

Answer 1) 891 12

(the dataset has 891 observations of 12 variables)

Answer 2) 29.69912 (Average Age)

Answer 3)

The screenshot shows the RStudio interface with the following components:

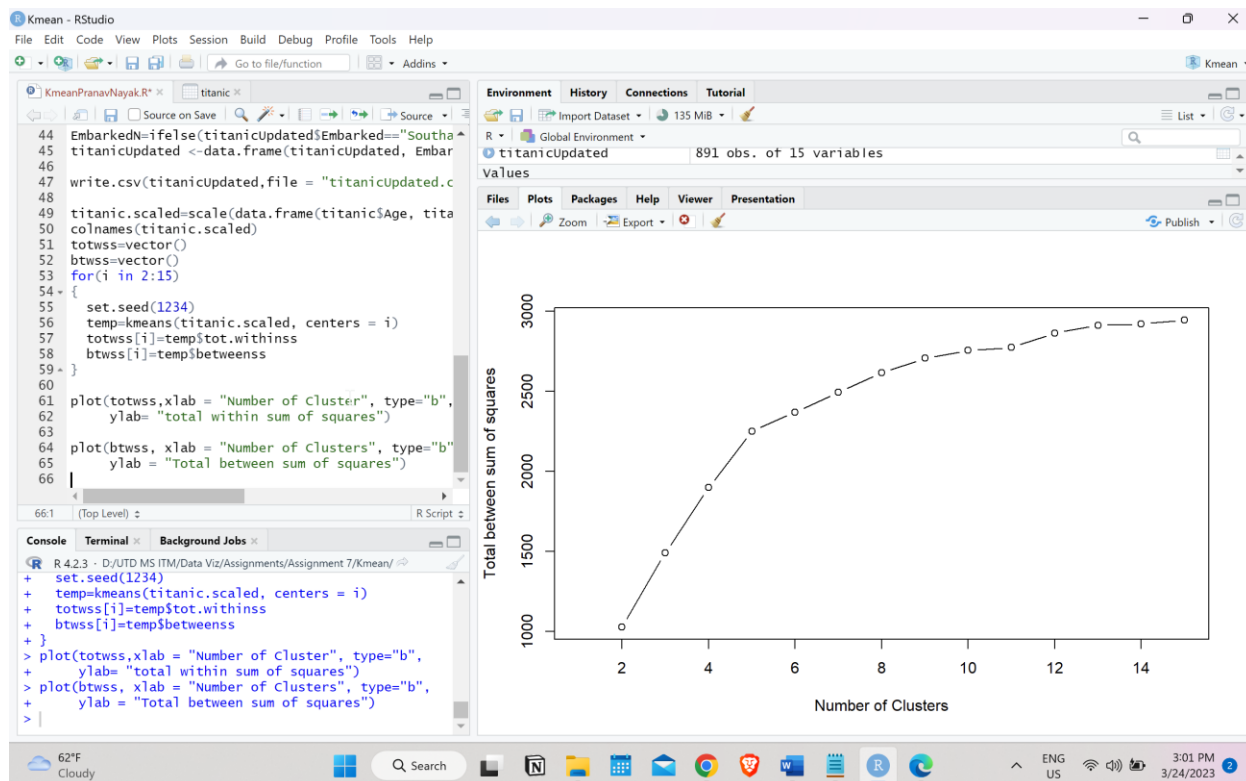
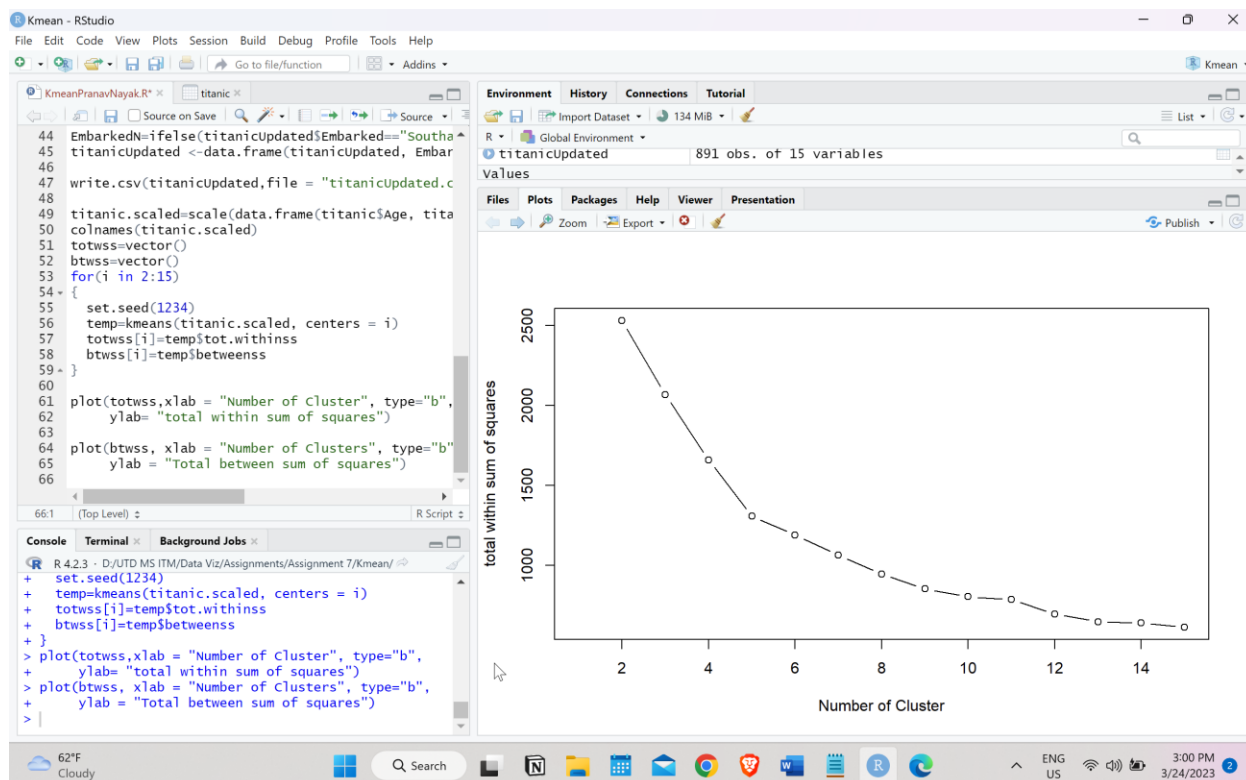
- Environment Pane:** Shows the 'titanic' dataset with 891 observations and 11 variables. The 'meanage' value is 29.6991176470588.
- Console:** Contains the following R code:
 

```

> titanic$Survived[titanic$Survived==1]="Survived"
> titanic$Pclass=factor(titanic$Pclass)
> titanic$AgeCat=factor(titanic$AgeCat)
> titanic$Survived=factor(titanic$Survived)
> titanic$Embarked=as.character(titanic$Embarked)
> titanic$Embarked[titanic$Embarked=="S"]="Southampton"
> titanic$Embarked[titanic$Embarked=="C"]="Cherbourg"
> titanic$Embarked[titanic$Embarked=="Q"]="Queenstown"
> titanic$Embarked=factor(titanic$Embarked)
> titanic=titanic[c(-9,-11)]
> View(titanic)
      
```
- Files Pane:** Shows the project files: 'Kmean.Rproj' (218 B), 'KmeanPranavNayak.R' (0 B), 'Kmeans code.R' (3.1 KB), and 'titanic.csv' (58.9 KB).

SSS

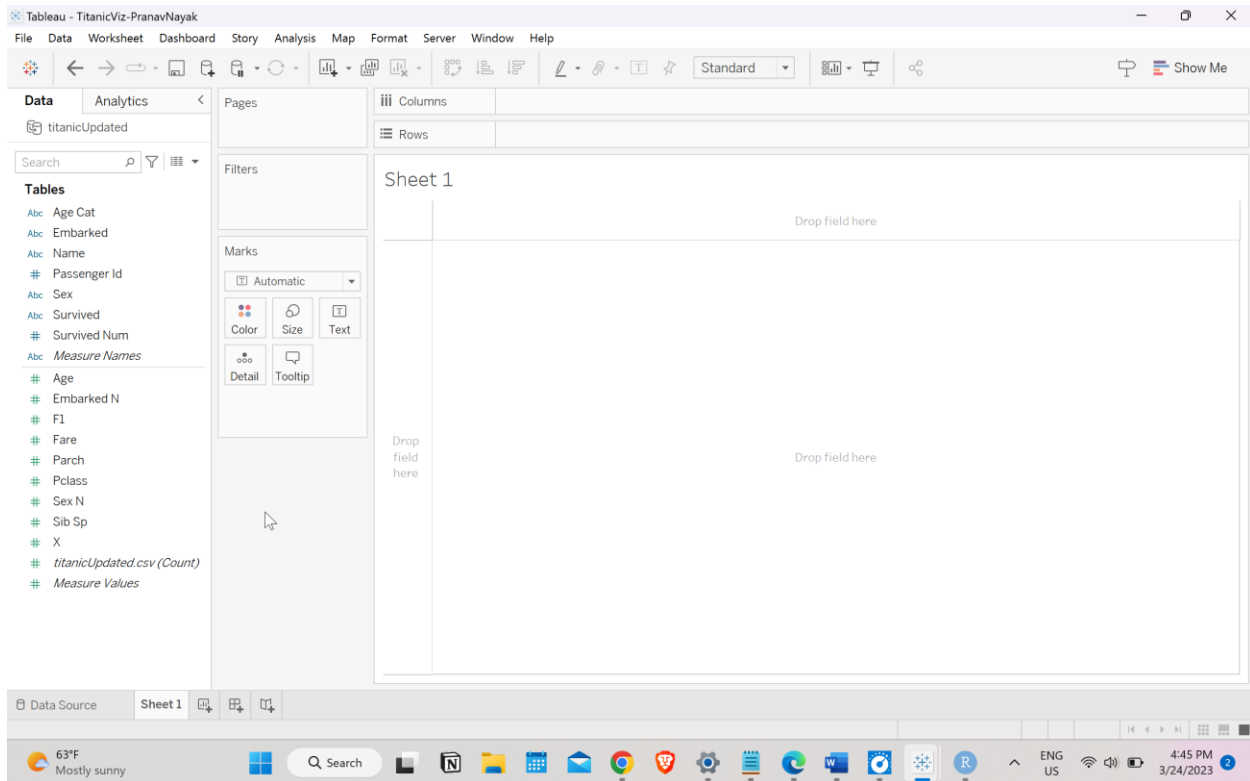
Answer 4)



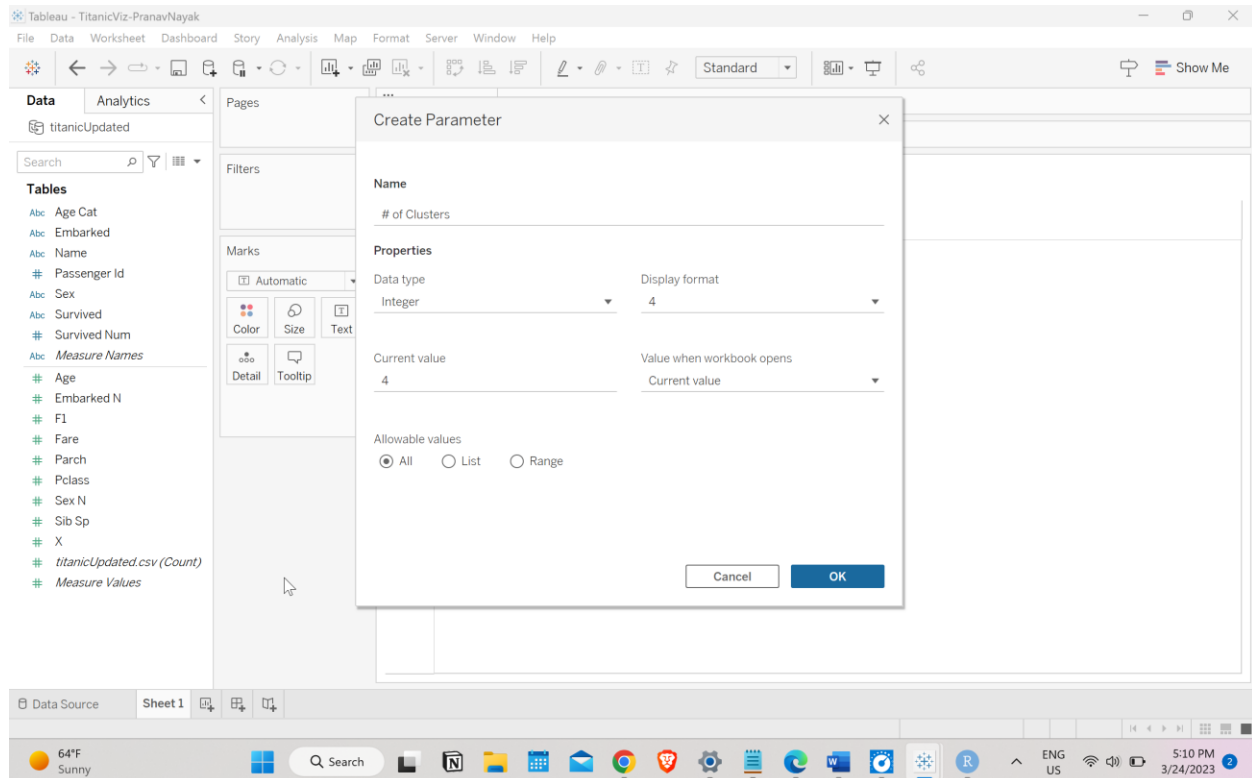
Answer 5)

Through the elbow method I have chosen 5 clusters

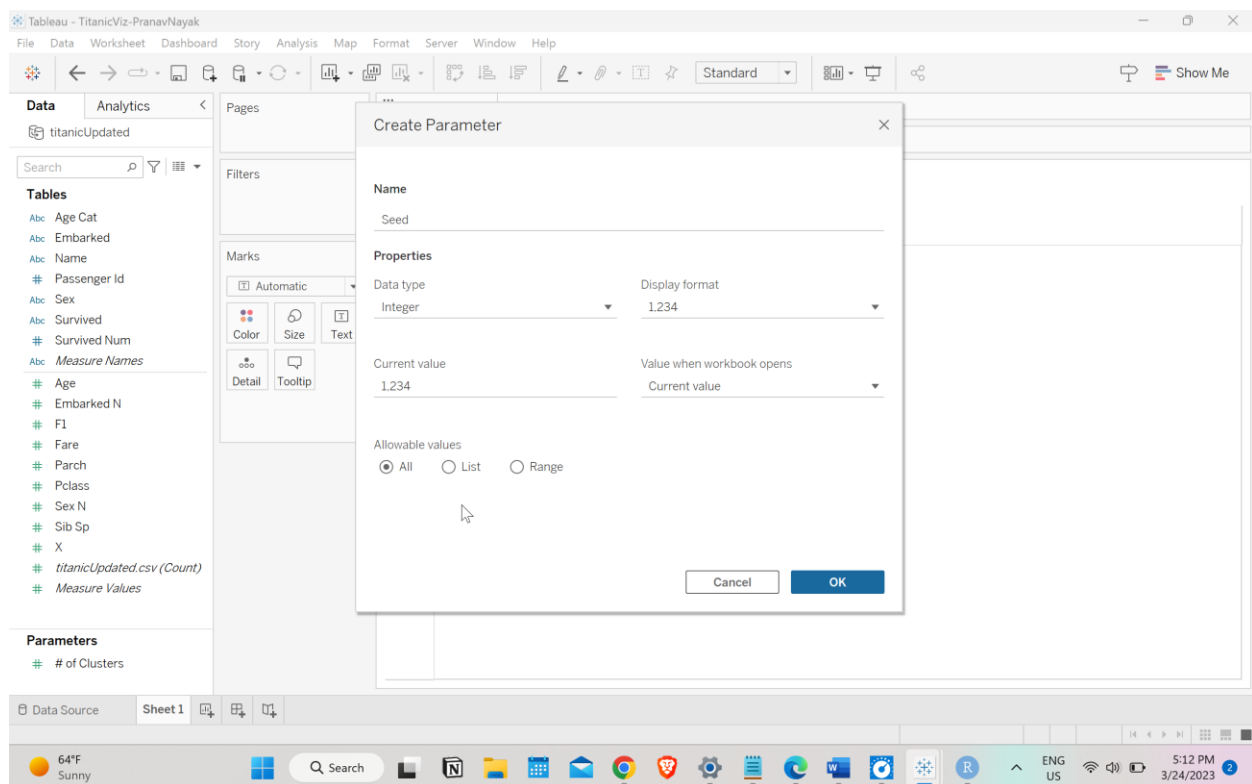
Answer 6)



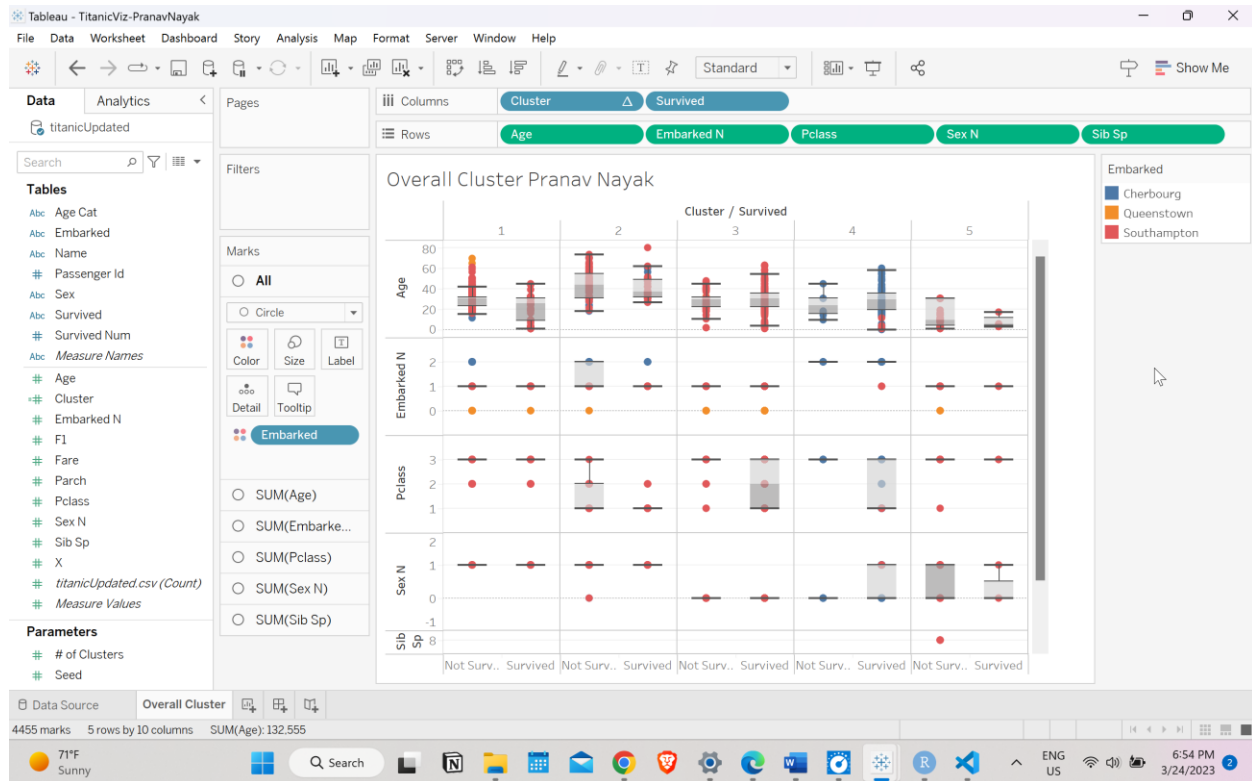
Answer 7)



## Answer 8)



## Answer 9)



## Answer 10)

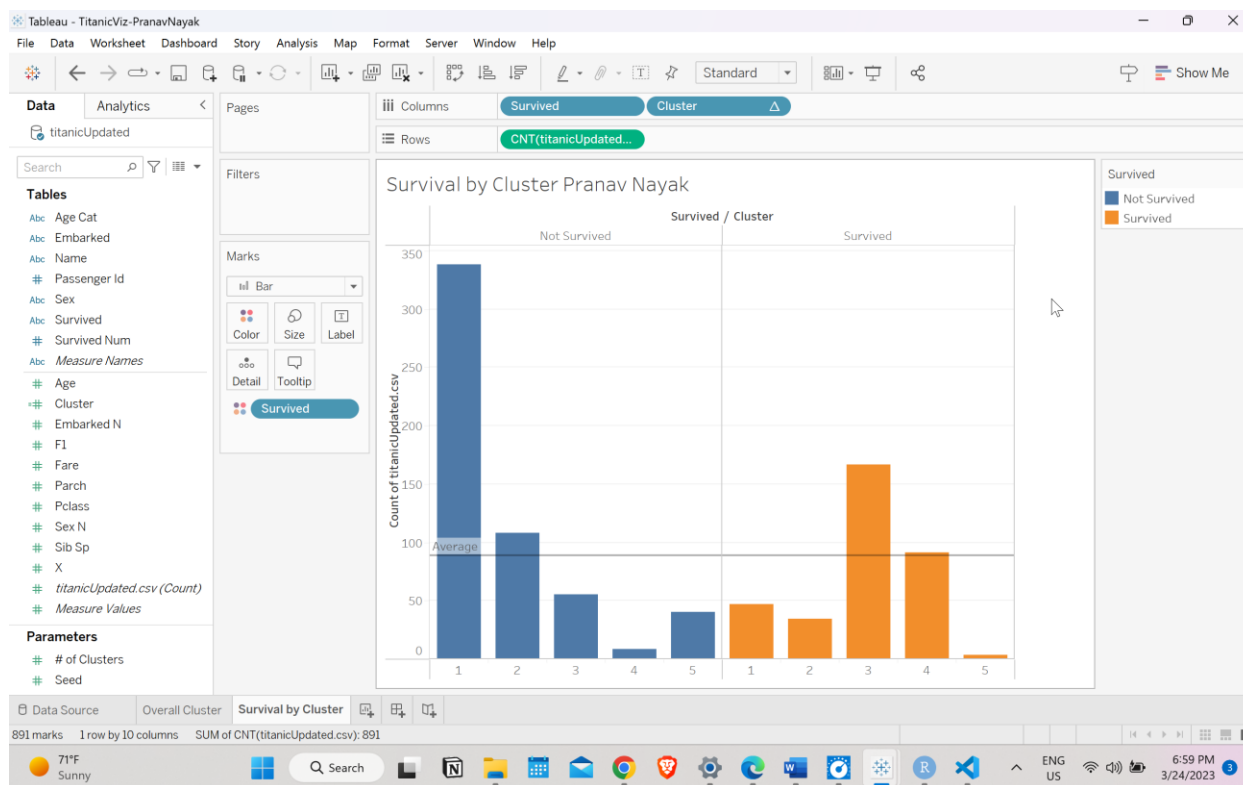
- From cluster 1 it can be observed that most people who are in the “not survived” section are in range of 11 to 41 with some outliers. Also, the people who survived had an age range of 1 to 45.
- People who survived mainly from Southampton and people who did not survive were from Southampton and Queenstown.
- Passengers from classes 2 and 3 had similar chances of both surviving and not surviving.
- the passengers who belonged to “not survived” category embarked from 3 ports namely Cherbourg, Southampton, and Queenstown while those who survived are from Southampton and Queenstown

## Answer 11)

- Most passengers that survived belonged to Cherbourg and Southampton and people who did not survive were mainly from Chebourg.
- In Cluster 4 the age distribution of people who did “not survived” is from 9 to 45 and for people who survived is 0 to 58

- The survivors belong to Pclass 1,2 and 3 and people who did not survive belonged to Pclass 3

Answer 12)



Answer 13)

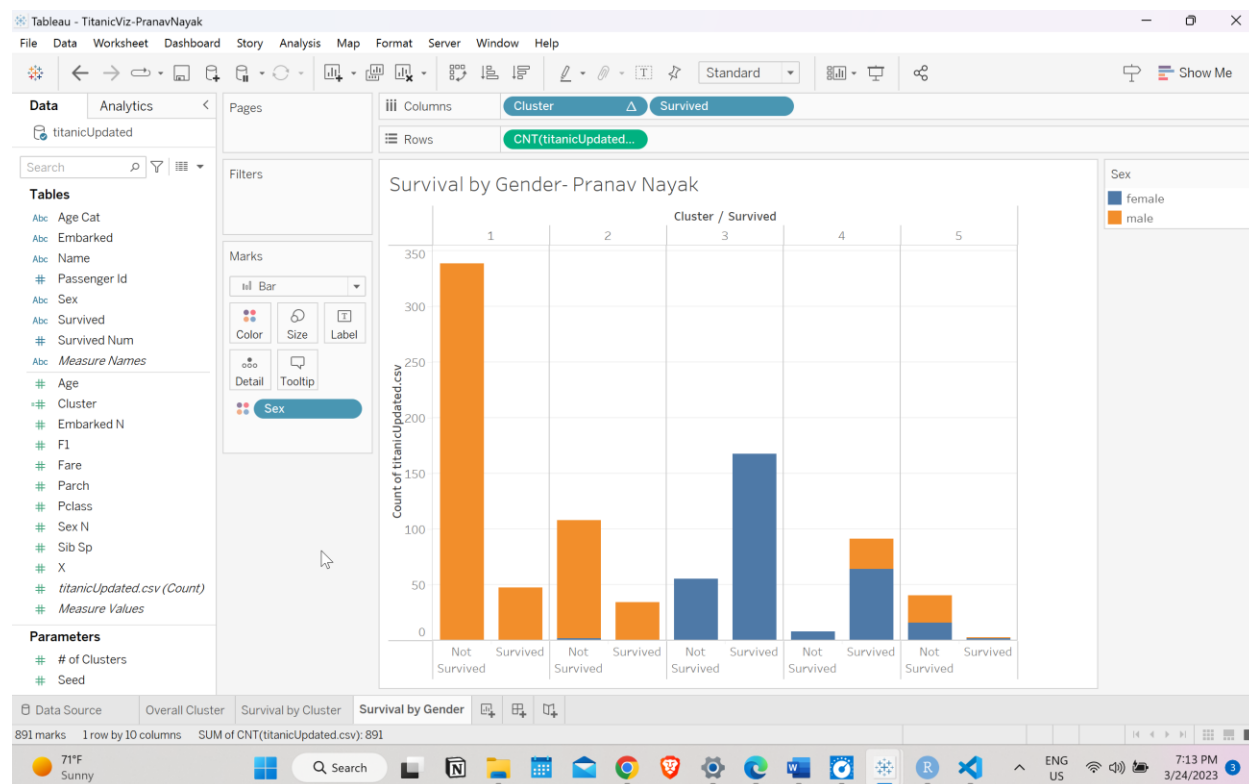
Survivability is highest in clusters 3 and 4.

Answer 14)

	Cluster 3	Cluster 4
Ideal Gender		
Ideal Passenger Class		
Ideal Age Category		
Ideal Embarked point		

Ideal number of siblings		
--------------------------	--	--

Answer 15)



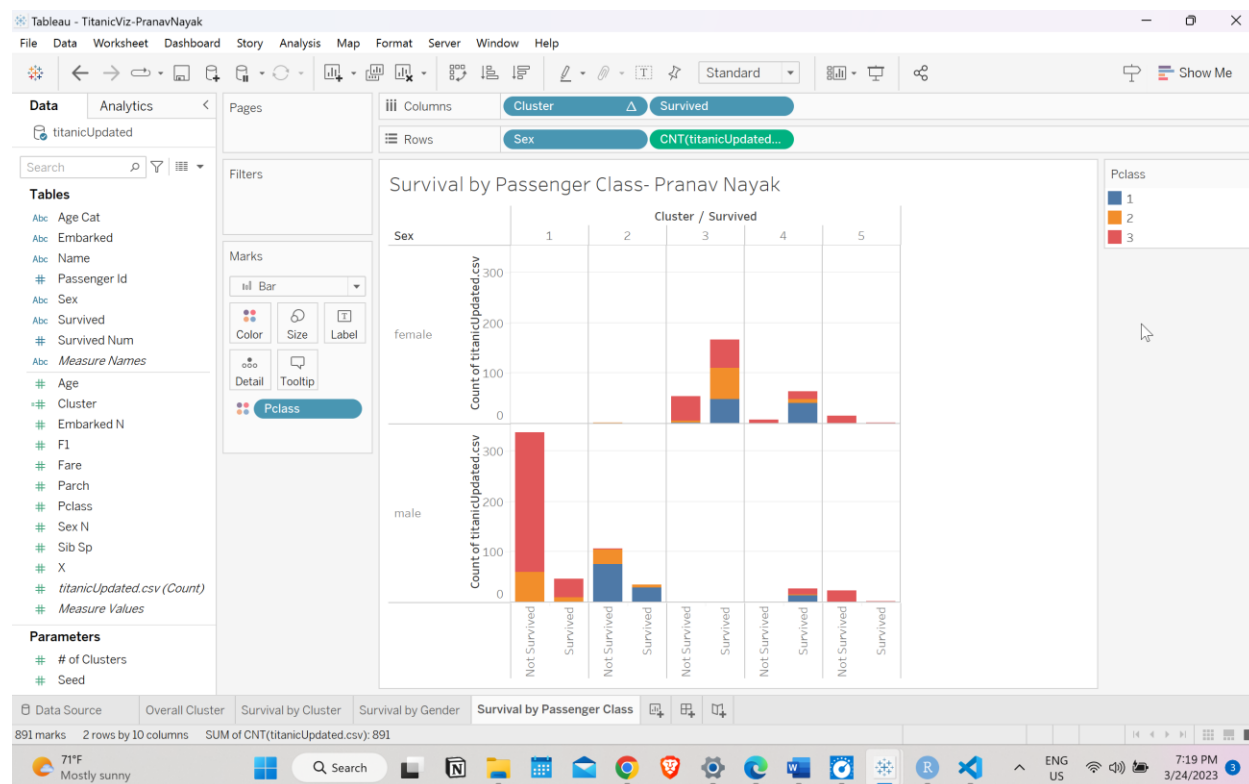
Answer 16)

For females, Cluster 3 has the highest survival probability.

Answer 17)

	Cluster 3	Cluster 4
Ideal Gender	Female	Female
Ideal Passenger Class		
Ideal Age Category		
Ideal Embarked point		
Ideal number of siblings		

## Answer 18)



## Answer 19)

	Cluster 3	Cluster 4
Ideal Gender	Female	Female
Ideal Passenger Class	Pclass 2	Pclass 1
Ideal Age Category		
Ideal Embarked point		
Ideal number of siblings		



Tableau - TitanicViz-PranavNayak

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Standard

Show Me

Data Analytics

Pages

Columns Cluster Survived

Rows Sex CNT(titanicUpdated.csv)

Search

Tables

- Age Cat
- Embarked
- Name
- Passenger Id
- Sex
- Survived
- Survived Num
- Measure Names
- Age
- Cluster
- Embarked N
- F1
- Fare
- Parch
- Pclass
- Sex N
- Sib Sp
- X
- titanicUpdated.csv (Count)
- Measure Values

Parameters

- # of Clusters
- Seed

Filters

Marks

Color Size Label

Detail Tooltip

Age Cat

Survival by Age Category- Pranav Nayak

Sex

Cluster / Survived

female

male

Count of titanicUpdated.csv

Age Cat

- 0-16
- 17-32
- 33-48
- 49-64
- 65 and above

Data Source Overall Cluster Survival by Cluster Survival by Gender Survival by Passenger Class Survival by Age Category

991 marks 2 rows by 10 columns SUM of CNT(titanicUpdated.csv): 891

71°F Mostly sunny

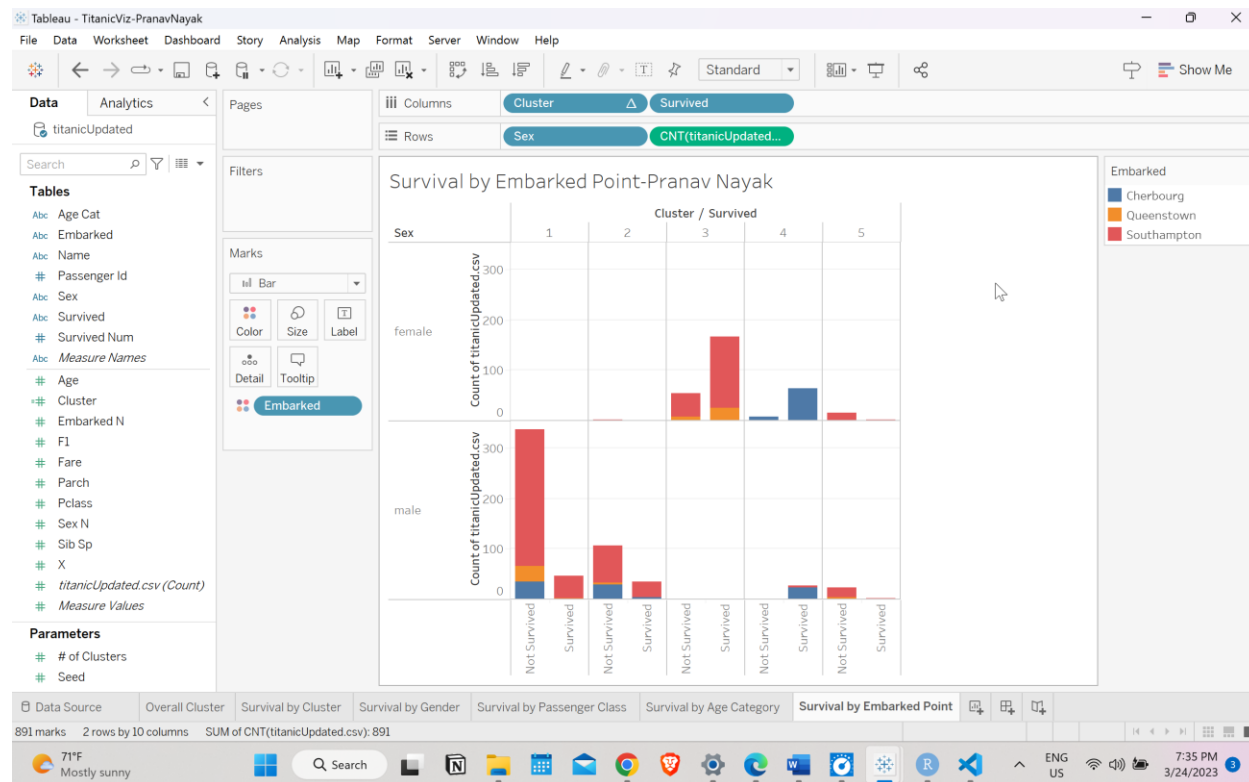
Search

ENG US

7:26 PM 3/24/2023

	Cluster 3	Cluster 4
Ideal Gender	Female	Female
Ideal Passenger Class	Pclass 2	Pclass 1
Ideal Age Category	17 -32 years	17 -32 years
Ideal Embarked point		
Ideal number of siblings		

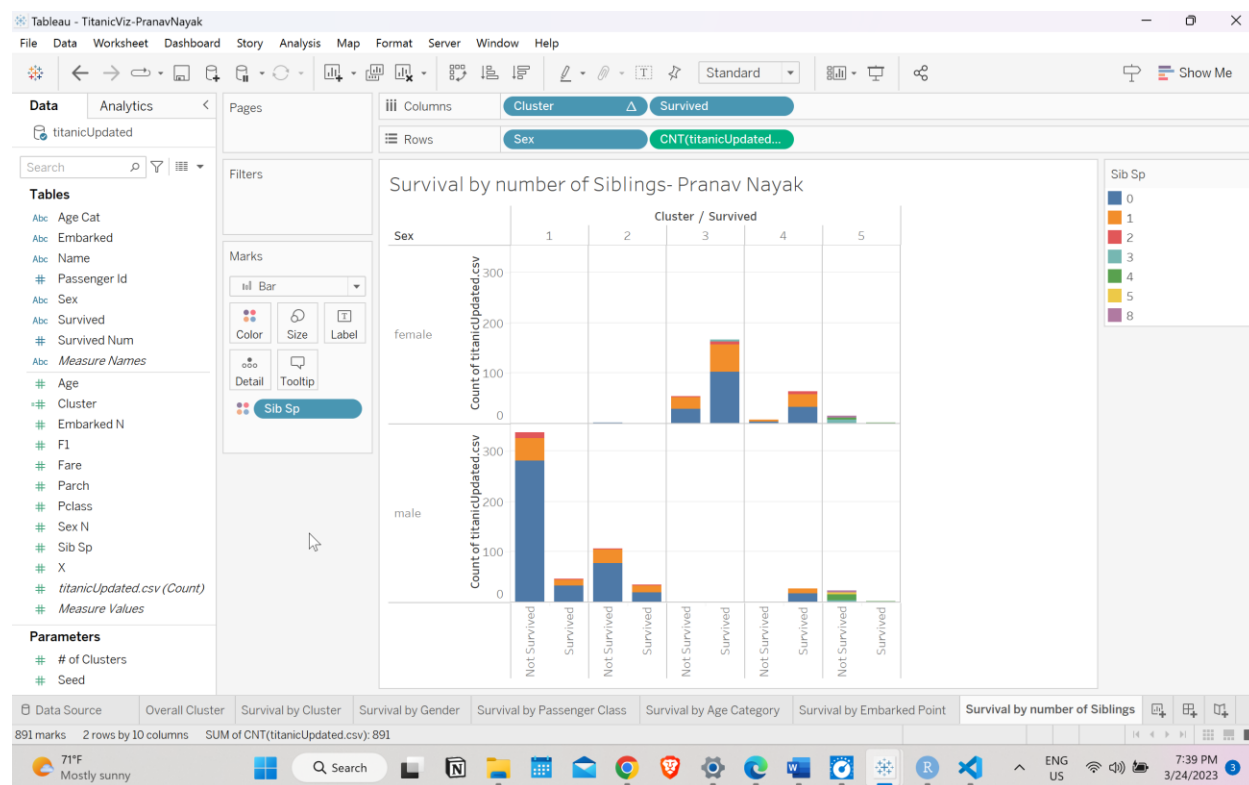
## Answer 22)



## Answer 23)

	Cluster 3	Cluster 4
Ideal Gender	Female	Female
Ideal Passenger Class	Pclass 2	Pclass 1
Ideal Age Category	17 -32 years	17 -32 years
Ideal Embarked point	Southampton	Cherbourg
Ideal number of siblings		

## Answer 24)



## Answer 25)

	Cluster 3	Cluster 4
Ideal Gender	Female	Female
Ideal Passenger Class	Pclass 2	Pclass 1
Ideal Age Category	17 -32 years	17 -32 years
Ideal Embarked point	Southampton	Cherbourg
Ideal number of siblings	Zero	Zero

Answer 26)

Cluster 3:

The data shows that passengers with the highest survival chance belong to cluster 3 that were majorly women in the 17-32 age group. They belonged to the second class and embarked from Southampton. They traveled alone without siblings.

Cluster 4:

Passengers with the highest survival probability and who belonged to cluster 4 were females between the age group of 17 to 32 years that were categorized in passenger class 1. The majority of them boarded from Cherbourg and didn't have any siblings