DATA ANALYTICS

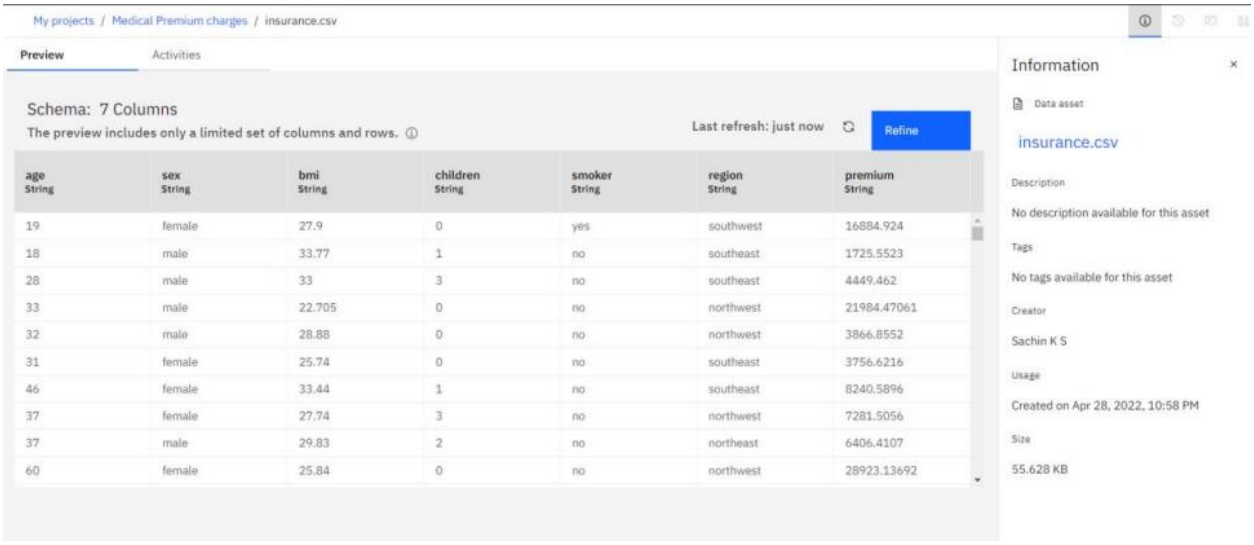
ASSIGNMENT – 4

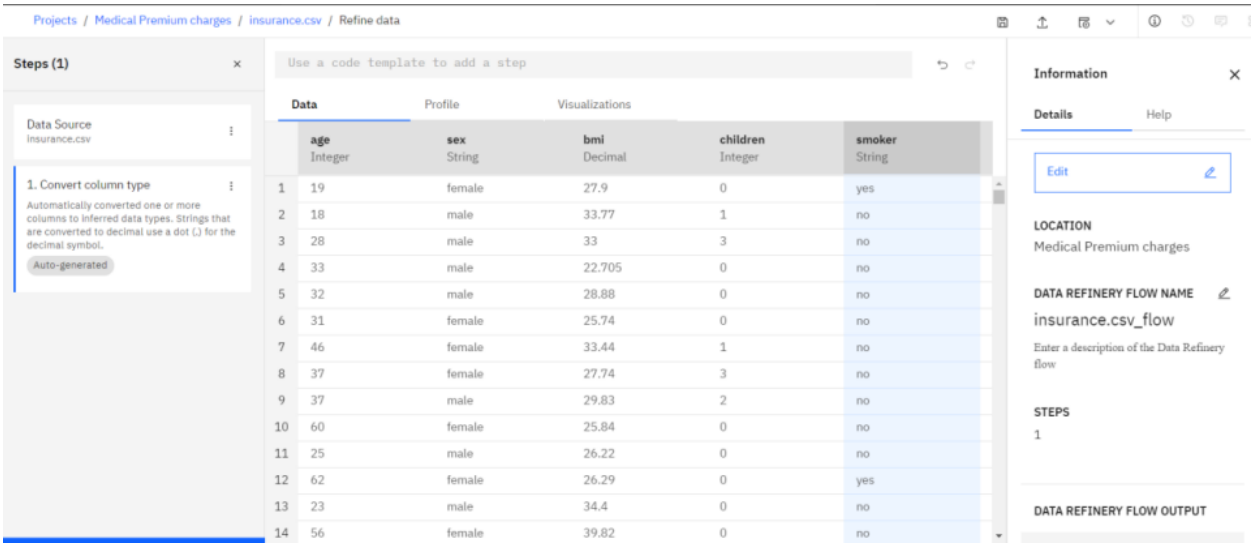
CLUSTERING MEDICAL PREMIUM CHARGES

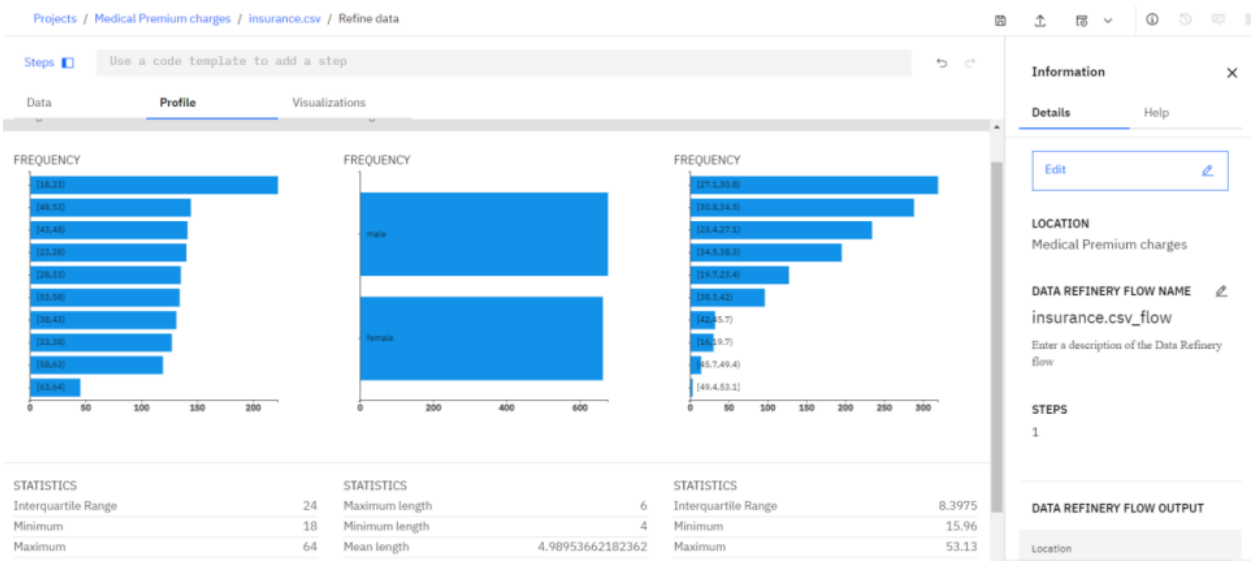
NAME : Harendhar K

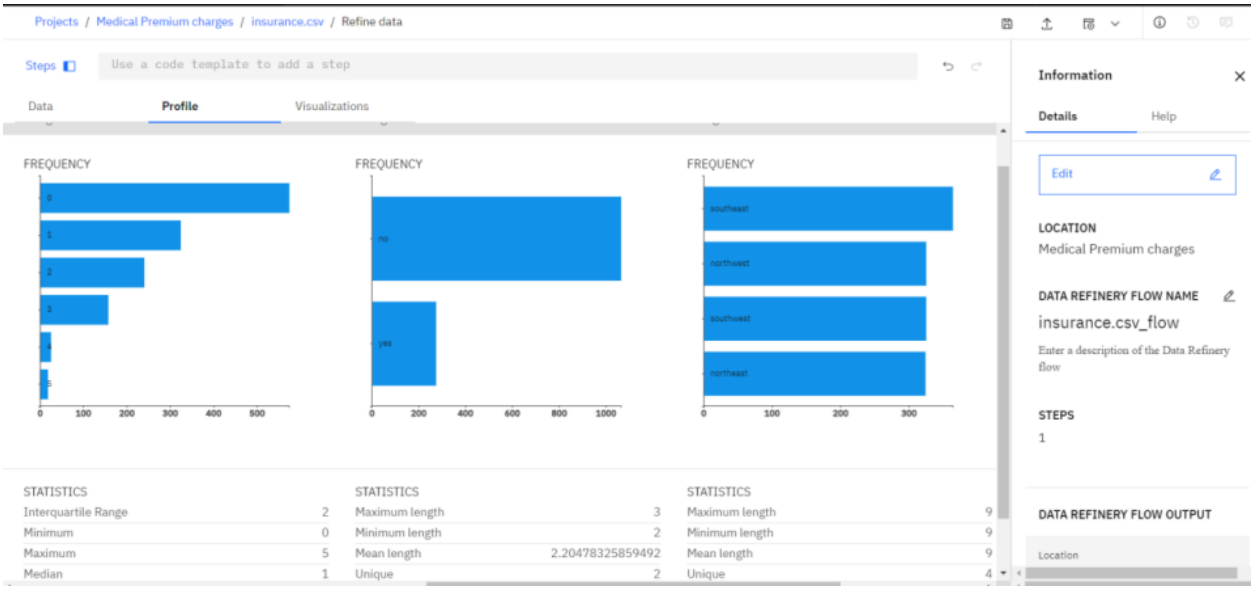
Reg no : 19BEE0140

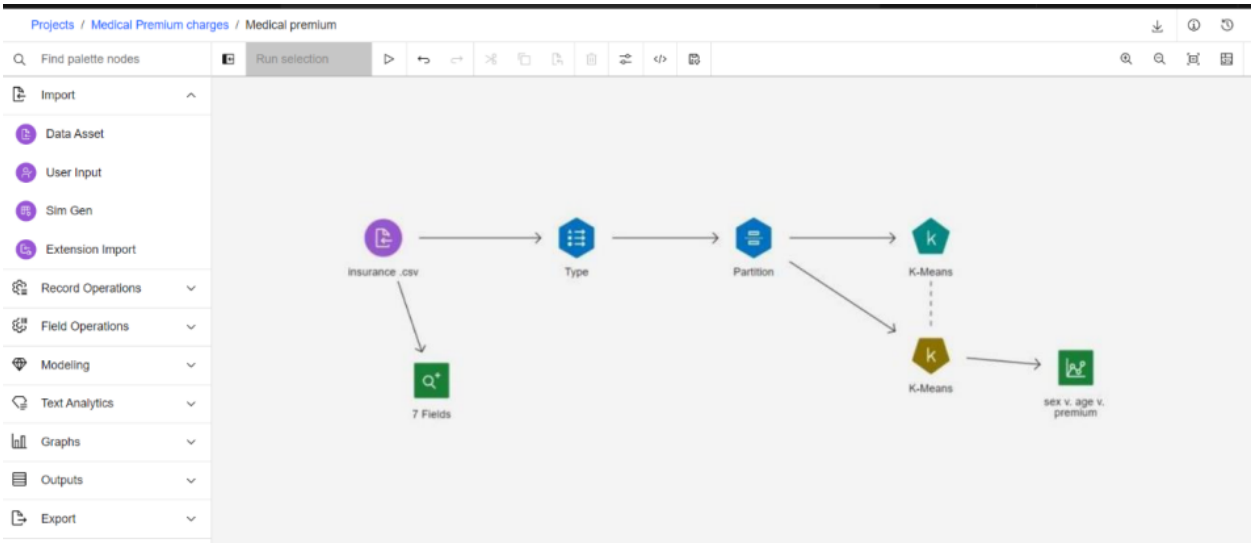
Campus: Vellore






Projects / Medical Premium charges / insurance.csv / Refine data
Steps
Use a code template to add a step
Data
Profile
Visualizations
FREQUENCY
FREQUENCY
FREQUENCY
male
female
STATISTICS
Interquartile Range 24
Minimum 18
Maximum 64
STATISTICS
Maximum length 6
Minimum length 4
Mean length 4.98953662182362
STATISTICS
Interquartile Range 8.3975
Minimum 15.96
Maximum 53.13
Information
Details
Help
Edit
LOCATION
Medical Premium charges
DATA REFINERY FLOW NAME
insurance.csv_flow
Enter a description of the Data Refinery flow
STEPS
1
DATA REFINERY FLOW OUTPUT
Location


Projects / Medical Premium charges / insurance.csv / Refine data
Steps
Use a code template to add a step
Data
Profile
Visualizations
FREQUENCY
FREQUENCY
FREQUENCY
no
yes
southeast
northwest
southwest
northeast
STATISTICS
Interquartile Range 2
Minimum 0
Maximum 5
Median 1
STATISTICS
Maximum length 3
Minimum length 2
Mean length 2.20478325859492
Unique 2
STATISTICS
Maximum length 9
Minimum length 9
Mean length 9
Unique 4
Information
Details
Help
Edit
LOCATION
Medical Premium charges
DATA REFINERY FLOW NAME
insurance.csv_flow
Enter a description of the Data Refinery flow
STEPS
1
DATA REFINERY FLOW OUTPUT
Location


Projects / Medical Premium charges / Medical premium
Find palette nodes
Run selection
Import
Data Asset
User Input
Sim Gen
Extension Import
Record Operations
Field Operations
Modeling
Text Analytics
Graphs
Outputs
Export
insurance .csv
Type
Partition
K-Means
K-Means
7 Fields
sex v. age v. premium

View Output: Data Audit of [7 fields]

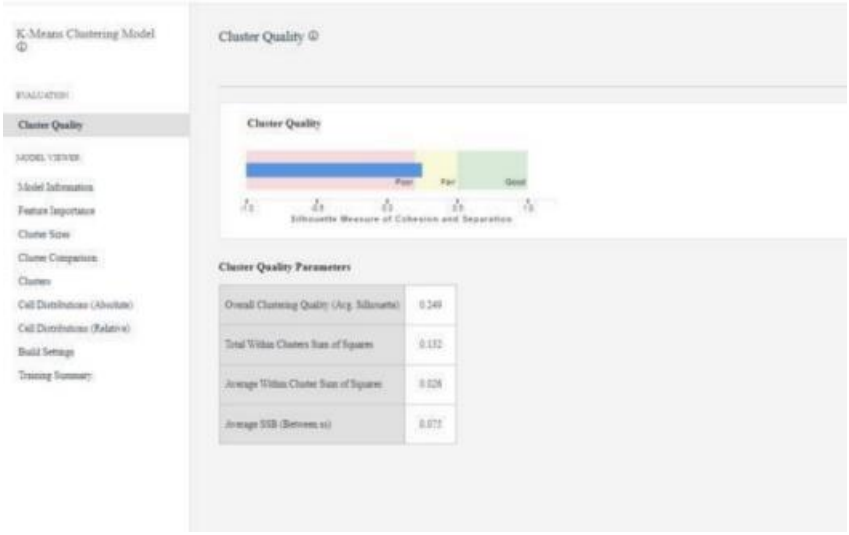
	Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	age		Continuous	18	64	39.207	14.050	0.056	--	1338
2	sex		Categorical	--	--	--	--	--	2	1338
3	bmi		Continuous	15.960	53.130	30.663	6.098	0.284	--	1338
4	children		Continuous	0	5	1.095	1.205	0.938	--	1338
5	smoker		Categorical	--	--	--	--	--	2	1338
6	region		Categorical	--	--	--	--	--	4	1338

View Output: Data Audit of [7 fields]

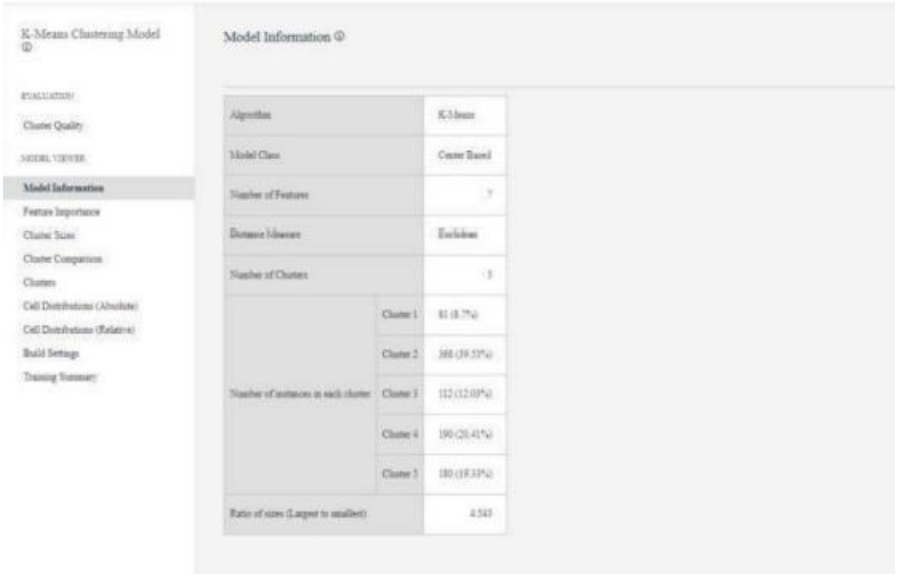
6	region		Categorical	--	--	--	--	--	4	1338
7	premium		Continuous	1121.874	63770.428	13270.422	12110.011	1.516	--	1338

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

View Model: K-Means



View Model: K-Means



View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

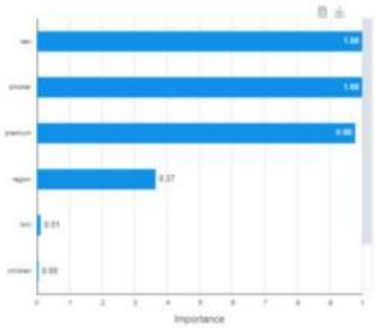
Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Feature Importance ⓘ



View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

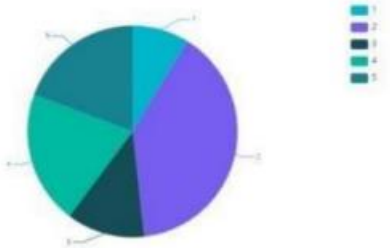
Cell Distributions (Absolute)

Cell Distributions (Relative)

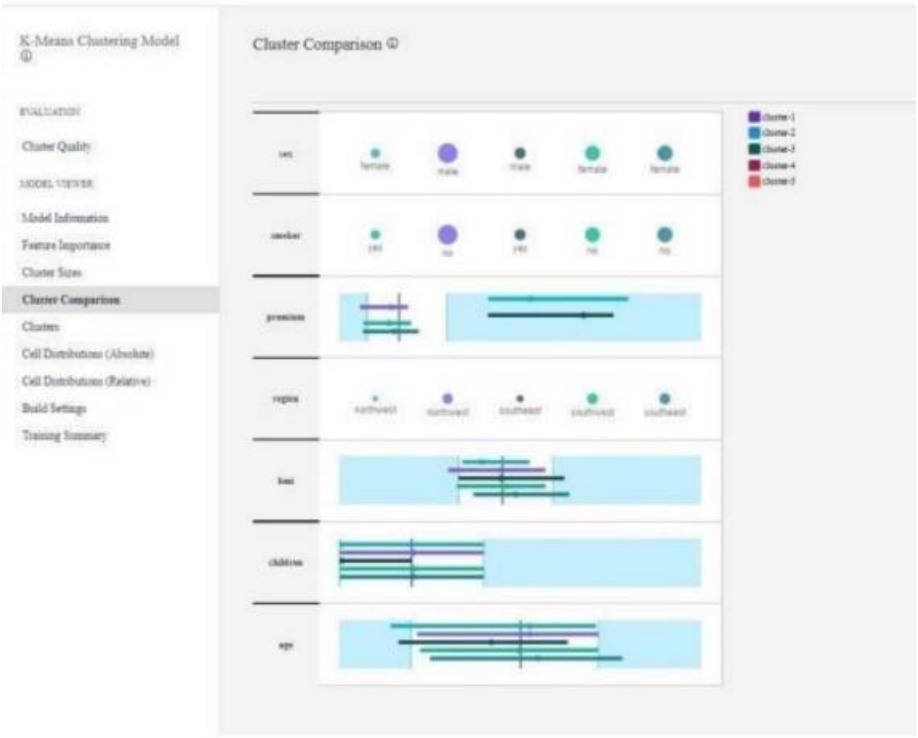
Build Settings

Training Summary

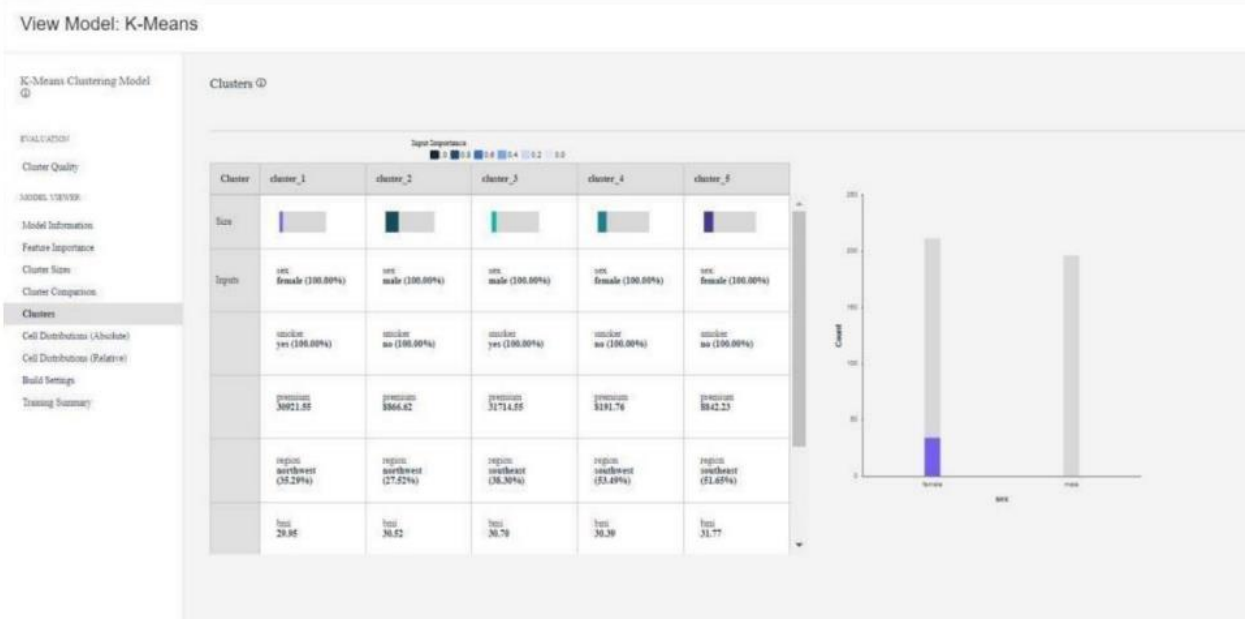
Cluster Sizes ⓘ



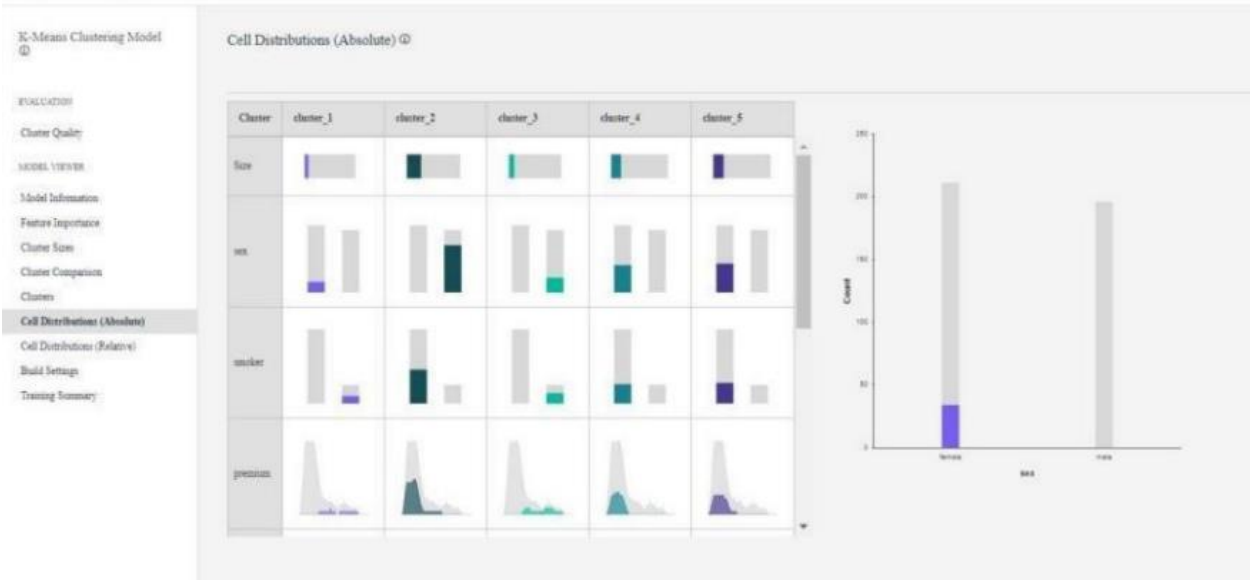
View Model: K-Means



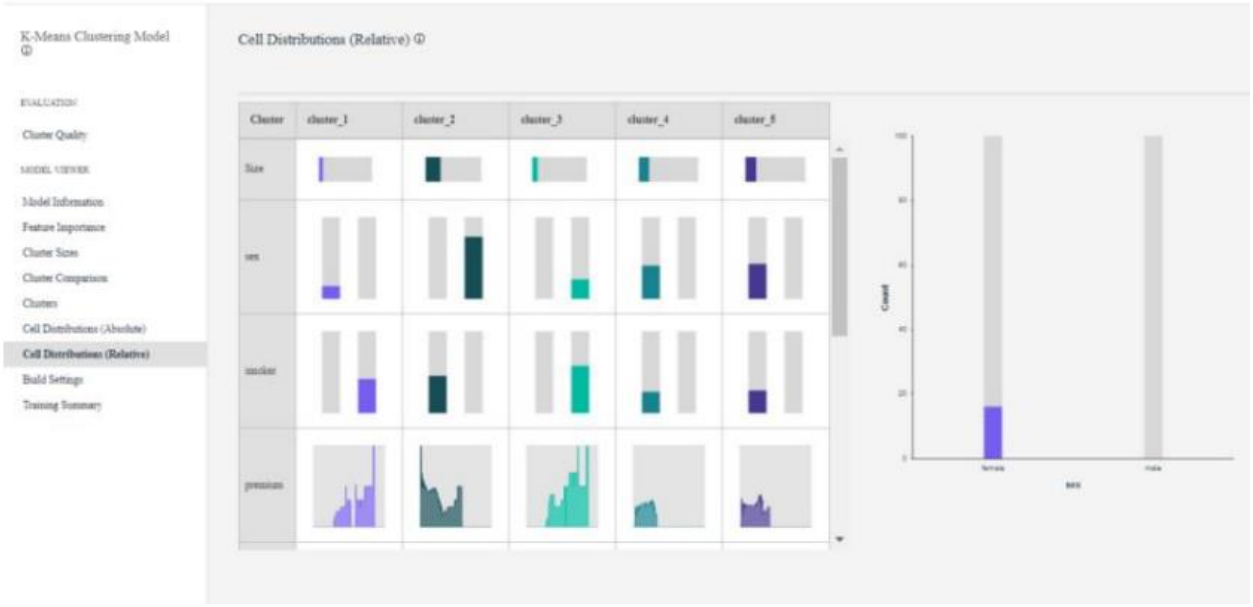
View Model: K-Means

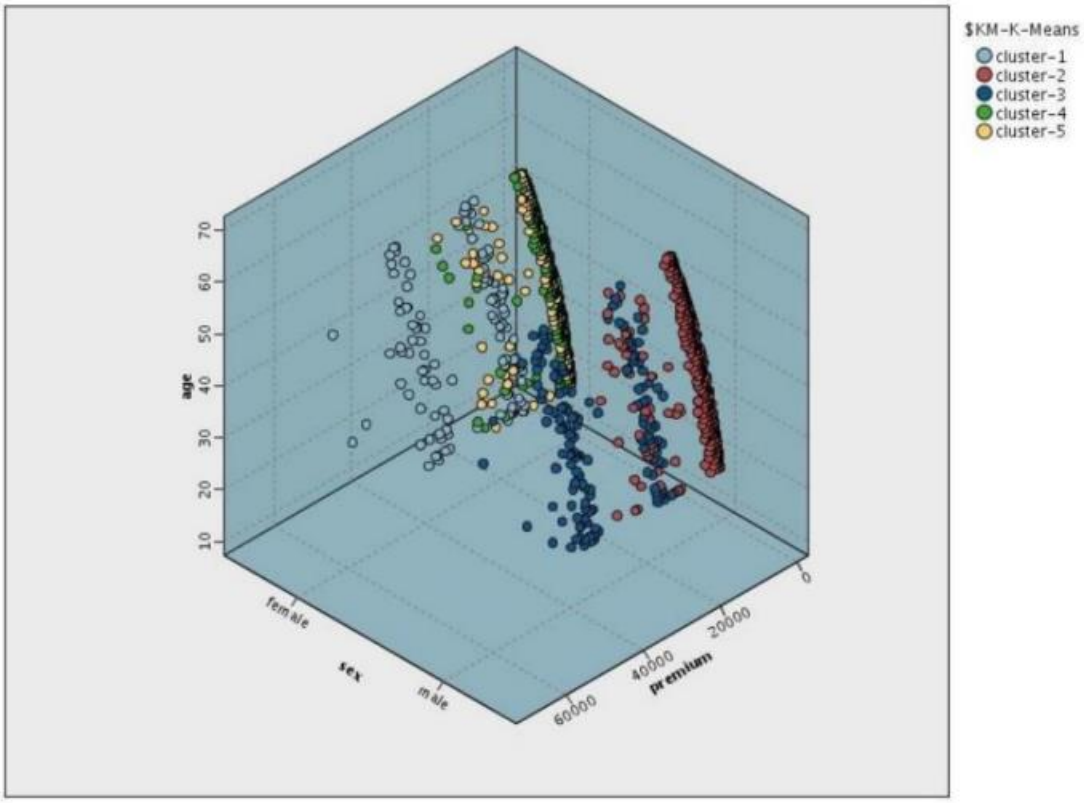


View Model: K-Means



View Model: K-Means




$KM-K-Means
cluster-1
cluster-2
cluster-3
cluster-4
cluster-5
age
70
60
50
40
30
20
10
sex
female
male
premium
60000
40000
20000
0