

Install Orange Data Mining. Include screenshots showing successful installation and the opening screen of the software. Explore Orange's interface. Take screenshots of the main workspace, widget panel, and canvas.

A white background with black border

Description automatically generated

Using the "Insurance.csv" or "Viral.csv" dataset, create a basic workflow in Orange. Include screenshots of your workflow, showing how data is loaded and visualized.

A screenshot of a computer

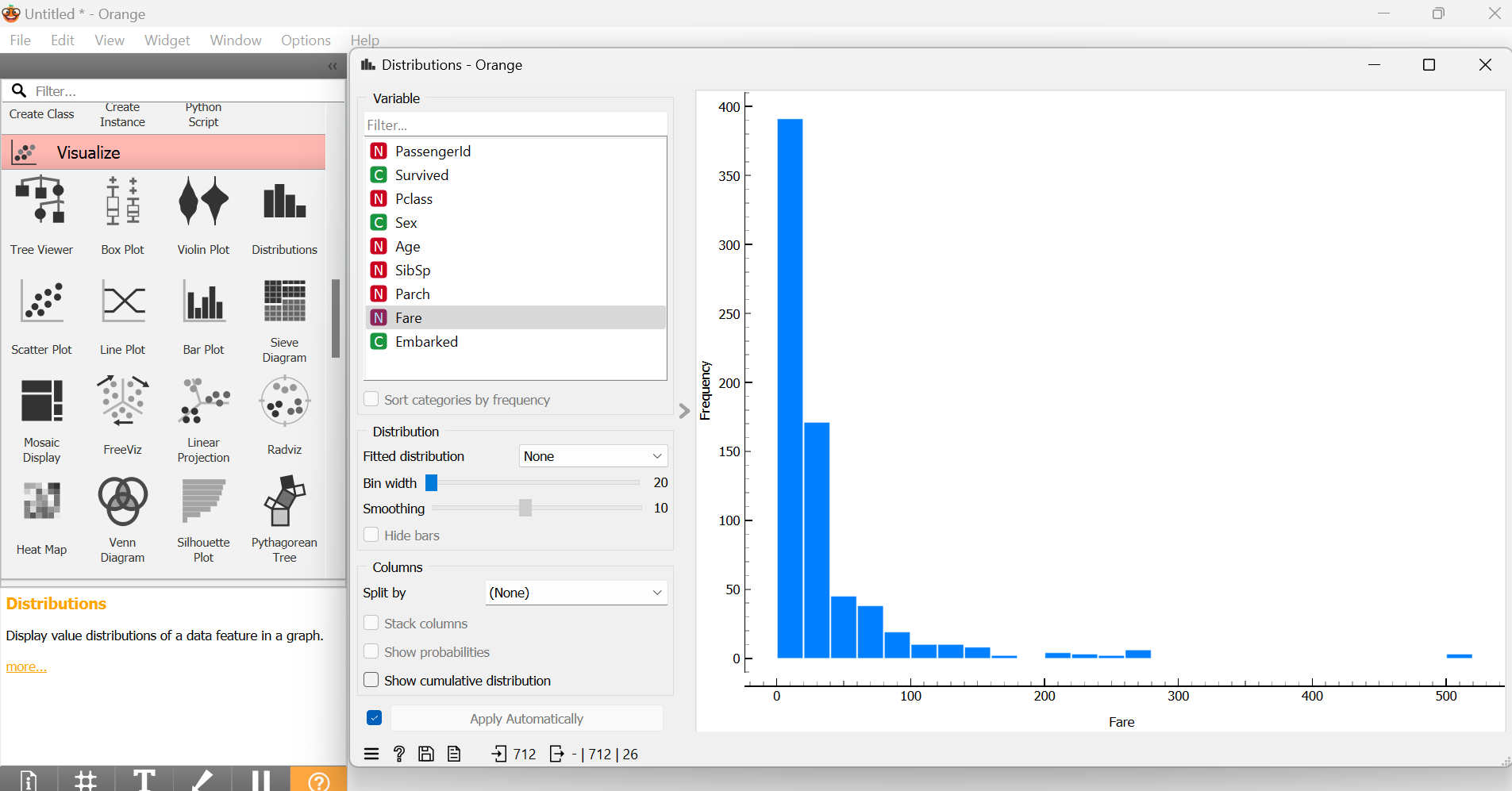
Description automatically generated

Analysis:

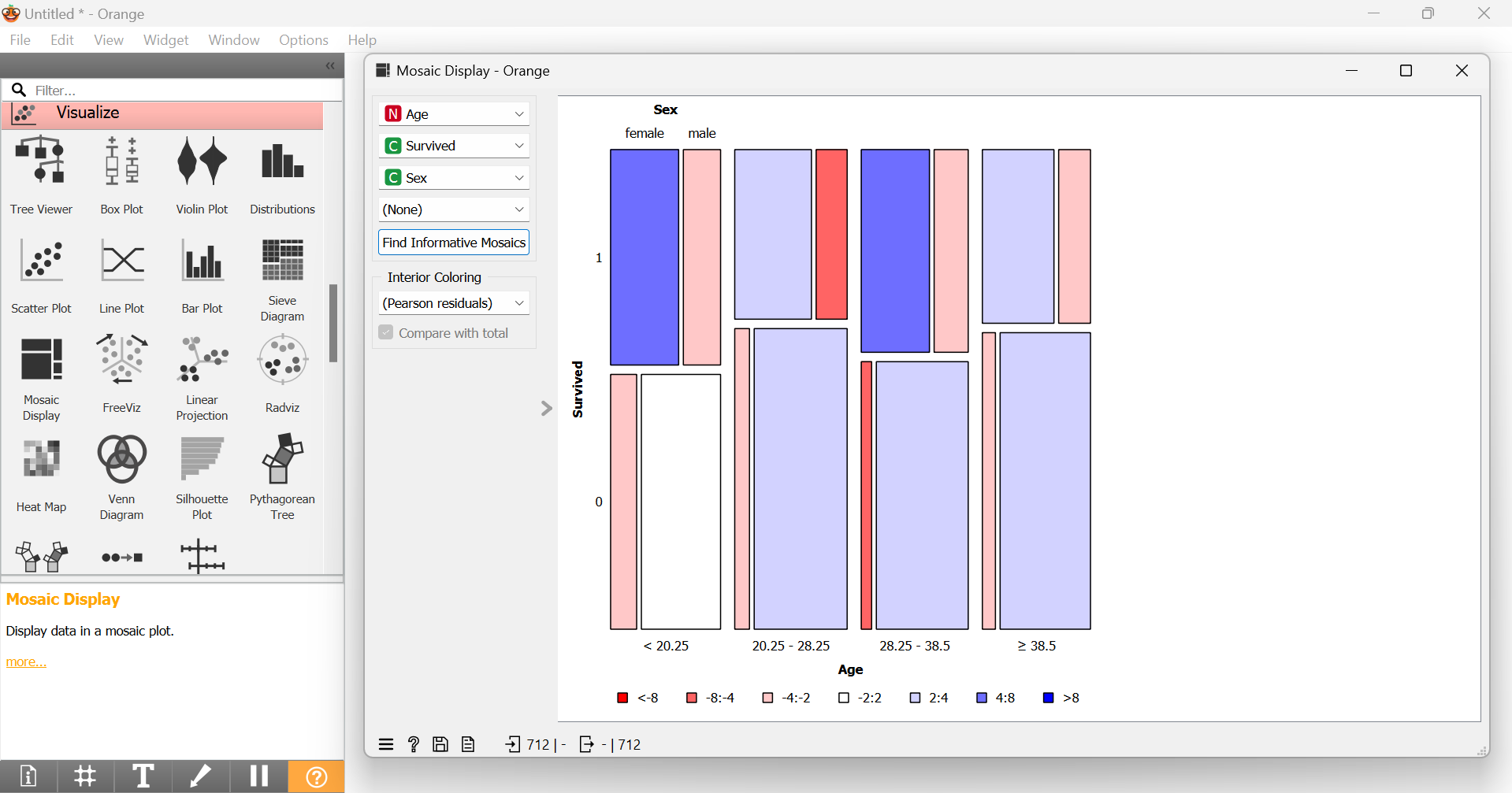
* Analyze the chosen dataset to identify patterns or insights. Document your process and findings with screenshots.



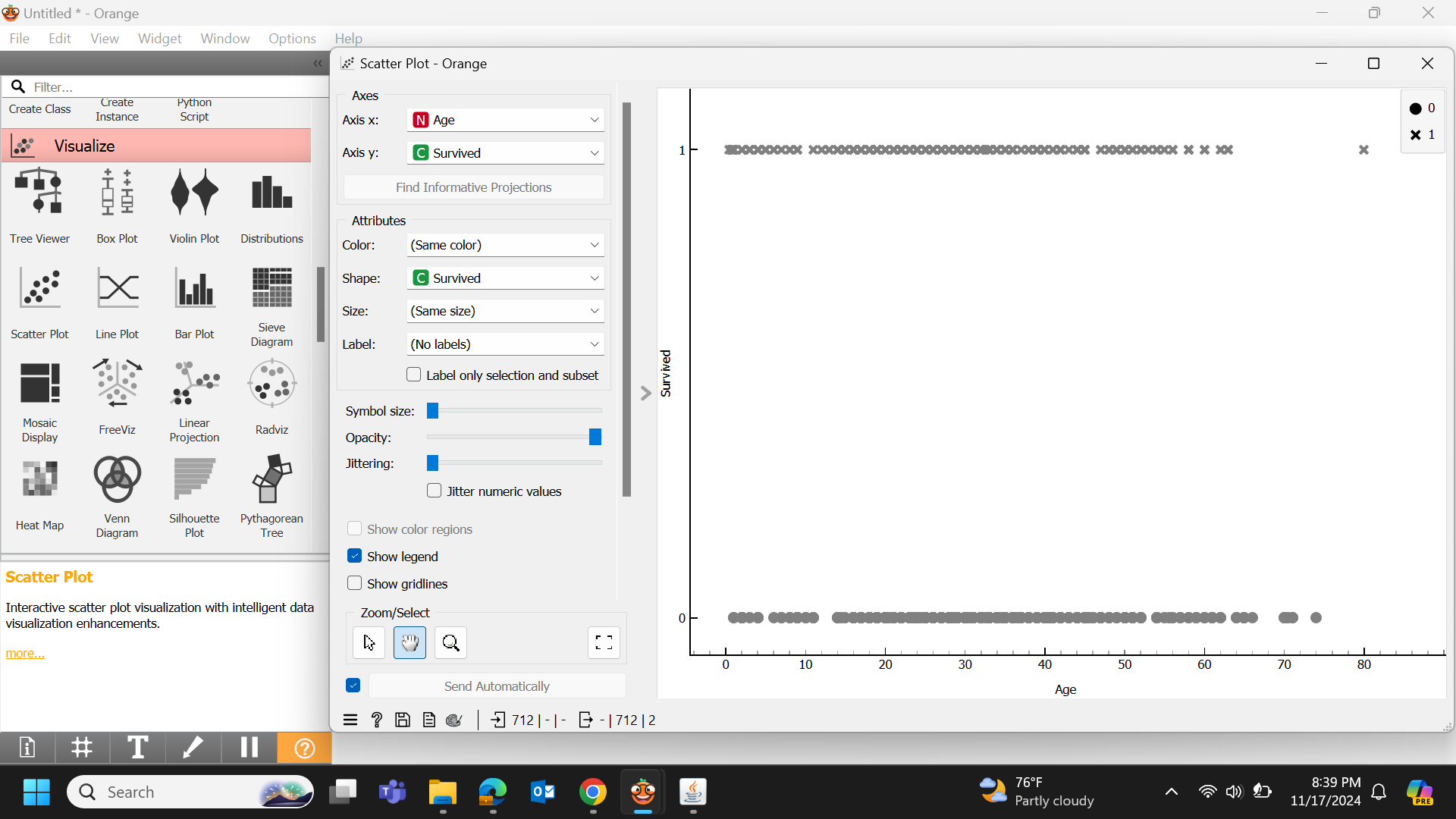
This flow shows the mean and median mode of the passengers on the Titanic. By adding the additional filter of the sex we can see the differences of the passengers' core demographic.



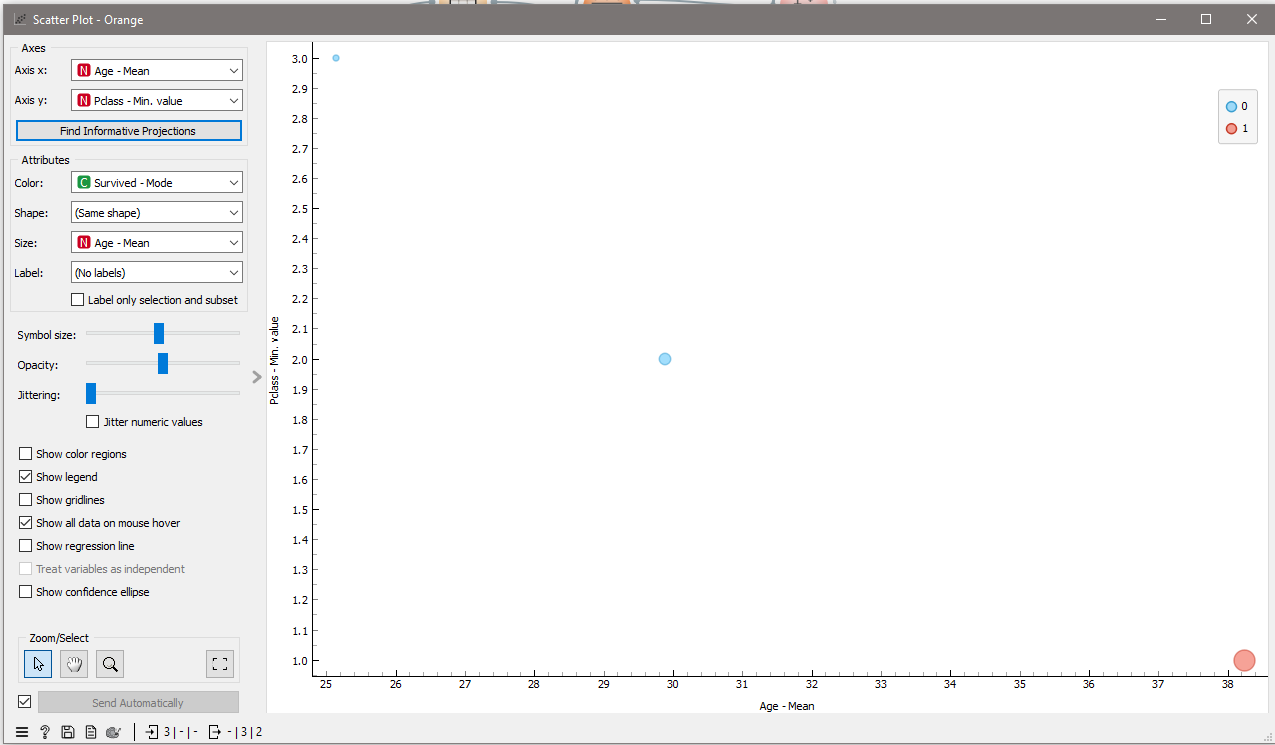
This flow shows the passengers on the ship and how much their fare was. The y-axis (frequency) is the number of passengers, and the x-axis (Fare) shows how much the passengers paid. This chart shows us the large differences in what the passengers paid.



This flow uses a mosaic display to see the different demographics of the passengers and who survived. From this visualization you can conclude that majority of those who survived the Titanic were females in their 20s-25s.



This flow uses a scatter plot, which is being used to compare the passengers who survived and their age. Those who survived have an X shape and those who passed away are represented as a circle.



A scatterplot shows the average age of passengers in the ship class. Further filtering shows that class 3 did not survive.

A screen shot of a computer

Description automatically generated

To further analyze the graph above, we then change the parameters to include fare and count all the ages rather than just the average. This gives us a different picture where some of the outliers include those who paid the highest did not survive and those who paid the lowest did survive.

A screenshot of a computer

Description automatically generated

Running the regression model, the lower R2 shows that the model is incomplete. The Roc analysis also failed.

A screenshot of a computer

Description automatically generated

Switching to a logical regression model, with the categorical target of “survived”. The target test scores are much higher, this model is more complete. The Roc analysis was completed.

A graph on a graph

Description automatically generated