**Integrated Analysis using R and Tableau**

**Name:** Venkata Sarath Chandra Muktevi

**Date:** 07/05/2019

**1. Data Retrieval:**

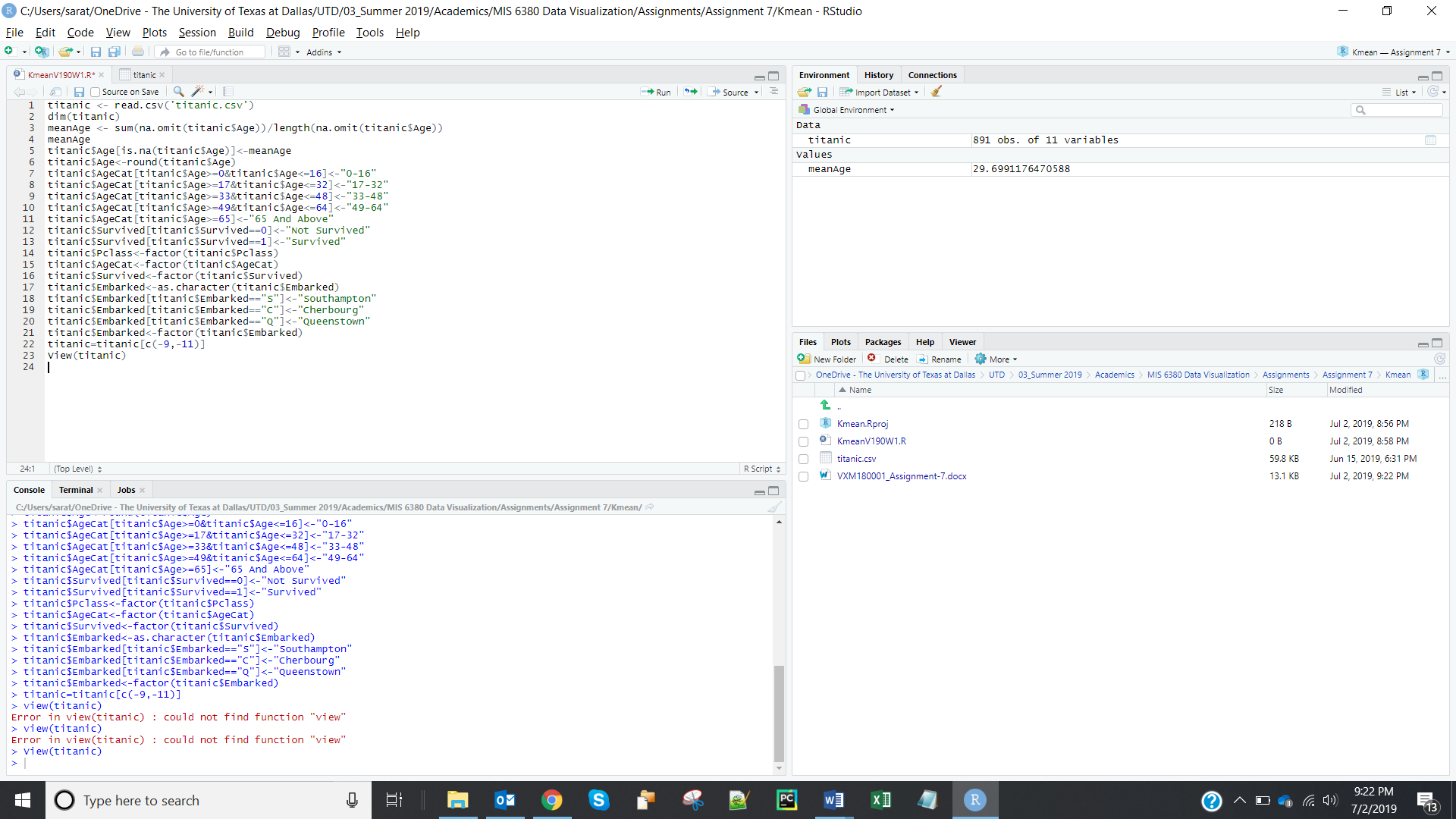
9. The output after executing dim(titanic) is:

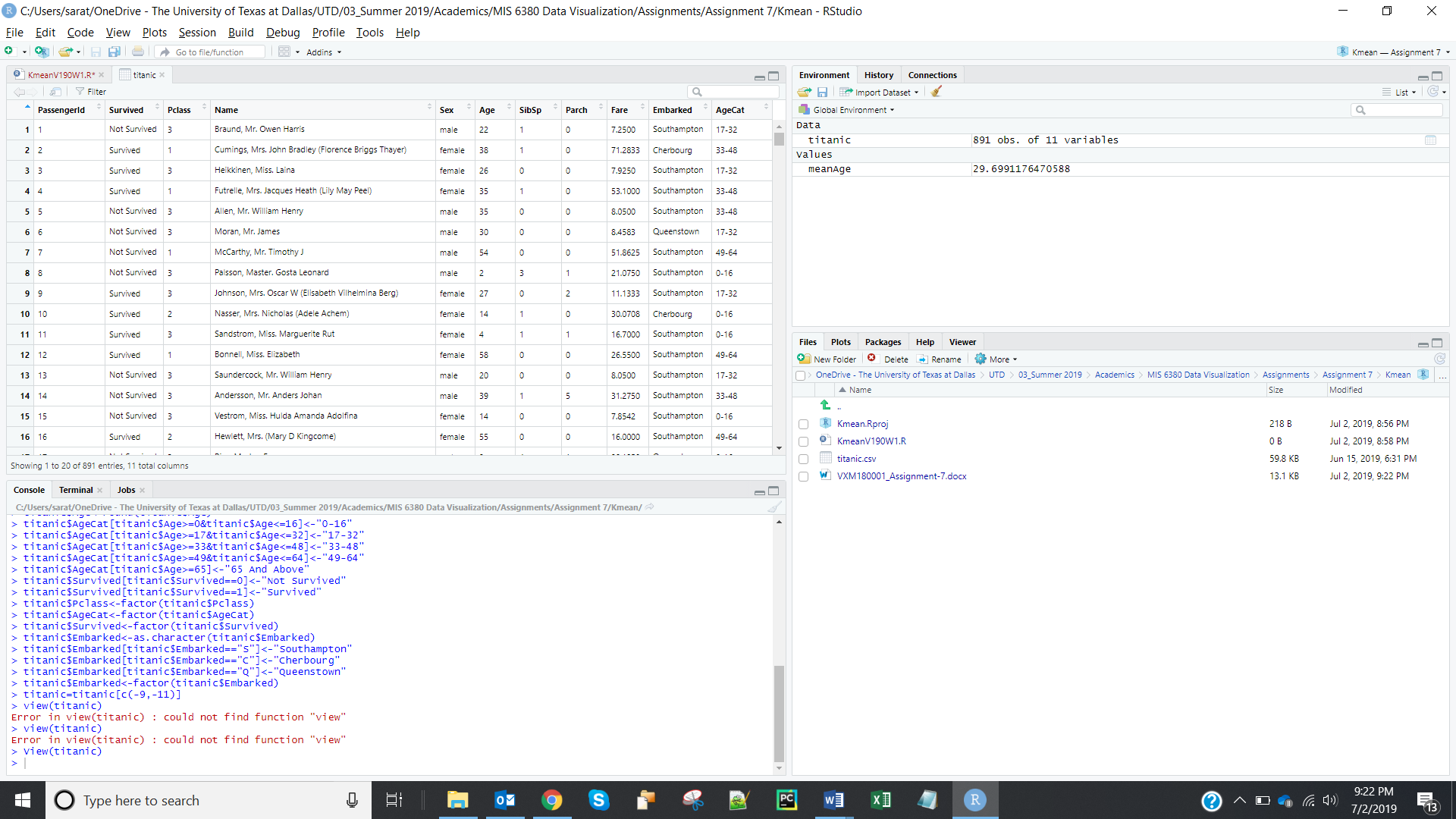
**891 12**

**2. Data Pre-processing:**

2. The average age is **29.69912**

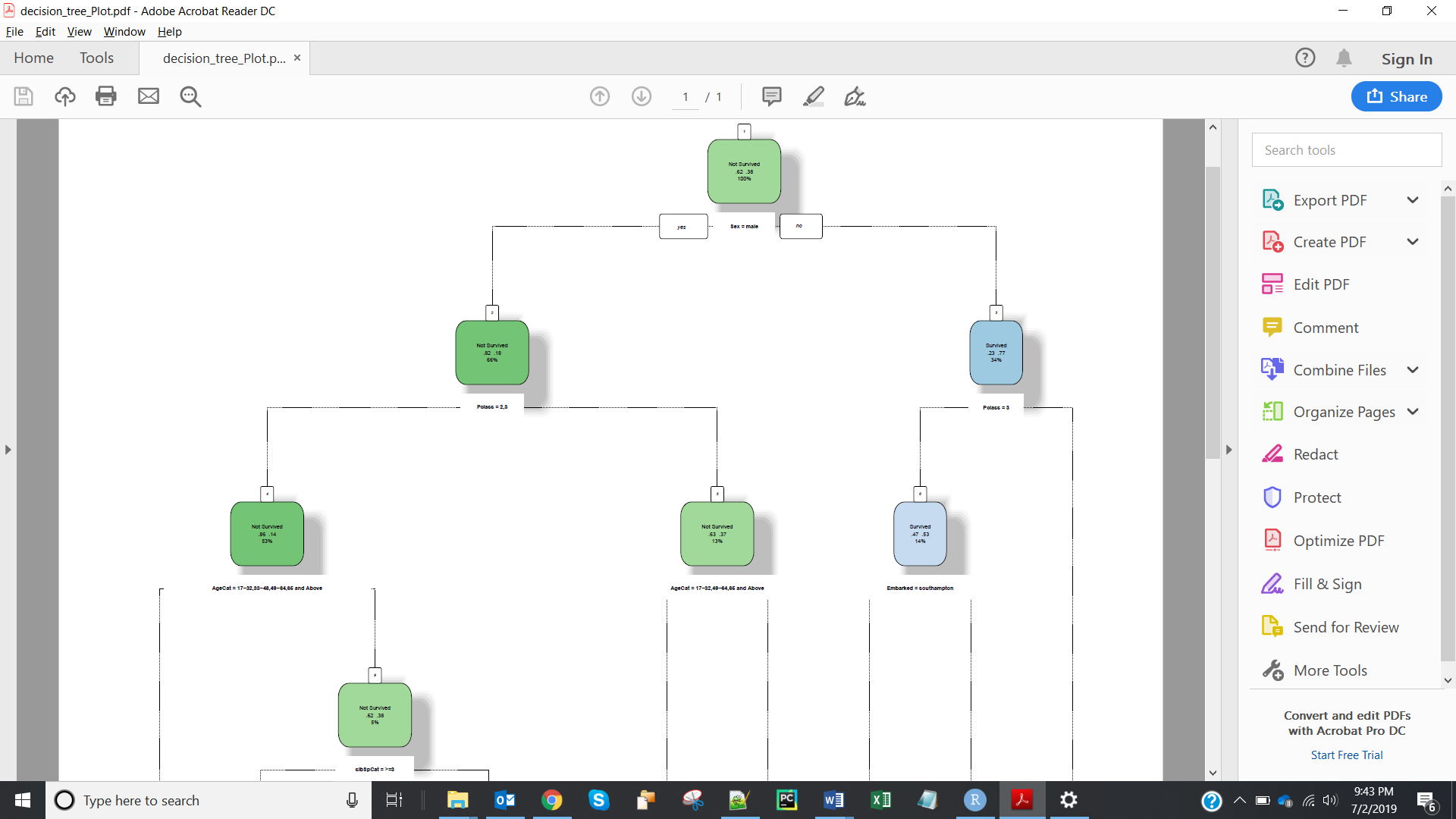
9. The screenshots after executing view command is:

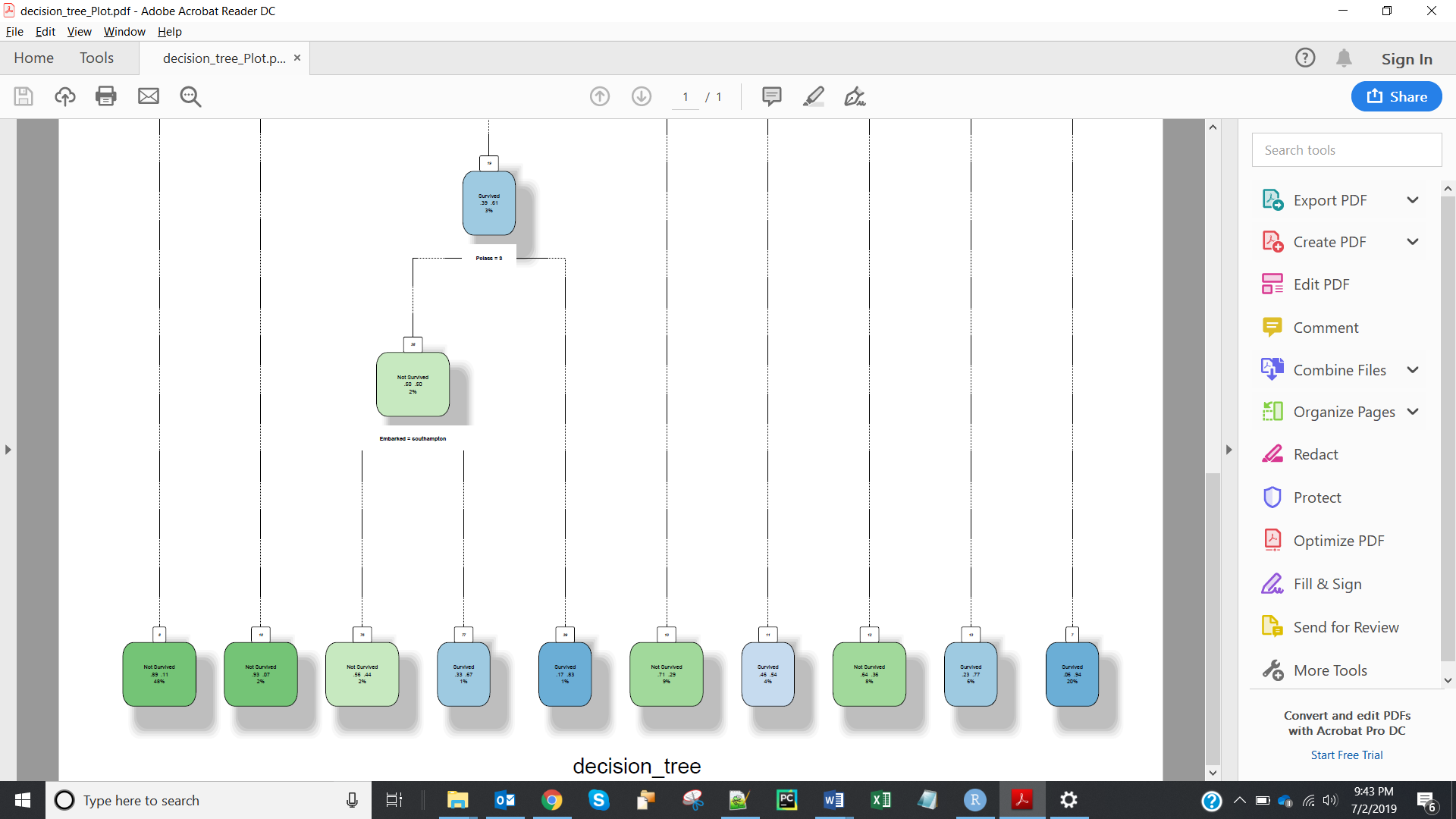




**3. Decision Tree:**

4. Screenshots of the decision tree is:





1) The misclassification rate for the current tree model is **0.22**

2) The first variable used for splitting is **Sex**

3) The ratio of Survived: Not survived initially is **38:62**

4)ratio of the Survived: not survived of Females is **77:23**

5)ratio of Survived: Not Survived of the males who are from Pclass 1 is **37:63**

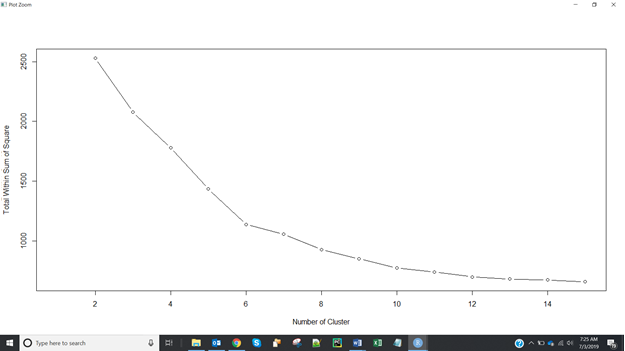
6) list of the top 6 variables from your decision tree in the order of importance are: **Survived, Sex, Pclass, Age Cat, SibSpCat, Embarked.**

**4. K-means Clustering:**

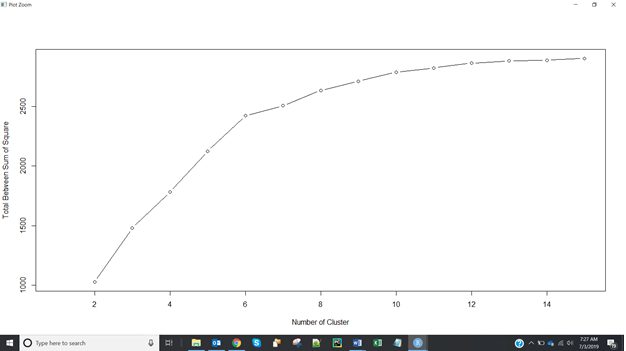
1. Key Variables.

3. The desired number of clusters I selected for the analysis is **6.**

4. Plot of Elbow Diagram- Within Cluster Sum of squares

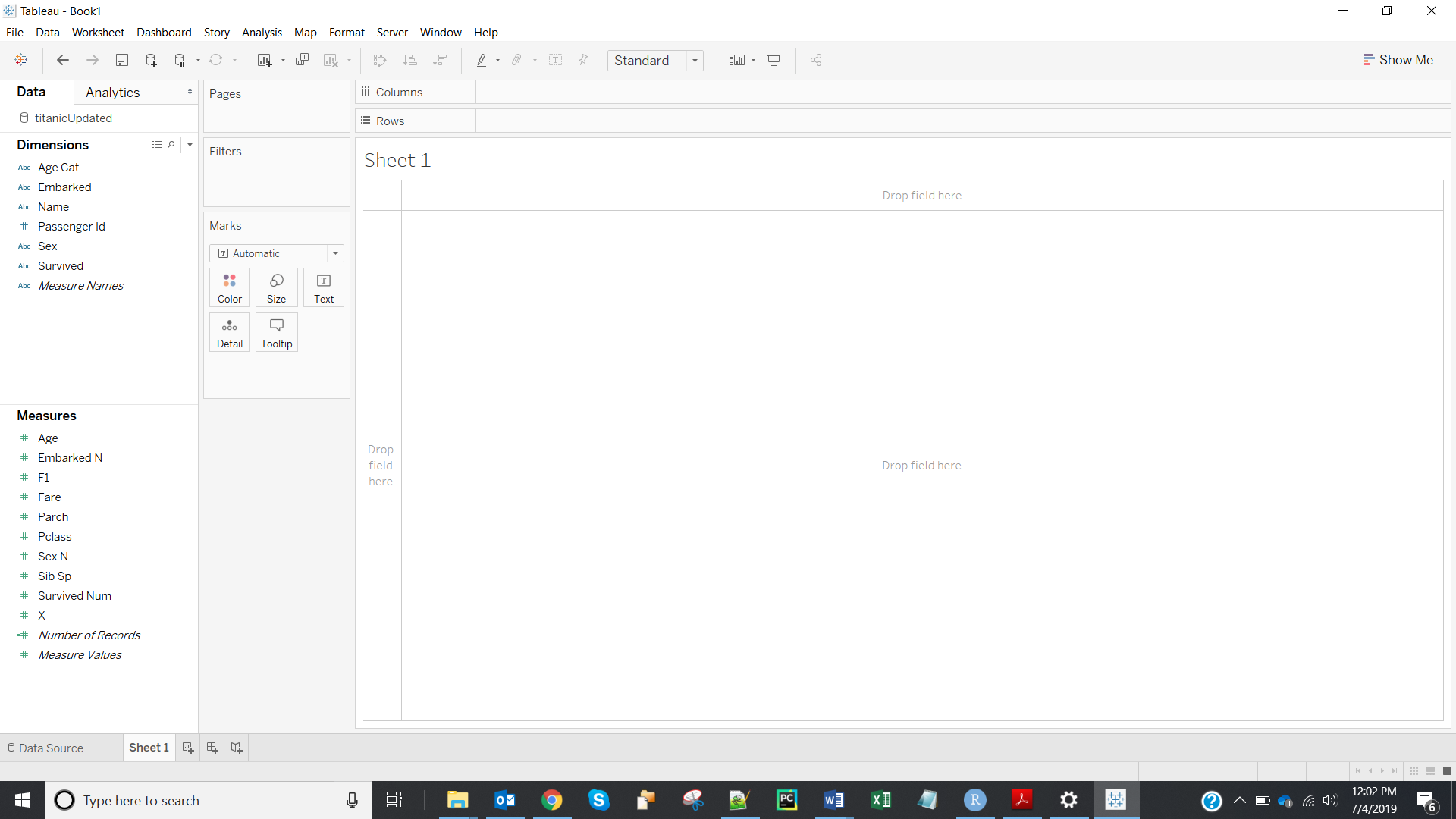


Plot of Elbow Diagram- Between Cluster Sum of squares:

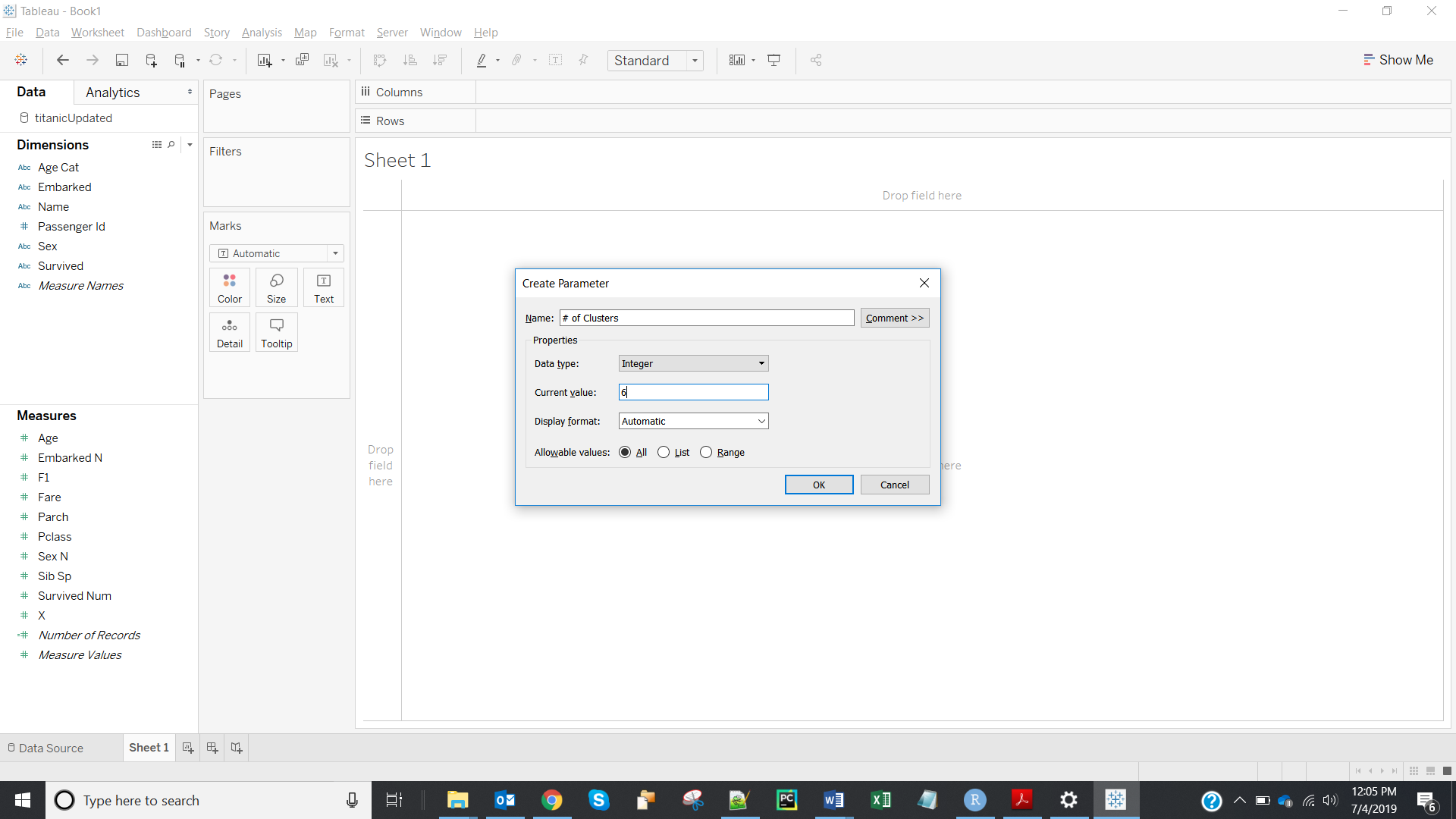


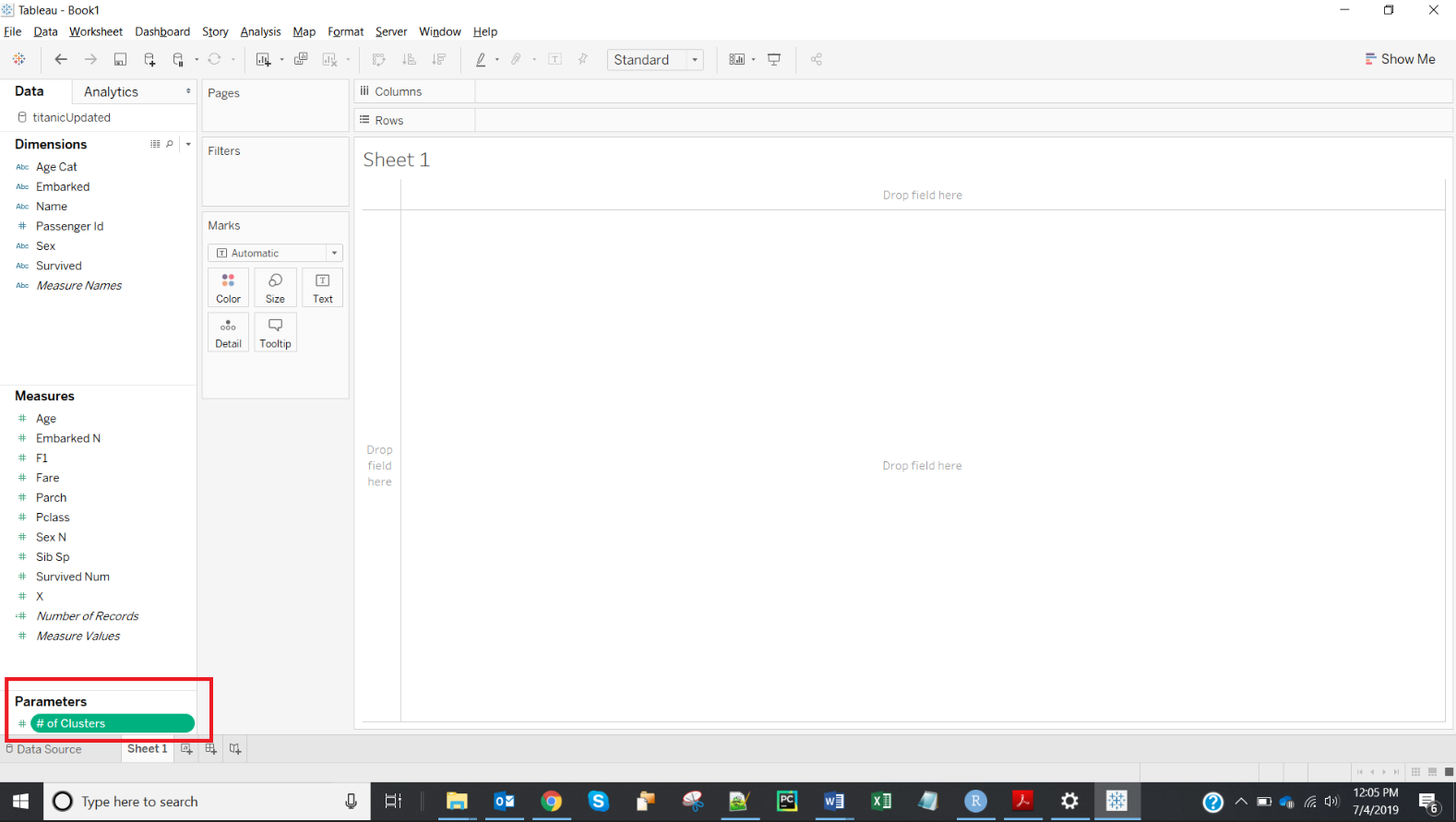
**5. Tableau/ R Integration:**

5. Screenshot of the current set of dimensions and measures is:

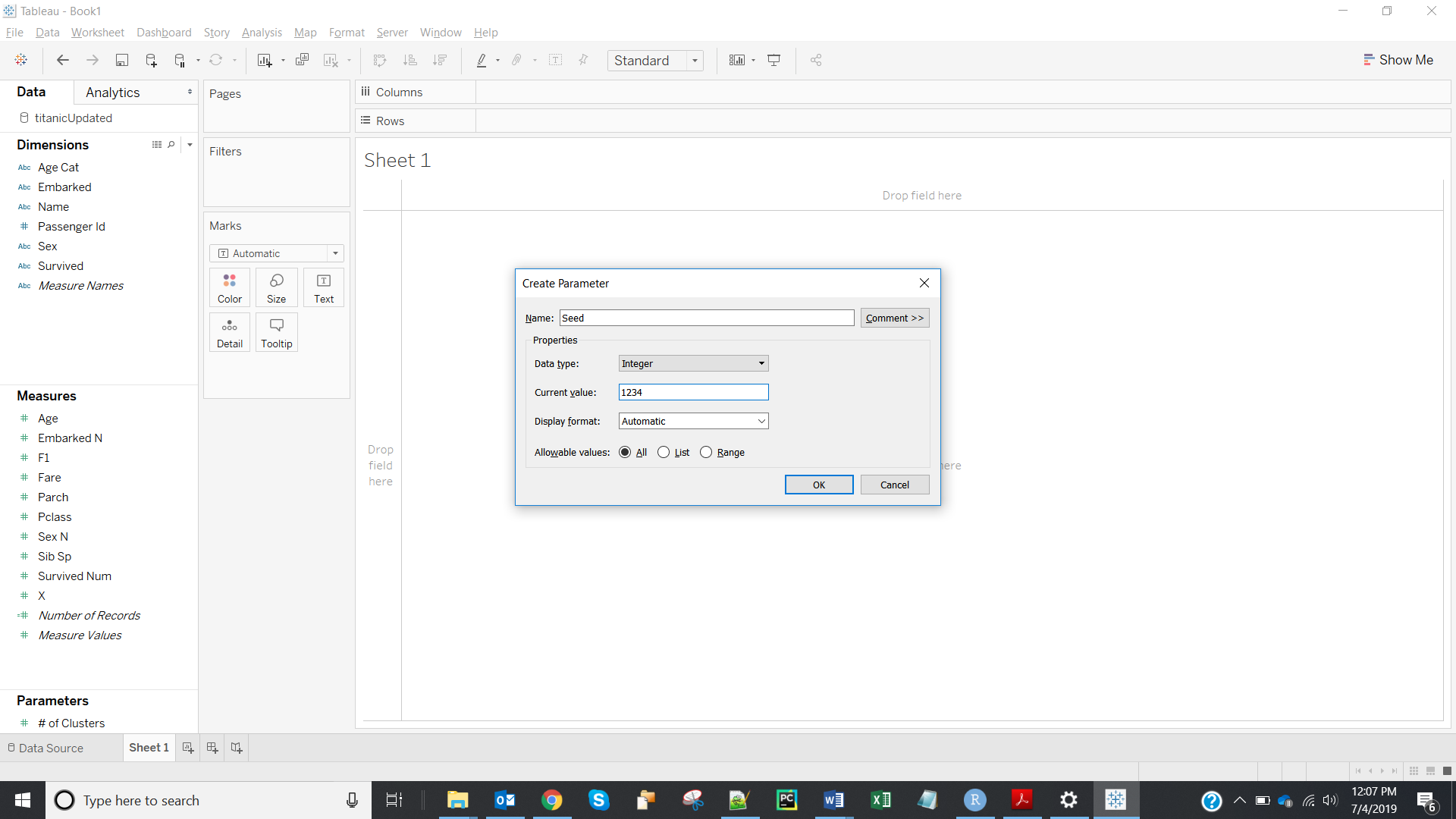


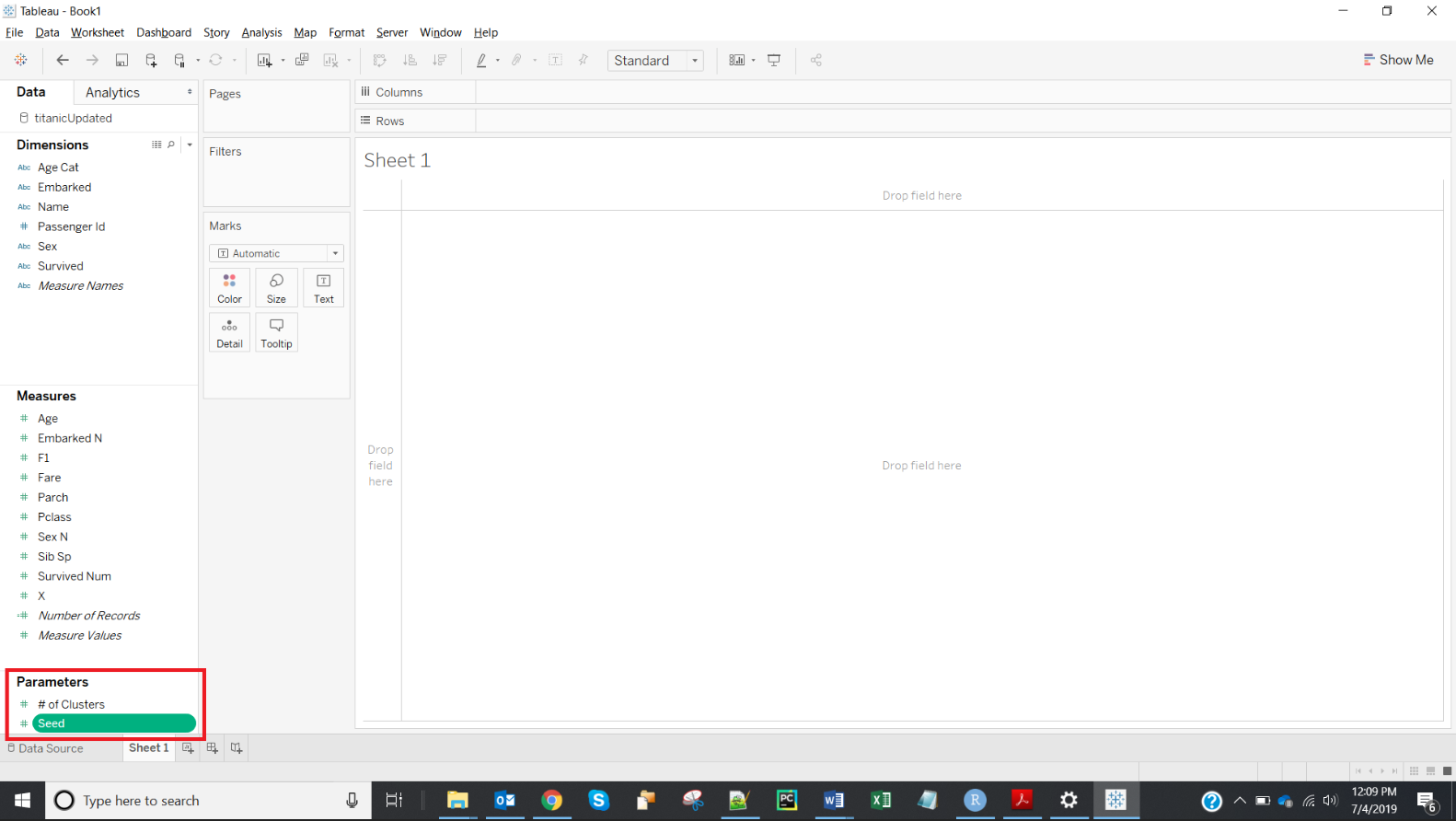
6. Screenshots after populating # of Clusters parameter is:



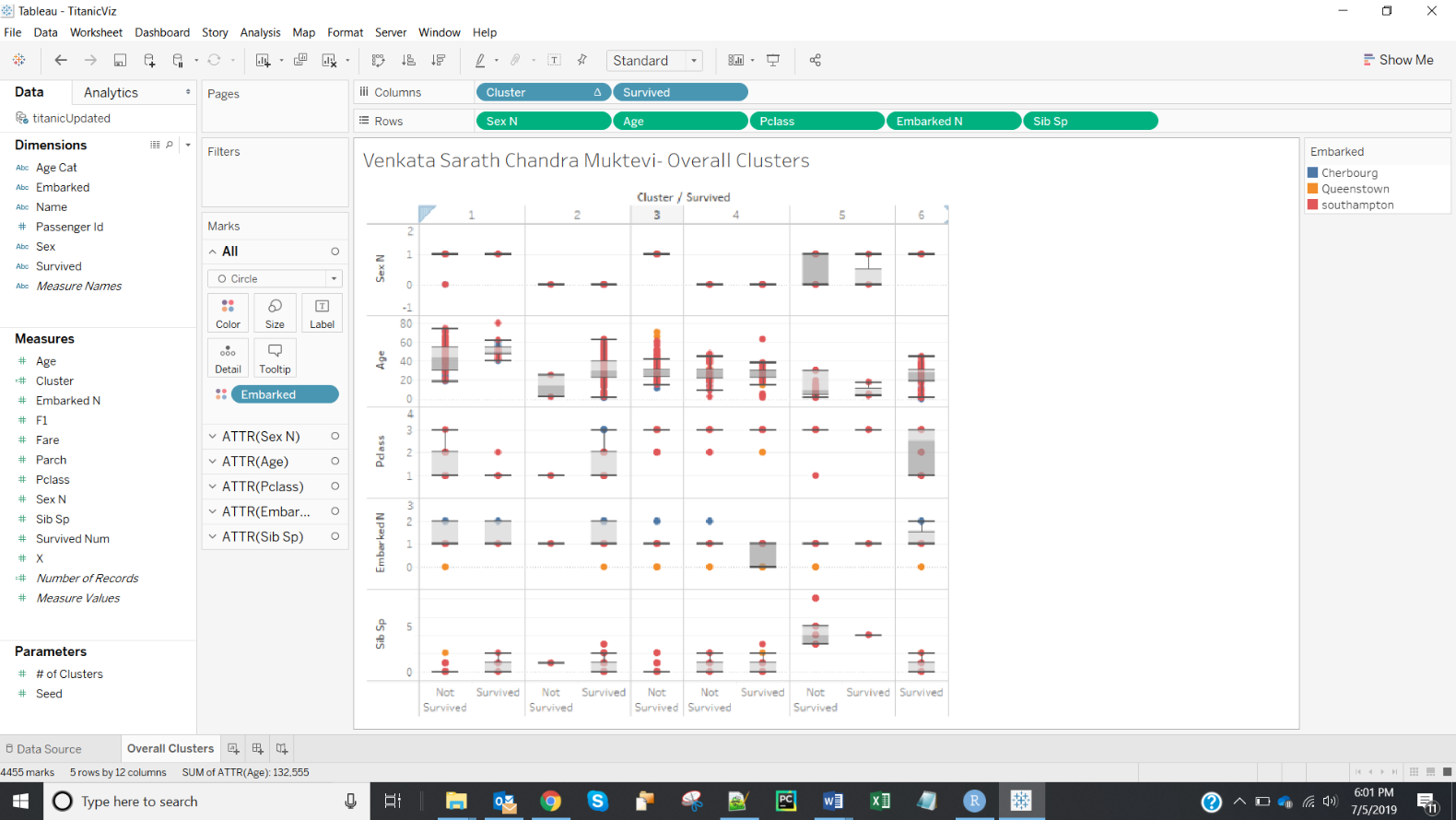


7. Screenshots for Seed parameter is:





8. Screenshot of ‘Overall Clusters’ is:



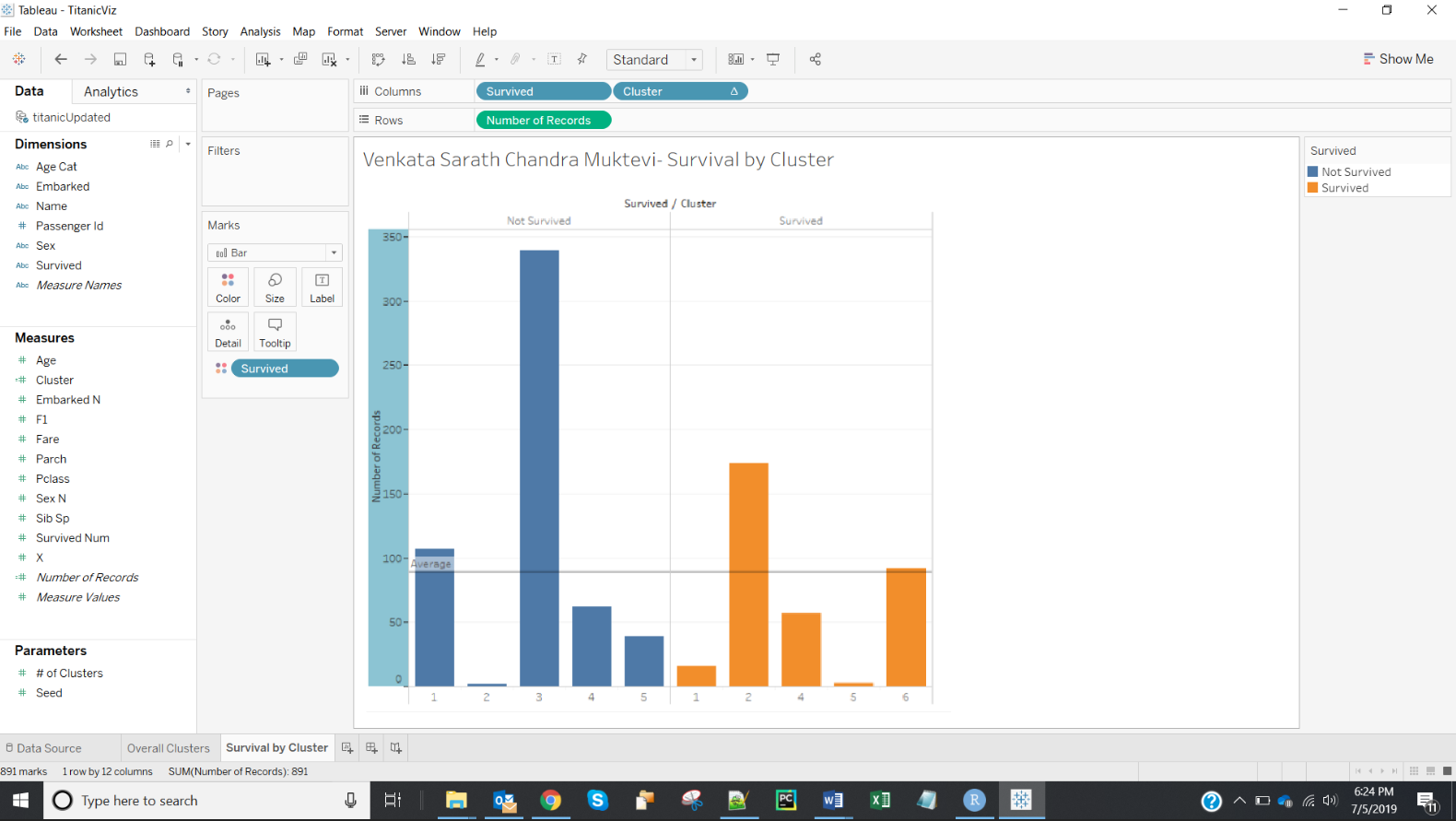
9. Insights from cluster 1:

There is no female survivor in this cluster. 50% of survivors in this cluster are from in between the age 50 – 65 with median around 55. All the survivors are from class 1 and 2. Most of the survivors have 0 or 1 siblings. All the survivors embarked from either Southampton or Cherbourg.

10. Insights from cluster 4:

This cluster contains only females. 50 percentage of the survivors and non-survivors in group are of age between 25 to 30. Passengers in this group are from class 2 and 3. All the survivors embarked from Queenstown or Cherbourg. All the passengers have 0 or 1 sibling.

11. Screenshot of ‘Survival by Cluster’ is:



12. Two clusters where survivability is the highest are: **Cluster 2 and Cluster 6**.

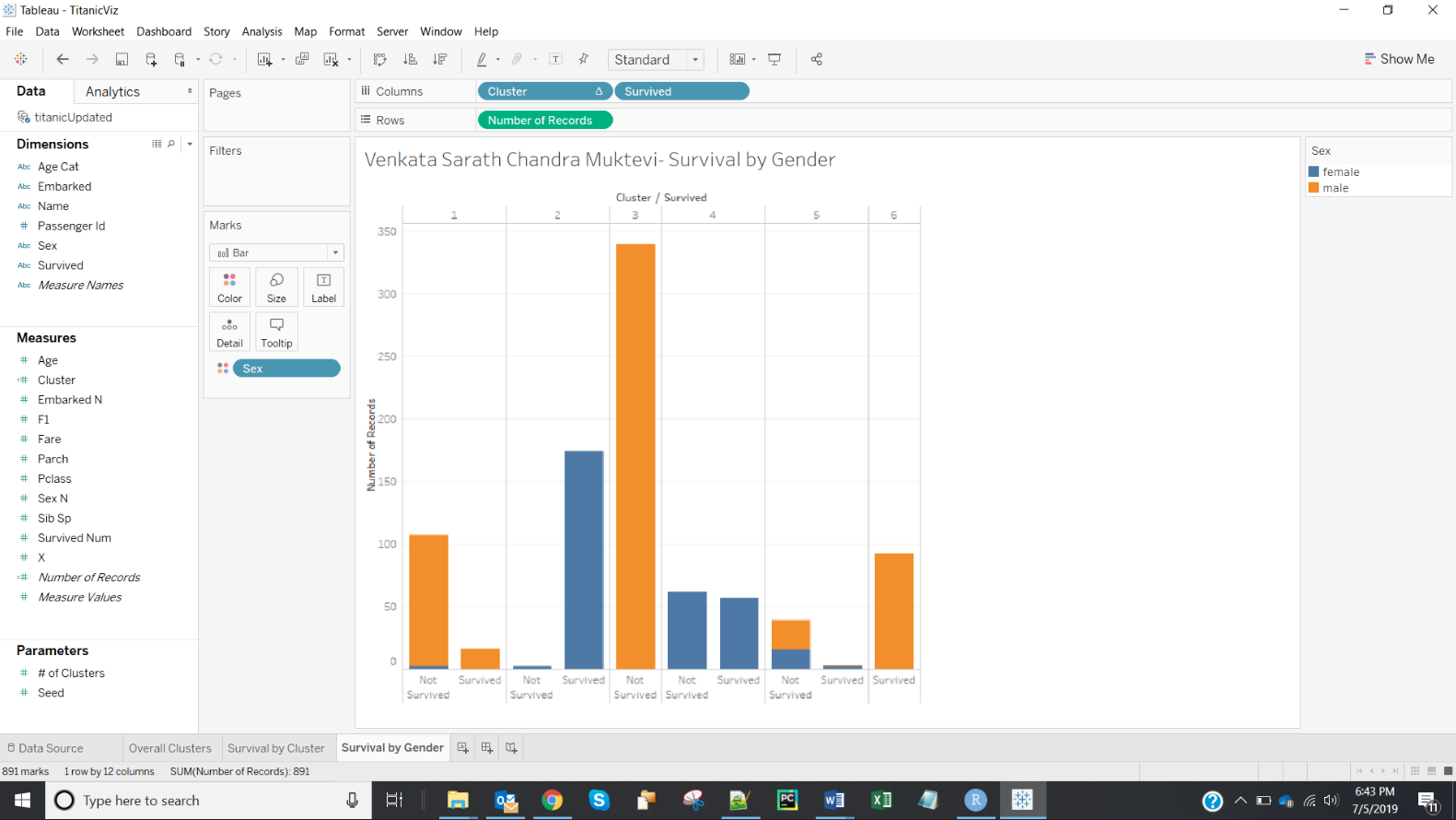
13.

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

14. The gender that has best chance to survive in cluster 2 and 6 is:

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

15. Screenshot of ‘Survival by Gender’ is:

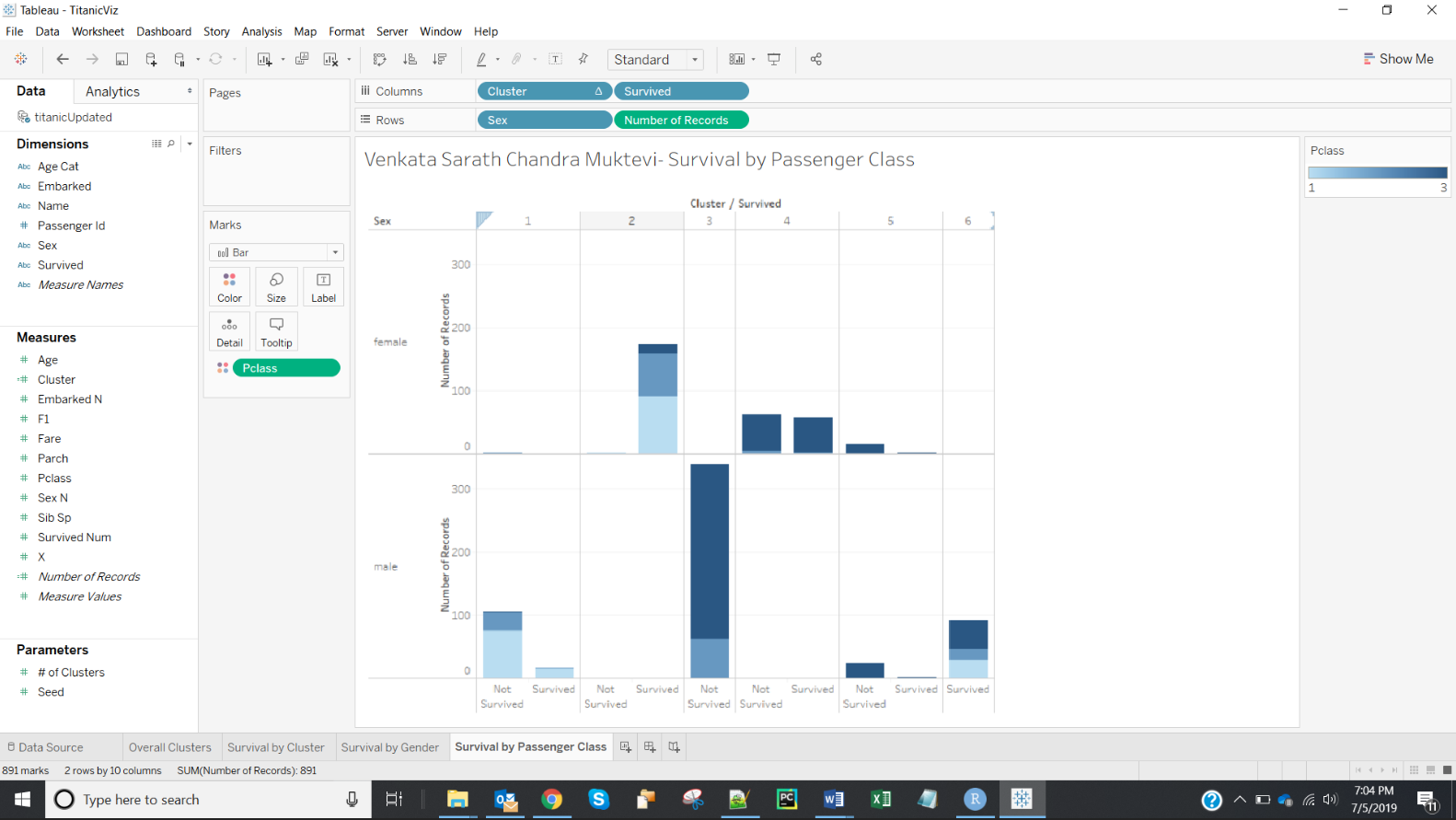


16. **Cluster 2** has more chance for females to survive.

17.

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

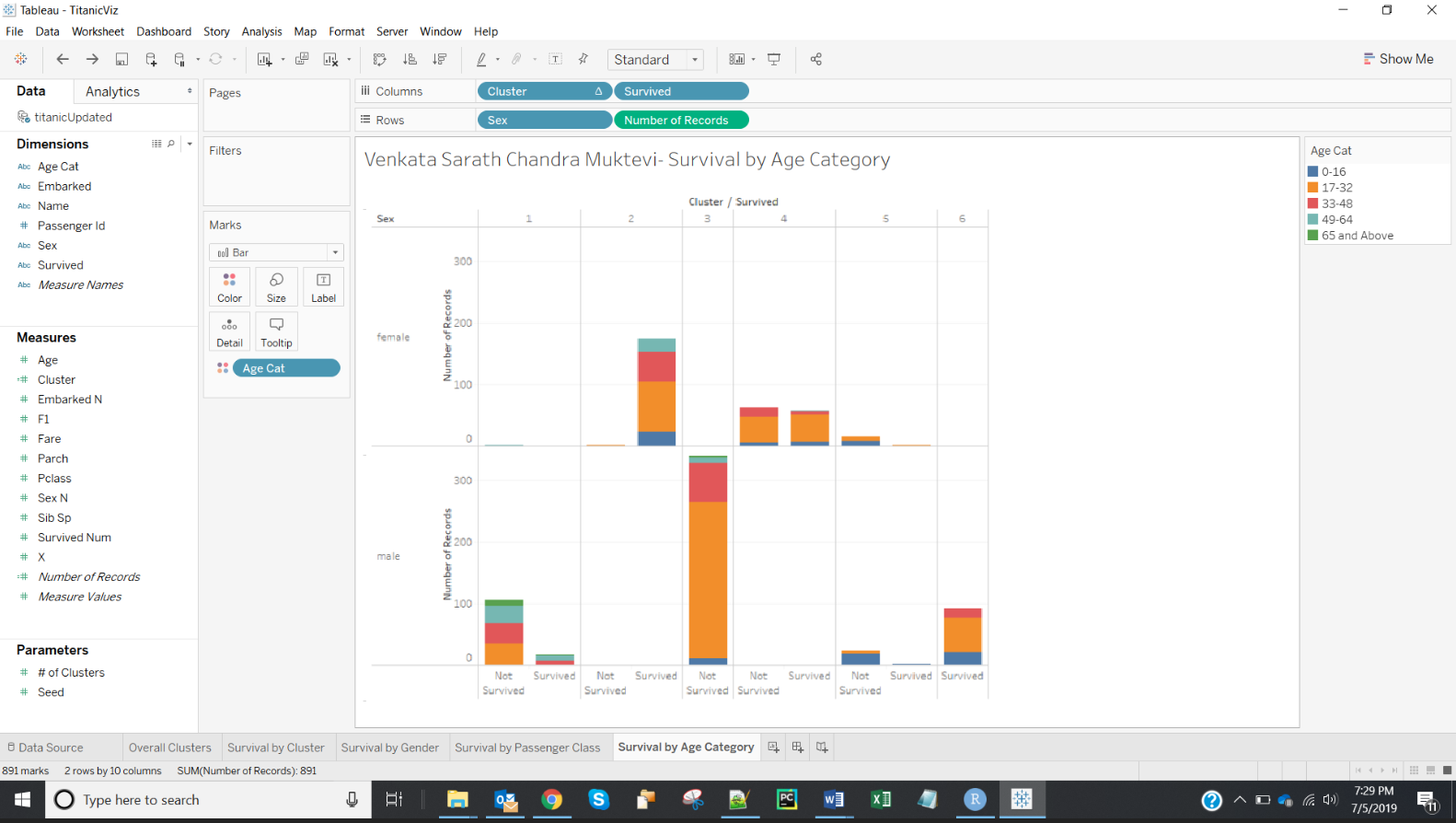
18. Screenshot of ‘Survival by Passenger Class’ is:



19.

|  |  |  |
| --- | --- | --- |
|  | 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 |  |  |

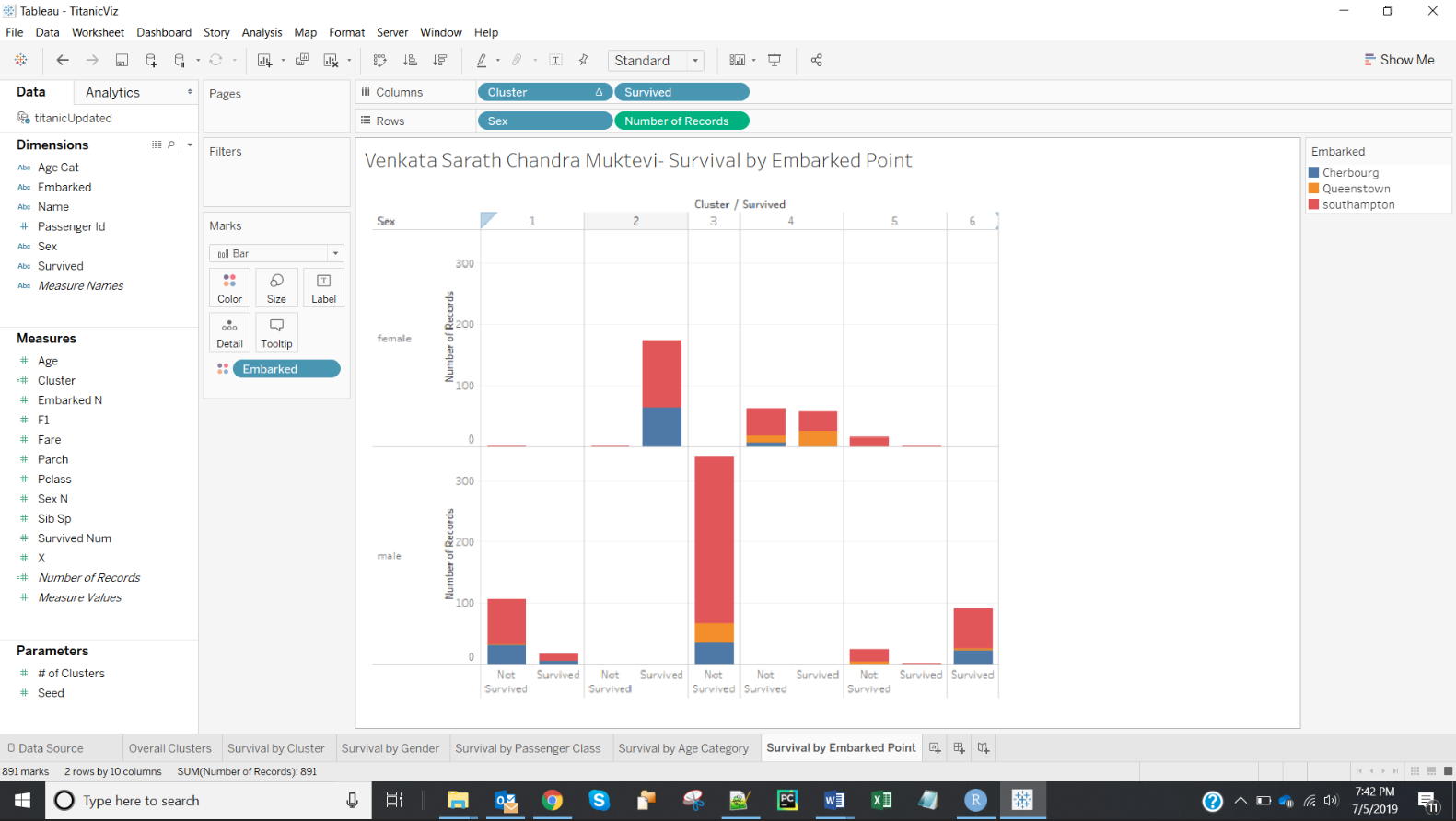
20. Screenshot of ‘Survival by Age Category’ is:



21.

|  |  |  |
| --- | --- | --- |
|  | 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 |  |  |

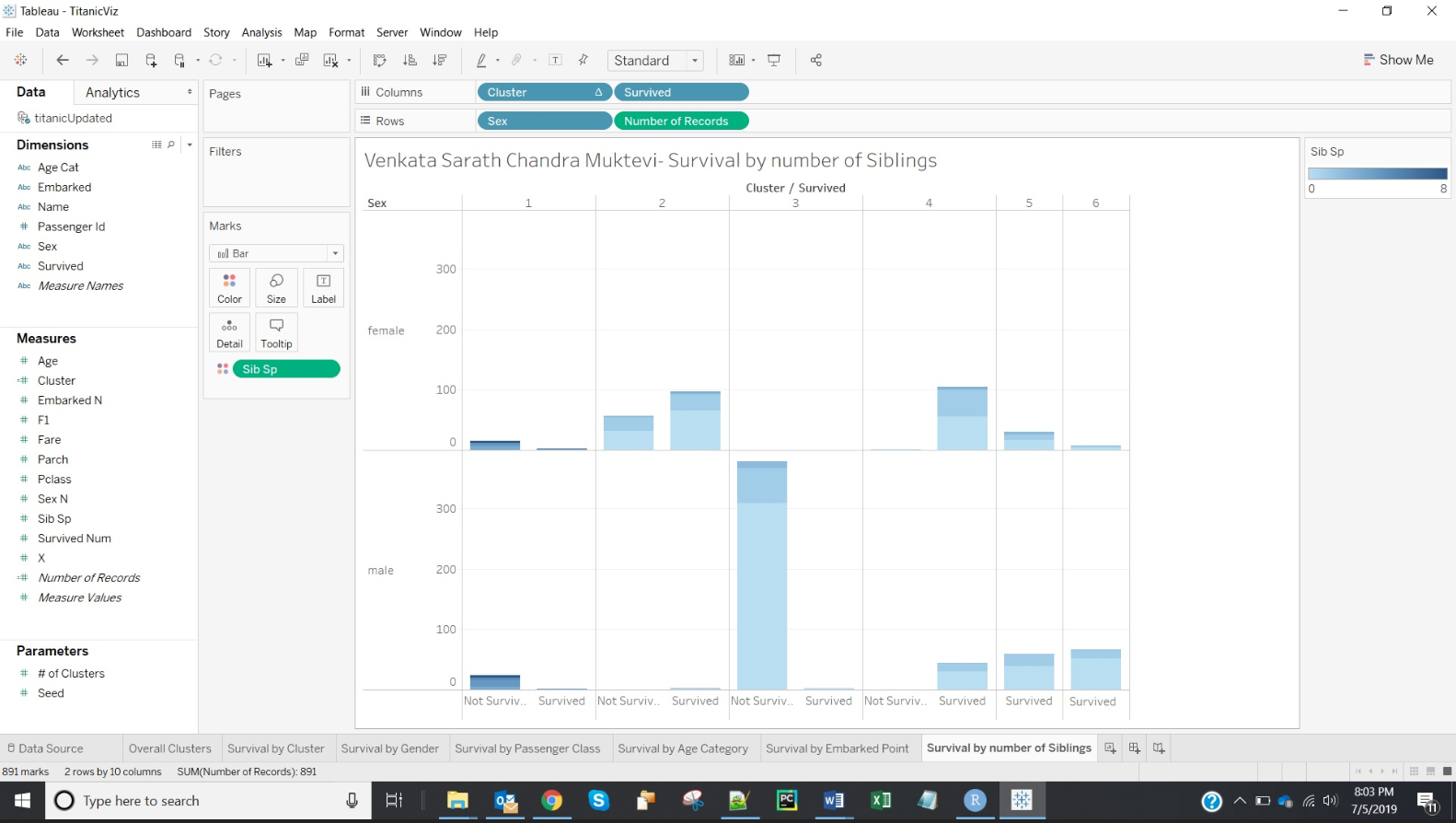
22. Screenshot of ‘Survival by Embarked Point’ is:



23.

|  |  |  |
| --- | --- | --- |
|  | 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** |

24. Screenshot of ‘Survival by number of Siblings’ is:



22. **The profile of passengers from cluster 2 is:** The passengers are female, embarked point is Southampton, travelling in passenger class Pclass 1, aged between 17 and 32 years, having number of siblings as 0 have most chance to survive.

**The profile of passengers from cluster 6 is:** The passengers are male, embarked point is Southampton, travelling in passenger class Pclass 3, aged between 17 and 32 years, having number of siblings as 0 have better chance to survive.