Sayak Dey 19BOE10072

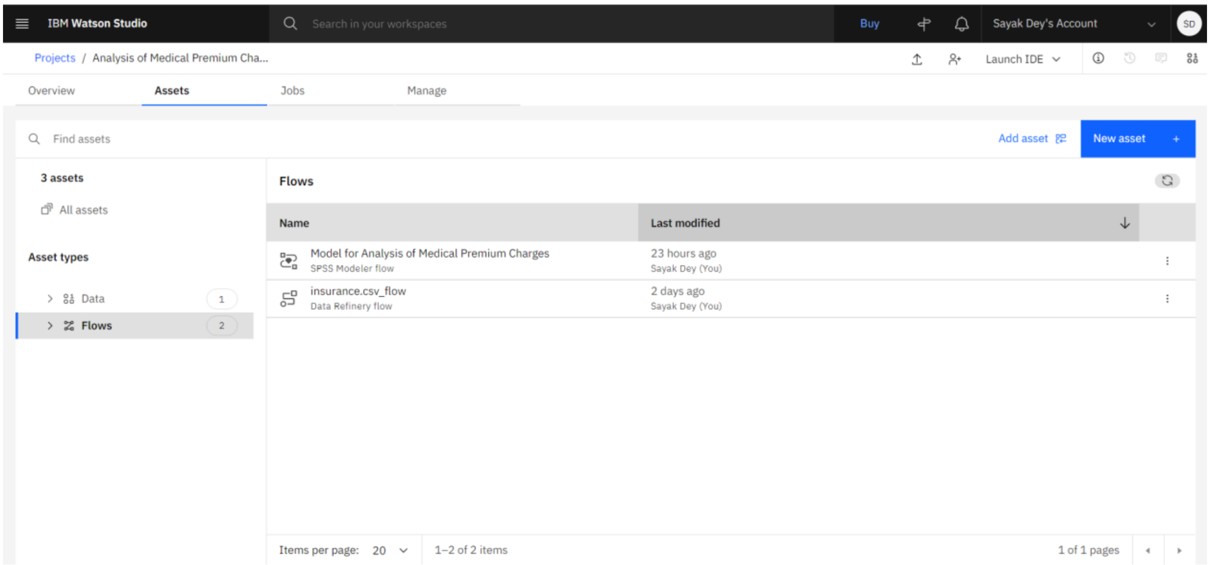
VIT Bhopal

SmartInternz

Assignment-4

Dataset : Insurance.csv

Overview:



IBM Watson Studio

Search in your workspaces

Buy

Sayak Dey's Account

Projects / Analysis of Medical Premium Cha...

Launch IDE

Overview Assets Jobs Manage

Find assets

Add asset

New asset +

3 assets

All assets

Asset types

Data 1

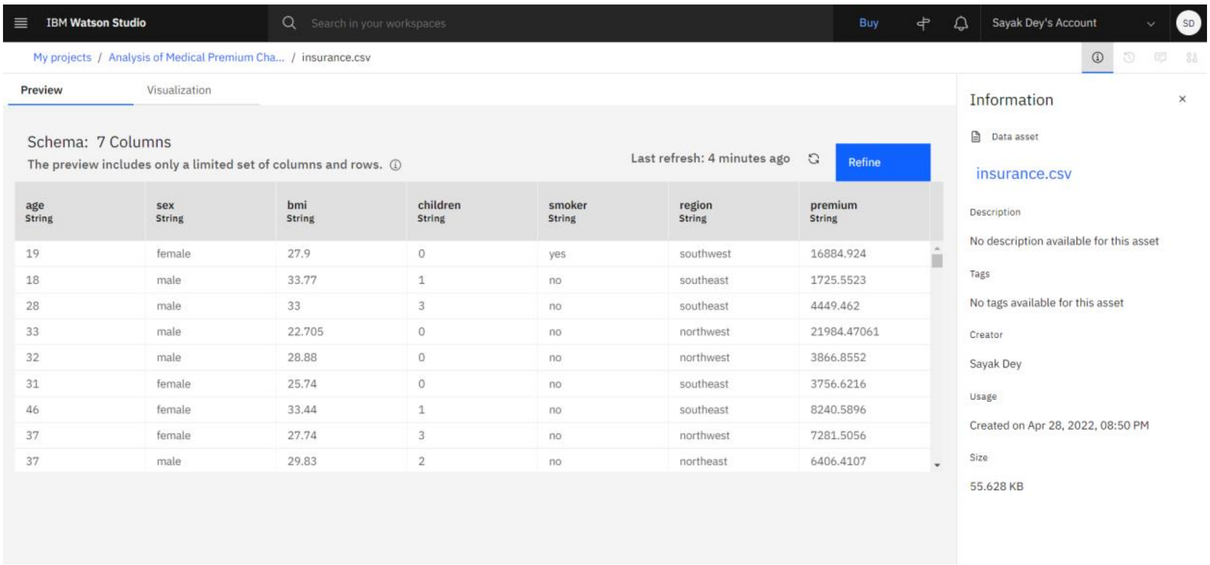
Flows 2

Flows

Name	Last modified
Model for Analysis of Medical Premium Charges SPSS Modeler flow	23 hours ago Sayak Dey (You)
insurance.csv_flow Data Refinery flow	2 days ago Sayak Dey (You)

Items per page: 20 1–2 of 2 items 1 of 1 pages

Uploaded Dataset:



IBM Watson Studio

Search in your workspaces

Buy

Sayak Dey's Account

My projects / Analysis of Medical Premium Cha... / insurance.csv

Preview Visualization

Schema: 7 Columns

The preview includes only a limited set of columns and rows.

Last refresh: 4 minutes ago

Refine

age String	sex String	bmi String	children String	smoker String	region String	premium String
19	female	27.9	0	yes	southwest	16884.924
18	male	33.77	1	no	southeast	1725.5523
28	male	33	3	no	southeast	4449.462
33	male	22.705	0	no	northwest	21984.47061
32	male	28.88	0	no	northwest	3866.8552
31	female	25.74	0	no	southeast	3756.6216
46	female	33.44	1	no	southeast	8240.5896
37	female	27.74	3	no	northwest	7281.5056
37	male	29.83	2	no	northeast	6406.4107

Information

Data asset

insurance.csv

Description

No description available for this asset

Tags

No tags available for this asset

Creator

Sayak Dey

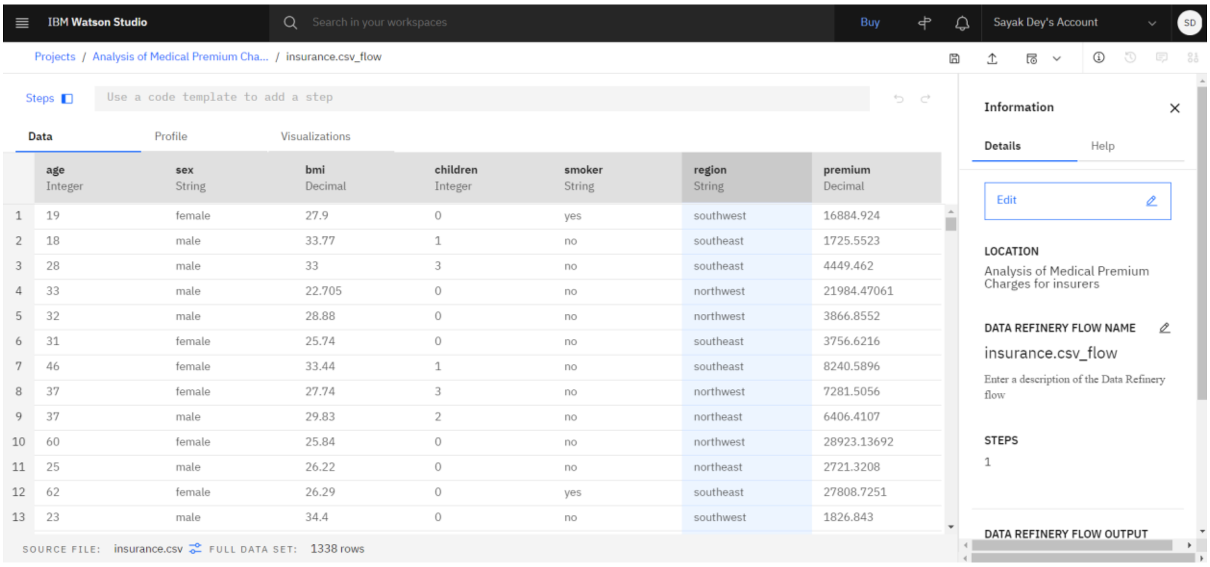
Usage

Created on Apr 28, 2022, 08:50 PM

Size

55.628 KB

Data refinery Flow:


IBM Watson Studio
Search in your workspaces
Buy
Sayak Dey's Account
SD
Projects / Analysis of Medical Premium Cha... / insurance.csv_flow
Steps
Use a code template to add a step
Data
Profile
Visualizations
Information
Details
Help
Edit

	age Integer	sex String	bmi Decimal	children Integer	smoker String	region String	premium Decimal
1	19	female	27.9	0	yes	southwest	16884.924
2	18	male	33.77	1	no	southeast	1725.5523
3	28	male	33	3	no	southeast	4449.462
4	33	male	22.705	0	no	northwest	21984.47061
5	32	male	28.88	0	no	northwest	3866.8552
6	31	female	25.74	0	no	southeast	3756.6216
7	46	female	33.44	1	no	southeast	8240.5896
8	37	female	27.74	3	no	northwest	7281.5056
9	37	male	29.83	2	no	northeast	6406.4107
10	60	female	25.84	0	no	northwest	28923.13692
11	25	male	26.22	0	no	northeast	2721.3208
12	62	female	26.29	0	yes	southeast	27808.7251
13	23	male	34.4	0	no	southwest	1826.843

SOURCE FILE: insurance.csv FULL DATA SET: 1338 rows

LOCATION

Analysis of Medical Premium Charges for insurers

DATA REFINERY FLOW NAME

insurance.csv_flow

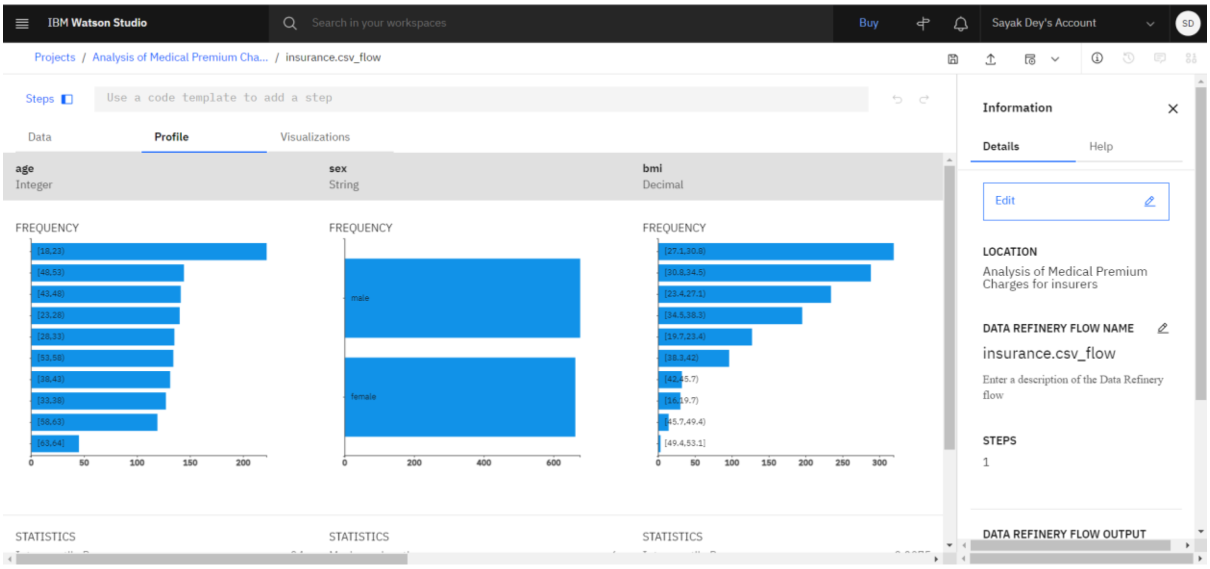
Enter a description of the Data Refinery flow

STEPS

1

DATA REFINERY FLOW OUTPUT

Data Profile:



LOCATION

Analysis of Medical Premium Charges for insurers

DATA REFINERY FLOW NAME

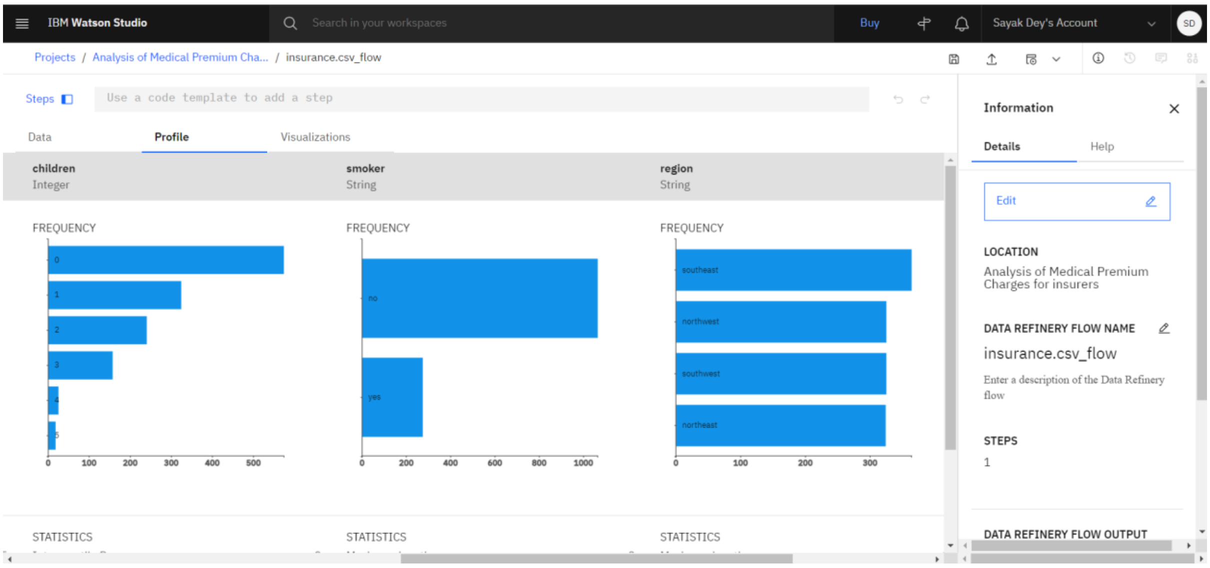
insurance.csv_flow

Enter a description of the Data Refinery flow

STEPS

1

DATA REFINERY FLOW OUTPUT



children
Integer

smoker
String

region
String

LOCATION
Analysis of Medical Premium Charges for insurers

DATA REFINERY FLOW NAME
insurance.csv_flow

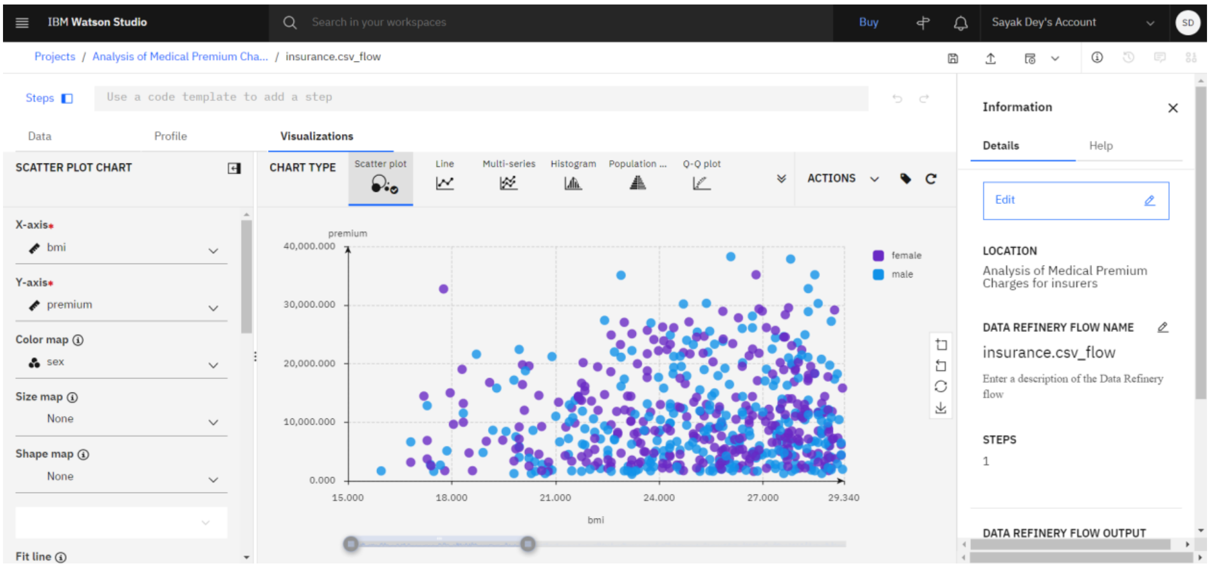
Enter a description of the Data Refinery flow

STEPS
1

DATA REFINERY FLOW OUTPUT

Data Visualization:

Bmi vs premium:



X-axis
bmi

Y-axis
premium

Color map
sex

Size map
None

Shape map
None

LOCATION
Analysis of Medical Premium Charges for insurers

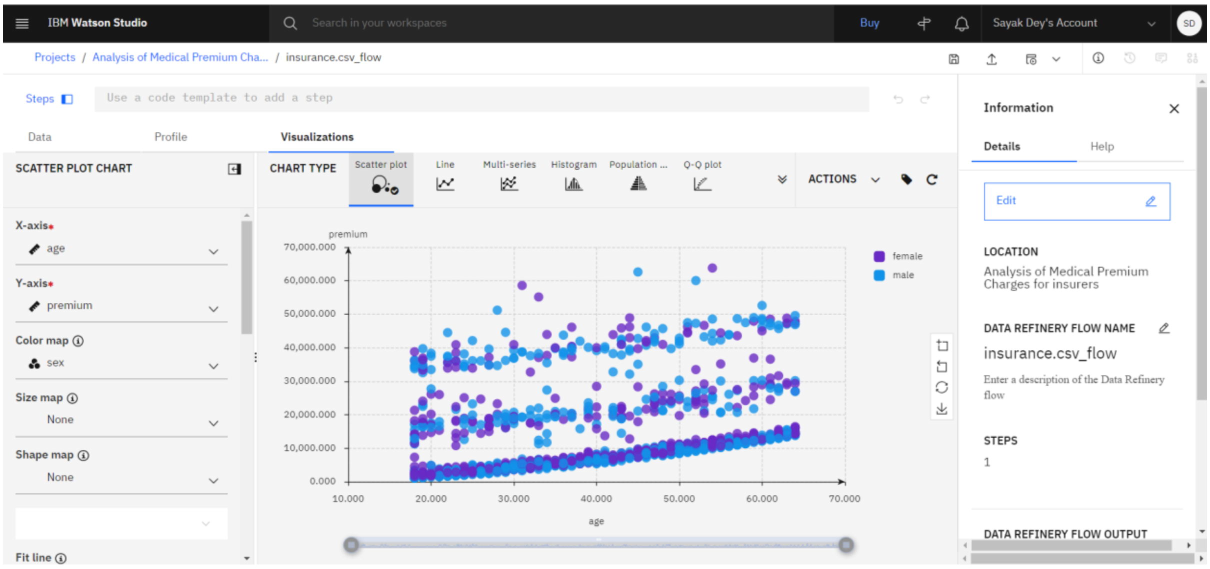
DATA REFINERY FLOW NAME
insurance.csv_flow

Enter a description of the Data Refinery flow

STEPS
1

DATA REFINERY FLOW OUTPUT

Age vs premium:



SCATTER PLOT CHART

X-axis• age

Y-axis• premium

Color map ⓘ sex

Size map ⓘ None

Shape map ⓘ None

Fit line ⓘ

LOCATION

Analysis of Medical Premium Charges for insurers

DATA REFINERY FLOW NAME

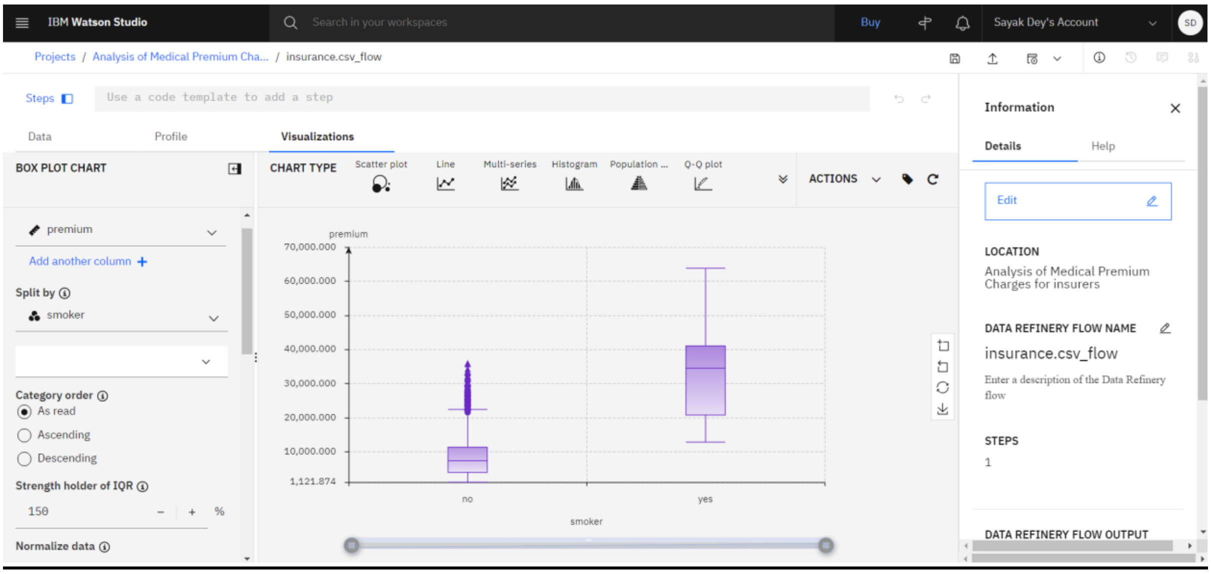
insurance.csv_flow

Enter a description of the Data Refinery flow

STEPS

1

Smoker vs premium:



BOX PLOT CHART

premium

Add another column +

Split by ⓘ smoker

Category order ⓘ

● As read

○ Ascending

○ Descending

Strength holder of IQR ⓘ 150 − + %

Normalize data ⓘ

LOCATION

Analysis of Medical Premium Charges for insurers

DATA REFINERY FLOW NAME

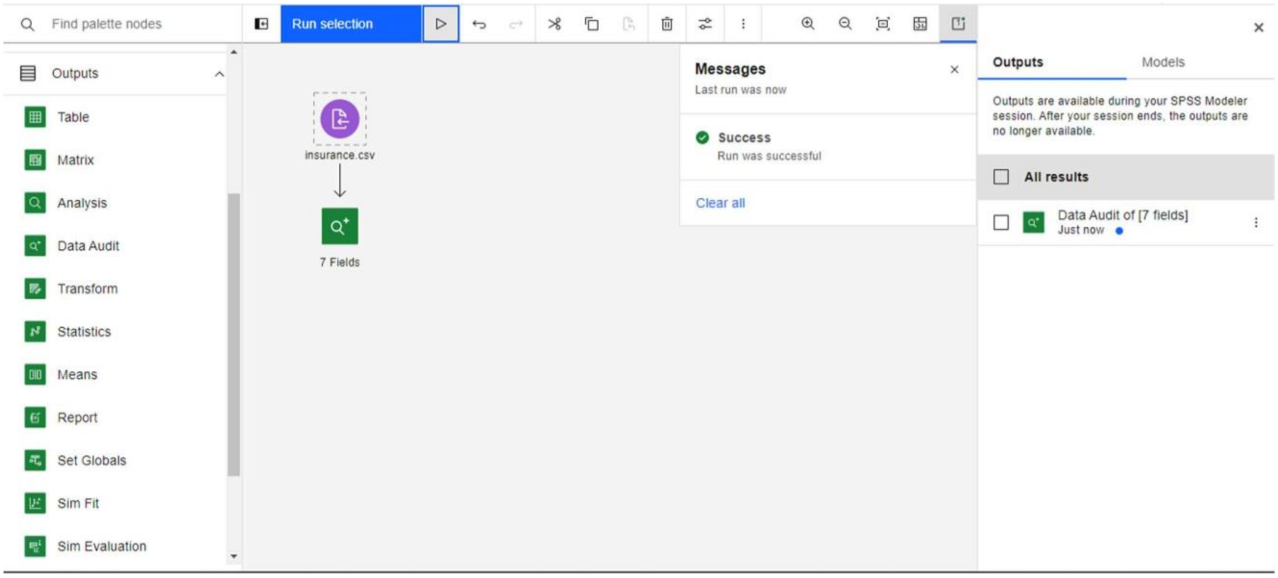
insurance.csv_flow

Enter a description of the Data Refinery flow

STEPS

1

Data Audit:

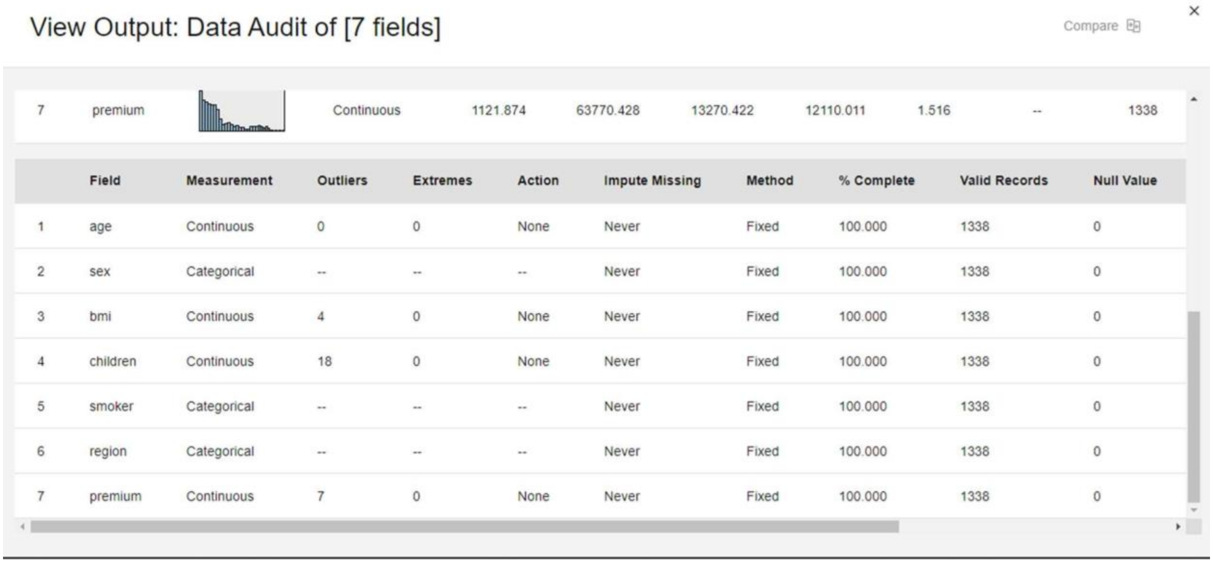


Messages
Last run was now

Success
Run was successful

Outputs are available during your SPSS Modeler session. After your session ends, the outputs are no longer available.

□ Data Audit of [7 fields]
Just now

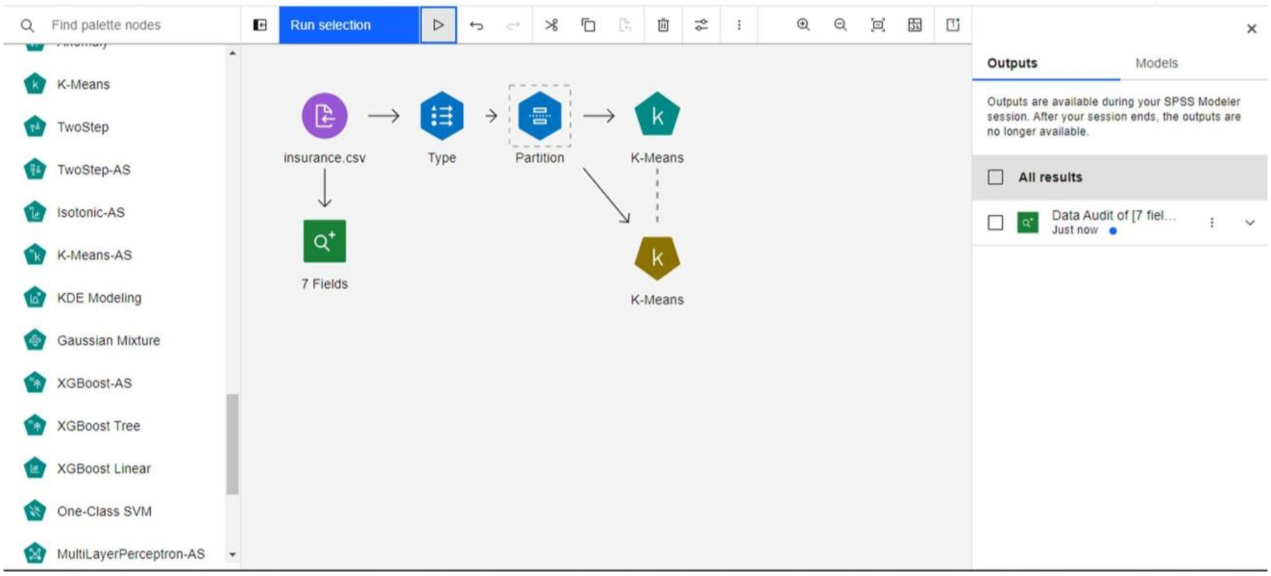


View Output: Data Audit of [7 fields]

7	premium		Continuous	1121.874	63770.428	13270.422	12110.011	1.516	--	1338

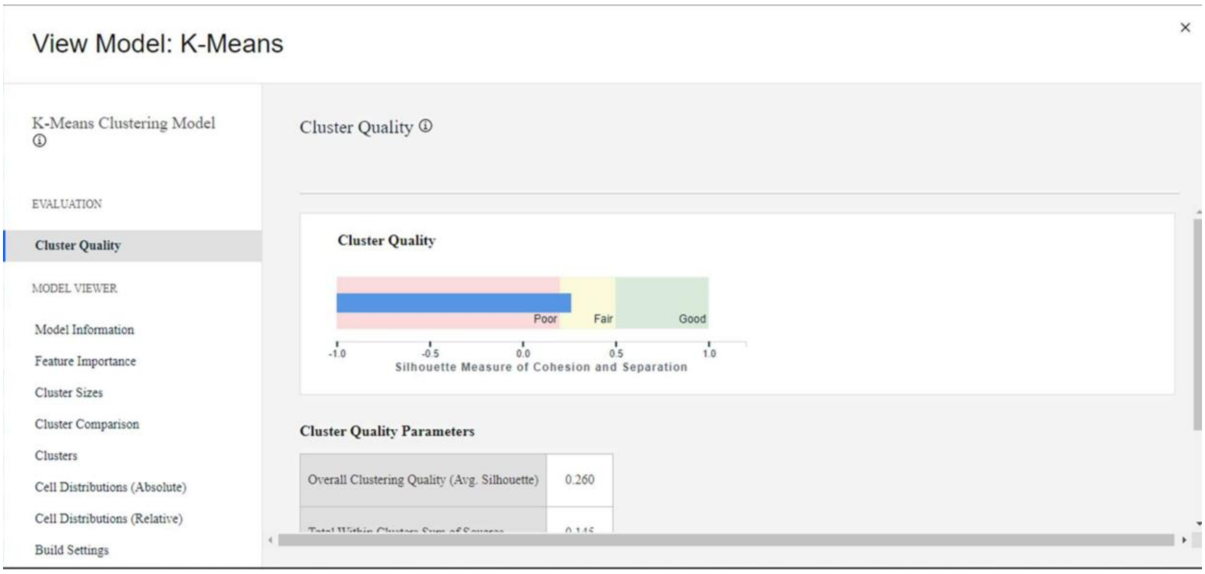
	Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value
1	age	Continuous	0	0	None	Never	Fixed	100.000	1338	0
2	sex	Categorical	--	--	--	Never	Fixed	100.000	1338	0
3	bmi	Continuous	4	0	None	Never	Fixed	100.000	1338	0
4	children	Continuous	18	0	None	Never	Fixed	100.000	1338	0
5	smoker	Categorical	--	--	--	Never	Fixed	100.000	1338	0
6	region	Categorical	--	--	--	Never	Fixed	100.000	1338	0
7	premium	Continuous	7	0	None	Never	Fixed	100.000	1338	0

Building K-Means Clustering Model:



Outputs are available during your SPSS Modeler session. After your session ends, the outputs are no longer available.

Model Output:



View Model: K-Means

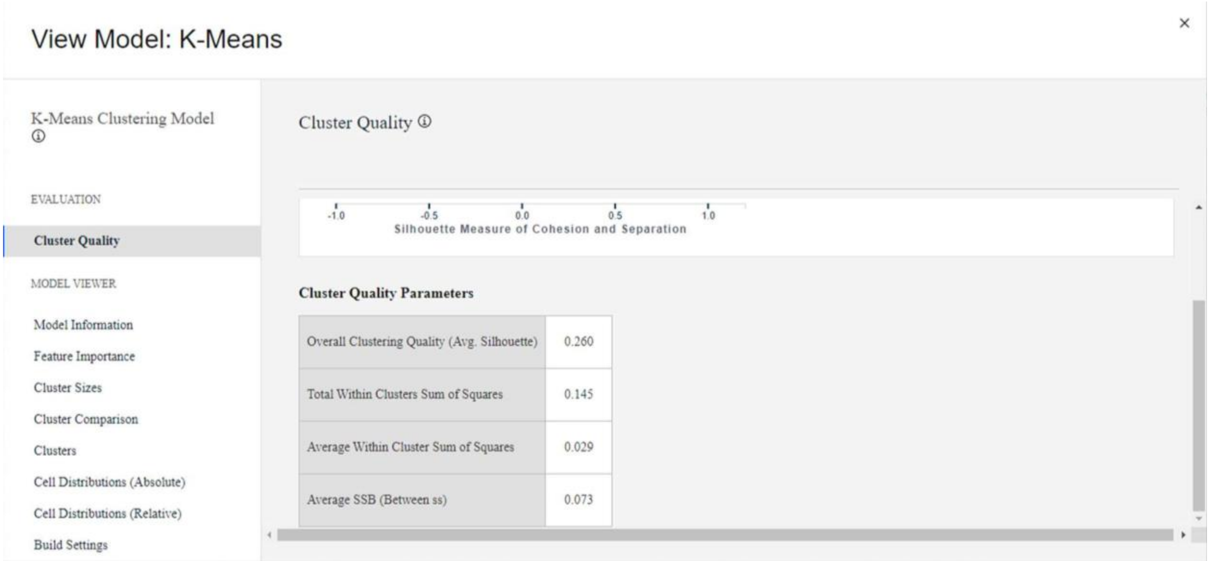
K-Means Clustering Model

Cluster Quality

Cluster Quality

Cluster Quality Parameters

Overall Clustering Quality (Avg. Silhouette)	0.260
Total Within Clusters Sum of Squares	0.145

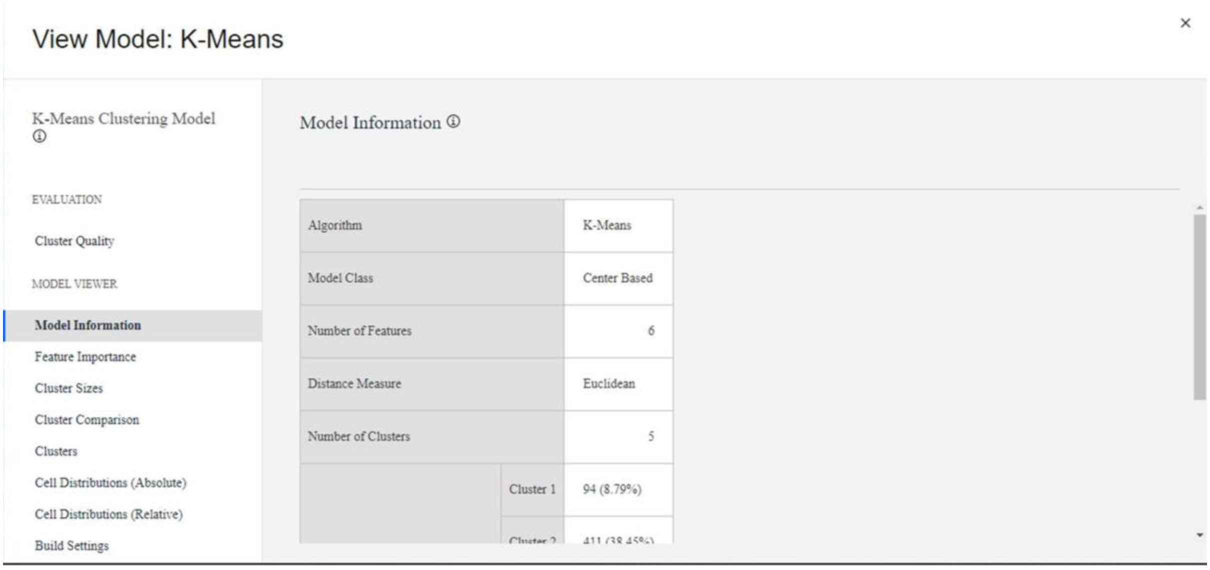


View Model: K-Means

Cluster Quality ⓘ

Cluster Quality Parameters

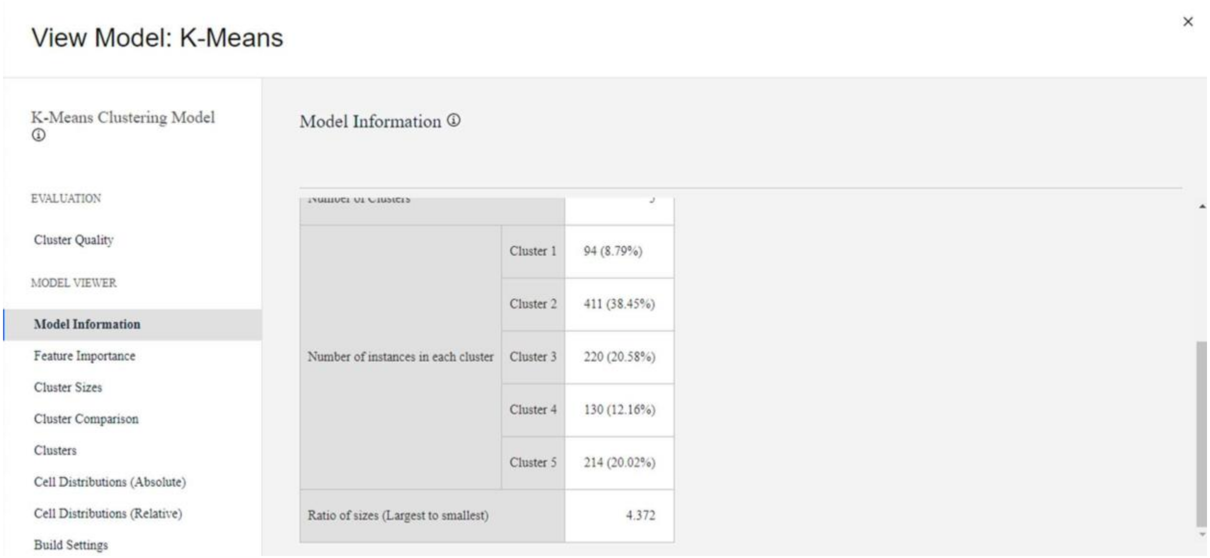
Overall Clustering Quality (Avg. Silhouette)	0.260
Total Within Clusters Sum of Squares	0.145
Average Within Cluster Sum of Squares	0.029
Average SSB (Between ss)	0.073



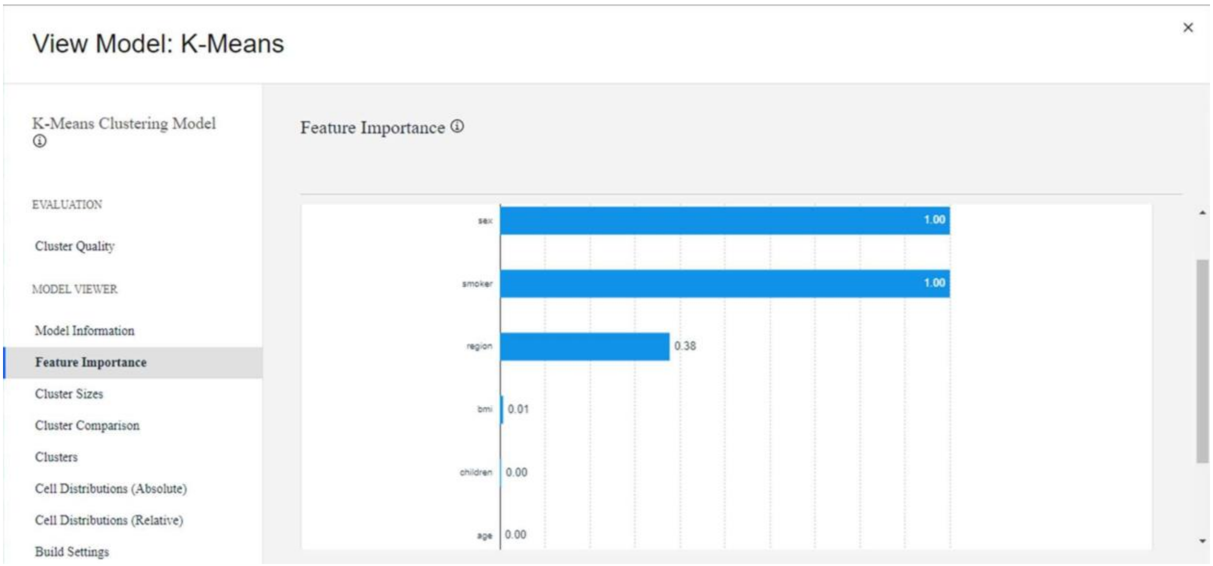
View Model: K-Means

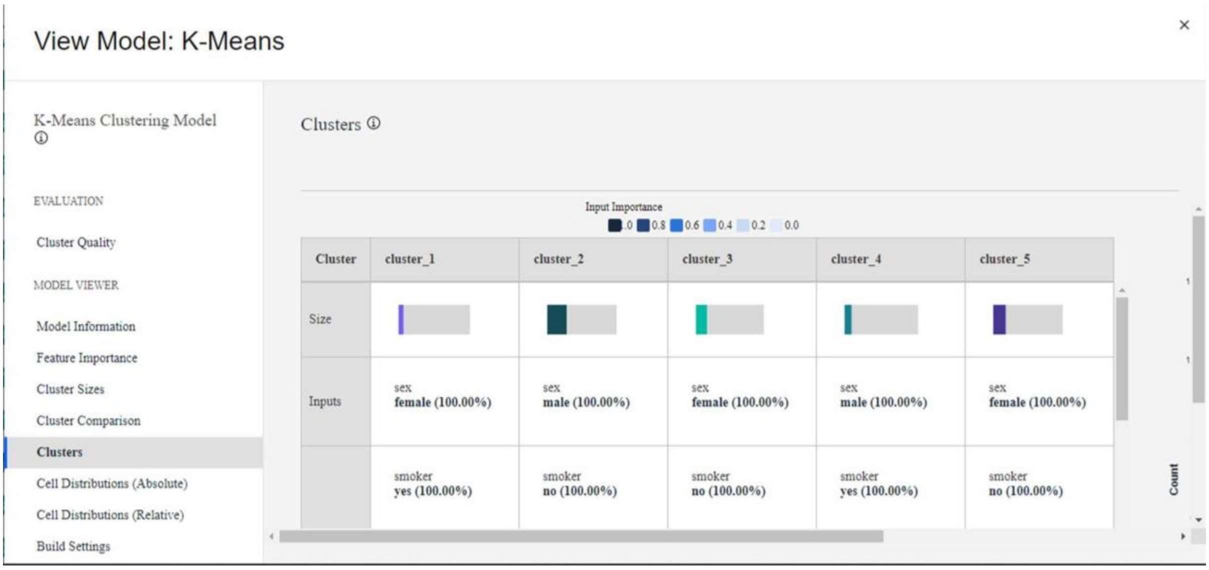
Model Information ⓘ

Algorithm		K-Means
Model Class		Center Based
Number of Features		6
Distance Measure		Euclidean
Number of Clusters		5
	Cluster 1	94 (8.79%)
	Cluster 2	411 (38.45%)



Number of Clusters		5
Number of instances in each cluster	Cluster 1	94 (8.79%)
	Cluster 2	411 (38.45%)
	Cluster 3	220 (20.58%)
	Cluster 4	130 (12.16%)
	Cluster 5	214 (20.02%)
Ratio of sizes (Largest to smallest)		4.372





View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

Cell Distributions (Absolute)

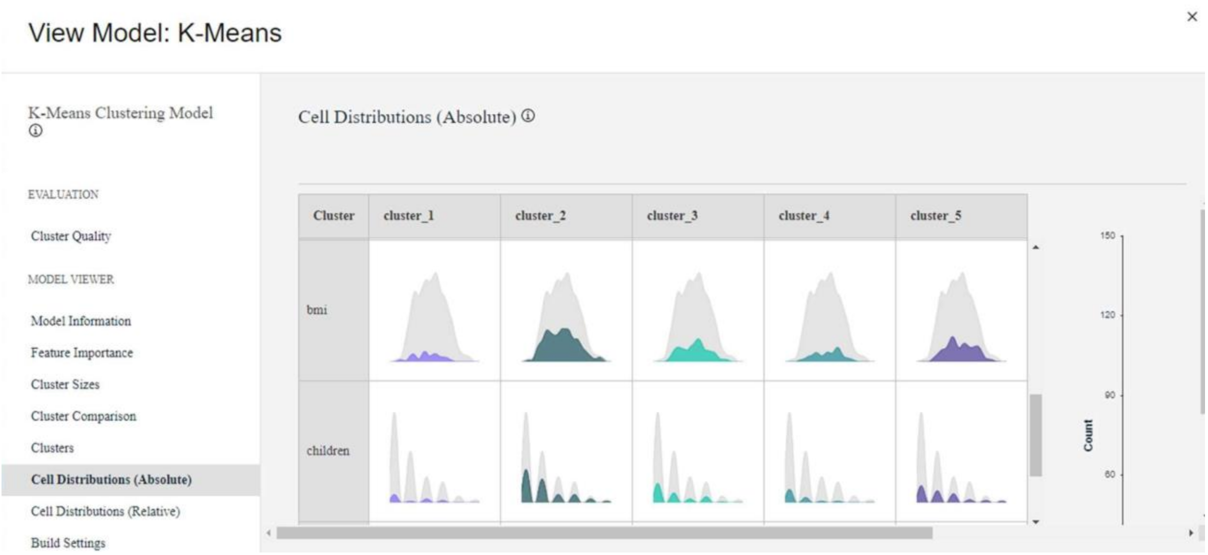
Cell Distributions (Relative)

Build Settings

Clusters ⓘ

Input Importance ■ .0 ■ 0.8 ■ 0.6 ■ 0.4 ■ 0.2 □ 0.0

Cluster	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5
Size					
Inputs	sex **female** (100.00%)	sex **male** (100.00%)	sex **female** (100.00%)	sex **male** (100.00%)	sex **female** (100.00%)
	smoker **yes** (100.00%)	smoker **no** (100.00%)	smoker **no** (100.00%)	smoker **yes** (100.00%)	smoker **no** (100.00%)



View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

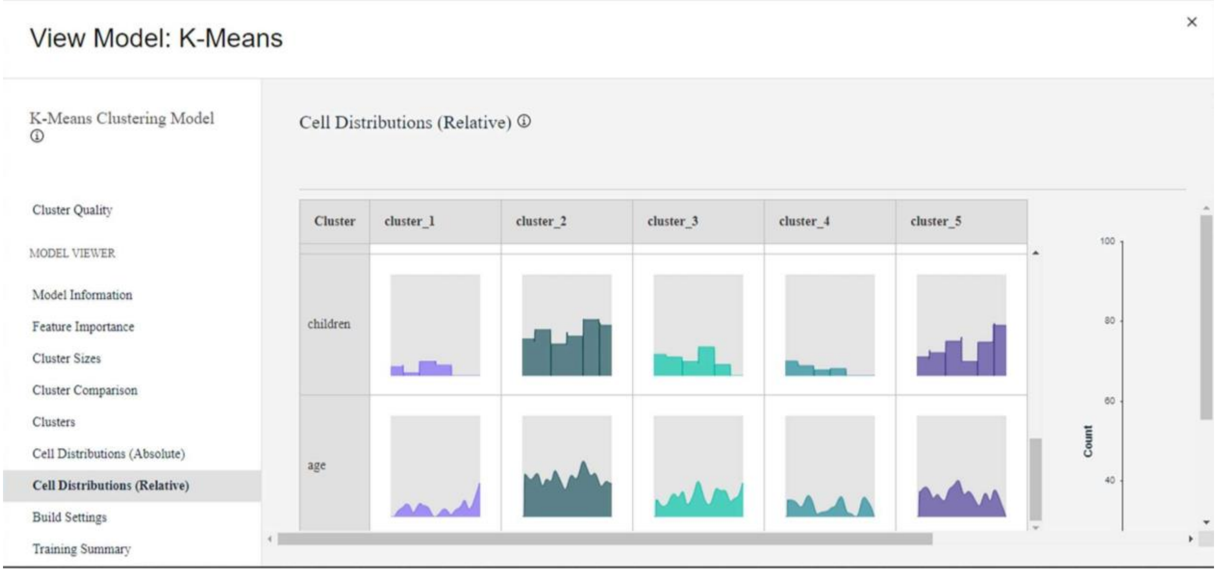
Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

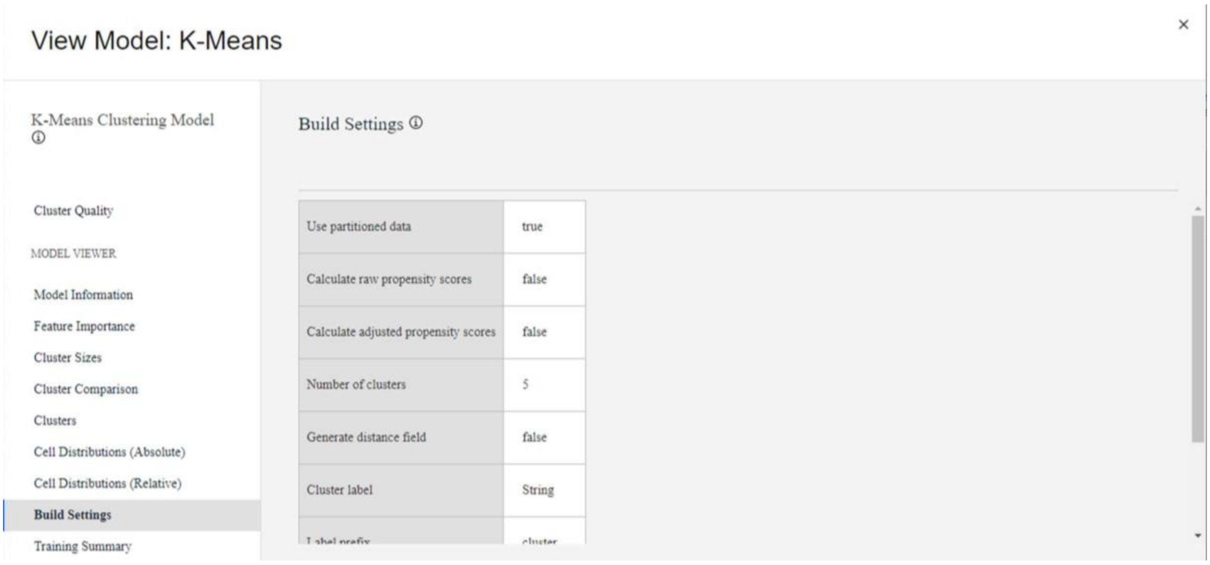
Cell Distributions (Absolute) ⓘ

Cluster	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5
bmi					
children					



View Model: K-Means

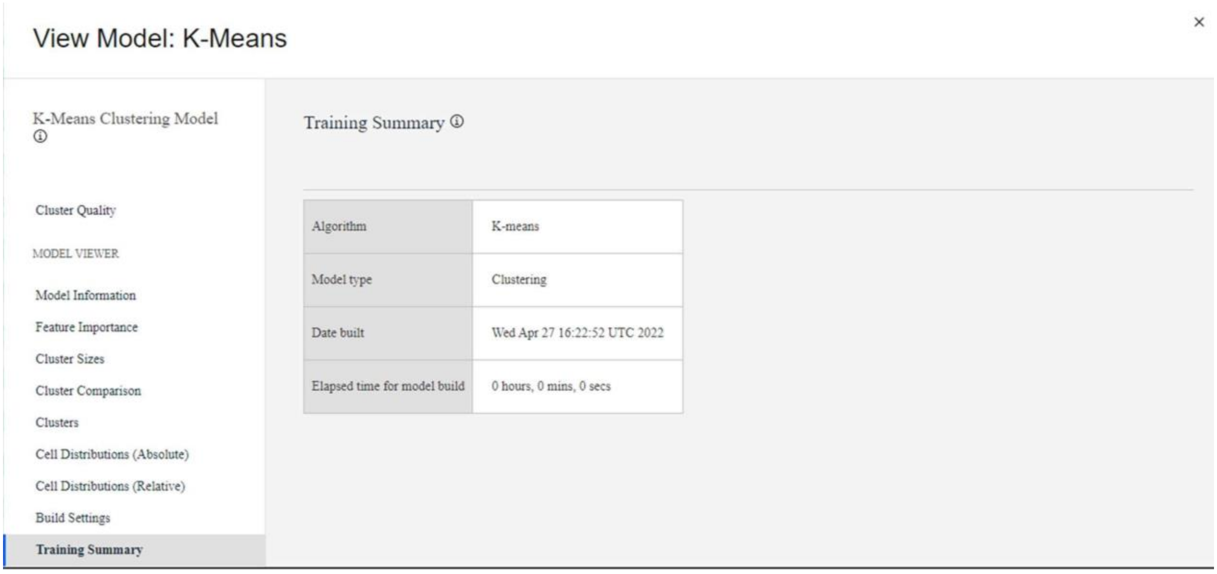
Cell Distributions (Relative) ⓘ



View Model: K-Means

Build Settings ⓘ

Use partitioned data	true
Calculate raw propensity scores	false
Calculate adjusted propensity scores	false
Number of clusters	5
Generate distance field	false
Cluster label	String
Label prefix	cluster

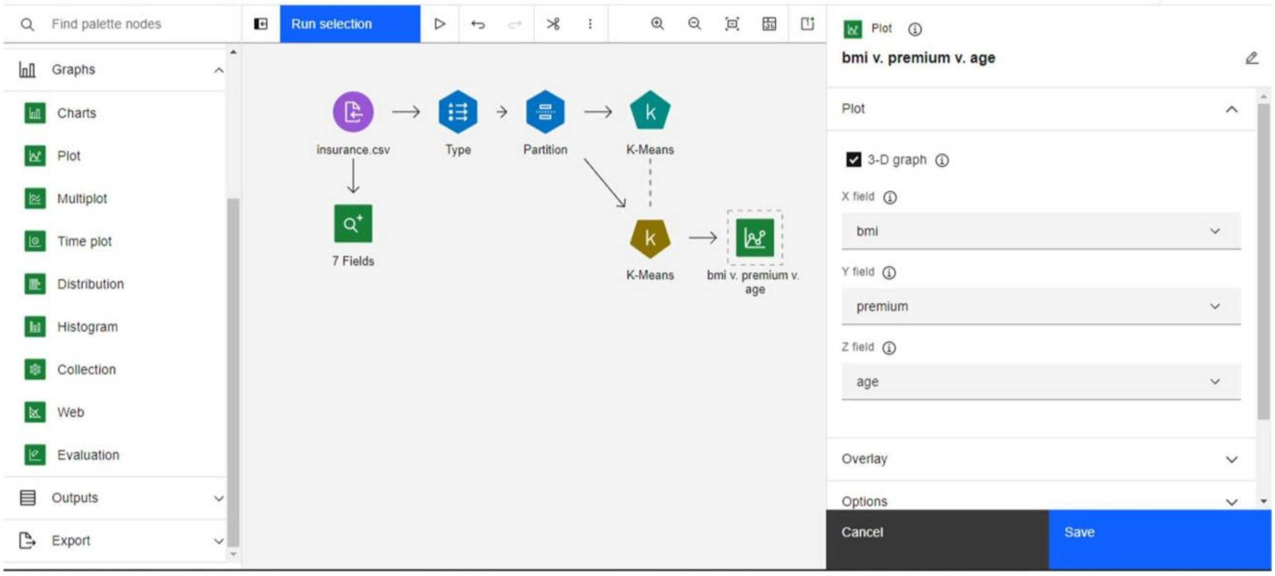


View Model: K-Means

Training Summary ⓘ

Algorithm	K-means
Model type	Clustering
Date built	Wed Apr 27 16:22:52 UTC 2022
Elapsed time for model build	0 hours, 0 mins, 0 secs

Building a Plot node with BMI vs Premium vs Age:



Plot ⓘ

bmi v. premium v. age

Plot

☑ 3-D graph ⓘ

X field ⓘ: bmi

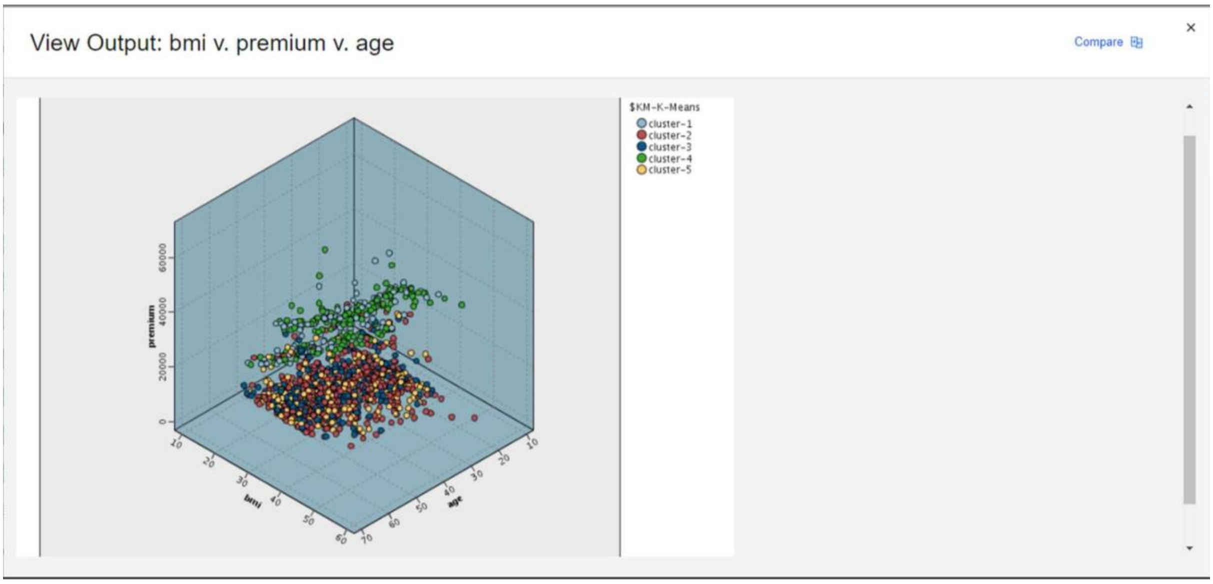
Y field ⓘ: premium

Z field ⓘ: age

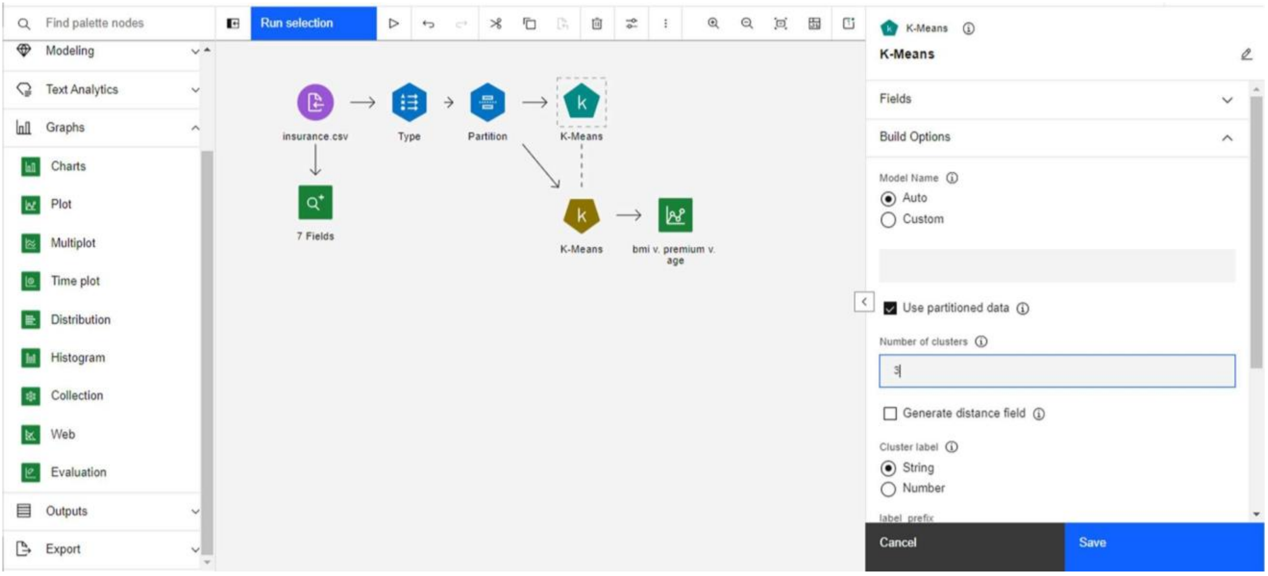
Overlay

Options

Output:


View Output: bmi v. premium v. age

Changing number of Clusters from 5 to 3:



K-Means

Fields

Build Options

Model Name
● Auto
○ Custom

☑ Use partitioned data

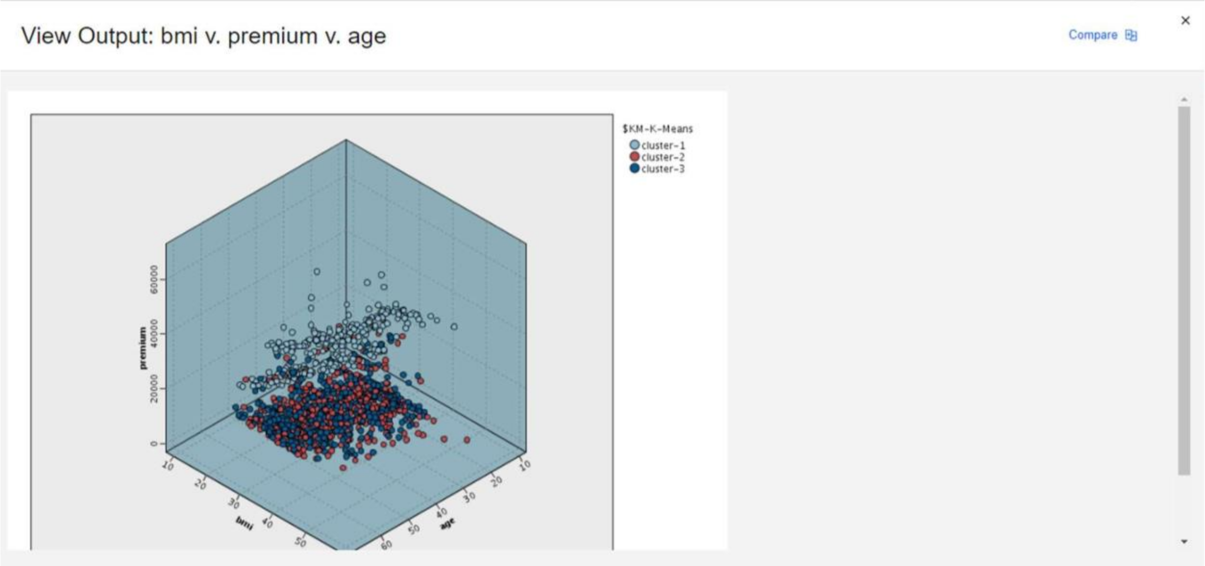
Number of clusters
3

☐ Generate distance field

Cluster label
● String
○ Number

label_prefix

Output:



View Output: bmi v. premium v. age