

Attributes	Type	Distinct	Unique	Missing	Minimum	Maximum
Country	Nominal	101	0	0	-	-
Year	Numeric	32	0	0	1985	2016
Sex	Nominal	2	0	0	-	-
Age	Numeric	6	0	0	15	75
No. of Suicides	Numeric	2084	1001	0	0	22338
Population	Numeric	25564	24110	0	278	43805214
Suicides/100K	Numeric	5298	2045	0	0	224.97
Country-Year	Nominal	2321	0	0	-	-
HDI for Year	Numeric	305	0	19456	0.483	0.944
GDP/year	Nominal	2321	0	0	-	-
GDP/Capita	Nominal	2233	0	0	251	126352
Generation	Nominal	6	0	0	-	-

Exploratory Sheet

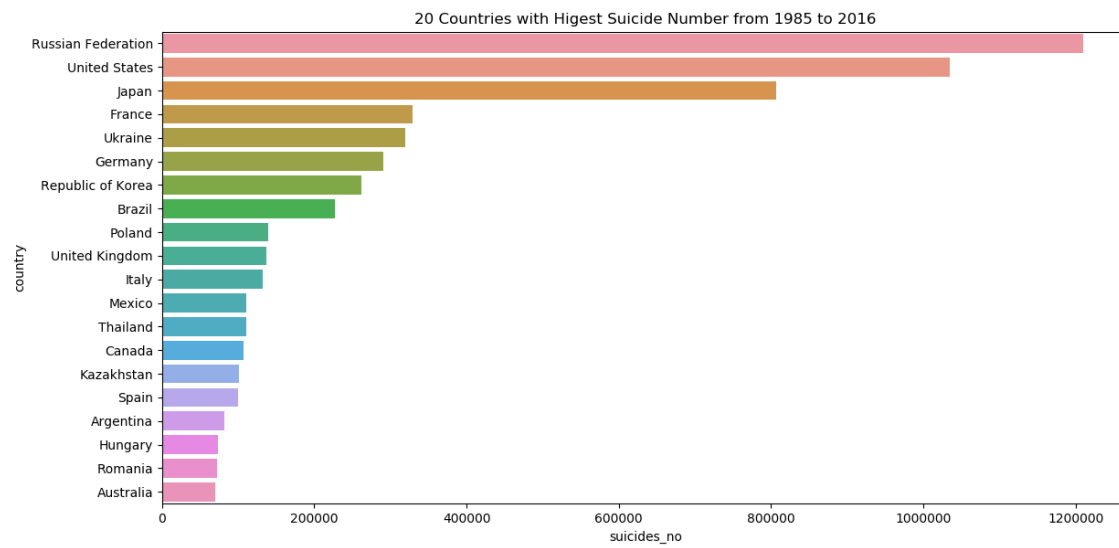
```

Python 3.6.6 Shell
File Edit Shell Debug Options Window Help
Python 3.6.6 (v3.6.6:cf1f54eb7, Jun 27 2018, 03:37:03) [MSC v.1900 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: D:\Ajay\input\BIL.py =====
--Table of Suicides according to Female Age Groups--
      age      sex  value
0  15-24 years  female  175437
1  25-34 years  female  208923
2  35-54 years  female  506233
3   5-14 years  female   16997
4  55-74 years  female  430036
5   75+ years  female  221984
|

```

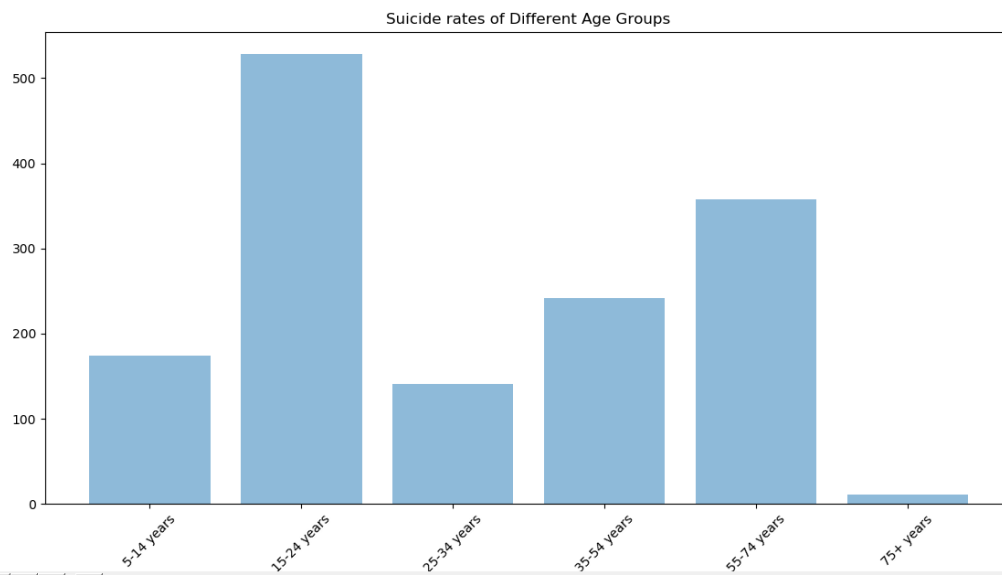
Python IDLE Output

Figure 1



Python Graphs – 1

Figure 1



Python Graphs – 2

Rapid Miner Execution

<new process> - RapidMiner Studio Trial 9.2.001 @ GOD

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Turbo Prep **Auto Model**

Find data, operators, etc. All Studio

Auto Model

Load Data Select Task Prepare Target Select Inputs Model Types Results

« RESTART < BACK > NEXT »

Predict
Want to predict the values of a column?

Clusters
Want to identify groups in your data?

Outliers
Want to detect outliers in your data?

country	year	sex	age	suicides_no	population	suicides/100k...	country-year	gdp_for_year...	gdp_per_cap...	generation
Category	Number	Category	Category	Number	Number	Number	Category	Category	Number	Category
Albania	1987	male	15-24 years	21	312900	6.710	Albania1987	2,156,624,900	796	Generati...
Albania	1987	male	35-54 years	16	308000	5.190	Albania1987	2,156,624,900	796	Silent
Albania	1987	female	15-24 years	14	289700	4.830	Albania1987	2,156,624,900	796	Generati...
Albania	1987	male	75+ years	1	21800	4.590	Albania1987	2,156,624,900	796	G.I. Gene...
Albania	1987	male	25-34 years	9	274300	3.280	Albania1987	2,156,624,900	796	Boomers
Albania	1987	female	75+ years	1	35600	2.810	Albania1987	2,156,624,900	796	G.I. Gene...
Albania	1987	female	35-54 years	6	278800	2.150	Albania1987	2,156,624,900	796	Silent
Albania	1987	female	15-24 years	4	267300	1.560	Albania1987	2,156,624,900	796	Boomers

27,820 rows - 11 columns (6 nominal, 5 numerical)

Information

Select Task

Perfect, you have selected a data set. Now you will decide what type of problem you would like to solve. Select one of the three tasks at the top before clicking Next. Here is some guidance:

- **Predict:** Select this task if you want to predict the values of one of the columns of your data, and decide the column by clicking on it. It will build a machine learning model which predicts the values of this column based on the values of the other columns.
- **Clusters:** Select this task if you want to group your data into clusters.

<new process> - RapidMiner Studio Trial 9.2.001 @ GOD

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Turbo Prep **Auto Model**

Find data, operators, etc. All Studio

Auto Model

Load Data Select Task Prepare Target Select Inputs Model Types Results

« RESTART < BACK > NEXT »

27,500
25,000
22,500
20,000
17,500
15,000
12,500
10,000
7,500
5,000
2,500
0

0 2,500 5,000 7,500 10,000 12,500 15,000 17,500 20,000 22,500

☐ Turn into Classification

Number of Classes: 2

Information

Prepare Target

OK, you are solving a regression problem. A histogram shows the distribution of values.

If this is indeed what you want to predict, and everything looks fine, click Next at the bottom of the screen.

Turn into Classification

If you're not interested in calculating an exact numerical value (e.g., 83.7 degrees Fahrenheit), and you'd rather know whether the value is "cold" or "hot", you can convert the numerical values in the target column to categorical values.

By selecting Turn into

<new process> - RapidMiner Studio Trial 9.2.001 @ GOD

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators, etc. All Studio

Auto Model

Load Data Select Task Prepare Target Select Inputs Model Types Results

RESTART BACK NEXT

Selected: 5 / Total: 10

Deselect Yellow Select All Deselect All

Selected	Status	Quality	Name	Correlation	ID-ness	Stability	Missing	Text-ness
<input checked="" type="checkbox"/>	●		country	1.41%	0.36%	1.37%	0.00%	9.90%
<input checked="" type="checkbox"/>	●		sex	2.09%	0.01%	50.00%	0.00%	2.22%
<input checked="" type="checkbox"/>	●		age	0.44%	0.02%	16.69%	0.00%	38.01%
<input checked="" type="checkbox"/>	●		population	37.97%	91.89%	0.07%	0.00%	0.00%
<input checked="" type="checkbox"/>	●		suicides/100k pop	9.40%	?	15.39%	0.00%	0.00%

Select Inputs

Here the focus is on the quality of your data, specifically the quality of each column of data. You may want to consider discarding data columns (Attributes) that provide less value.

How do you know which Attributes are valuable, and which are worthless? A key point is that you're looking for patterns. Without some variation in the data and some discernible patterns, the data is not likely to be useful. A quick summary of things to look out for (more details below) includes:

- (C) Columns that too closely mirror the target column,
- (I) Columns

<new process> - RapidMiner Studio Trial 9.2.001 @ GOD

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators, etc. All Studio

Auto Model

Load Data Select Task Prepare Target Select Inputs Model Types Results

RESTART BACK RUN

Data Preparation

☐ Extract Date Information

☐ Extract Text Information

Number of Extracted Features: 1,000

☐ Add Language Column

☐ Add Sentiment Column (Only English)

☐ Automatic Feature Selection

Additional Time (in Minutes): 60

Final Feature Set should be: Accurate

☐ Automatic Feature Generation

Models

☒ Generalized Linear Model

☒ Use Regularization ☐ Calculate p-Values

☒ Deep Learning

☒ Decision Tree

☒ Automatically Optimize Maximal Depth: 20

☐ Random Forest

☒ Automatically Optimize Number of Trees: 20 Maximal Depth: 20

☒ Gradient Boosted Trees - Warning: long computation time on this data!

☒ Automatically Optimize Number of Trees: 20 Maximal Depth: 20

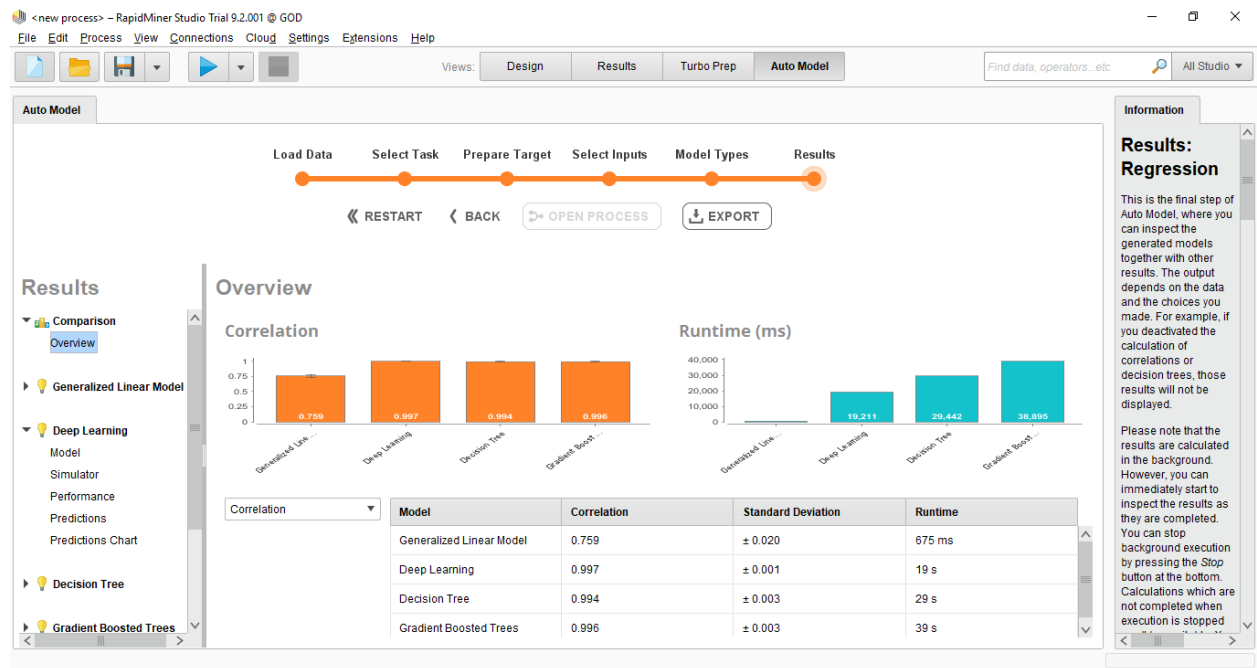
Select Model Types

Based on your data and your choices, relevant machine learning models are now displayed. If you select more than one model, the results will include a performance comparison. You need to select at least one of the models before you can click on the Run button at the bottom of the screen.

Depending on the data set, some models may be deselected, to avoid long runtimes. If you are willing to wait, feel free to include these models before pressing Run.

Data Preparation

- Extract Data If seasonal info should be extracted



<new process> - RapidMiner Studio Trial 9.2.001 @ GOD

File Edit Process View Connections Cloud Settings Extensions Help

Views: Design Results Turbo Prep Auto Model

Find data, operators, etc. All Studio

Auto Model

Load Data Select Task Prepare Target Select Inputs Model Types Results

RESTART BACK OPEN PROCESS EXPORT

Results

- Overview
- Generalized Linear Model
- Deep Learning
 - Model
 - Simulator
 - Performance
 - Optimal Parameters
 - Predictions
 - Predictions Chart
- Decision Tree
 - Model
 - Simulator
 - Performance
 - Optimal Parameters
 - Predictions
 - Predictions Chart
- Gradient Boosted Trees

Decision Tree - Predictions

Row No.	suicides_no	prediction(s...	age	l>zcountry	population	sex	suicides/10...
1	21	21.500	15-24 years	Albania	312900	male	6.710
2	16	15.500	35-54 years	Albania	308000	male	5.190
3	14	14.333	15-24 years	Albania	289700	female	4.830
4	1	1	75+ years	Albania	21800	male	4.590
5	1	2.111	75+ years	Albania	35600	female	2.810
6	0	0	55-74 years	Albania	144600	female	0
7	2	2.111	75+ years	Albania	36400	female	5.490
8	17	14.333	15-24 years	Albania	319200	male	5.330
9	14	15.500	35-54 years	Albania	314100	male	4.460
10	8	7	15-24 years	Albania	295600	female	2.710
11	16	15.500	35-54 years	Albania	308000	male	5.190

Information

Results: Regression

This is the final step of Auto Model, where you can inspect the generated models together with other results. The output depends on the data and the choices you made. For example, if you deactivated the calculation of correlations or decision trees, those results will not be displayed.

Please note that the results are calculated in the background. However, you can immediately start to inspect the results as they are completed. You can stop background execution by pressing the Stop button at the bottom. Calculations which are not completed when execution is stopped

