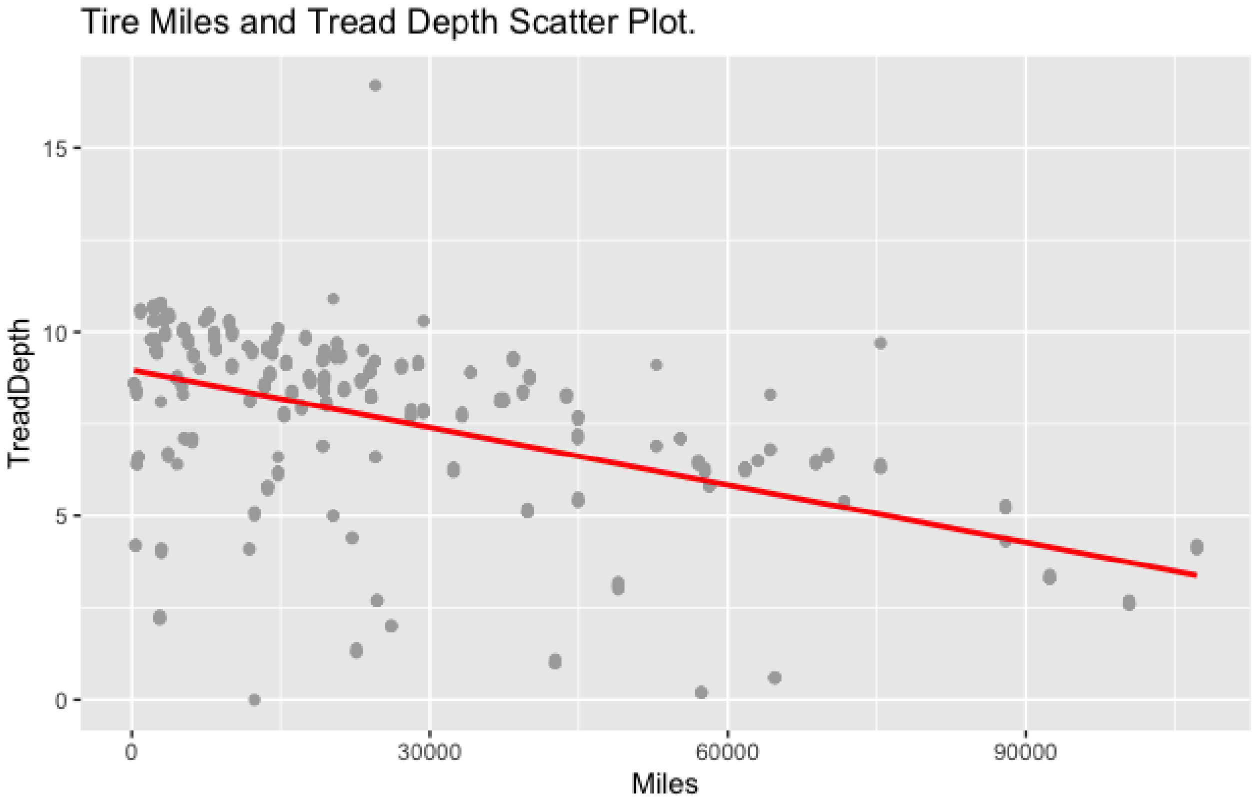
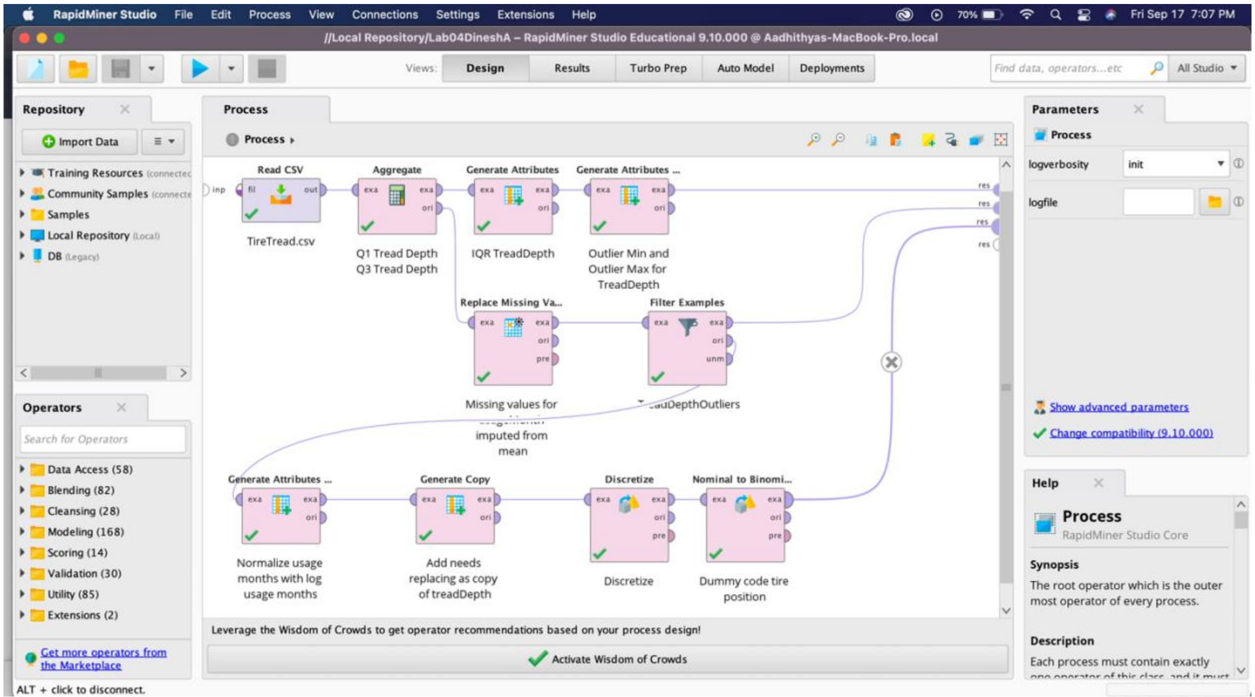
R script code:

```
# Aadhithya Dinesh
# MIS 545 Section 02
# Lab02DineshA.R
# Import and prepare a dataset of automobile tire usage and perform data
# preprocessing tasks like imputing missing data, identifying outliers,
# normalizing features, discretizing features and dummy coding.

# intsall.packages("tidyverse")
# install.packages("dummies")

library(tidyverse)
library(dummies)
library(scales)

# set the working directory
setwd("~/MIS/Classes/MIS545/Assignments/Lab04")

# read the csv file with column types specified
tireTread1 <- read_csv(file = "TireTread.csv",
                       col_types = "cfnni",
                       col_names = TRUE)

# print the tire Tread data along with summary
print(tireTread1)
str(tireTread1)
print(summary(tireTread1))

# Impute missing values with the mean value
tireTread2 <- tireTread1 %>%
  mutate(UsageMonths = ifelse(is.na(UsageMonths), mean(UsageMonths, na.rm = TRUE), UsageMonths))

# printe the summary
print(summary(tireTread2))

# outliers are separately stored in treadDepthOutliers
outlierMin <- quantile(tireTread2$TreadDepth, .25) -
  (IQR(tireTread2$TreadDepth) * 1.5)
outlierMax <- quantile(tireTread2$TreadDepth, .75) +
  (IQR(tireTread2$TreadDepth) * 1.5)

treadDepthOutliers <- tireTread2 %>%
  filter(TreadDepth < outlierMin | TreadDepth > outlierMax)

# normalize the data by taking the log
tireTread3 <- tireTread2 %>%
  mutate(LogUsageMonths = log(UsageMonths))

# discretization by setting values to true and false based on TreadDepth
tireTread4 <- tireTread3 %>%
  mutate(NeedsReplacing = TreadDepth <=1.6)
```

```
tireTread4DataFrame <- data.frame(tireTread4)
# storing the values in a tibble
tireTread5 <- as_tibble(dummy.data.frame(data = tireTread4DataFrame,
                        names = "Position"))
# creating a scatter plot visualization
scatterPlotMilesTreadDepth <- ggplot(data = tireTread5,
                      aes(x = Miles,
                        y = TreadDepth
                      ))
scatterPlotMilesTreadDepth + geom_point(color = "dark gray") +
 scale_y_continuous() +
 geom_smooth(method = lm,
        level = 0,
        color = "red") +
 ggtitle("Tire Miles and Tread Depth Scatter Plot.")
```



Rapid Miner Process and Result screenshots:



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

Activate Wisdom of Crowds

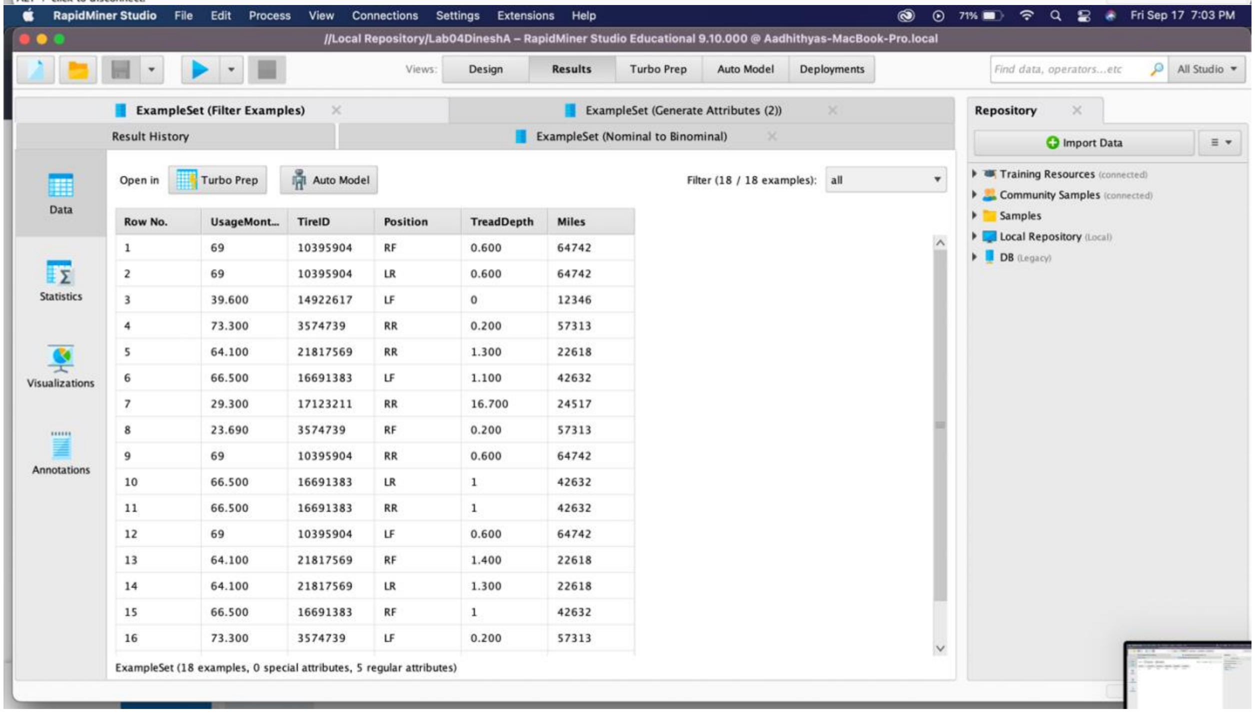
Synopsis

The root operator which is the outer most operator of every process.

Description

Each process must contain exactly

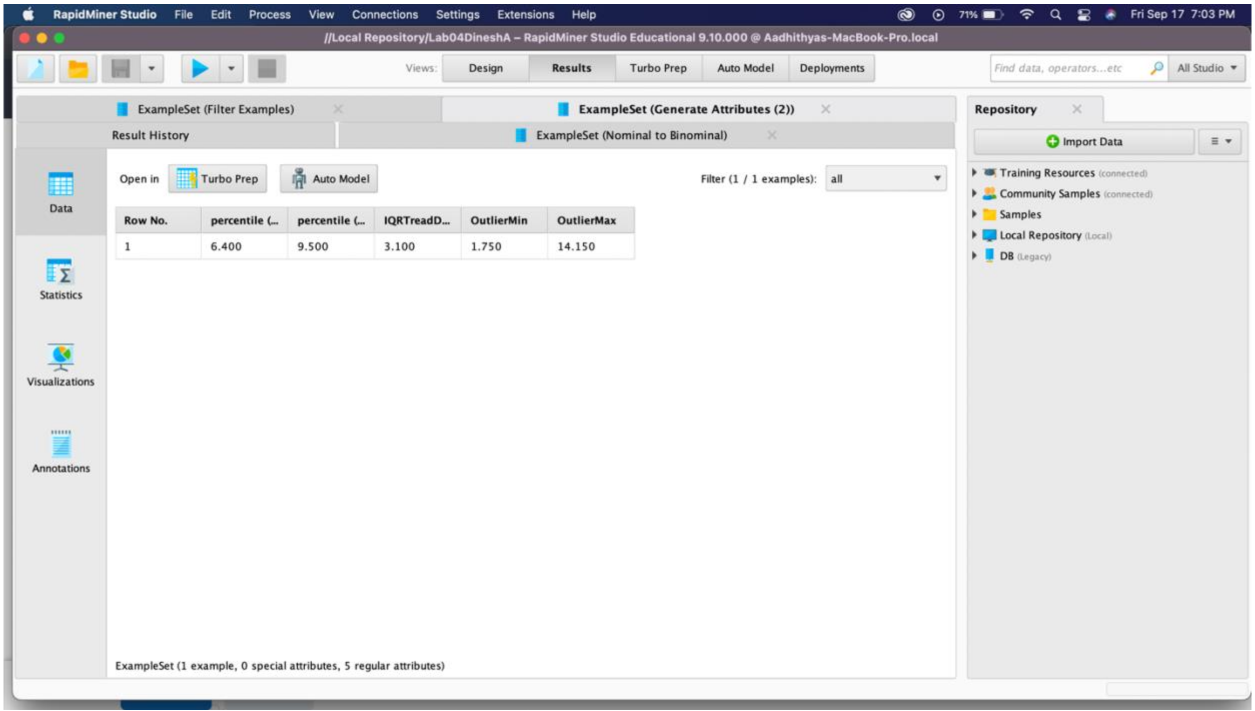
ALT + click to disconnect.



Filter (18 / 18 examples): all

Row No.	UsageMont...	TireID	Position	TreadDepth	Miles
1	69	10395904	RF	0.600	64742
2	69	10395904	LR	0.600	64742
3	39.600	14922617	LF	0	12346
4	73.300	3574739	RR	0.200	57313
5	64.100	21817569	RR	1.300	22618
6	66.500	16691383	LF	1.100	42632
7	29.300	17123211	RR	16.700	24517
8	23.690	3574739	RF	0.200	57313
9	69	10395904	RR	0.600	64742
10	66.500	16691383	LR	1	42632
11	66.500	16691383	RR	1	42632
12	69	10395904	LF	0.600	64742
13	64.100	21817569	RF	1.400	22618
14	64.100	21817569	LR	1.300	22618
15	66.500	16691383	RF	1	42632
16	73.300	3574739	LF	0.200	57313

ExampleSet (18 examples, 0 special attributes, 5 regular attributes)

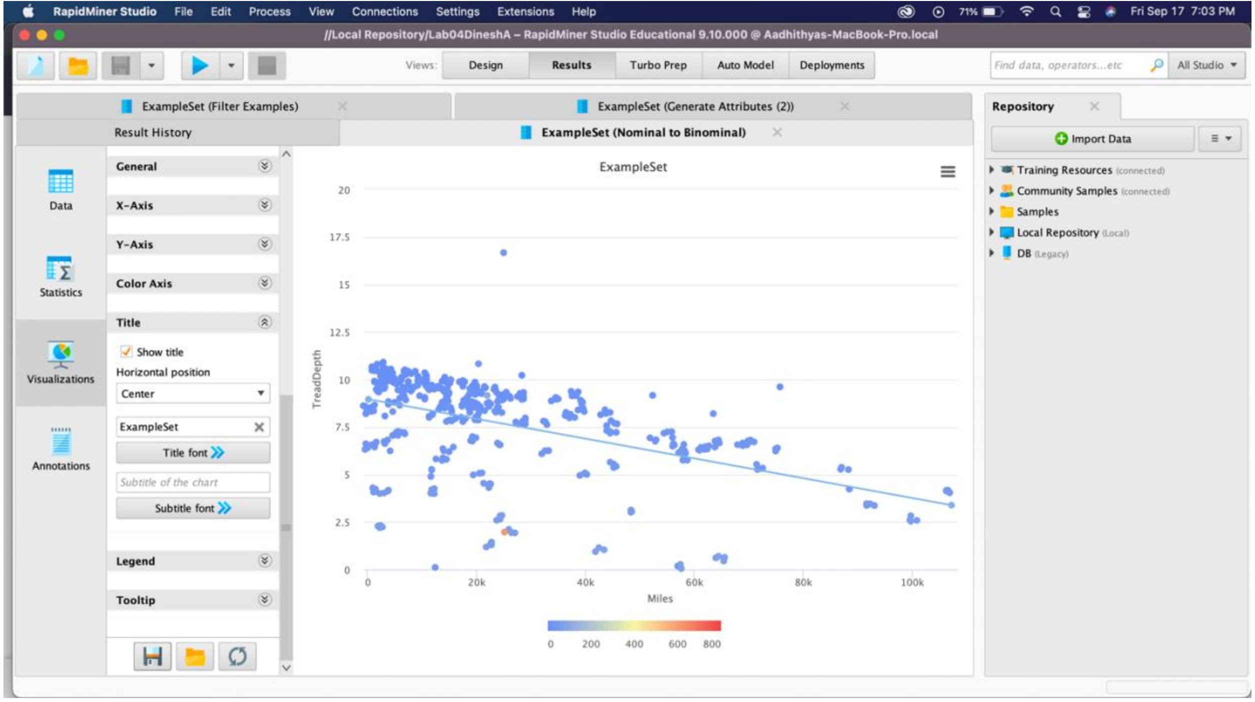


RapidMiner Studio File Edit Process View Connections Settings Extensions Help
Fri Sep 17 7:03 PM
//Local Repository/Lab04DineshA – RapidMiner Studio Educational 9.10.000 @ Aadhithyas-MacBook-Pro.local
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators...etc
All Studio
ExampleSet (Filter Examples)
ExampleSet (Generate Attributes (2))
Result History
ExampleSet (Nominal to Binominal)
Data
Statistics
Visualizations
Annotations
Open in Turbo Prep Auto Model
Repository
Import Data
Training Resources (connected)
Community Samples (connected)
Samples
Local Repository (Local)
DB (Legacy)

Filter (1 / 1 examples): all

Row No.	percentile (...	percentile (...	IQRTreadD...	OutlierMin	OutlierMax
1	6.400	9.500	3.100	1.750	14.150

ExampleSet (1 example, 0 special attributes, 5 regular attributes)



RapidMiner Studio File Edit Process View Connections Settings Extensions Help
Fri Sep 17 7:03 PM
//Local Repository/Lab04DineshA – RapidMiner Studio Educational 9.10.000 @ Aadhithyas-MacBook-Pro.local
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators...etc
All Studio
ExampleSet (Filter Examples)
ExampleSet (Generate Attributes (2))
Result History
ExampleSet (Nominal to Binominal)
Data
Statistics
Visualizations
Annotations
General
X-Axis
Y-Axis
Color Axis
Title
Legend
Tooltip
Title font
Subtitle of the chart
Subtitle font
ExampleSet
20
17.5
15
12.5
10
7.5
5
2.5
0
TreadDepth
0
20k
40k
60k
80k
100k
Miles
0
200
400
600
800
Repository
Import Data
Training Resources (connected)
Community Samples (connected)
Samples
Local Repository (Local)
DB (Legacy)

☑ Show title

Horizontal position: Center

ExampleSet

Yes, a correlation exists as the data points follow a linear regression model except for the outliers.