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

**Student Name: Pulkit Agrawal Student Id: PXA180017**

**Date: 10-19-2018**

**1) Data Retrieval**

9)Write the output of the above executed command.

Answer: 891 rows 12 columns

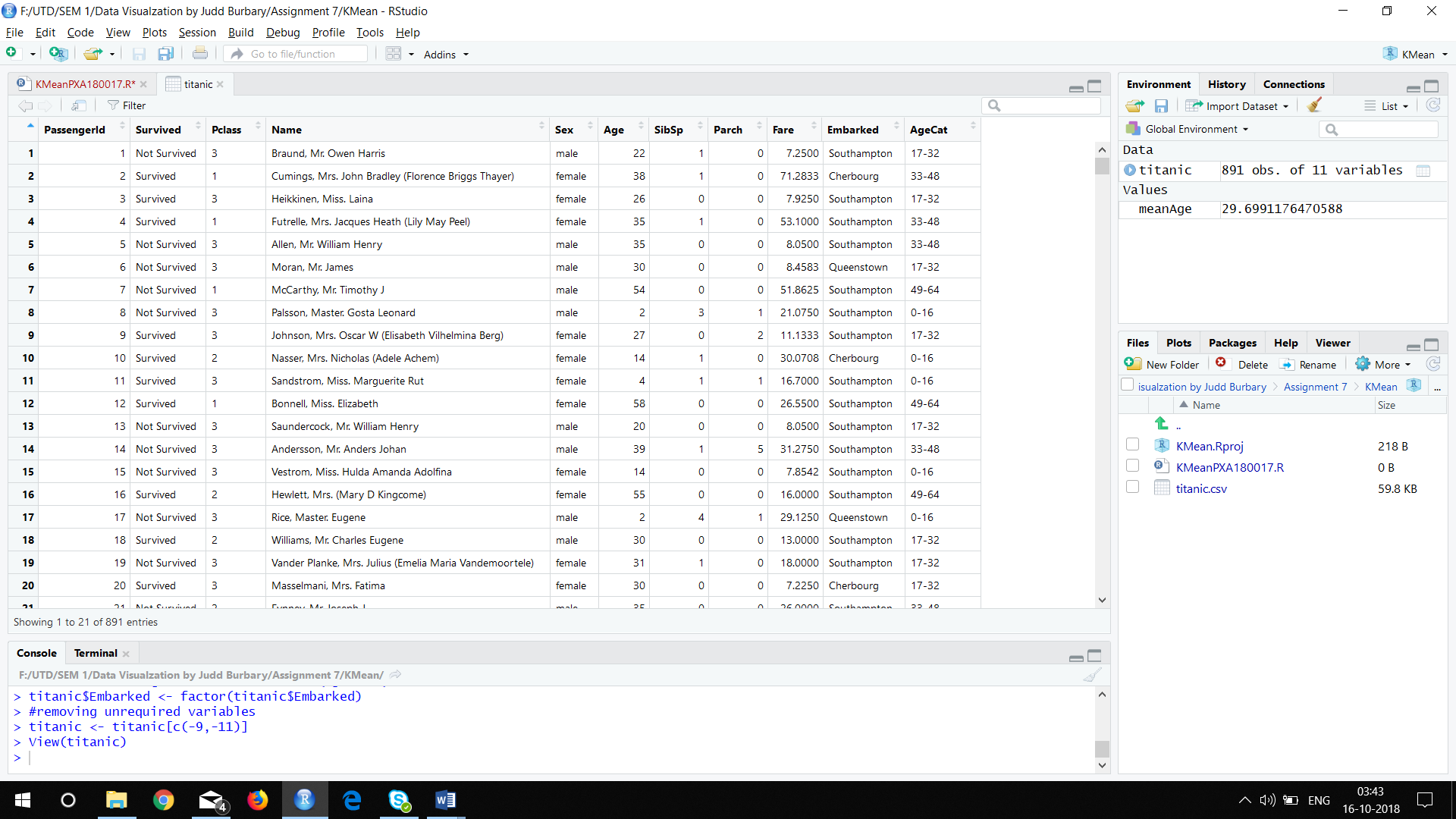
**2) Data Pre-processing**

2) What is the average age that you get?

Answer: 29.69912

9) **Paste the screenshot** of the above executed command.

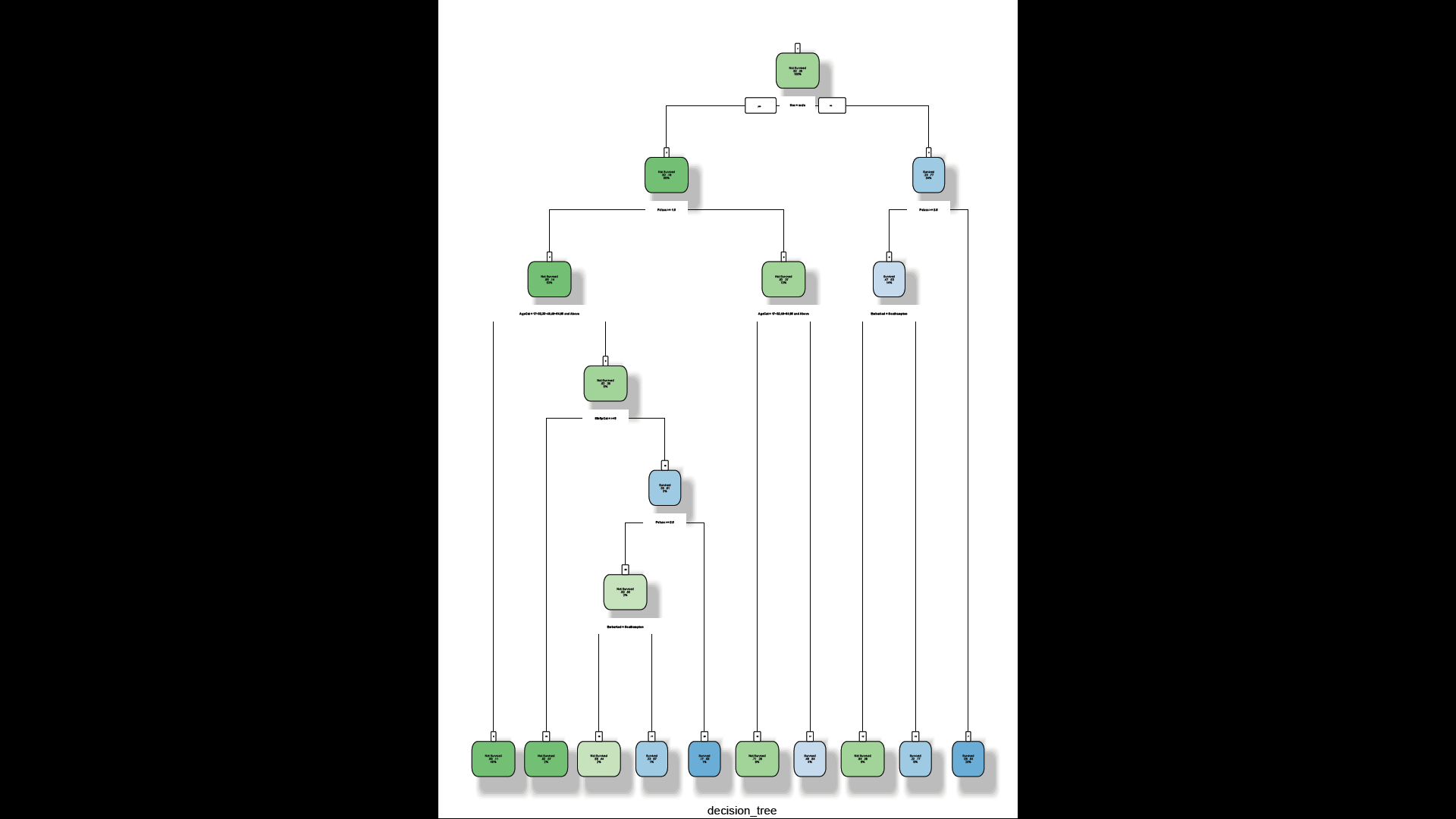
Answer:



**3) Decision Tree**

Export the tree plot and save it as a PDF with a name **decision\_tree\_Plot**. Paste the screenshot of the decision tree below. Note that the screenshot should be readable.

Answer:



**Question:**

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

Answer: 0.22

**2)Which is the first variable used for splitting?**

Answer: Survived

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

Answer: 38:62 = 19:31

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

Answer: 23:77

**5)What is the ratio of Survived: Not Survived of the males who are from Pclass1?**

Answer: 37:63

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

Answer: 1. Survived/not survived

2. Sex

3. Pclass

4. AgeCategory

5. Embarked

6. SibSpCat

**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.

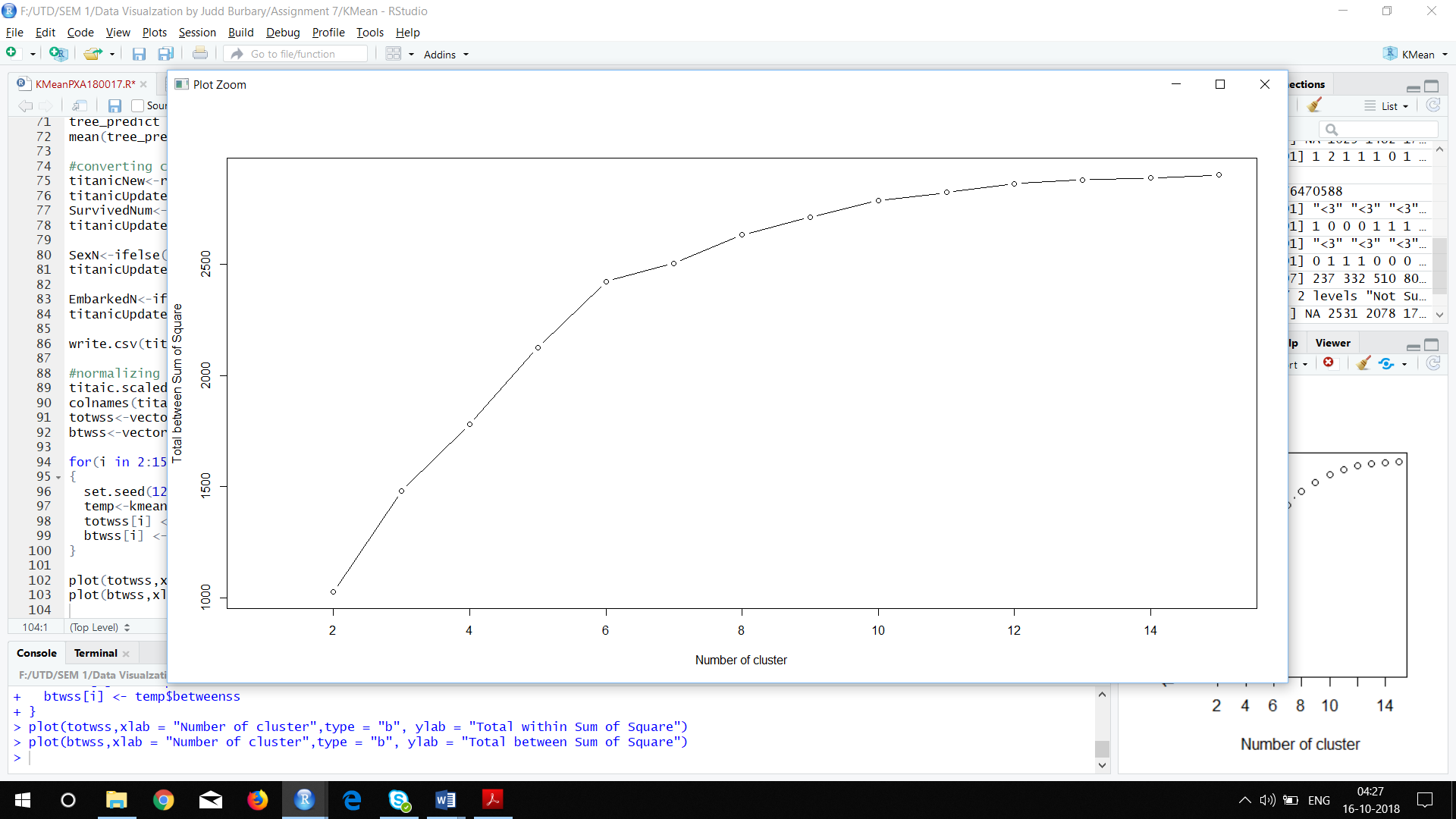
Answer: Key Variables.

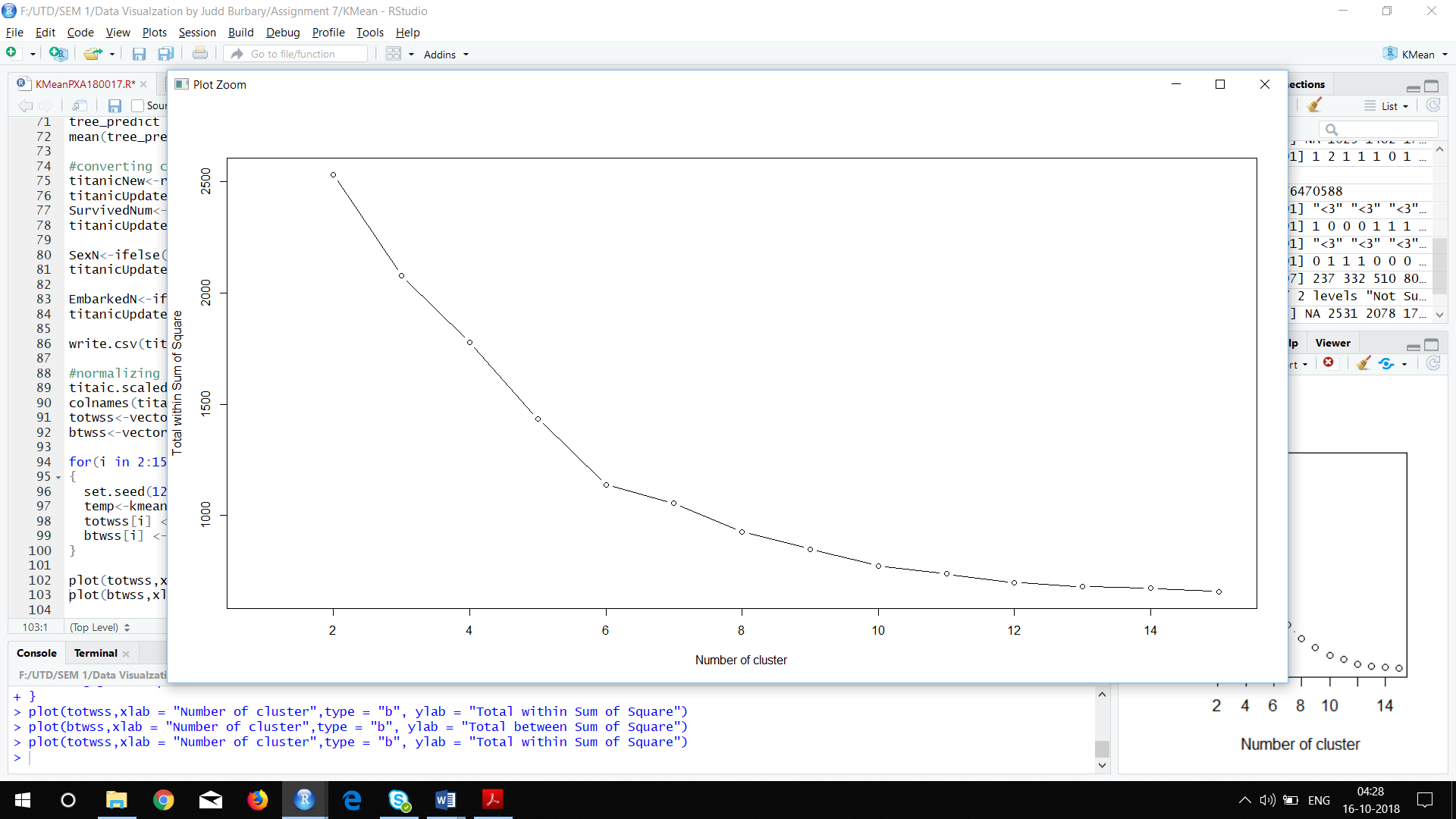
3) Write the desired number of cluster you selected for the analysis

Answer: 6

4) **Paste the screenshot of the two plots**

Answer:

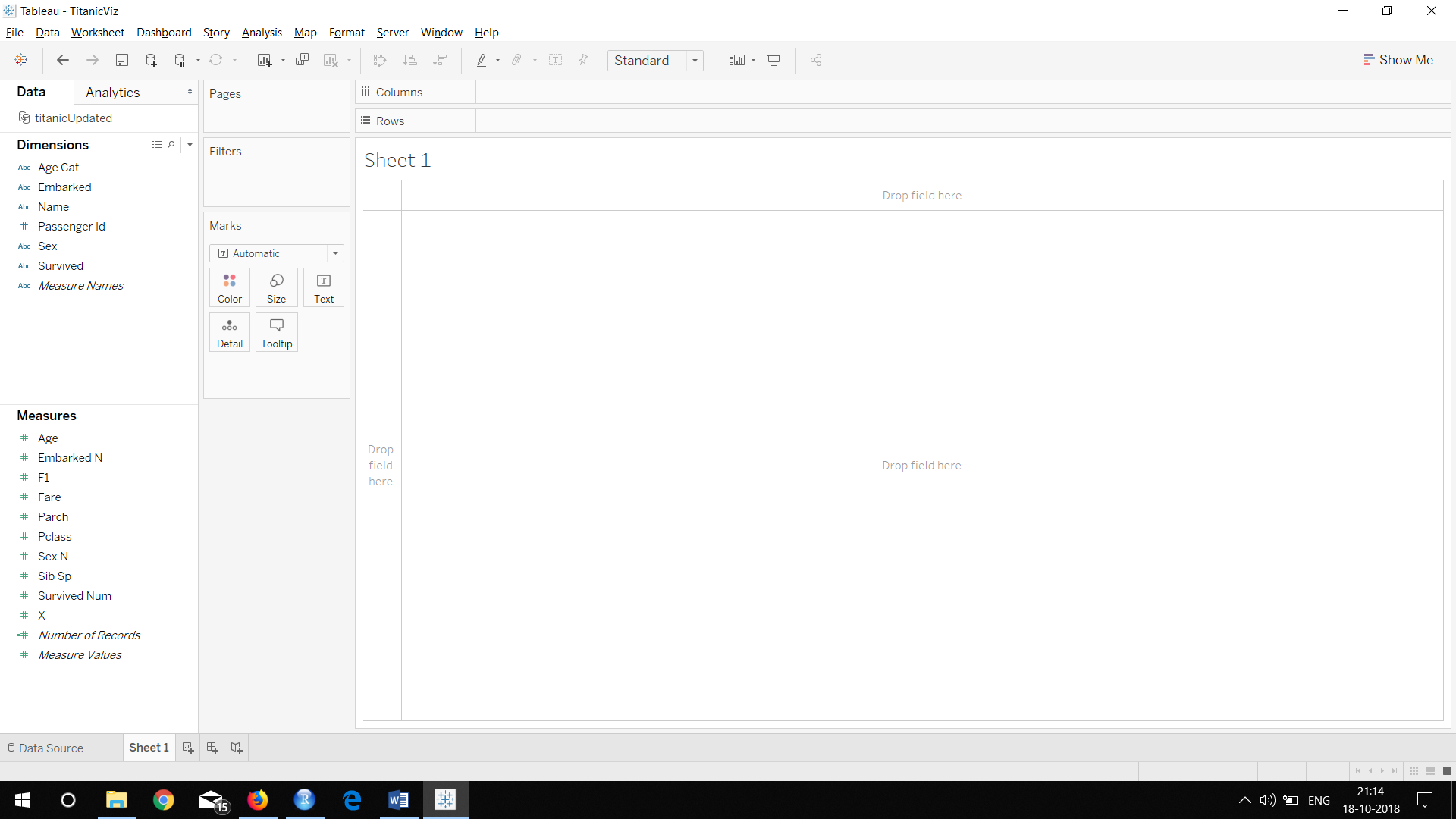




**5) TABLEAU / R INTEGRATION**

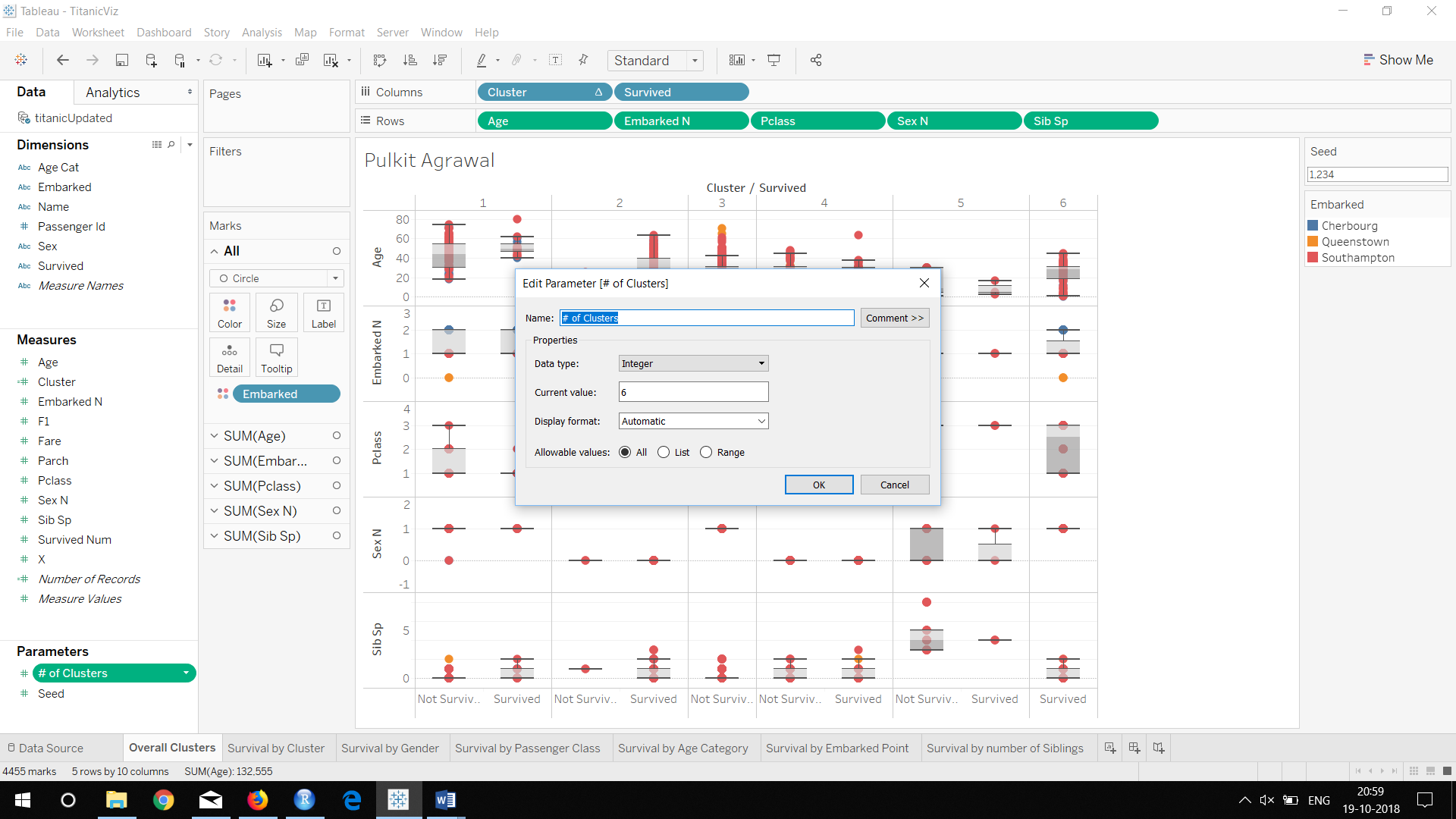
Q 5) Paste a screenshot of the current set of dimensions and measures.

Answer:



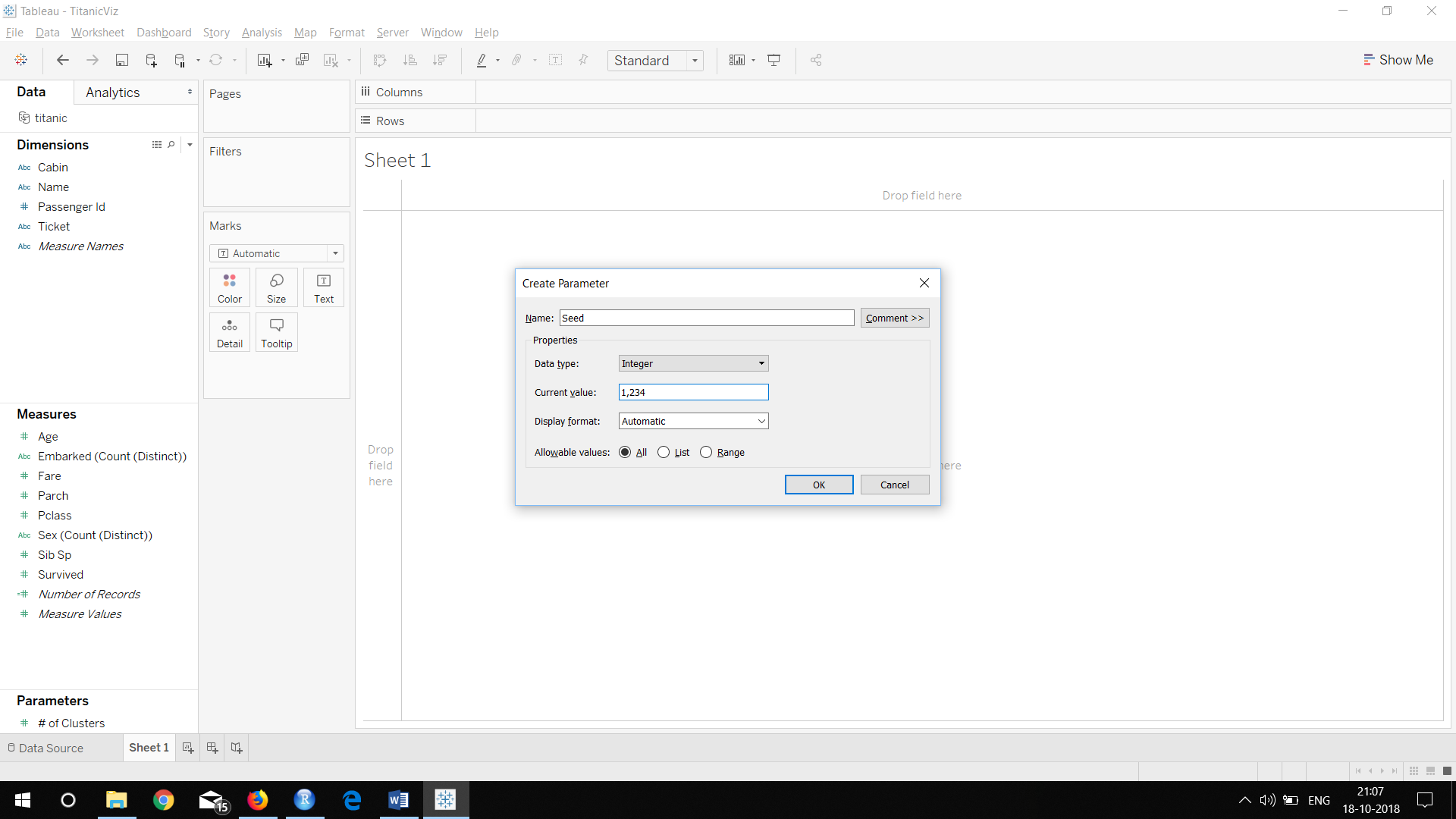
Q6) Please paste screenshot once you populate all fields for this parameter.

Answer:



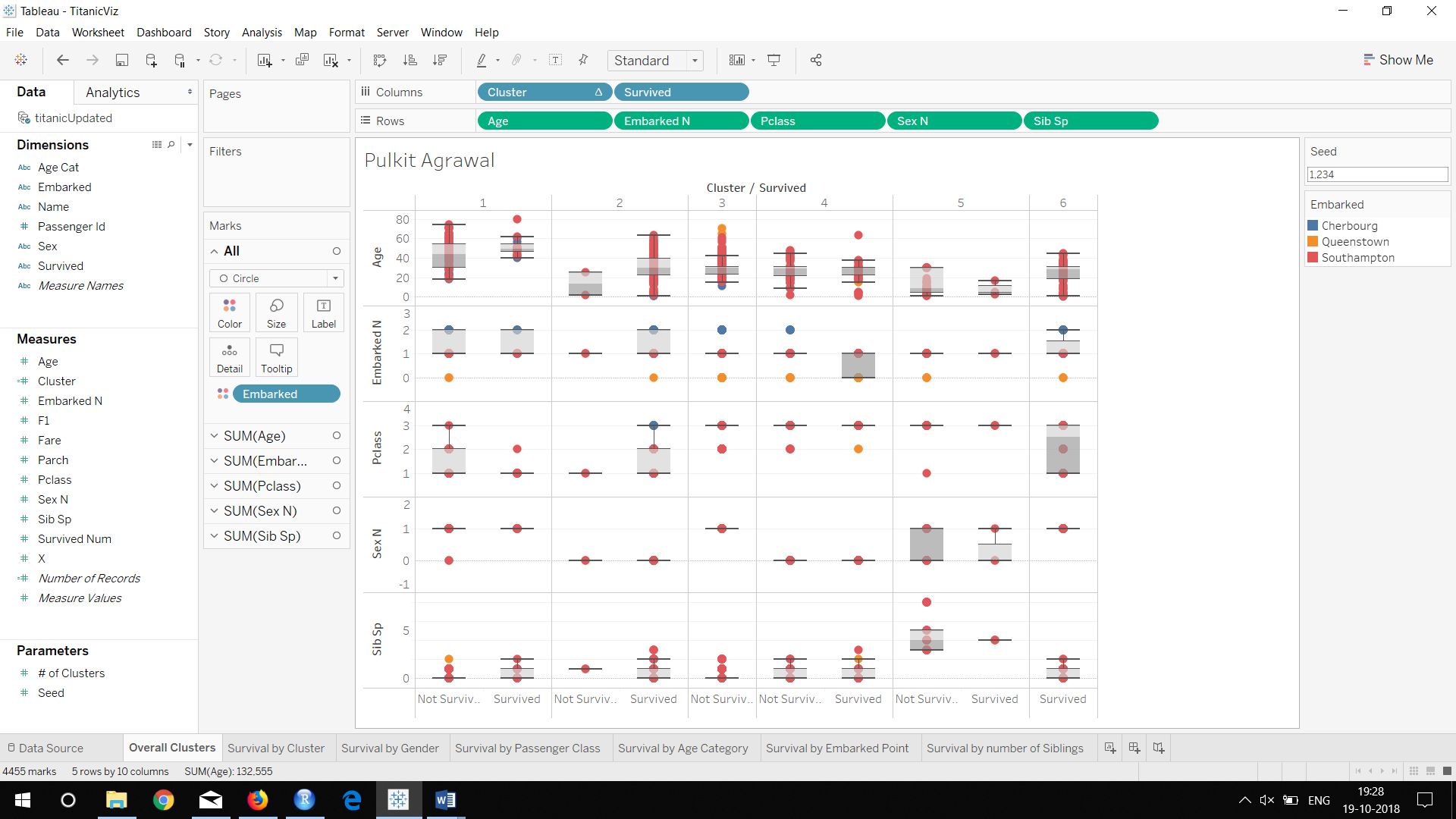
Q7) Paste a screenshot once you populate all fields for this “Seed” parameter.

Answer:



Q8) Paste the screenshot of the worksheet “Overall Clusters” below

Answer:



Q9) Provide your insights from cluster 1.

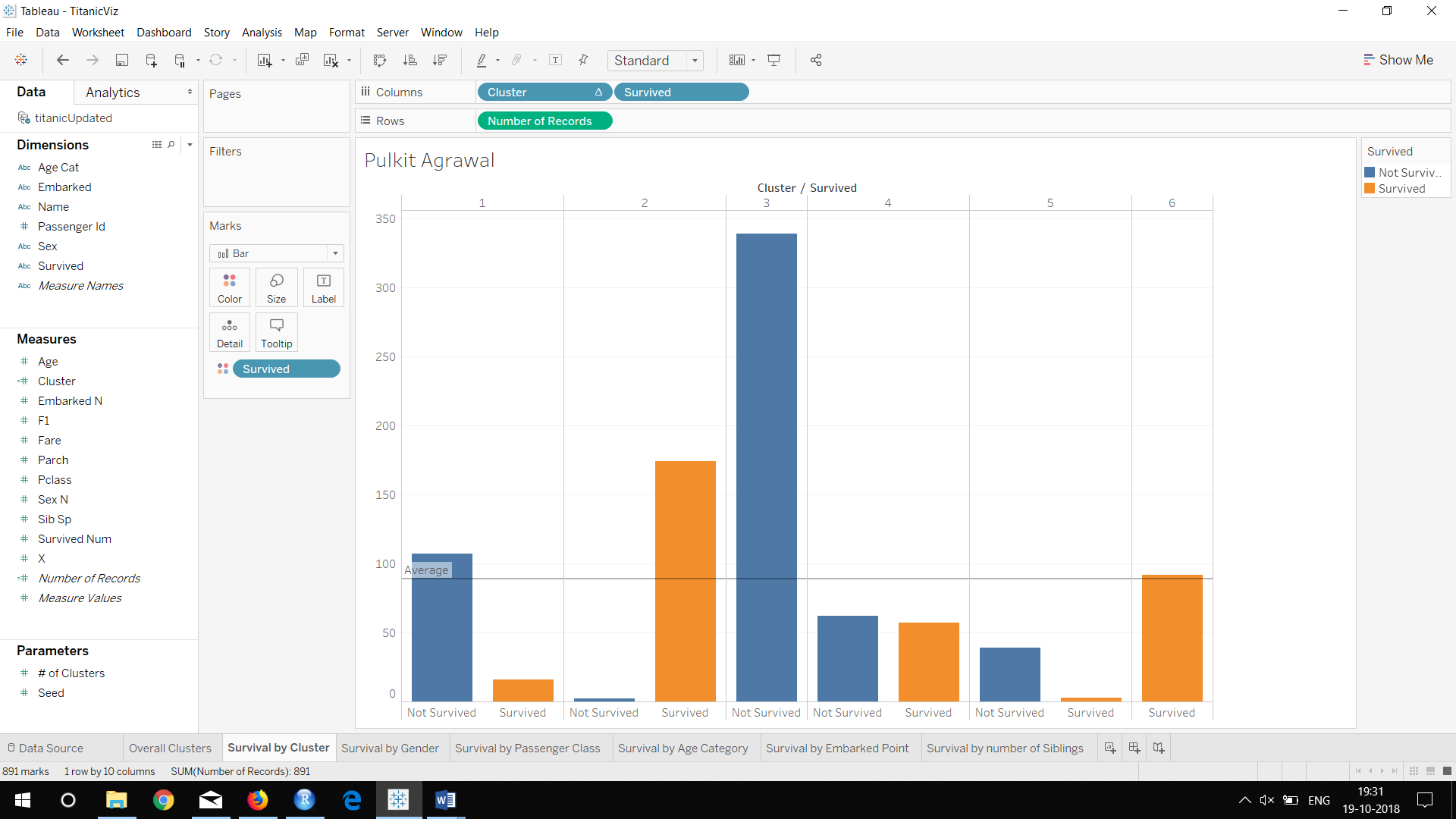
Answer: Only males have survived in cluster 1. Median age of Not Survived is 44 and median age of Survived is 49. Passenger Classes present are 1,2 and 3 for Not Survived and 1 , 2 for survived. Embarked points available are Cherbourg, Queenstown, Southampton for not survived and Southampton and Cherbourg for Survived. Maximum number of siblings present in cluster 1 is 2.

Q10) Provide your insights from cluster 4.

Answer: No Males have survived in Cluster 4. Median age of Not Survived is 29.5 and median age of Survived is 30. Passenger Classes present are 2 and 3 for Survived and Not Survived. Embarked points available are Southampton and Queenstown for Survived and, Southampton, Queenstown and Cherbourg for Not Survived. Maximum number of siblings present in cluster 4 is 3

Q 11) Paste the screenshot of your worksheet.

Answer:



Q 12) From the bar chart, could you name two clusters where survivability is the highest?

Answer: 2 and 6

Q 13) Enter the top two clusters you’ve identified and enter them against <Cluster 1> and <Cluster 2>. Leave the other rows blank for now.

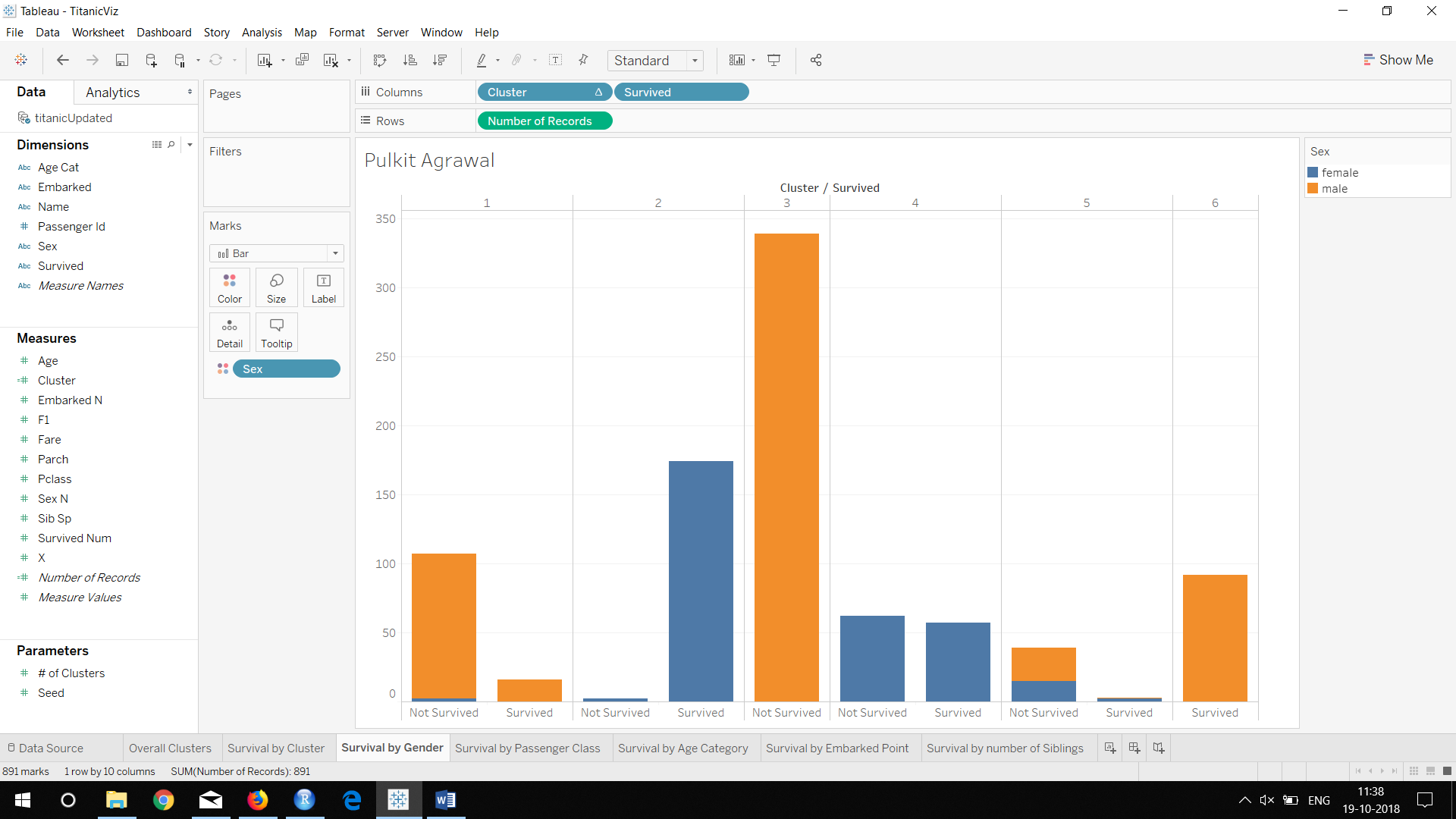
|  |  |  |
| --- | --- | --- |
|  | Cluster 2 | Cluster 6 |
| Ideal Gender |  |  |
| Ideal Passenger Class |  |  |
| Ideal Age Category |  |  |
| Ideal Embarked point |  |  |
| Ideal number of siblings |  |  |

Q14) Based on your findings, enter the ideal Gender (that has the best chance to survive) in your top two cluster.

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

Q15) Paste your screenshot of the Tableau worksheet below. Make sure it carries the Title with your name.

Answer:



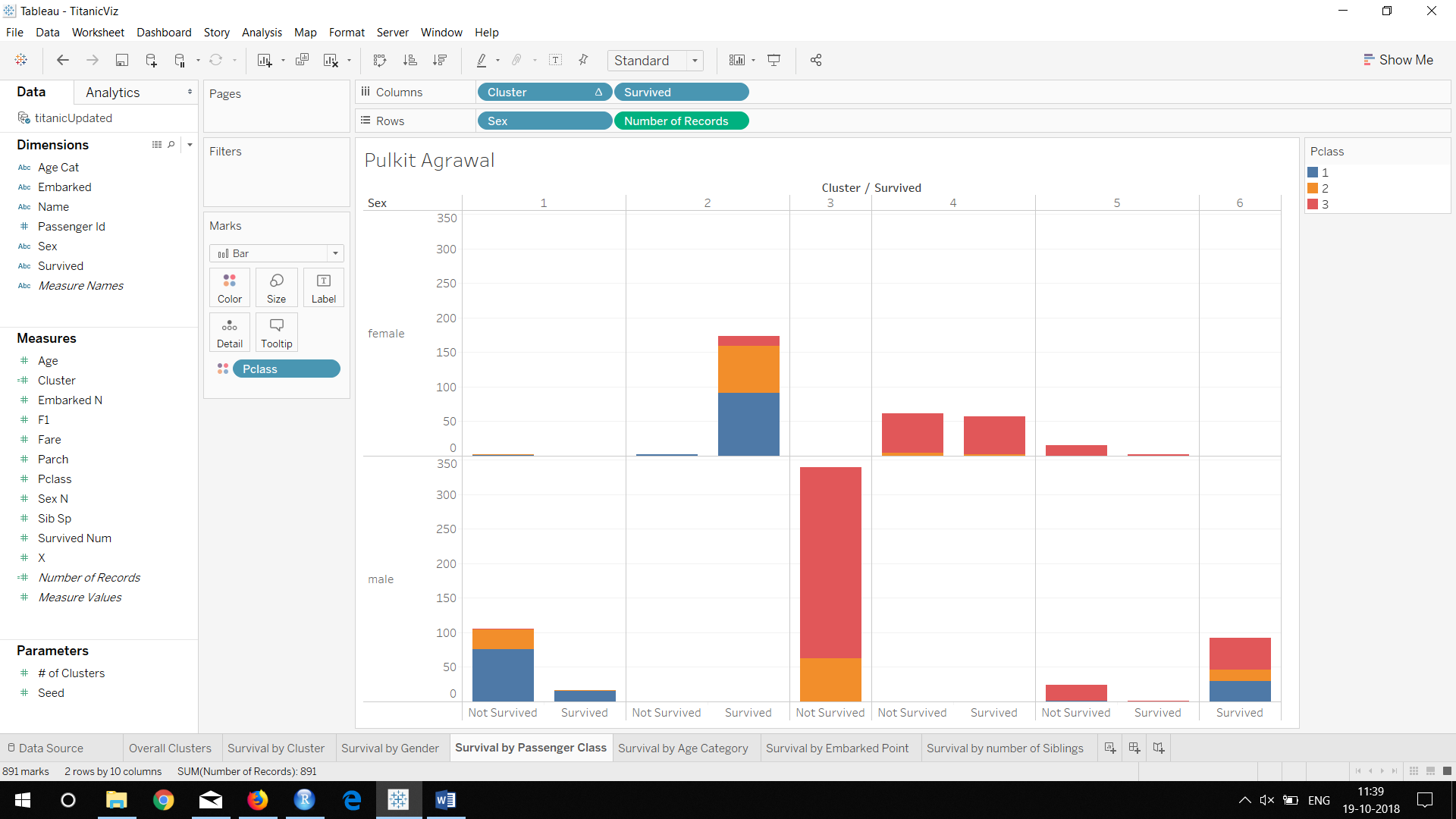
Q16) If you are Female, which cluster should you belong, to ensure higher probability of survival?

Answer: Cluster 2

Q17) Based on your findings, enter the ideal Gender/Passenger Class (that has the best chance to survive) in your top two cluster.

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

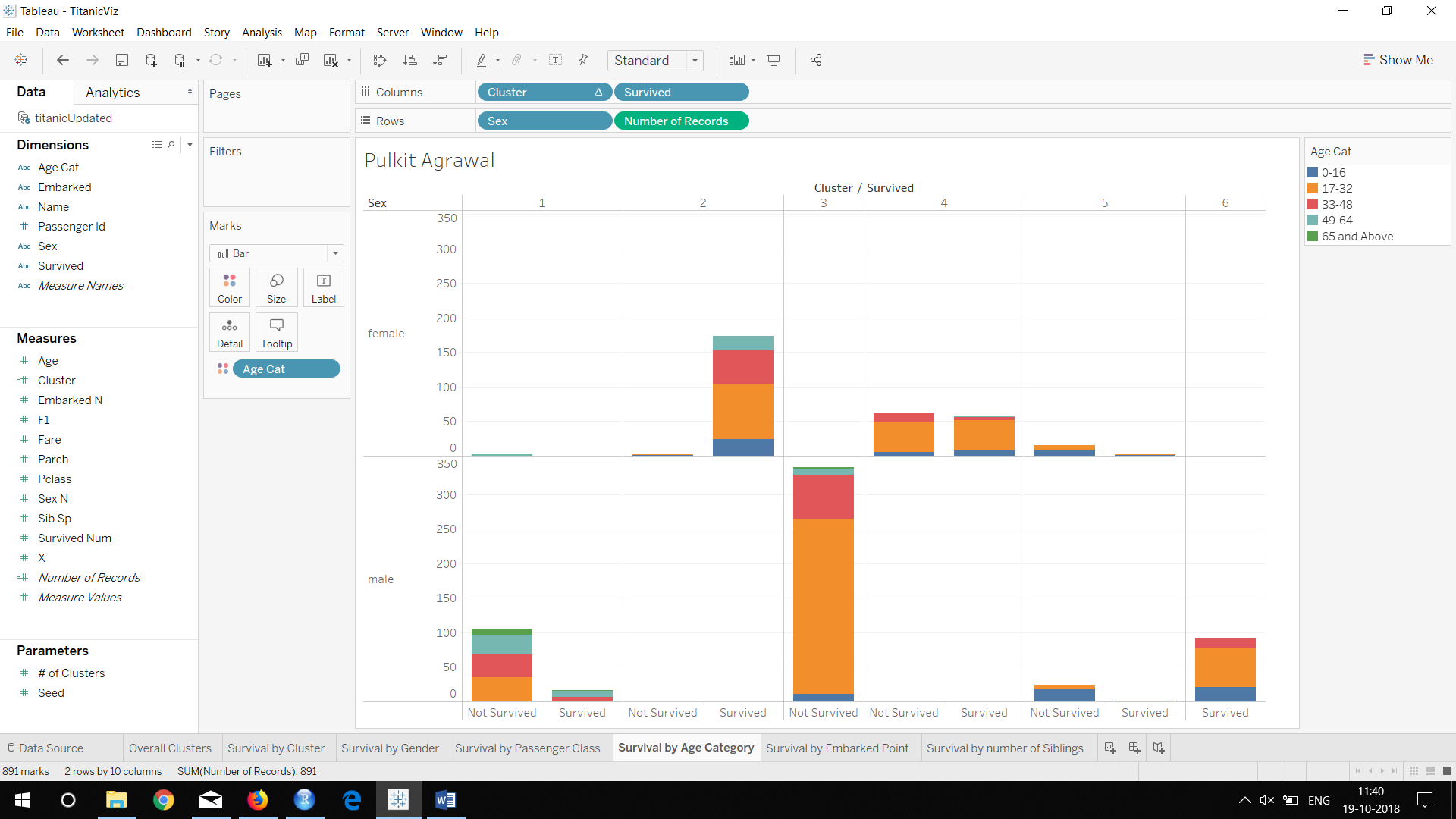
Q18) Paste your screenshot of the Tableau worksheet below. Ensure you have the title with your name on it.



Q19): Based on your findings, enter the ideal Gender/Age Category Class (that has the best chance to survive) in your top two cluster.

|  |  |  |
| --- | --- | --- |
|  | Cluster 2 | Cluster 6 |
| Ideal Gender | Female | Male |
| Ideal Passenger Class | 1 | 3 |
| Ideal Age Category | 17-32 | 17-32 |
| Ideal Embarked point |  |  |
| Ideal number of siblings |  |  |

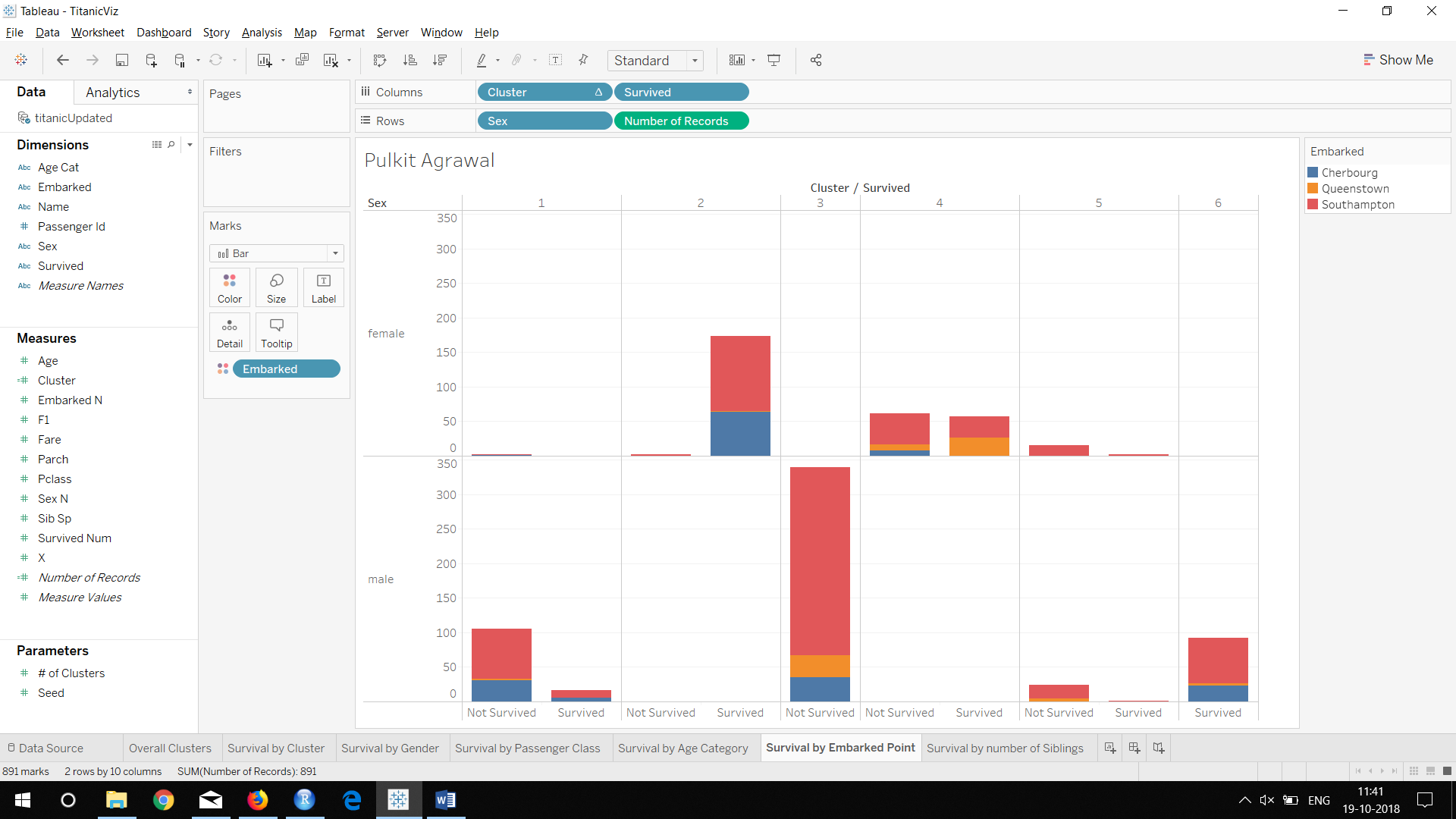
Q20) Paste your screenshot of the Tableau worksheet below. Ensure you add the title with your name.



Q21) Output: Based on your findings, enter the ideal Gender/Embarked point (that has the best chance to survive) in your top two cluster.

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

Q22) Paste your screenshot of the Tableau worksheet below. Ensure you have the title added with your name.

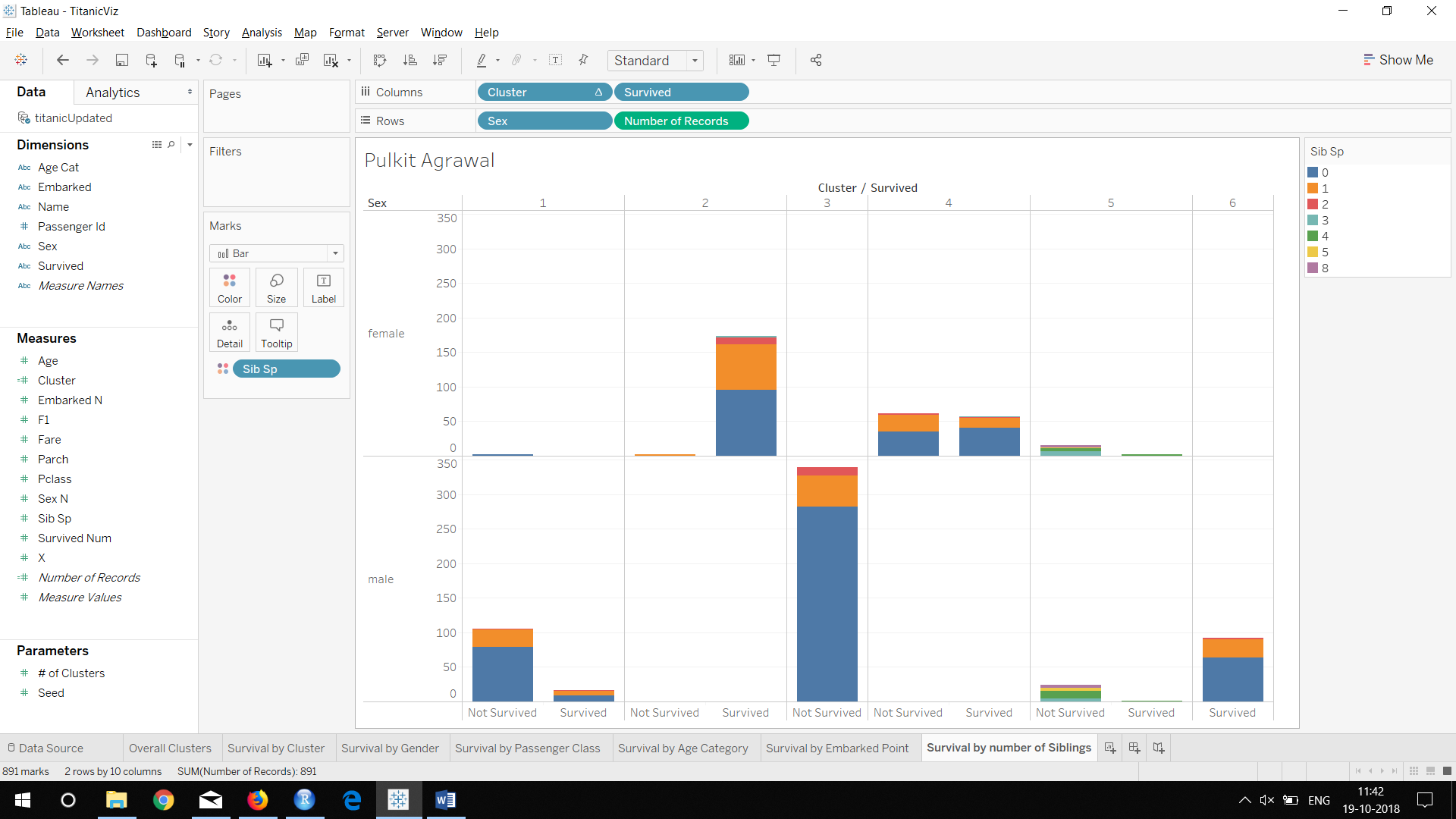


Q23) Output: Based on your findings, enter the ideal Gender/# of Sib Sp (that has the best chance to survive) in your top two cluster.

|  |  |  |
| --- | --- | --- |
|  | Cluster 2 | Cluster 6 |
| Ideal Gender | Female | Male |
| Ideal Passenger Class | 1 | 3 |
| Ideal Age Category | 17-32 | 17-32 |
| Ideal Embarked point | Southampton | Southampton |
| Ideal number of siblings | 0 | 0 |

Q24) Paste your screenshot of the Tableau worksheet below. Ensure you have the title with your name on it.

Answer:



Q22) Please put the profile of passengers from each of the top two clusters in words. Use the above filled in table to form your passenger profile. Please write two separate sentences explaining each cluster profile.

Answer: **Cluster 2 profile:**

Females belonging to passenger class 1 in the age group 17 – 32 embarking at Southampton and having zero siblings have the best chance for survival in cluster 2.

**Cluster 6 profile:**

Males belonging to passenger class 3 in the age group 17 – 32 embarking at Southampton and having zero siblings have the best chance for survival in cluster 6.