

Date – 09-09-2021

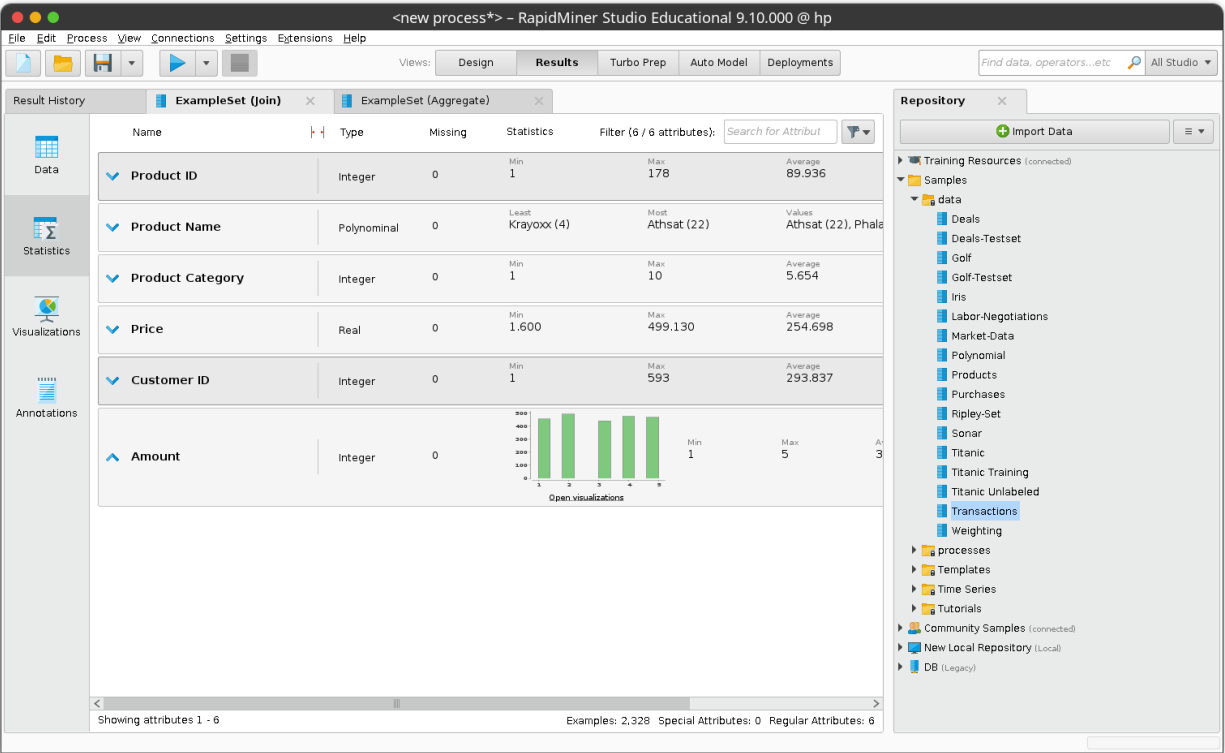
Activity 1 : To perform operations on data

Data sets used : Products, Transactions

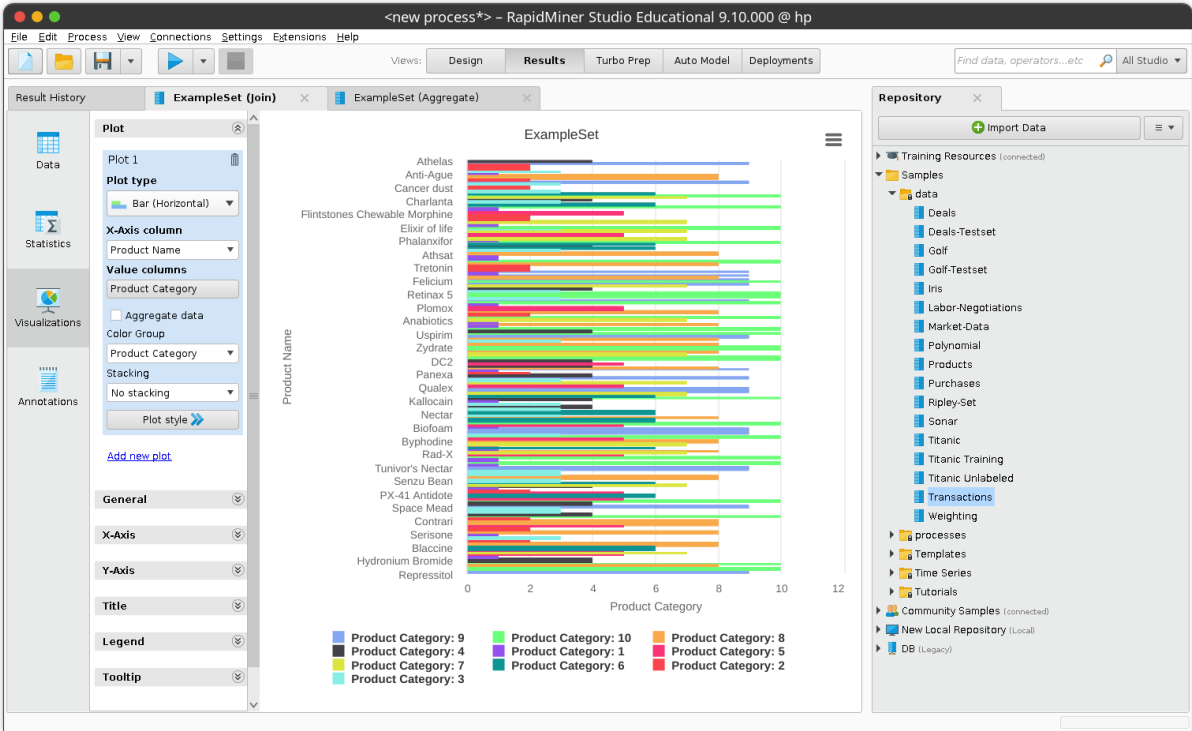
Steps:

- Expand the **Samples** repository in the **Repository** panel.
 - Next, expand the **data** folder within the samples repository to retrieve the **Products** and **Transactions** data.
 - Drag the **Products** data and the **Transactions** data from the **Samples - Data** folder into the **Process** panel.
 - Search for the **Join** operator in the search box at the top of the **Operator** panel. Drag **Join** into the **Process** panel.
 - Connect the output port of **Retrieve Products** to an input port of **Join** (it doesn't matter which one).
 - Connect **Retrieve Transactions** to the other **Join** input port.
 - Click on **Join** to select it. In the **Parameters** panel find the **key attributes** field.
 - Click **Edit List**. Select *Product ID* for the left and right key attributes. Then, click **Apply**.
 - Drag the **Aggregate** operator into the process. Connect it to the output of **Join**.
 - Click **Aggregate** to select. Make the following changes in the **Parameters** panel:
 - Click on **aggregation attributes**.
 - Select *Customer ID* in the left box and set **function** to *count* in the right box.
 - Stay in this dialog and add another entry *Product Name* with **function** set to *mode*. Click **Apply**.
 - Click on **group by attributes**. Then, select the *Product ID* by moving it to the right. Click **Apply**.
 - Connect **Aggregate** to the result port on the right.
- Press **Run** to execute the process.

Statistics



Visualization



Interpretation

In this activity, I took two different datasets of Products and Transactions and joined them to view the transactions of a particular product by using the common attribute Product ID. After joining the results, they were grouped using aggregate.

Date – 09-09-2021

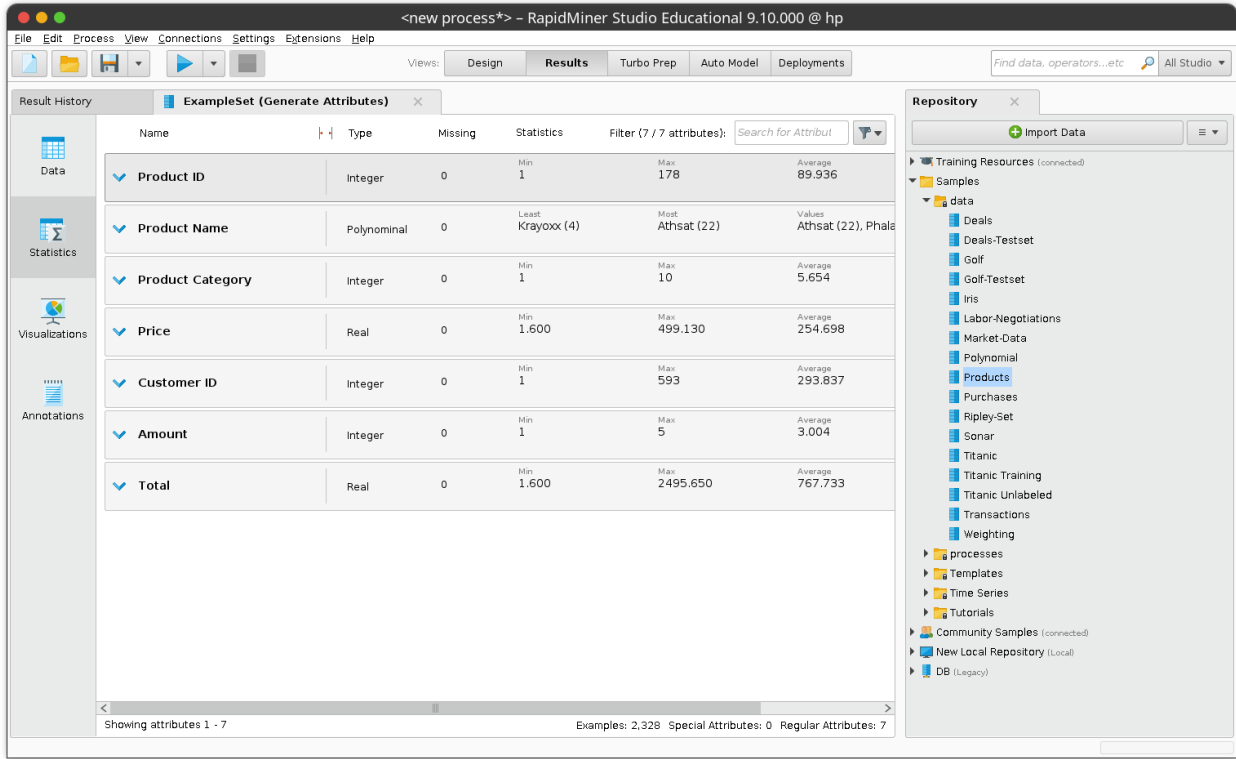
Activity 2 : Working with Attributes

Data sets used : Products, Transactions

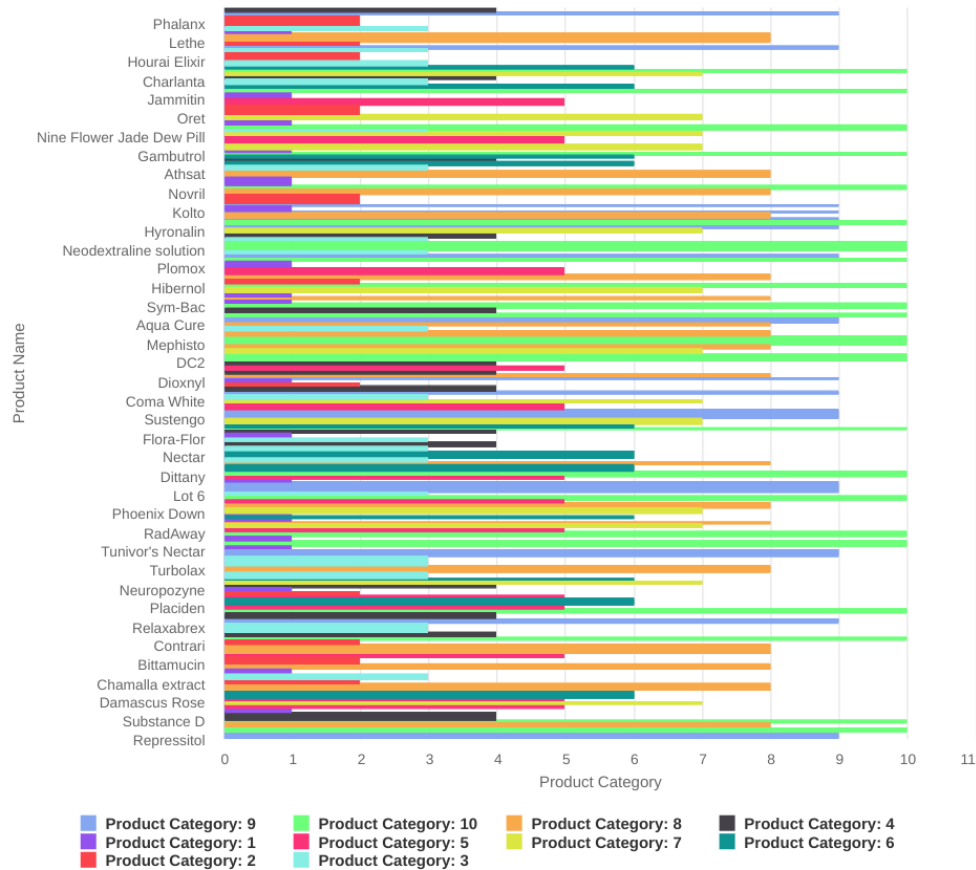
Steps:

- **Drag the Transactions and Products datasets into the Process.**
- Add a **Join** operator.
- Connect all operators.
- Specify the columns to use in the **Parameters** for **Join**. That is, click **Edit List**, and use *Product ID* for the left and right key attributes.
- Add the **Generate Attributes** operator.
- Connect the operator with **Join**.
- Click **Edit List** in the **Parameters** for **Generate Attributes** to define the new attribute (column). A dialog will pop up.
- In the left column of the dialog, enter *Total* for the **attribute name**.
- In the right column, type *Amount*Price* for the **function expression**.
- Add the **Select Attributes** operator to the process and connect it. Make the following changes in **Parameters**:
- Set **attribute filter type** to *subset*. In general, this means that the operator will be applied only to those attributes (columns) you specify. Here this allows you to choose a subset of columns to keep in the data - all other columns will be removed.
- Click **Select Attributes**.
- In the resulting dialog, select the *Customer ID*, *Product Name*, and *Total* attributes. If the list is empty, you forgot to connect the operator first...
- **Run** the process.

Statistics



Visualization



Interpretation