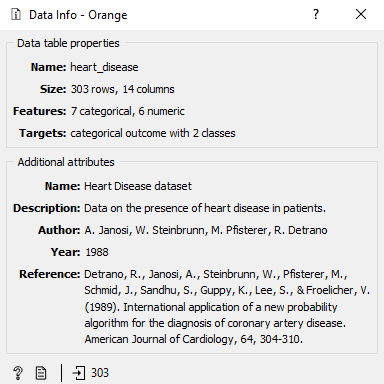
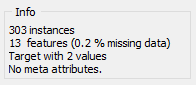
**Section 1**

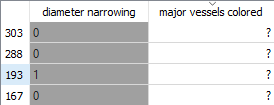
**a Preprocessing (Orange Tool) Class Works**

1) Perform imputation on Heart Disease dataset.

**Input**



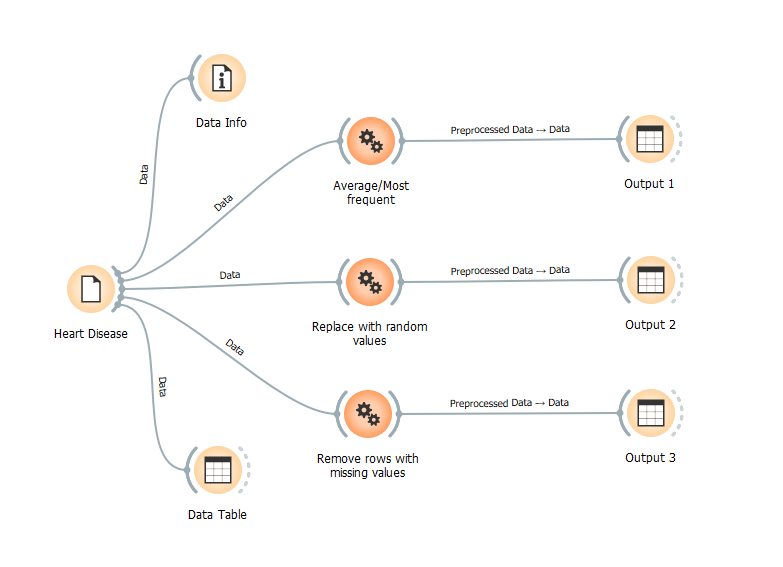




**Process**

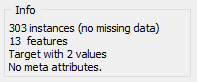
Imputation

* Replace Missing values with average or most frequent value.
* Replace with a random value.
* Remove Rows with missing values.

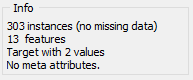


**Output**

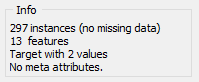
Output 1 (Average/Most frequent value)



Output 2 (Replace with random value)

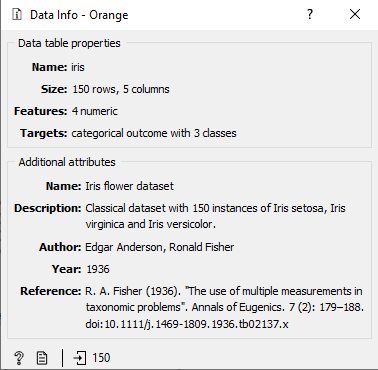


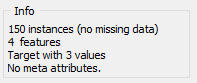
Output 3- Remove rows with missing values.

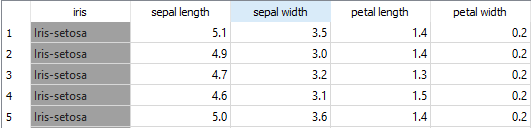


2) Perform Discretization on Iris dataset

**Input**



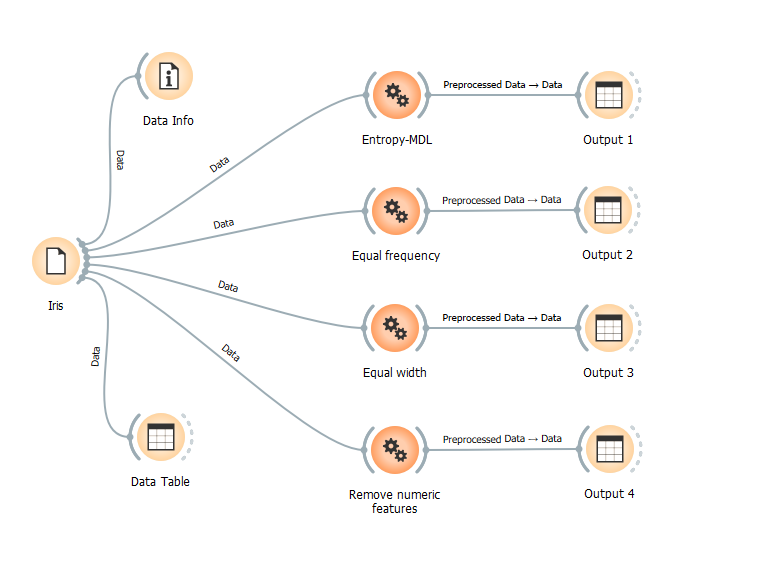




**Process**

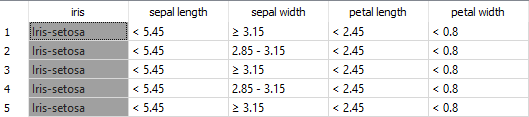
Discretization

* Equal frequency discretization
* Equal width discretization
* Remove numeric values

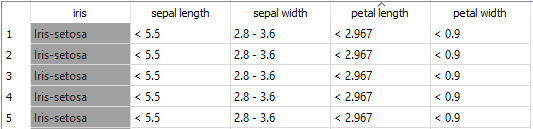


**Output**

Output 1 (Equal frequency discretization)



Output 2 (Equal width discretization)

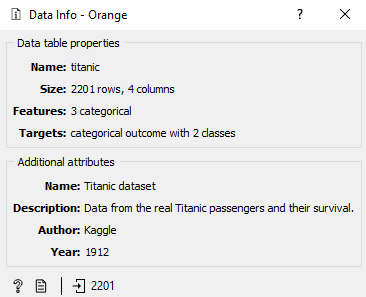


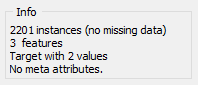
Output 3 (Remove numeric values)

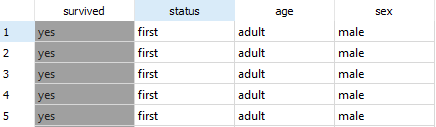


3. Perform continuization on Titanic dataset.

**Input**



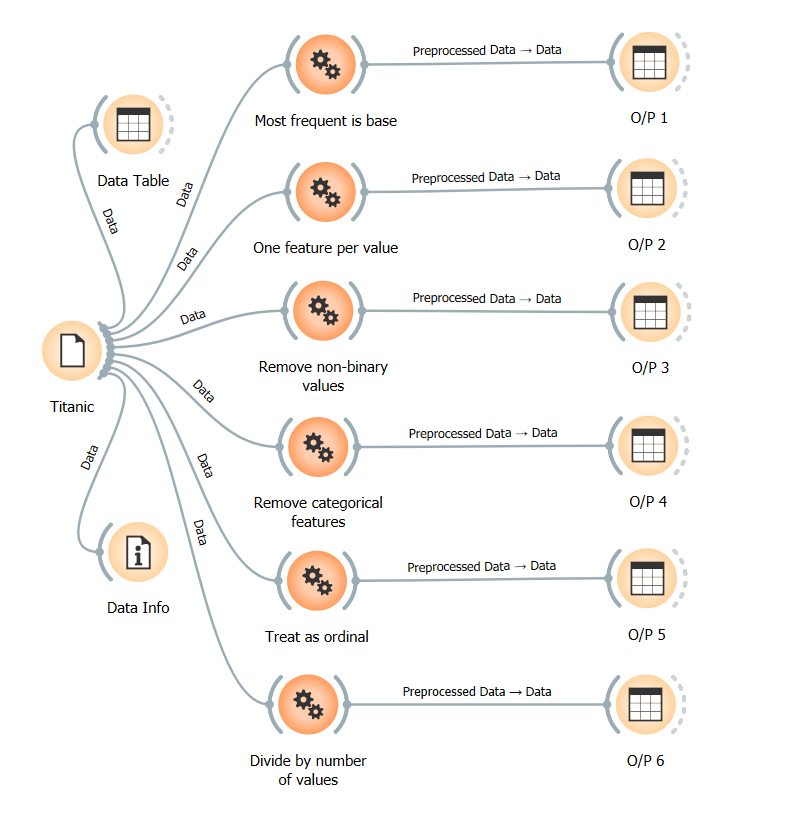




**Process**

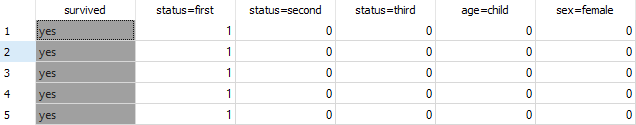
Continuization

* Most frequent is base.
* One feature per value.
* Remove non-binary values.
* Remove categorical features.
* Treat as ordinal.
* Divide by number of values.

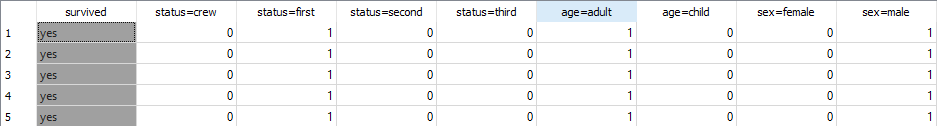


**Output**

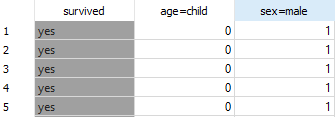
Output 1 (Most frequent is base)



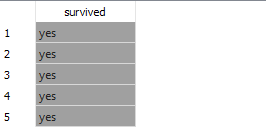
Output 2 (One feature per value)



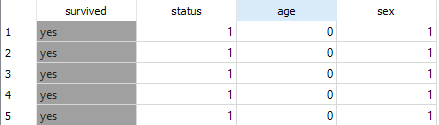
Output 3 (Remove non-binary values)



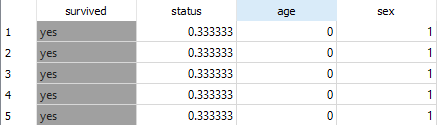
Output 4 (Remove categorical features)



Output 5 (Treat as ordinal)

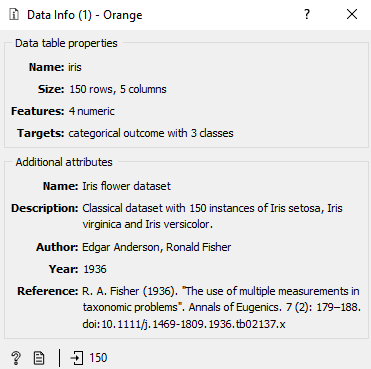


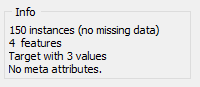
Output 6 (Divide by number of values)

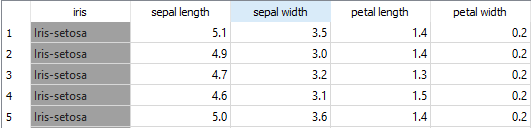


4. Perform normalization on Iris dataset

**Input**



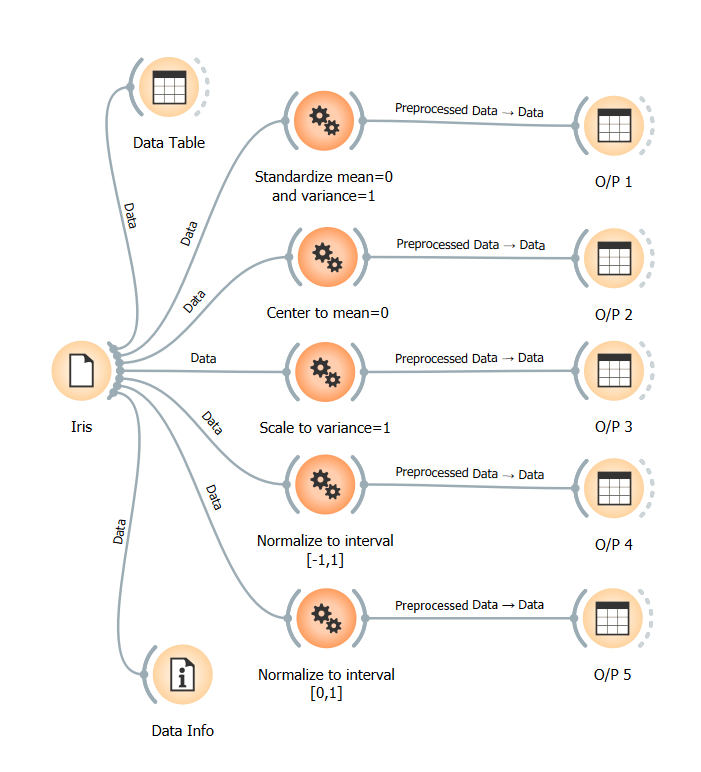




**Process**

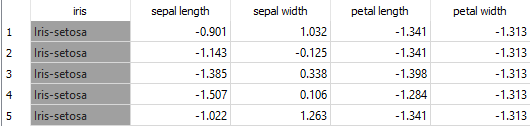
Normalization

* Standardize µ=0 and variance = 1.
* Center to
* Scale to variance = 1.
* Normalize to interval [-1,1].
* Normalize to interval [0,1].

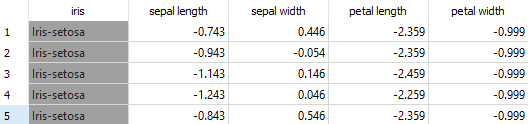


**Output**

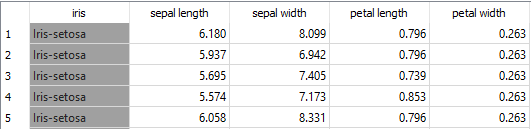
Output 1 (Standardize µ=0 and variance = 1)



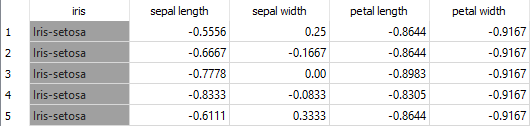
Output 2 (Center to µ=0.)



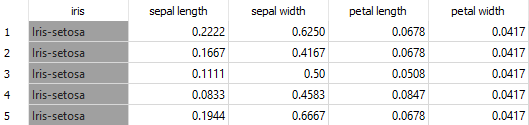
Output 3 (Scale to variance = 1)



Output 4 (Normalize to interval [-1,1])

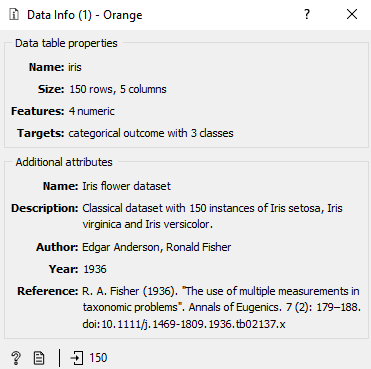


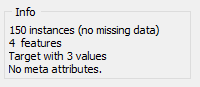
Output 5 (Normalize to interval [0,1])

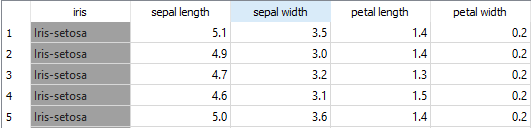


5) Perform Randomization on Iris dataset.

**Input**

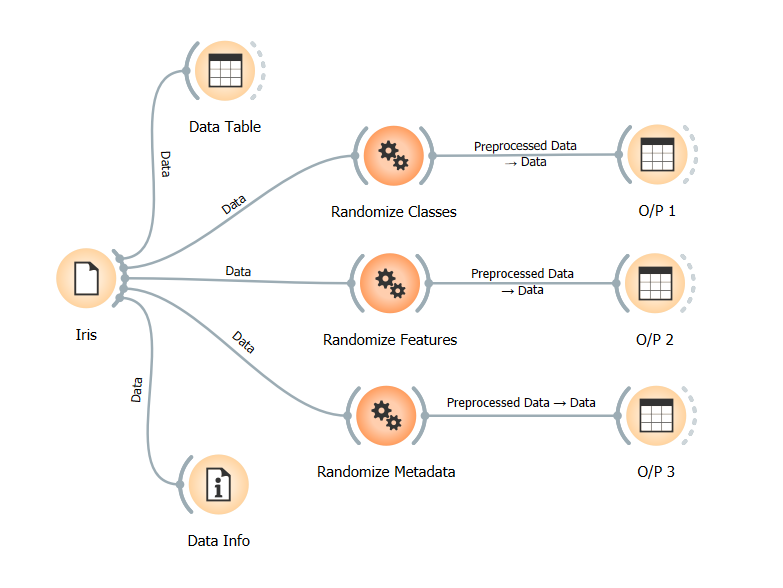






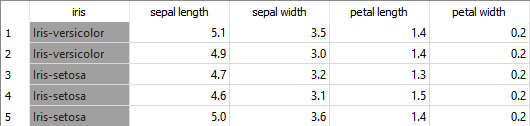
**Process**

Randomization

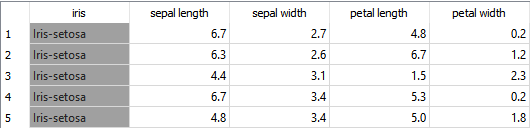


**Output**

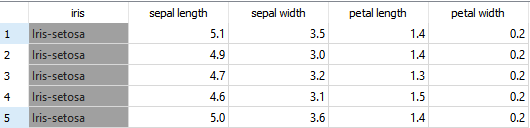
Output 1 (Randomize Classes)



Output 2 (Randomize Features)

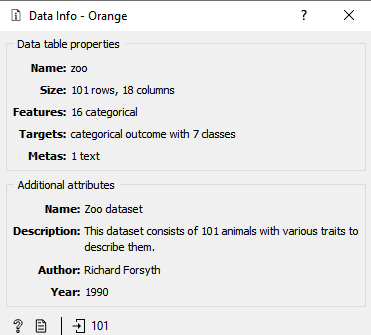


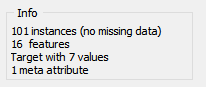
Output 3 (Randomize Metadata)

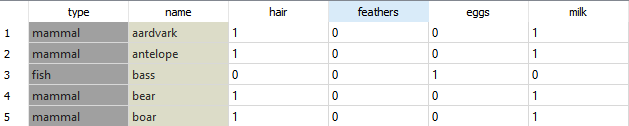


6) Perform Remove Sparse on zoo data set

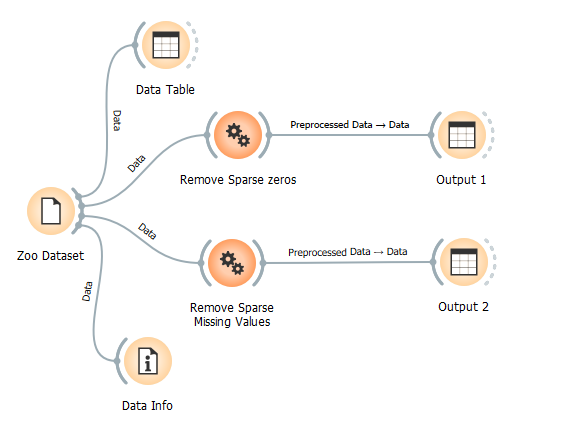
**Input**





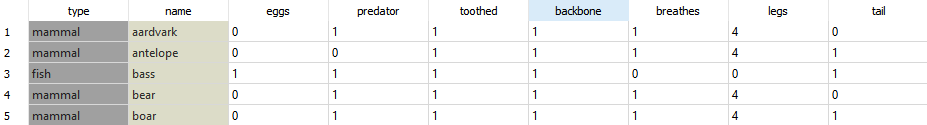


**Process**

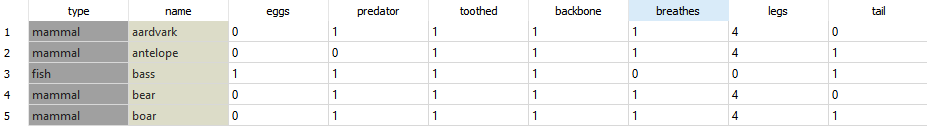


**Output**

Output 1 (Remove sparse zeros)

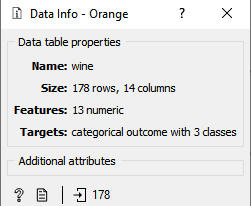


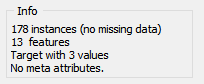
Output 2 (Remove sparse missing values)

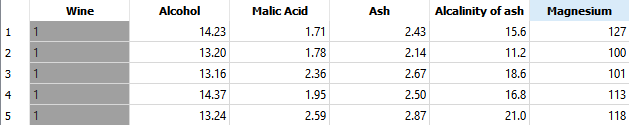


7) Perform Feature Selection on Wine dataset.

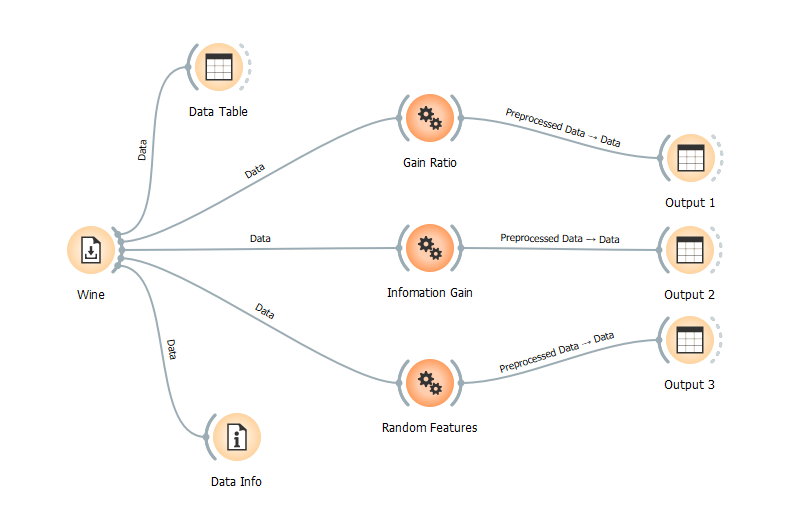
**Input**





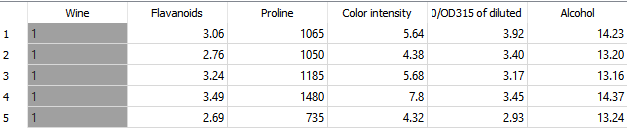


**Process**

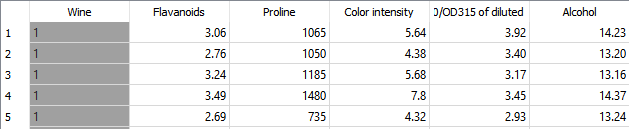


**Output**

Output 1 (Gain Ratio)



Output 2 (Information Gain)

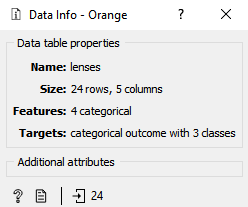


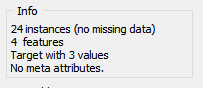
Output 3 (Random Features)

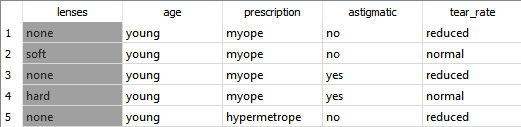


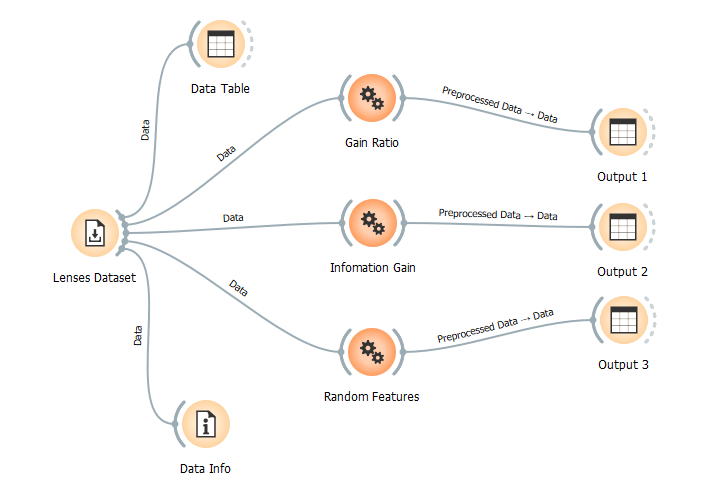
8) Perform Feature Selection on Lenses dataset

**Input**



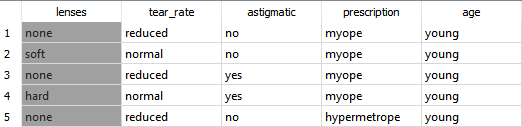




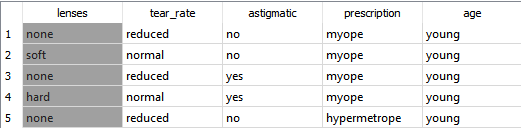


**Output**

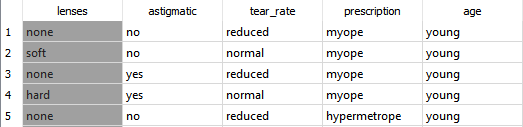
Output 1 (Gain Ratio)



Output 2 (Information Gain)



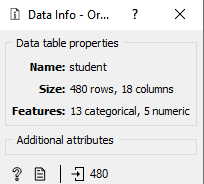
Output 3 (Random Features)

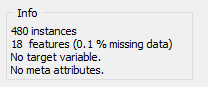


**b. Preprocessing (Orange Tool) Exercises**

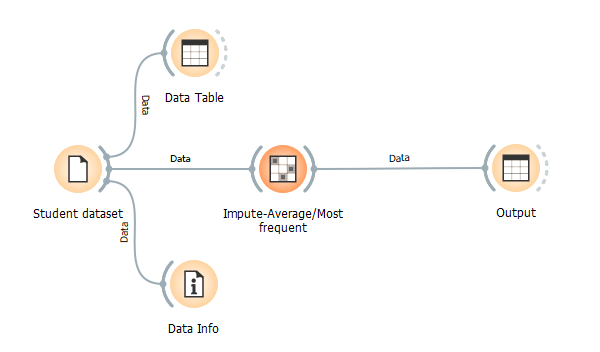
1) Replace missing values by the mean of the values of records having same class value. Display the entire data after replacement.

**Input**

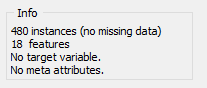




**Process**

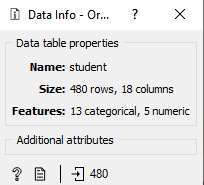


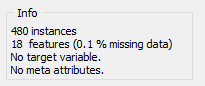
**Output**

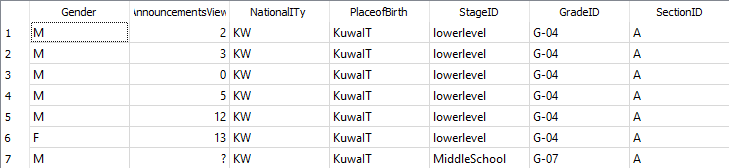


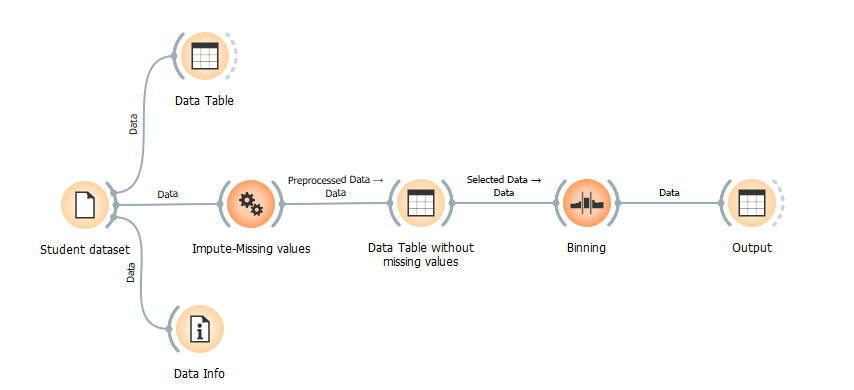
2) Perform binning(3 bins) for the attribute AnnouncementsView.

**Input**

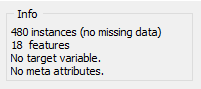


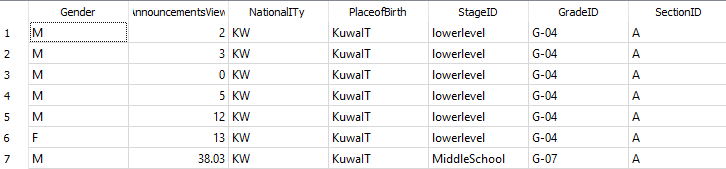


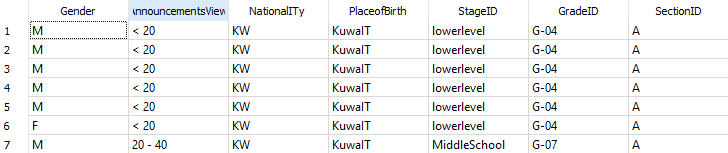


**Process**

**Output**

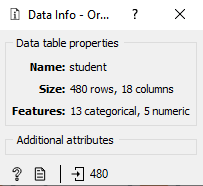


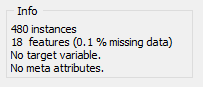


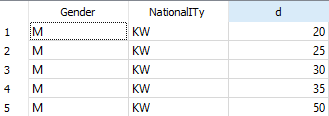


3) Remove redundant variables/features having high correlation.

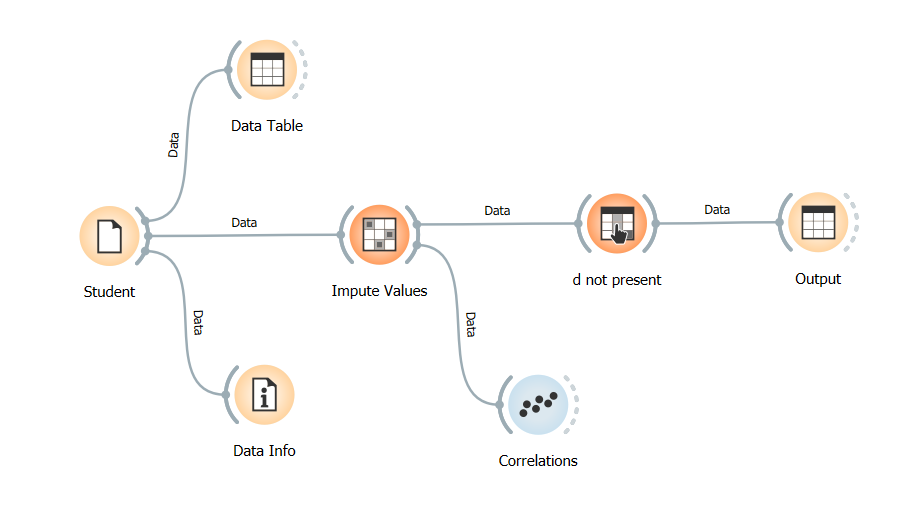
**Input**



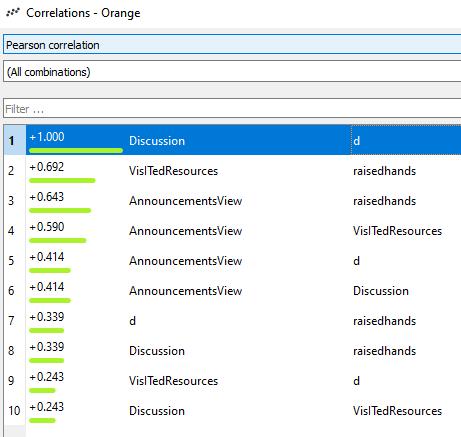


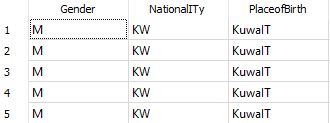


**Process**

****

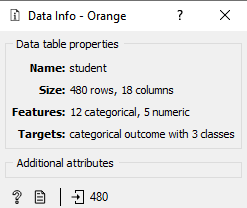
**Output**



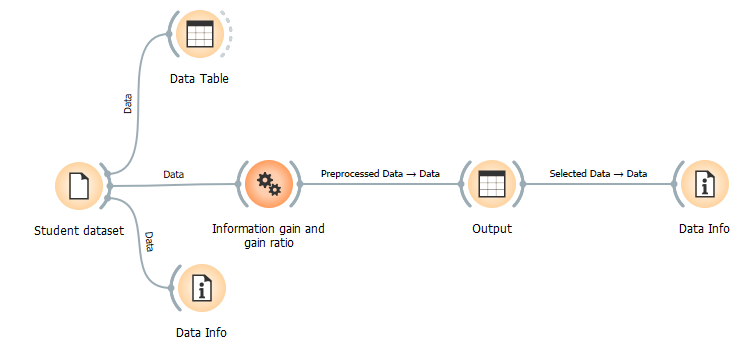


4 ) Select important variables/features using Information gain and gain ratio.

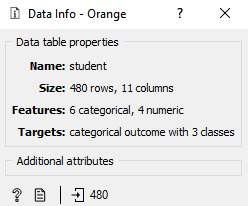
**Input**



**Process**



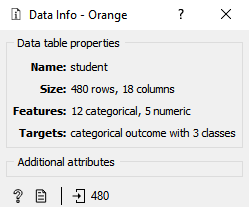
**Output**

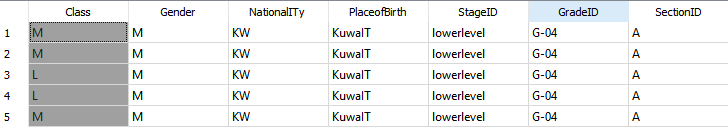


5) Perform normalization [-1,1] on the attribute raisedhands.

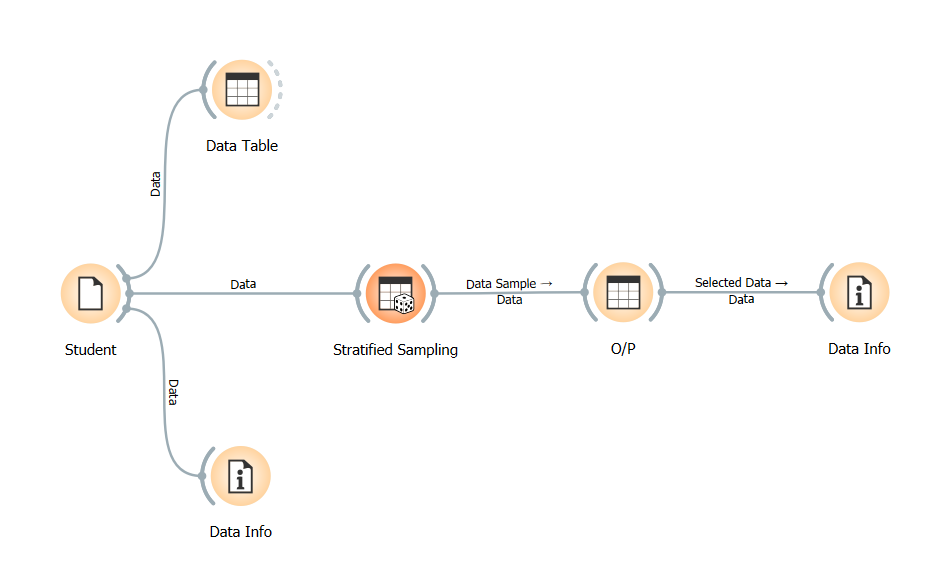
6) Do a stratified random sampling to draw a sample size of approximately 100 out of the total records.

**Input**

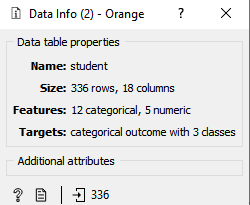




**Process**



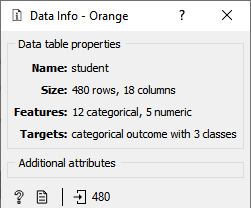
**Output**



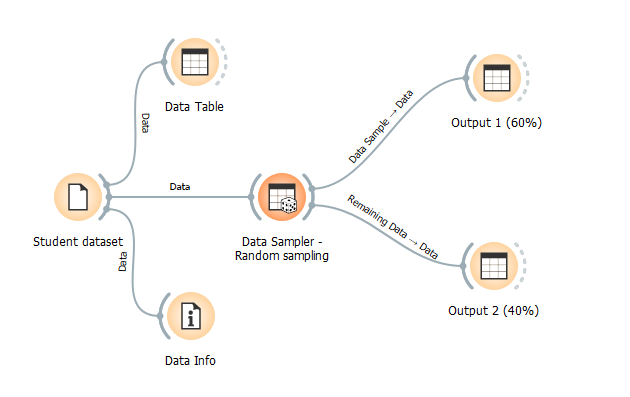


7) Partition the data into 2 data sets (60:40) using random partitioning.

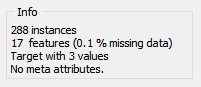
**Input**

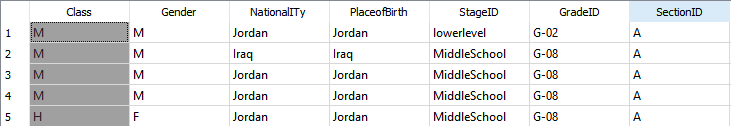


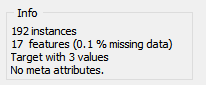
**Process**



**Output**





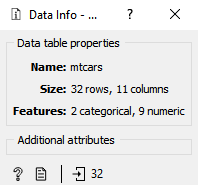


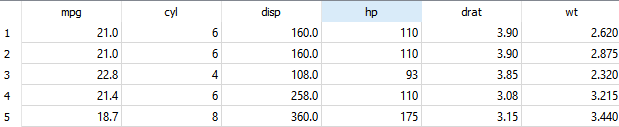


**Use mtcars data set to**

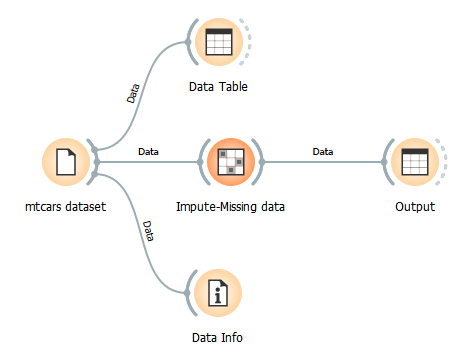
1) Replace the missing data with the average/median of the feature wt.

**Input**

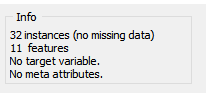


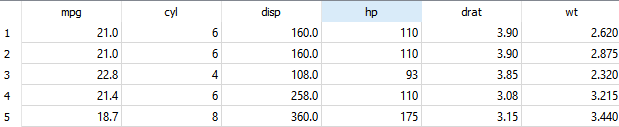


**Process**



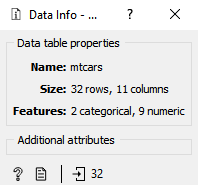
**Output**

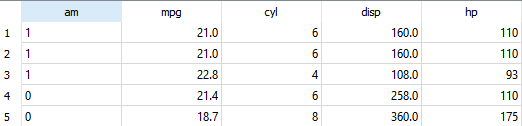




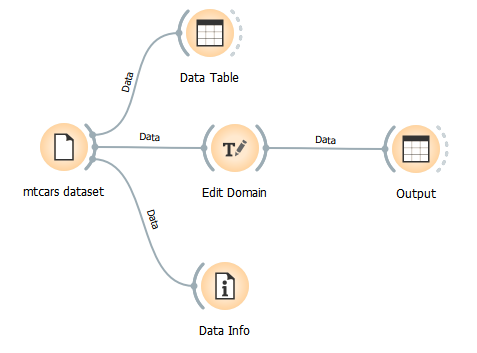
2) Transform the numerical variable am to manual-0 and automatic-1.

**Input**

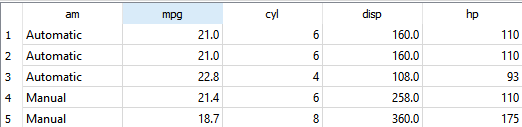




**Process**

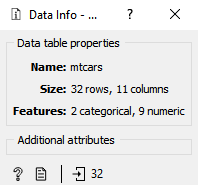


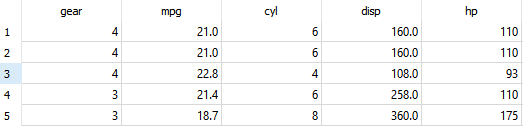
**Output**



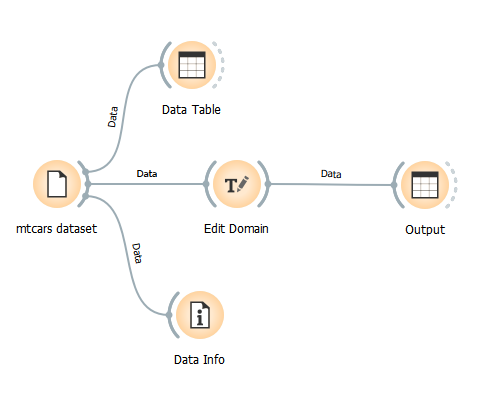
3) Transform the numerical variable gear by appending “gear” to the number of gears given in the feature.

**Input**

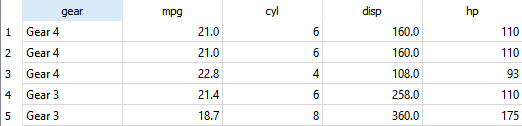




**Process**

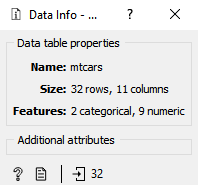


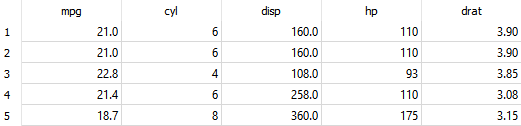
**Output**



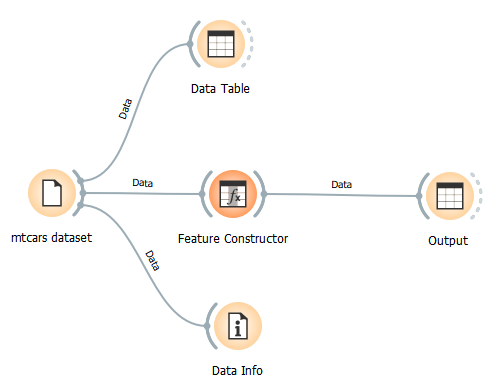
4) Add a new attribute Engine type based on the condition for the attribute vs (0 = V-shaped, 1 = straight).

**Input**

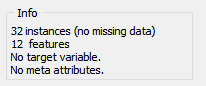


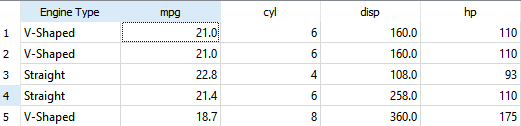


**Process**



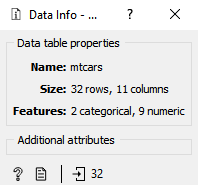
**Output**

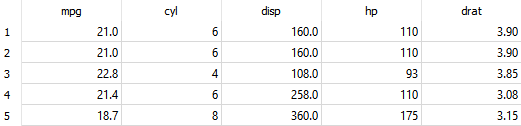




5) Scale the feature disp.

**Input**





Process

6) Split the dataset into 70%training data set and 30% test dataset

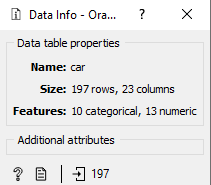
**Section II**

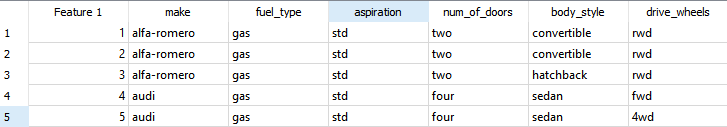
**a) Data Visualization (Orange Tool) Class Work**

***Use car.csv data set to***

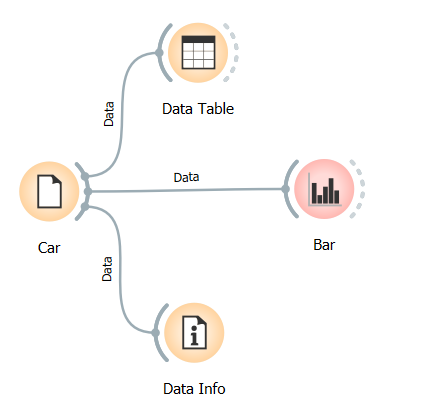
1) Plot a bar chart to compare the price of different makes of car.

**Input**

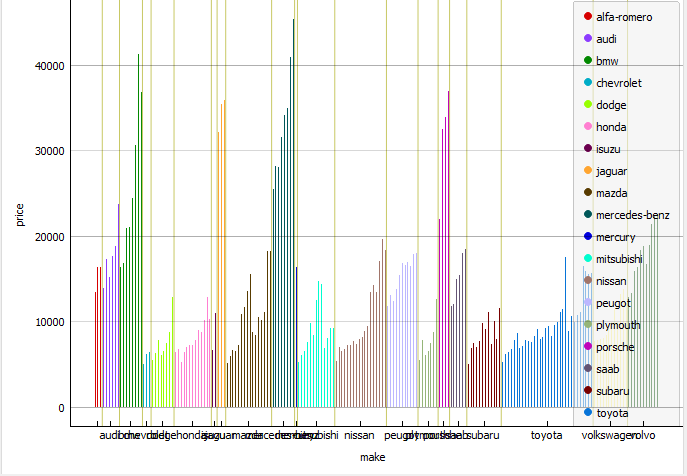




**Process**

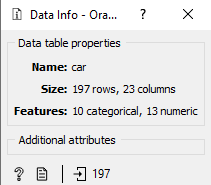


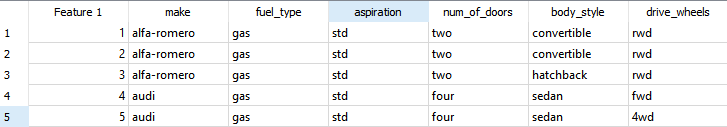
**Output**



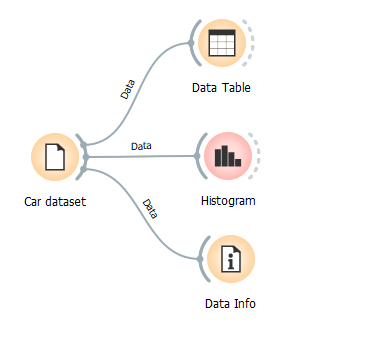
2) Create a histogram for analyzing city mileage.

**Input**

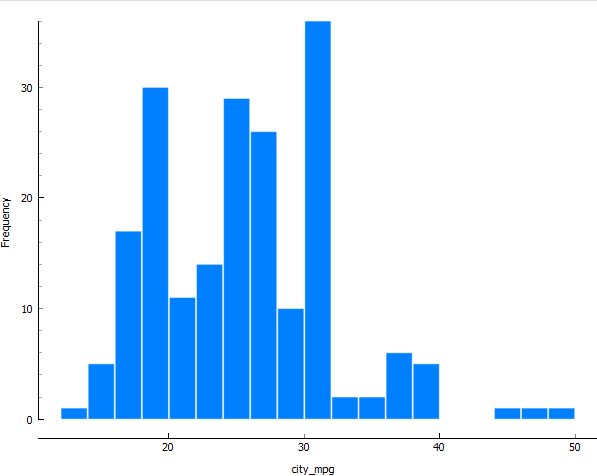




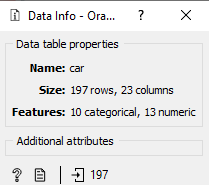
**Process**

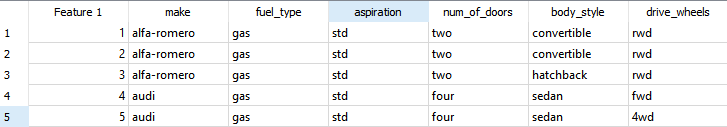


**Output**

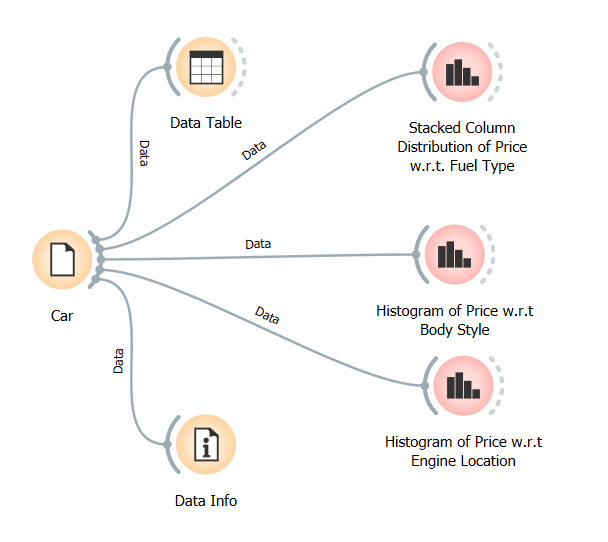
3) Create a histogram for analyzing price. Show a stacked column distribution with respect to fuel\_type. Similarly create a histogram for price w.r.t body\_style and price w.r.t engine\_location. Write your inferences for price of cars w.r.t the above variables.

**Input**



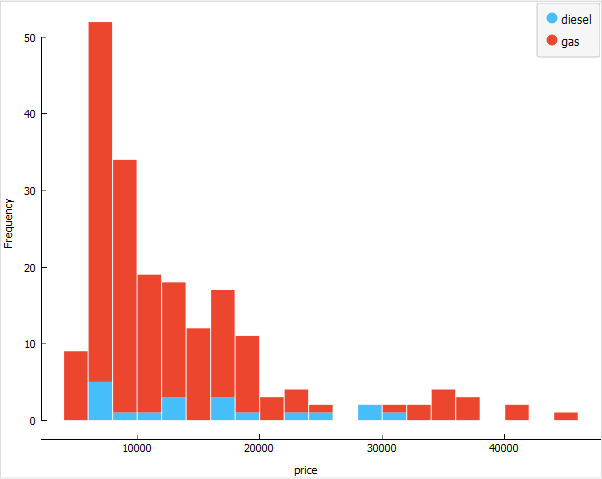


**Process**

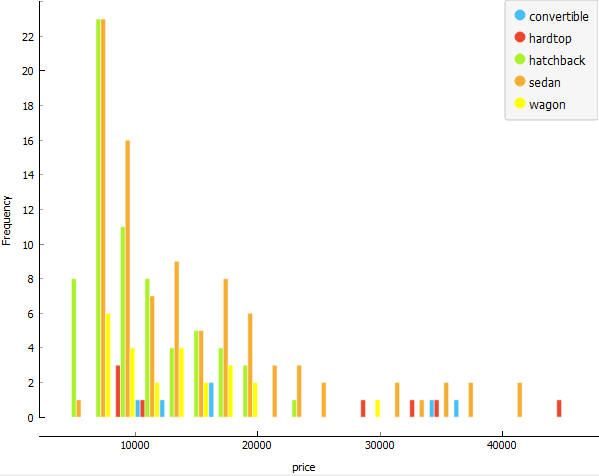


**Output**

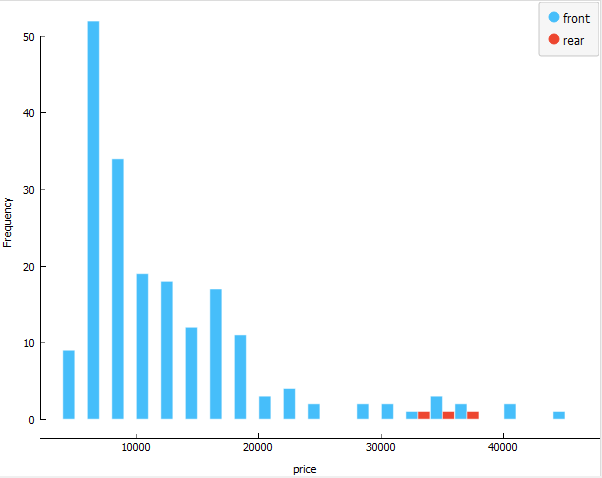
a) Stacked column distribution of price w.r.t fuel type.



Histogram for price w.r.t body style.

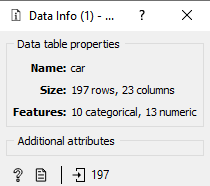


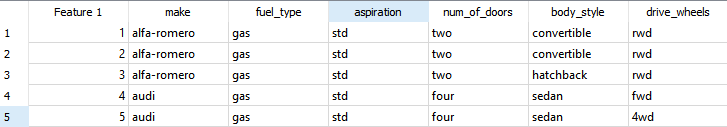
Histogram of price with respect to engine location.



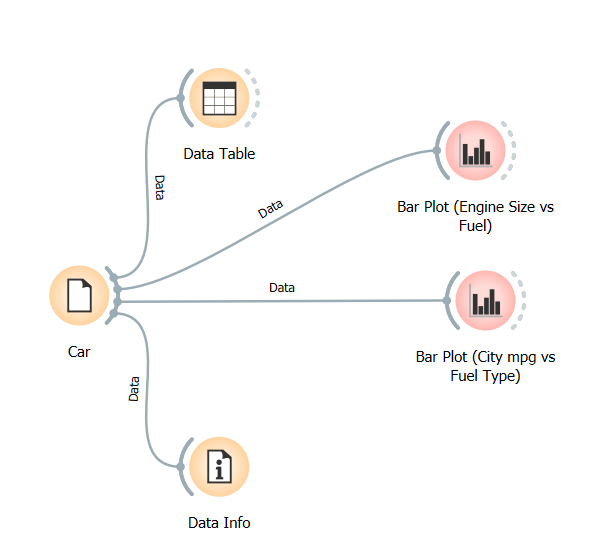
4) Visualize a bar plot for engine\_size Vs make. Similarly visualize a bar plot for city\_mpg vs fuel\_type and write your inferences.

**Input**



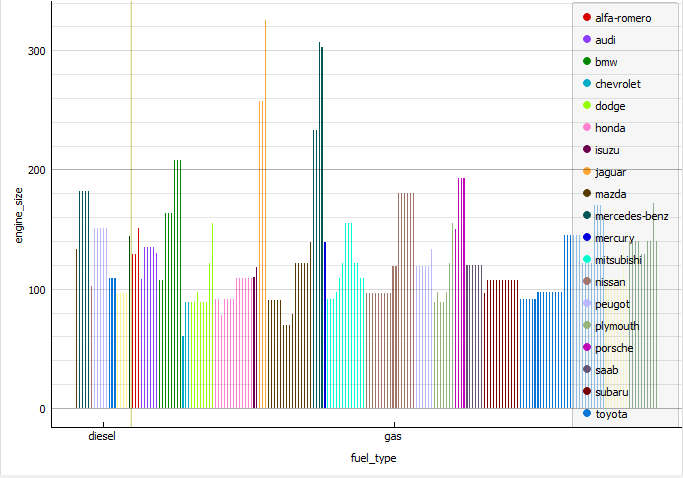


**Process**

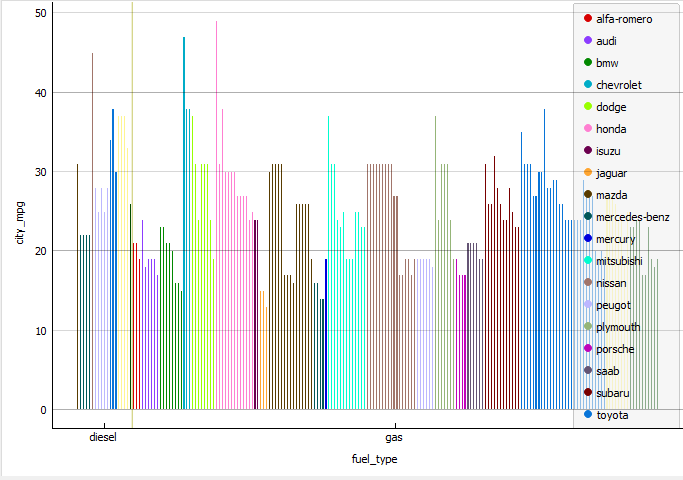
****

**Output**

Output 1 (Bar plot for engine\_size vs fuel\_type.)

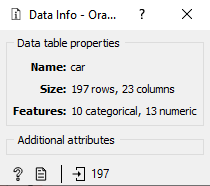


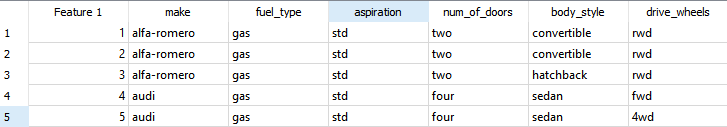
Output 2 (Bar plot for city\_mpg vs fuel\_type.)



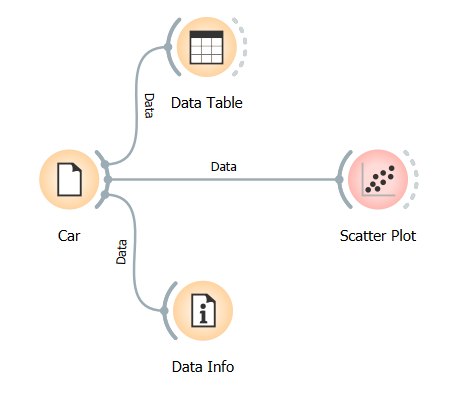
5) Create a scatter plot for price, vs engine\_size, w.r.t num\_of\_cylinders(color), aspiration(shape), wheel\_base(size).

**Input**





**Process**

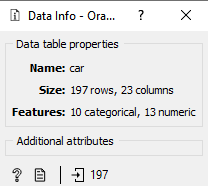


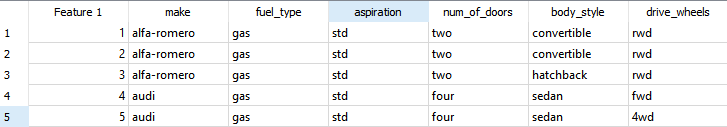
**Output**



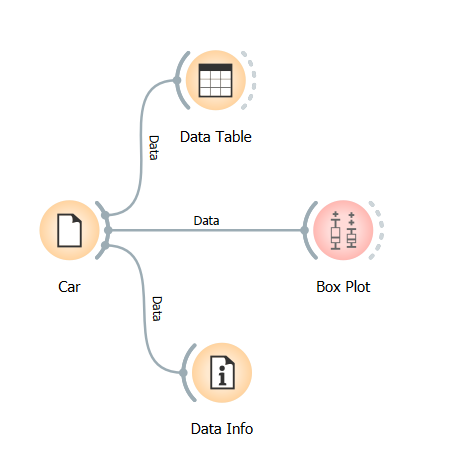
6) Create a boxplot for price w.r.t body\_styles.

**Input**

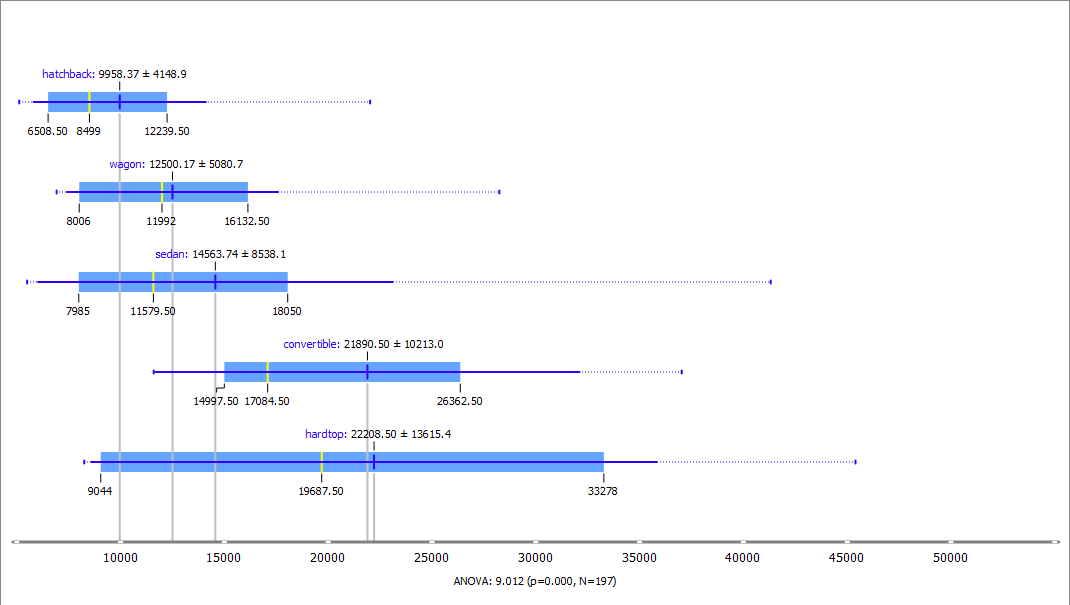




**Process**

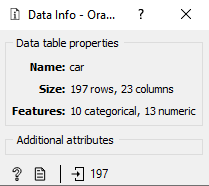


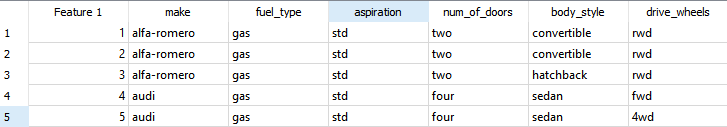
**Output**



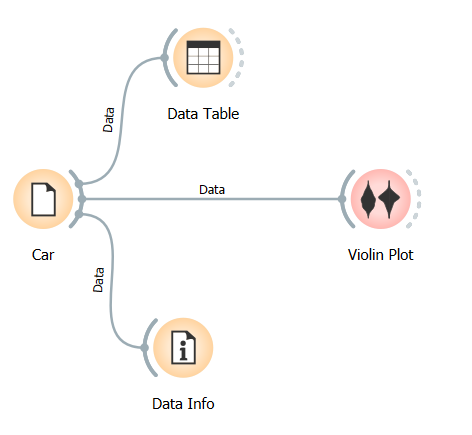
7) Create a violin plot for price w.r.t aspiration.

**Input**

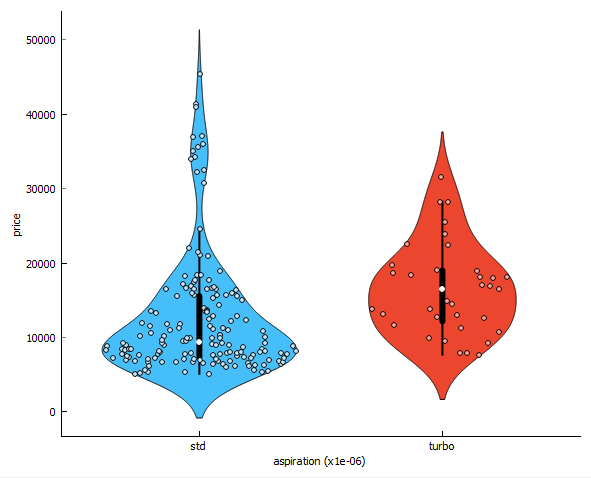




**Process**

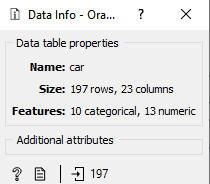


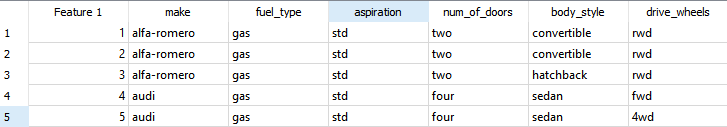
**Output**



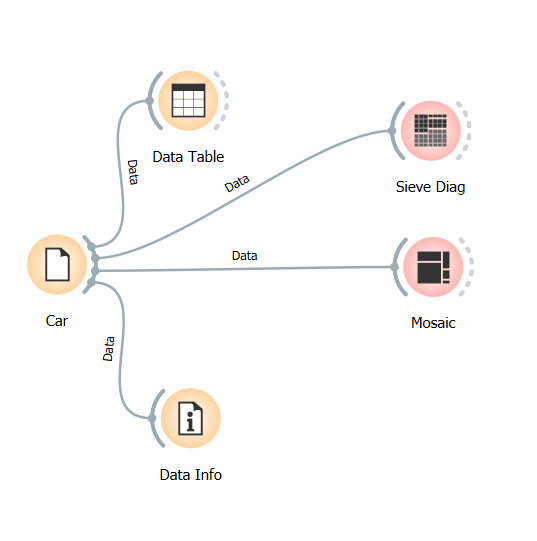
8) Illustrate sieve diagram and mosaic display for city\_mpg vs highway\_mpg.

**Input**



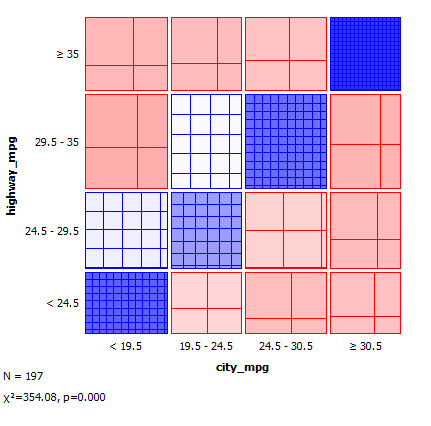


**Process**

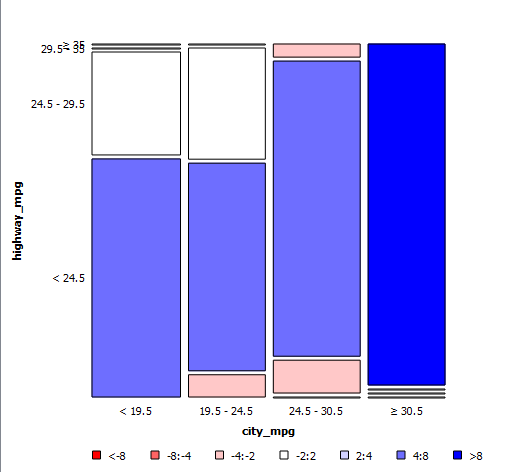


**Output**

Output 1 (Sieve Diagram for city\_mpg vs highway\_mpg)



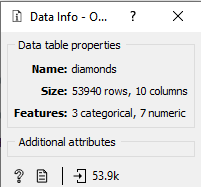
Output 2 (Mosaic Display for city\_mpg vs highway\_mpg)

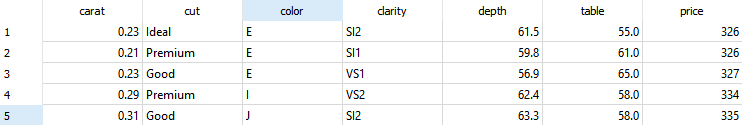


**Illustrate the following using *diamonds* data set**

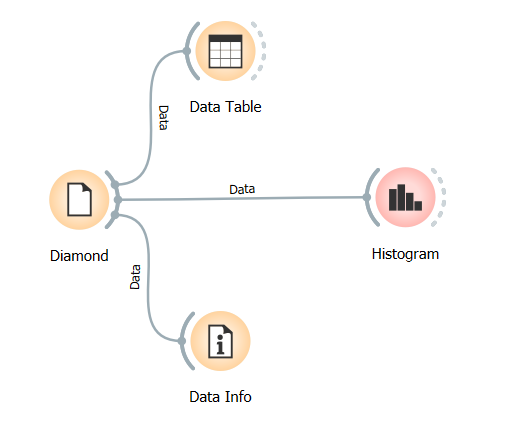
1) Create a histogram of & “carat” w.r.t cut.

**Input**

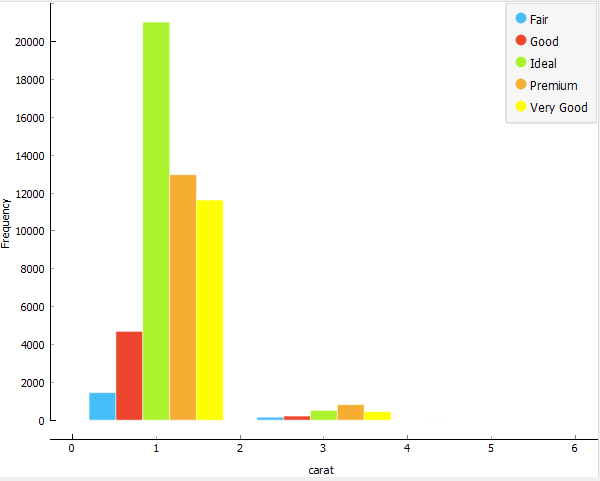




**Process**

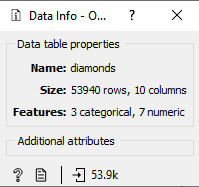


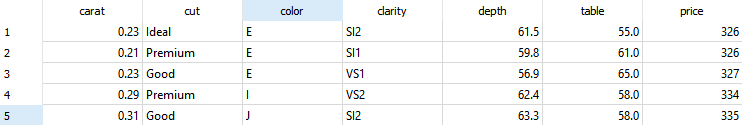
**Output**



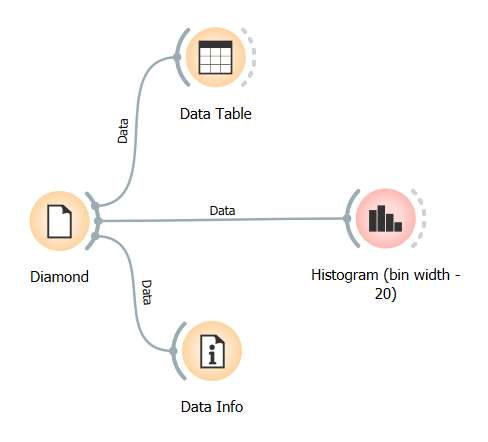
2) Set the bin width of the histogram to 20.

**Input**

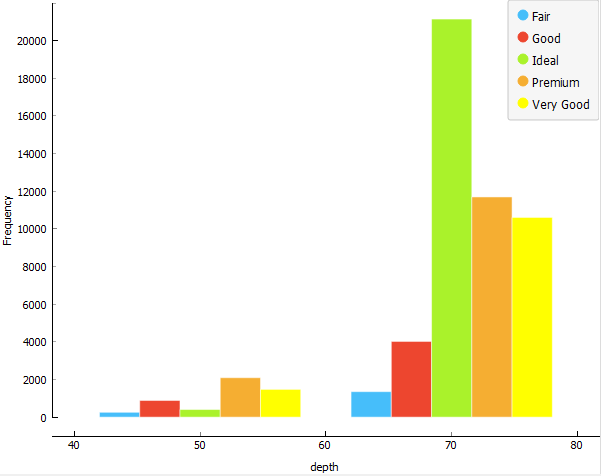




**Process**

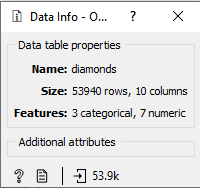


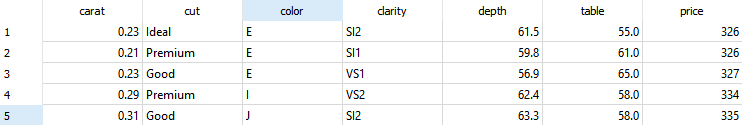
**Output**



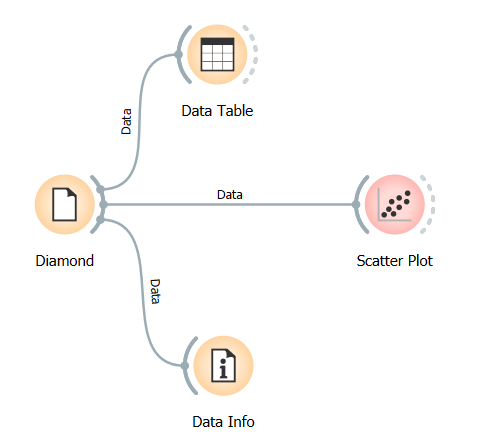
3) Make a scatterplot: carat vs price, set the color to clarity.

**Input**

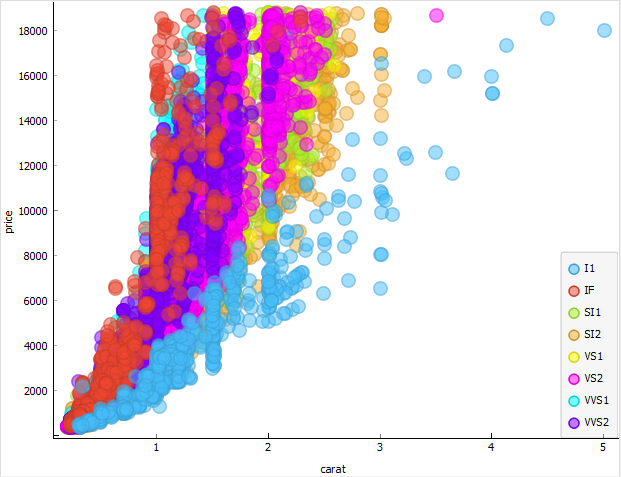




**Process**

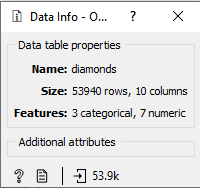


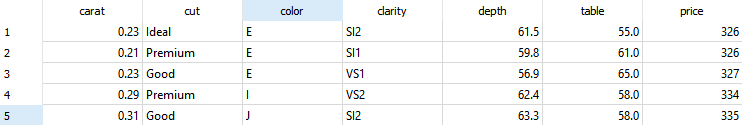
**Output**



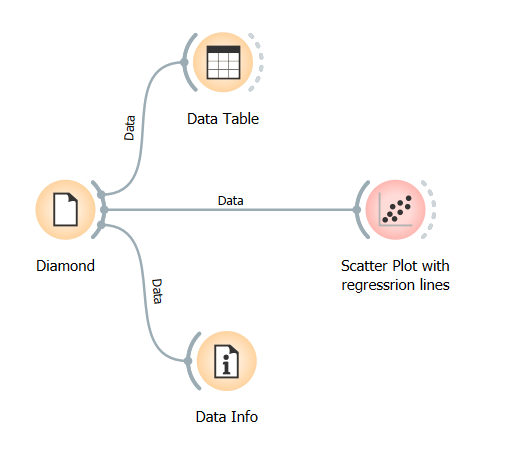
4) Make a scatterplot: carat vs price, set the color to clarity. Also add regression line to the plot.

**Input**

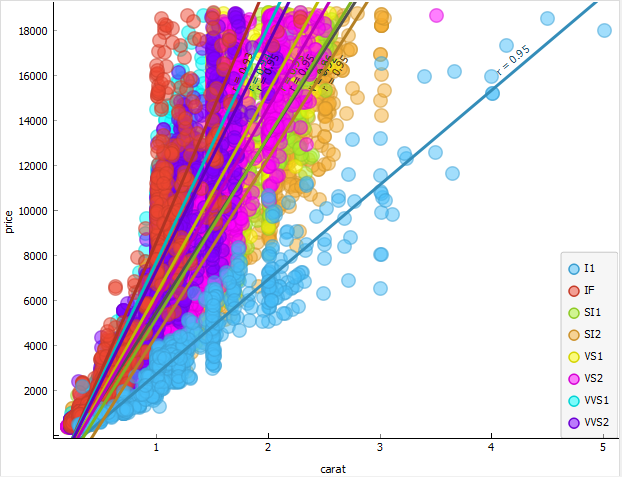




**Process**

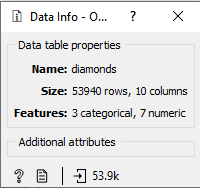


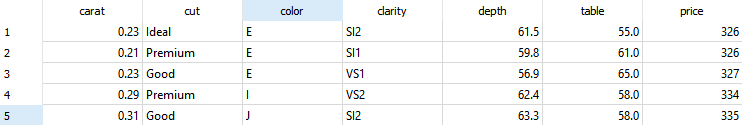
**Output**



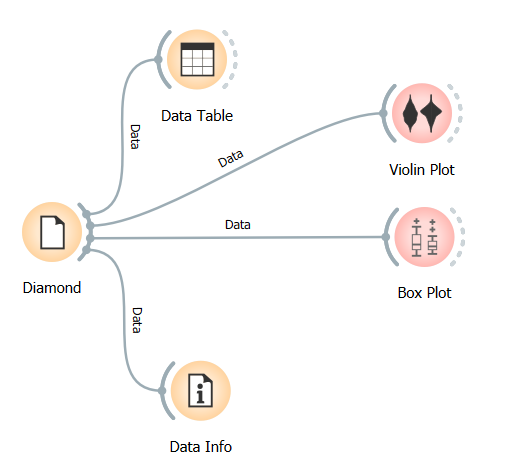
5) For carat vs cut, make a violin and a boxplot.

**Input**



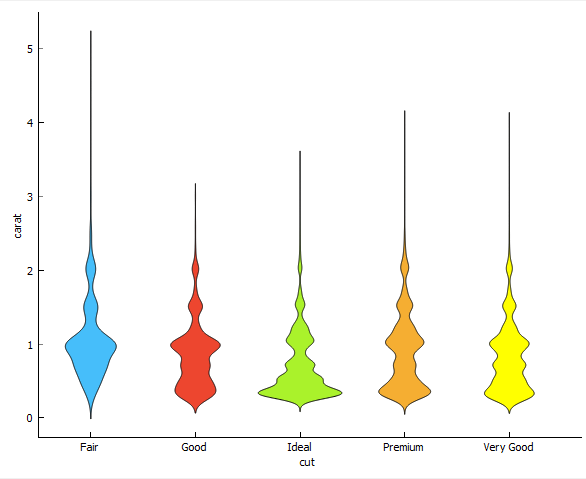


**Process**

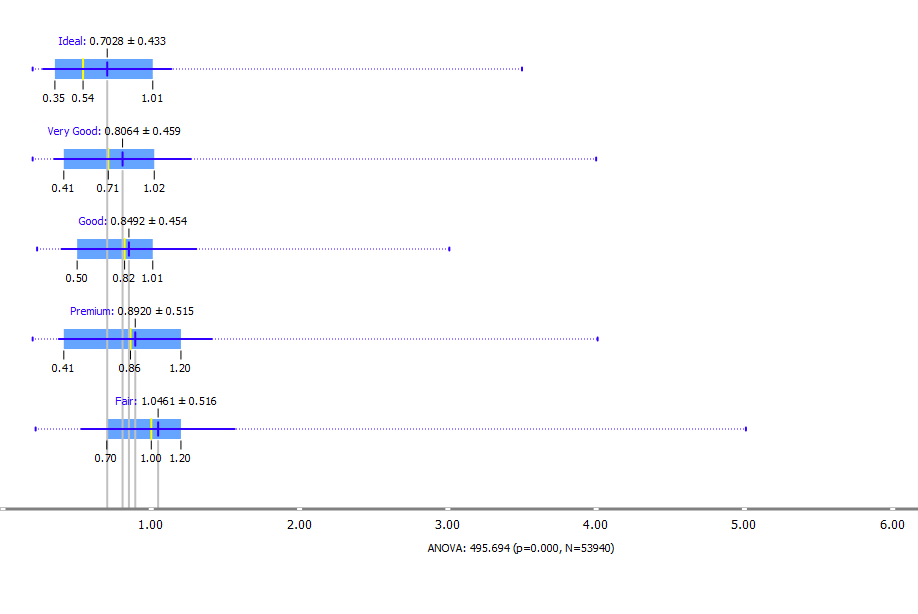


**Output**

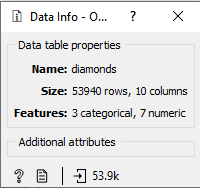
Output 1 (Violin plot)

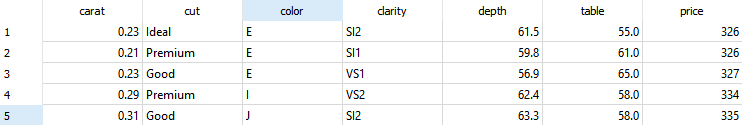


Output 2 (Box plot)

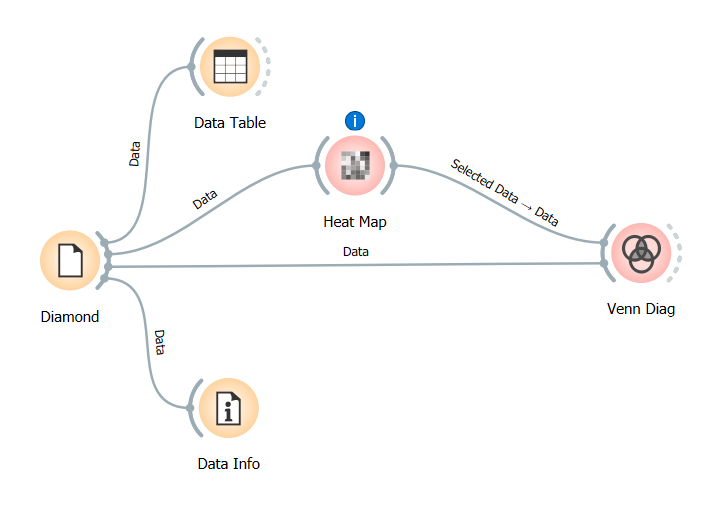
6) Illustrate Heat map and Venn Diagram using the data set.

**Input**



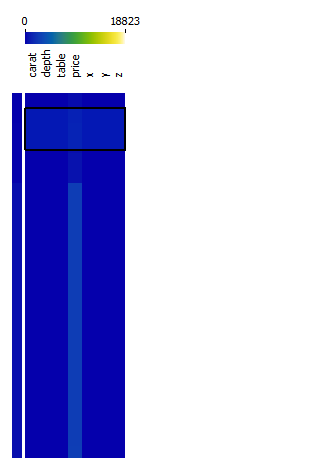


**Process**

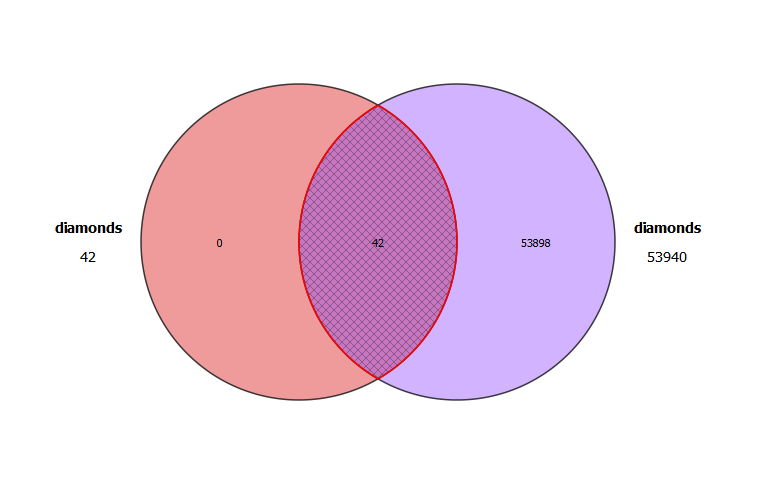


**Output**

Output 1 (Heat map)

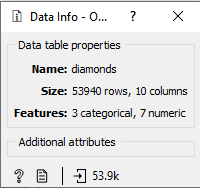


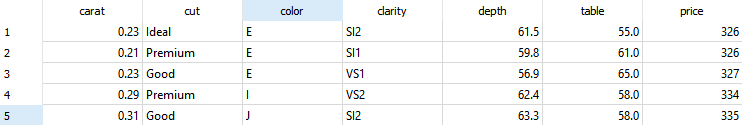
Output 2 (Venn Diagram)



7) Illustrate freeviz, linear projection and radviz using the data set.

**Input**





**Process**

**a Data Visualization (Weka)**

1) Give a visualization of the distribution of Iris dataset w.r.t all the features

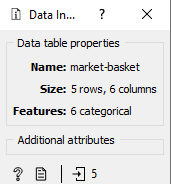
2) Display the plot matrix for the Iris data set

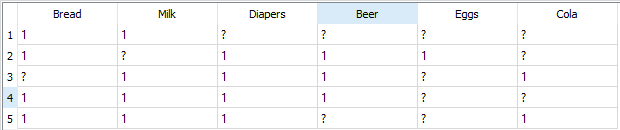
**Section III**

**a (Association Rule Mining -Class Work)**

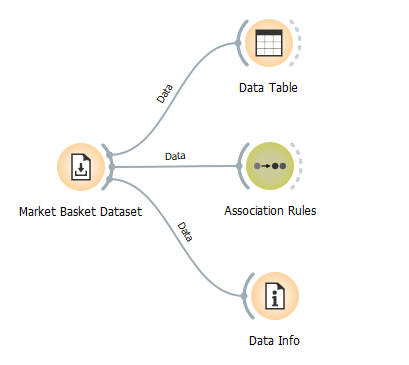
1) Generate association rules using Market Basket Data set in Orange Tool. Compare the different measures to assess the quality of rules.

**Input**

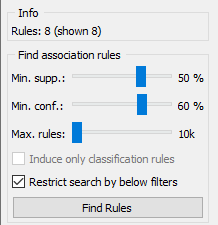


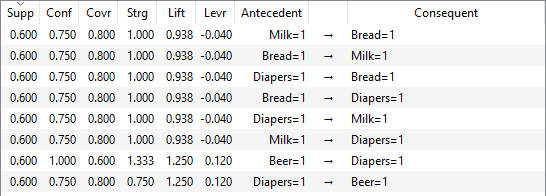


**Process**



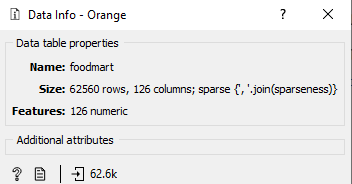
**Output**





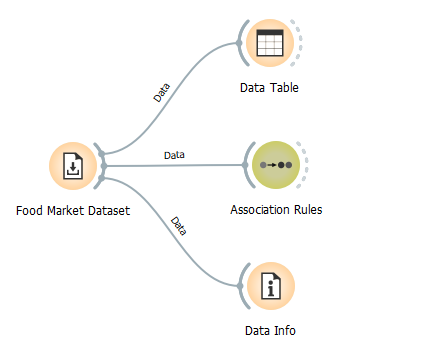
2) Generate association rules using Food mart Data set in Orange Tool. Compare the different measures to assess the quality of rules.

**Input**

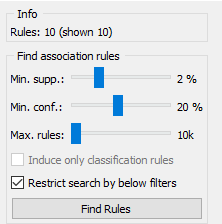


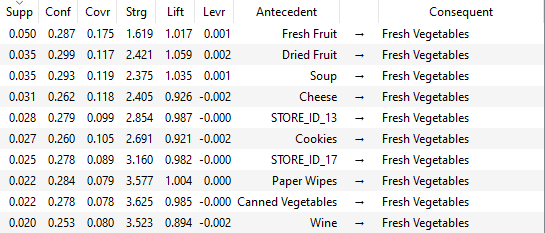


**Process**



**Output**





3) Generate association rules using supermarket Data set in WEKA using Apriori algorithm.

4) Generate association rules using supermarket Data set in WEKA using FP – growth Algorithm

**b (Association Rule Mining -Class Work)**

1) Generate association rules using Lenses Data set in Orange Tool

2) Generate association rules using Lenses Data set in WEKA using Apriori algorithm.

3) Generate association rules using Lenses Data set in WEKA using FP – growth Algorithm