titanic” dataset are:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 891 entries, 0 to 890

Data columns (total 15 columns):

survived 891 non-null int64

pclass 891 non-null int64

sex 891 non-null object

age 714 non-null float64

sibsp 891 non-null int64

parch 891 non-null int64

fare 891 non-null float64

embarked 889 non-null object

class 891 non-null category

who 891 non-null object

adult\_male 891 non-null bool

deck 203 non-null category

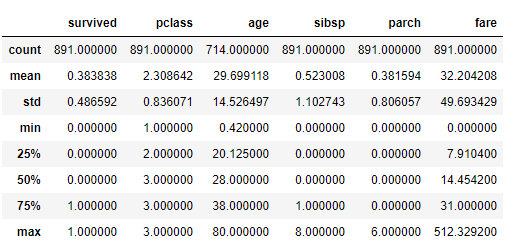
embark\_town 889 non-null object

alive 891 non-null object

alone 891 non-null bool

dtypes: bool(2), category(2), float64(2), int64(4), object(5)

memory usage: 80.6+ KB



Now we will start our data analysis and visualization on Titanic dataset with the answering of these questions:

Table of contents:

1.) Who were the passengers on the Titanic? (Ages,Gender,Class,..etc)

2.) What deck were the passengers on and how does that relate to their class?

3.) Where did the passengers come from?

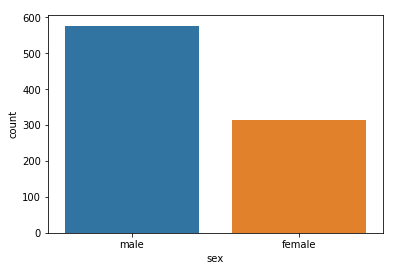
4.) Who was alone and who was with family?

5.) What factors helped someone survive the sinking?

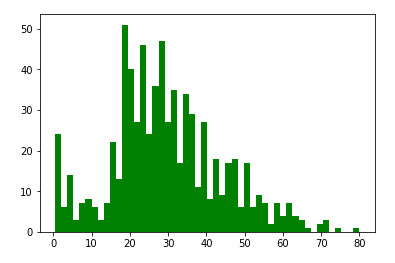
Answering the questions:

1.) Who were the passengers on the Titanic? (Ages,Gender,Class,..etc)

Counts of passengers based on gender.

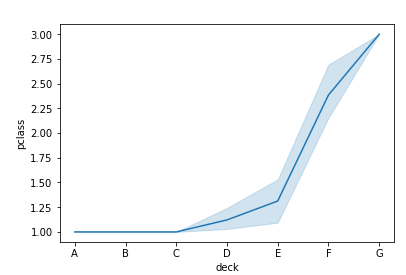


Counts of passengers based on age



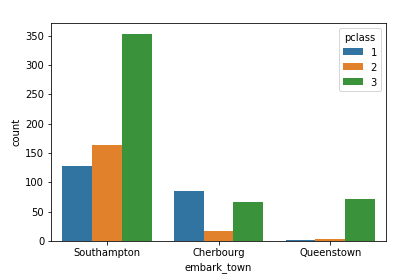
2.) What deck were the passengers on and how does that relate to their class?

The following figure answers the above question. There were A,B,C,D,E,F and G decks are available and class as 1,2,3.



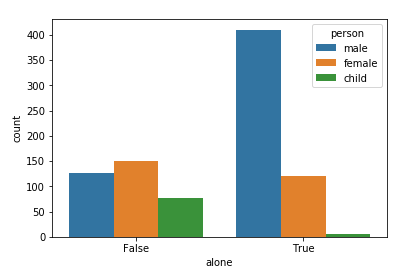
3.) Where did the passengers come from?

In the below figure we can see that passengers were from Southampton, Cherbourg and Queenstown.The figure shows the Number of passengers from different town differentiated by class.

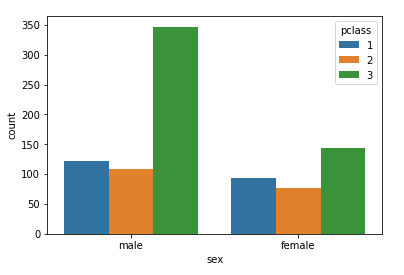


4.) Who was alone and who was with family?

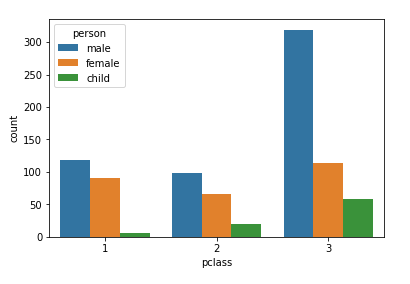
From figure we can see that most of the male passengers were alone.



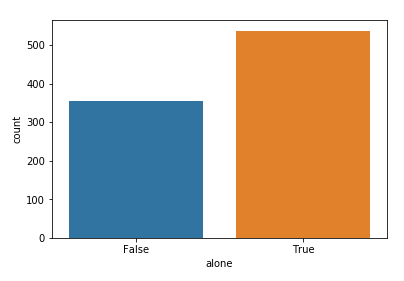
In the following figure we can see the number of male and female passengers according to the different class.



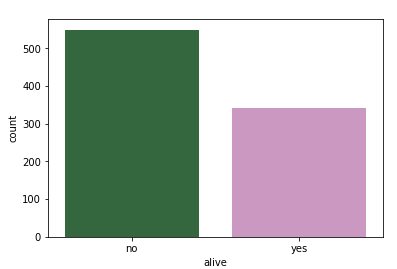
In the below figure we can see the different aspect of the above figure i.e. the different number of passengers in class according to as they are male,female and child.



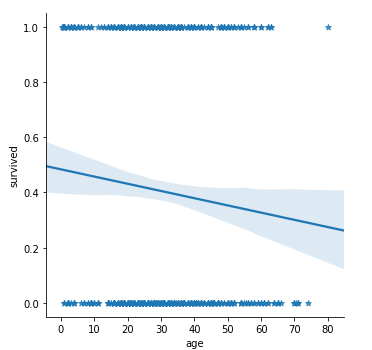
From the below figure we can draw the conclusion that there are more alone passengers on titanic than passengers with family.



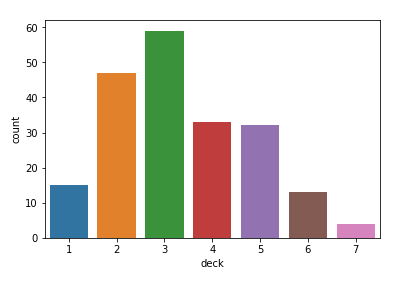
From the figure plotted below we can see that the survival rate is less on titanic. There are more passengers who could not survive.



From the below figure, we can deduce that the old aged passenger are less in number of survival whereas less aged passengers are more survived.

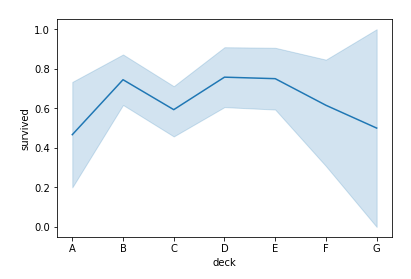


The following figure shows the number of passengers on the different decks on titanic.



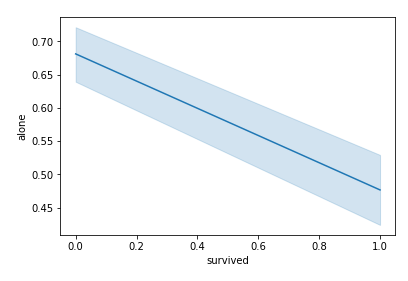
Did the deck have an effect on the passengers survival rate? Did this answer match up with your intuition?

The below figure answers the above question. We can see that Deck E and F have more survival rate than any other deck.



Did having a family member increase the odds of surviving the crash?

The below figure reveals the truth that the passengers who were alone have less survival rate whereas the passengers with family have more survival rate.



Conclusion:

So we can say that with data analytics and data visualization in python, we can reveal many hidden facts inside the raw data if proper calculation and analytical applications are applied on datasets.