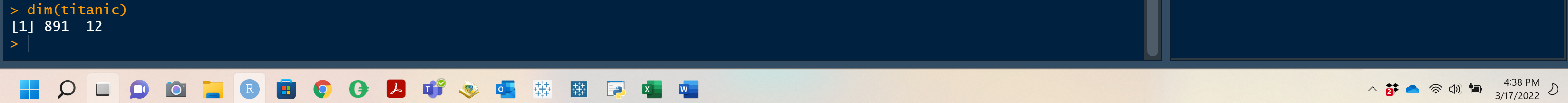
Exercise 7

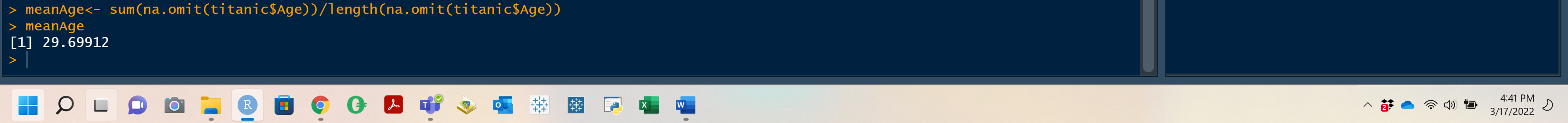
**DATA RETRIEVAL**

9)



**DATA PRE-PROCESSING**

**2)**

****

**9)**

**Graphical user interface, text, application

Description automatically generated**

**3) DECISION TREE**

Decision Tree Plot:

Diagram

Description automatically generated

1)*Misclassification rate:*

17.508%

**Diagram

Description automatically generated**2)*First Variable Used For Splitting*:

**Sex = Male**

3)Ratio of Survived:Not Survived Initially

**38:62**

*Diagram

Description automatically generated*4)*For Females: Ratio of Survived: Not Survived*

**74:26**

Diagram

Description automatically generated5) For Pclass1 Males: Ratio of Survived: Not Survived

**40:60**

6)*Top Variables from the decision tree for* ***Survival*** *are:*

**Sex, Pclass, AgeCat, SibSpCat, Embarked**

**ParchCat does not show up in the decision tree with minsplit values set to 10 but would show up if minsplit was 4**

**4) K-MEANS CLUSTERING**

1) You should focus on just the **Key Variable** for clustering

3) The elbow appears to begin forming at 5

4)

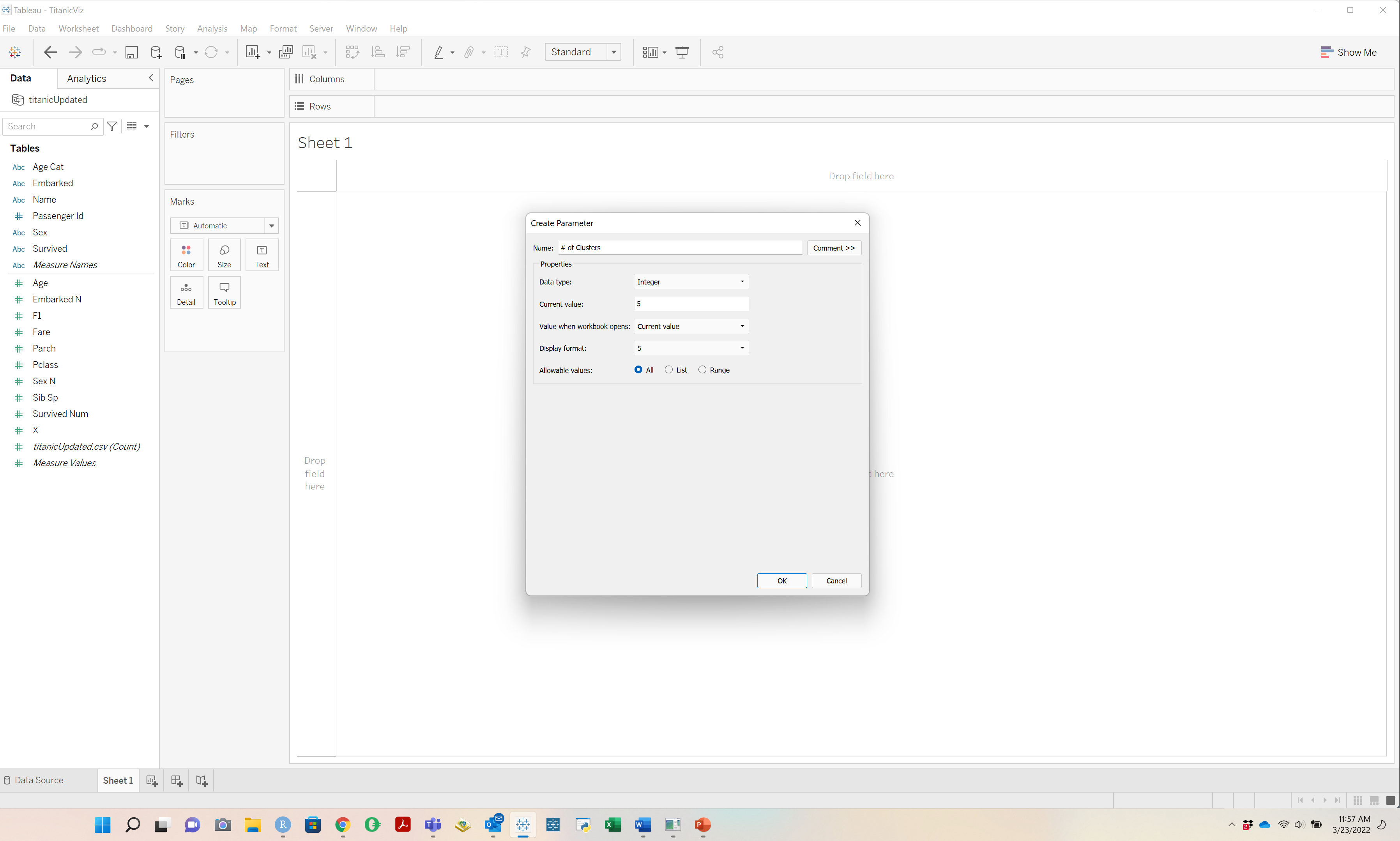
Graphical user interface, chart

Description automatically generated

5) Tableau / R INTEGRATION

Graphical user interface, application, Word

Description automatically generated



Graphical user interface, application

Description automatically generated

Q8:

A picture containing chart

Description automatically generated

9) Insights from Cluster 1: Most passengers embarked in SouthHampton and most of the surviving passengers were from Southhampton as well with no survivors from Cherbourg. The nonsurvivors were typically between the ages of 23 and 31 with several outliers while survivers were typically between the age of 8.5 and 30. Nobody over the age of 45 survived from this cluster. They were all the same sex, male, and noone was a first class passenger in this cluster. No passengers in this cluster had more than two Sibblings.

10)Insights from Cluster 4: Nobody from Queenstown was in this cluster. Most passengers were from Cherbourg and most suvived. Survivors could be found from all three passenger classes, but the only passengers who didn’t survive in this cluster were 3rd class passengers. The typical age of surviving passengers was 18.5 to 35 in this cluster, but there are many passengers who suvived outside of this age group. Nobody in this cluster was from Queenstown. Both sexes were represented in survival but all of the males in this cluster survived. Any passenger who had two sibblings also survived in this cluster.

Q 11)Chart, bar chart

Description automatically generated

Q 12) The two clusters with the highest survivability are clusters 4 & 5.

|  |  |  |
| --- | --- | --- |
|  | Cluster 4 | Cluster 5 |
| Ideal Gender | Female | Male |
| Ideal Passenger Class | 1 | 3 |
| Ideal Age Category | 17-32  (but virtually every age group survived in this cluster) | 17-32  (but all age groups survived) |
| Ideal Embarked point | Southampton  (but every Cherbourg embarkment survived as well) | Southampton  (but every Cherbourg embarkment survived as well) |
| Ideal number of siblings | 0 (and 1 is very close) | 0 (and 1 although not as close of a ratio as cluster 4) |

Q 15)

Chart, bar chart

Description automatically generated

Q 16) If you are female, you want to belong to Cluster 4 in order to have the highest probability of survival.

Q18)

Chart, waterfall chart

Description automatically generated

Q20)

Chart, waterfall chart

Description automatically generated

Q22)

Chart, waterfall chart

Description automatically generated

Q24)

Chart, waterfall chart

Description automatically generated

|  |  |  |
| --- | --- | --- |
|  | **Cluster 4** | **Cluster 5** |
| Ideal Gender | **Female** | **Male** |
| Ideal Passenger Class | **1** | **3** |
| Ideal Age Category | **17-32**  *(but virtually every age group survived in this cluster)* | **17-32**  *(but all age groups survived)* |
| Ideal Embarked point | **Southampton**  (but every Cherbourg embarkment survived as well) | **Southampton**  *(but every Cherbourg embarkment survived as well)* |
| Ideal number of siblings | **0**  *(and 1 is very close)* | **0**  *(and 1 although not as close of a ratio as cluster 4)* |

Q25)

Cluster 4: The ideal survival passenger was a 1st class female between the ages of 17 and 32 embarking from Southhampton and having 0 siblings.

Cluster 5: The ideal survival passenger was a 3rd class male between the ages of 17 and 32 embarking from Southampton and having 0 siblings.