



# IA 1

UT 01 – PD2

Felipe Mestre

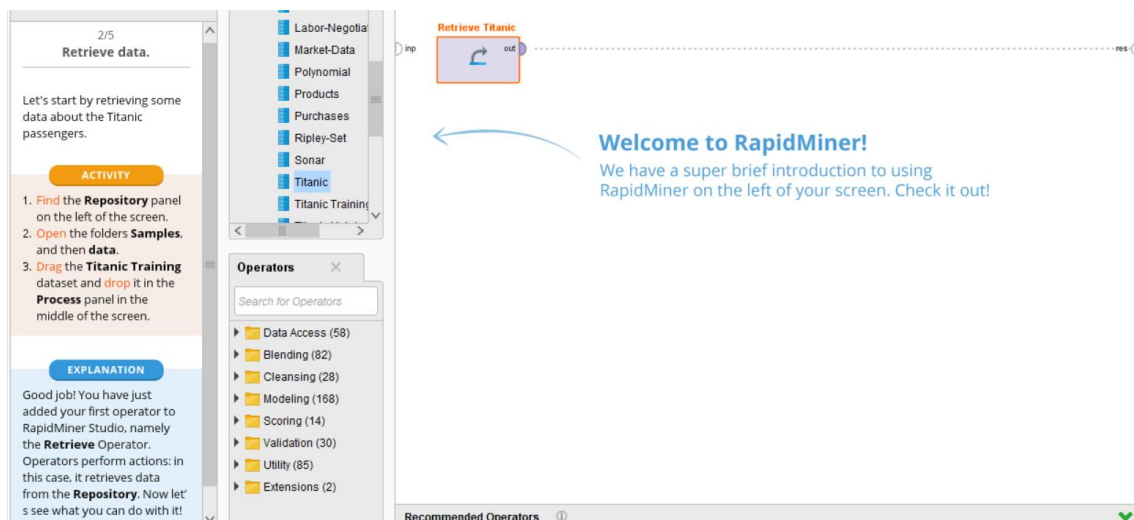
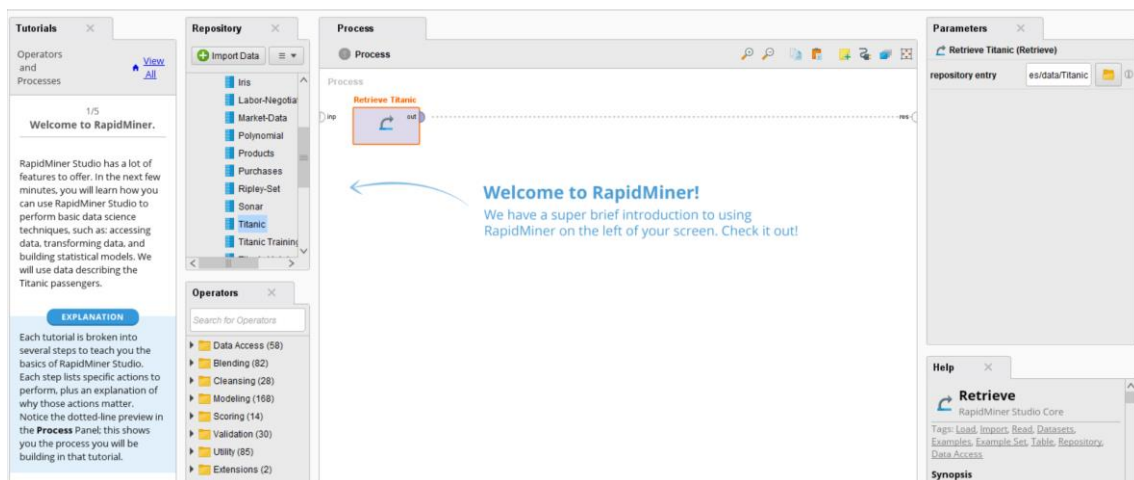
# Ejercicio 1

Sepal length		Sepal Width		Petal Length		Petal Width	
Media	5.843333333	Media	3.054	Media	3.75866667	Media	1.19866667
Error típico	0.067611316	Error típico	0.035402827	Error típico	0.14406432	Error típico	0.06231181
Mediana	5.8	Mediana	3	Mediana	4.35	Mediana	1.3
Moda	5	Moda	3	Moda	1.5	Moda	0.2
Desviación estándar	0.828066128	Desviación estándar	0.433594311	Desviación estándar	1.76442042	Desviación estándar	0.76316074
Varianza de la muestra	0.685693512	Varianza de la muestra	0.188004027	Varianza de la muestra	3.11317942	Varianza de la muestra	0.58241432
Curtosis	-0.552064041	Curtosis	0.290781062	Curtosis	-1.4019208	Curtosis	-1.33975417
Coefficiente de asimetría	0.314910957	Coefficiente de asimetría	0.334052662	Coefficiente de asimetría	-0.27446425	Coefficiente de asimetría	-0.10499656
Rango	3.6	Rango	2.4	Rango	5.9	Rango	2.4
Mínimo	4.3	Mínimo	2	Mínimo	1	Mínimo	0.1
Máximo	7.9	Máximo	4.4	Máximo	6.9	Máximo	2.5
Suma	876.5	Suma	458.1	Suma	563.8	Suma	179.8
Cuenta	150	Cuenta	150	Cuenta	150	Cuenta	150

Hay 49 valores de iris setosa, 50 de iris virginica y 50 de iris versicolor.

# Ejercicio 2

A continuación se detalla con capturas todos los pasos que seguí para completar los 8 tutoriales



create processes by adding and connecting operators. Operators are connected together via their Ports:

**ACTIVITY**

1. Connect the output port ("out") of **Retrieve Titanic Training** with the result port ("res") on the right side of the **Process** panel.
2. Make the connection either by **dragging a line** between the ports, or by first **clicking** on one port and then on the other port.

**EXPLANATION**

Excellent! You have built your first process in RapidMiner Studio. We can now look at the output of the Retrieve operator. Whenever you want to see the output of an operator, make sure it is connected to the 'res' port.

Labor-Negotia

Market-Data

Polynomial

Products

Purchases

Ripley-Set

Sonar

Titanic

Titanic Training

Operators

Search for Operators

Data Access (58)

Blending (82)

Cleansing (28)

Modeling (168)

Scoring (14)

Validation (30)

Utility (85)

Extensions (2)

Retrieve Titanic

inp

out

res

res

Welcome to RapidMiner!

We have a super brief introduction to using RapidMiner on the left of your screen. Check it out!

**ACTIVITY**

1. Click **Run** (top left) to execute the process.

**EXPLANATION**

Nice! You have run your first process in RapidMiner Studio. When you click "Run", the operators in your process will perform their actions. RapidMiner executes the process and shows the data connected to the result port. In the center of the **Results** view, you can see our raw data about the Titanic passengers such as their family size or their age. On the **Statistics** tab we get a statistics summary, which provides useful information: for example, 349 passengers survived the Titanic accident. If you get different results, please make sure you dragged in the correct dataset: **Titanic Training**.

Recommended Operators

Result History

ExampleSet (Retrieve Titanic)

Open in

Turbo Prep

Auto Model

Filter (1,309 / 1,309 examples): all

Row No.	Passenger ...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Num...	Passenger F...
1	First	Allen, Miss. E...	Female	29	0	0	24160	211.338
2	First	Allison, Mast...	Male	0.917	1	2	113781	151.550
3	First	Allison, Miss. ...	Female	2	1	2	113781	151.550
4	First	Allison, Mr. H...	Male	30	1	2	113781	151.550
5	First	Allison, Mrs. ...	Female	25	1	2	113781	151.550
6	First	Anderson, Mr...	Male	48	0	0	19952	26.550
7	First	Andrews, Mis...	Female	63	1	0	13502	77.958
8	First	Andrews, Mr. ...	Male	39	0	0	112050	0
9	First	Appleton, Mrs...	Female	53	2	0	11769	51.479
10	First	Artagaveytia, ...	Male	71	0	0	PC 17609	49.504
11	First	Astor, Col. Jo...	Male	47	1	0	PC 17757	227.525
12	First	Astor, Mrs. Jo...	Female	18	1	0	PC 17757	227.525
13	First	Aubart, Mme. ...	Female	24	0	0	PC 17477	69.300
14	First	Barber, Miss. ...	Female	26	0	0	19877	78.850
15	First	Barkworth, Mr...	Male	80	0	0	27042	30
16	First	Baumann, Mr...	Male	?	0	0	PC 17318	25.925
17	First	Baxter, Mr. Qu...	Male	24	0	1	PC 17558	247.521
18	First	Baxter, Mrs. I	Female	50	0	1	PC 17558	247.521

ExampleSet (1,309 examples, 0 special attributes, 12 regular attributes)

3/5

Build a decision tree.

Decision Trees are a popular statistical modeling technique that finds hidden patterns in the data. Let's build one!

ACTIVITY

- Find the **Operators** panel on the left.
- Open the folder **Modeling**, **Predictive**, and finally **Trees**.
- Drag in the **Decision Tree** operator and drop it after **Retrieve Titanic Training**.

EXPLANATION

Good, we now have the operator in the process but we still need to connect the two operations and define what we want to see as output before we can run it.

Back

Next

Iris

Labor-Negotia

Market-Data

Polynomial

Products

Purchases

Ripley-Set

Sonar

Titanic

Titanic Training

Operators

decision

Modeling (8)

Predictive (8)

Trees (8)

Decision Tree

Random Forest

Gradient Boosted

ID3

We found "WhiBo" and "Weka Extension" in the Marketplace. [Show me!](#)

Process

Retrieve Titanic

Decision Tree

inp

out

tra

mod

exa

wei

res

Recommended Operators

Apply Model

44%

Set Role

43%

Select Attributes

33%

Filter Examples

26%

Modeling

4/5

Connect & execute.

The **Decision Tree** operator will build a Decision Tree for us based on our Titanic Training data. We still need to connect these operators together to build a Process, though.

ACTIVITY

- Connect the output port ("out") of **Retrieve Titanic Training** with the input port ("tra") of **Decision Tree**. You can click on the ports, or you can drag on them!
- Connect the first output ("mod") of the **Decision Tree** with the result port ("res") on the right side of the **Process** panel.
- Press the **Run** button in the toolbar to execute the process.

Import Data

Iris

Labor-Negotiations

Market-Data

Polynomial

Products

Purchases

Ripley-Set

Sonar

Titanic

Titanic Training

Titanic Unlabeled

Transactions

Weighting

processes

Templates

Time Series

Operators

decision

Modeling (8)

Predictive (8)

Trees (8)

Decision Tree

Process

Retrieve Titanic

Decision Tree

inp

out

tra

mod

exa

wei

res

Decision Tree

criterion

gain\_ratio

maximal depth

10

apply pruning

confidence

0.1

apply prepruning

minimal gain

0.01

minimal leaf size

2

minimal size for split

4

number of prepruning alternatives

3

Hide advanced parameters

Change compatibility (0-10,000)

Help

Repository

Import Data

Iris

Labor-Negotiations

Market-Data

Polynomial

Products

Purchases

Ripley-Set

Sonar

Titanic

Titanic Training

Titanic Unlabeled

Transactions

Weighting

processes

Templates

Time Series

Process

Retrieve Titanic

Decision Tree

Set Role

inp

out

tra

mod

exa

wei

res

El tutorial tiene un error, porque pretende usar un decision tree sin setear un label.



El árbol resultante es ilegible, pero por ahora me voy a apegar al tutorial.

The screenshot displays the RapidMiner software interface during a data import process. On the left, a 'Tutorials' panel shows step 3 of a guide: 'When you complete the import, store the data as Titanic in your Local Repository.' The main workspace is divided into three panels: 'Repository', 'Process', and 'Process' (labeled as such in the interface). The 'Repository' panel on the left lists various data sources like 'Iris', 'Titanic', and 'Titanic Training'. The 'Process' panel on the right shows a list of operators, including 'Blending', 'Attributes', and 'Names & Roles'. The central area is a 'Select the data location.' dialog box. It shows a file explorer view with a table of files. The file 'Titanic+Data.xls' is selected. Below the table, it says 'The selected file will be imported as: Excel'. At the bottom, there are 'Previous', 'Next', and 'Cancel' buttons. Below this dialog, another 'Where to store the data?' dialog box is open. It shows a tree view of the 'Local Repository' with folders like 'data', 'processes', and 'housing'. The file 'titanic' is selected under the 'data' folder. At the bottom, there is a 'Name' field containing 'Titanic' and a 'Location' field showing 'J:/Local Repository/Titanic'. There are 'Previous', 'Finish', and 'Cancel' buttons at the bottom right.

2/6  
Download your data.

**ACTIVITY**

1. Download this Excel file to your computer.
2. To import the downloaded data into RapidMiner, click **Import Data** in the **Repository** panel and follow the steps in the wizard.
3. When you complete the import, store the data as **Titanic** in your **Local Repository**.

**EXPLANATION**

The **Repository** panel, in the upper left corner by default, is the place to store all your data, processes, and results. You should always import data into the repository, especially when it comes from files like XLS or CSV. This will simplify the design of analytical processes a lot since RapidMiner's repository stores describing meta data together with the data.

**Operators**

role

- Blending (6)
- Attributes (6)
  - Names & Roles (6)
    - Rename
    - Rename by Replacing
    - Rename by Example Value
    - Rename by Generic Name
    - Rename by Construction
    - Set Role

**Repository**

**Process**

Import Data - Where to store the data?

**Where to store the data?**

- Local Repository (Local)
  - data
    - Class 2 ( 8/13/21 7:43 PM - 9 KB)
    - Class 3 ( 8/24/21 7:09 PM - 4 KB)
    - housing ( 9/3/21 6:49 PM - 61 KB)
    - iris ( 9/5/21 12:22 AM - 7 KB)
    - titanic ( 8/20/21 8:38 PM - 234 KB)
    - wine ( 8/20/21 10:45 PM - 25 KB)

Name: Titanic

Location: J:/Local Repository/Titanic

Importo el csv del titánico

3/6

Add data to the process.

ACTIVITY

1. Click the **Design** tab to return to the **Process** panel.

2. Drag the imported **Titanic data** from the **Repository** panel into the **Process** panel.

EXPLANATION

When you drag data from the repository into the process, it transforms into a data-loading operator (in this case, **Retrieve Titanic**). Data is not actually loaded (or delivered at the round output ports of each operator) until you run the process.

Time Series

Tutorials

Community Samples (connected)

Local Repository (Local)

Connections

data

processes

Clase 2 ( 8/13/21 7:43 PM – 8 kB)

Clase 3 ( 8/24/21 7:09 PM – 4 kB)

housing ( 9/3/21 6:49 PM – 61 kB)

iris ( 9/5/21 12:22 AM – 7 kB)

titanic ( 8/20/21 6:38 PM – 234 kB)

Titanic+data ( 9/5/21 12:48 AM – 185 kB)

wine ( 8/26/21 10:45 PM – 25 kB)

DB (Legacy)

Operators

role

Process

Retrieve Titanic+data

inp

out

Result History

ExampleSet (Retrieve Titanic+data)

ExampleSet (/Local Repository/Titanic+data)

Open in Turbo Prep Auto Model

Filter (1,309 / 1,309 examples): all

Data

Statistics

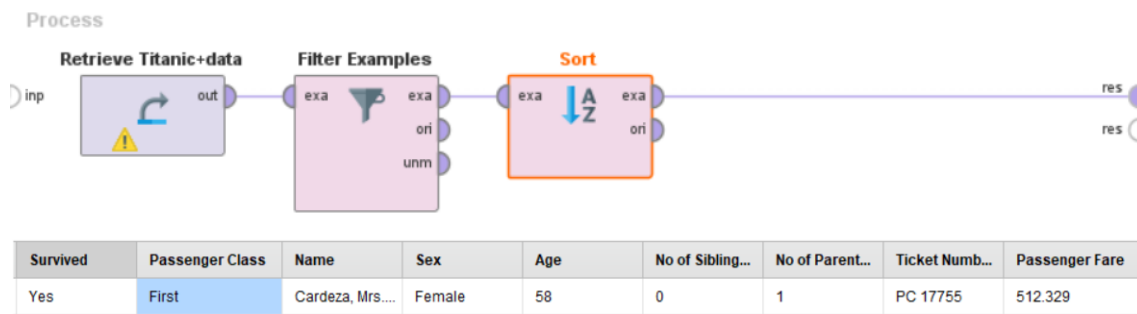
Visualizations

Annotations

Row No.	Survived	Passenger ...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Num...	Passenger F...	Cabin	Port of
1	Yes	First	Allen, Miss. E.	Female	29	0	0	24160	211.338	B5	Southa
2	Yes	First	Allison, Mast...	Male	0.917	1	2	113781	151.550	C22 C26	Southa
3	No	First	Allison, Miss. ...	Female	2	1	2	113781	151.550	C22 C26	Southa
4	No	First	Allison, Mr. H.	Male	30	1	2	113781	151.550	C22 C26	Southa
5	No	First	Allison, Mrs. ...	Female	25	1	2	113781	151.550	C22 C26	Southa
6	Yes	First	Anderson, Mr...	Male	48	0	0	19952	26.550	E12	Southa
7	Yes	First	Andrews, Mis...	Female	63	1	0	13502	77.958	D7	Southa
8	No	First	Andrews, Mr. ...	Male	39	0	0	112050	0	A36	Southa
9	Yes	First	Appleton, Mrs...	Female	53	2	0	11769	51.479	C101	Southa
10	No	First	Artagaveyfla, ...	Male	71	0	0	PC 17609	49.504	?	Cherbo
11	No	First	Astor, Col. Jo...	Male	47	1	0	PC 17757	227.525	C62 C64	Cherbo
12	Yes	First	Astor, Mrs. Jo...	Female	18	1	0	PC 17757	227.525	C62 C64	Cherbo
13	Yes	First	Aubart, Mme. ...	Female	24	0	0	PC 17477	69.300	B35	Cherbo
14	Yes	First	Barber, Miss. ...	Female	26	0	0	19877	78.850	?	Southa
15	Yes	First	Barkworth, Mr...	Male	80	0	0	27042	30	A23	Southa
16	No	First	Baumann, Mr...	Male	?	0	0	PC 17318	25.925	?	Southa
17	No	First	Baxter, Mr. Qu...	Male	24	0	1	PC 17558	247.521	B58 B60	Cherbo
18	Yes	First	Baxter, Mrs. J...	Female	50	0	1	PC 17558	247.521	B58 B60	Cherbo
19	Yes	First	Bazzani, Miss...	Female	32	0	0	11813	76.292	D15	Cherbo
20	No	First	Beattie, Mr. T...	Male	36	0	0	13050	75.242	C6	Cherbo
21	Yes	First	Beckwith, Mr. ...	Male	37	1	1	11751	52.554	D35	Southa
22	Yes	First	Beckwith, Mrs...	Female	47	1	1	11751	52.554	D35	Southa
23	Yes	First	Behr, Mr. Karl...	Male	26	0	0	111369	30	C148	Cherbo
24	Yes	First	Bidois, Miss. ...	Female	42	0	0	PC 17757	227.525	?	Cherbo
25	Yes	First	Bird, Miss. Ell...	Female	29	0	0	PC 17483	221.779	C97	Southa

ExampleSet (1,309 examples, 0 special attributes, 12 regular attributes)





Esa es la mujer que pago más

1/6  
Say hello to two new data sets...

Let's take a break from the Titanic and learn about some other frequently used tasks in data preparation, specifically merging and grouping data together. We will deal with two data sets: one containing the products sold by an organization and one with the transactions (information about which customer purchased which product). After combining these sets, we can answer questions about the most frequently purchased product or who is your most loyal customer. Let's get started...

processes

- Class 2 ( 8/13/21 7:43 PM)
- Class 3 ( 8/24/21 7:09 PM)
- housing ( 8/3/21 6:49 PM)
- iris ( 9/5/21 12:22 AM - 7)
- titanic ( 8/20/21 6:38 PM)
- Titanic+data ( 9/5/21 12:4)
- wine ( 8/25/21 10:45 PM)
- DB (Legacy)

Operators

- sort
- Blending (4)
- Attributes (1)
- Reorder Attributes
- Examples (3)

Repository

- Market-Data
- Polynomial
- Products
- Purchases
- Ripley-Set
- Sonar
- Titanic
- Titanic Training
- Titanic Unlabeled
- Transactions

Operators

- join
- Blending (8)
- Table (8)
- Joins (8)
- Append
- Join
- Set Minus

Process

Retrieve Products

Join

Aggregate

res

Retrieve Transactions

Process

Process

1. Search for the **Join** operator in the search box at the top of the **Operator** panel. Drag **Join** into the **Process** panel.

2. Connect the output port of **Retrieve Products** to an input port of **Join** (it doesn't matter which one).

3. Connect **Retrieve Transactions** to the other **Join** input port.

4. Click on **Join** to select it. In the **Parameters** panel find the **key attributes** field.

5. Click **Edit List**. Select **Product ID** for the left and right key attributes. Then, click **Apply**.



### Edit Parameter List: key attributes

The attributes which shall be used for join. Attributes which shall be matched must be of the same type.

left key attributes	right key attributes
Product ID	Product ID

Add Entry

Remove Entry

Apply

Cancel

**ACTIVITY**

1. Drag the **Aggregate** operator into the process. Connect it to the output of **Join**.
2. Click **Aggregate** to select. Make the following changes in the **Parameters** panel:
3. Click on **aggregation attributes**.
4. Select **Customer ID** in the left box and set **function** to **count** in the right box.
5. Stay in this dialog and add another entry **Product Name** with function set to **mode**. Click **Apply**.
6. Click on **group by attributes**. Then, select the **Product ID** by moving it to the right. Click **Apply**.

Polynomial

Products

Purchases

Ripley-Set

Sonar

Titanic

Titanic Training

Titanic Unlabeled

Transactions

Retrieve Products

Join

Retrieve Transactions

Aggregate

Operators

agreg

Blending (4)

Attributes (1)

Generation (1)

Generate Aggregate

Table (3)

Grouping (1)

aggregation attributes

Edit List (2)

group by attributes

Select Attributes

count all combinations

only distinct

ignore missings

Hide advanced parameters

Change compatibility (9.10.000)

Help

Aggregate

RapidMiner Studio Core

Tags: Groupby, Group by, Grouping, Sum, Count, Min, Max, Average, Avg, Mean, Pivot, Cross-table, Crosstable, Distinct, Percentile, Standard Deviation

6/6

### Congratulations!

Great, you just made your first steps for data blending with RapidMiner! Before you move on to the next tutorial, try to answer the questions below:

**CHALLENGE**

- Which product has been most often sold? And which product was sold only 5 times?
- Can you find out in the **Statistics** tab what the average number of transactions was? Can you also see the visual distribution of values in this tab?
- The *count* function counts the number of transactions for each product, but each product can also be purchased multiple times in each transaction. Can you change the parameters of **Aggregate** so that the total sum for each product is calculated? Which products have been sold more than 350 times?

[Back](#) [Next Tutorial](#)

Data

Statistics

Visualizations

Annotations

Row No.	Product ID	Customer ID	count(Custo...	mode(Product Name)
1	1	31	1	Repressitol
2	1	47	1	Repressitol
3	1	62	1	Repressitol
4	1	64	1	Repressitol
5	1	79	1	Repressitol
6	1	82	1	Repressitol
7	1	99	1	Repressitol
8	1	125	1	Repressitol
9	1	191	1	Repressitol
10	1	220	1	Repressitol
11	1	231	1	Repressitol
12	1	256	1	Repressitol
13	1	334	1	Repressitol
14	1	352	1	Repressitol
15	1	365	1	Repressitol
16	1	408	2	Repressitol
17	1	447	1	Repressitol
18	1	566	1	Repressitol

Tutorials

Creating and Removing Columns

1/5

Working with Attributes.

You are almost ready to build your first predictive model in RapidMiner! But first, we need to deal with two very important operations for transforming your data sets into a format more suitable for training your model. The beginning of this process is the same as the previous one; this is a great opportunity for you to implement what you have learned so far. We will then create a new data column i.e. generate an attribute and remove some unused/unnecessary columns from the data.

Back

Next

Repository

Import Data

Market-Data

Polynomial

Products

Purchases

Ripley-Set

Sonar

Titanic

Titanic Training

Titanic Unlabeled

Transactions

Operators

agreg

Blending (4)

Attributes (1)

Generation (1)

Generate Aggreg

Table (3)

Grouping (1)

Aggregate

Rotation (2)

Pivot

No results were found.

Process

Process

Retrieve Products

Join

Generate Attributes

Select Attributes

Retrieve Transactions

Recommended Operators

Retrieve

Select Attributes

Set Role

Edit Parameter List: function descriptions  
List of functions to generate.

attribute name	function expressions
Total	Amount*Price

Add Entry

Remove Entry

Apply

Cancel

Tutorials

Creating and Removing Columns

[View All](#)

5/5

Congratulations!

You are becoming a master of data blending! You have seen now some of the most important operators for data preprocessing: Join, Aggregate, Filter, Sort, Generate Attributes, and Select Attributes. RapidMiner has a lot more operators, but those six will be used most frequently.

CHALLENGE

○ From the **Results** view, can you find out the customer ID of the customer who paid most for a single product? How much did they pay? Remember that you can sort data by clicking on a column header.

○ Can you answer the same question by using an operator instead?

○ How would you describe the

Result History

Data

Statistics

Visualizations

Annotations

ExampleSet (Select Attributes)

Open in

Turbo Prep

Auto Model

Row No.	Product Na...	Customer ID	Total
1	Repressitol	220	142.160
2	Repressitol	125	106.620
3	Repressitol	352	71.080
4	Repressitol	408	106.620
5	Repressitol	334	35.540
6	Repressitol	31	71.080
7	Repressitol	62	106.620
8	Repressitol	79	106.620
9	Repressitol	64	35.540
10	Repressitol	191	106.620
11	Repressitol	365	142.160
12	Repressitol	231	142.160
13	Repressitol	566	142.160
14	Repressitol	99	71.080
15	Repressitol	447	177.700
16	Repressitol	82	71.080
17	Repressitol	47	177.700
18	Repressitol	256	71.080

Tutorials

Changing Types and Roles

View All

1/5

What to predict?

To learn about building predictive models, let's go back to the Titanic data from earlier. This data is somewhat simpler and does not need a lot of preparation, but we still need to specify which column we want to predict!

Repository

Import Data

Polynomial  
Products  
Purchases  
Ripley-Set  
Sonar  
Titanic  
Titanic Training  
Titanic Unlabeled  
Transactions  
Weighting

Operators

select

Blending (9)

Attributes (7)

Selection (7)

Select Attributes  
Select by Weights  
Select by Random  
Remove Attribute

We found "Information Selection", "Repository Subset Selector" and one

Process

Process

inp

Retrieve Titanic

Discretize

Set Role

out

Recommended Operators

Tutorials

Changing Types and Roles

View All

2/5

Add product details to transactions.

ACTIVITY

1. Drag the **Titanic** data into the process.
2. Now, hover the mouse over the output port of the **Retrieve** operator and wait for a small window to pop up and display some meta data about the Titanic dataset. Some of the information you know from the **Statistics** tab.
3. Note the two columns **Role** and **Type** in the table at the bottom.

Repository

Import Data

Polynomial  
Products  
Purchases  
Ripley-Set  
Sonar  
Titanic  
Titanic Training  
Titanic Unlabeled  
Transactions  
Weighting

Operators

select

Blending (9)

Attributes (7)

Selection (7)

Select Attributes  
Select by Weights  
Select by Random  
Remove Attribute

We found "Information Selection", "Repository Subset Selector" and one

Process

Process

inp

Retrieve Titanic

Discretize

Set Role

out

Tutorials

Changing Types and Roles

View All

3/5

Define attribute types and roles.

ACTIVITY

1. Add a **Discretize by Binning** operator and connect it. In its **Parameters**:
  - Set the **attribute filter type** to **single** (i.e. you only work on one of the attributes).
  - Set **attribute** to **Age**.
  - Set **number of bins** to **3**.

Repository

Import Data

Polynomial  
Products  
Purchases  
Ripley-Set  
Sonar  
Titanic  
Titanic Training  
Titanic Unlabeled  
Transactions  
Weighting

Operators

discretize

Cleansing (5)

Binning (5)

Discretize by Size  
Discretize by Binning  
Discretize by Threshold

Process

Process

inp

Retrieve Titanic

Discretize

Set Role

out

Parameters

Discretize (Discretize by Binning)

create view

attribute filter type

single

attribute

Age

invert selection

include special attributes

number of bins

3

define boundaries

Hide advanced parameters

Change compatibility (9.10.000)

Help

Discretize by Binning
RapidMiner Studio Core
Tag: Continuous, Categorical, Nominal, Polynomial

attribute name: Survived

target role: regular

set additional roles: Edit List (0)...

Row No. Age Passenger ... Name Sex No of Sibling... No of Parent... Ticket Num... Passenger F... Cabin

1	range1 [-= - 4...	First	Allen, Miss. E...	Female	0	0	24160	211.338	B5
2	range1 [-= - 4...	First	Allison, Mast...	Male	1	2	113781	151.550	C22 C26
3	range1 [-= - 4...	First	Allison, Miss...	Female	1	2	113781	151.550	C22 C26
4	range1 [-= - 4...	First	Allison, Mr. H...	Male	1	2	113781	151.550	C22 C26
5	range1 [-= - 4...	First	Allison, Mrs. ...	Female	1	2	113781	151.550	C22 C26
6	range2 [40.0...	First	Anderson, Mr...	Male	0	0	19952	26.550	E12
7	range2 [40.0...	First	Andrews, Mis...	Female	1	0	13502	77.958	D7
8	range1 [-= - 4...	First	Andrews, Mr. ...	Male	0	0	112050	0	A36

Get started (8/8)

## Ejercicio 3

1-

Row No.	id	label	a1	a2	a3	a4
1	id_1	Iris-setosa	5.100	3.500	1.400	0.200

Se agrega una columna id para cada fila, la label es la variable que queremos predecir que en este caso es el tipo de la planta. Las columnas mencionadas son de tipo nominal y el resto de tipo real. Los nombres de las cuatro últimas columnas no corresponden con los nombres dados por el archivo de descripción del dataset de uci.

2- Usando el operador rename cambié los nombres de las variables a los que están en iris.name

Process

Retrieve Iris

Rename

Edit Parameter List: rename attributes

Use this list to define the renaming of the attributes.

old name	new name
a1	sepal length in cm
a2	sepal width in cm
a3	petal length in cm
a4	petal width in cm

Add Entry Remove Entry Apply Cancel

3-

Id	Nominal	0	Least id_99 (1)	Most id_1 (1)	Values id_1 (1), id_10 (1), ...[148 more]
Label	Nominal	0	Least Iris-virginica (50)	Most Iris-setosa (50)	Values Iris-setosa (50), Iris-versicolor (50), ...[1 more]
sepal length in cm	Real	0	Min 4.300	Max 7.900	Average 5.843
sepal width in cm	Real	0	Min 2	Max 4.400	Average 3.054
petal length in cm	Real	0	Min 1	Max 6.900	Average 3.759
petal width in cm	Real	0	Min 0.100	Max 2.500	Average 1.199

Sepal length		Sepal Width		Petal Length		Petal Width	
Media	5.843333333	Media	3.054	Media	3.75866667	Media	1.19866667
Error típico	0.067611316	Error típico	0.035402827	Error típico	0.14406432	Error típico	0.06231181
Mediana	5.8	Mediana	3	Mediana	4.35	Mediana	1.3
Moda	5	Moda	3	Moda	1.5	Moda	0.2
Desviación estándar	0.828066128	Desviación estándar	0.433594311	Desviación estándar	1.76442042	Desviación estándar	0.76316074
Varianza de la muestra	0.685693512	Varianza de la muestra	0.188004027	Varianza de la muestra	3.11317942	Varianza de la muestra	0.58241432
Curtosis	-0.552064041	Curtosis	0.290781062	Curtosis	-1.4019208	Curtosis	-1.33975417
Coeficiente de asimetría	0.314910957	Coeficiente de asimetría	0.334052662	Coeficiente de asimetría	-0.27446425	Coeficiente de asimetría	-0.10499656
Rango	3.6	Rango	2.4	Rango	5.9	Rango	2.4
Mínimo	4.3	Mínimo	2	Mínimo	1	Mínimo	0.1
Máximo	7.9	Máximo	4.4	Máximo	6.9	Máximo	2.5
Suma	876.5	Suma	458.1	Suma	563.8	Suma	179.8
Cuenta	150	Cuenta	150	Cuenta	150	Cuenta	150

Podemos ver que todos los datos obtenidos son congruentes con el primer ejercicio.