Preview   Activities   Visualization

Schema: 7 Columns
Preview: First 1000 rows

Last refresh: 10 seconds 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
60	female	25.84	0	no	northwest	28923.13692
25	male	26.22	0	no	northeast	2721.3208
62	female	26.29	0	yes	southeast	27808.7251
23	male	34.4	0	no	southwest	1826.843

Information ×

Data Asset

insurance .csv

Description

No description is available for this asset.

Tags

No description is available for this asset.
Added: Dec 20, 2021, 10:26 PM
Size: 55.628 KB

Steps ▣   Use a code template to add a step   ↶ ↷

Data   Profile   Visualizations

	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
ʼʼ	56	female	39.82	0	no	southeast	11090.7178

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

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

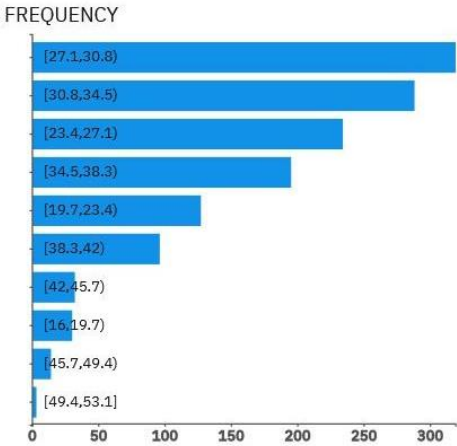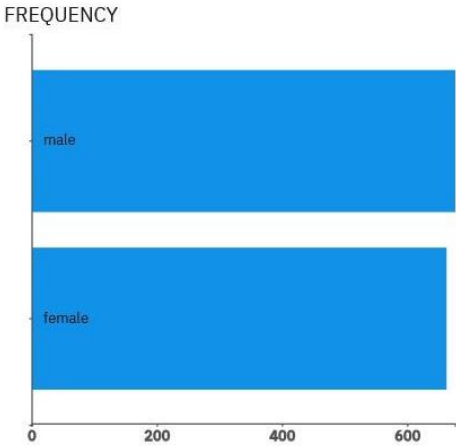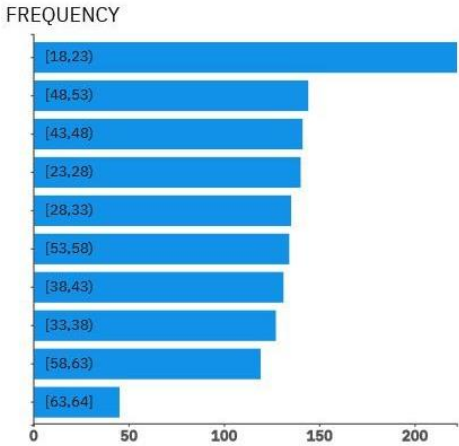Location

Steps ▣

Use a code template to add a step

| Data | **Profile** | Visualizations |

age
Integer

sex
String

bmi
Decimal

FREQUENCY

[18,23)
[48,53)
[43,48)
[23,28)
[28,33)
[53,58)
[38,43)
[33,38)
[58,63)
[63,64)

0   50   100   150   200

FREQUENCY

male

female

0   200   400   600

FREQUENCY

[27.1,30.8)
[30.8,34.5)
[23.4,27.1)
[34.5,38.3)
[19.7,23.4)
[38.3,42)
[42,45.7)
[16,19.7)
[45.7,49.4)
[49.4,53.1)

0   50   100   150   200   250   300

STATISTICS

| Interquartile Range | 24 |

STATISTICS

| Maximum length | 6 |

STATISTICS

| Interquartile Range | 8.3975 |

Information

Details   Help

Edit   ✎

LOCATION
Medical Premium Charges

DATA REFINERY FLOW NAME

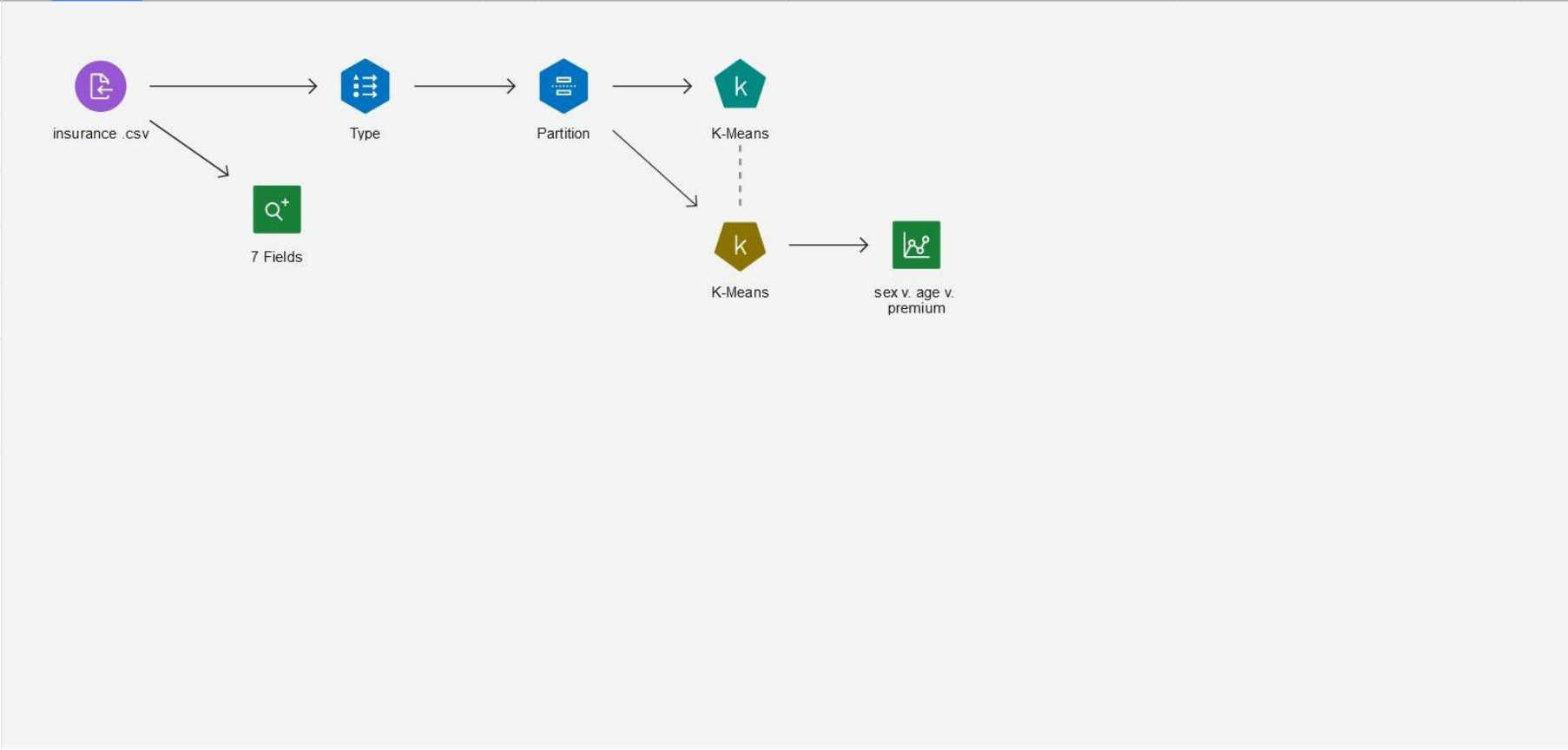insurance .csv_flow

Enter a description of the Data
Refinery flow

STEPS
1

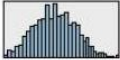**DATA REFINERY FLOW OUTPUT**

Location

Find palette nodes
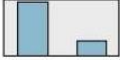
Run ▷

Import

Record Operations

Field Operations

Modeling

Text Analytics

Graphs

Outputs

Export

insurance .csv

7 Fields

Type

Partition

K-Means

K-Means

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
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 | White Space | E |

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
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	White Space	Blank Value
1	age	Continuous	0	0	None	Never	Fixed	100.000	1338	0	0	0	0
2	sex	Categorical	--	--	--	Never	Fixed	100.000	1338	0	0	0	0
3	bmi	Continuous	4	0	None	Never	Fixed	100.000	1338	0	0	0	0
4	children	Continuous	18	0	None	Never	Fixed	100.000	1338	0	0	0	0
5	smoker	Categorical	--	--	--	Never	Fixed	100.000	1338	0	0	0	0
6	region	Categorical	--	--	--	Never	Fixed	100.000	1338	0	0	0	0
7	premium	Continuous	7	0	None	Never	Fixed	100.000	1338	0	0	0	0

View Output: sex v. age v. premium
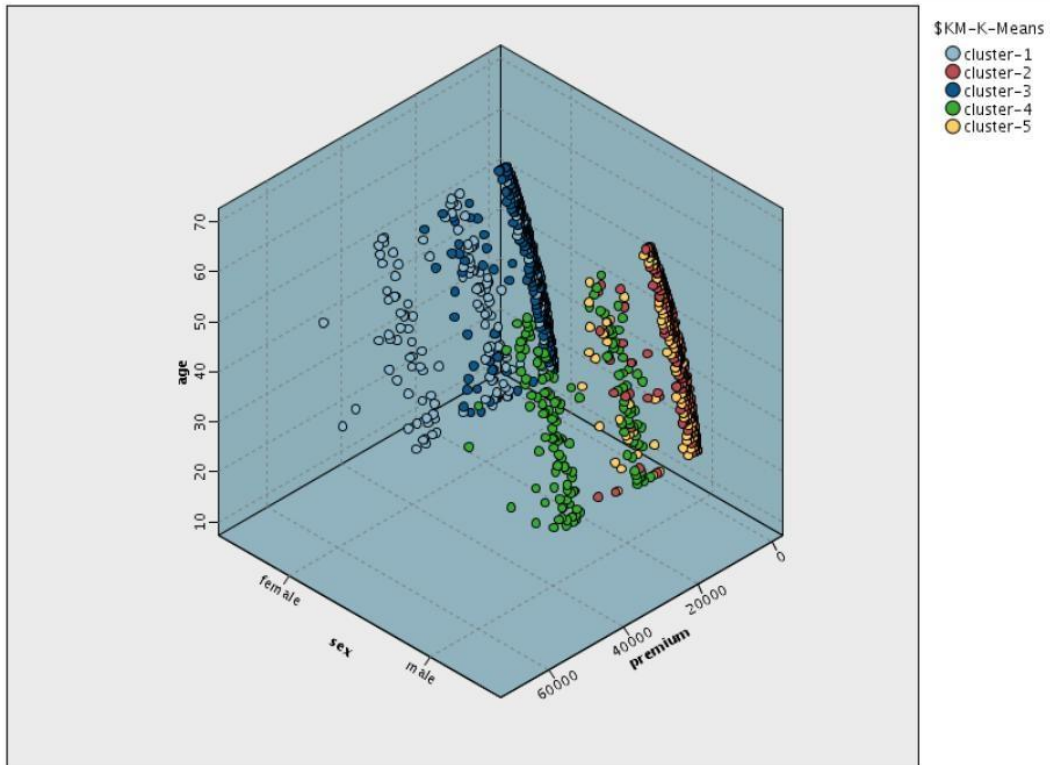
View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance
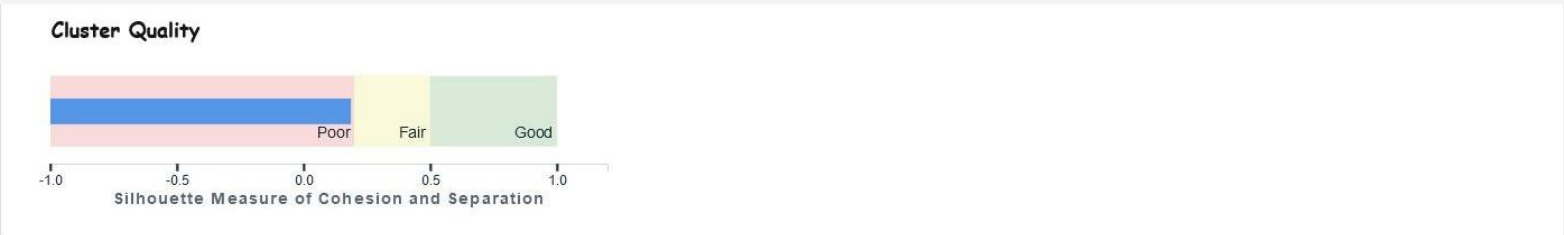
Cluster Sizes

Cluster Comparison

Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Cluster Quality ⓘ

Cluster Quality

Silhouette Measure of Cohesion and Separation

Cluster Quality Parameters

Overall Clustering Quality (Avg. Silhouette)	0.187
Total Within Clusters Sum of Squares	0.140
Average Within Cluster Sum of Squares	0.028
Average SSB (Between ss)	0.069

View Model: K-Means

K-Means Clustering
Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information
Feature Importance
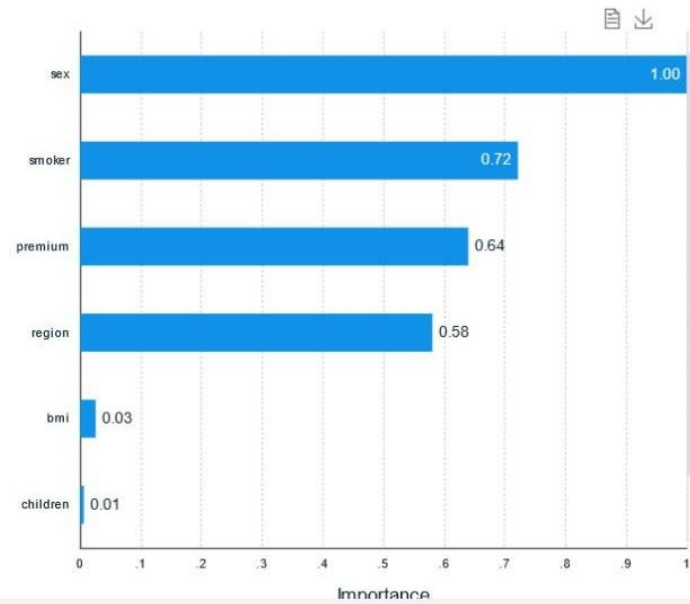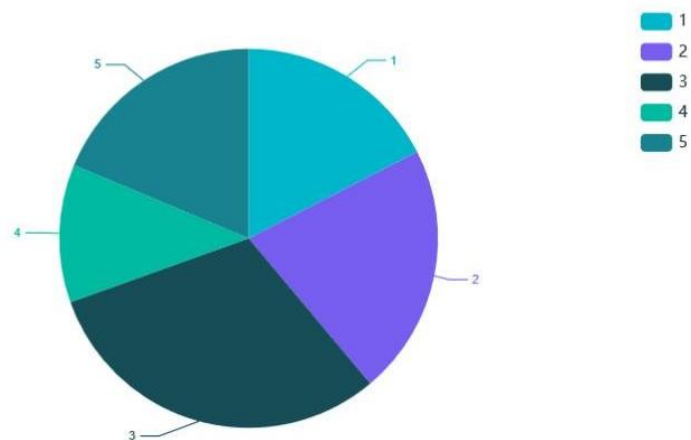Cluster Sizes
Cluster Comparison
Clusters
Cell Distributions (Absolute)
Cell Distributions (Relative)
Build Settings
Training Summary

Model Information ⓘ

Algorithm		K-Means
Model Class		Center Based
Number of Features		7
Distance Measure		Euclidean
Number of Clusters		5
Number of instances in each cluster	Cluster 1	114 (17.46%)
	Cluster 2	140 (21.44%)
	Cluster 3	200 (30.63%)
	Cluster 4	77 (11.79%)
	Cluster 5	122 (18.68%)
Ratio of sizes (Largest to smallest)		2.597

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 ⓘ

Feature Importance bar chart:

Feature	Importance
sex	1.00
smoker	0.72
premium	0.64
region	0.58
bmi	0.03
children	0.01

Importance

View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

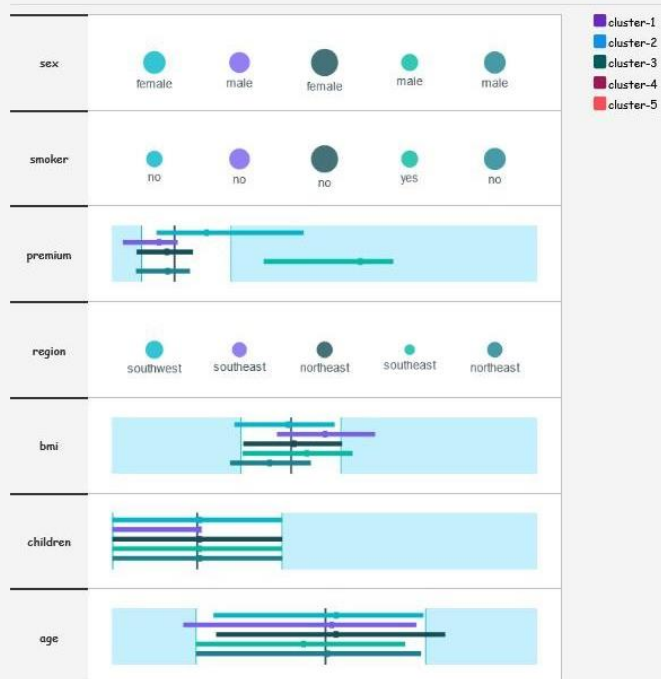**Cluster Sizes**

Cluster Comparison

Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Cluster Sizes ⓘ

View Model: K-Means

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 Comparison ⓘ

Legend:
- cluster-1
- cluster-2
- cluster-3
- cluster-4
- cluster-5

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

Clusters ⓘ

Input Importance
■ 1.0 ■ 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 **male (100.00%)**
	smoker **no (54.23%)**	smoker **no (100.00%)**	smoker **no (100.00%)**	smoker **yes (100.00%)**	smoker **no (100.00%)**
	premium **19023.57**	premium **7653.39**	premium **9187.54**	premium **32910.81**	premium **9261.61**
	region **southwest (64.08%**	region **southeast (52.50%)**	region **northeast (35.92%)**	region **southeast (39.02%)**	region **northeast (50.37%)**
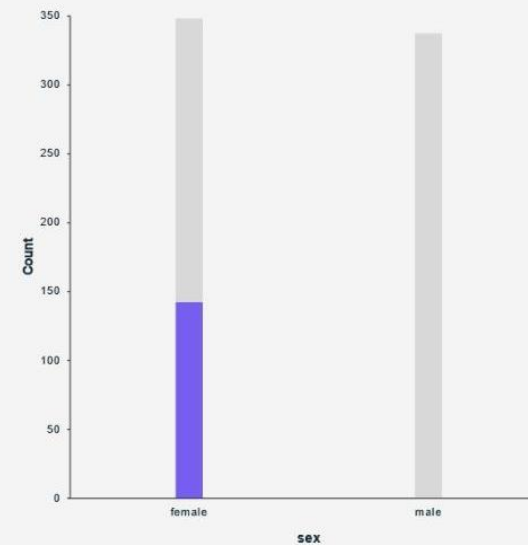
Count (y-axis): 350, 300, 250, 200, 150, 100, 50, 0

sex (x-axis): female, male

View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

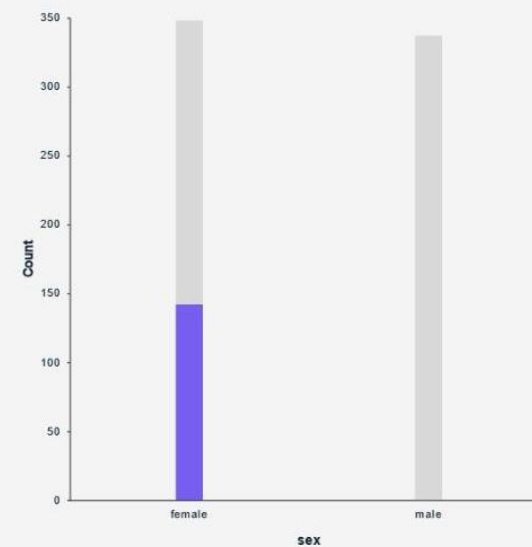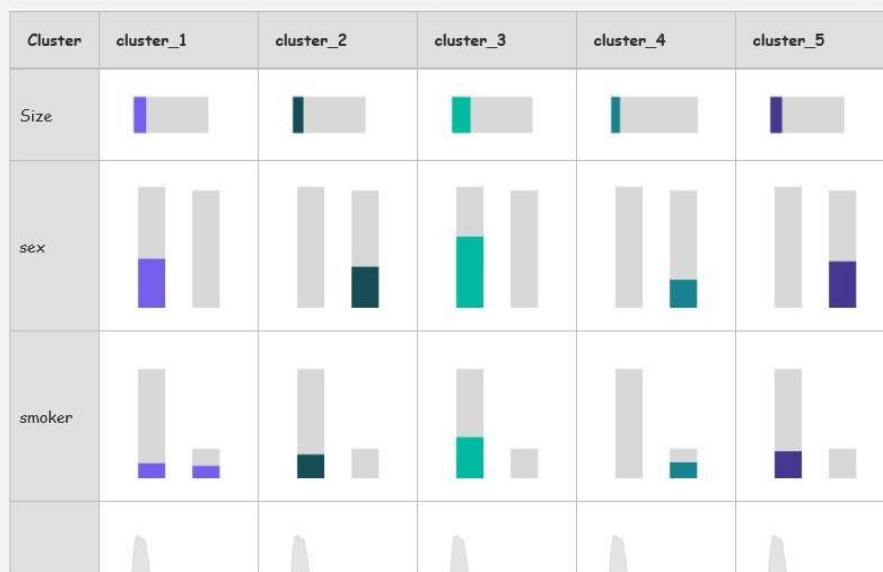Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Cell Distributions (Absolute) ⓘ

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

Cell Distributions (Relative) ⓘ

View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

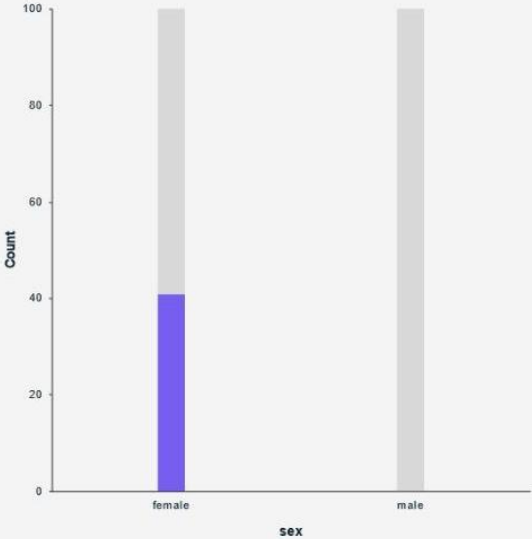Feature Importance
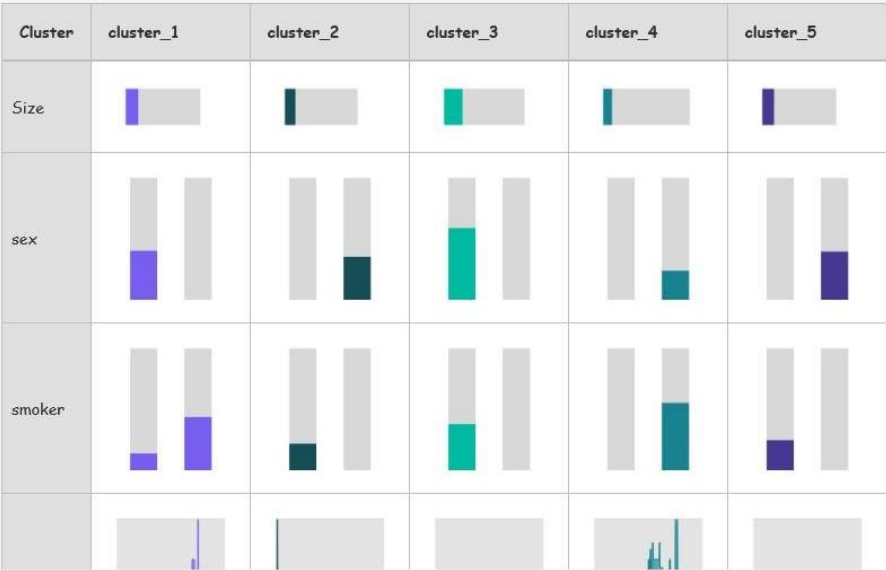
Cluster Sizes

Cluster Comparison

Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

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
Optimize	Memory
Mode	Simple

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

Training Summary ⓘ

Algorithm	K-means
Model type	Clustering
Date built	Sun Jan 09 17:40:44 UTC 2022
Elapsed time for model build	0 hours, 0 mins, 0 secs