

# TRATAMENTO DE DADOS



MANOELA KOHLER

[Prof.manoela@ica.ele.puc-rio.br](mailto:Prof.manoela@ica.ele.puc-rio.br)

# TÓPICOS

## R

### Análise exploratória

### Pré-processamento

- Balanceamento
- *Outliers*
- *Missing values*
- Normalização
- Seleção de atributos (Filtros, Wrappers, PCA)

### Associação:

- Apriori
- *FP-Growth*
- *Eclat*

### Classificação:

- Regressão logística
- *Support Vector Machine (SVM)*
- Árvores de Decisão
- *Random Forest*
- Redes Neurais
- *K nearest neighbors*

### Regressão

- Regressão linear simples
- Regressão linear múltipla
- Regressão não linear simples
- Regressão não linear múltipla

### Agrupamento

- Particionamento (K-means, K-medoids)
- Hierárquico (DIANA, AGNES)
- Densidade (DBSCAN)

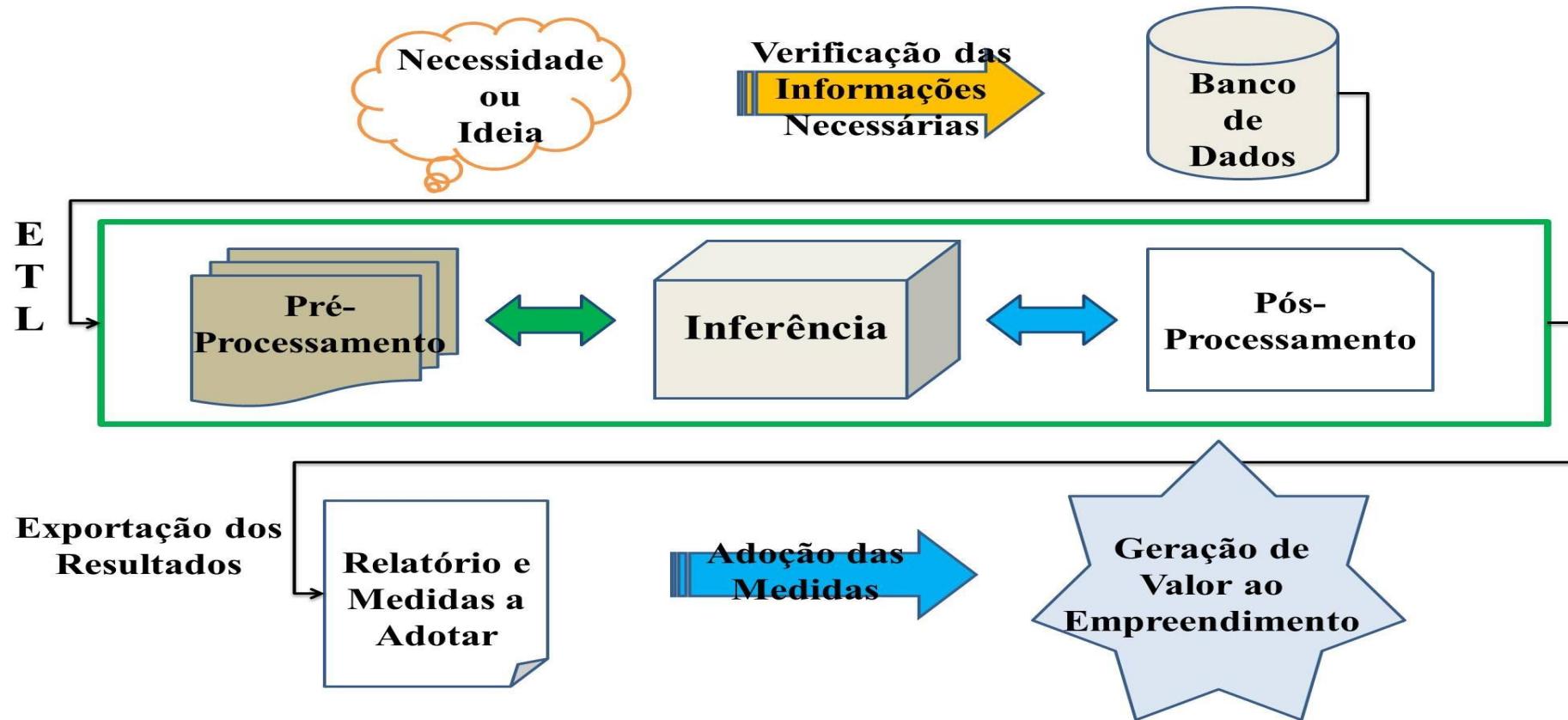
### Séries Temporais

- Naive
- Média Móvel
- Amortecimento exponencial
- Auto-regressivo integrados de média móvel
- Auto regressivo não linear

# Recapitulação

# ETAPAS DE UM PROJETO DE DATA MINING

# ESQUEMA BÁSICO DE UM PROJETO DE DM



# Análise Exploratória de Dados

# Tratamento de Dados

## **Pré-processamento**

Balanceamento

*Outliers*

*Missing values*

Normalização

Seleção de atributos (Filtros, Wrappers, PCA)



# ESTUDO DE CASO

<http://archive.ics.uci.edu/ml/datasets/SECOM>

Um complexo processo de fabricação de semi-condutores moderno é normalmente feito sob vigilância consistente através do monitoramento de sinais, variáveis coletadas de sensores e pontos de medição de processo.

No entanto, nem todos estes sinais são igualmente valiosos num sistema de monitoramento específico. Os sinais medidos contêm uma combinação de informações úteis, informações irrelevantes, bem como ruído.

Engenheiros tipicamente têm um número muito maior de sinais do que são realmente necessários. Se considerarmos cada tipo de sinal como uma característica, então a seleção da característica pode ser aplicada para identificar os sinais mais relevantes.

Os Engenheiros de Processo podem então usar estes sinais para determinar os fatores chave para o processo de produção. Isso poderá trazer benefícios para o processo, como redução do tempo e diminuição nos custos de produção.

**Base de dados SECOM: 1567 exemplos e 591 atributos**



# RESULTADOS

SVM	Acurácia	Kappa	Recall	Precisão
Missing	93.43%	0.018	50.48%	96.71%
+Balanceamento	80.29%	0.606	80.29%	80.31%
+Outliers	86.47%	0.729	86.44%	87.39%
Wrapper	84.54%	0.690	84.48%	86.53%
Ranker	71.98%	0.439	71.91%	74.30%
PCA	<b>98.55%</b>	<b>0.971</b>	<b>98.55%</b>	<b>98.55%</b>
SMOTE	90.87%	0.817	90.87%	91.00%

# R Script

# IBM ANALYTICS EMPLOYEE ATTRITION & PERFORMANCE

Predict attrition of your valuable employees

- Base de dados

Descobrir os fatores que levam ao desgaste dos funcionários e explorar questões importantes, tais como "mostrar-me a influência da distância de casa em função do cargo e atrito" ou "comparar renda média mensal por educação e atrito". Este é um conjunto de dados fictícios criado por cientistas de dados da IBM.

# IBM ANALYTICS EMPLOYEE ATTRITION & PERFORMANCE

Predict attrition of your valuable employees

- [Base de datos](#)

## **34 atributos:**

**Education:** 1 'Below College' 2 'College' 3 'Bachelor' 4 'Master' 5 'Doctor'

**EnvironmentSatisfaction:** 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

**JobInvolvement:** 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

**JobSatisfaction:** 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

**PerformanceRating:** 1 'Low' 2 'Good' 3 'Excellent' 4 'Outstanding'

**RelationshipSatisfaction:** 1 'Low' 2 'Medium' 3 'High' 4 'Very High'

**WorkLifeBalance:** 1 'Bad' 2 'Good' 3 'Better' 4 'Best'

Gender, MaritalStatus, MonthlyIncome, Age, DistanceFromHome, OverTime, JobRole, YearsWithCurrentManager, Over18, etc...



# IBM ANALYTICS EMPLOYEE ATTRITION & PERFORMANCE

Predict attrition of your valuable employees

- Base de dados

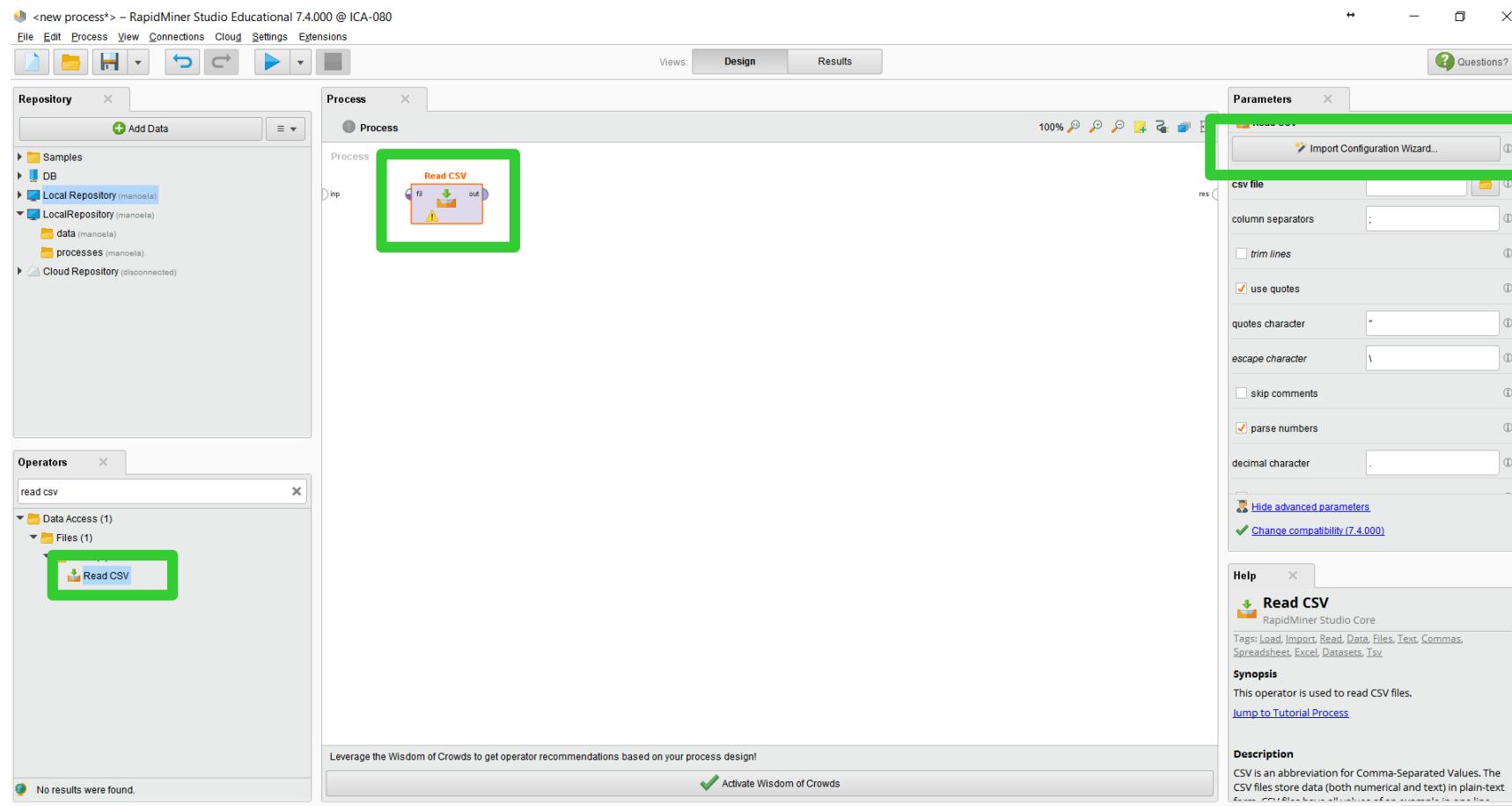
**2 classes:**

Attrition: Yes / No

# **TRATAMENTO DE VALORES FALTANTES**

## **(Missing Values)**

# 1. IMPORTAR A BASE



<new process\*> – RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results Questions?

**Repository**

+ Add Data

- Samples
- DB
- Local Repository (manoela)
- LocalRepository (manoela)
  - data (manoela)
  - processes (manoela)
- Cloud Repository (disconnected)

**Process**

Process Data import wizard - Step 1 of 4

This wizard guides you to import your data.  
Step 1: Please select the file that should be imported.

Aula02

Bookmarks	File Name	Size	Type	Last Modified
★ -- Last Directory	R	222 KB	File Folder	May 2, 2017
	WA_Fn-UseC_-HR-Employee-Attrition.csv	221 KB	Microsoft Excel Comma Se...	Mar 31, 2017
	WA_Fn-UseC_-HR-Employee-Attrition_Missing.csv	221 KB	Microsoft Excel Comma Se...	May 2, 2017

WA\_Fn-UseC\_-HR-Employee-Attrition\_Missing.csv

Delimiter separated files (.csv, .tsv)

Previous Next Finish Cancel

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

**Parameters**

Read CSV

Import Configuration Wizard...

csv file:

column separators:

trim lines

use quotes

quotes character:

escape character:

skip comments

parse numbers

decimal character:

Hide advanced parameters

Change compatibility (7.4.000)

**Operators**

read csv

Data Access (1)

Files (1)

Read (1)

Read CSV

No results found.

**Help**

Read CSV

RapidMiner Studio Core

Tags: Load, Import, Read, Data, Files, Text, Commas, Spreadsheet, Excel, Datasets, Tsv

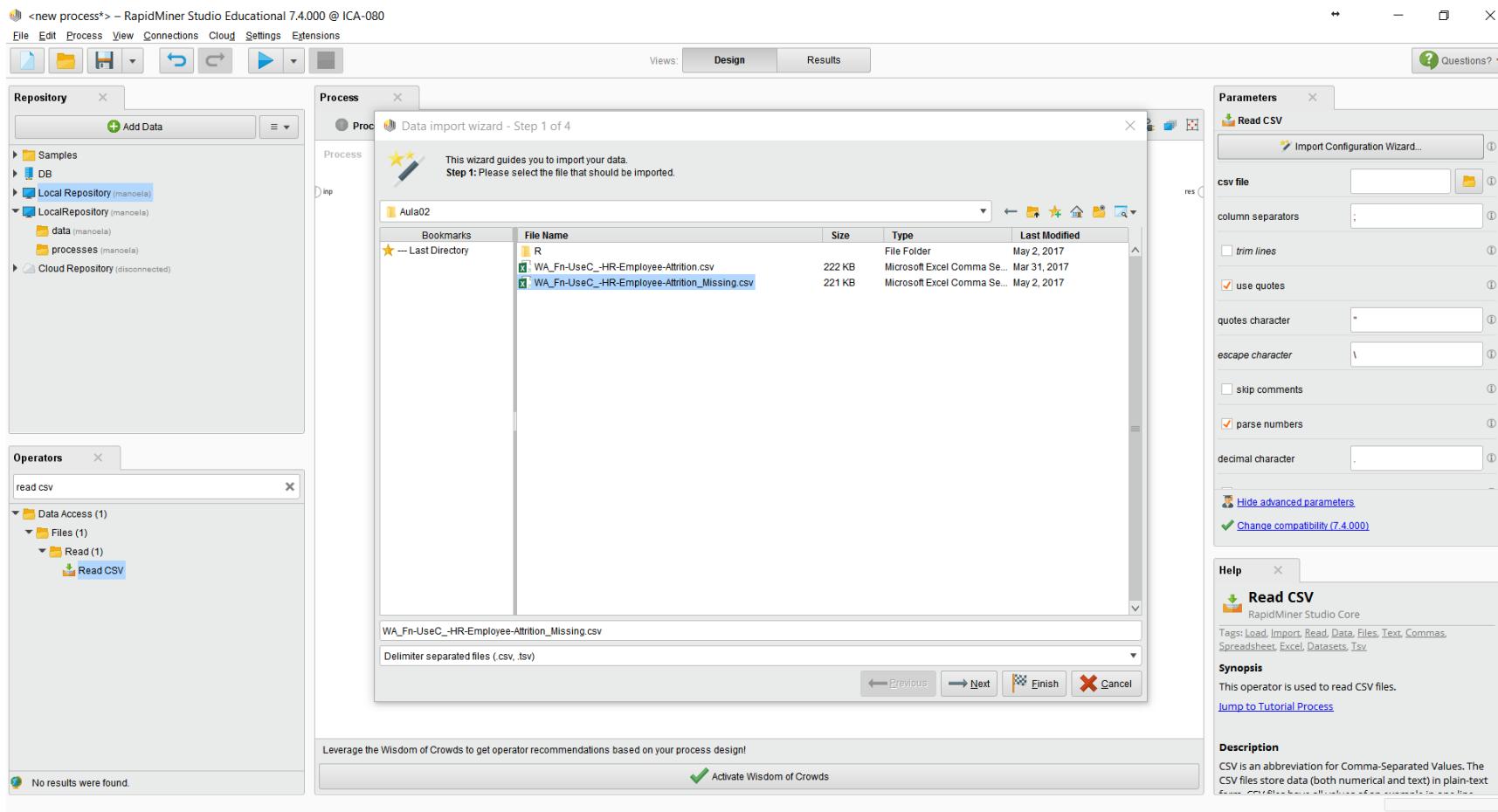
**Synopsis**

This operator is used to read CSV files.

[Jump to Tutorial Process](#)

**Description**

CSV is an abbreviation for Comma-Separated Values. The CSV files store data (both numerical and text) in plain-text



 Data import wizard - Step 2 of 4

This wizard guides you to import your data.  
Step 2: Please specify how the file should be parsed and how columns are separated.

**File Reading**

File Encoding: windows-1252

Trim Lines

Skip Comments: #

**Column Separation**

Comma ","

Space

Semicolon ":"

Tab

Regular Expression: ,\s\*|\s\*

Escape Character: \

Use Quotes: "

**Data Preview**

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EmployeeType	EnvironmentSatisfaction	Gender	HourlyRate
41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2	Female	94	
53	No	Travel_Rarely	102	Research & Development	23	4	Life Sciences		901	4	Female	72	
37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other		4	4	Male	92	
39	No	Non-Traveler	105	Research & Development	9	3	Life Sciences		2022	4	Male	87	
31	No	Travel_Rarely	106	Human Resources	2	3	Human Resources		424	1	Male	62	
40	No	Travel_Rarely	107	Sales	10	3	Technical		1239	2	Female	84	
55	No		111	Sales	1	2	Life Sciences		106	1	Male	70	
30	No	Non-Traveler	111	Research & Development	9	3	Medical		239	3	Male	66	
44	No	Non-Traveler	111	Research & Development	17	3	Life Sciences		1206	4	Male	74	
32	No	Travel_Frequently	116	Research & Development	13	3	Other		1234	3	Female	77	
39	No		116	Research & Development	24	1	Life Sciences		2014	1	Male	52	

**Error Log**

Row, Column	Error	Original value	Message
1,1			

**Buttons**

← Previous    → Next     Finish     Cancel

 Data import wizard - Step 3 of 4

This wizard guides you to import your data.

Step 3: In RapidMiner Studio, each attribute can be annotated. The most important annotation of an attribute is its name - a row with this annotation defines the names of the attributes. If your data does not contain attribute names, do not set this property. If further annotations are contained in the rows of your data file, you can assign them here.

Annotat...	att1	att2	att3	att4	att5	att6	att7	att8	att9	att10	att11	att12
Name	Age	Attrition	Business...	DailyRate	Departm...	Distance...	Education	Edutatio...	Employee...	Employee...	Environ...	Gender
-	41	Yes	Travel_R...	1102	Sales	1	2	Life Scie...	1	1	2	Female
-	53	No	Travel_R...	102	Researc...	23	4	Life Scie...		901	4	Female
-	37	Yes	Travel_R...	1373	Researc...	2	2	Other		4	4	Male
-	39	No	Non-Trav...	105	Researc...	9	3	Life Scie...		2022	4	Male
-	31	No	Travel_R...	106	Human ...	2	3	Human ...		424	1	Male
-	40	No	Travel_R...	107	Sales	10	3	Technica...		1239	2	Female
-	55	No		111	Sales	1	2	Life Scie...		106	1	Male
-	30	No	Non-Trav...	111	Researc...	9	3	Medical		239	3	Male
-	44	No	Non-Trav...	111	Researc...	17	3	Life Scie...		1206	4	Male
-	32	No	Travel_F...	116	Researc...	13	3	Other		1234	3	Female
-	39	No		116	Researc...	24	1	Life Scie...		2014	1	Male
-	39	No		117	Researc...	10	1	Medical		429	3	Male
-	32	No	Travel_R...	117	Sales	13	4	Life Scie...		859	2	Male
-	33	No	Travel_R...	117	Researc...	9	3	Medical		1238	1	Male
-	28	Yes	Travel_R...	103	Researc...	24	3	Life Scie...		19	3	Male
-	48	No	Travel_F...	117	Researc...	22	3	Medical		1900	4	Female
-	40	No	Travel_R...	118	Sales	14	2	Life Scie...		1598	4	Female
-	38	No	Travel_R...	119	Sales	3	3	Life Scie...		307	1	Male
-	39	No	Travel_R...	119	Sales	15	4	Marketing		1975	2	Male
-	28	No	Non-Trav...	120	Sales	4	3	Medical		129	2	Male

< > ← Previous → Next Finish Cancel

 Data import wizard - Step 4 of 4

This wizard guides you to import your data.

Step 4: RapidMiner Studio uses strongly typed attributes. In this step, you can define the data types of your attributes. Furthermore, RapidMiner Studio assigns roles to the attributes, defining what they can be used for by the individual operators. These roles can be also defined here. Finally, you can rename attributes or deselect them entirely.

Preview uses only first 100 rows

**3. Binomial**

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate
integer	polynomial	polynomial	integer	polynomial	integer	integer	polynomial	integer	integer	polynomial	integer	integer
attribute	label	attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute	attribute
41	Yes	Travel_Freq										
53	No	Travel_Rate	102	Research_Loc	23	4						
37	Yes	Travel_Rate	1373	Research_Loc	2	2	Other	?	4	4	Male	92
39	No	Non-Travel_Rate	105	Research_Loc	9	3	Life_Sciences	?	2022	4	Male	87
31	No	Travel_Rate	106	Human_Relations	2	3	Human_Relations	?	424	1	Male	62
40	No	Travel_Rate	107	Sales	10	3	Technical_Department	?	1239	2	Female	84
55	No	?	111	Sales	1	2	Life_Sciences	?	106	1	Male	70
30	No	Non-Travel_Rate	111	Research_Loc	9	3	Medical	?	239	3	Male	66
44	No	Non-Travel_Rate	111	Research_Loc	17	3	Life_Sciences	?	1206	4	Male	74
32	No	Travel_Freq	116	Research_Loc	13	3	Other	?	1234	3	Female	77
39	No	?	116	Research_Loc	24	1	Life_Sciences	?	2014	1	Male	52

0 errors.

Ignore errors  Show only errors

Row, Column Error Original value Message

← Previous → Next  **Finish**  Cancel

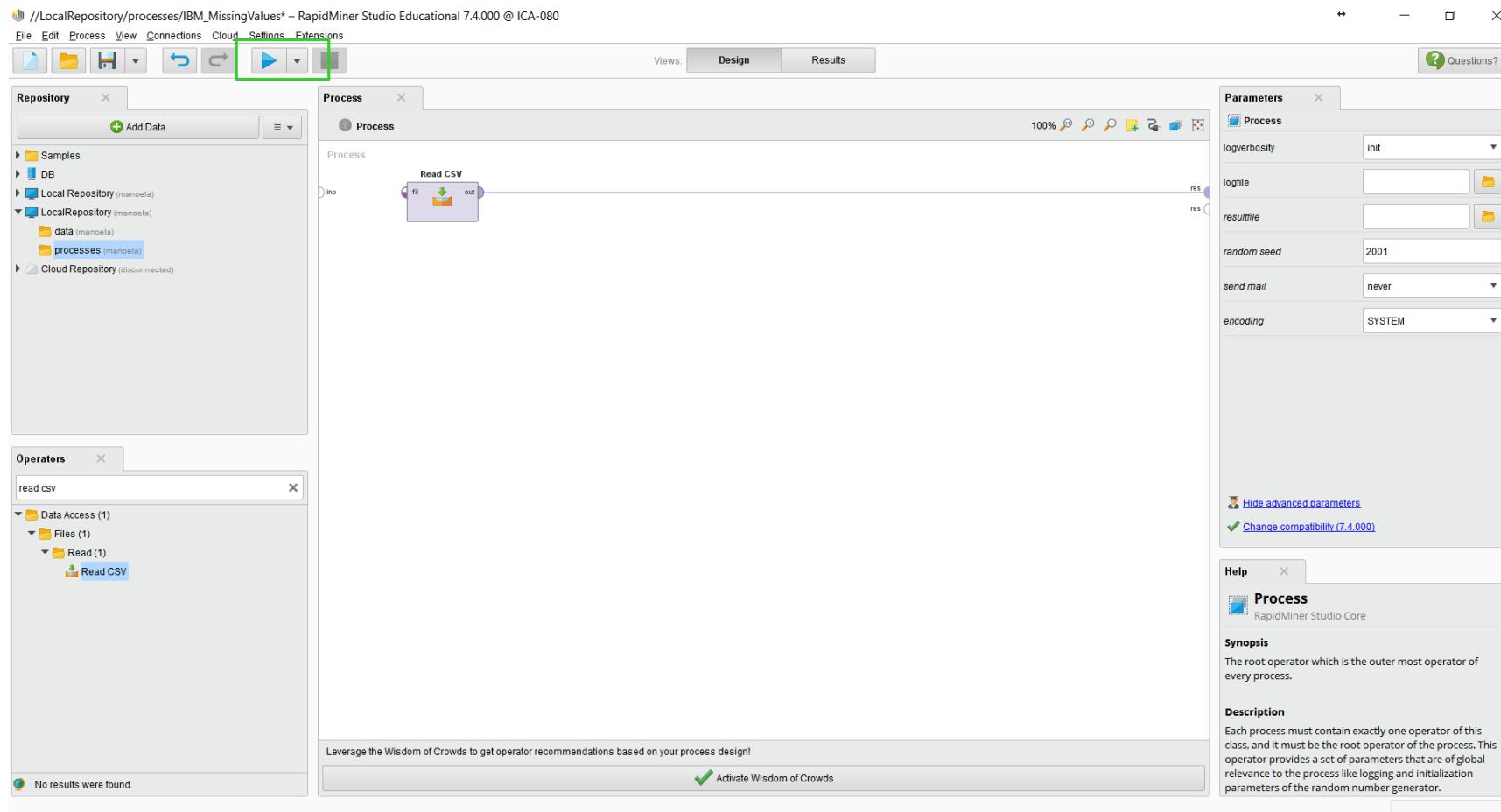
**1. Colunas que serão importadas.**

**2. Tipos dos atributos**

The screenshot shows the RapidMiner Studio interface. In the center, the 'Process' tab displays a flow with a single operator, 'Read CSV', which has an 'inp' port on the left and an 'out' port on the right. To the left of the process is the 'Repository' panel, which lists 'Samples', 'DB', and 'Local Repository (mancela)'. Below it is the 'Operators' panel, which shows a search bar for 'read csv' and a list under 'Data Access (1)' containing 'Files (1)' and 'Read (1)', with 'Read CSV' selected. A message at the bottom of the operators panel says 'No results were found.' On the right side of the screen is the 'Parameters' panel for the 'Read CSV' operator. It includes fields for 'csv file' (set to 'Attrition\_Missing.csv'), 'column separators' (set to ','), 'use quotes' (checked), 'quotes character' (set to '\"'), 'escape character' (set to '\\'), 'parse numbers' (checked), and 'decimal character' (set to '.'). Below these are 'Help' and 'Synopsis' sections. The 'Synopsis' section states: 'This operator is used to read CSV files.' and 'Jump to Tutorial Process'. At the bottom of the parameters panel, there is a note: 'CSV is an abbreviation for Comma-Separated Values. The CSV files store data (both numerical and text) in plain-text format. CSV files can be used to store data from various sources such as databases, spreadsheets, and text files.' A large callout box in the center-right area contains the text: 'Obs: base é carregada em memória, ou seja, o arquivo tem que estar disponível no caminho especificado para o processo funcionar!'

**Obs: base é carregada em memória, ou seja, o arquivo tem que estar disponível no caminho especificado para o processo funcionar!**

# 2. VISUALIZAR BASE



IBM\_MissingValues\* - RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Read CSV)

ExampleSet (1470 examples, 1 special attribute, 34 regular attributes)

Data Statistics Charts Advanced Charts Annotations

Row No.	Attrition	Age	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender	HourlyRate
1	Yes	41	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	2	Female	94
2	No	53	Travel_Rarely	102	Research & Development	23	4	Life Sciences	?	901	4	Female	72
3	Yes	37	Travel_Rarely	1373	Research & Development	2	2	Other	?	4	4	Male	92
4	No	39	Non-Travel	105	Research & Development	9	3	Life Sciences	?	2022	4	Male	87
5	No	31	Travel_Rarely	106	Human Resources	2	3	Human Resources	?	424	1	Male	62
6	No	40	Travel_Rarely	107	Sales	10	3	Technical Debt	?	1239	2	Female	84
7	No	55	?	111	Sales	1	2	Life Sciences	?	106	1	Male	70
8	No	30	Non-Travel	111	Research & Development	9	3	Medical	?	239	3	Male	66
9	No	44	Non-Travel	111	Research & Development	17	3	Life Sciences	?	1206	4	Male	74
10	No	32	Travel_Frequent	116	Research & Development	13	3	Other	?	1234	3	Female	77
11	No	39	?	116	Research & Development	24	1	Life Sciences	?	2014	1	Male	52
12	No	39	?	117	Research & Development	10	1	Medical	?	429	3	Male	99
13	No	32	Travel_Rarely	117	Sales	13	4	Life Sciences	?	859	2	Male	73
14	No	33	Travel_Rarely	117	Research & Development	9	3	Medical	?	1238	1	Male	60
15	Yes	28	Travel_Rarely	103	Research & Development	24	3	Life Sciences	?	19	3	Male	50
16	No	48	Travel_Frequent	117	Research & Development	22	3	Medical	?	1900	4	Female	58
17	No	40	Travel_Rarely	118	Sales	14	2	Life Sciences	?	1598	4	Female	84
18	No	38	Travel_Rarely	119	Sales	3	3	Life Sciences	?	307	1	Male	76
19	No	39	Travel_Rarely	119	Sales	15	4	Marketing	?	1975	2	Male	77
20	No	28	Non-Travel	120	Sales	4	3	Medical	?	129	2	Male	43
21	No	32	Travel_Rarely	120	Research & Development	6	5	Life Sciences	?	231	3	Male	43
22	Yes	36	Travel_Rarely	1218	Sales	9	4	Life Sciences	?	27	3	Male	82
23	No	34	Travel_Rarely	121	Research & Development	2	4	Medical	?	804	3	Female	86
24	No	53	Travel_Frequent	124	Sales	2	3	Marketing	?	1050	3	Female	38
25	Yes	34	Travel_Rarely	699	Research & Development	6	1	Medical	?	31	2	Male	83
26	No	37	Travel_Rarely	124	Research & Development	3	3	Other	?	1062	4	Female	35

Filter (1,470 / 1,470 examples): all

Repository

- Add Data
- Samples
- DB
- Local Repository (mancala)
- LocalRepository (mancala)
  - data (mancala)
  - processes (mancala)
- Cloud Repository (disconnected)

# 3. ORDENAR POR QUANTIDADE DE *MISSING VALUES*

The screenshot shows the RapidMiner Studio interface with the title bar //LocalRepository/processes/IBM\_MissingValues\* – RapidMiner Studio Educational 7.4.000 @ ICA-080. The main window displays the 'ExampleSet (Read CSV)' view, which lists various attributes with their types and missing value counts. A green box highlights the 'Missing' column header. The 'Statistics' section for each attribute provides numerical summaries like Min, Max, and Average. The 'Repository' panel on the right shows the Local Repository structure, including 'Samples', 'DB', 'Local Repository (manoela)', 'LocalRepository (manoela)', and 'Cloud Repository (disconnected)'. The bottom status bar indicates Examples: 1,470, Special Attributes: 1, and Regular Attributes: 34.

Name	Type	Missing	Statistics
EmployeeCount	Integer	1012	Min: 1, Max: 1, Average: 1
BusinessTravel	Polynomial	17	Least: Non-Travel (150), Most: Travel_Rarely (1026), Values: Travel_Rarely (1026), Travel_Frequently (277), ... [1 more]
HourlyRate	Integer	17	Min: 30, Max: 100, Average: 65.902
MonthlyIncome	Integer	15	Min: 1009, Max: 19999, Average: 6503.056
JobLevel	Integer	1	Min: 1, Max: 5, Average: 2.065
Attrition	Polynomial	0	Least: Yes (237), Most: No (1233), Values: No (1233), Yes (237)
Age	Integer	0	Min: 18, Max: 60, Average: 36.924
DailyRate	Integer	0	Min: 102, Max: 1499, Average: 802.486
Department	Polynomial	0	Least: Human Resources (63), Most: Research & Development (961), Values: Research & Development (961), Sales (446), ... [1 more]
DistanceFromHome	Integer	0	Min: 1, Max: 29, Average: 9.193
Education	Integer	0	Min: 1, Max: 5, Average: 2.913
EducationField	Polynomial	0	Least: Human Resources (27), Most: Life Sciences (606), Values: Life Sciences (606), Medical (464), ... [4 more]

# 4. ANALISAR

//LocalRepository/processes/IBM\_MissingValues\* – RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Read CSV)

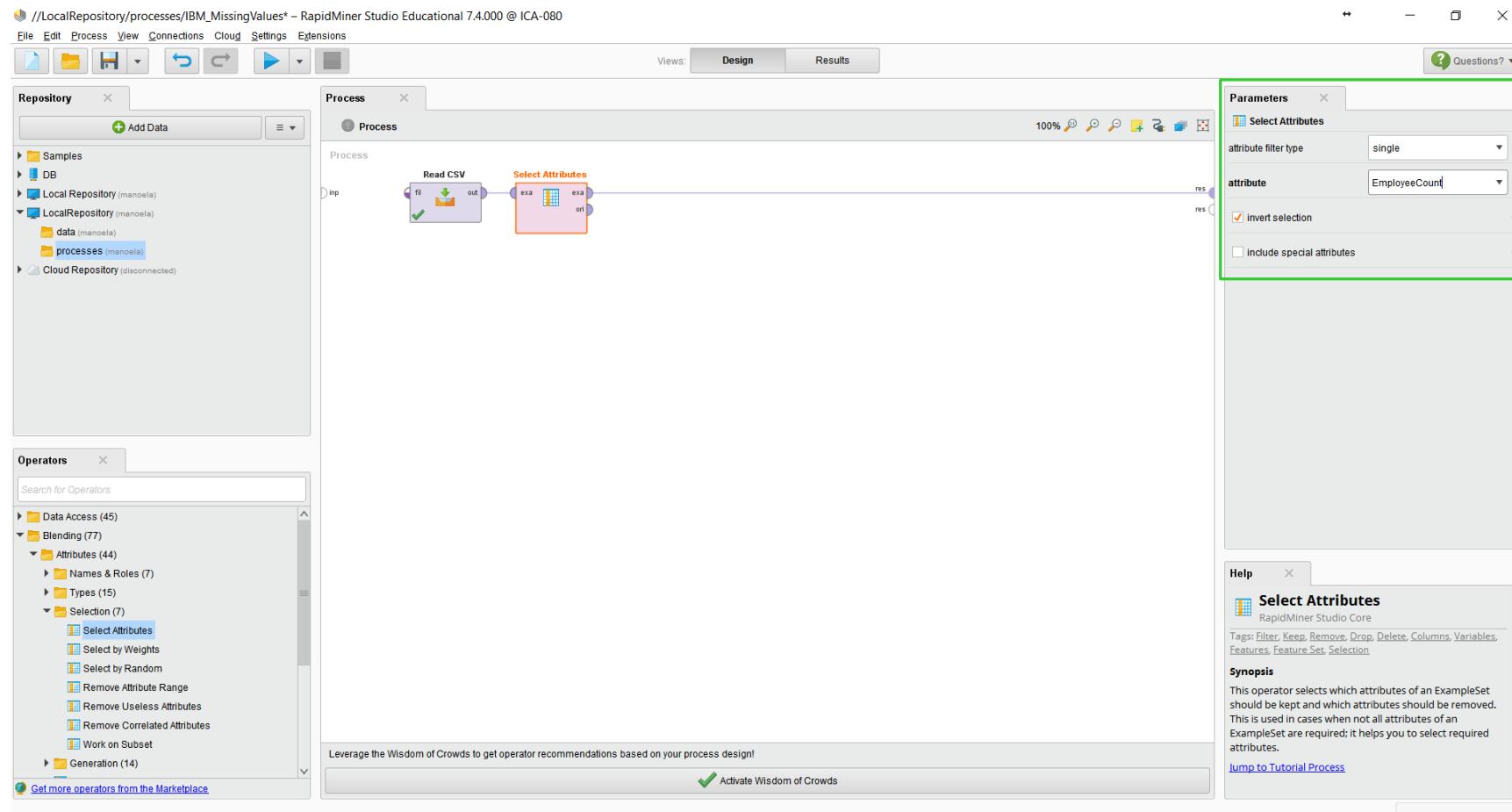
Name	Type	Missing	Statistics
EmployeeCount	Integer	1012	Min 1 Max 1 Average 1
BusinessTravel	Polynomial	17	Least Non-Travel (150) Most Travel_Rarely (1026) Values Travel_Rarely (1026), Travel_Frequently (277), ... [more]
HourlyRate	Integer	17	Min 30 Max 100 Average 65.902
MonthlyIncome	Integer	15	Min 1009 Max 19999 Average 6503.056
JobLevel	Integer	1	Min 1 Max 5 Average 2.065
Attrition	Polynomial	0	Least Yes (237) Most No (1233) Values No (1233), Yes (237)
Age	Integer	0	Min 18 Max 60 Average 36.924
DailyRate	Integer	0	Min 102 Max 1499 Average 802.486
Department	Polynomial	0	Least Human Resources (63) Most Research [...] ent (961) Values Research & Development (961), Sales (446), ... [more]
DistanceFromHome	Integer	0	Min 1 Max 29 Average 9.193
Education	Integer	0	Min 1 Max 5 Average 2.913
EducationField	Polynomial	0	Least Human Resources (27) Most Life Sciences (606) Values Life Sciences (606), Medical (464), ... [more]

Showing attributes 1 - 35 Examples: 1,470 Special Attributes: 1 Regular Attributes: 34

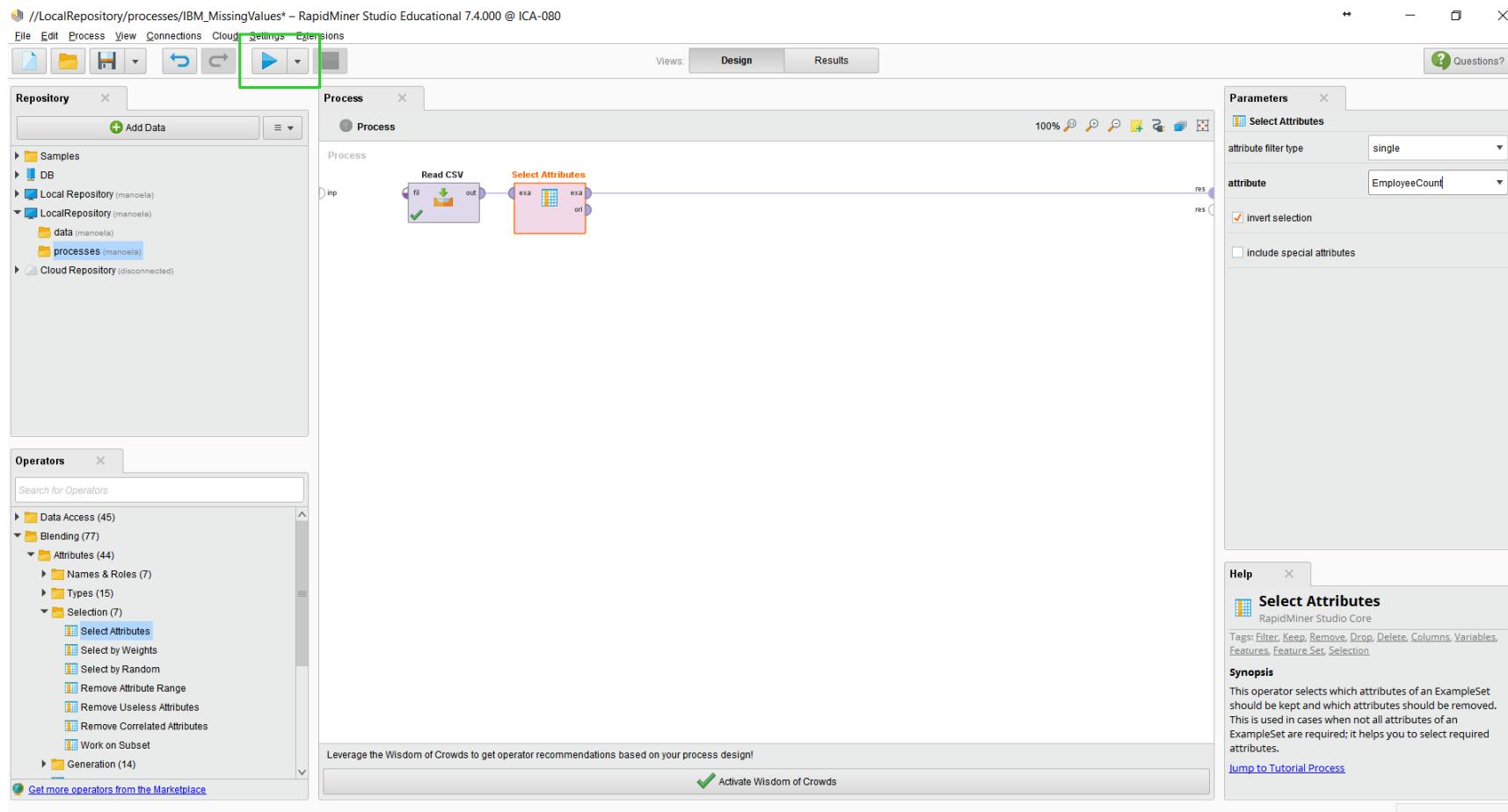
Repository

- Add Data
- Samples
- DB
- Local Repository (mancela)
- LocalRepository (mancela)
  - data (mancela)
  - processes (mancela)
- Cloud Repository (disconnected)

# 5. REMOVER ATRIBUTOS QUE POSSAM ATRAPALHAR



# 6. ANALISAR A BASE



File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Select Attributes)

Name	Type	Missing	Statistics
<b>BusinessTravel</b>	Polynomial	17	Least Non-Travel (150) Most Travel_Rarely (1026) Values Travel_Rarely (1026), Travel_Frequently (277), ...[1 more]
<b>HourlyRate</b>	Integer	17	Min 30 Max 100 Average 65.902
<b>MonthlyIncome</b>	Integer	15	Min 1009 Max 19999 Average 6503.056
<b>JobLevel</b>	Integer	1	Min 1 Max 5 Average 2.065
<b>Attrition</b>	Polynomial	0	Least Yes (237) Most No (1233) Values No (1233), Yes (237)
<b>Age</b>	Integer	0	Min 18 Max 60 Average 36.924
<b>DailyRate</b>	Integer	0	Min 102 Max 1499 Average 802.486
<b>Department</b>	Polynomial	0	Least Human Resources (63) Most Research & Development (961) Values Research & Development (961), Sales (446), ...[1 more]
<b>DistanceFromHome</b>	Integer	0	Min 1 Max 29 Average 9.193
<b>Education</b>	Integer	0	Min 1 Max 5 Average 2.913
<b>EducationField</b>	Polynomial	0	Least Human Resources (27) Most Life Sciences (606) Values Life Sciences (606), Medical (464), ...[4 more]
<b>EmployeeNumber</b>	Integer	0	Min 1 Max 2068 Average 1024.865

Showing attributes 1 - 34 Examples: 1,470 Special Attributes: 1 Regular Attributes: 33

Repository

- Samples
- DB
- Local Repository (manoela)
  - LocalRepository (manoela)
    - data (manoela)
    - processes (manoela)
- Cloud Repository (disconnected)

# 7. TRATAR *MISSING VALUES*

The screenshot shows the RapidMiner Studio interface with the following components:

- Repository View:** Shows the Local Repository (manoela) containing various processes like 01\_DecisionTree, IBM\_AttributeSelection, IBM\_Limpeza, etc.
- Process View:** Displays a process flow:
  - Read CSV operator (Input: data/manoela)
  - Select Attributes operator
  - Replace Missing Values operator (highlighted with a green box)
- Operators View:** Shows the Operators palette with the Replace Missing Values operator selected.
- Parameters View (highlighted with a green box):** Configuration for the Replace Missing Values operator:
  - attribute filter type: subset
  - attributes: BusinessTravel, HourlyRate, MonthlyIncome (selected from the Select Attributes dialog)
  - default: average
- Select Attributes Dialog (highlighted with a green box):** Shows the list of attributes and the selected attributes:
  - Attributes: Age, Attrition, BusinessTravel, DailyRate, Department, DistanceFromHome, Education, EducationField, EmployeeNumber, EnvironmentSatisfaction, Gender, JobInvolvement, JobLevel.
  - Selected Attributes: BusinessTravel, HourlyRate, MonthlyIncome
- Help View:** Provides details about the Replace Missing Values operator, including its synopsis and description.

# 8. ANALISAR A BASE

The screenshot shows the RapidMiner Studio interface with a process flow for data cleaning. The process starts with a 'Read CSV' operator, followed by a 'Select Attributes' operator, and finally a 'Replace Missing Values' operator. The 'Replace Missing Values' operator is highlighted with a red border. The 'Parameters' panel on the right is configured for this operator, with 'subset' selected for 'attribute filter type' and 'average' for 'default'. The 'Operators' palette on the left shows the 'Replace Missing Values' operator selected. The 'Help' panel provides a synopsis and description of the 'Replace Missing Values' operator.

File Edit Process View Connections Cloud Settings Extensions

Repository X Add Data

Process X Process

Process

Read CSV Select Attributes Replace Missing Values

Parameters X

Replace Missing Values

create view

attribute filter type: subset

attributes: Select Attributes...

invert selection

include special attributes

default: average

columns: Edit List (0)...

Hide advanced parameters

Change compatibility (7.4.000)

Operators X

Search for Operators

Data Access (45)

Blending (77)

Cleansing (26)

Normalization (3)

Binning (5)

Missing (6)

- Replace Missing Values
- Impute Missing Values
- Declare Missing Value
- Replace Infinite Values
- Remove Unused Values
- Fill Data Gaps

Duplicates (1)

Outliers (4)

Get more operators from the Marketplace

Leverage the Wisdom of Crowds to get operator recommendations based on your process design!

Activate Wisdom of Crowds

Help X

Replace Missing Values

RapidMiner Studio Core

Tags: Nulls, Empty, Cleansing, Quality, Missing, Handle, Impute, Na, Nan, Fill na, Missing

Synopsis

This operator replaces missing values in examples of selected attributes by a specified replacement.

Jump to Tutorial Process

Description

This operator replaces missing values in examples of selected attributes by a specified replacement. Missing values can be replaced by the minimum, maximum or average value of that attribute. Zero can also be placed in place of missing values. Any replenishment value can also be specified as a replacement of missing values.

//LocalRepository/processes/Copy of IBM\_MissingValues\* - RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Replace Missing Values)

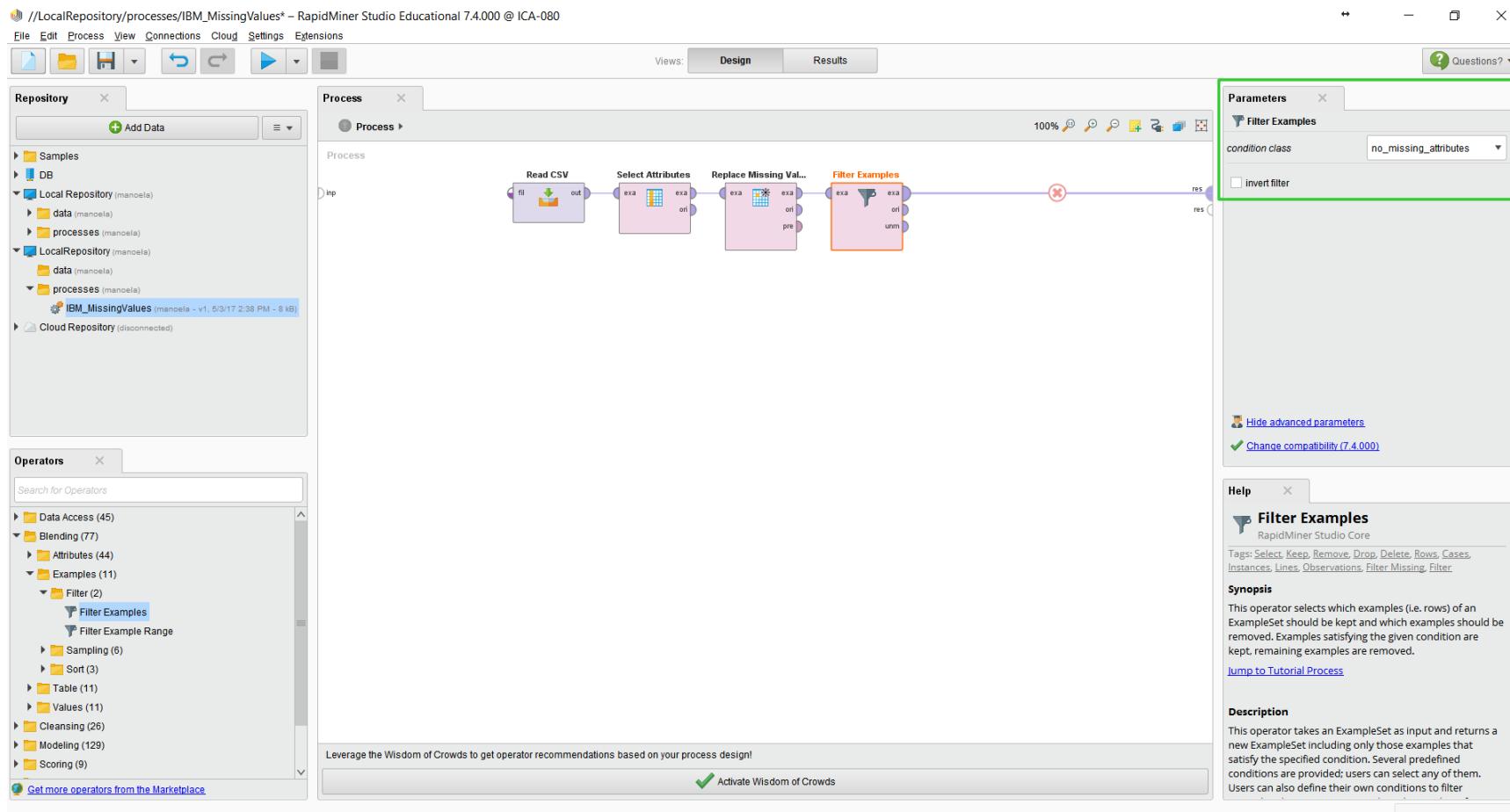
Name	Type	Missing	Statistics			Filter (34 / 34 attributes): Search for Attributes
JobLevel	Integer	1	Min 1	Max 5	Average 2.065	
Attrition	Polynomial	0	Least Yes (237)	Most No (1233)	Values No (1233), Yes (237)	
BusinessTravel	Polynomial	0	Least Non-Travel (150)	Most Travel_Rarely (1043)	Values Travel_Rarely (1043), Travel_Frequently (277), ...[1 more]	
HourlyRate	Integer	0	Min 30	Max 100	Average 65.903	
MonthlyIncome	Integer	0	Min 1009	Max 19999	Average 6503.055	
Age	Integer	0	Min 18	Max 60	Average 36.924	
DailyRate	Integer	0	Min 102	Max 1499	Average 802.486	
Department	Polynomial	0	Least Human Resources (63)	Most Research [...] ent (961)	Values Research & Development (961), Sales (446), ...[1 more]	
DistanceFromHome	Integer	0	Min 1	Max 29	Average 9.193	
Education	Integer	0	Min 1	Max 5	Average 2.913	
EducationField	Polynomial	0	Least Human Resources (27)	Most Life Sciences (606)	Values Life Sciences (606), Medical (464), ...[4 more]	
EmployeeNumber	Integer	0	Min 1	Max 2068	Average 1024.865	
			Min	Max	Average	
			~ 700	~ 700	~ 700	

Showing attributes 1 - 34 Examples: 1,470 Special Attributes: 1 Regular Attributes: 33

Repository

- Samples
- DB
- Local Repository (manoela)
  - data (manoela)
  - processes (manoela)
- LocalRepository (manoela)
  - data (manoela)
  - processes (manoela)
- Cloud Repository (disconnected)

# 9. FILTRAR EXEMPLOS



# 10. ANALISAR A BASE

The screenshot shows the RapidMiner Studio interface with the following details:

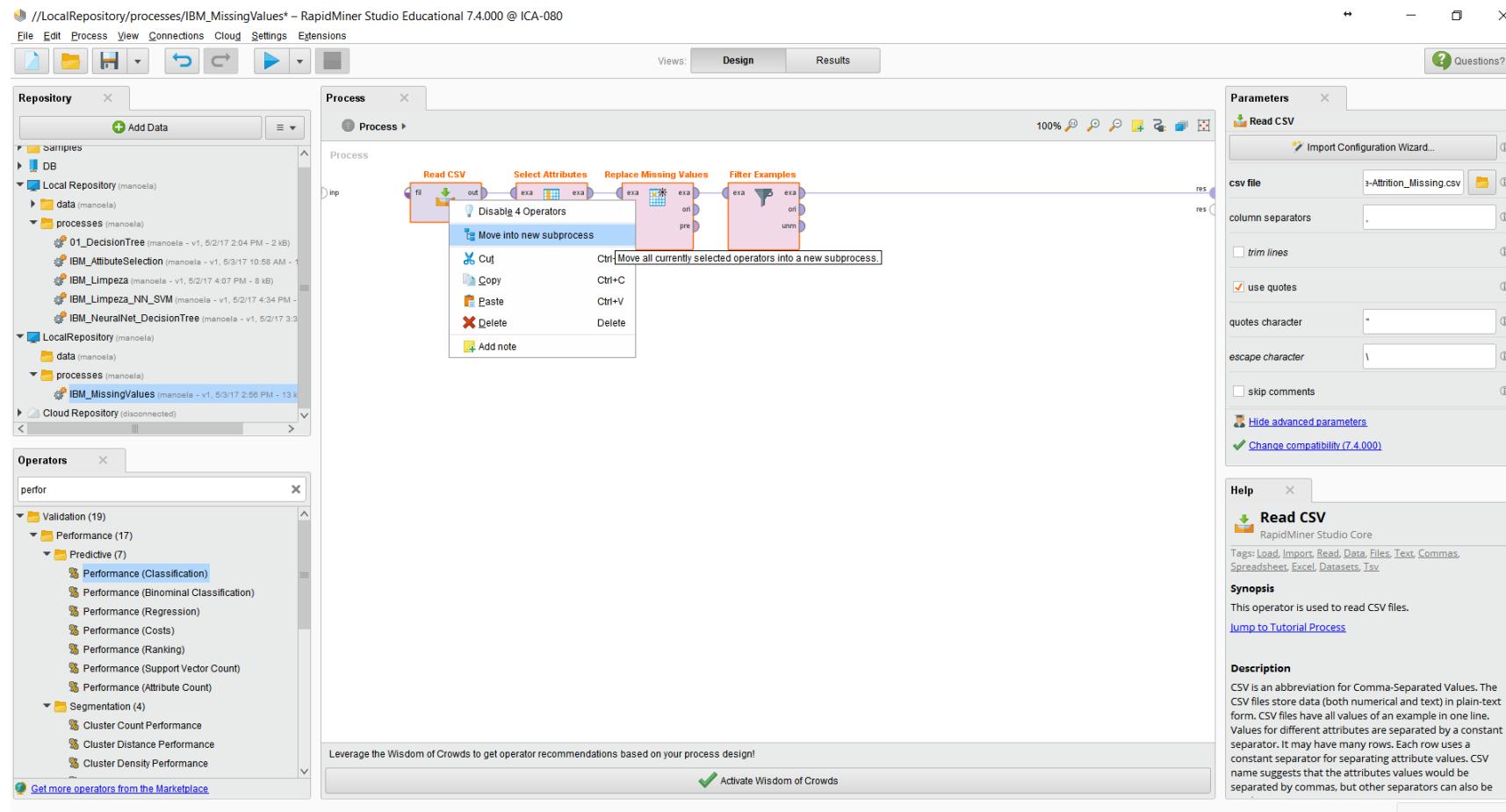
**Result History View:** Displays a list of attributes from the 'IBM\_MissingValues' dataset. The columns include Name, Type, Missing, Statistics, and Values.

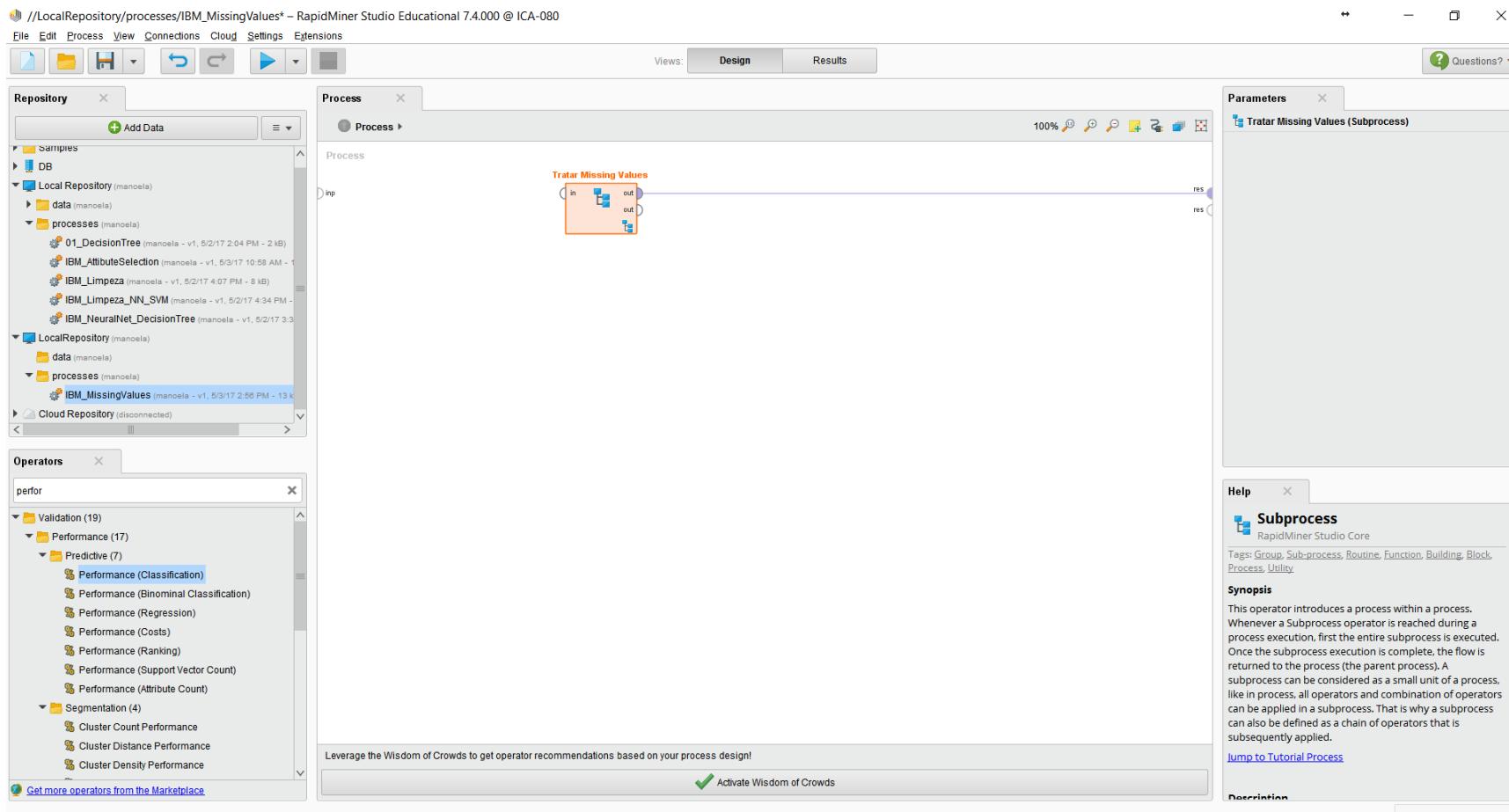
Name	Type	Missing	Statistics	Values
Attrition	Polynomial	0	Least Yes (237) Most No (1232)	No (1232), Yes (237)
BusinessTravel	Polynomial	0	Least Non-Travel (150) Most Travel_Rarely (1043)	Travel_Rarely (1043), Travel_Frequently (276), ...[1 more]
HourlyRate	Integer	0	Min 30 Max 100	Average 65.880
MonthlyIncome	Integer	0	Min 1009 Max 19999	Average 6505.744
Age	Integer	0	Min 18 Max 60	Average 36.926
DailyRate	Integer	0	Min 102 Max 1499	Average 802.554
Department	Polynomial	0	Least Human Resources (63) Most Research & Development (960), Sales (446), ...[1 more]	Research & Development (960), Sales (446), ...[1 more]
DistanceFromHome	Integer	0	Min 1 Max 29	Average 9.188
Education	Integer	0	Min 1 Max 5	Average 2.912
EducationField	Polynomial	0	Least Human Resources (27) Most Life Sciences (605)	Life Sciences (605), Medical (464), ...[4 more]
EmployeeNumber	Integer	0	Min 1 Max 2068	Average 1024.993
EnvironmentSatisfaction	Integer	0	Min 1 Max 4	Average 2.722

**Repository View:** Shows the local repository structure, including samples, DB, Local Repository, and Cloud Repository. The 'IBM\_MissingValues' process is selected in the Local Repository.

Showing attributes 1 - 34 Examples: 1,469 Special Attributes: 1 Regular Attributes: 33

# 11. CRIAR SUBPROCESO





# 12. CONVERTER ATRIBUTOS CATEGÓRICOS EM NUMÉRICOS

The screenshot shows the RapidMiner Studio interface with the following components:

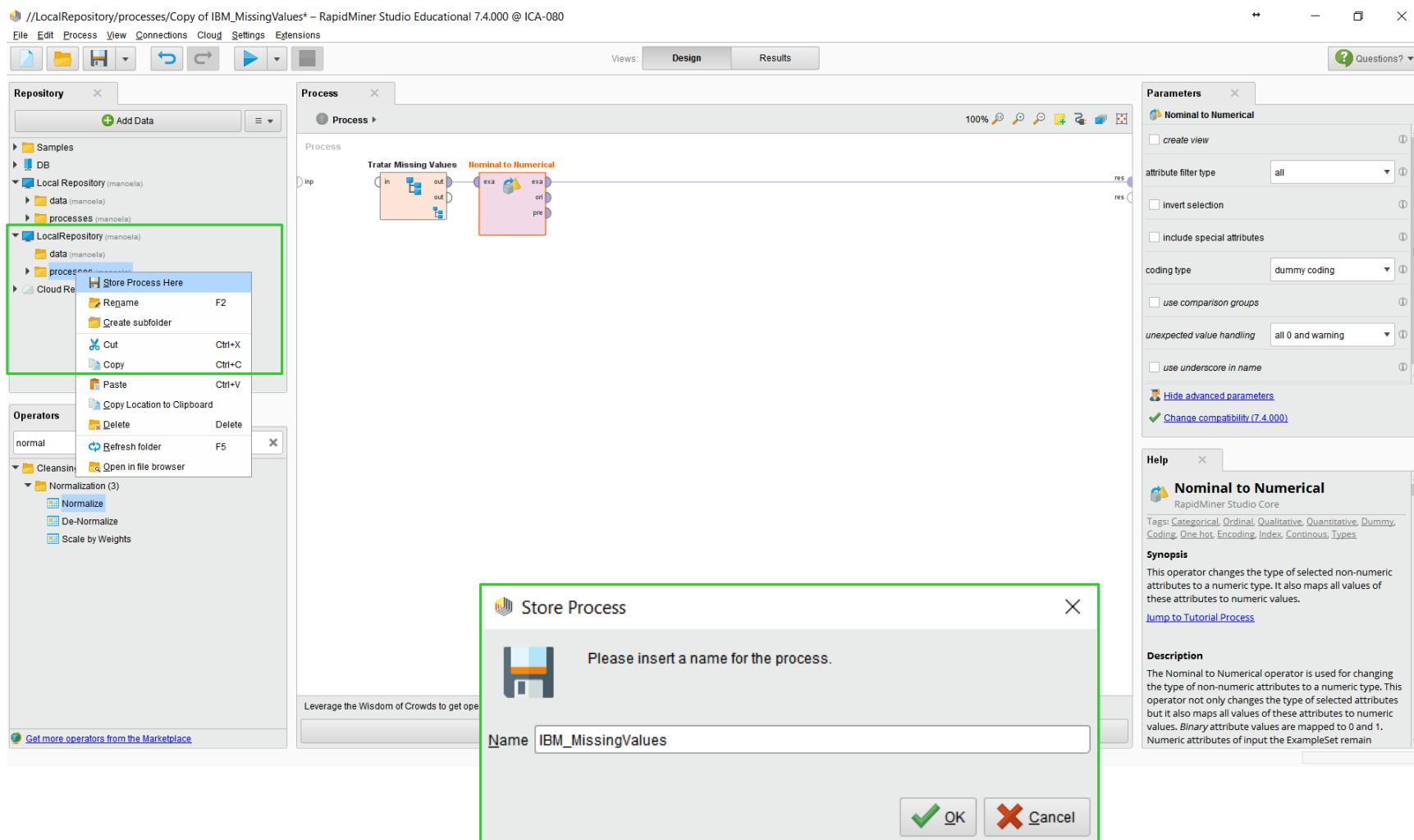
- Repository:** Shows the local repository with samples, DB, and processes.
- Process:** The main workspace where a process is being designed. It contains two operators: "Tratar Missing Values" and "Nominal to Numerical".
- Operators:** A palette on the left containing various operators categorized under "nominal". The "Nominal to Numerical" operator is highlighted.
- Parameters:** A panel on the right showing the configuration for the "Nominal to Numerical" operator. It includes fields for "attribute filter type" (set to "all"), "coding type" (set to "dummy coding"), and "unexpected value handling" (set to "all 0 and warning").
- Help:** A panel at the bottom right providing a synopsis and description of the "Nominal to Numerical" operator.

The process flow consists of the following steps:

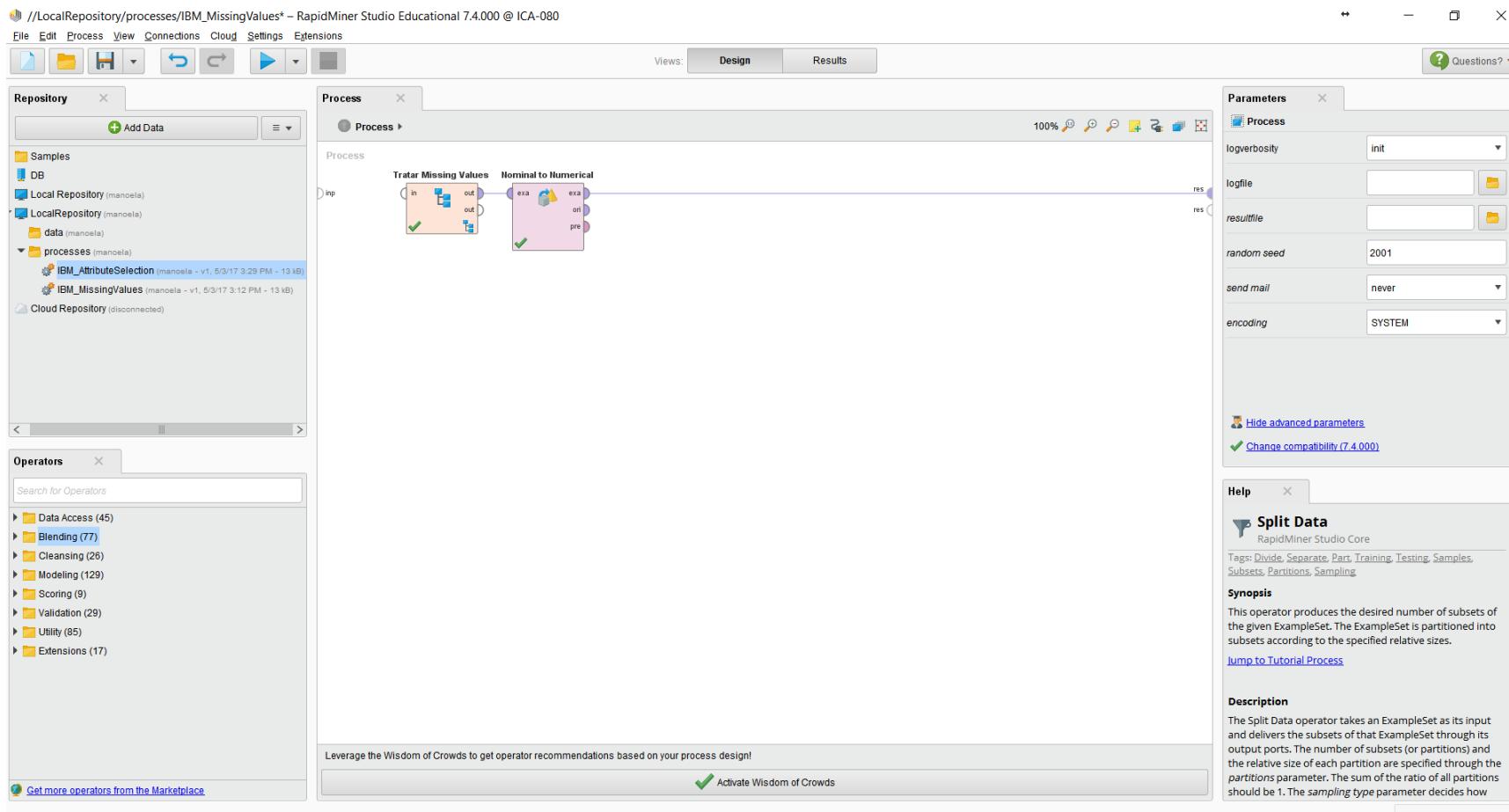
```
graph LR; TratarMissingValues[Tratar Missing Values] --> NominalToNumerical[Nominal to Numerical]; NominalToNumerical --> res[res]
```

A message at the bottom of the workspace says: "Leverage the Wisdom of Crowds to get operator recommendations based on your process design!"

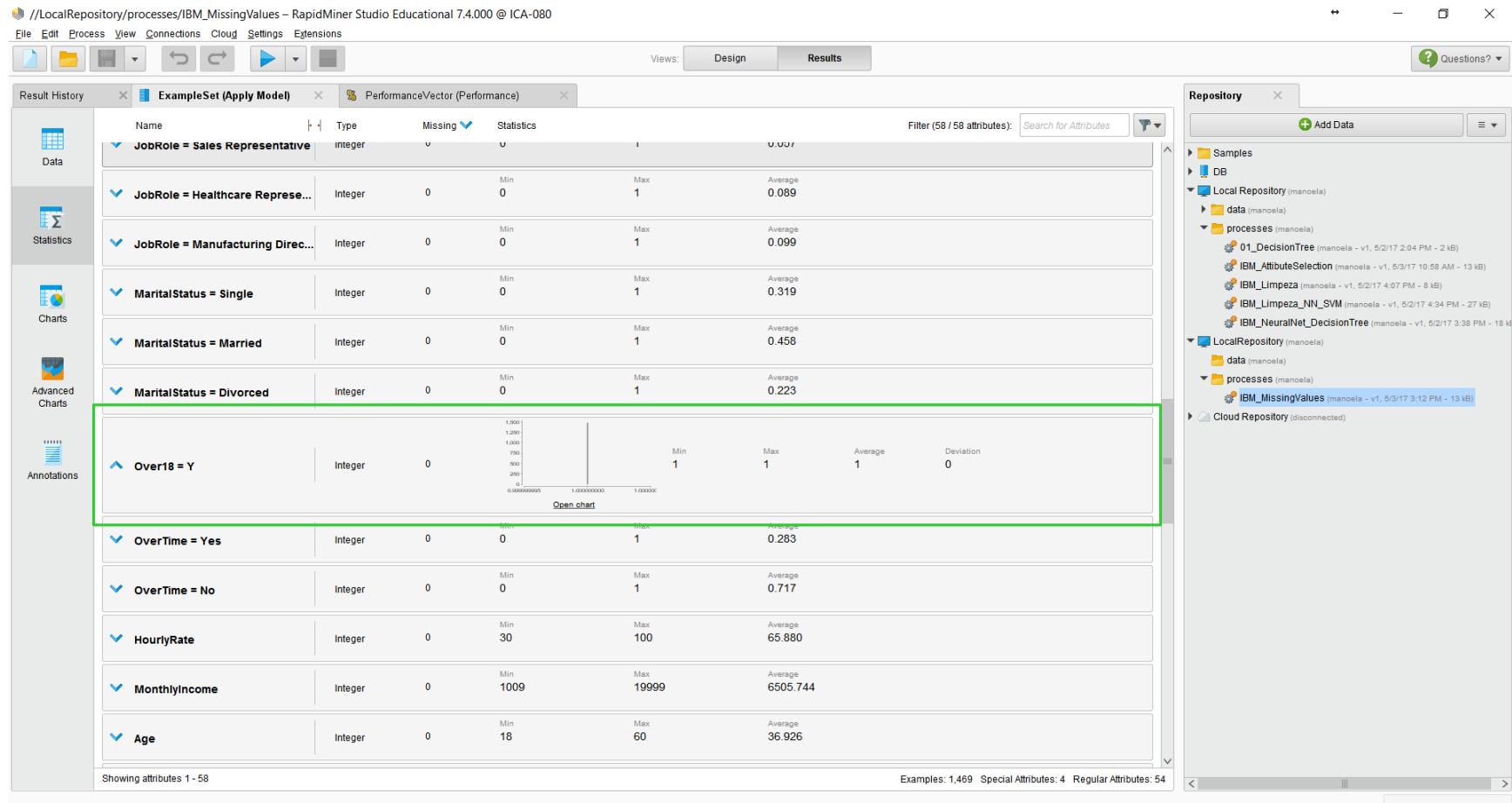
# 13. SALVAR PROCESSO



# **Seleção de atributos**



# 1. EXCLUSÃO DE ATRIBUTOS DESNECESSÁRIOS



The screenshot shows the RapidMiner Studio interface with the title bar //LocalRepository/processes/IBM\_MissingValues – RapidMiner Studio Educational 7.4.000 @ ICA-080. The main window displays the 'PerformanceVector (Performance)' view under the 'ExampleSet (Apply Model)' tab. On the left, there is a sidebar with tabs: Data, Statistics, Charts, Advanced Charts, and Annotations. The 'Annotations' tab is selected and has a green border around it. The main table lists attributes with their types, missing values, and statistics. One row, 'Over18 = Y', is highlighted with a green box. Below this row is a histogram chart for the 'Over18' attribute. The histogram shows a single peak at value 1, with the x-axis ranging from 0.0000000 to 1.0000000 and the y-axis ranging from 0 to 1,500. The chart has labels for Min, Max, Average, and Deviation. The repository panel on the right shows the local repository structure, including samples, DB, Local Repository, and processes. The 'IBM\_MissingValues' process is selected.

Name	Type	Missing	Statistics
JobRole = Sales Representative	integer	0	0 0.097
JobRole = Healthcare Represe...	Integer	0 0	Min 1 Max 1 Average 0.089
JobRole = Manufacturing Direct...	Integer	0 0	Min 1 Max 1 Average 0.099
MaritalStatus = Single	Integer	0 0	Min 1 Max 1 Average 0.319
MaritalStatus = Married	Integer	0 0	Min 1 Max 1 Average 0.458
MaritalStatus = Divorced	Integer	0 0	Min 1 Max 1 Average 0.223
Over18 = Y	Integer	0	Min 1 Max 1 Average 1 Deviation 0
OverTime = Yes	Integer	0 0	Average 0.283
OverTime = No	Integer	0 0	Min 1 Max 1 Average 0.717
HourlyRate	Integer	0 30	Max 100 Average 65.880
MonthlyIncome	Integer	0 1009	Max 19999 Average 6505.744
Age	Integer	0 18	Max 60 Average 36.926

Showing attributes 1 - 58 Examples: 1,469 Special Attributes: 4 Regular Attributes: 54

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Apply Model) PerformanceVector (Performance)

Name Type Missing Statistics

**MonthlyRate** Integer 0 Min 2094 Max 26999 Average 14317.193

**NumCompaniesWorked** Integer 0 Min 0 Max 9 Average 2.694

**PercentSalaryHike** Integer 0 Min 11 Max 25 Average 15.209

**PerformanceRating** Integer 0 Min 3 Max 4 Average 3.154

**RelationshipSatisfaction** Integer 0 Min 1 Max 4 Average 2.712

**StandardHours** Integer 0 Min 80 Max 80 Average 80 Deviation 0

**StockOptionLevel** Integer 0 Min 0 Max 3 Average 0.794

**TotalWorkingYears** Integer 0 Min 0 Max 40 Average 11.283

**TrainingTimesLastYear** Integer 0 Min 0 Max 6 Average 2.799

**WorkLifeBalance** Integer 0 Min 1 Max 4 Average 2.761

**YearsAtCompany** Integer 0 Min 0 Max 40 Average 7.010

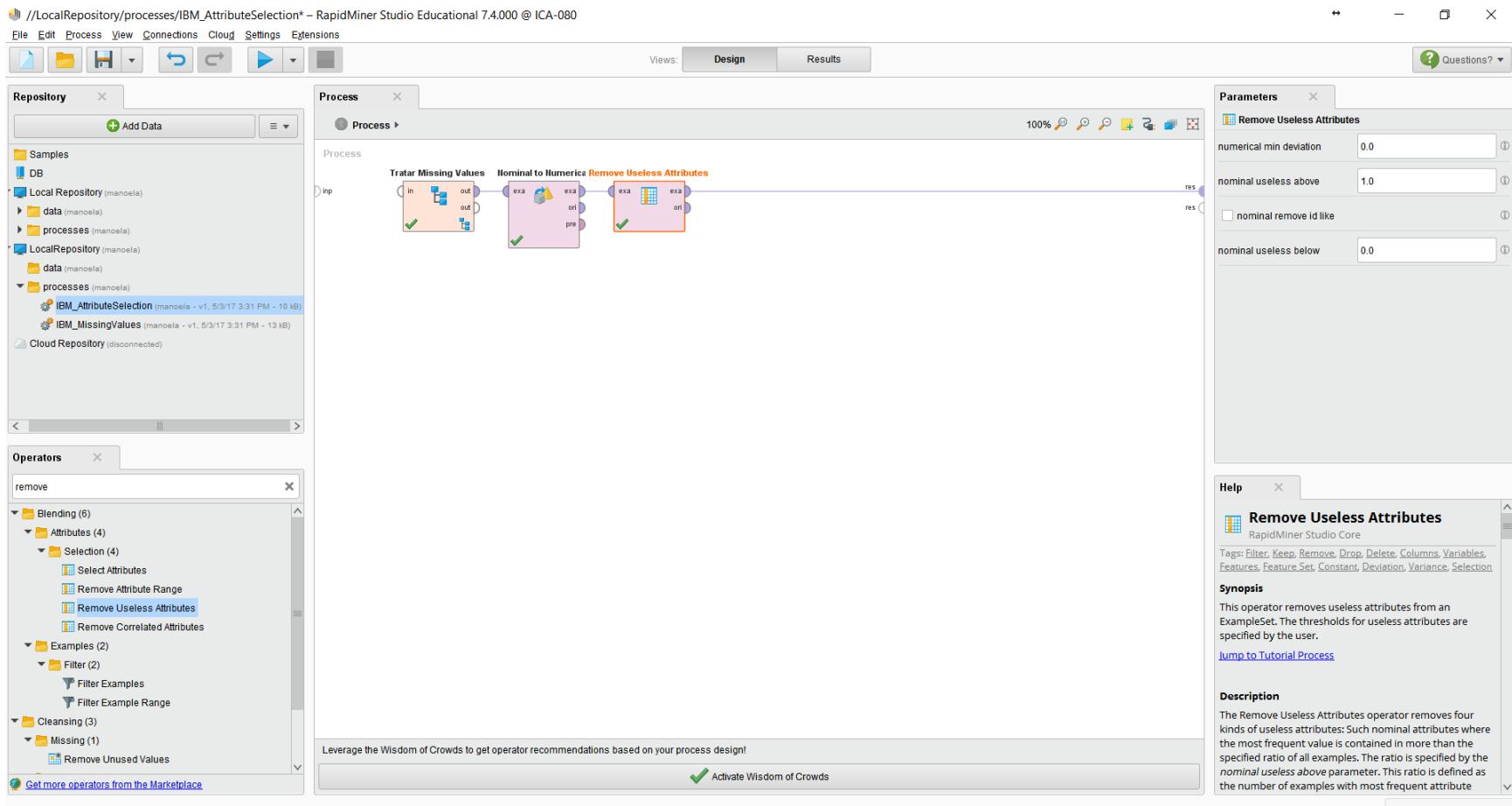
**YearsInCurrentRole** Integer 0 Min 0 Max 18 Average 4.231

Showing attributes 1 - 58 Examples: 1,469 Special Attributes: 4 Regular Attributes: 54

Repository

- Samples
- DB
- Local Repository (manoela)
  - 01\_DecisionTree (manoela - v1, 5/2/17 2:04 PM - 2 kB)
  - IBM\_AttributeSelection (manoela - v1, 5/3/17 10:58 AM - 13 kB)
  - IBM\_Limpeza (manoela - v1, 5/2/17 4:07 PM - 8 kB)
  - IBM\_Limpeza\_NN\_SVM (manoela - v1, 5/2/17 4:34 PM - 27 kB)
  - IBM\_NeuralNet\_DecisionTree (manoela - v1, 5/2/17 3:38 PM - 18 kB)
- LocalRepository (manoela)
  - data (manoela)
  - processes (manoela)
    - IBM\_MissingValues (manoela - v1, 5/3/17 3:12 PM - 13 kB)
- Cloud Repository (disconnected)





//LocalRepository/processes/IBM\_AttributeSelection\* – RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Remove Useless Attributes)

ExampleSet (1469 examples, 1 special attribute, 52 regular attributes)

Dois atributos foram excluídos

Repository

Samples DB Local Repository (manoela) LocalRepository (manoela) LocalRepository (manoela) Cloud Repository (disconnected)

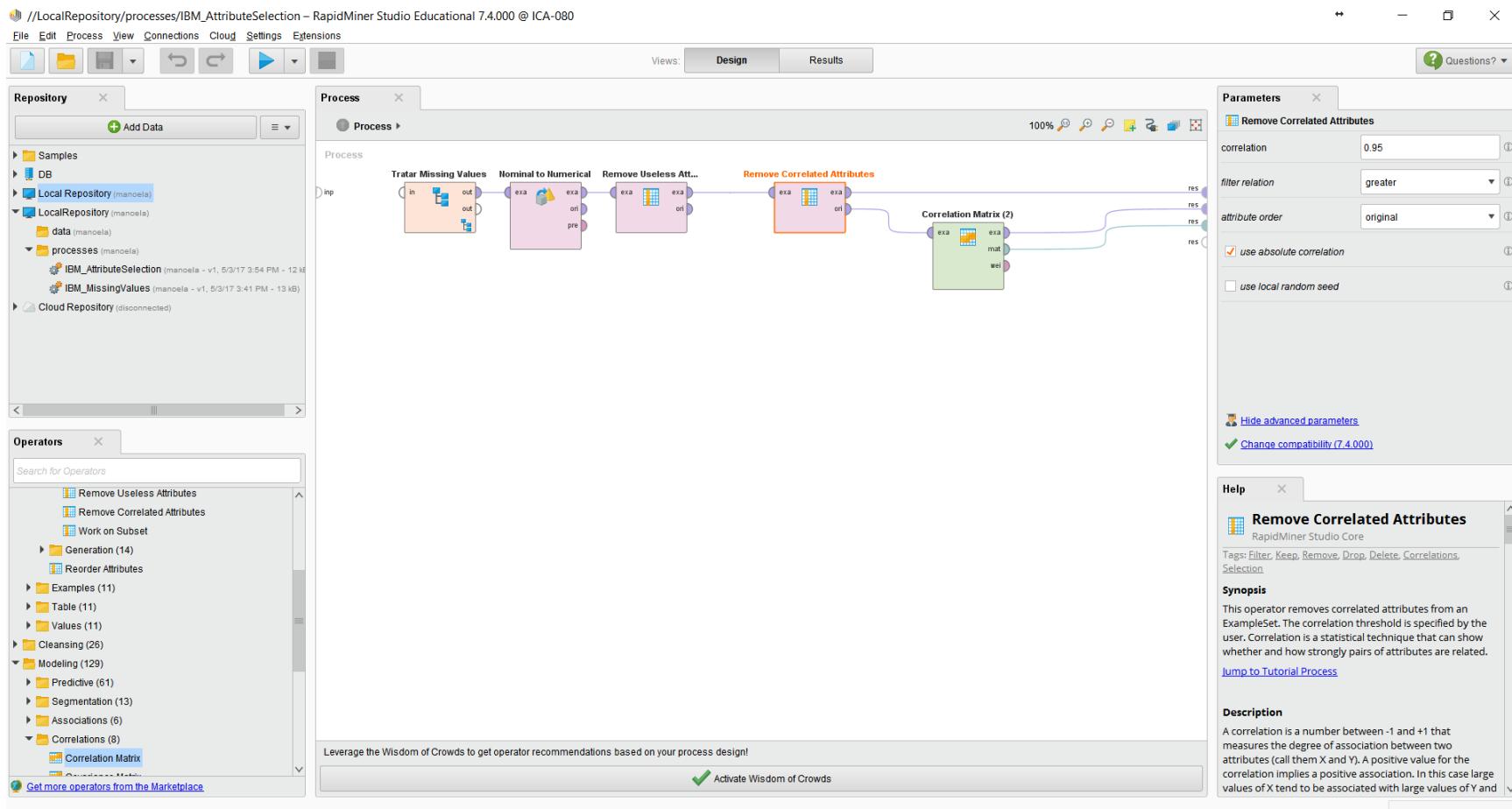
Advanced Charts Annotations

Data Statistics Charts

Row No. Attrition BusinessTr... BusinessTr... Busi...

Row No.	Attrition	BusinessTr...	BusinessTr...	Busi...	...	...	...	...	...	...	...
1	Yes	1	0	0	0	0	0	0	0	0	0
2	No	1	0	0	0	0	0	0	0	0	0
3	Yes	1	0	0	0	0	0	0	0	0	0
4	No	0	1	0	0	0	1	0	0	0	0
5	No	1	0	0	0	0	0	1	0	0	0
6	No	1	0	0	0	1	0	0	0	0	0
7	No	1	0	0	1	0	0	0	1	0	0
8	No	0	1	0	0	1	0	0	0	0	0
9	No	0	1	0	0	0	1	0	0	0	0
10	No	0	0	1	0	1	0	0	1	0	0
11	No	1	0	0	0	1	0	1	0	0	0
12	No	1	0	0	0	1	0	0	0	0	1
13	No	1	0	0	1	0	0	1	0	0	0
14	No	1	0	0	0	1	0	0	0	0	1
15	Yes	1	0	0	0	1	0	1	0	0	0
16	No	0	0	1	0	1	0	0	0	0	1
17	No	1	0	0	1	0	0	1	0	0	0
18	No	1	0	0	1	0	0	1	0	0	0
19	No	1	0	0	1	0	0	0	0	0	1
20	No	0	1	0	1	0	0	0	0	0	1
21	No	1	0	0	0	1	0	1	0	0	0
22	Yes	1	0	0	1	0	0	1	0	0	0
23	No	1	0	0	0	1	0	0	0	0	1
24	No	0	0	1	1	0	0	0	0	0	1
25	Yes	1	0	0	0	1	0	0	0	0	1
26	No	1	0	0	0	1	0	0	1	0	0

# 2. REMOVER ATRIBUTOS ALTAMENTE CORRELACIONADOS



//LocalRepository/processes/IBM\_AttributeSelection\* – RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History Correlation Matrix (Correlation Matrix) ExampleSet (Remove Useless Attributes) ExampleSet (Remove Correlated Attributes)

?

Questions? ▾

Data Pairwise Table Charts Annotations

Attributes	BusinessTravel = Travel... BusinessTravel = Non-Tra...	BusinessTravel = Travel_F...	BusinessTravel = Non-Travel...	Department = Sales Department = Research &...	Department = Human Res...	EducationField = Life Scie...	EducationField = Other EducationField = Human ...	EducationField = Technical...	EducationField = Medical EducationField = Marketing	Gender = Female Gender = Male	JobRole = Sales Executive JobRole = Research Direct...	JobRole = Laboratory Tec...	JobRole = Manager JobRole = Human Resour...	JobRole = Research Scie...	JobRole = Sales Represe...	JobRole = Healthcare Rep...	JobRole = Manufacturing ... MaritalStatus = Single
BusinessTravel = Travel...	1	-0.528	-0.753	-0.005	0.001	0.009	-0.029	0.018	-0.024	0.017	-0.005	0.034	0.016	-0.016	-0.012	0.034	
BusinessTravel = Non-Tra...	-0.528	1	-0.162	0.007	-0.005	-0.005	0.006	-0.013	0.021	0.004	0.013	-0.031	-0.050	0.050	0.031	-0.021	
BusinessTravel = Travel_F...	-0.753	-0.162	1	0.001	0.002	-0.007	0.029	-0.011	0.012	-0.023	-0.004	-0.016	0.020	-0.020	-0.009	-0.023	
Department = Sales	-0.005	0.007	0.001	1	-0.907	-0.140	-0.101	-0.064	-0.090	-0.031	-0.168	0.528	0.033	-0.033	0.809	-0.158	
Department = Research &...	0.001	-0.005	0.002	-0.907	1	-0.291	0.127	0.065	-0.188	0.039	0.184	-0.478	-0.016	0.016	-0.733	0.175	
Department = Human Res...	0.009	-0.005	-0.007	-0.140	-0.291	1	-0.068	-0.008	0.646	-0.020	-0.050	-0.074	-0.035	0.035	-0.113	-0.051	
EducationField = Life Scie...	-0.029	0.006	0.029	-0.101	0.127	-0.068	1	-0.203	-0.115	-0.263	-0.569	-0.292	-0.008	0.008	-0.091	0.019	
EducationField = Other	0.018	-0.013	-0.011	-0.064	0.065	-0.008	-0.203	1	-0.033	-0.076	-0.165	-0.085	-0.023	0.023	-0.037	-0.006	
EducationField = Human ...	-0.024	0.021	0.012	-0.090	-0.188	0.646	-0.115	-0.033	1	-0.043	-0.093	-0.048	-0.029	0.029	-0.073	-0.033	
EducationField = Technical...	0.017	0.004	-0.023	-0.031	0.039	-0.020	-0.263	-0.076	-0.043	1	-0.213	-0.109	-0.004	0.004	-0.059	-0.023	
EducationField = Medical	-0.005	0.013	-0.004	-0.168	0.184	-0.050	-0.569	-0.165	-0.093	-0.213	1	-0.237	0.014	-0.014	-0.134	0.063	
EducationField = Marketing	0.034	-0.031	-0.016	0.528	-0.478	-0.074	-0.292	-0.085	-0.048	-0.109	-0.237	1	0.024	-0.024	0.457	-0.084	
Gender = Female	0.016	-0.050	0.020	0.033	-0.016	-0.035	-0.008	-0.023	-0.029	-0.004	0.014	0.024	1	-1	0.006	0.006	
Gender = Male	-0.016	0.050	-0.020	-0.033	0.016	0.035	0.008	0.023	0.029	0.004	-0.014	-0.024	-1	1	-0.006	-0.006	
JobRole = Sales Executive	-0.012	0.031	-0.009	0.809	-0.733	-0.113	-0.091	-0.037	-0.073	-0.059	-0.134	0.457	0.006	-0.006	1	-0.128	
JobRole = Research Director...	0.034	-0.021	-0.023	-0.158	0.175	-0.051	0.019	-0.006	-0.033	-0.023	0.063	-0.084	0.006	-0.006	-0.128	1	
JobRole = Laboratory Technician	-0.015	0.009	0.011	-0.305	0.337	-0.098	0.045	0.059	-0.063	-0.027	0.066	-0.161	-0.067	0.067	-0.247	-0.111	
JobRole = Manager	0.027	0.014	-0.042	0.035	-0.071	0.088	-0.011	-0.008	0.082	-0.039	-0.001	0.026	0.034	-0.034	-0.146	-0.066	
JobRole = Human Resources	0.009	-0.016	0.002	-0.126	-0.263	0.905	-0.063	0.002	0.550	-0.009	-0.043	-0.067	-0.036	0.036	-0.102	-0.046	
JobRole = Research Scientist	0.013	-0.010	-0.007	-0.328	0.362	-0.105	0.042	0.006	-0.068	0.077	0.041	-0.173	-0.011	0.011	-0.265	-0.119	
JobRole = Sales Representative	-0.026	-0.034	0.056	0.371	-0.336	-0.052	-0.043	-0.034	-0.033	0.057	-0.052	0.133	0.029	-0.029	-0.131	-0.059	
JobRole = Healthcare Representative	-0.016	0.013	0.008	-0.207	0.228	-0.066	0.029	0.018	-0.043	0.019	0.034	-0.109	-0.007	0.007	-0.167	-0.075	
JobRole = Manufacturing Worker	0.000	-0.014	0.010	-0.219	0.241	-0.070	0.052	-0.011	-0.045	0.008	0.035	-0.115	0.066	-0.066	-0.177	-0.079	
MaritalStatus = Single	-0.019	-0.004	0.026	0.034	-0.011	-0.051	0.020	0.005	-0.072	0.015	-0.004	-0.013	0.032	-0.032	0.007	-0.042	

//LocalRepository/processes/IBM\_AttributeSelection.r - RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History Correlation Matrix (Correlation Matrix) ExampleSet (Remove Useless Attributes) ExampleSet (Remove Correlated Attributes)

Questions?

**Data**

**Pairwise Table**

**Charts**

**Annotations**

Attributes	BusinessTravel = Travel... 1	BusinessTravel = Non-Tra... -0.528	BusinessTravel = Travel_F... -0.753	Department = Sales -0.005	Department = Research &... 0.001	Department = Human Res... 0.009	EducationField = Life Scie... -0.029	EducationField = Other 0.018	EducationField = Human ... -0.024	EducationField = Technic... 0.017	EducationField = Medical -0.005	EducationField = Marketing 0.034	Gender = Female 0.016	Gender = Male -0.016	JobRole = Sales Executive -0.012	JobRole = Research Direct... 0.034	JobRole = Laboratory Tec... -0.015	JobRole = Manager 0.027	JobRole = Human Resour... 0.009	JobRole = Research Scie... 0.013	JobRole = Sales Represe... -0.026	JobRole = Healthcare Rep... -0.016	JobRole = Manufacturing ... 0.000	MaritalStatus = Single -0.019	
BusinessTravel = Travel... 1	-0.528	-0.753	-0.005	0.001	0.009	-0.029	0.018	-0.024	0.017	-0.005	0.034	0.016	-0.016	-0.012	0.034	-0.021	-0.021	-0.023	-0.033	0.809	-0.158	-0.009	-0.175	-0.051	
BusinessTravel = Non-Tra... -0.528	1	-0.162	0.007	-0.005	-0.005	0.006	-0.013	0.021	0.004	0.013	-0.031	-0.050	0.050	0.031	-0.021	-0.021	-0.021	-0.020	-0.020	-0.009	-0.158	-0.009	-0.175	-0.051	
BusinessTravel = Travel_F... -0.753	-0.162	1	0.001	0.002	-0.007	0.029	-0.011	0.012	-0.023	-0.004	-0.016	0.020	-0.020	-0.009	-0.021	-0.021	-0.021	-0.020	-0.020	-0.009	-0.158	-0.009	-0.175	-0.051	
Department = Sales -0.005	0.007	0.001	1	-0.907	-0.140	-0.101	-0.064	-0.090	-0.031	-0.168	0.528	0.033	-0.033	0.809	-0.158	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
Department = Research &... 0.001	-0.005	0.002	-0.907	1	-0.291	0.127	0.065	-0.188	0.039	0.184	-0.478	-0.016	0.016	-0.733	0.175	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
Department = Human Res... 0.009	-0.005	-0.007	-0.140	-0.291	1	-0.068	-0.008	0.646	-0.020	-0.050	-0.074	-0.035	0.035	-0.113	-0.051	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
EducationField = Life Scie... -0.029	0.006	0.029	-0.101	0.127	-0.068	1	-0.203	-0.115	-0.263	-0.569	-0.292	-0.008	0.008	-0.091	0.019	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
EducationField = Other 0.018	-0.013	-0.011	-0.064	0.065	-0.008	-0.203	1	-0.033	-0.076	-0.165	-0.085	-0.023	0.023	-0.037	-0.006	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
EducationField = Human ... -0.024	0.021	0.012	-0.090	-0.188	0.646	-0.115	-0.033	1	-0.043	-0.093	-0.048	-0.029	0.029	-0.073	-0.033	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
EducationField = Technic... 0.017	0.004	-0.023	-0.031	0.039	-0.020	-0.263	-0.076	-0.043	1	-0.213	-0.109	-0.004	0.004	-0.059	-0.023	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
EducationField = Medical -0.005	0.013	-0.004	-0.168	0.184	-0.050	-0.569	-0.165	-0.093	-0.213	1	-0.237	0.014	-0.014	-0.134	0.063	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
EducationField = Marketing 0.034	-0.031	-0.016	0.528	-0.478	-0.074	-0.292	-0.085	-0.048	-0.109	-0.237	1	0.024	-0.024	0.157	-0.084	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
Gender = Female 0.016	-0.050	0.020	0.033	-0.016	-0.035	-0.008	-0.023	-0.029	-0.004	0.014	0.024	1	-1	0.001	0.006	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
Gender = Male -0.016	0.050	-0.020	-0.033	0.016	0.035	0.008	-0.024	-1	1	-0.006	-0.024	-1	1	-0.006	-0.006	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
JobRole = Sales Executive -0.012	0.031	-0.009	0.809	-0.733	-0.113	-0.091	-0.457	0.006	-0.006	1	-0.128	-0.084	0.006	-0.006	-0.128	1	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
JobRole = Research Direct... 0.034	-0.021	-0.023	-0.158	0.175	-0.051	0.019	-0.161	-0.067	0.067	-0.247	-0.111	-0.161	-0.067	0.034	-0.146	-0.066	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
JobRole = Laboratory Tec... -0.015	0.009	0.011	-0.305	0.337	-0.098	0.045	-0.026	-0.067	-0.036	0.036	-0.102	-0.046	-0.067	-0.036	-0.102	-0.046	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
JobRole = Manager 0.027	0.014	-0.042	0.035	-0.071	0.088	-0.011	-0.057	-0.024	-0.006	0.011	-0.265	-0.119	-0.057	-0.029	-0.131	-0.059	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
JobRole = Human Resour... 0.009	-0.016	0.002	-0.126	-0.263	0.905	-0.063	-0.067	-0.036	0.036	-0.102	-0.046	-0.067	-0.036	-0.102	-0.046	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
JobRole = Research Scie... 0.013	-0.010	-0.007	-0.328	0.362	-0.105	0.042	0.006	-0.068	0.077	0.041	-0.173	-0.011	0.011	-0.265	-0.119	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
JobRole = Sales Represe... -0.026	-0.034	0.056	0.371	-0.336	-0.052	-0.043	-0.034	-0.033	0.057	-0.052	0.133	0.029	-0.029	-0.131	-0.059	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
JobRole = Healthcare Rep... -0.016	0.013	0.008	-0.207	0.228	-0.066	0.029	0.018	-0.043	0.019	0.034	-0.109	-0.007	0.007	-0.167	-0.075	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
JobRole = Manufacturing ... 0.000	-0.014	0.010	-0.219	0.241	-0.070	0.052	-0.011	-0.045	0.008	0.035	-0.115	0.066	-0.066	-0.177	-0.079	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	
MaritalStatus = Single -0.019	-0.004	0.026	0.034	-0.011	-0.051	0.020	0.005	-0.072	0.015	-0.004	-0.013	0.032	-0.032	-0.032	0.007	-0.042	-0.009	-0.175	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051	-0.051

Dummy Variable

Repository

Add Data

- Samples
- DB
- Local Repository (manoela)
  - data (manoela)
  - processes (manoela)
    - IBM\_AttributeSelection (manoela - v1, 5/3/17 3:54 PM - 12 kB)
    - IBM\_MissingValues (manoela - v1, 5/3/17 3:41 PM - 13 kB)
- Cloud Repository (disconnected)



File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History Correlation Matrix (Correlation Matrix) ExampleSet (Remove Useless Attributes) ExampleSet (Remove Correlated Attributes)

Repository

	JobRole...	Marital...	Marital...	Marital...	Overtime = Yes	Overtime = No	HourlyR...	MonthlyIncome	Age	DailyRate	Distance...	Educati...	Employ...	Environ...	JobInvo...	JobLevel	J
-0.164	0.052	-0.039	-0.011	0.055	-0.055	0.018	-0.345	-0.146	-0.002	-0.012	-0.001	-0.017	0.002	0.049	-0.387	0	
-0.081	0.073	-0.024	-0.053	0.003	-0.003	-0.018	-0.202	-0.176	0.005	-0.016	-0.091	0.006	0.003	-0.027	-0.217	0	
-0.104	-0.030	0.005	0.028	-0.001	0.001	0.015	0.068	0.099	0.040	0.023	0.025	0.026	0.014	0.001	0.116	0	
1	-0.021	0.003	0.020	-0.010	0.010	-0.014	0.056	0.050	-0.005	0.012	-0.005	-0.014	0.059	-0.022	0.115	-1	
-0.021	1	-0.630	-0.366	-0.006	0.006	-0.035	-0.089	-0.119	-0.076	-0.028	0.003	-0.035	0.009	-0.044	-0.086	0	
0.003	-0.630	1	-0.492	-0.14	0.014	0.038	0.056	0.084	0.040	0.031	-0.001	0.054	-0.022	0.028	0.050	-1	
0.020	-0.366	-0.492	1	0.023	-0.023	-0.006	0.032	0.033	0.037	-0.005	-0.002	-0.025	0.017	0.016	0.037	-1	
-0.010	-0.006	-0.014	0.023	1	-1	-0.007	0.006	0.028	0.009	0.026	-0.020	-0.024	0.070	-0.004	0.000	0	
0.010	0.006	0.014	0.023	-1	1	0.007	-0.006	-0.028	-0.009	-0.026	0.020	0.024	-0.070	0.004	-0.000	-1	
-0.014	-0.035	0.038	-0.006	-0.007	0.007	1	-0.015	0.025	0.024	0.030	0.015	0.036	-0.050	0.044	-0.027	-1	
0.056	-0.089	0.056	0.032	0.006	-0.006	0.015	1	0.498	0.008	-0.017	0.096	-0.015	-0.006	-0.016	0.950	-1	
0.050	-0.119	0.084	0.033	0.028	-0.020	0.025	0.498	1	0.011	-0.002	0.208	-0.010	0.010	0.030	0.510	-1	
-0.005	-0.076	0.040	0.037	0.009	-0.009	0.024	0.008	0.011	1	-0.005	-0.017	-0.051	0.018	0.046	0.003	0	
0.012	-0.028	0.031	-0.005	0.026	-0.026	0.030	-0.017	-0.002	-0.005	1	0.020	0.033	-0.016	0.009	0.006	-1	
-0.005	0.003	-0.001	-0.002	-0.020	0.020	0.015	0.096	0.208	-0.017	0.020	1	0.042	-0.027	0.043	0.102	-1	
-0.014	-0.035	0.054	-0.025	-0.024	0.024	0.036	-0.015	-0.010	-0.051	0.033	0.042	1	0.018	-0.007	-0.019	-1	
0.059	0.009	-0.022	0.017	0.070	-0.070	-0.050	-0.006	0.010	0.018	-0.016	-0.027	0.018	1	-0.008	0.001	-1	
-0.022	-0.044	0.028	0.016			-0.016	0.030	0.046	0.009	0.043	-0.007	-0.008	1	-0.013		-1	
0.115	-0.086	0.050	0.037			0.950	0.510	0.003	0.006	0.102	-0.019	0.001	-0.013	1			-1
-0.013	0.023	-0.010	-0.015			-0.006	-0.005	0.031	-0.004	-0.012	-0.046	-0.007	-0.021	-0.001	1		-1
0.008	0.038	-0.035	-0.001	0.021	-0.021	-0.014	0.034	0.003	-0.032	0.028	-0.025	0.012	0.038	-0.017	0.039	0	
0.009	-0.019	-0.017	0.041	-0.021	0.021	0.023	0.149	0.300	0.038	-0.029	0.127	-0.001	0.013	0.015	0.142	-1	
0.035	-0.002	0.021	-0.023	-0.005	0.005	-0.009	-0.027	0.004	0.023	0.040	-0.011	-0.013	-0.032	-0.017	-0.035	0	
-0.030	-0.001	0.009	-0.010	0.004	-0.004	-0.002	-0.017	0.002	0.000	0.027	-0.024	-0.020	-0.029	-0.029	-0.022	0	

//LocalRepository/processes/IBM\_AttributeSelection\* – RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History Correlation Matrix (Correlation Matrix) ExampleSet (Remove Useless Attributes) ExampleSet (Remove Correlated Attributes)

Três atributos foram excluídos

Repository

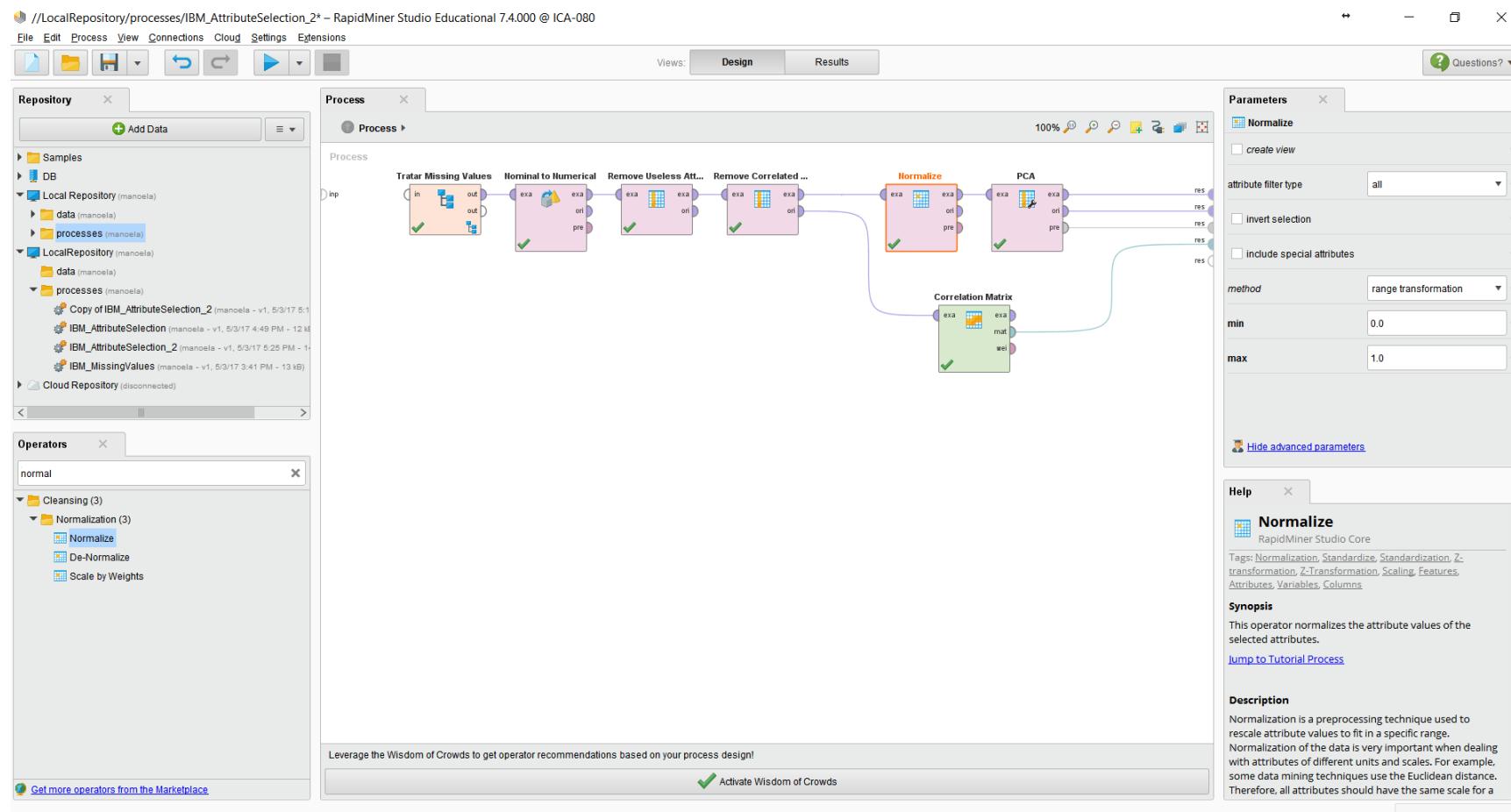
Samples DB Local Repository (manoela) LocalRepository (manoela) data (manoela) processes (manoela) IBM\_AttributeSelection (manoela - v1, 5/3/17 3:54 PM - 12 kB) IBM\_MissingValues (manoela - v1, 5/3/17 3:41 PM - 13 kB) Cloud Repository (disconnected)

The screenshot shows the RapidMiner Studio interface. In the center, there is a table titled 'Correlation Matrix (Correlation Matrix)' with 26 rows and 10 columns. A green box highlights the title bar of this table, which contains the text 'ExampleSet (1469 examples, 1 special attribute, 49 regular attributes)'. An annotation box with a black border and white background is overlaid on the table, containing the text 'Três atributos foram excluídos' (Three attributes were excluded). On the left side, there is a sidebar with icons for Data, Statistics, Charts, Advanced Charts, and Annotations. The 'Annotations' icon is highlighted. On the right side, there is a 'Repository' panel showing a tree structure of local and cloud repositories. The 'Local Repository (manoela)' node is expanded, showing 'data (manoela)', 'processes (manoela)', 'IBM\_AttributeSelection (manoela - v1, 5/3/17 3:54 PM - 12 kB)', and 'IBM\_MissingValues (manoela - v1, 5/3/17 3:41 PM - 13 kB)'. The 'Cloud Repository' node is shown as '(disconnected)'.

# **Seleção de atributos**

## **PCA**

# 3. SELEÇÃO DE ATRIBUTOS: PCA



//LocalRepository/processes/IBM\_AttributeSelection\_2\* – RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History Correlation Matrix (Correlation Matrix) ExampleSet (Remove Useless Attributes) PCA (PCA) ExampleSet (Normalize) ExampleSet (PCA)

Eigenvalues

Component	Standard Deviation	Proportion of Variance	Cumulative Variance
PC 1	0.798	0.121	0.121
PC 2	0.642	0.078	0.199
PC 3	0.602	0.069	0.268
PC 4	0.571	0.062	0.330
PC 5	0.533	0.054	0.384
PC 6	0.502	0.048	0.431
PC 7	0.471	0.042	0.474
PC 8	0.452	0.039	0.512
PC 9	0.429	0.035	0.547
PC 10	0.423	0.034	0.581
PC 11	0.385	0.028	0.609
PC 12	0.370	0.026	0.635
PC 13	0.361	0.025	0.660
PC 14	0.359	0.024	0.685
PC 15	0.354	0.024	0.708
PC 16	0.347	0.023	0.731
PC 17	0.325	0.020	0.751
PC 18	0.315	0.019	0.770
PC 19	0.306	0.018	0.788
PC 20	0.305	0.018	0.805
PC 21	0.295	0.017	0.822
PC 22	0.291	0.016	0.838
PC 23	0.287	0.016	0.854
PC 24	0.284	0.015	0.869
PC 25	0.275	0.014	0.883

Eigenvectors

Cumulative Variance

Annotations

Repository

- Samples
- DB
- Local Repository (manoela)
  - data (manoela)
  - processes (manoela)
- LocalRepository (manoela)
  - data (manoela)
  - processes (manoela)
    - Copy of IBM\_AttributeSelection\_2 (manoela - v1, 5/3/17 5:19 PM - 1 kB)
    - IBM\_AttributeSelection (manoela - v1, 5/3/17 4:49 PM - 12 kB)
    - IBM\_AttributeSelection\_2 (manoela - v1, 5/3/17 5:01 PM - 13 kB)
    - IBM\_MissingValues (manoela - v1, 5/3/17 3:41 PM - 13 kB)
- Cloud Repository (disconnected)

//LocalRepository/processes/IBM\_AttributeSelection\_2\* - RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Normalize) Correlation Matrix (Correlation Matrix) PCA (PCA) ExampleSet (PCA)

Eigenvalues

Component	Standard Deviation	Proportion of Variance	Cumulative Variance
PC 17	0.325	0.020	0.751
PC 18	0.315	0.019	0.770
PC 19	0.306	0.018	0.788
PC 20	0.305	0.018	0.805
PC 21	0.295	0.017	0.822
PC 22	0.291	0.016	0.838
PC 23	0.287	0.016	0.854
PC 24	0.284	0.015	0.869
PC 25	0.275	0.014	0.883
PC 26	0.266	0.013	0.897
PC 27	0.248	0.012	0.908
PC 28	0.246	0.011	0.920
PC 29	0.242	0.011	0.931
PC 30	0.235	0.010	0.941
PC 31	0.229	0.010	0.951
PC 32	0.227	0.010	0.961
PC 33	0.209	0.008	0.969
PC 34	0.190	0.007	0.976
PC 35	0.173	0.006	0.982
PC 36	0.152	0.004	0.986
PC 37	0.139	0.004	0.990
PC 38	0.110	0.002	0.992
PC 39	0.108	0.002	0.995
PC 40	0.098	0.002	0.996
PC 41	0.092	0.001	0.998

Eigenvectors

Cumulative Variance

Annotations

Repository

- Samples
- DB
- Local Repository (manoela)
  - data (manoela)
  - processes (manoela)
- LocalRepository (manoela)
  - data (manoela)
  - processes (manoela)
    - Copy of IBM\_AttributeSelection\_2 (manoela - v1, 5/3/17 5:19 PM - 1)
    - IBM\_AttributeSelection (manoela - v1, 5/3/17 4:49 PM - 12 kB)
    - IBM\_AttributeSelection\_2 (manoela - v1, 5/3/17 5:25 PM - 14 kB)
    - IBM\_MissingValues (manoela - v1, 5/3/17 3:41 PM - 13 kB)
- Cloud Repository (disconnected)

//LocalRepository/processes/IBM\_AttributeSelection\_2\* - RapidMiner Studio Educational 7.4.000 @ ICA-080

File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Normalize) Correlation Matrix (Correlation Matrix) PCA (PCA) ExampleSet (PCA)

ExampleSet (1469 examples, 1 special attribute, 31 regular attributes)

Data Statistics Charts Advanced Charts Annotations

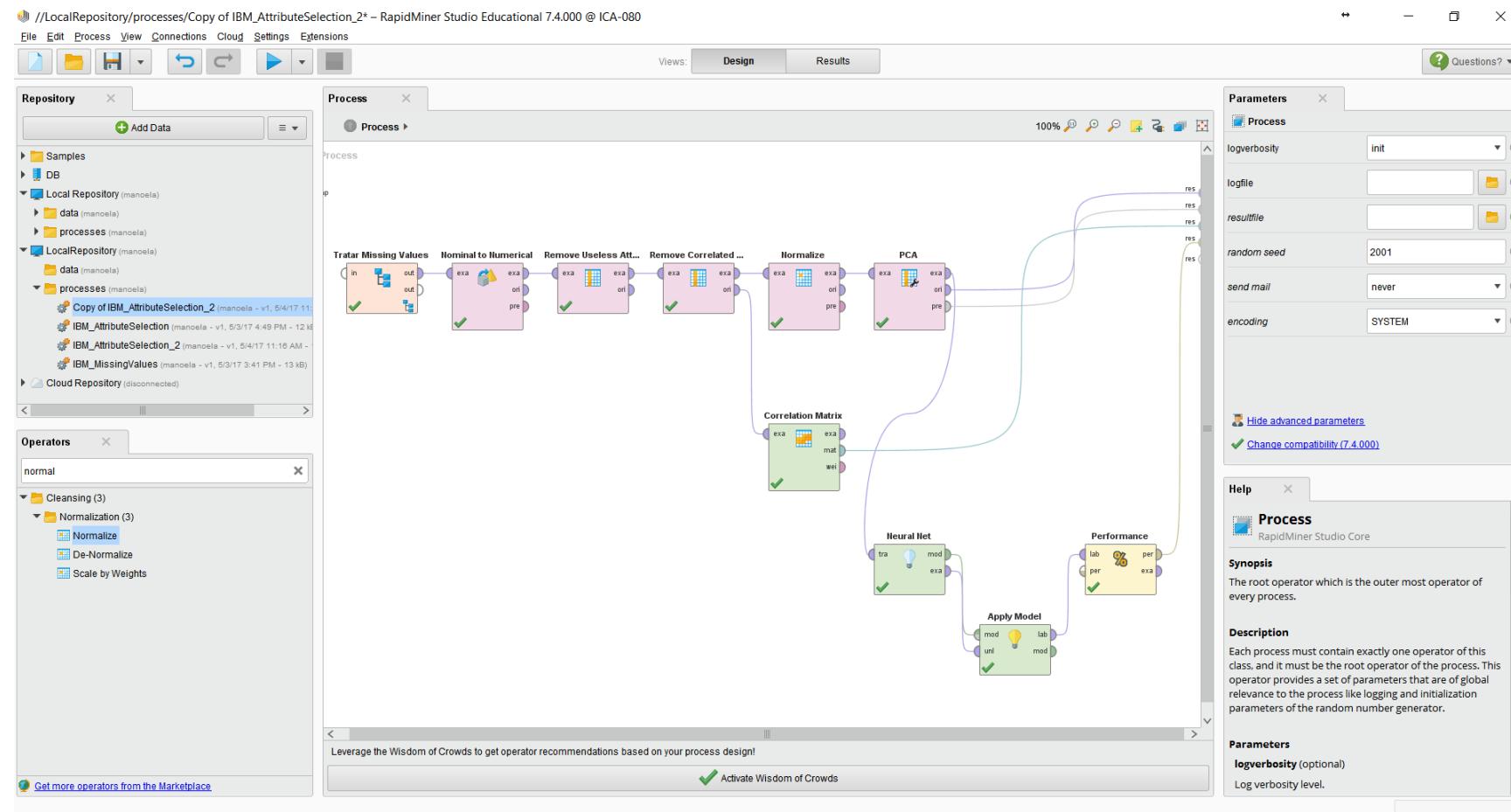
Redução de 18 atributos

Row No.	Attrition	pc_1	pc_2	pc_3	pc_4	pc_5	pc_6	pc_7	pc_8	pc_9	pc_10	pc_11	pc_12
1	Yes	1.141	-0.844	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.714	0.252
2	No	-0.402	-0.668	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.041	0.373
3	Yes	-0.564	-0.745	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.555	0.314
4	No	-0.389	-0.885	0.641	0.308	0.150	0.115	0.922	0.500	0.063	0.275	0.021	0.427
5	No	0.284	0.610	-0.069	-0.319	0.187	0.589	0.210	-0.023	-0.090	0.382	0.511	-1.07
6	No	1.218	-0.092	-0.241	-0.214	-0.853	-0.426	-0.209	-0.743	-0.002	0.361	0.427	-0.15
7	No	1.160	0.521	0.711	-0.322	0.029	0.434	0.288	0.062	0.037	0.116	-0.802	-0.19
8	No	-0.663	-0.086	-0.725	0.957	-0.895	0.420	0.430	-0.433	-0.507	-0.031	0.062	-0.26
9	No	-0.641	-1.091	0.599	0.277	0.321	0.557	-0.118	0.122	0.418	0.574	0.007	0.048
10	No	-0.521	0.389	0.078	0.978	0.782	-0.117	-0.159	-0.821	-0.129	-0.841	0.865	-0.14
11	No	-0.607	-0.787	0.549	-0.651	0.135	0.407	0.095	0.080	0.603	0.450	-0.473	0.168
12	No	-0.389	0.839	-0.708	-0.189	0.101	-0.195	0.491	1.243	-0.258	-0.012	0.037	-0.64
13	No	1.098	-0.239	0.589	-0.178	-1.069	0.434	-0.043	-0.279	0.001	0.164	-0.631	0.063
14	No	-0.619	0.740	-0.620	-0.201	0.236	0.498	-0.015	0.048	0.626	0.518	-0.647	0.175
15	Yes	-0.606	-0.882	0.485	-0.594	0.119	0.435	-0.223	0.347	-0.904	-0.520	-0.075	0.100
16	No	-0.404	-0.034	-0.648	1.224	-0.587	-1.072	0.733	0.047	0.020	0.375	0.002	0.575
17	No	1.170	0.534	0.651	-0.404	0.251	-0.288	-0.180	-0.537	-0.153	0.136	0.018	-0.35
18	No	1.139	-0.073	0.587	-0.118	-1.221	0.150	0.407	-0.104	0.131	0.154	-0.672	-0.21
19	No	1.531	-0.605	-0.259	-0.532	-0.003	0.213	0.727	0.192	0.134	0.276	0.362	-0.21
20	No	1.115	0.398	-0.585	0.738	0.416	0.693	0.162	-0.183	0.161	0.093	-0.573	0.188
21	No	-0.656	-0.867	0.551	-0.655	0.198	0.403	-0.082	0.024	1.116	-0.368	0.071	0.216
22	Yes	0.718	-0.841	0.479	-0.603	0.137	0.382	0.239	-0.186	0.612	-0.863	0.285	-0.49
23	No	-0.641	-0.695	-0.904	-0.575	0.411	-0.370	-0.209	-0.337	0.320	0.654	-0.017	-0.47
24	No	1.543	0.473	0.059	0.957	0.529	-0.612	0.267	-0.238	-0.015	0.392	0.417	0.237
25	Yes	-0.689	-0.698	-0.889	-0.513	0.224	0.505	-0.086	0.013	0.392	0.577	-0.217	-0.61
26	No	-0.413	-0.698	-0.267	-0.634	0.291	-0.444	0.155	-0.338	-0.111	0.207	0.783	0.490

Repository

- Samples
- DB
- Local Repository (manoela)
  - data (manoela)
  - processes (manoela)
- LocalRepository (manoela)
  - data (manoela)
  - processes (manoela)
    - Copy of IBM\_AttributeSelection\_2 (manoela - v1, 5/3/17 5:19 PM - 1 kB)
    - IBM\_AttributeSelection (manoela - v1, 5/3/17 4:49 PM - 12 kB)
    - IBM\_AttributeSelection\_2 (manoela - v1, 5/3/17 5:25 PM - 14 kB)
    - IBM\_MissingValues (manoela - v1, 5/3/17 3:41 PM - 13 kB)
- Cloud Repository (disconnected)

# 4. APLICAR MODELO DE *MACHINE LEARNING*: REDES NEURAIS



# 5. AVALIAR

The screenshot shows the RapidMiner Studio interface with the following details:

- Title Bar:** //LocalRepository/processes/Copy of IBM\_AttributeSelection\_2\* – RapidMiner Studio Educational 7.4.000 @ ICA-080
- Menu Bar:** File Edit Process View Connections Cloud Settings Extensions
- Toolbar:** Includes icons for New, Open, Save, Undo, Redo, Run, Stop, and Help.
- Views:** Design (selected), Results
- PerformanceVector (Performance) View:** Criterion: accuracy  
accuracy: 97.28%

	true Yes	true No	class precision
pred. Yes	197	0	100.00%
pred. No	40	1232	96.86%
class recall	83.12%	100.00%	
- Correlation Matrix (Correlation Matrix) View:** Not visible in the screenshot.
- PCA (PCA) View:** Not visible in the screenshot.
- ExampleSet (Normalize) View:** Not visible in the screenshot.
- Repository View:** Shows the local repository structure:
  - Samples
  - DB
  - Local Repository (mancela)
    - data (mancela)
    - processes (mancela)
  - LocalRepository (mancela)
    - data (mancela)
    - processes (mancela)
      - Copy of IBM\_AttributeSelection\_2 (mancela - v1, 5/4/17 11:36 AM - 12 kB)
      - IBM\_AttributeSelection (mancela - v1, 5/3/17 4:49 PM - 12 kB)
      - IBM\_AttributeSelection\_2 (mancela - v1, 5/4/17 11:16 AM - 18 kB)
      - IBM\_MissingValues (mancela - v1, 5/3/17 3:41 PM - 13 kB)
  - Cloud Repository (disconnected)

# 6. MODIFICAR PARÂMETROS

The screenshot shows the RapidMiner Studio interface with the following components:

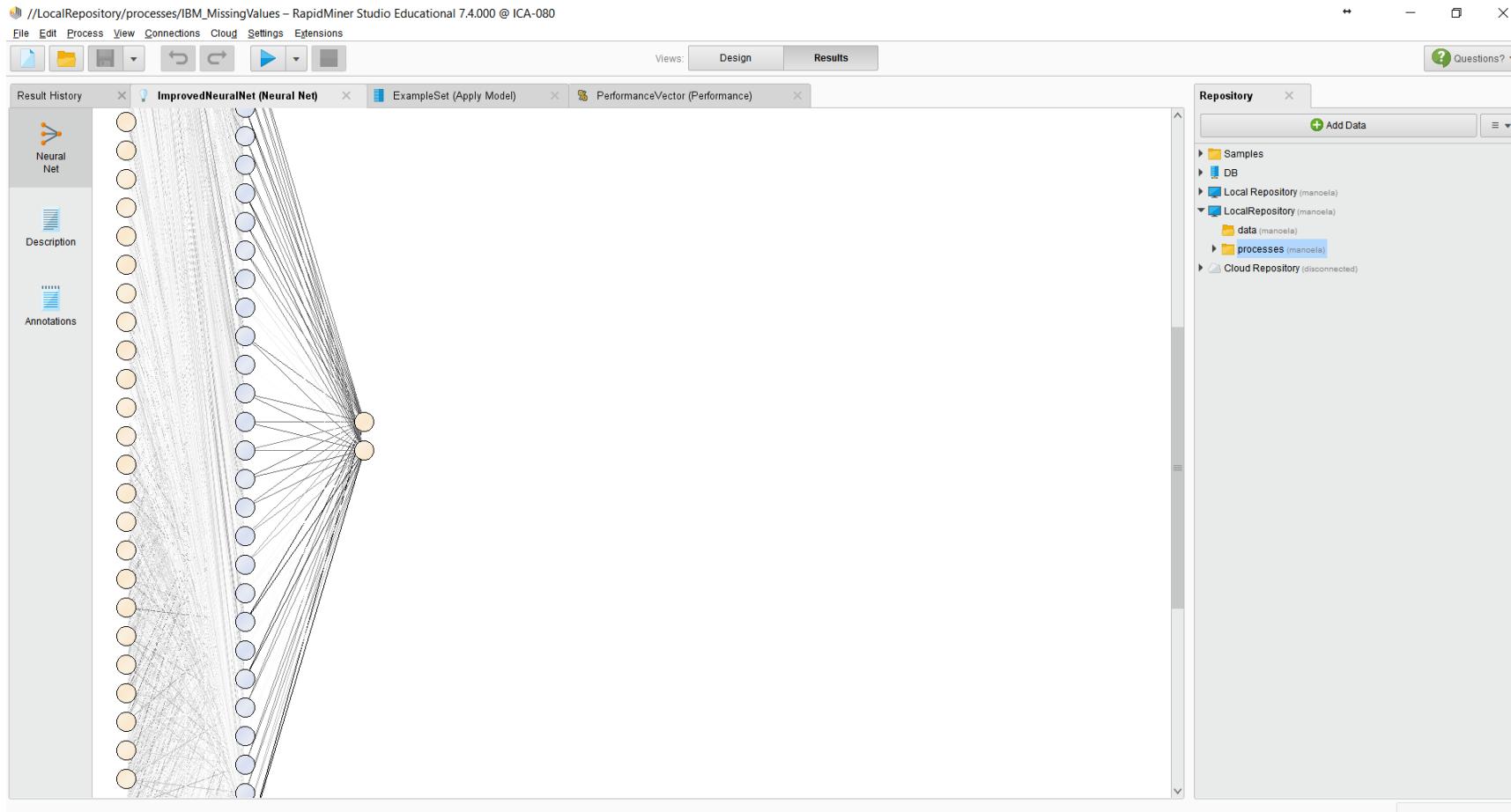
- Repository View:** Shows local and cloud repositories with various samples and processes.
- Process View:** Displays a data processing workflow starting with "Tratar Missing Values", followed by "Nominal to Numerical", "Remove Useless Attr...", "Remove Correlated Attr...", "Normalize", and "PCA". The output of PCA feeds into a "Correlation Matrix" operator.
- Operators View:** Shows the "Neural Net" operator selected, with its configuration dialog open.
- Parameters View:** Shows the configuration for the "Neural Net" operator. A green box highlights the "hidden layers" field, which is set to "2000". An arrow points from this field to a callout box containing the formula:  $\text{No de neurônios} = (\text{No de atributos} + \text{No de classes}) / 2$ ,  $= (31 + 2) / 2 = 16.5$ .
- Help View:** Provides detailed documentation for the "Neural Net" operator, including its synopsis, description, and tags.

# 7. AVALIAR

The screenshot shows the RapidMiner Studio interface with the following details:

- Title Bar:** //LocalRepository/processes/Copy of IBM\_AttributeSelection\_2\* – RapidMiner Studio Educational 7.4.000 @ ICA-080
- Menu Bar:** File Edit Process View Connections Cloud Settings Extensions
- Toolbar:** Includes icons for file operations, process steps like "Copy", and navigation.
- Views:** Design and Results (selected).
- Result History:** Shows the current process step: PerformanceVector (Performance).
- Performance View:** Criterion accuracy. Accuracy: 97.75%. Confusion matrix table:

	true Yes	true No	class precision
pred. Yes	204	0	100.00%
pred. No	33	1232	97.39%
class recall	86.08%	100.00%	
- Correlation Matrix (Correlation Matrix) View:** Not currently selected.
- PCA (PCA) View:** Not currently selected.
- ExampleSet (Normalize) View:** Not currently selected.
- Repository:** Shows the local repository structure:
  - Samples
  - DB
  - Local Repository (manoela)
    - data (manoela)
    - processes (manoela)
  - LocalRepository (manoela)
    - data (manoela)
    - processes (manoela)
      - Copy of IBM\_AttributeSelection\_2 (manoela - v1, 5/4/17 11:36 AM - 12 kB)
      - IBM\_AttributeSelection (manoela - v1, 5/3/17 4:49 PM - 12 kB)
      - IBM\_AttributeSelection\_2 (manoela - v1, 5/4/17 11:16 AM - 18 kB)
      - IBM\_MissingValues (manoela - v1, 5/3/17 3:41 PM - 13 kB)
  - Cloud Repository (disconnected)



File Edit Process View Connections Cloud Settings Extensions

Views: Design Results

Result History ExampleSet (Apply Model) PerformanceVector (Performance)

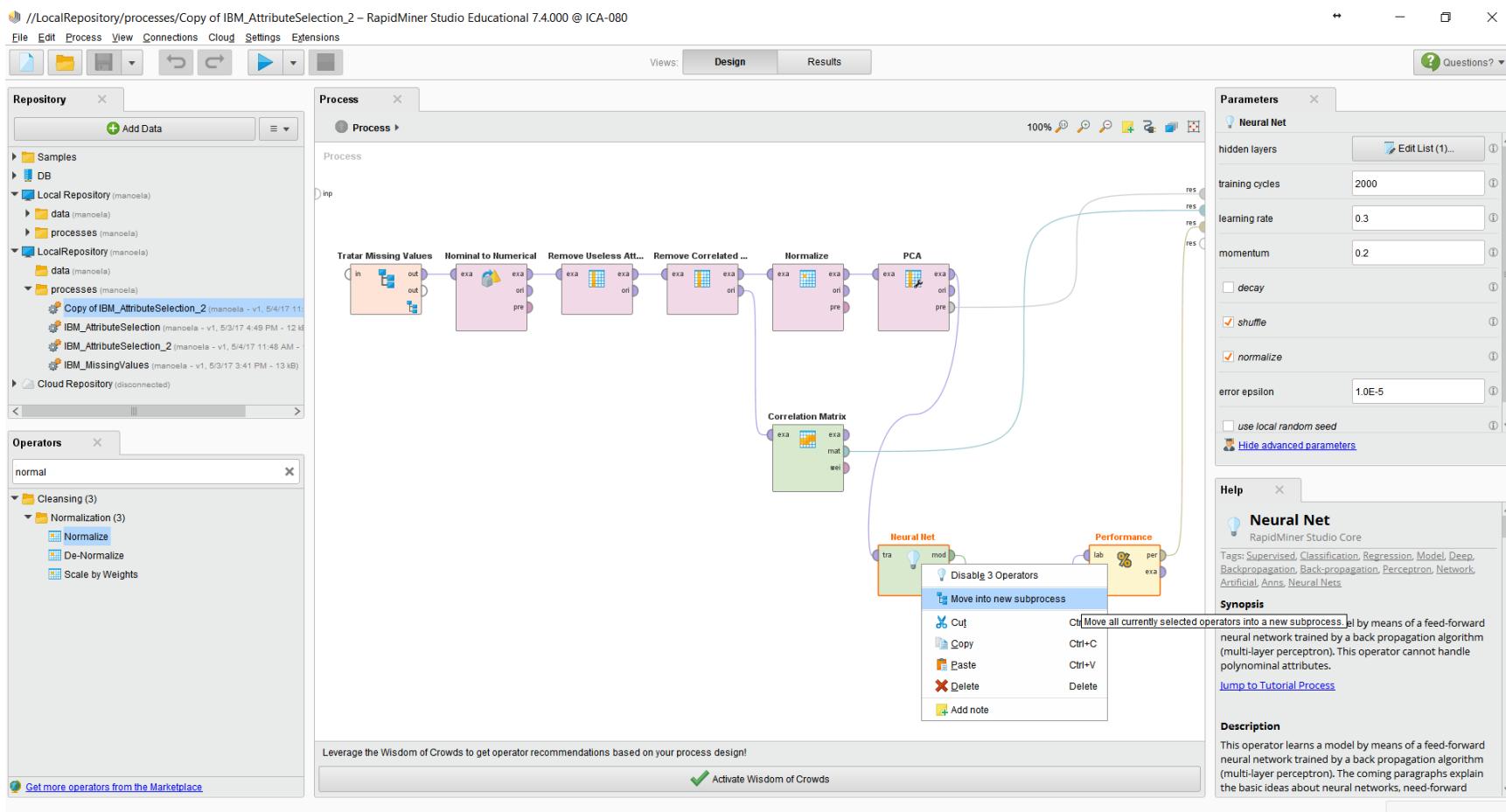
ExampleSet (1469 examples, 4 special attributes, 54 regular attributes)

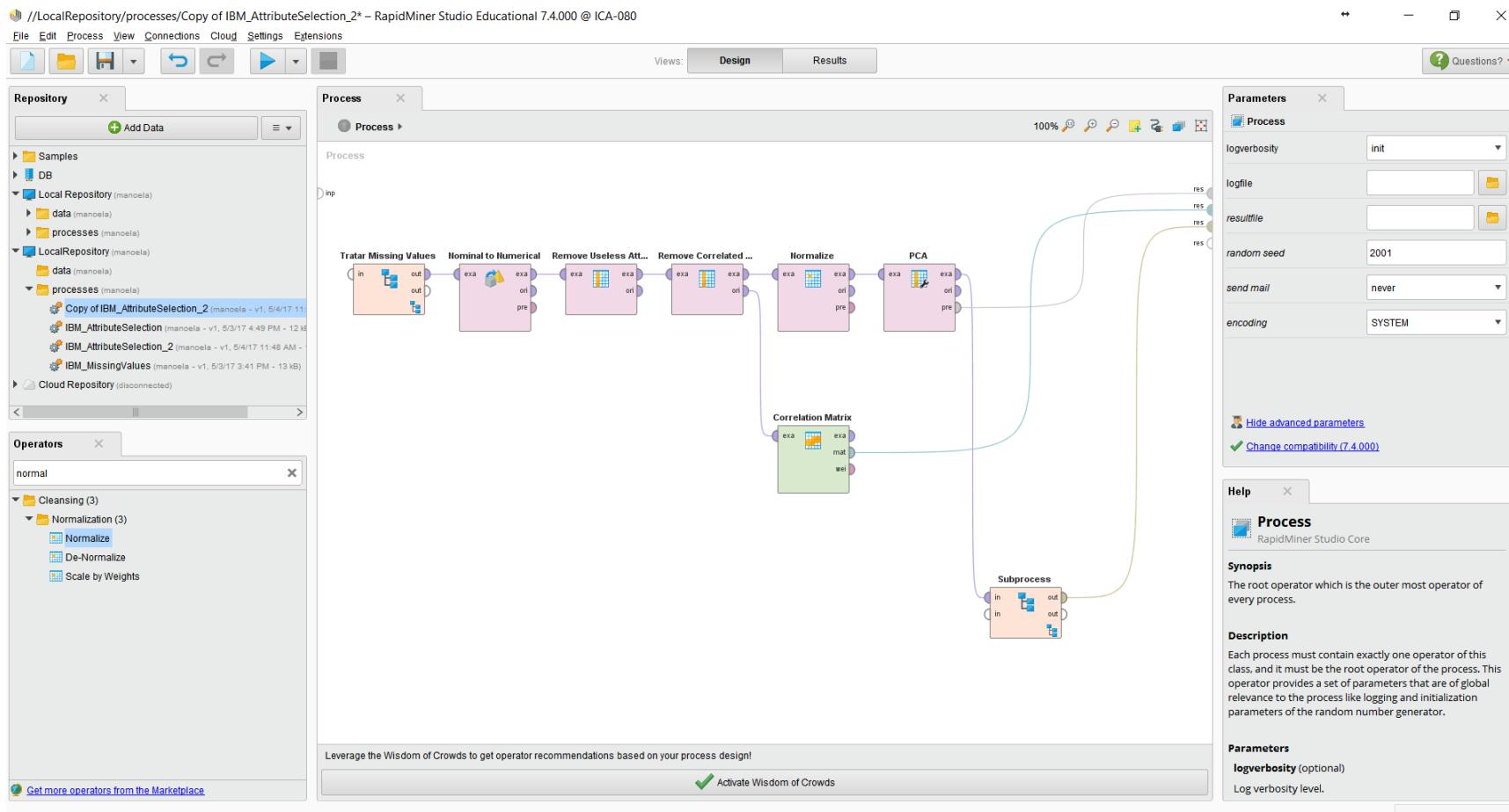
Row No.	Attrition	prediction(Attriti...	confidence(...	confidence(...	BusinessTr...	BusinessTr...	BusinessTr...	Department ...	Department ...	Department ...	EducationFl...	EducationFl...	
1	Yes	Yes	1.000	0.000	1	0	0	1	0	0	1	0	0
2	No	No	0.000	1.000	1	0	0	0	1	0	1	0	0
3	Yes	Yes	1.000	0.000	1	0	0	0	1	0	0	1	0
4	No	No	0.000	1.000	0	1	0	0	1	0	1	0	0
5	No	No	0.022	0.978	1	0	0	0	0	1	0	0	0
6	No	No	0.001	0.999	1	0	0	1	0	0	0	0	0
7	No	No	0.000	1.000	1	0	0	1	0	0	1	0	0
8	No	No	0.000	1.000	0	1	0	0	1	0	0	0	0
9	No	No	0.000	1.000	0	1	0	0	1	0	1	0	0
10	No	No	0.000	1.000	0	0	1	0	1	0	0	1	0
11	No	No	0.000	1.000	1	0	0	0	1	0	1	0	0
12	No	No	0.004	0.996	1	0	0	0	1	0	0	0	0
13	No	No	0.000	1.000	1	0	0	1	0	0	1	0	0
14	No	No	0.000	1.000	1	0	0	0	1	0	0	0	0
15	Yes	Yes	1.000	0.000	1	0	0	0	1	0	1	0	0
16	No	No	0.000	1.000	0	0	1	0	1	0	0	0	0
17	No	No	0.000	1.000	1	0	0	1	0	0	1	0	0
18	No	No	0.000	1.000	1	0	0	1	0	0	1	0	0
19	No	No	0.000	1.000	1	0	0	1	0	0	0	0	0
20	No	No	0.000	1.000	0	1	0	1	0	0	0	0	0
21	No	No	0.000	1.000	1	0	0	0	1	0	1	0	0
22	Yes	Yes	1.000	0.000	1	0	0	1	0	0	1	0	0
23	No	No	0.000	1.000	1	0	0	0	1	0	0	0	0
24	No	No	0.000	1.000	0	0	1	1	0	0	0	0	0
25	Yes	Yes	0.971	0.029	1	0	0	0	1	0	0	0	0
26	No	No	0.000	1.000	1	0	0	0	1	0	0	1	0

Repository

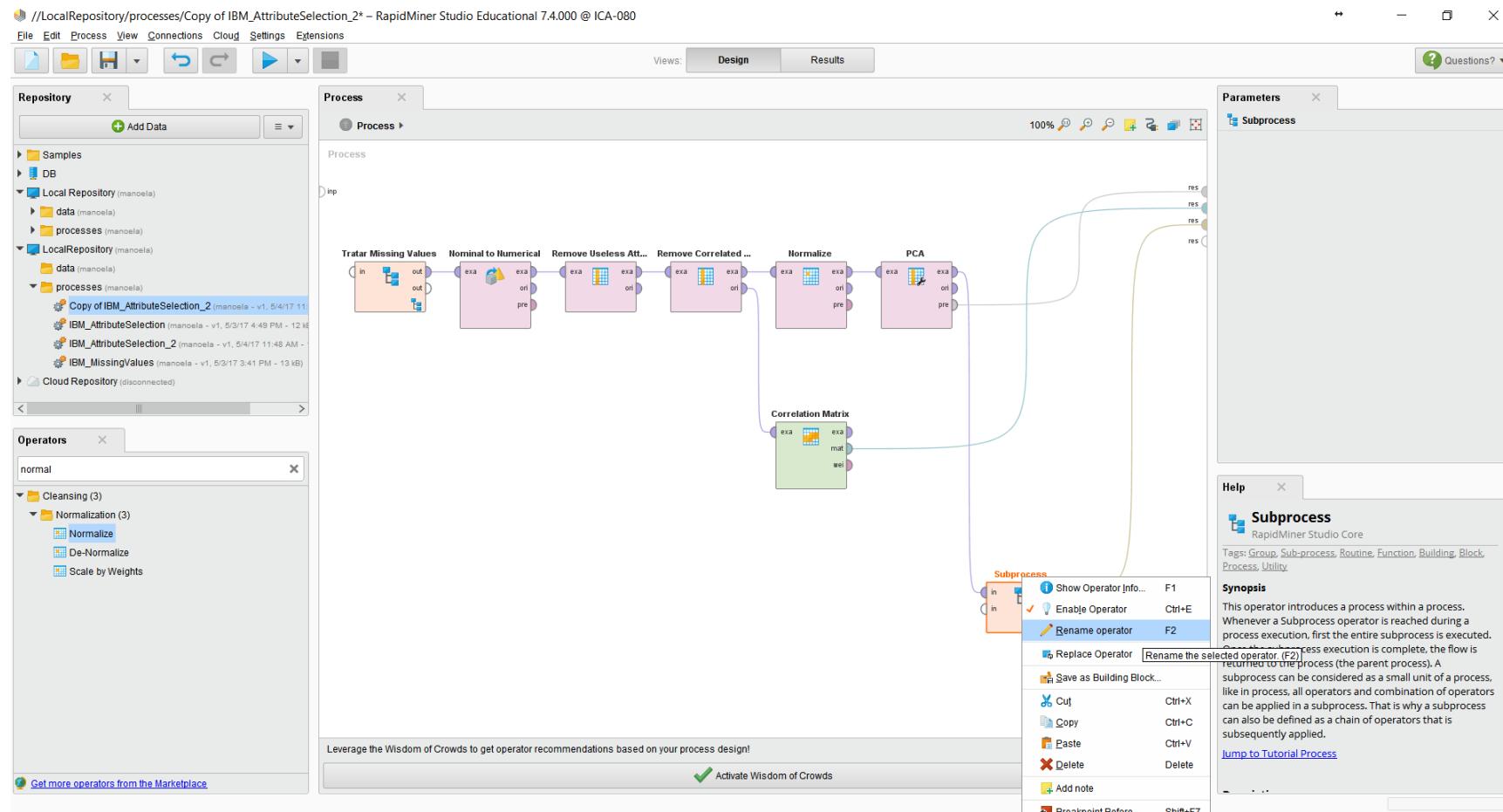
- Samples
- DB
- Local Repository (manoela)
  - data (manoela)
  - processes (manoela)
    - 01\_DecisionTree (manoela - v1, 5/2/17 2:04 PM - 2 kB)
    - IBM\_AttributeSelection (manoela - v1, 5/3/17 10:58 AM - 13 kB)
    - IBM\_Limpeza (manoela - v1, 5/2/17 4:07 PM - 8 kB)
    - IBM\_Limpeza\_NN\_SVM (manoela - v1, 5/2/17 4:34 PM - 27 kB)
    - IBM\_NeuralNet\_DecisionTree (manoela - v1, 5/2/17 3:38 PM - 18 kB)
- LocalRepository (manoela)
  - data (manoela)
  - processes (manoela)
    - IBM\_MissingValues (manoela - v1, 5/3/17 3:08 PM - 13 kB)
- Cloud Repository (disconnected)

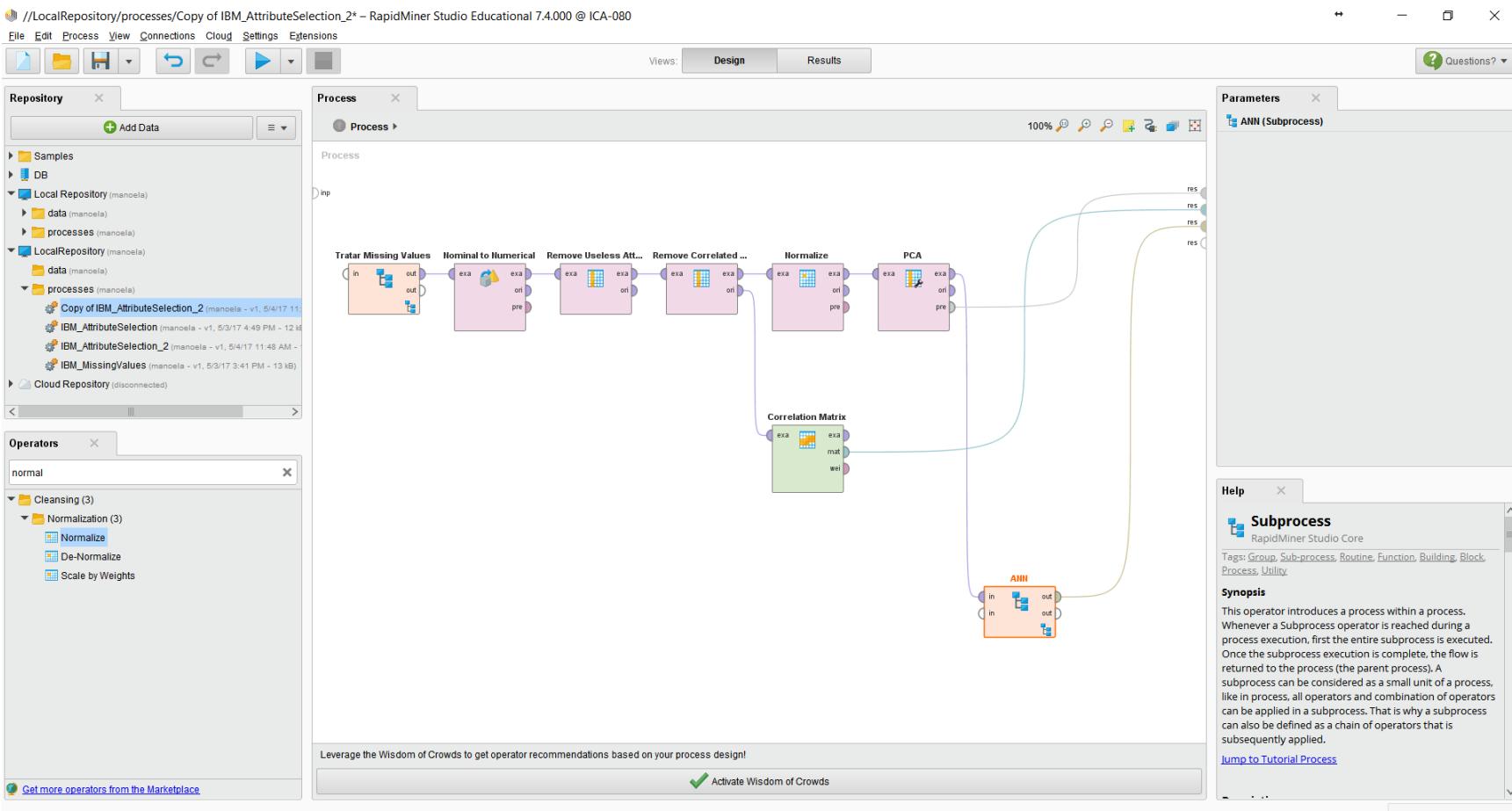
# 8. CRIAR SUBPROCESSO

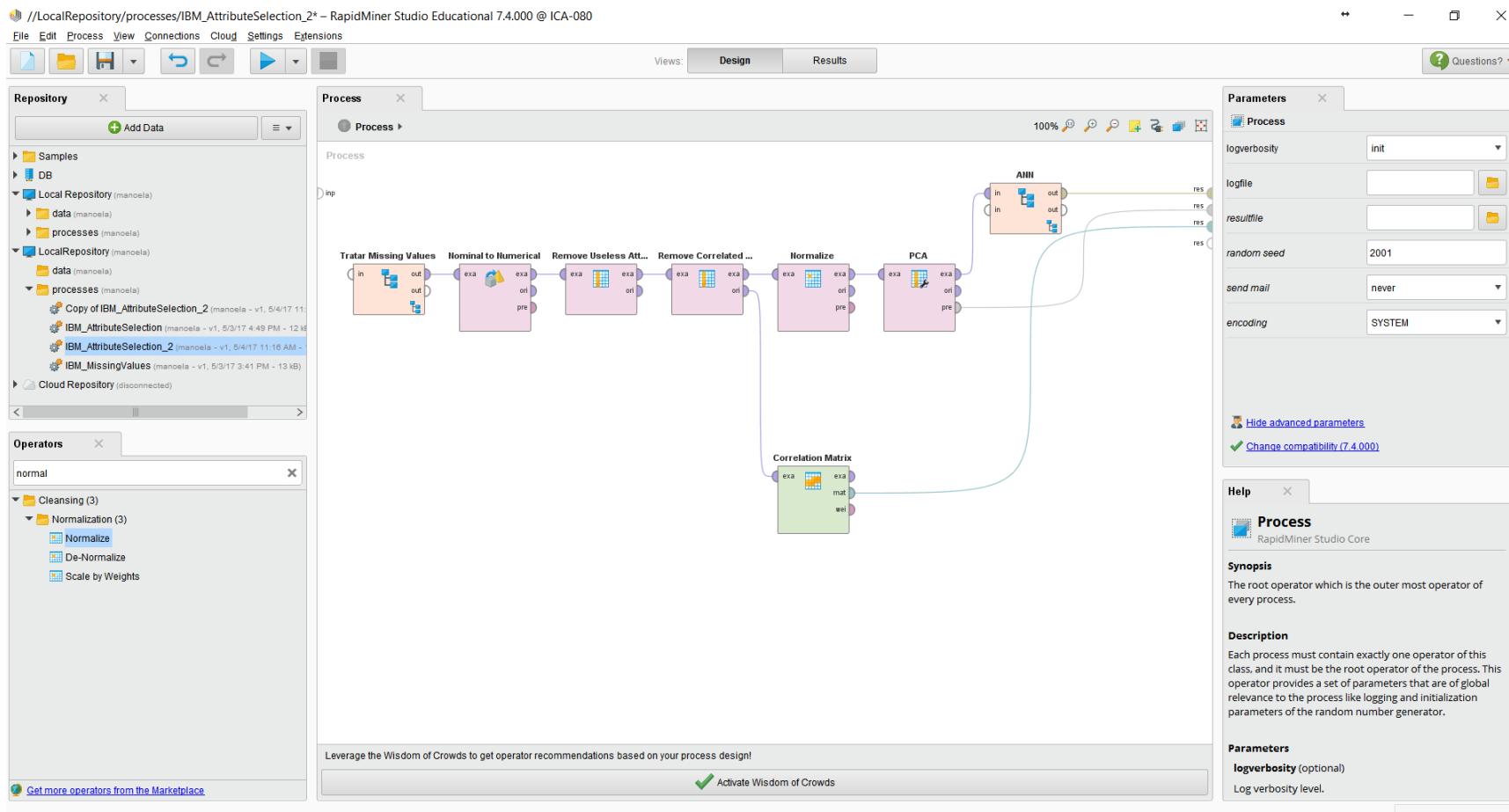




# 9. RENOMEAR SUBPROCESSO







Balanceamento?

Outro algoritmo de classificação?

Outra parametrização do algoritmo classificador escolhido?