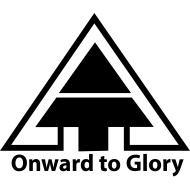
**DSA MINI PROJECT REPORT**

**TITANIC: PREDICTING SURVIVAL**



ARMY INSTITUTE OF TECHNOLOGY

(DEPARTEMENT OF COMPUTER ENGINEERING)

SE-COMPUTER SEM-1 2018-2019

SUBMITTED BY-

K. CHAITANYA KUMAR (3248)

KUMAR SHIVAM (3249)

**1. ABSTRACT**-

In this project , we complete the analysis of what sorts of people were likely to survive. In particular , we are applying the tools of machine learning to predict which passengers survived the tragedy.

**2. Software requirement.**

* Jupyter notebook
* Word Editor (Notepad++)

**Hardware requirement.**

* Network Connection

**3. INTRODUCTION-**

Machine Learning is an idea to learn from examples and experience, without being explicitly programmed. Instead of writing code, you feed data to the generic algorithm, and it builds logic based on the data given.

“A computer program is said to learn from experience E with some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.”

The sinking of the RMS Titanic is one of the most infamous shipwrecks in history.  On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class.

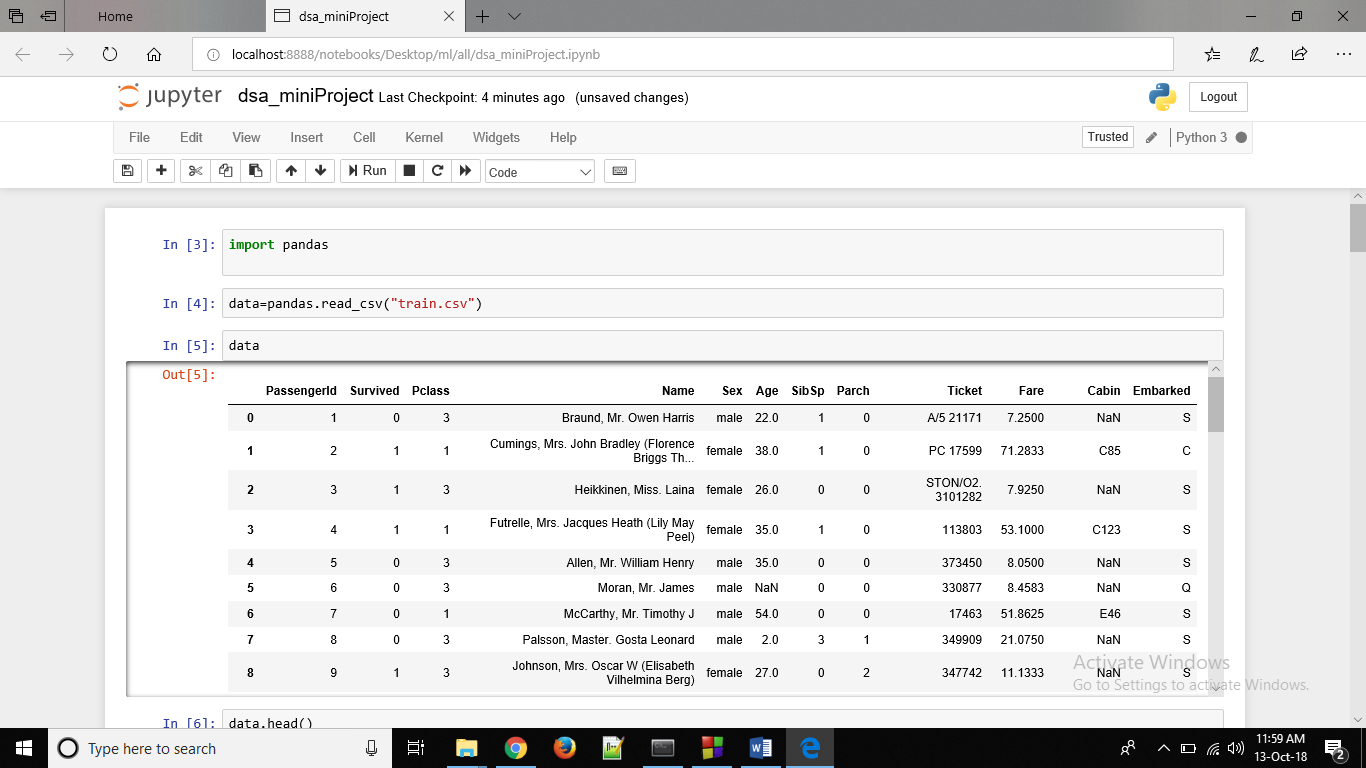
Thus, we have used a lazy algorithm to train the model called KNN. It is also known as

k-Neighbors algorithm. By providing some value of k that is number of nearest neighbors

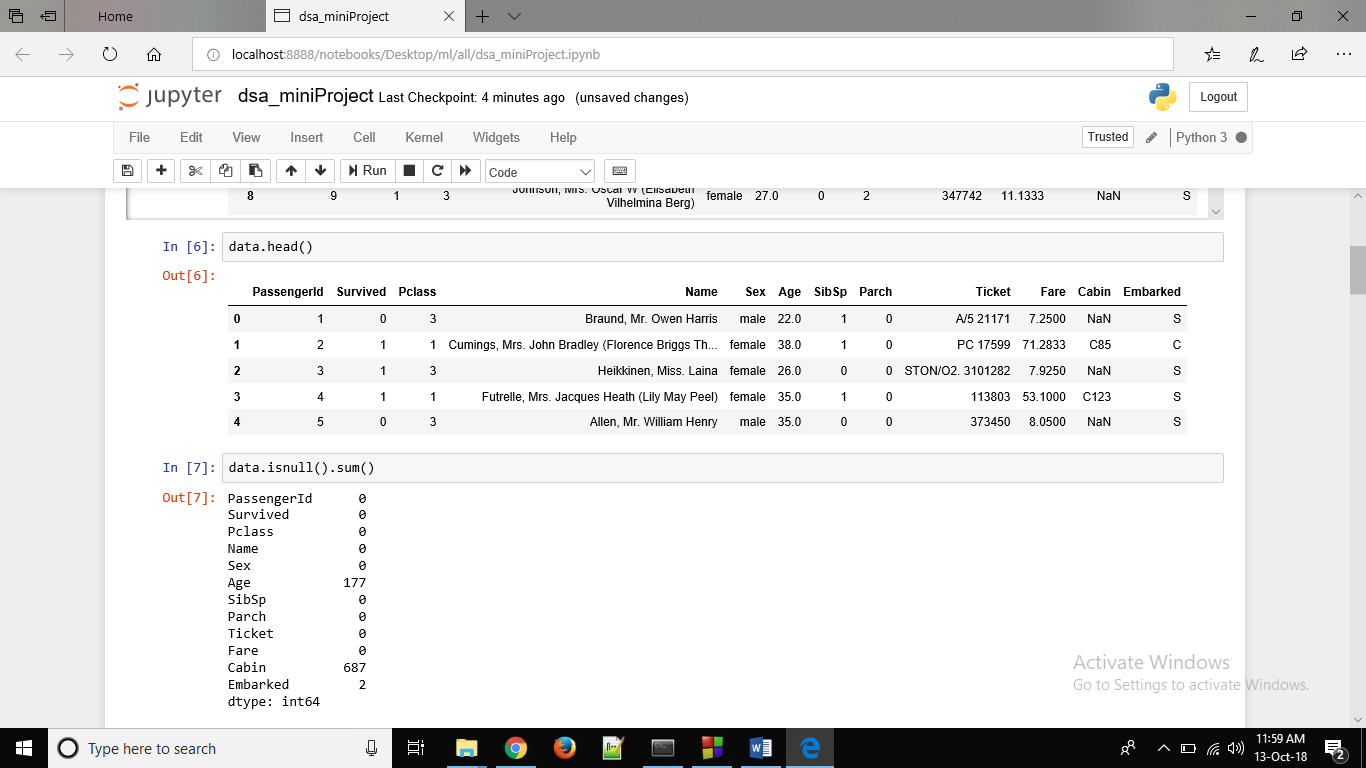
we calculate the Euclidian distance between them and classify them accordingly.

**4. SNAPSHOTS-**

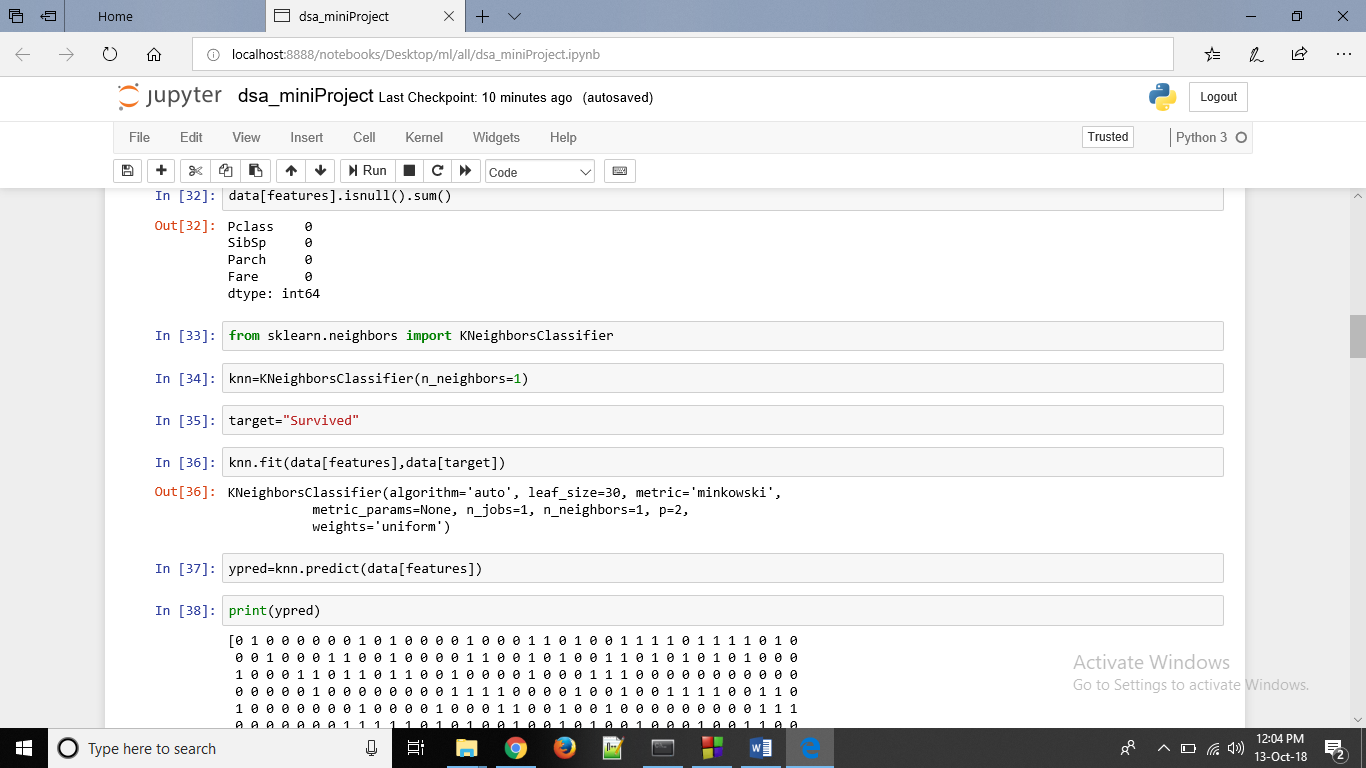
Training Data



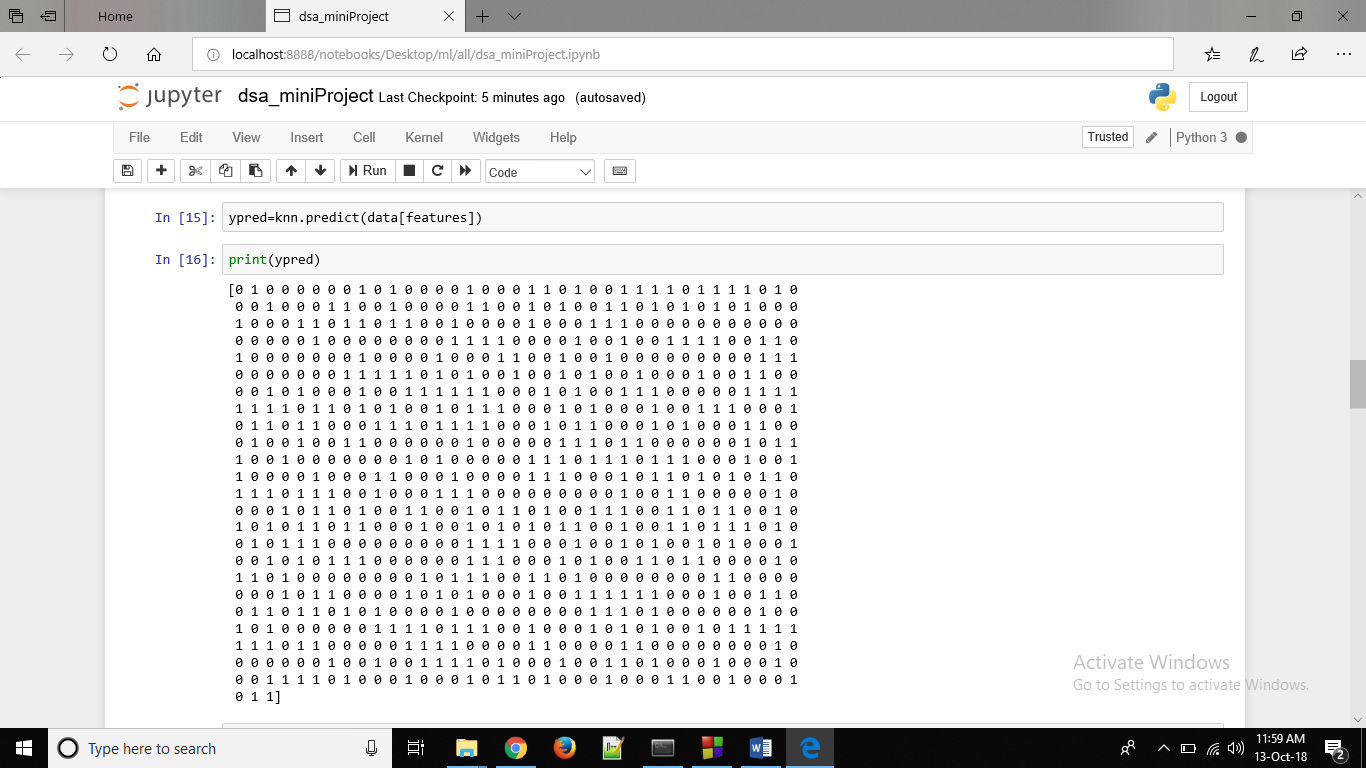
General View of dataset



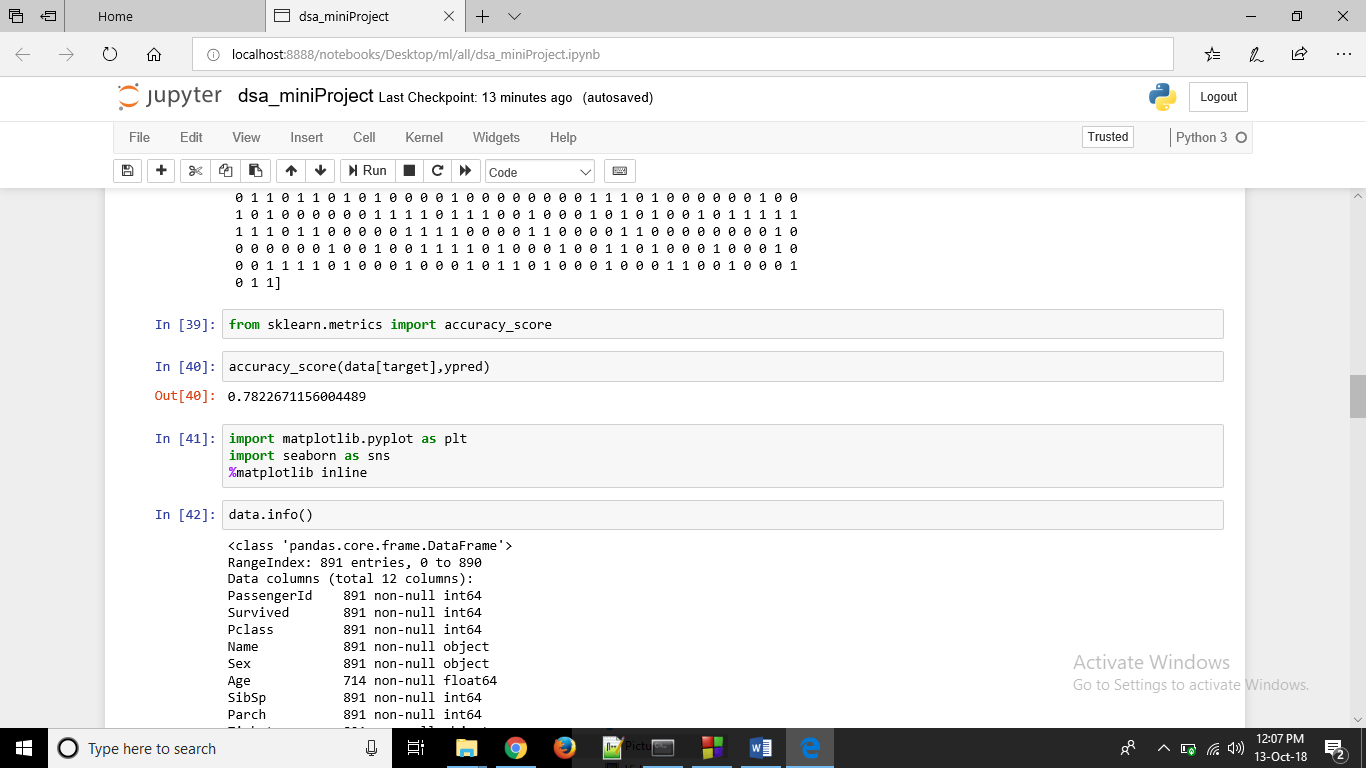
Applying Algorithm in model (KNN ALGORITHM)



Predicting the Survival using KNN

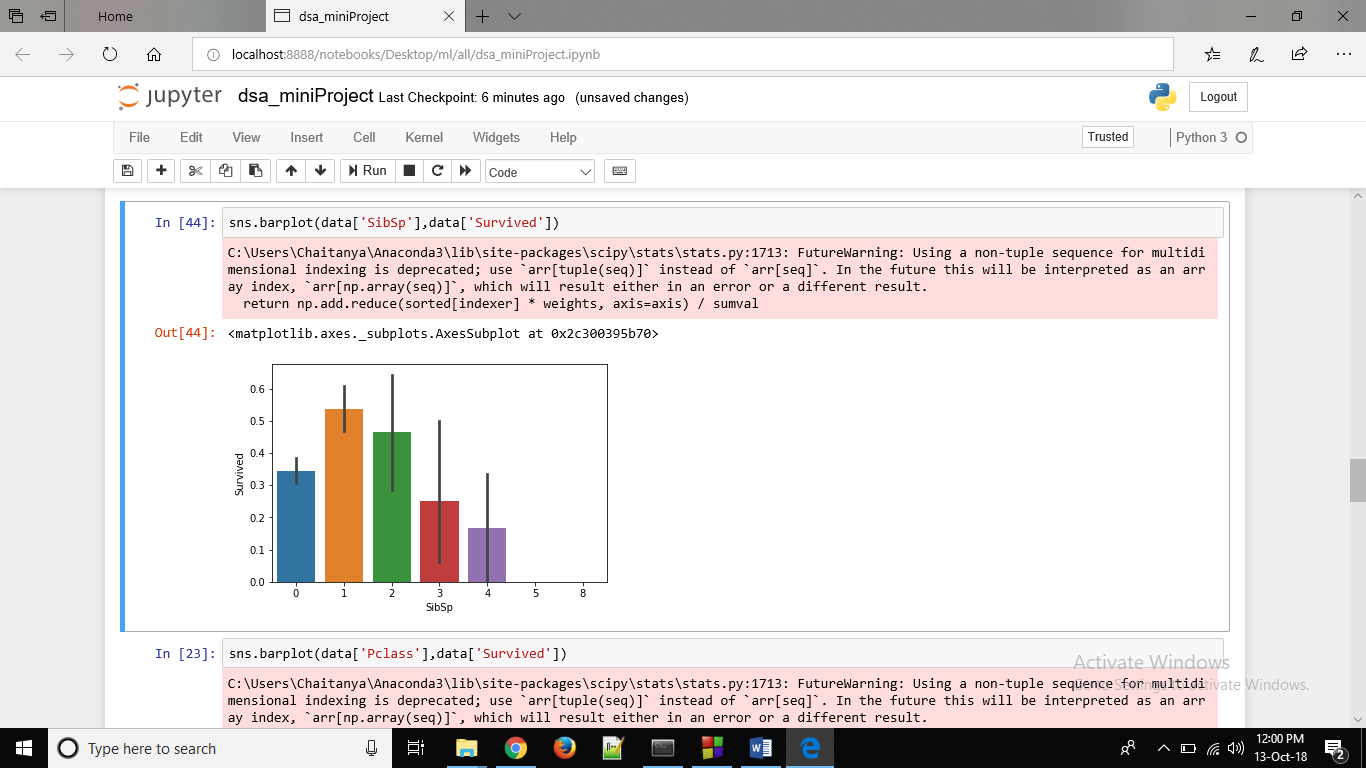


Predicting the accuracy using KNN

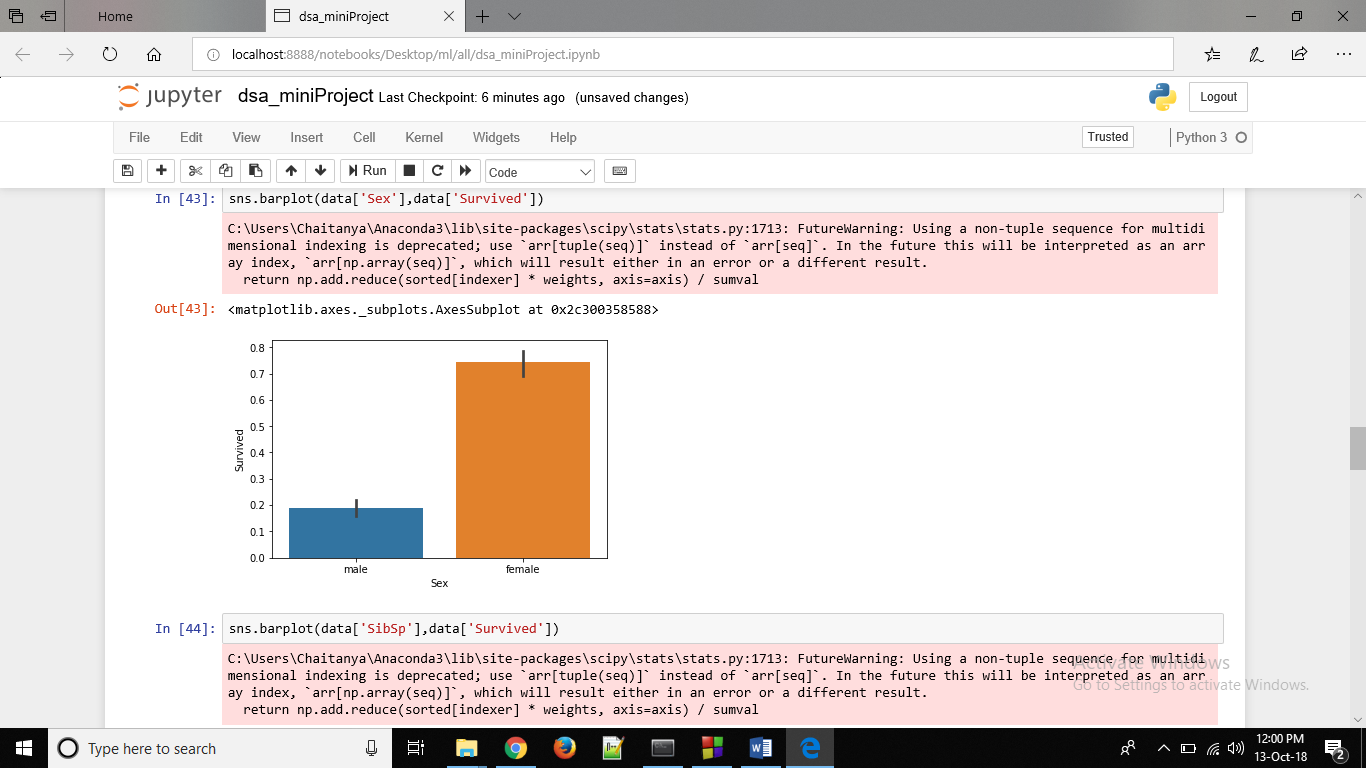
****

**Graphs being plotted to analyze the train data with the Survival**

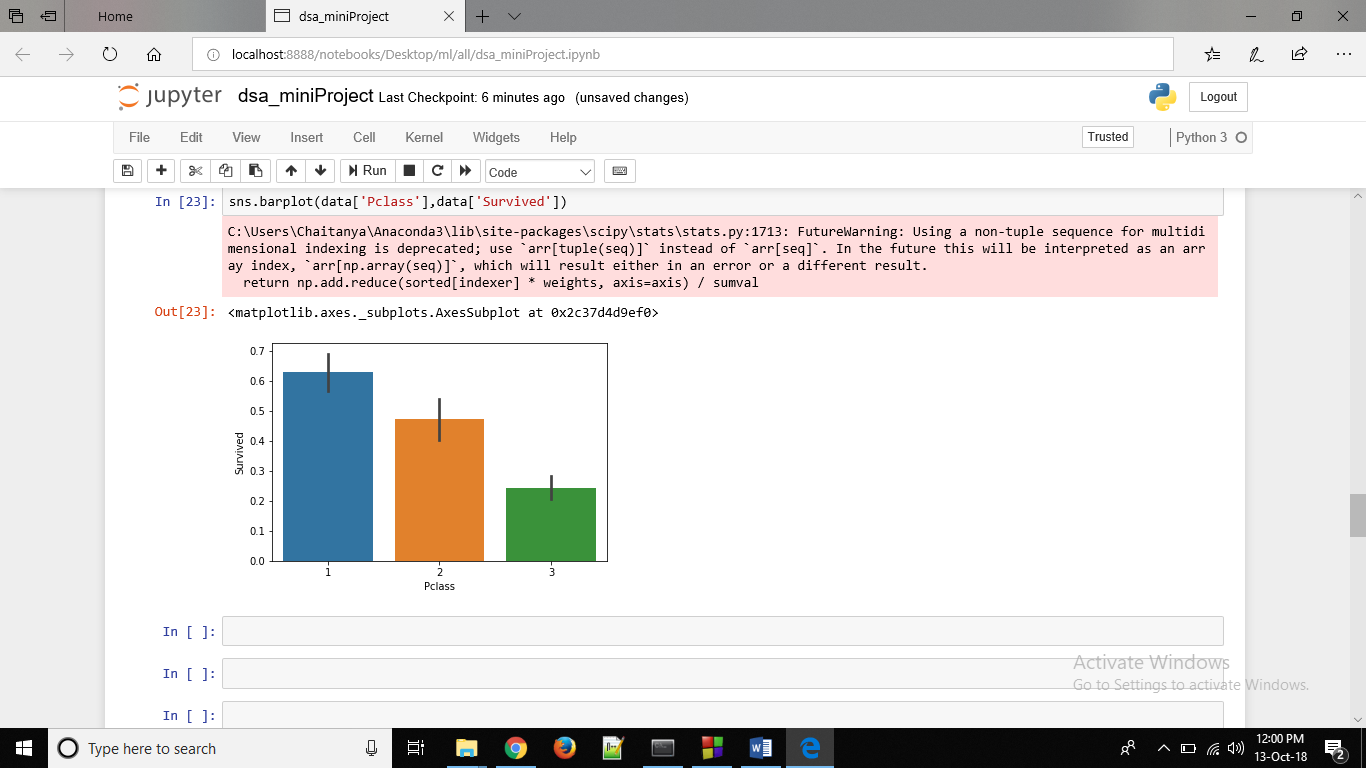
1. **Siblings vs survived**



1. **Sex vs Survived**



1. **Passenger class vs survived**



1. **ALGORITHMS AND DATA STRUCTURE USED.**

***k*-nearest neighbors algorithm** (***k*-NN**)

Definition

In [pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition), the ***k*-nearest neighbors algorithm** (***k*-NN**) is non – parametric  method used for classification and [regression](https://en.wikipedia.org/wiki/Regression_analysis). In both cases, the input consists of the *k* closest training examples in the feature space. The output depends on whether *k*-NN is used for classification or regression:

* In *k-NN classification*, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer, typically small). If *k* = 1, then the object is simply assigned to the class of that single nearest neighbor.
* In *k-NN regression*, the output is the property value for the object. This value is the average of the values of its *k* nearest neighbors.

The algorithm can be summarized as:

1. A positive integer k is specified, along with a new sample
2. We select the k entries in our database which are closest to the new sample
3. We find the most common classification of these entries
4. This is the classification we give to the new sample

A few other features of KNN:

* KNN stores the entire training dataset which it uses as its representation.
* KNN does not learn any model.
* KNN makes predictions just-in-time by calculating the similarity between an input sample and each training instance.

**6. REFERENCES**

The source code of this project can be found on:

**https://github.com/TheForeverLost/AIT-Hostel**

**7. REFERENCES**

* <https://en.wikipedia.org/wiki/KNN>
* <https://www.kaggle.com/c/titanic/data>