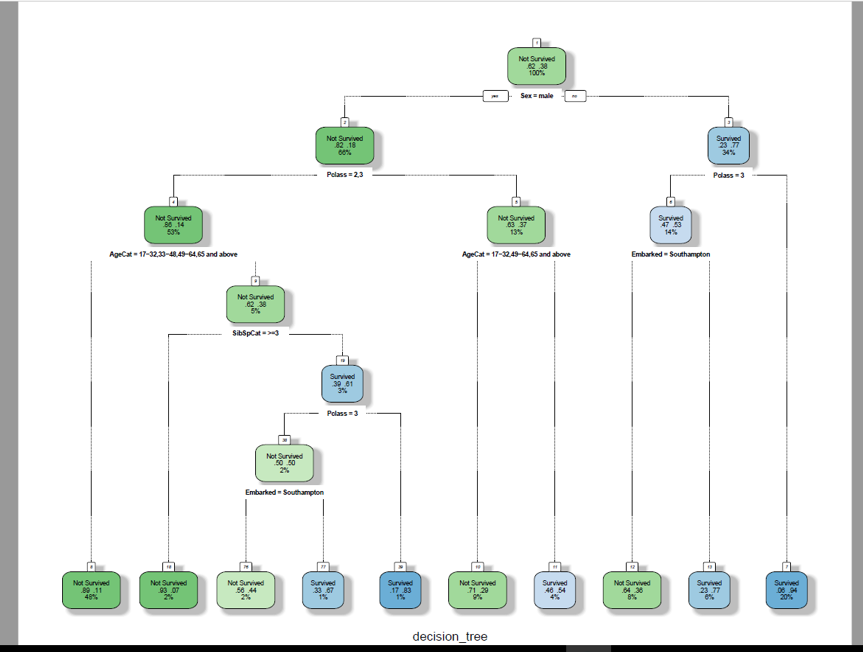
# **Integrated Analysis - Decision Tree and K-means Clustering using Tableau & R**

1. **Decision Tree**



**Misclassification rate for the current tree model (up to 2 decimal places):** **0.2188**

**The first variable used for splitting:** **Sex**

**The ratio of Survived: Not survived initially:** **38: 62**

**Ratio of the Survived: not survived of Females:** **77:23**

**The ratio of Survived: Not Survived of the males who are from Pclass 1:** **37:63**

**Top 6 variables from your decision tree in the order of importance: Sex, Pclass, AgeCat, Embarked, SibSpCat and ParchCat**

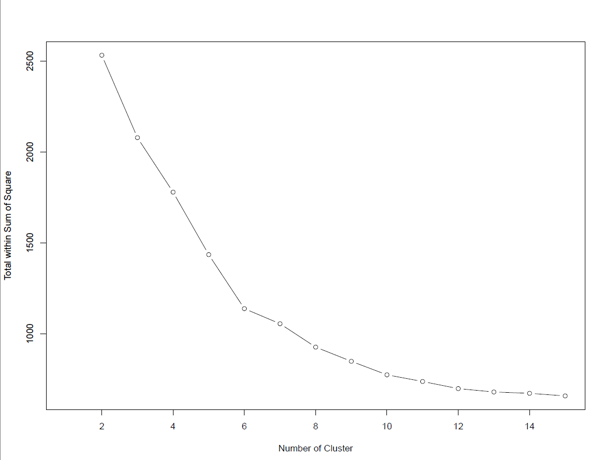
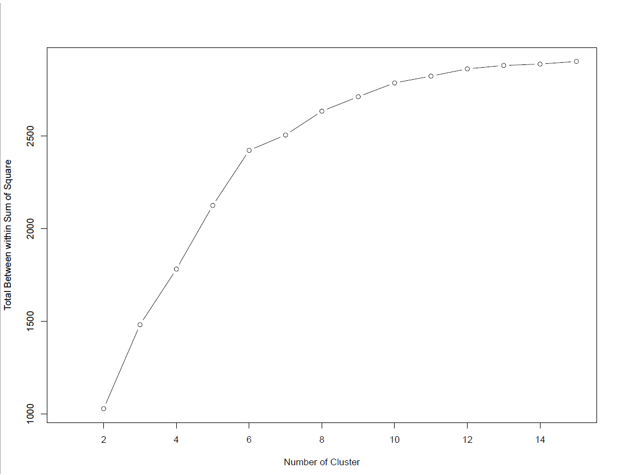
**2) K-means Clustering**

Focusing on the key decision variables to cluster the passengers would be a good strategy

The desired number of cluster selected for the analysis: 6

We need to keep adding clusters to the point where further addition of cluster won’t do much of explanation of the variation. This is also the point where the slope of the curve changes suddenly and gives an angle to the graph.

**Screenshot of the two plot**

Let’s try to analyze this now. To see my visualizations, you need to do the following steps:

**Initiate Rserve and Making Connection**

Type the following code on your R window

install.Packages("Rserve")

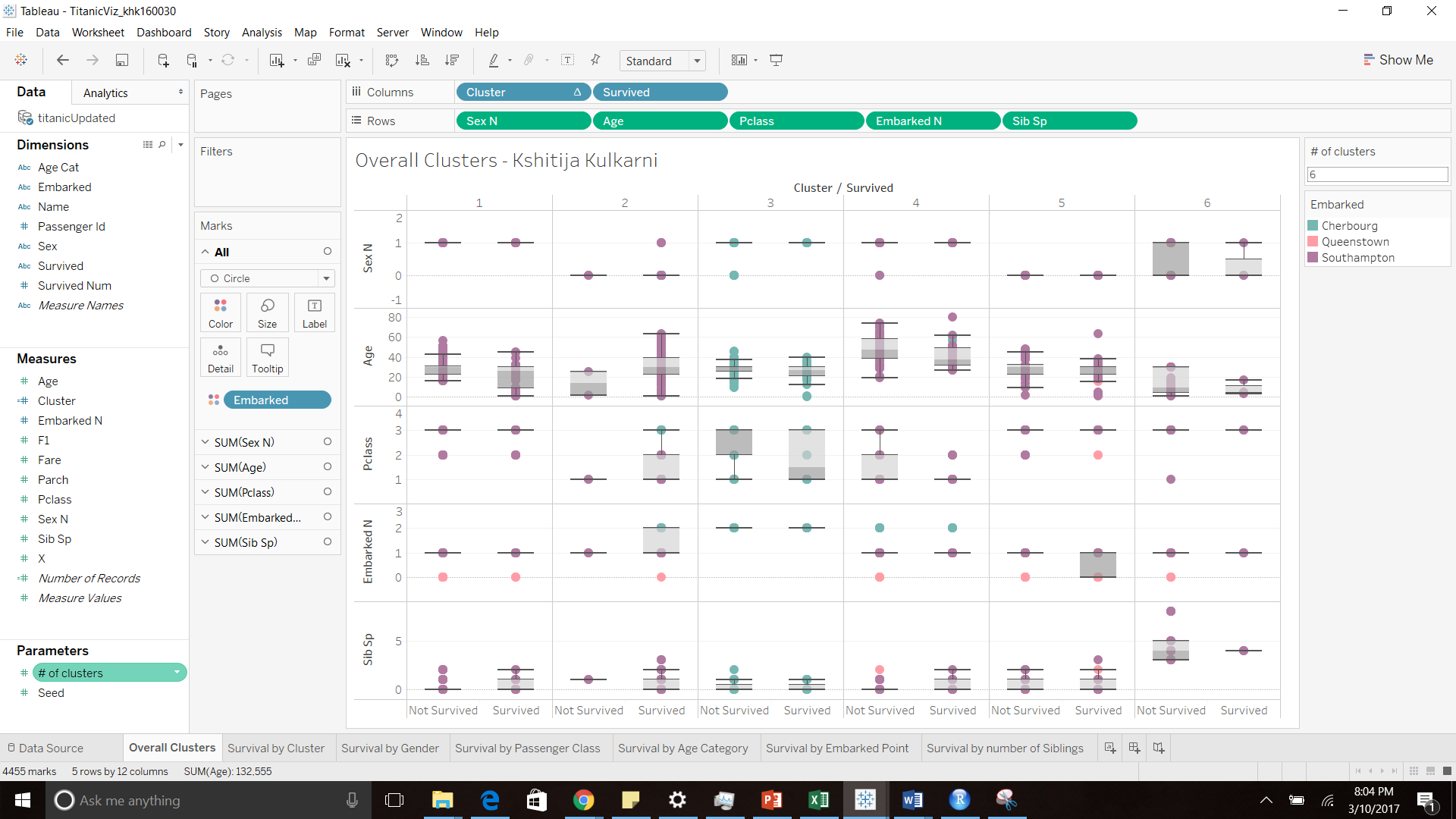
library(Rserve)

Rserve()

In order for us to be able to make a connection between Tableau and R, we need to do it through Rserve. Rserve itself is the server that is a program that responds to requests from clients. It listens for any incoming connections and processes incoming requests.

1. Open Tableau
2. Go to the Help menu and select “Manage R Connection”.
3. Enter a server name of “Localhost” (or “127.0.0.1”) and a port of “6311”.
4. Click on the “Test Connection” button to make sure everything runs smoothly. You should see a successful message. Click OK to close.

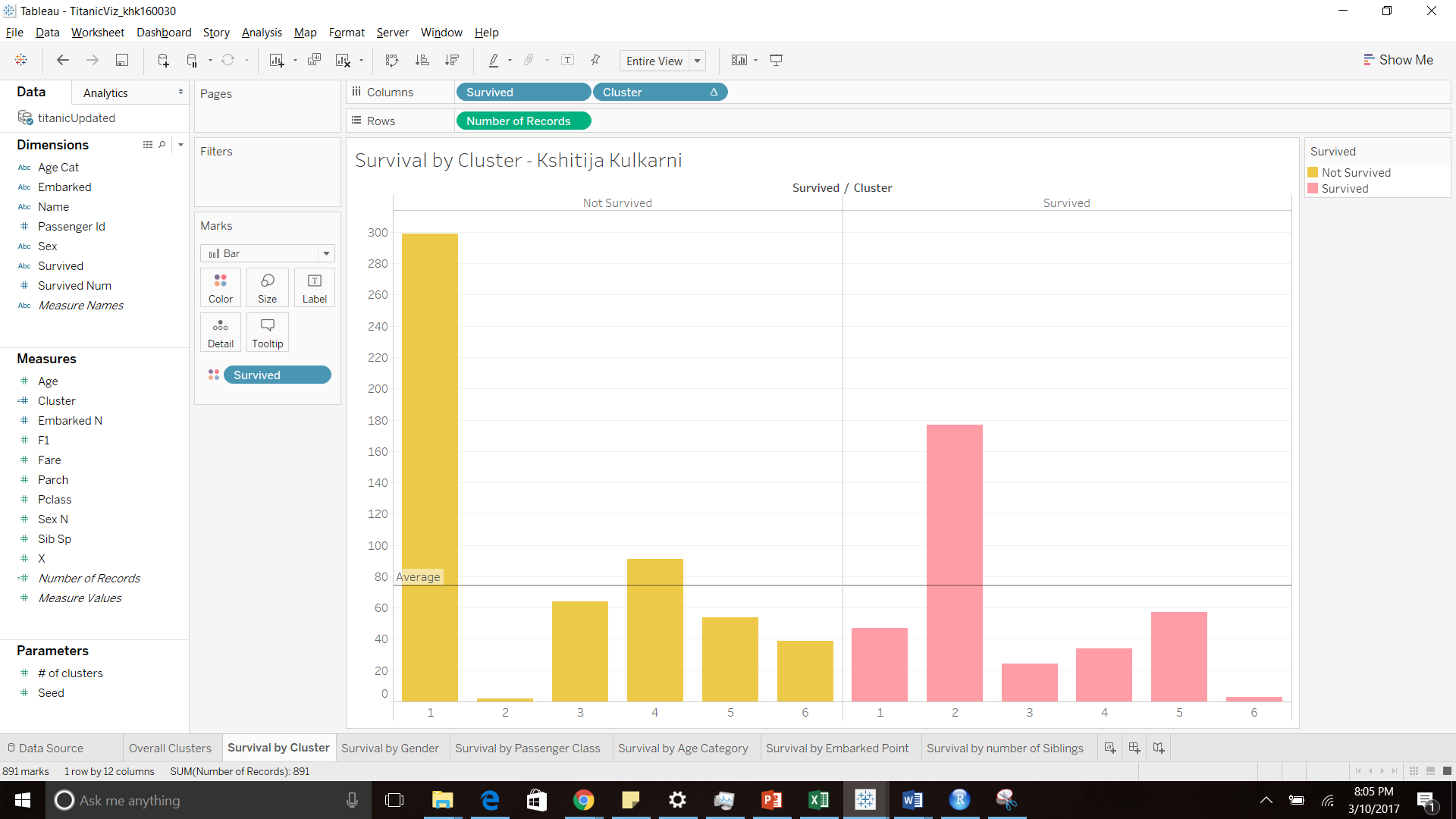
After successful connection of R and Tableau, you can see the following visualizations:

Overall Clusters:  


Cluster 1 insights: Passengers in Cluster 1 are Male belonging to Passenger class 2 and 3 and most of which embarked from Southampton.

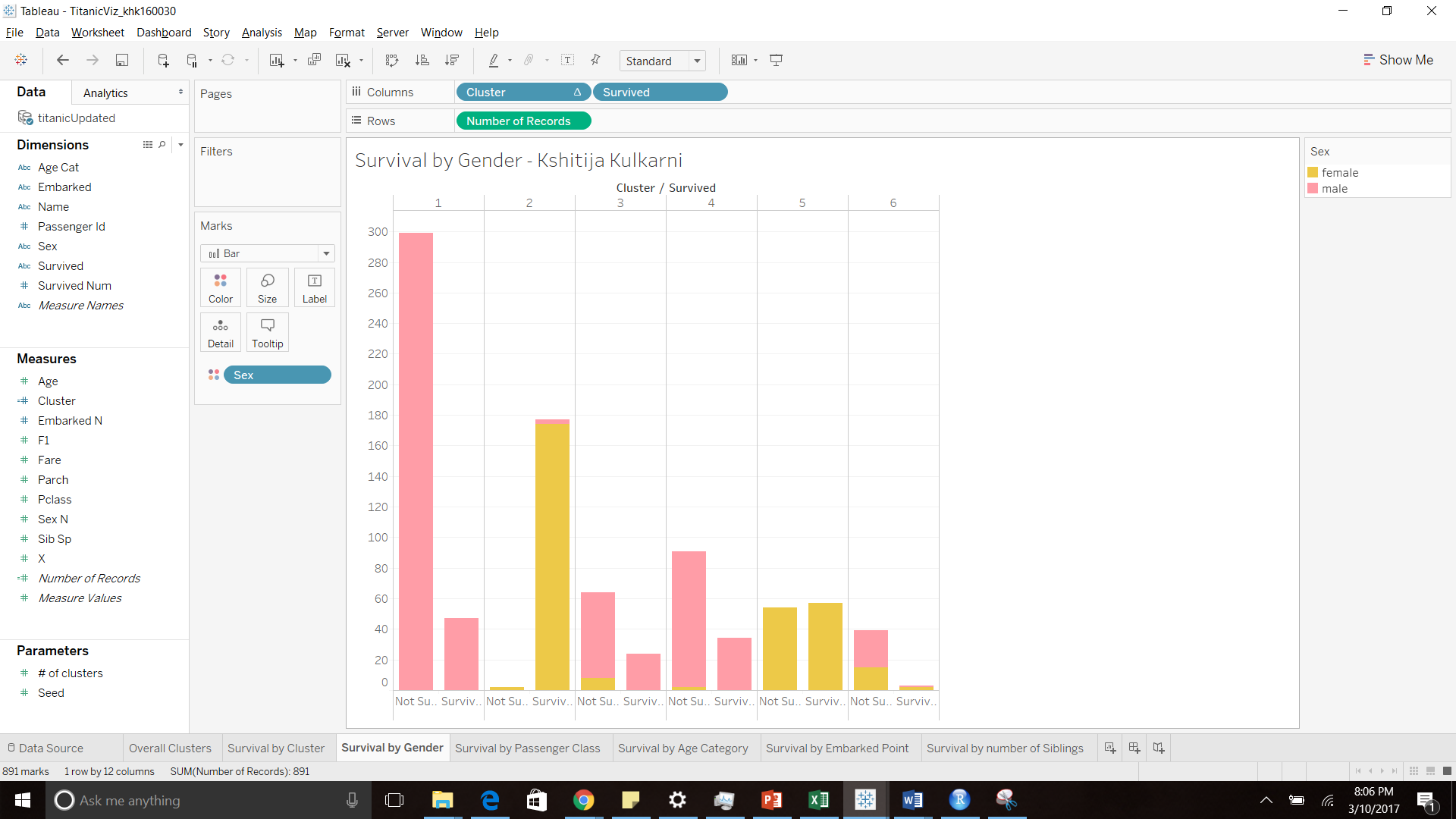
Of which who survived had median age 26 as compared to 30 for those who did not survive. SibSp is mostly 0 and 1 with a few having value 2.

Cluster 4 insights: Those who survived were Male from Passenger class 1 and 2 with median Age 37.5 and Embarked mostly from Southampton and Cherbourg. Large number of passengers did not survive in this category and belonged to Higher Age groups. SibSp is mostly 0 and 1 with a few having value 2.



Cluster 2 and cluster 5 are the two clusters where survivability is the highest.

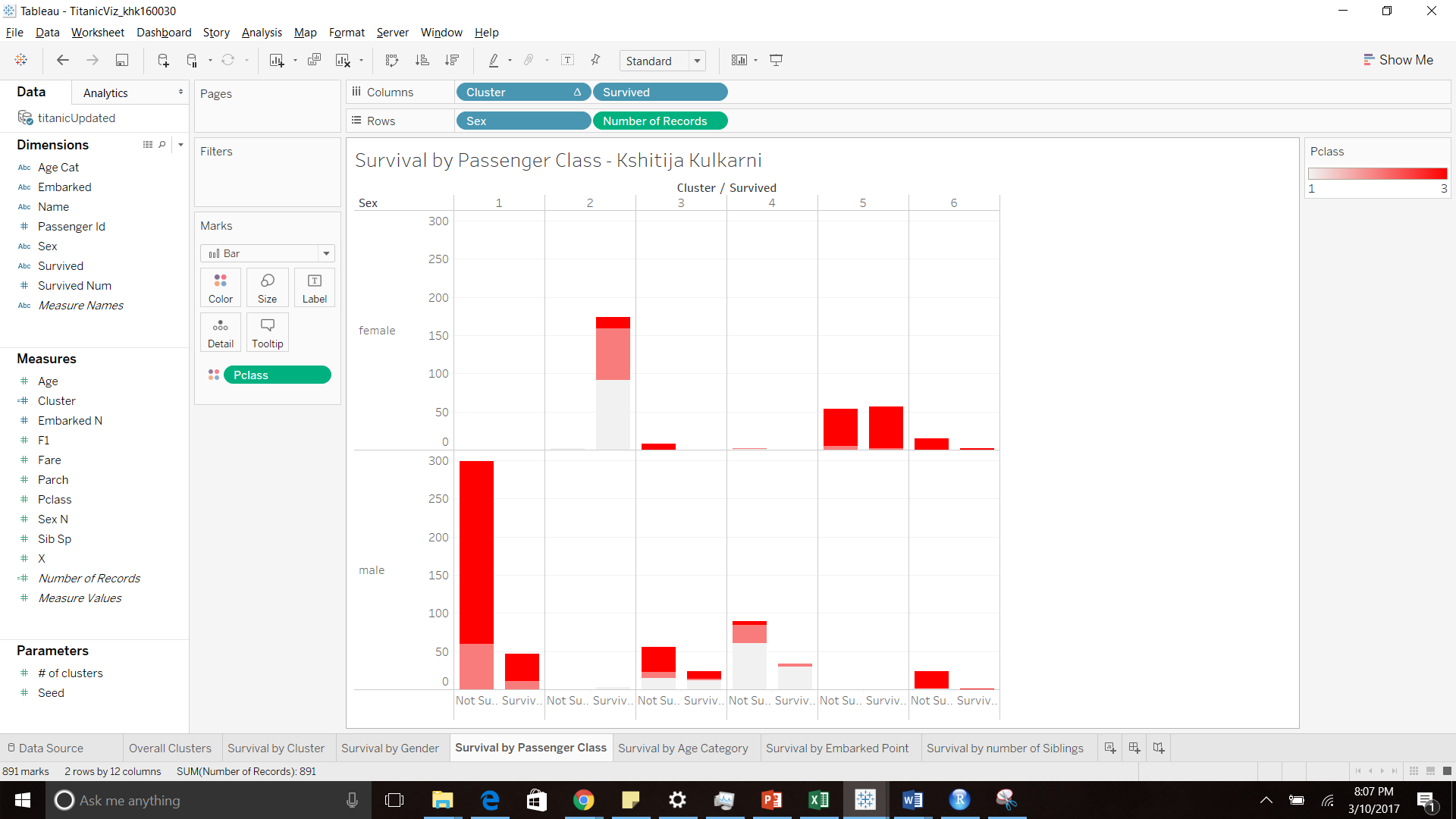
|  |  |  |
| --- | --- | --- |
|  | Cluster 2 | Cluster 5 |
| Ideal Gender | Female | Female |
| Ideal Passenger Class |  |  |
| Ideal Age Category |  |  |
| Ideal Embarked point |  |  |
| Ideal number of siblings |  |  |



If you are Female, to ensure higher probability of survival, you should belong to Cluster 2

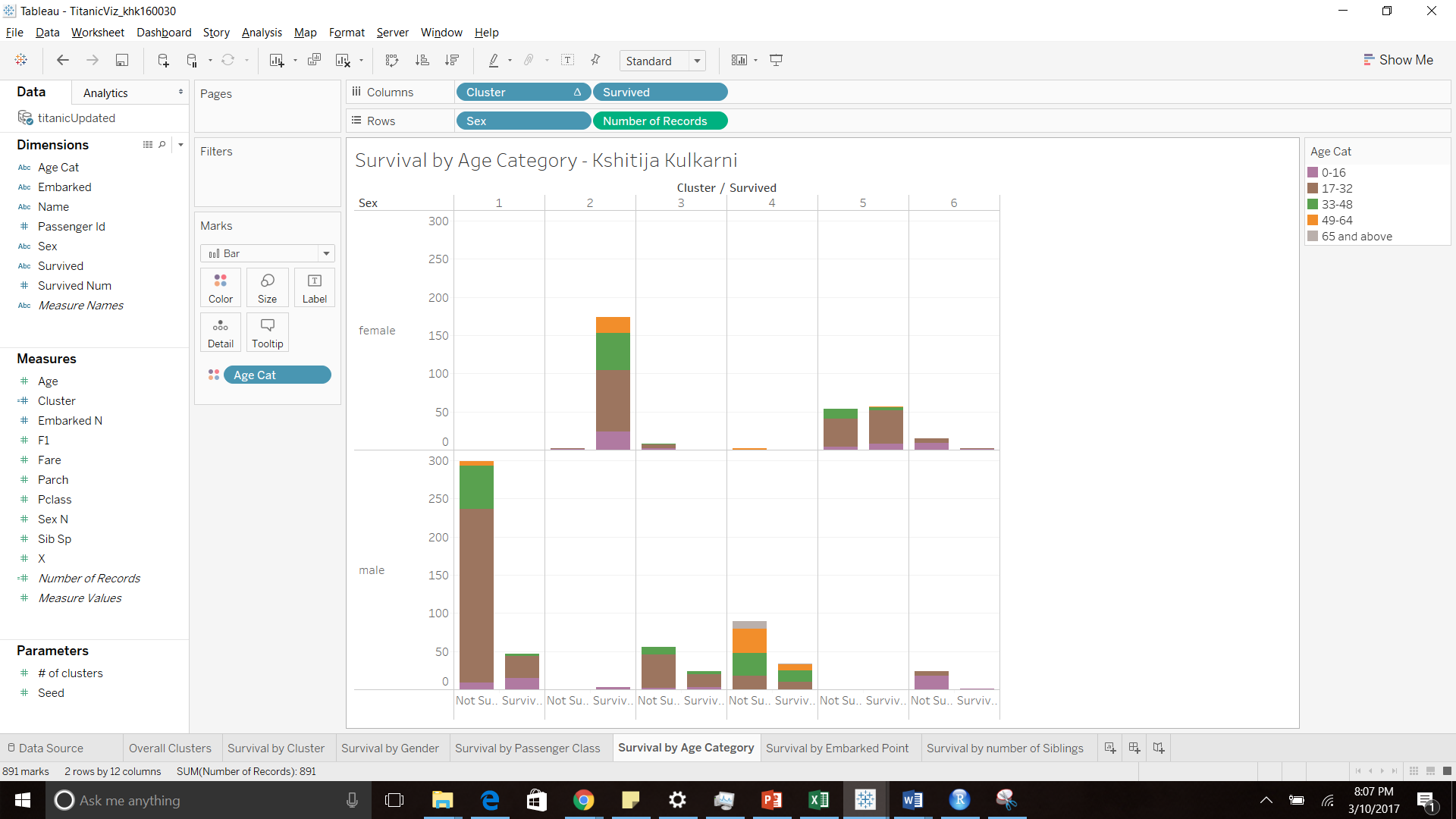
Based on my findings, following is the ideal Gender/Passenger Class (that has the best chance to survive) in top two cluster.

|  |  |  |
| --- | --- | --- |
|  | Cluster 2 | Cluster 5 |
| Ideal Gender | Female | Female |
| Ideal Passenger Class | Class 1 | Class 3 |
| Ideal Age Category |  |  |
| Ideal Embarked point |  |  |
| Ideal number of siblings |  |  |



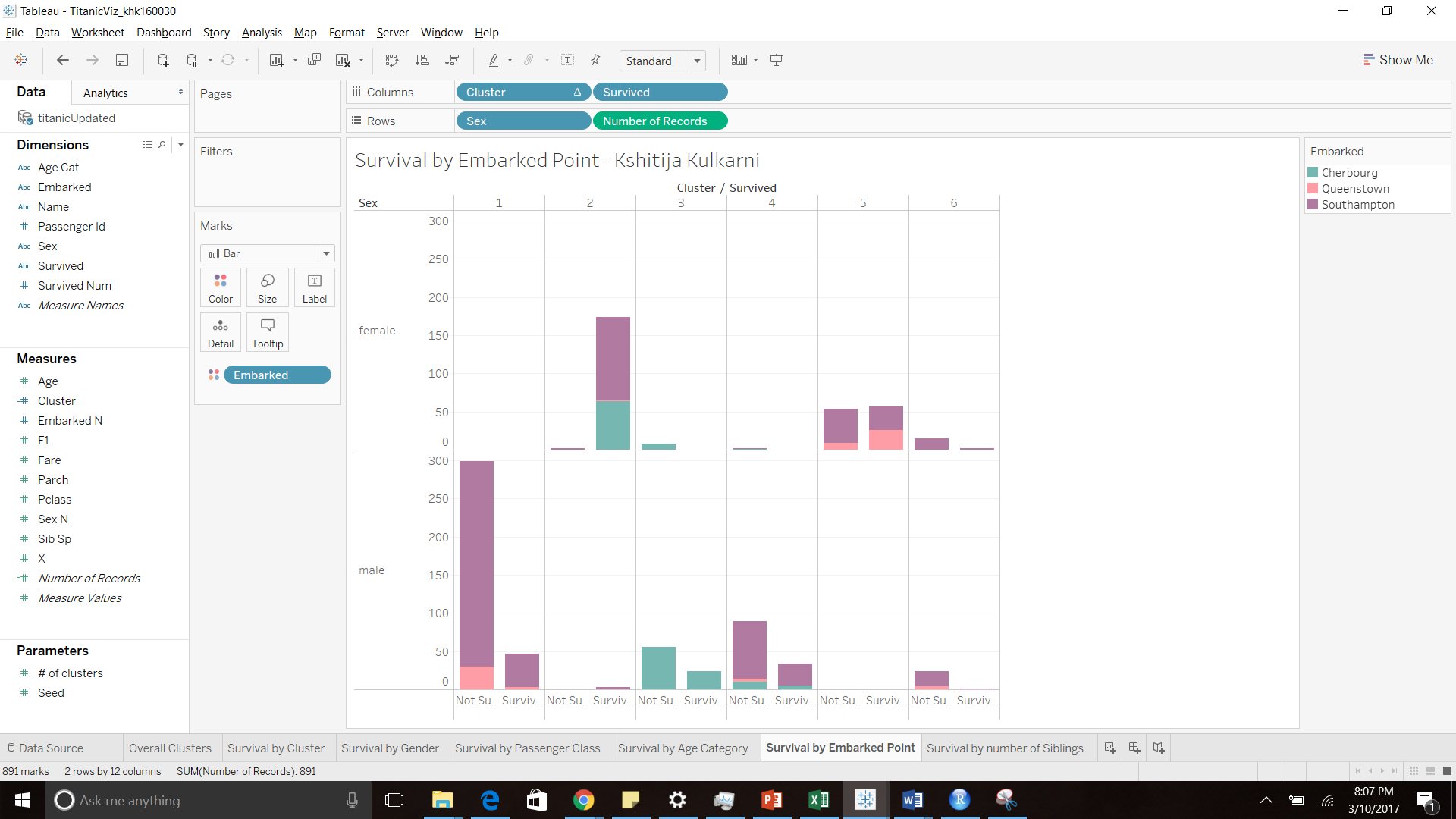
The ideal Gender/Age Category Class (that has the best chance to survive) in top two clusters:

|  |  |  |
| --- | --- | --- |
|  | Cluster 2 | Cluster 5 |
| Ideal Gender | Female | Female |
| Ideal Passenger Class | Class 1 | Class 3 |
| Ideal Age Category | 17-32 | 17-32 |
| Ideal Embarked point |  |  |
| Ideal number of siblings |  |  |



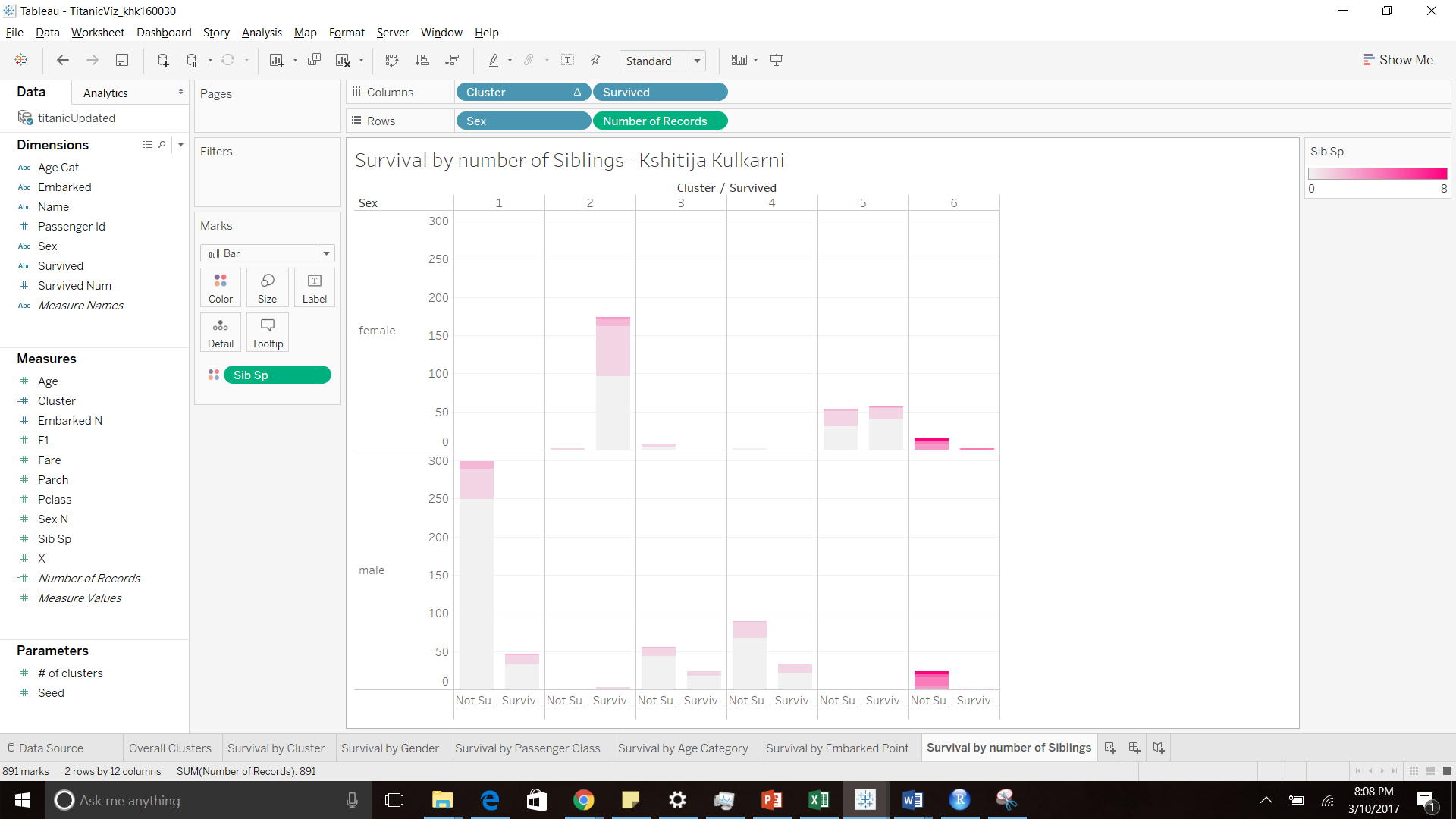
The ideal Gender/Embarked point (that has the best chance to survive) in top two clusters:

|  |  |  |
| --- | --- | --- |
|  | Cluster 2 | Cluster 5 |
| Ideal Gender | Female | Female |
| Ideal Passenger Class | Class 1 | Class 3 |
| Ideal Age Category | 17-32 | 17-32 |
| Ideal Embarked point | Southampton | Southampton |
| Ideal number of siblings |  |  |



The ideal Gender/# of Sib Sp (that has the best chance to survive) in top two clusters:

|  |  |  |
| --- | --- | --- |
|  | Cluster 2 | Cluster 5 |
| Ideal Gender | Female | Female |
| Ideal Passenger Class | Class 1 | Class 3 |
| Ideal Age Category | 17-32 | 17-32 |
| Ideal Embarked point | Southampton | Southampton |
| Ideal number of siblings | 0 | 0 |



Cluster 2 has chance of survival is dominated by Female from Class 1, 17 - 32 years of age who Embarked from Southampton and had no sibling. Cluster 5 has chance of survival in this cluster is better for Female from Class 3 ,17 - 32 years of age who Embarked from Southampton and had no sibling.