3/13/2020

Suseel K, Rohit

rxs190016

Integrated Analysis with R

Exercise 7

**Data Retrieval**

**1.**Write the output of the above executed command.

dim(titanic)

[1] 891 12

Number of Rows = 891

Number of Columns = 12

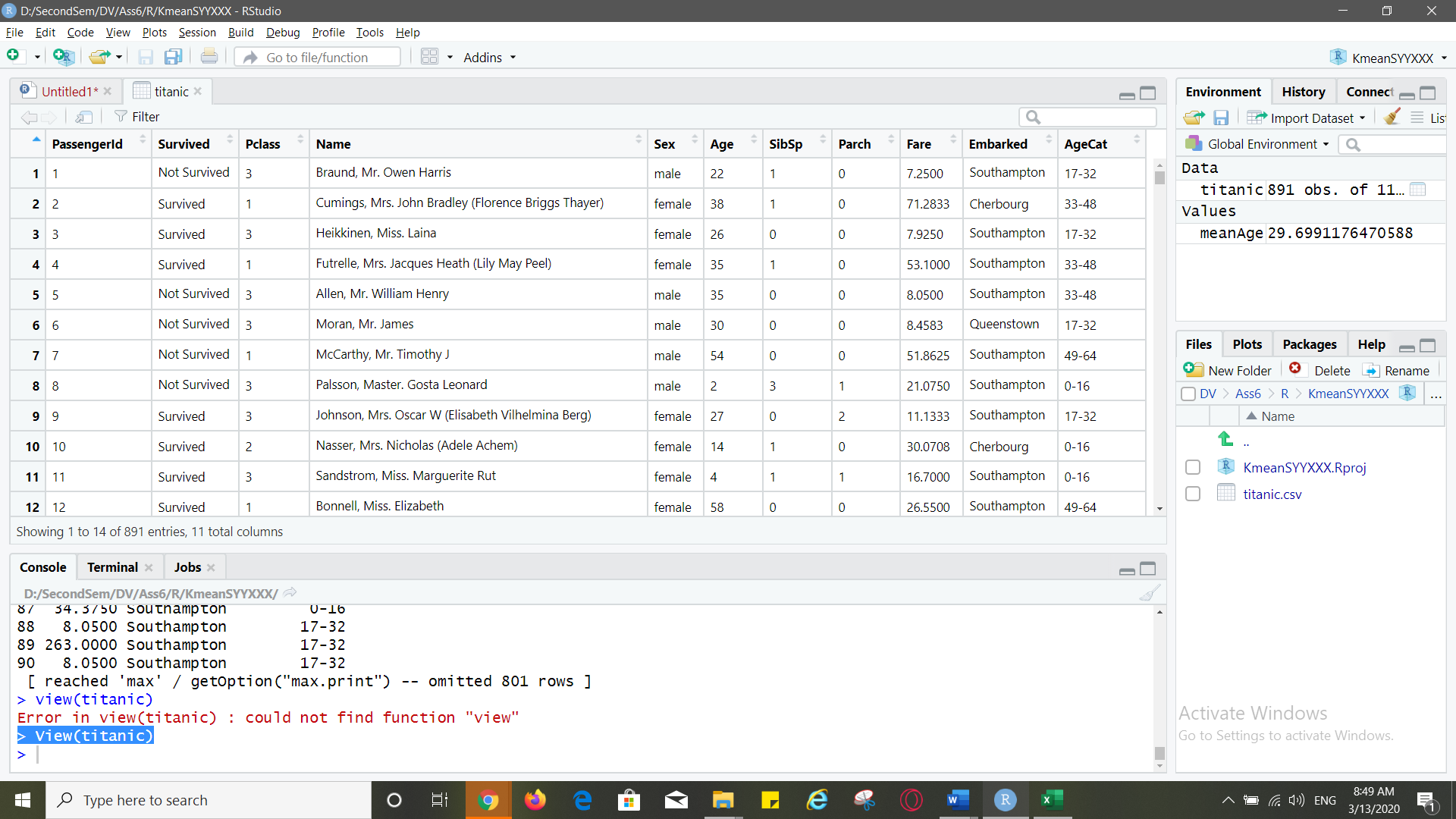
**Data Pre-processing**

2) What is the average age that you get?

29.69912

3.Paste the screenshot of the below executed command.

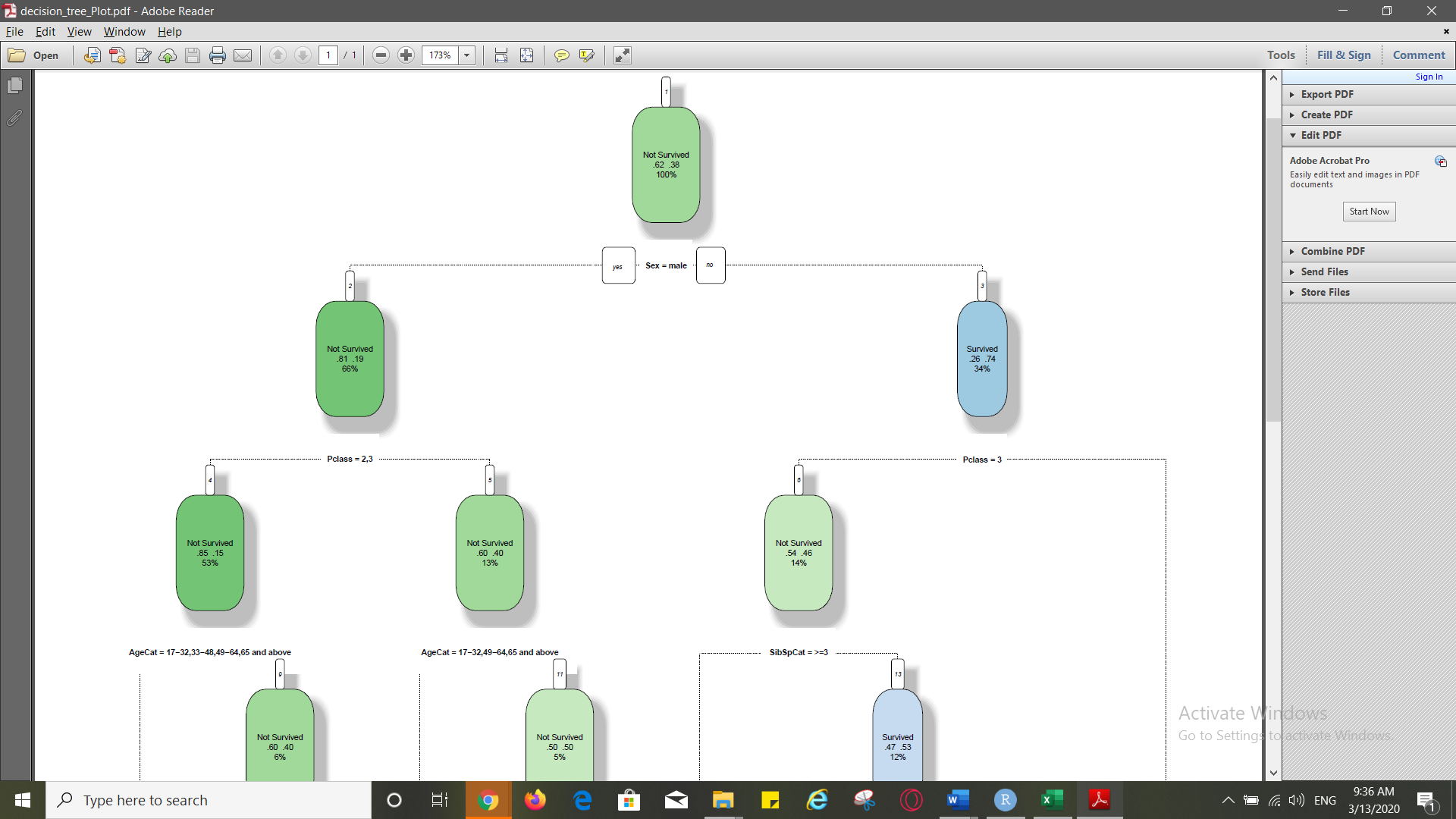
> View(titanic)

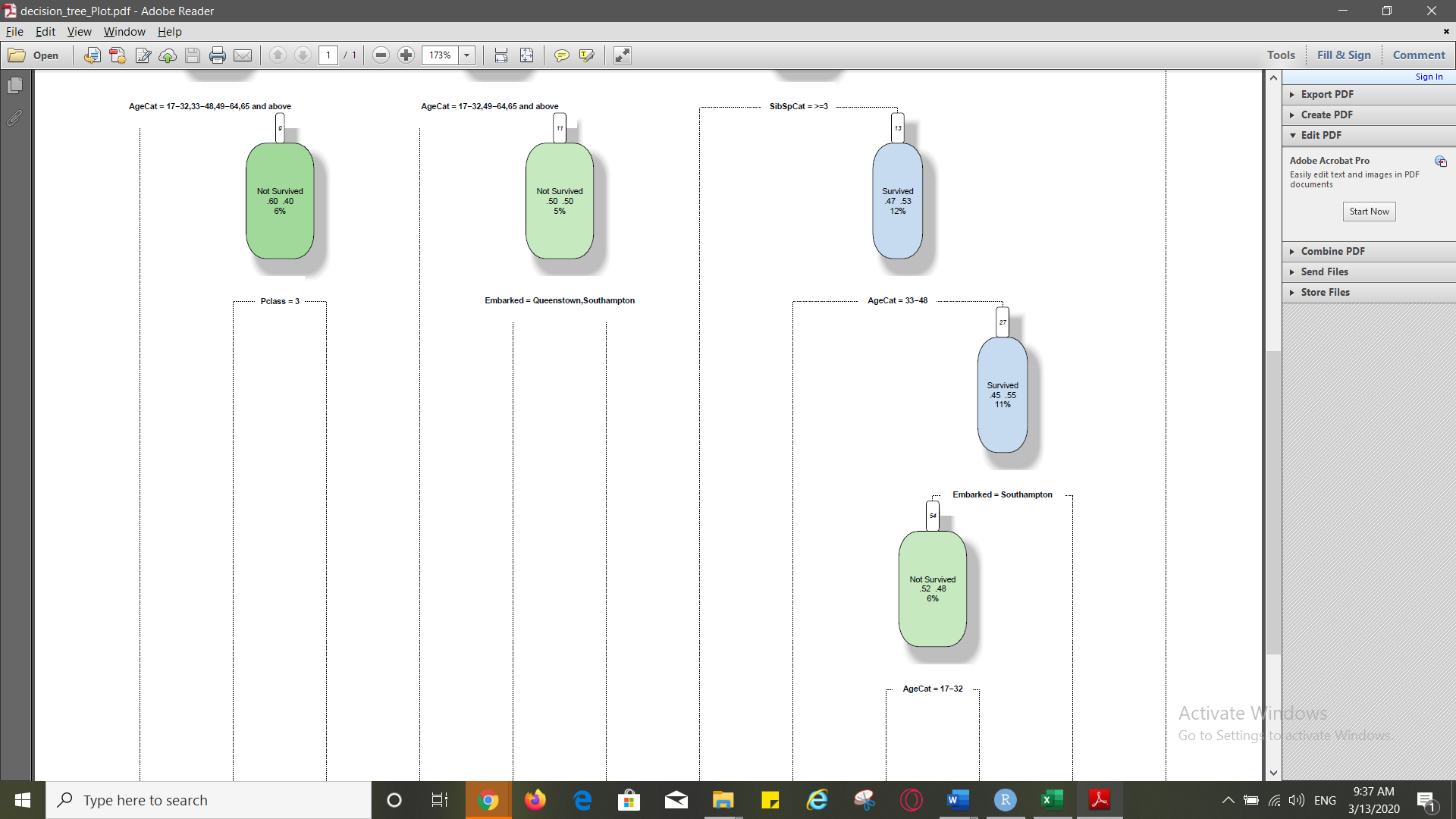


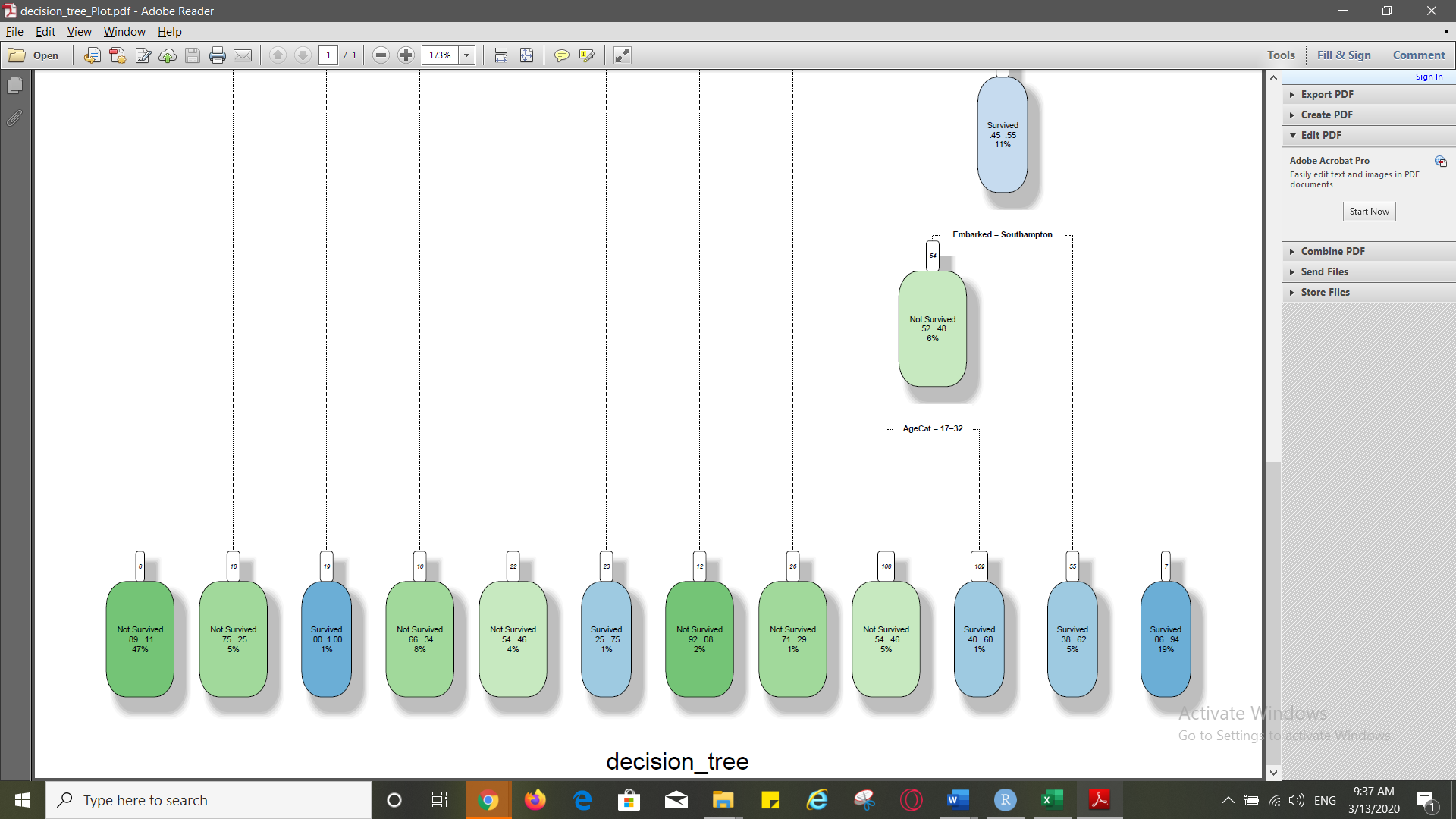
**Decision Tree**

4.Paste the screenshot of the decision tree below (for better visibility, attached the plot in 3 screenshots)

Tree plot:







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

Misclassification Rate:

0.18

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

**Sex**

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

**38/62**

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

**74/26**

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

**40/60**

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

1. Sex
2. Pclass
3. AgeCat
4. SibSpCat
5. Embarked
6. Survived- dependent variable

**K-means Clustering**

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:

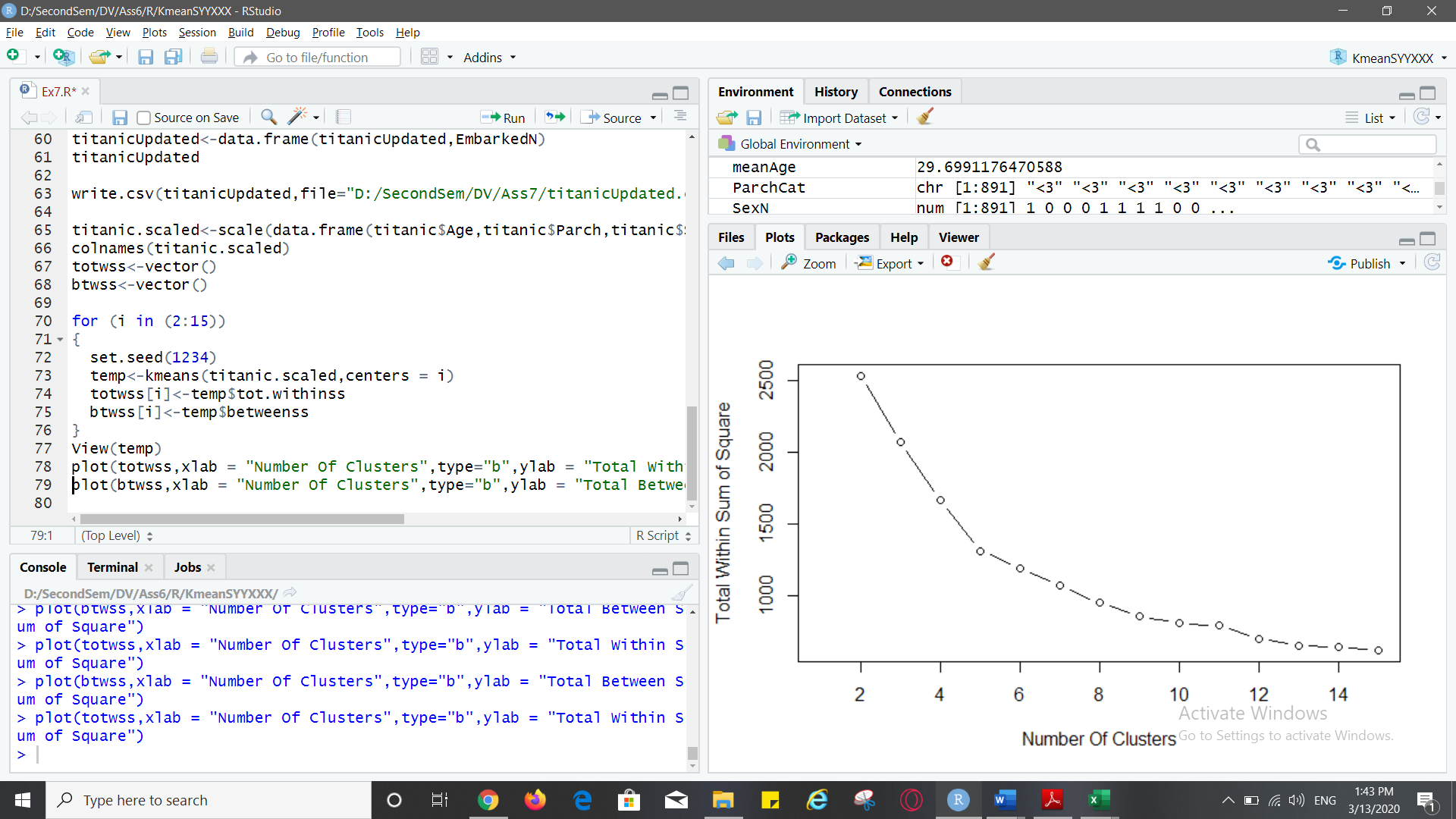
Key Variables

1.Write the desired number of clusters you selected for the analysis

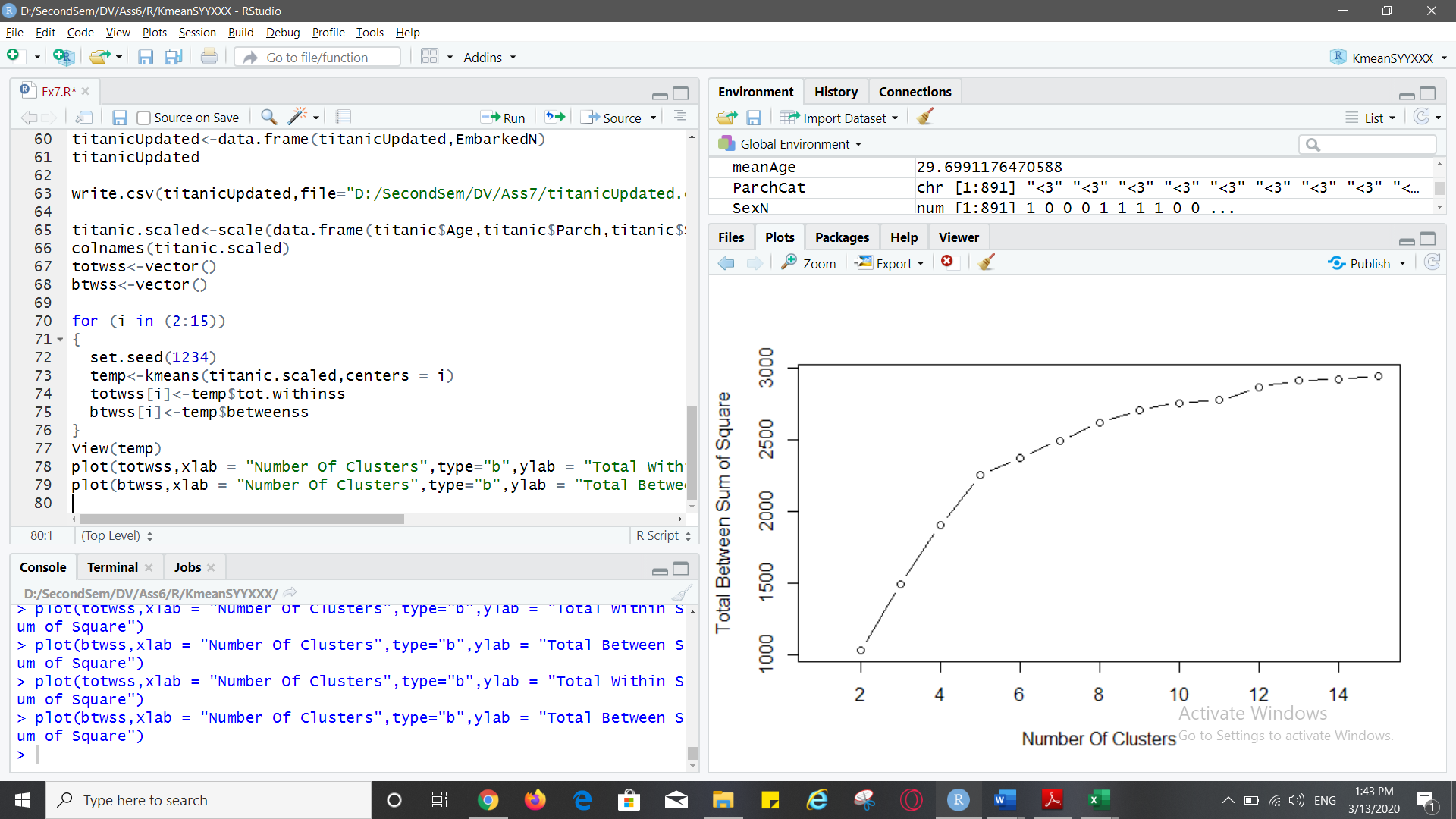
5

2. **Paste the screenshot of the two plots**

TOTWSS

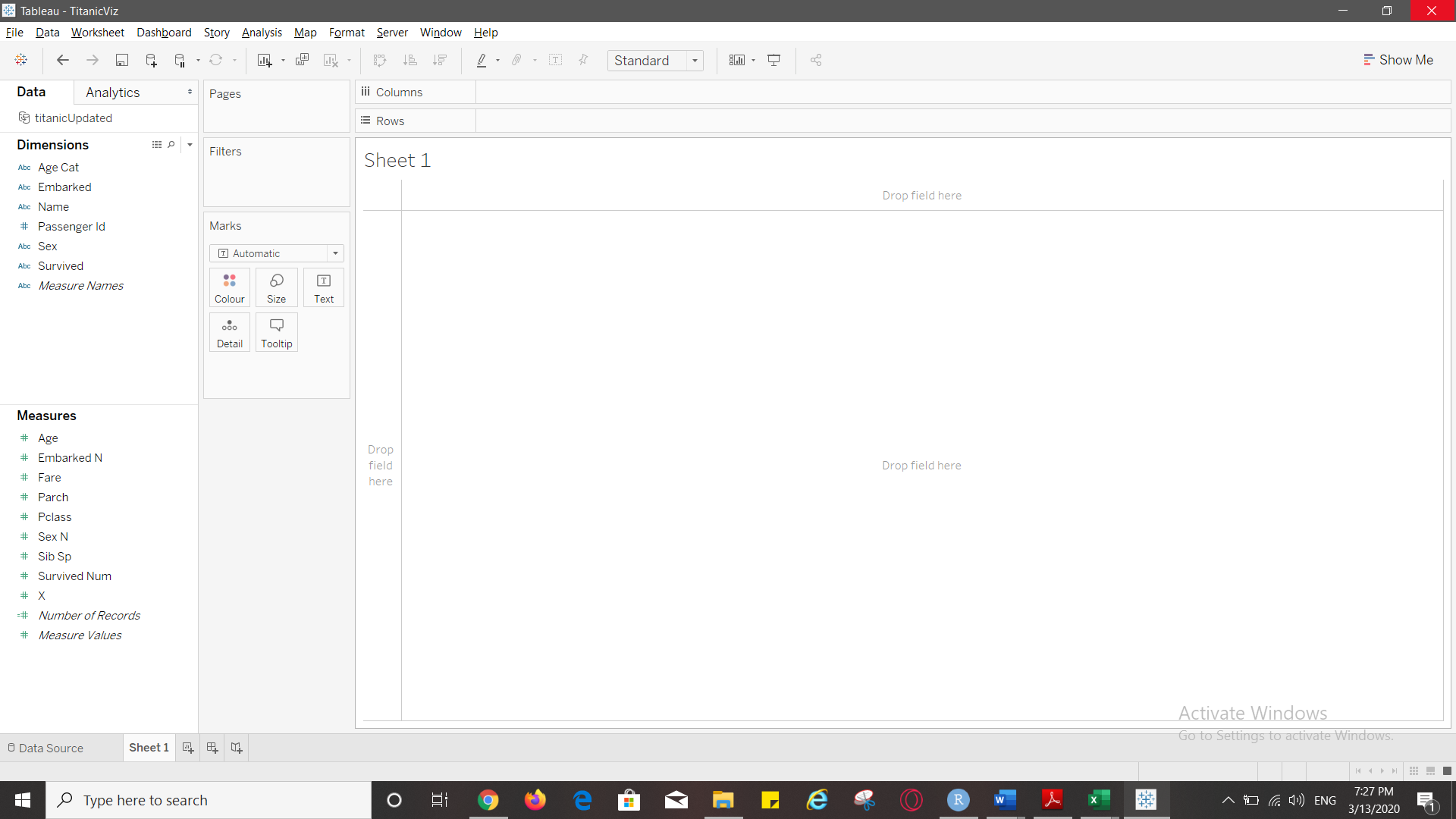


BTWSS

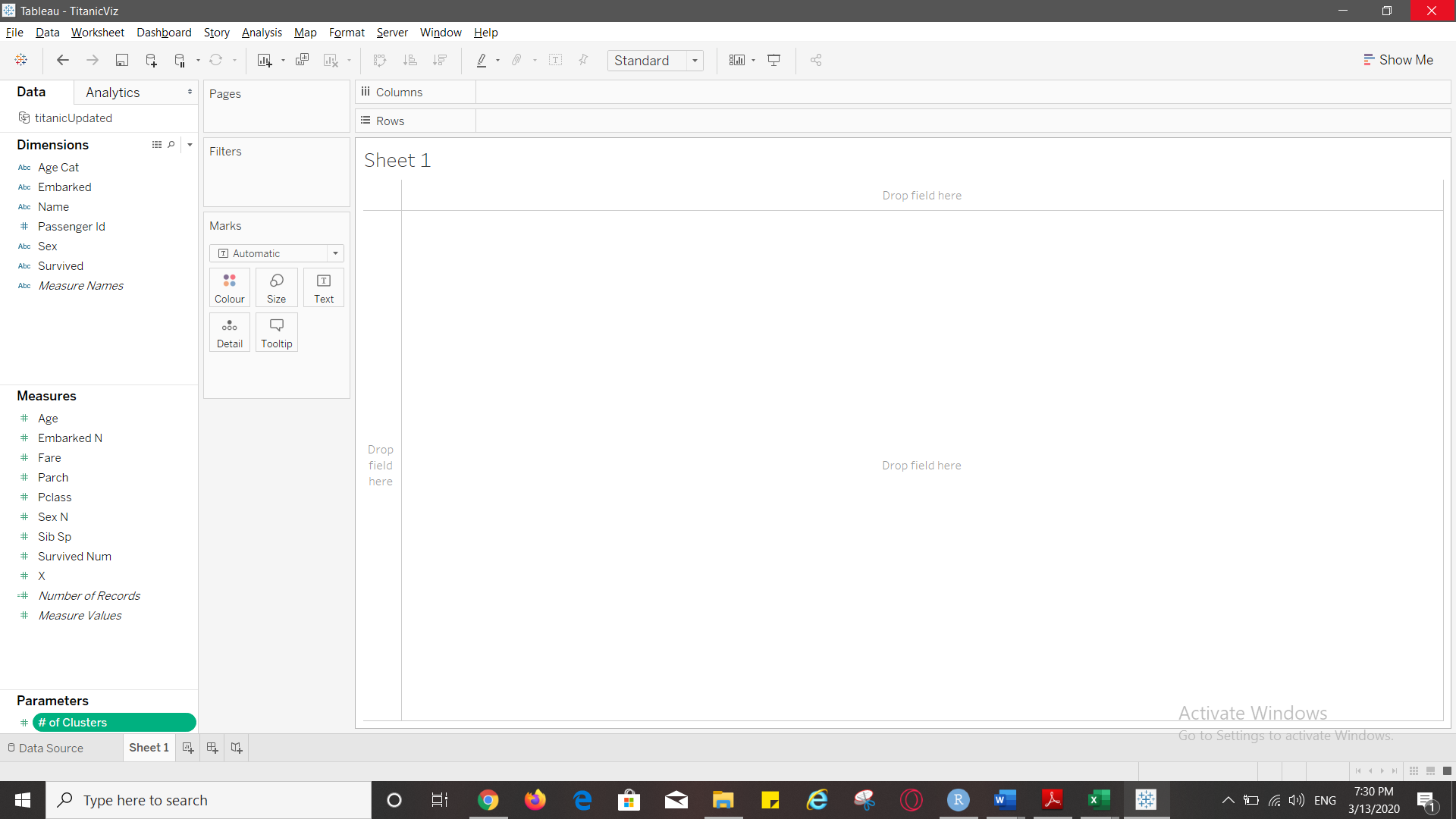
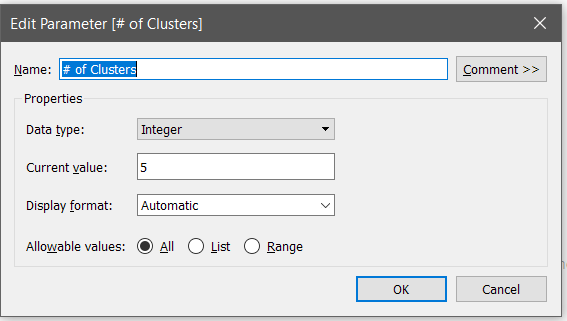


**TABLEAU / R INTEGRATION**

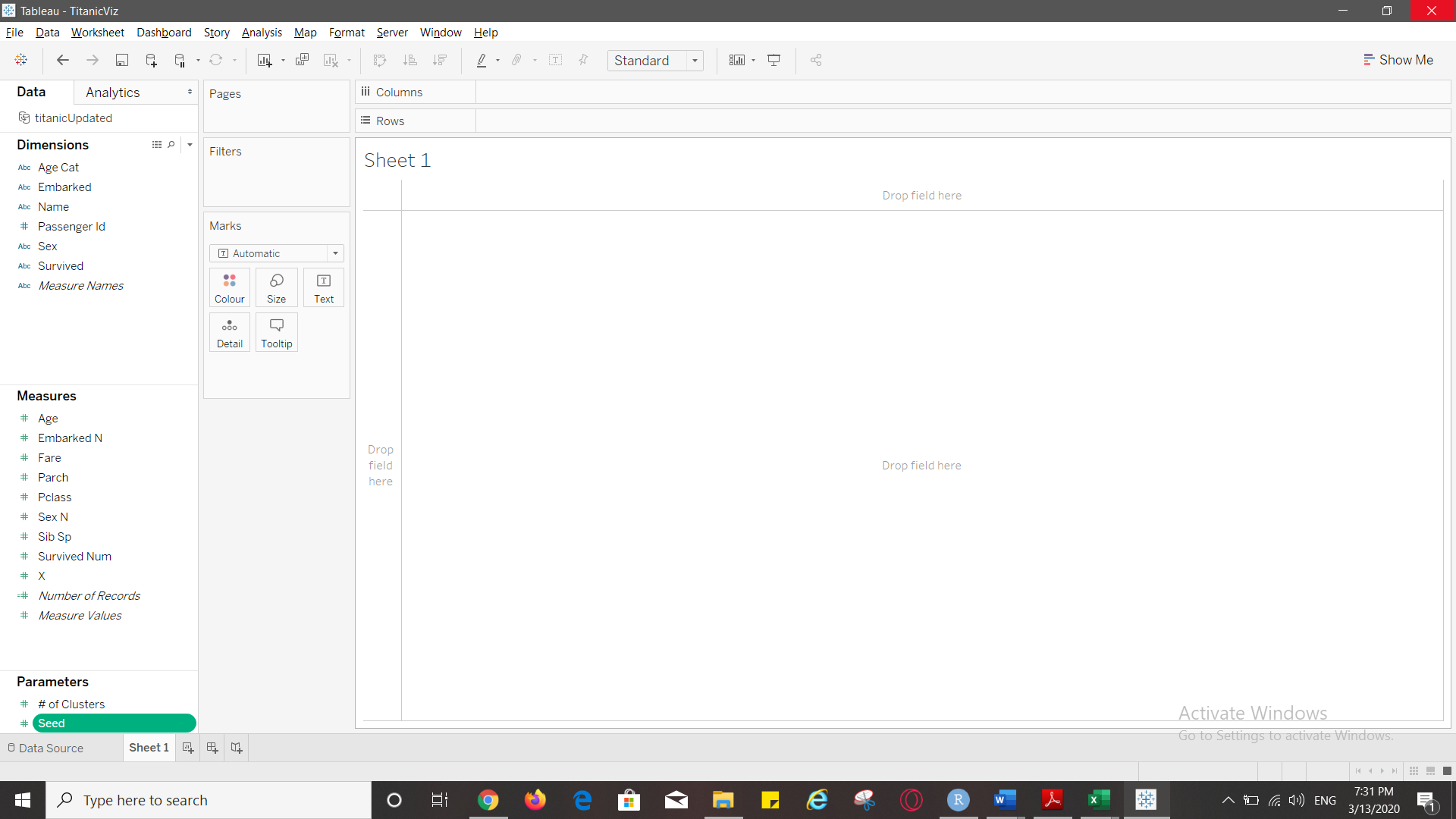
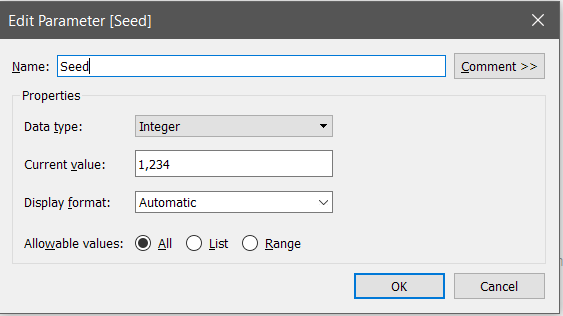
Q 5) Paste a screenshot of the current set of dimensions and measures.



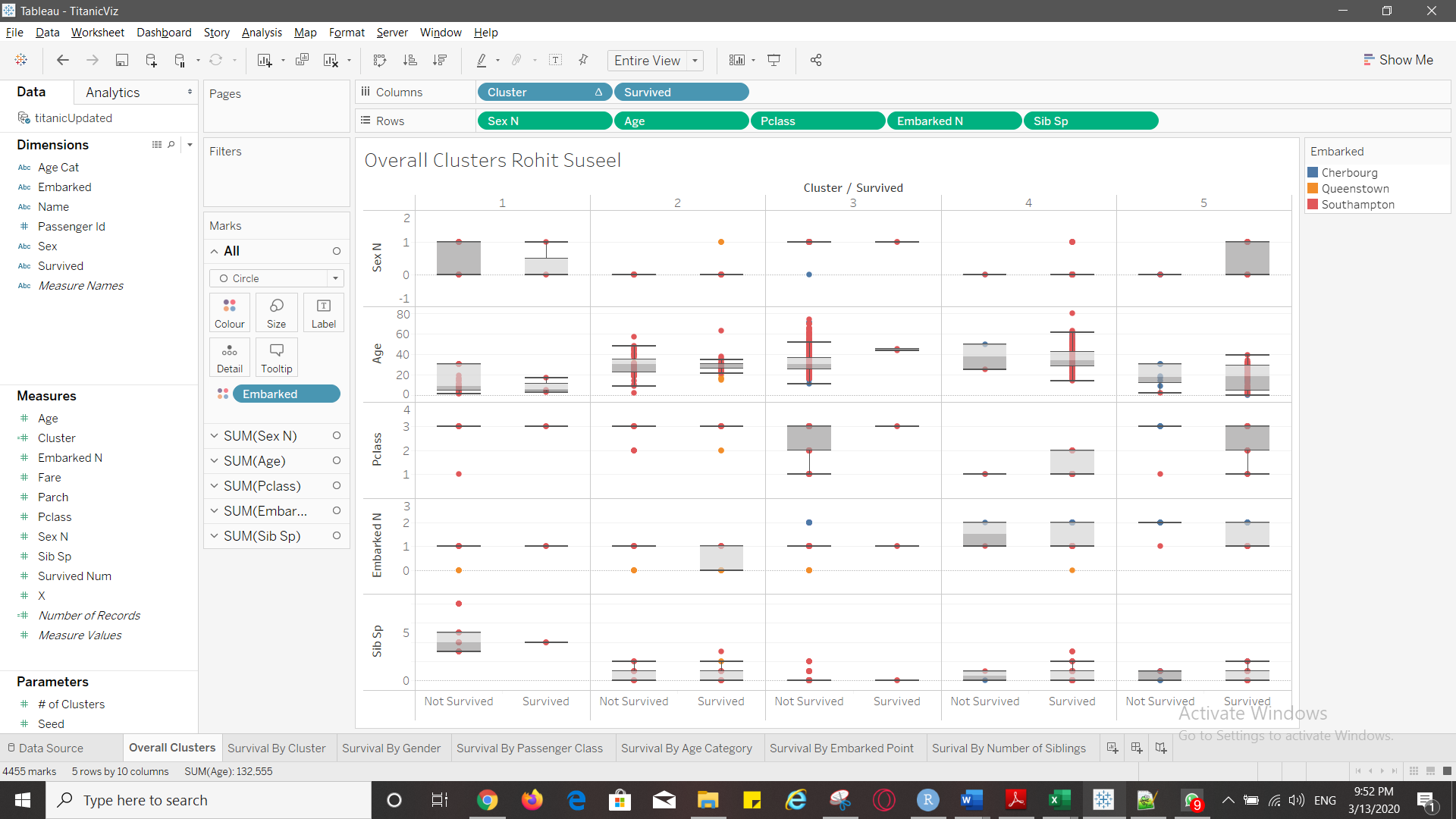
Q6) Please paste screenshot once you populate all fields for this parameter



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



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



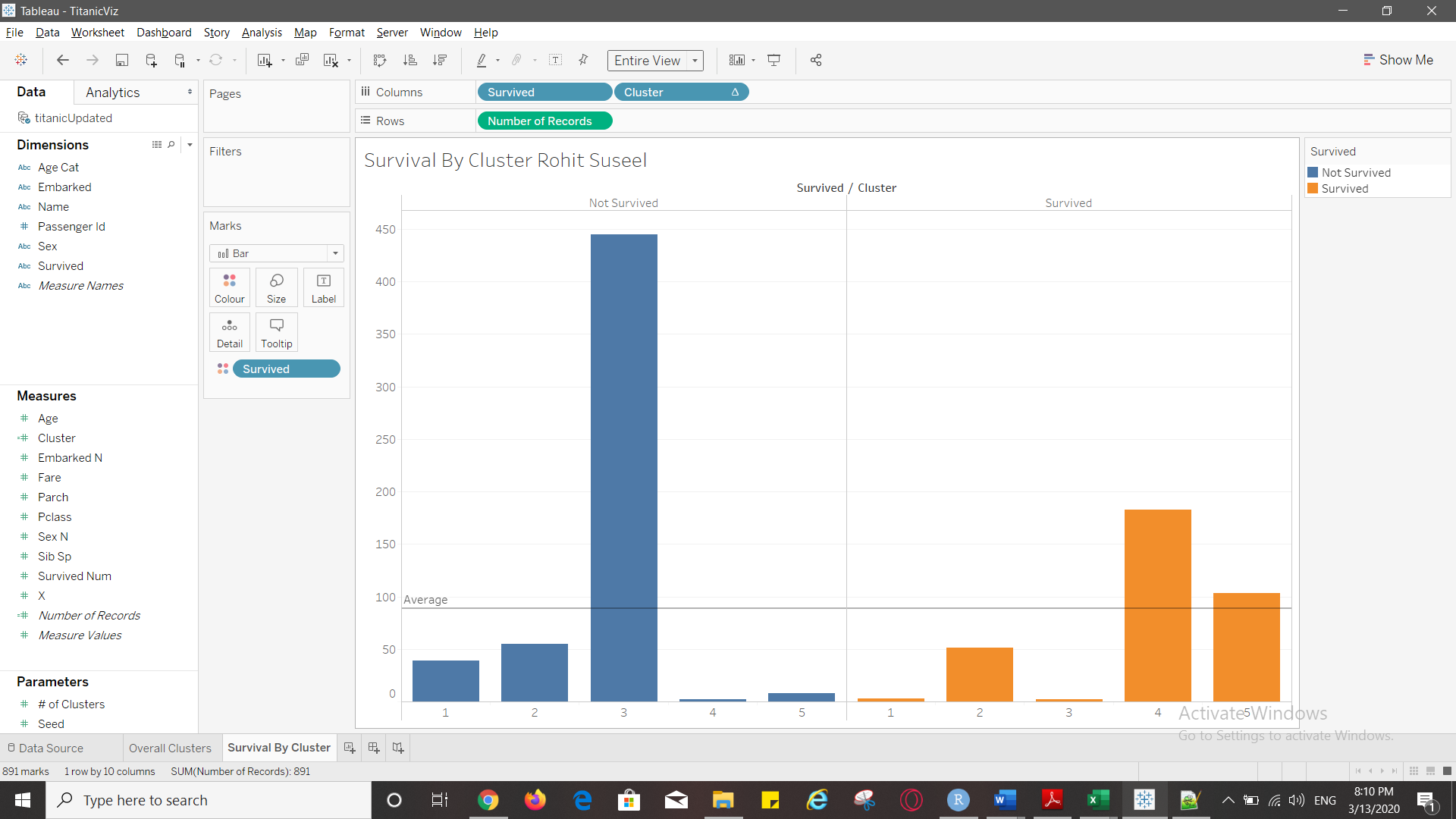
Q9) Provide your insights from cluster 1.

* Passengers from class 1 did not survive in this cluster.
* No passengers embarked from Cherbourg in this cluster.
* There were passengers as young as 1 and as old as 30
* Median age of passengers who could not survive is 9

Q10) Provide your insights from cluster 4.

* The median age of passengers who survived is 34.
* There are no class 3 passengers
* Median age of passengers who didn’t survive is 37.5

Q 11) Paste the screenshot of your worksheet. Ensure that you have added a Title and your name is contained in the title.



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

4 and 5

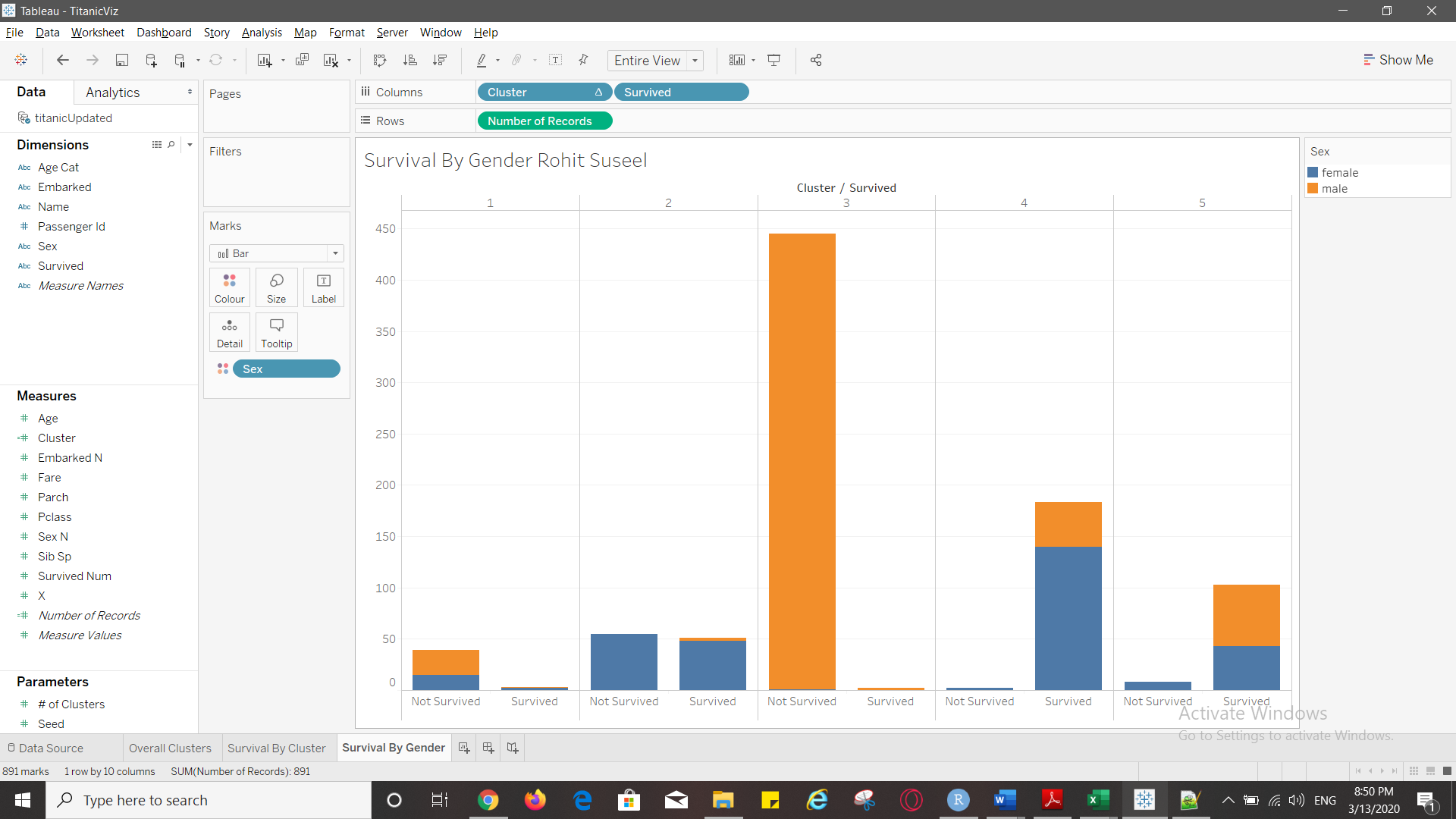
Q 13) Enter the top two clusters you’ve identified and enter them against <Cluster 1> and <Cluster 2>. Leave the other rows blank for now.

|  |  |  |
| --- | --- | --- |
|  | Cluster 4 | Cluster 5 |
| 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 4 | Cluster 5 |
| 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.



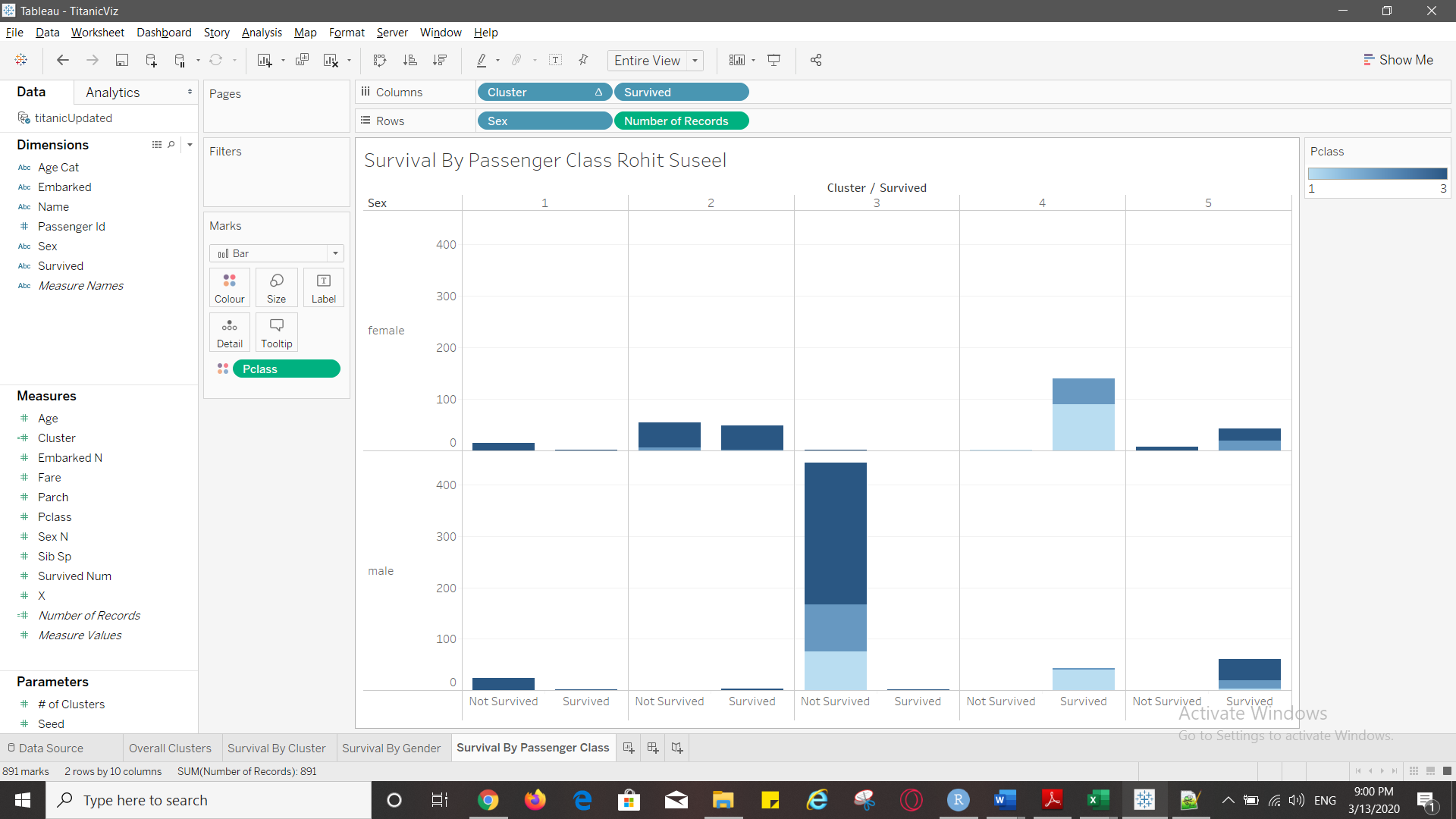
Q16) If you are Female, which cluster should you belong, to ensure higher probability of survival?

Cluster 4

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

|  |  |  |
| --- | --- | --- |
|  | Cluster 4 | Cluster 5 |
| Ideal Gender | Female | Male |
| Ideal Passenger Class | 1 | 3 |
| Ideal Age Category |  |  |
| Ideal Embarked point |  |  |
| Ideal number of siblings |  |  |

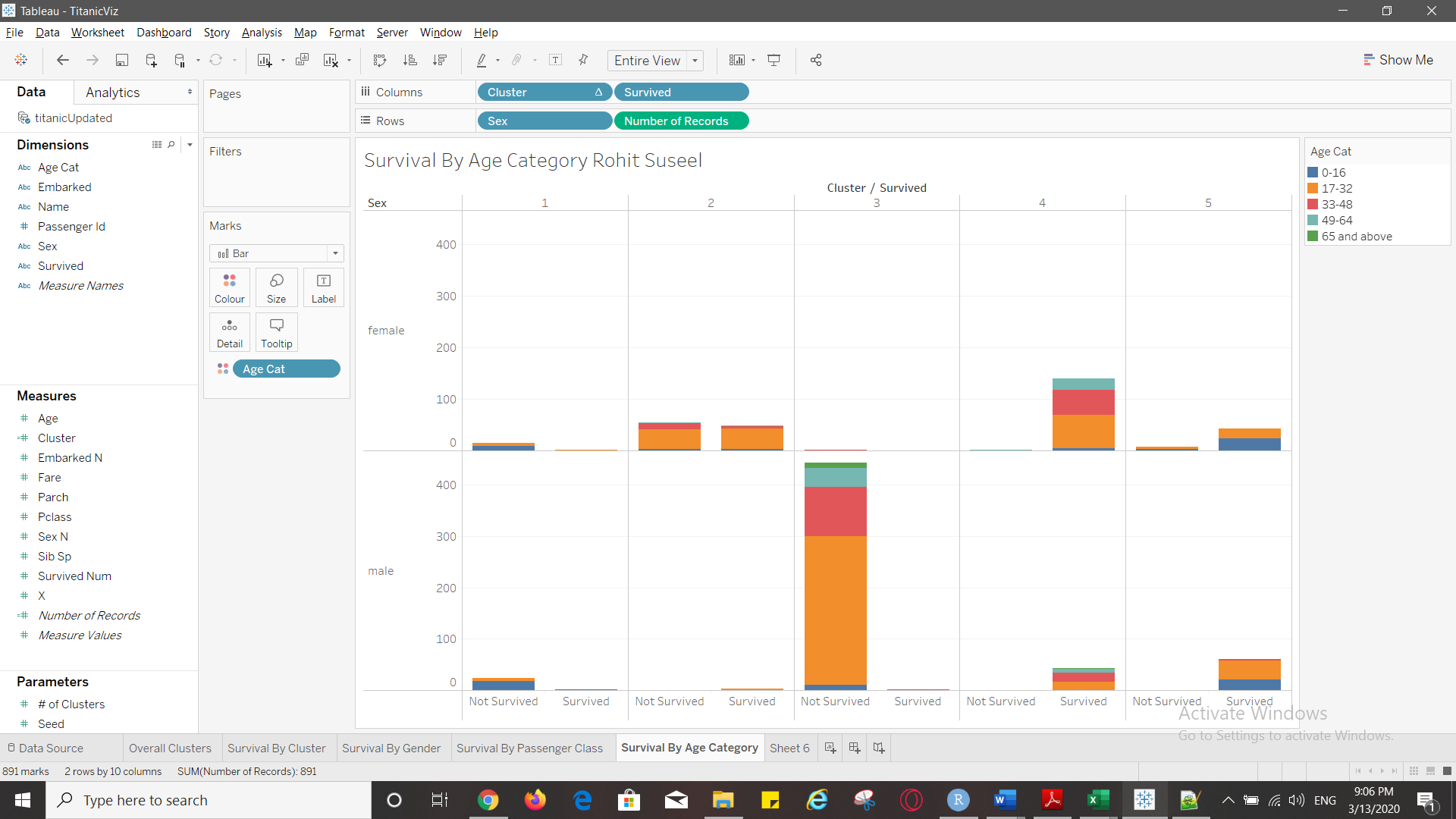
Q18) Paste your screenshot of the Tableau worksheet below. Ensure you have the title with your name on it.



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

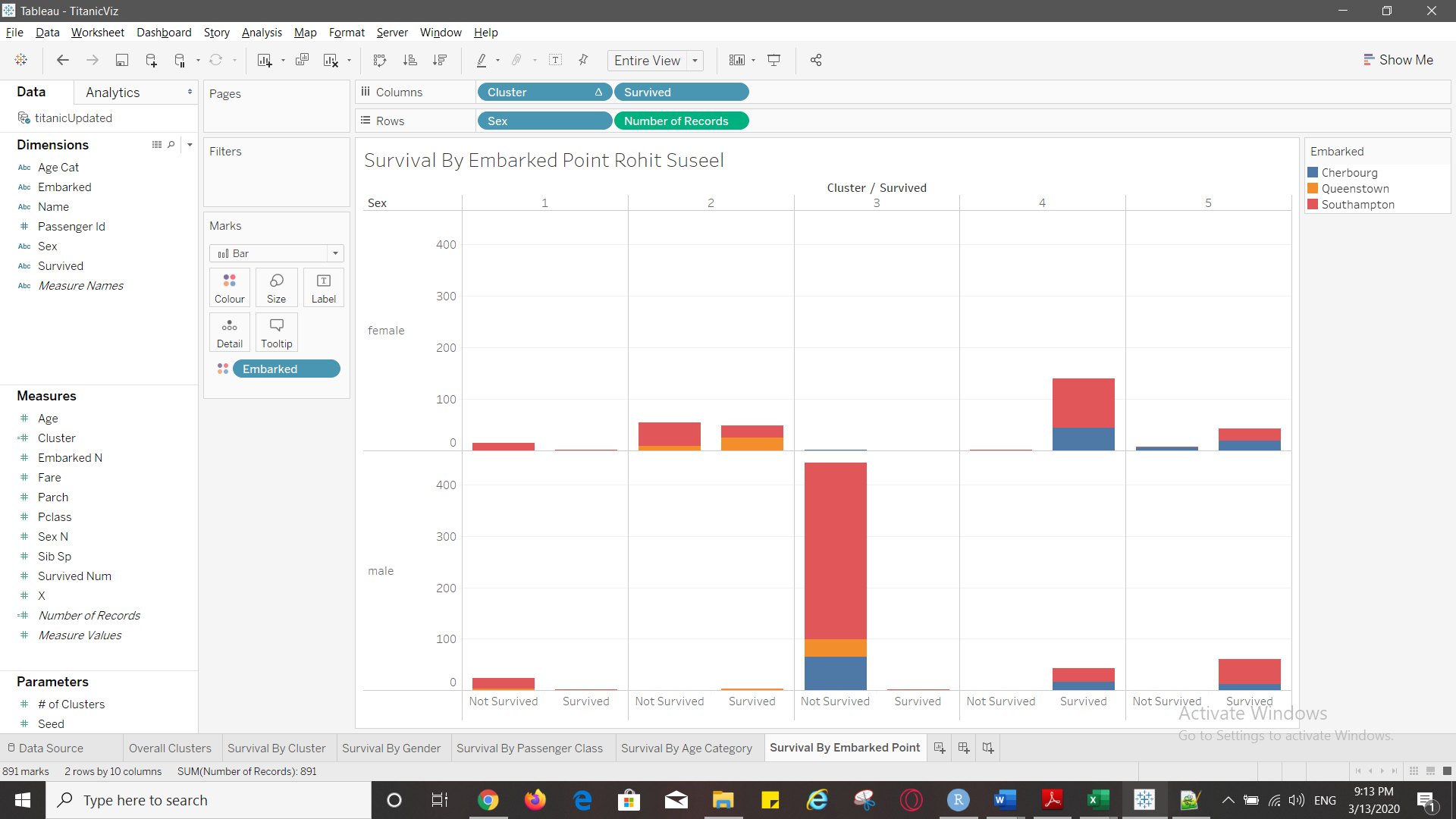
|  |  |  |
| --- | --- | --- |
|  | 4 | 5 |
| Ideal Gender | Female | Male |
| Ideal Passenger Class | 1 | 3 |
| Ideal Age Category | 17-32 | 17-32 |
| Ideal Embarked point |  |  |
| Ideal number of siblings |  |  |

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.

|  |  |  |
| --- | --- | --- |
|  | 4 | 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 |  |  |

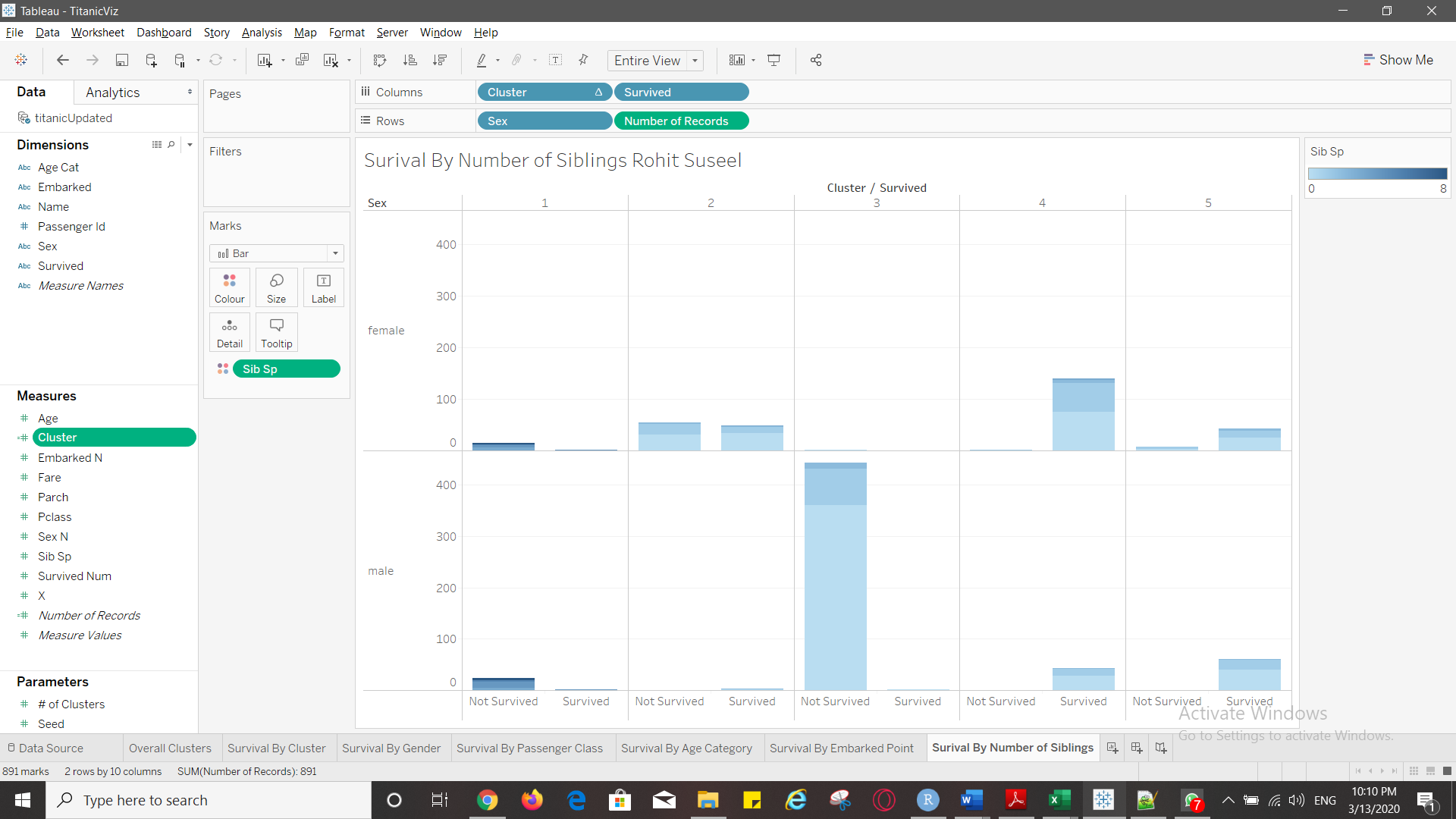
Q22) Paste your screenshot of the Tableau worksheet below. Ensure you have the title added with your name.



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.

|  |  |  |
| --- | --- | --- |
|  | 4 | 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 | 0 | 0 |

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



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.

|  |  |  |
| --- | --- | --- |
|  | 4 | 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 | 0 | 0 |

For survivability in this cluster 4, the passenger should be female travelling in class 1 belonging to the age group 17-32, who embarked from Southampton and has no siblings.

For survivability in this cluster 5, the passenger should be male travelling in class 3 belonging to the age group 17-32, who embarked from Southampton and has no siblings.