# Exercise 7: Integrated Analysis - Decision Tree and K-means Clustering using Tableau & R

---

**Student Name: Ravi Dawar**

**Date: 3/2/2018**

## 1) Data Retrieval

Q9) Write the output of the command: dim(titanic)

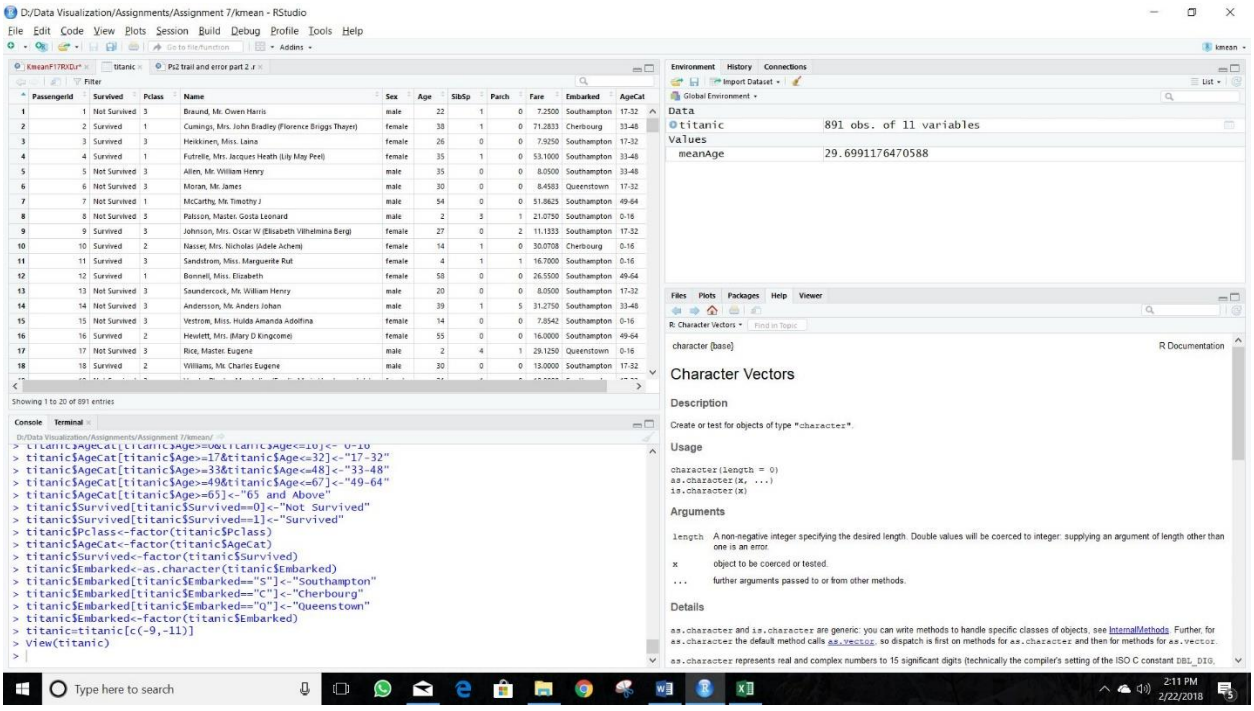
Answer: 891 12 which means 891 Rows and 12 Columns

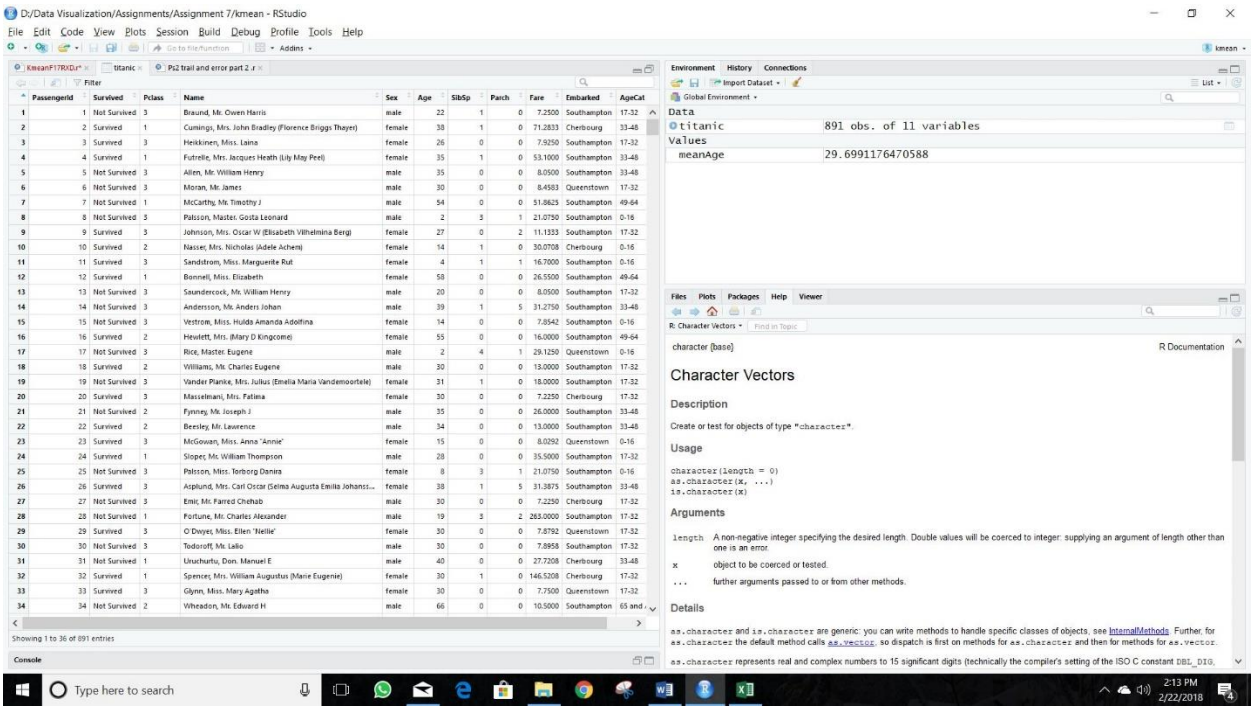
## 2) Data Pre-processing

Q2) What is the average age that you get?

Answer 2: 29.69912

Answer 3:

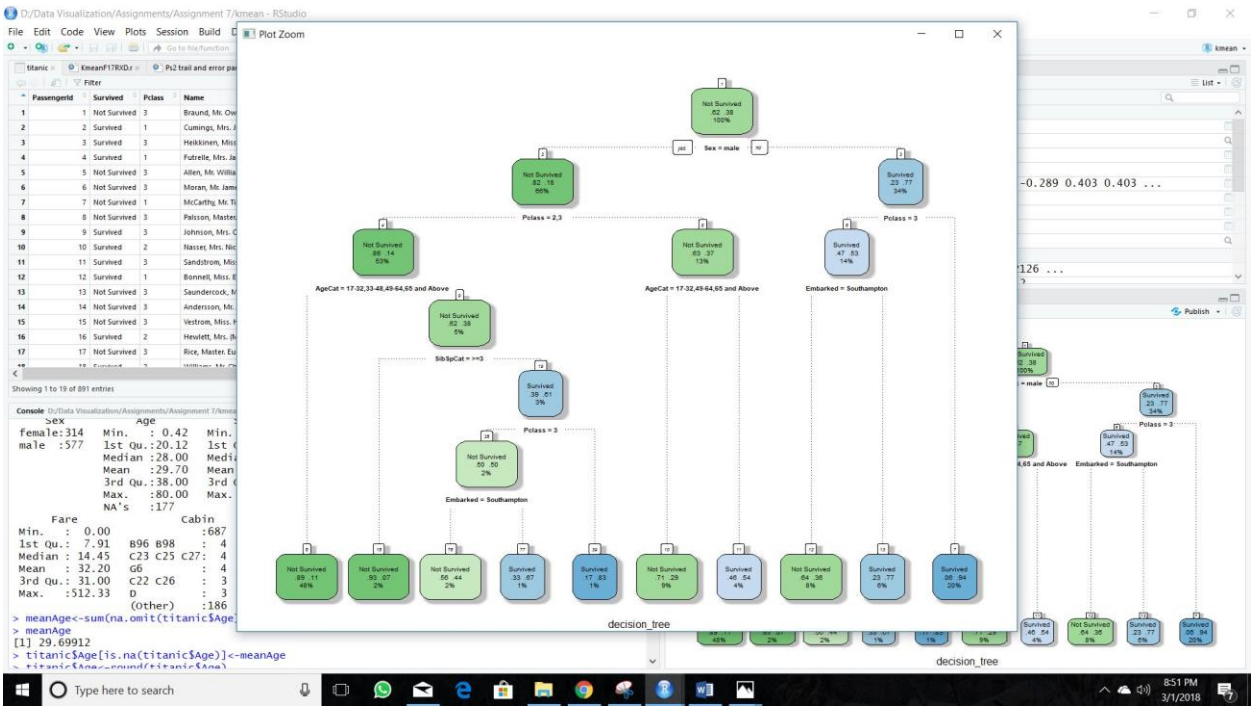





# 3) Decision Tree

## Q4) Screenshot of decision Tree is below

**Answer:**



Question:-

1)What is the misclassification rate for the current tree model(up to 2 decimal places)?

Answer: 0.21

2)Which is the first variable used for splitting?

Answer: Sex

3)What is the ratio of Survived: Not survived initially?

Answer: 38:62

4)What is the ratio of the Survived: not survived of Females?

Answer: 77:23

5)What is the ratio of Survived: Not Survived of the males who are from Pclass 1 ?

Answer: 37:63

6) Please list the top 6 variables from your decision tree in the order of importance

Answer:
1. Sex
2. PClass
3. AgeCat
4. Embarked
5. SibSpcat
6. ParchCat

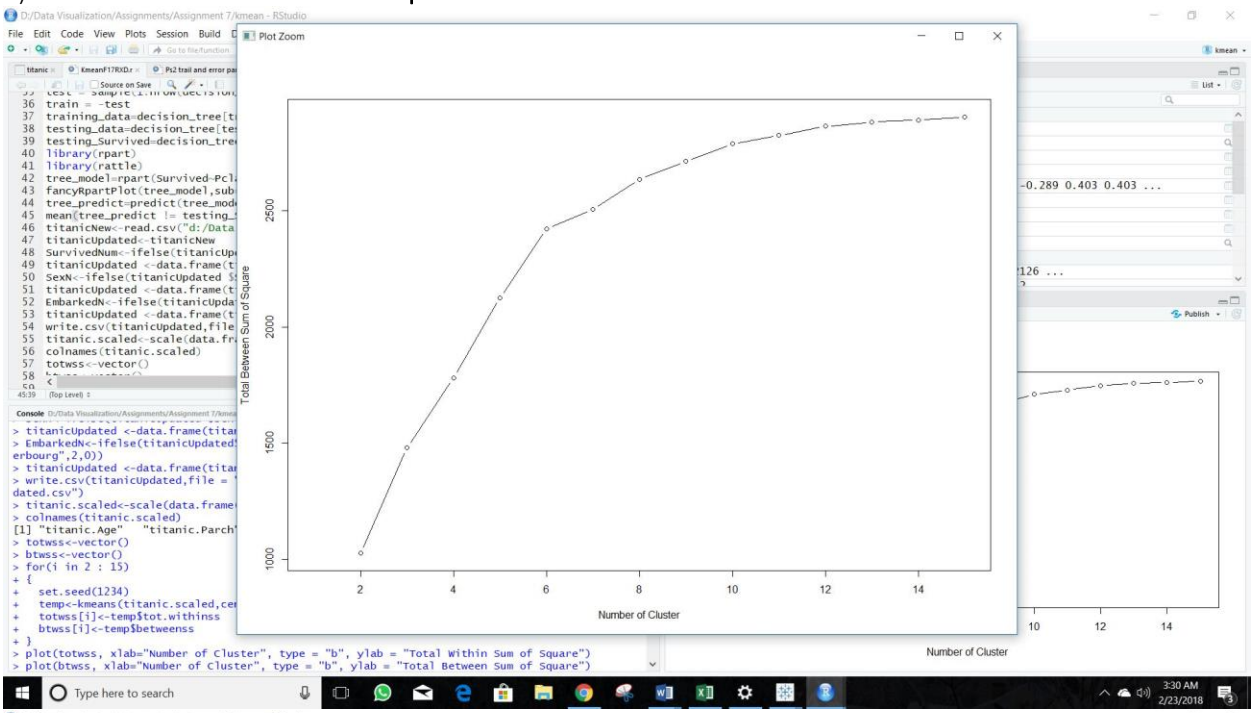
# 4) K-means Clustering

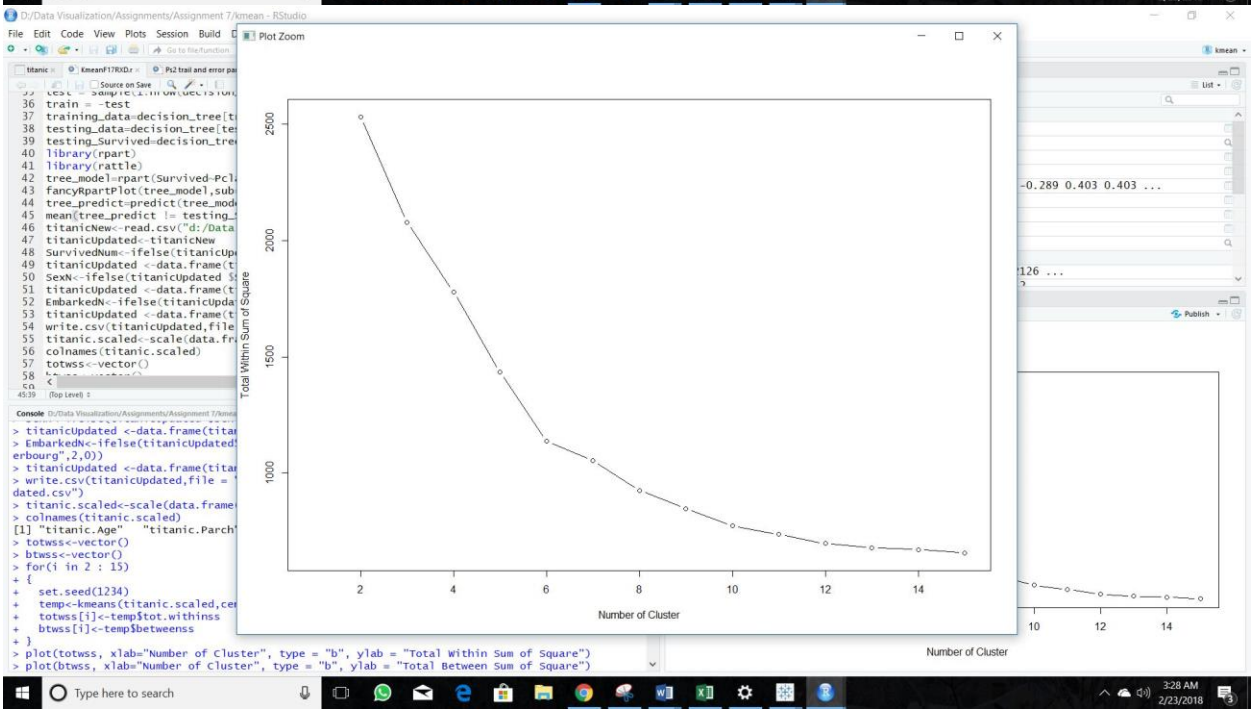
Q1 ) Based on the decision tree from Part 3) of the exercise, do you feel that using all the variables to cluster the passengers would be a good strategy or focusing on the key decision variables (as obtained in Section 3, Answer 6) would be a good approach? Select one of the two choices below.

Answer 1) Considering all the key variables, would be a good approach for clustering passengers.

3) Write the desired number of cluster you selected for the analysis
Answer: 6

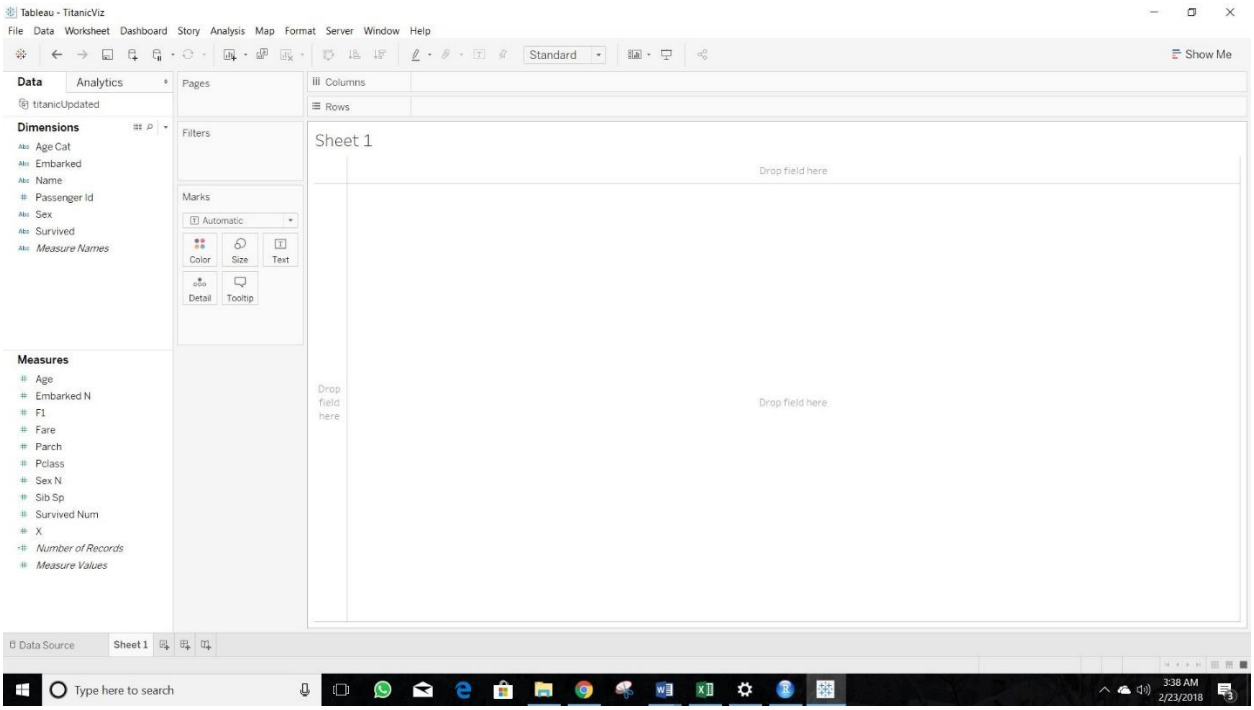
4) Paste the screenshot of the two plot



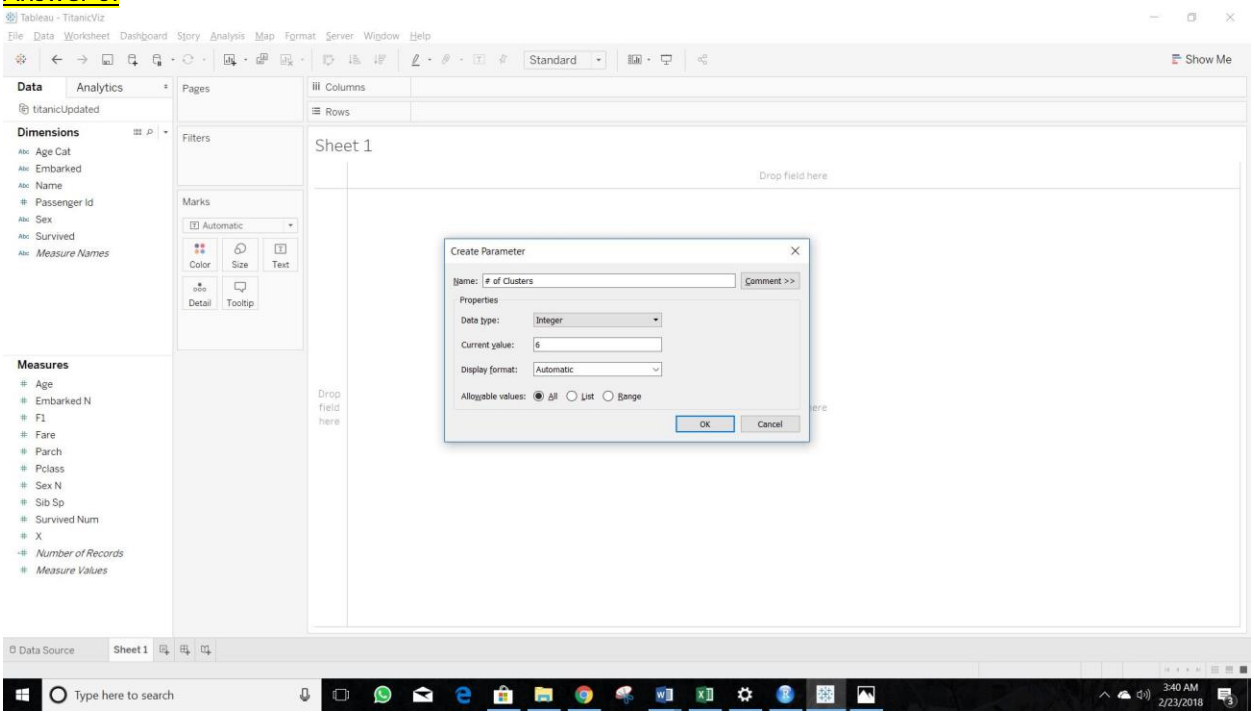



# 5) TABLEAU / R INTEGRATION

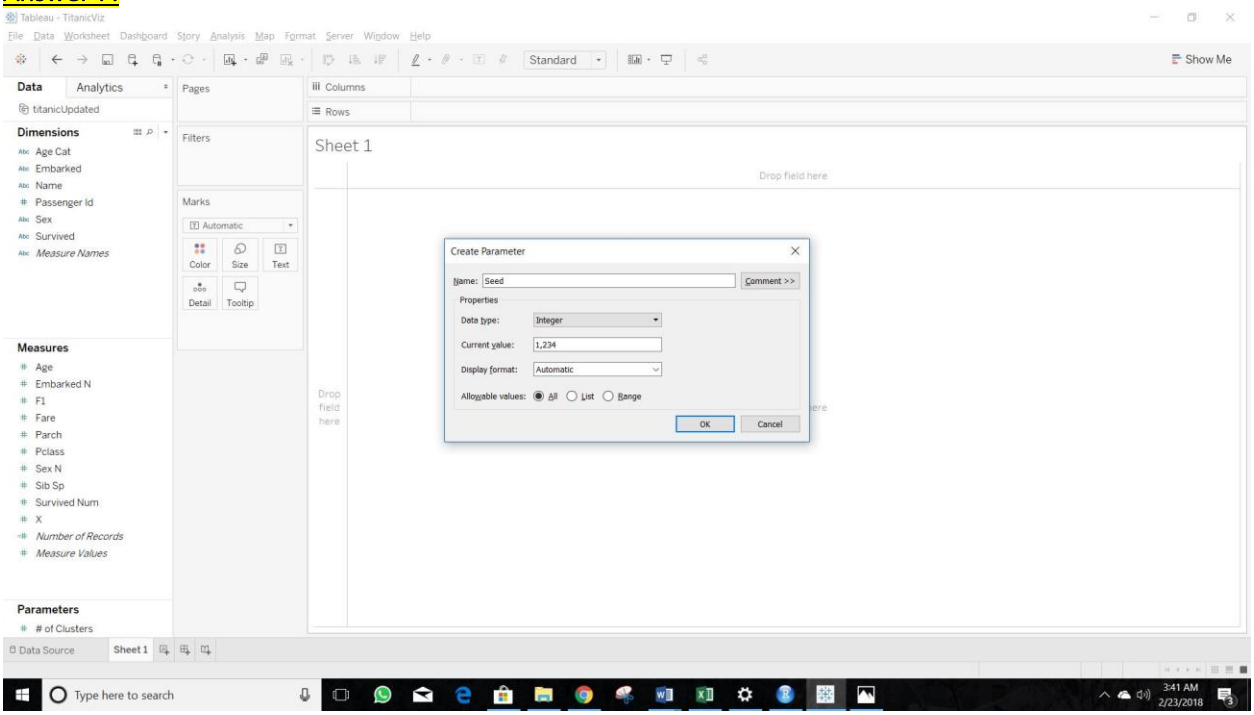
**Answer 5:**



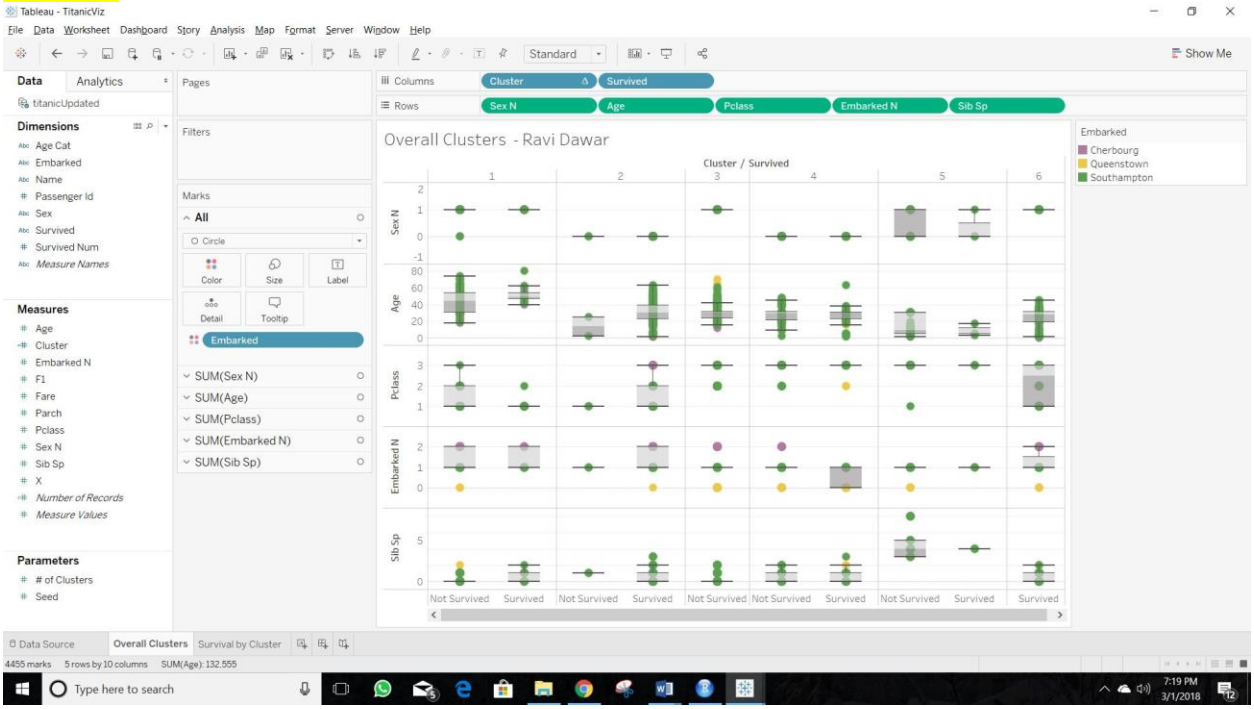

**Answer 6:**



## Answer 7:




## Answer 8:



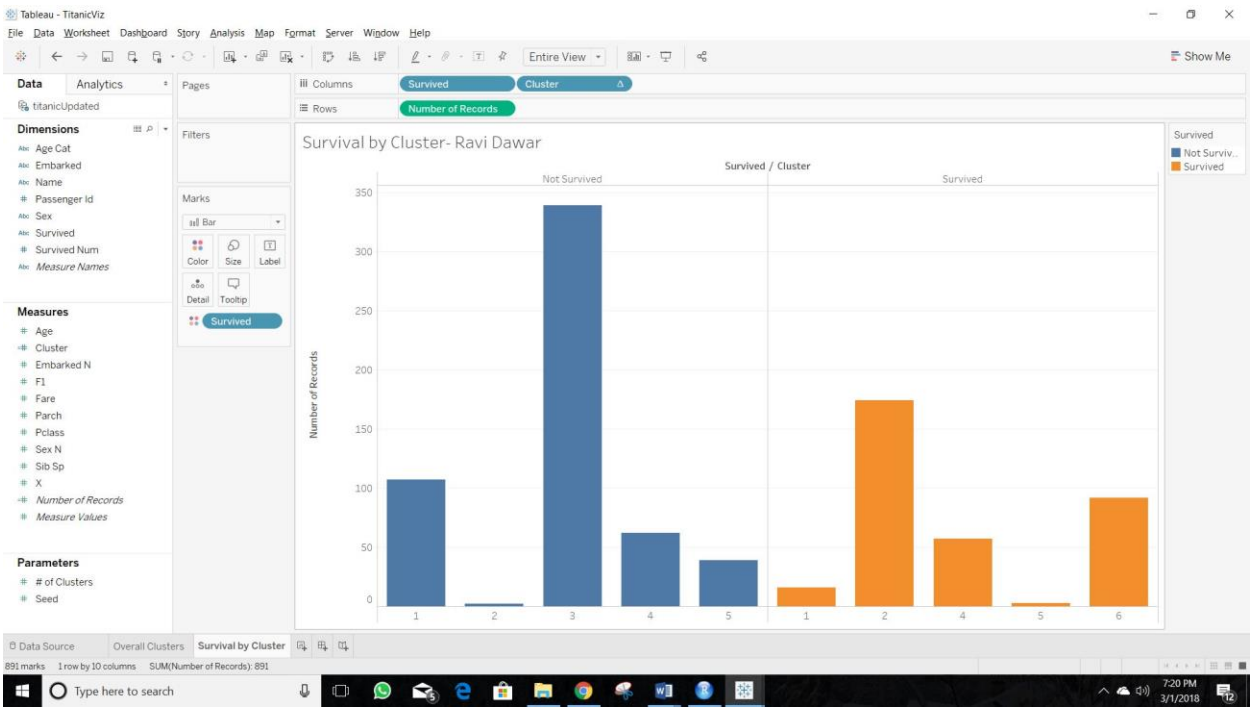
## Answer 9:

On analyzing Cluster 1 we can infer that all male passenger survived. No passenger survived below age 40 and above age 18 and those belonging from Pclass 3. It has passengers who embarked for all three Cities but passengers who embarked for Queenstown did not survived. Maximum number of siblings of passengers are 2 in this clusters

## Answer 10:

On analyzing Cluster 4 we can infer that there are only female passengers. Passengers age range is from 0 to 48 with outlier 1 and 63. No passenger survived below age 15 and above age 2. There was no passenger in Pclass 1. It has passenger who embarked for all three cities but passenger who embarked for Cherbourg did not survived. Maximum number of siblings of passenger who survived are 3 in this clusters.

## Answer 11:




## Answer 12:

Cluster 2 and Cluster 6 have highest survivability.

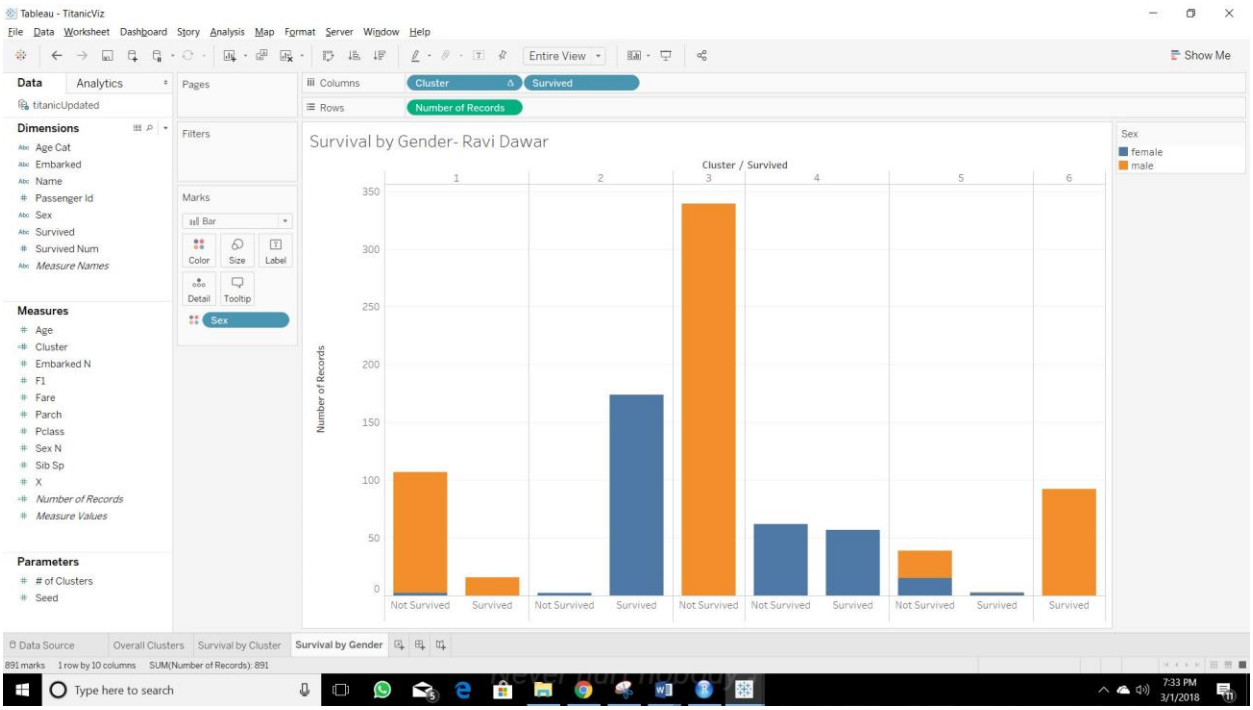
## Answer 13:

| | Cluster 2 | Cluster 6 |
|---|---|---|
| Ideal Gender | | |
| Ideal Passenger Class | | |
| Ideal Age Category | | |
| Ideal Embarked point | | |
| Ideal number of siblings | | |

**Answer 14:**

| | Cluster 2 | Cluster 6 |
|---|---|---|
| Ideal Gender | Female | Male |
| Ideal Passenger Class | | |
| Ideal Age Category | | |
| Ideal Embarked point | | |
| Ideal number of siblings | | |

**Answer 15 :**

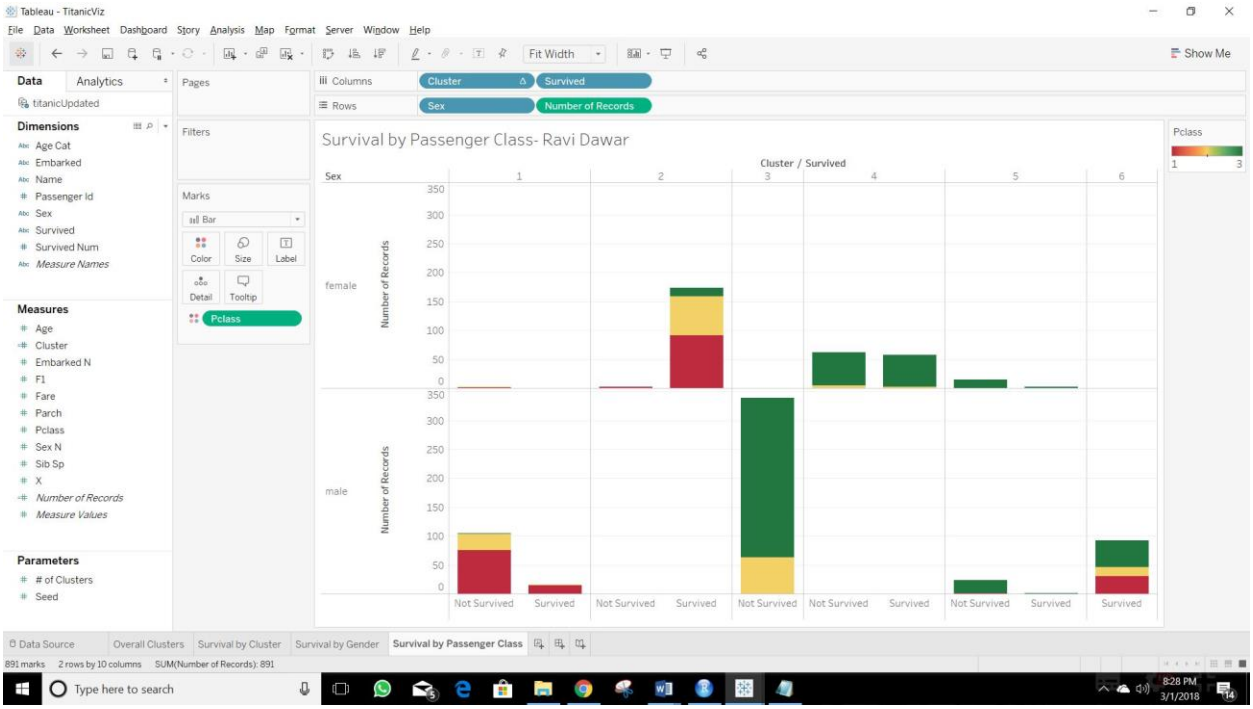



**Answer 16**: Cluster 2

**Answer 17:**

| | Cluster 2 | Cluster 6 |
|---|---|---|
| Ideal Gender | **Female** | **Male** |
| Ideal Passenger Class | **PClass 1** | **PClass 3** |
| Ideal Age Category | | |
| Ideal Embarked point | | |
| Ideal number of siblings | | |

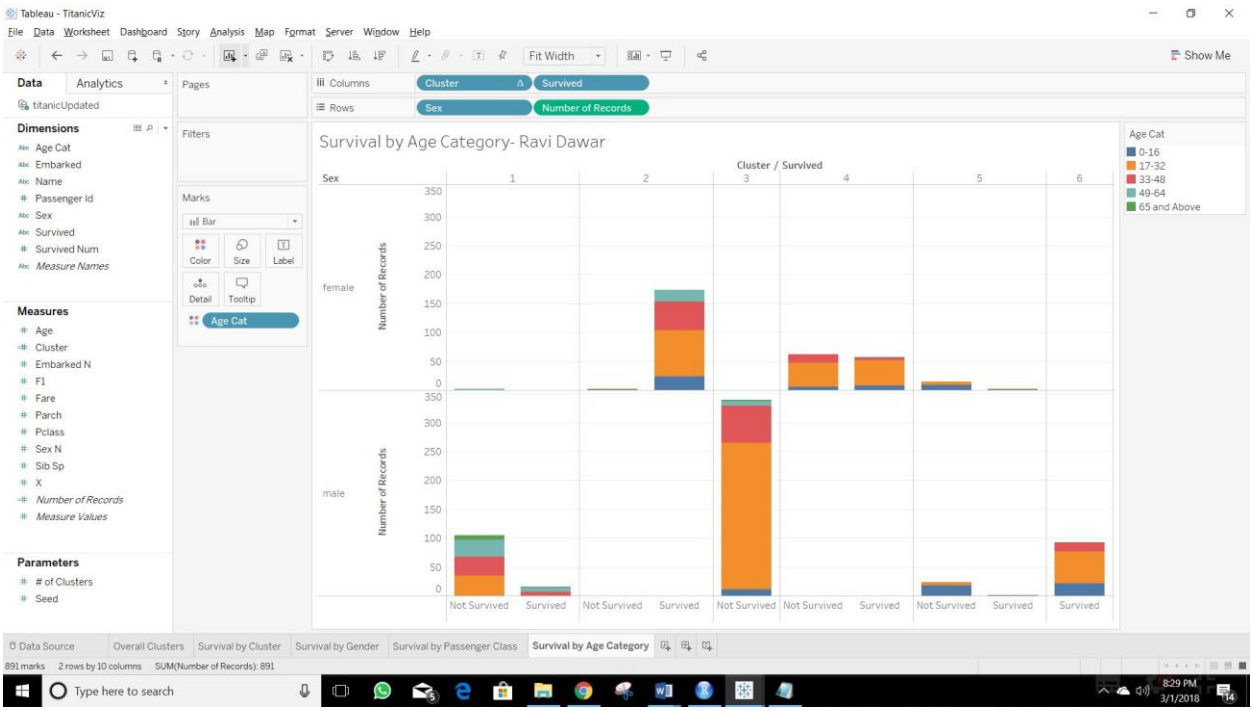
## Answer 18:




## Answer 19:

| | Cluster 2 | Cluster 6 |
|---|---|---|
| Ideal Gender | **Female** | **Male** |
| Ideal Passenger Class | | |
| Ideal Age Category | **17-32** | **17-32** |
| Ideal Embarked point | | |
| Ideal number of siblings | | |

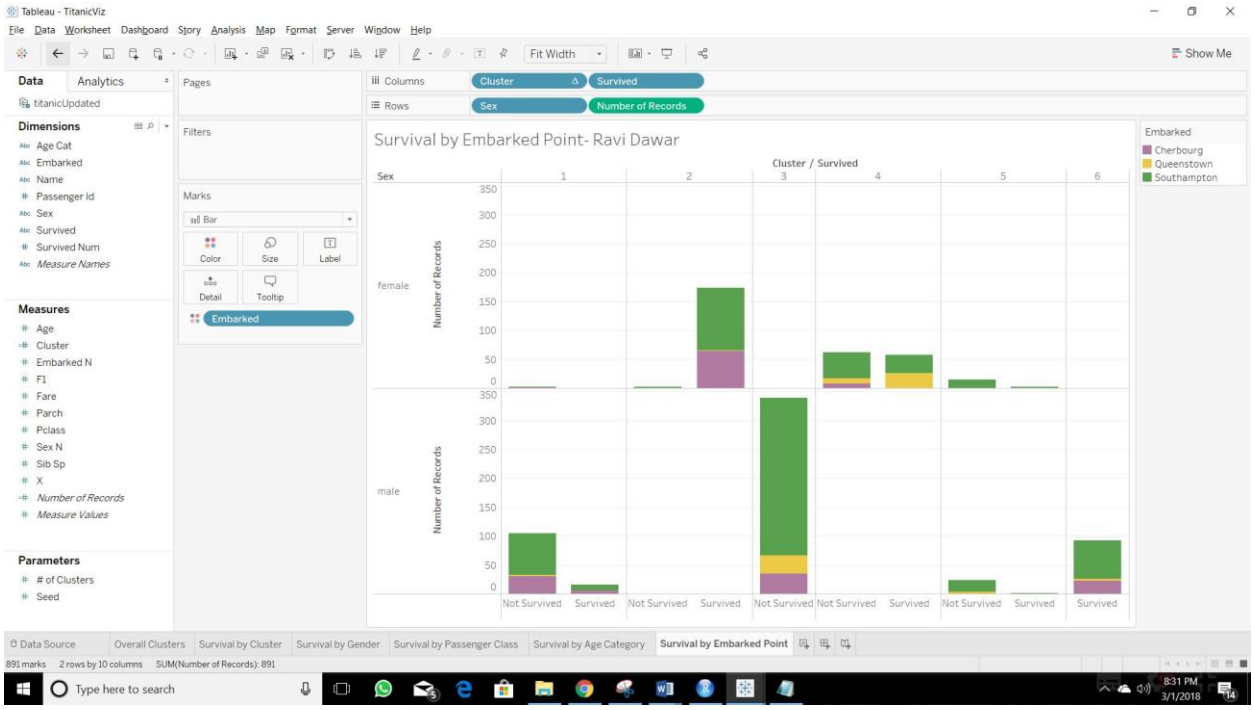
## Answer 20:




## Answer 21 :

| | Cluster 2 | Cluster 6 |
|---|---|---|
| Ideal Gender | Female | Male |
| Ideal Passenger Class | | |
| Ideal Age Category | | |
| Ideal Embarked point | Southampton | Southampton |
| Ideal number of siblings | | |

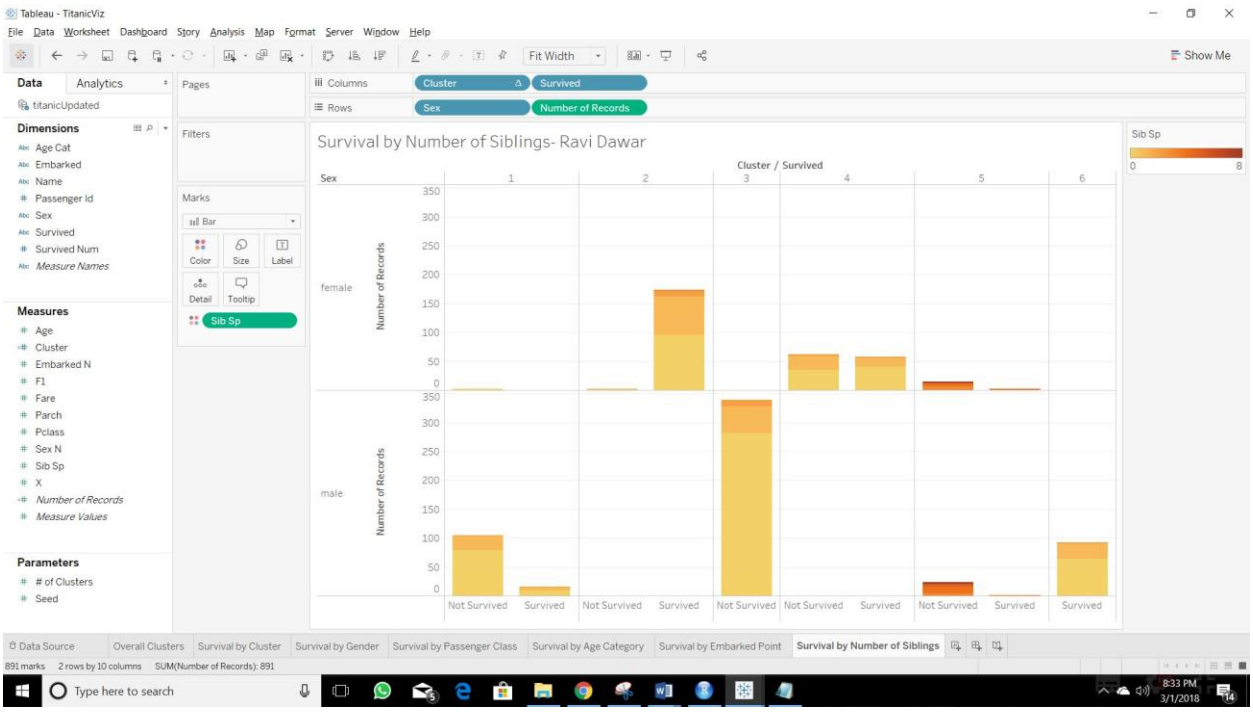
## Answer 22:




## Answer 23:

| | Cluster 2 | Cluster 6 |
|---|---|---|
| Ideal Gender | Female | Male |
| Ideal Passenger Class | | |
| Ideal Age Category | | |
| Ideal Embarked point | | |
| Ideal number of siblings | 0 | 0 |

## Answer 24 :




## Answer 22:

Top two clusters which have the best chance of survivability in comparison to other cluster are:

Cluster 2: Ideal passenger profile is Female passenger which belongs to age range between 17 to 32, travelling in Pclass 1 who embarked from Southampton with no siblings.

Cluster 6: Ideal passenger profile is Male passenger which belongs to age range between 17 to 32, travelling in Pclass 3 who embarked from Southampton with no siblings.