Projects / Medical Premium Charges / insurance .csv_flow

Steps | Use a code template to add a step

Data | Profile | Visualizations

	age String	sex String	bmi String	children String	smoker String	region String	premium String
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
14	56	female	39.82	0	no	southeast	11090.7178

Information

Details | Help

Edit

LOCATION

Medical Premium Charges

DATA REFINERY FLOW NAME

insurance .csv_flow

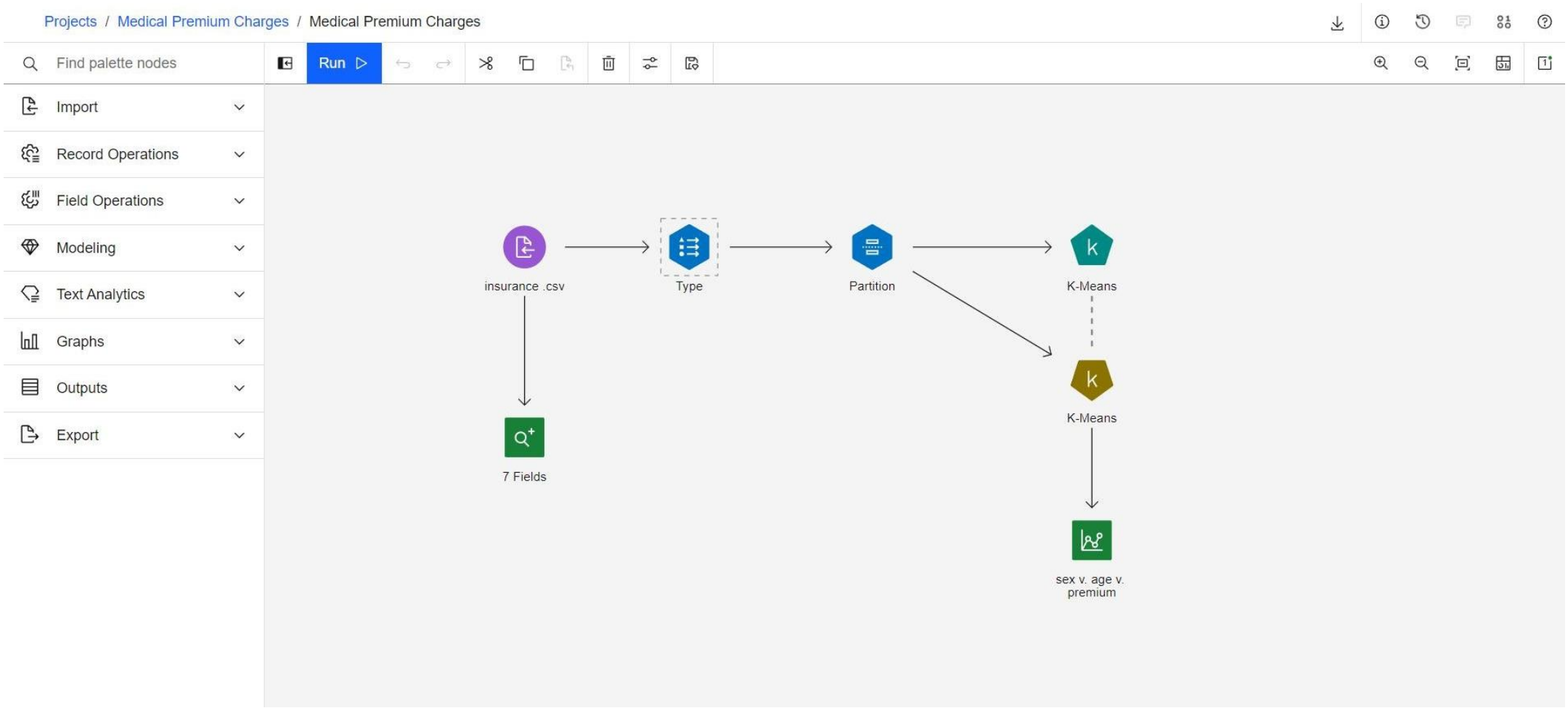
Enter a description of the Data Refinery flow

STEPS

0

DATA REFINERY FLOW OUTPUT

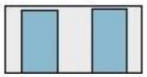
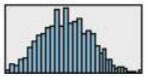
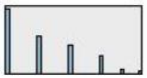
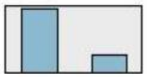
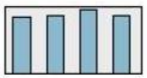
Location


Projects / Medical Premium Charges / Medical Premium Charges
Find palette nodes
Run
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

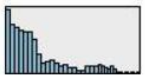
Find in column Field

Field	Measure	Role	Value mode	Values	Check
age	Continuous	Input	Instantiated	18, 64	None
sex	Flag	Input	Instantiated	female, male	None
bmi	Continuous	Input	Instantiated	15.96, 53.13	None
children	Continuous	Input	Instantiated	0, 5	None
smoker	Flag	Input	Instantiated	no, yes	None
region	Nominal	Input	Instantiated	northeast, northwe…	None
premium	Continuous	Input	Instantiated	1121.8739, 63770.…	None

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

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

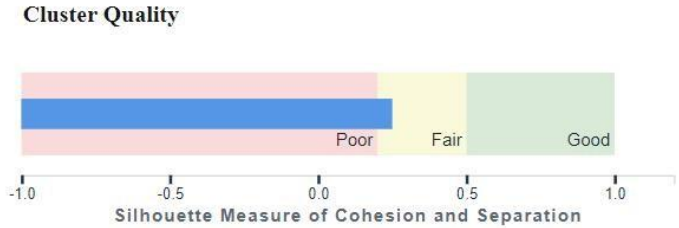
Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Cluster Quality ⓘ



Cluster Quality Parameters

Overall Clustering Quality (Avg. Silhouette)	0.249
Total Within Clusters Sum of Squares	0.132
Average Within Cluster Sum of Squares	0.026

View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

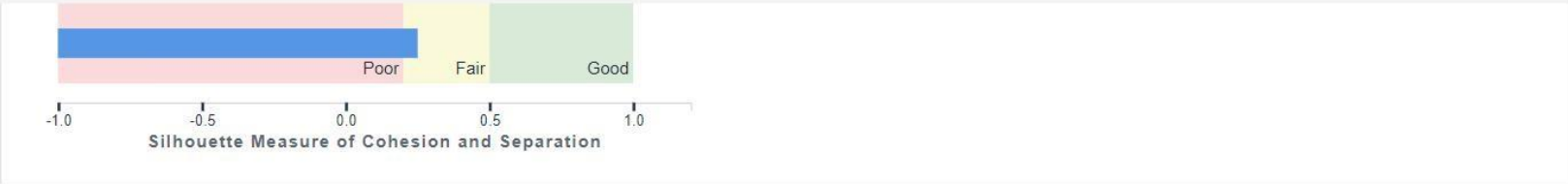
Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Cluster Quality ⓘ



Cluster Quality Parameters

Overall Clustering Quality (Avg. Silhouette)	0.249
Total Within Clusters Sum of Squares	0.132
Average Within Cluster Sum of Squares	0.026
Average SSB (Between ss)	0.075

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
	Cluster 1	81 (8.7%)
	Cluster 2	368 (39.53%)

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)

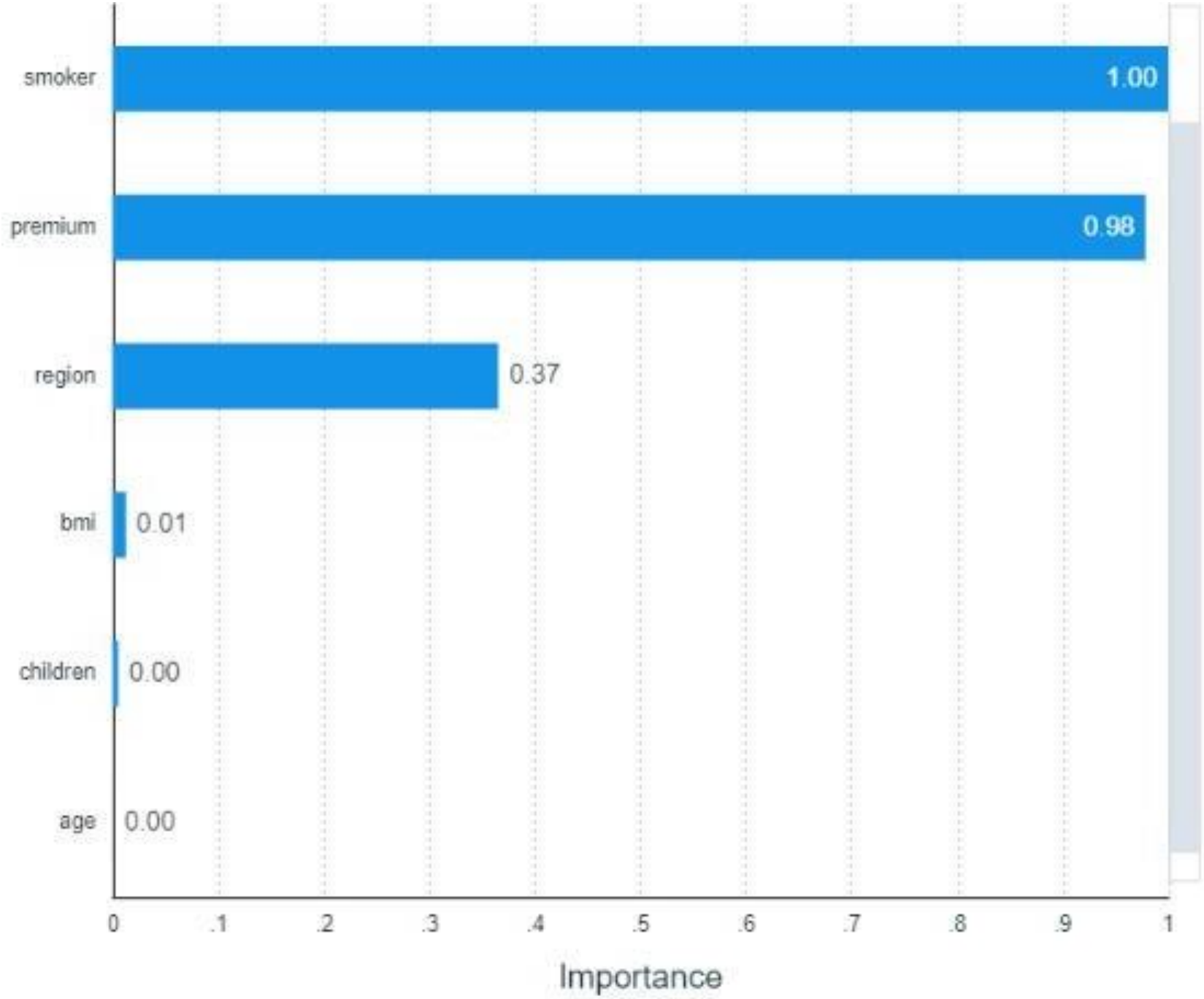
Cell Distributions (Relative)

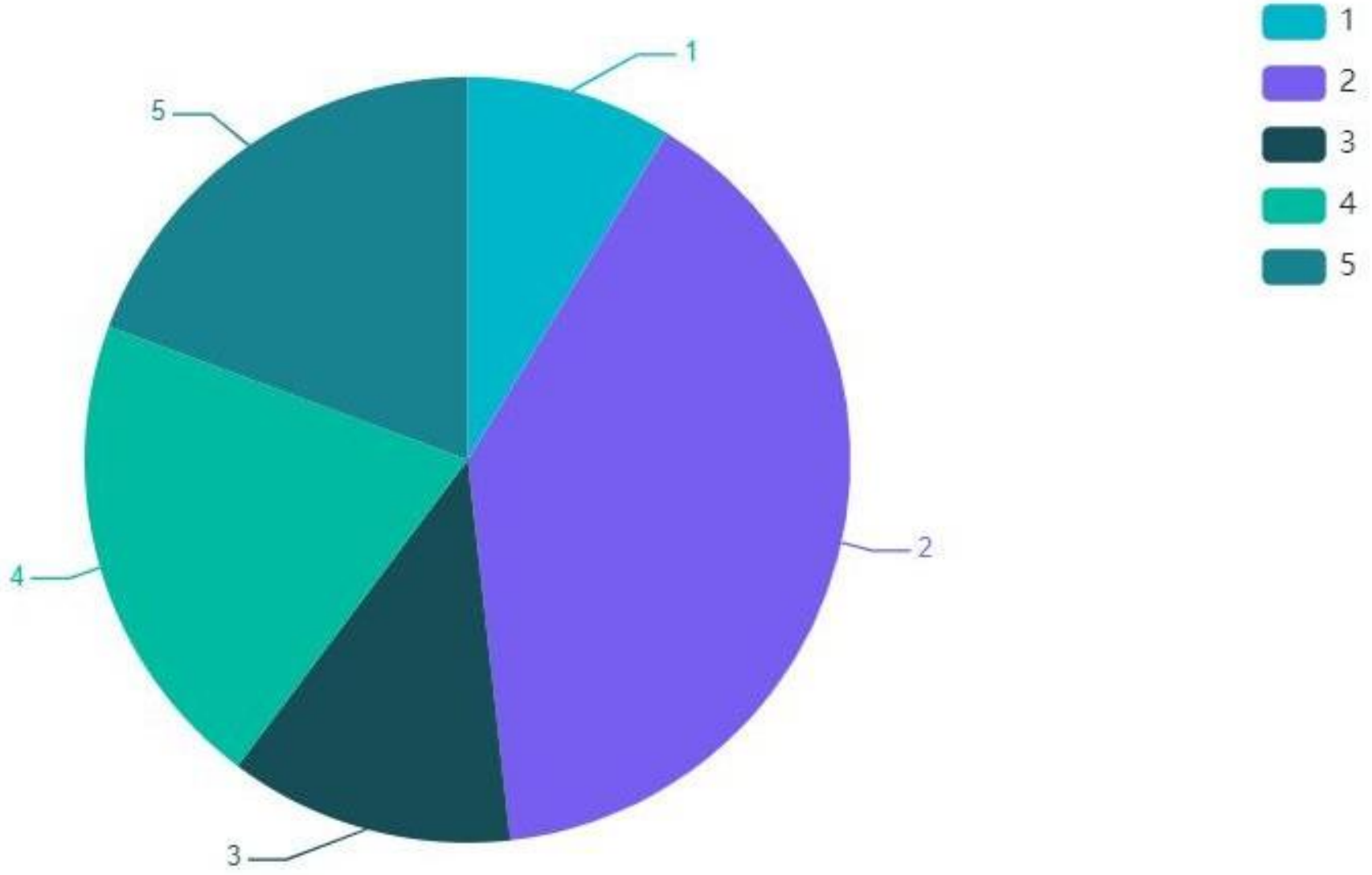
Build Settings

Training Summary

Model Information ⓘ

Number of Clusters		5
Number of instances in each cluster	Cluster 1	81 (8.7%)
	Cluster 2	368 (39.53%)
	Cluster 3	112 (12.03%)
	Cluster 4	190 (20.41%)
	Cluster 5	180 (19.33%)
Ratio of sizes (Largest to smallest)		4.543


smoker
1.00
premium
0.98
region
0.37
bmi
0.01
children
0.00
age
0.00
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Importance


1
2
3
4
5
1
2
3
4
5

View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

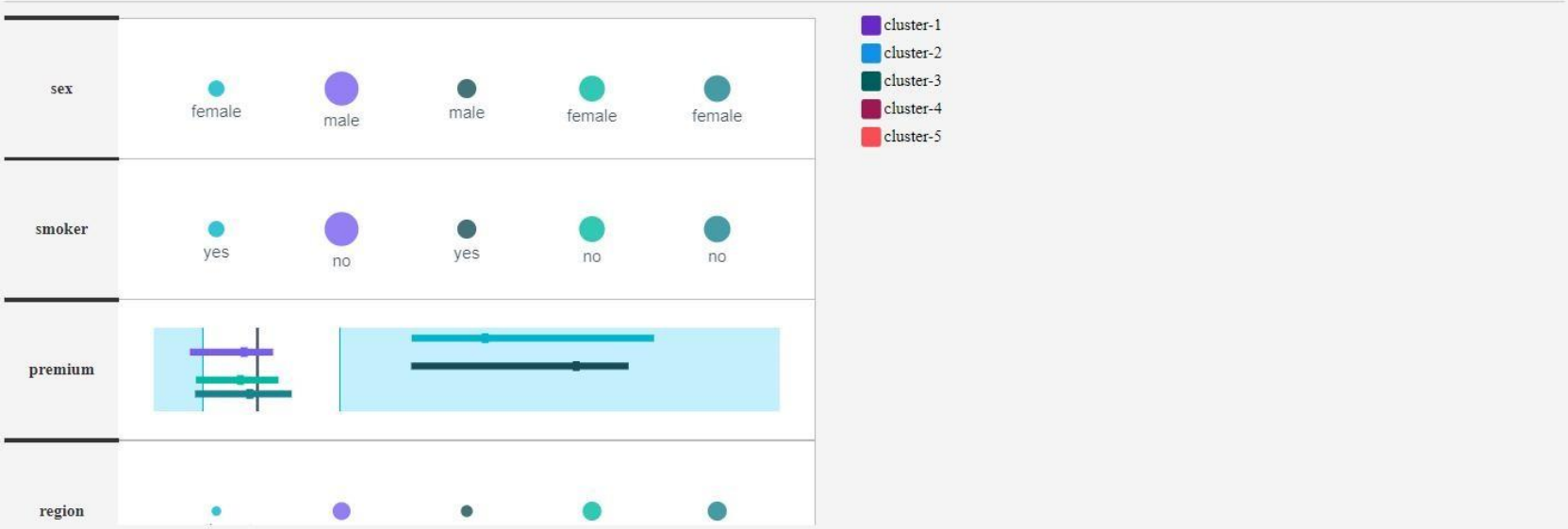
Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Cluster Comparison ⓘ



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

Cluster Comparison ⓘ

View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

- Cluster Quality

MODEL VIEWER

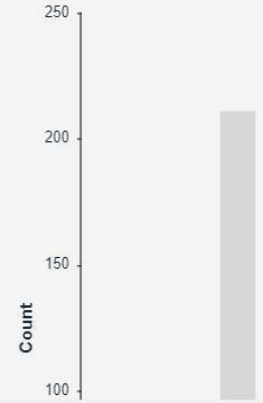
- Model Information
- Feature Importance
- Cluster Sizes
- Cluster Comparison
- **Clusters**
- Cell Distributions (Absolute)
- Cell Distributions (Relative)
- Build Settings

Clusters ⓘ

Input Importance

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 **male (100.00%)**	sex **female (100.00%)**	sex **female (100.00%)**
	smoker **yes (100.00%)**	smoker **no (100.00%)**	smoker **yes (100.00%)**	smoker **no (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)

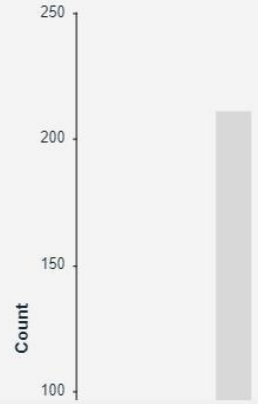
Build Settings

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
	premium **30921.55**	premium **8866.62**	premium **31714.55**	premium **8191.76**	premium **8842.23**
	region **northwest (35.29%)**	region **northwest (27.52%)**	region **southeast (38.30%)**	region **southwest (53.49%)**	region **southeast (51.65%)**
	bmi **29.95**	bmi **30.52**	bmi **30.70**	bmi **30.39**	bmi **31.77**



View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

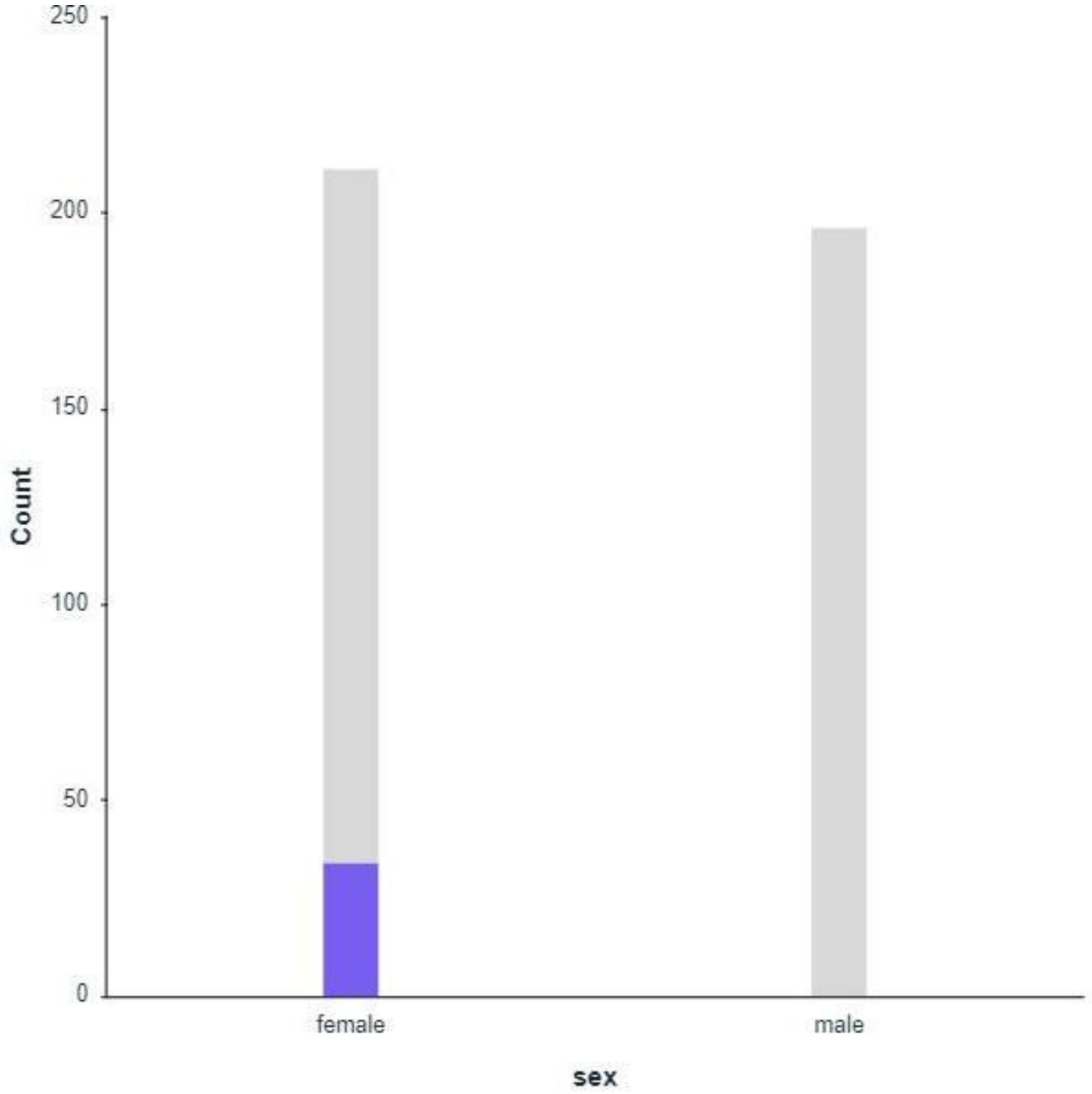
- Cluster Quality

MODEL VIEWER

- Model Information
- Feature Importance
- Cluster Sizes
- Cluster Comparison
- **Clusters**
- Cell Distributions (Absolute)
- Cell Distributions (Relative)

Clusters ⓘ

	bmi **29.95**	bmi **30.52**	bmi **30.70**	bmi **30.39**	bmi **31.77**
	children **.91**	children **1.11**	children **.77**	children **1.16**	children **1.07**
	age **39.65**	age **40.22**	age **37.79**	age **40.50**	age **41.43**


250
200
150
100
50
0
Count
female
male
sex

View Model: K-Means

K-Means Clustering Model ⓘ

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

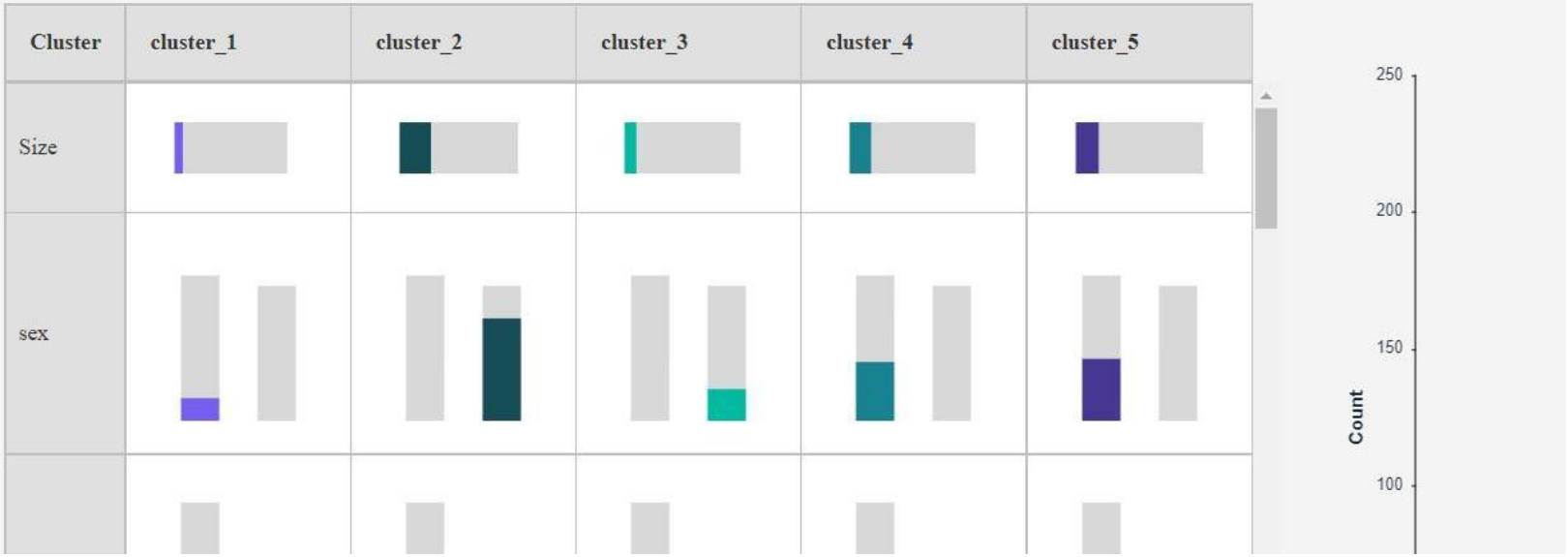
Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Cell Distributions (Absolute) ⓘ



View Model: K-Means

K-Means Clustering Model ⓘ

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

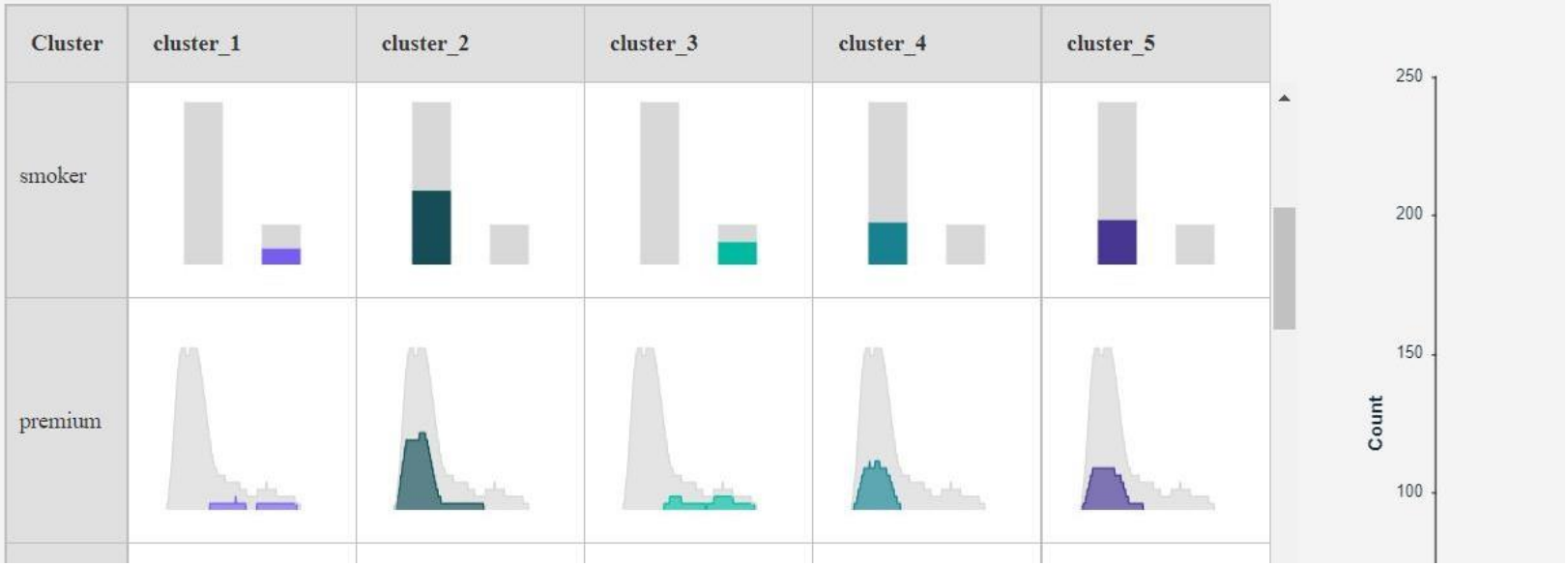
Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Cell Distributions (Absolute) ⓘ



View Model: K-Means

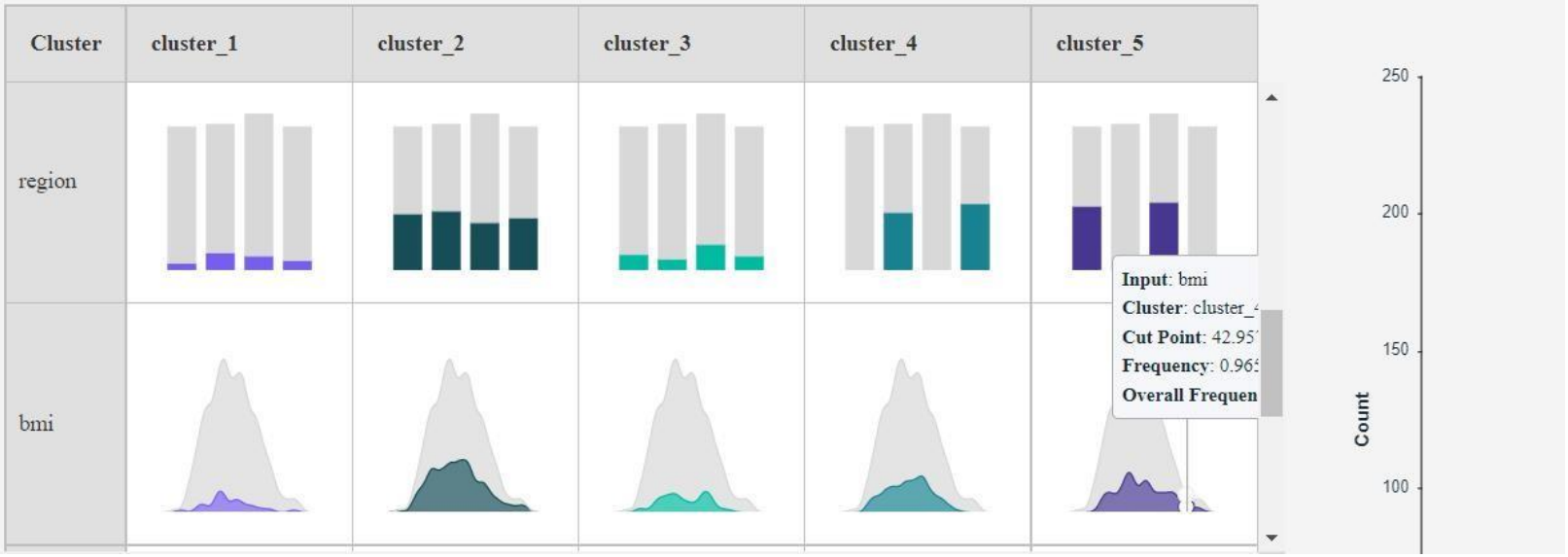
K-Means Clustering Model ⓘ

- Cluster Quality

MODEL VIEWER

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

Cell Distributions (Absolute) ⓘ



View Model: K-Means

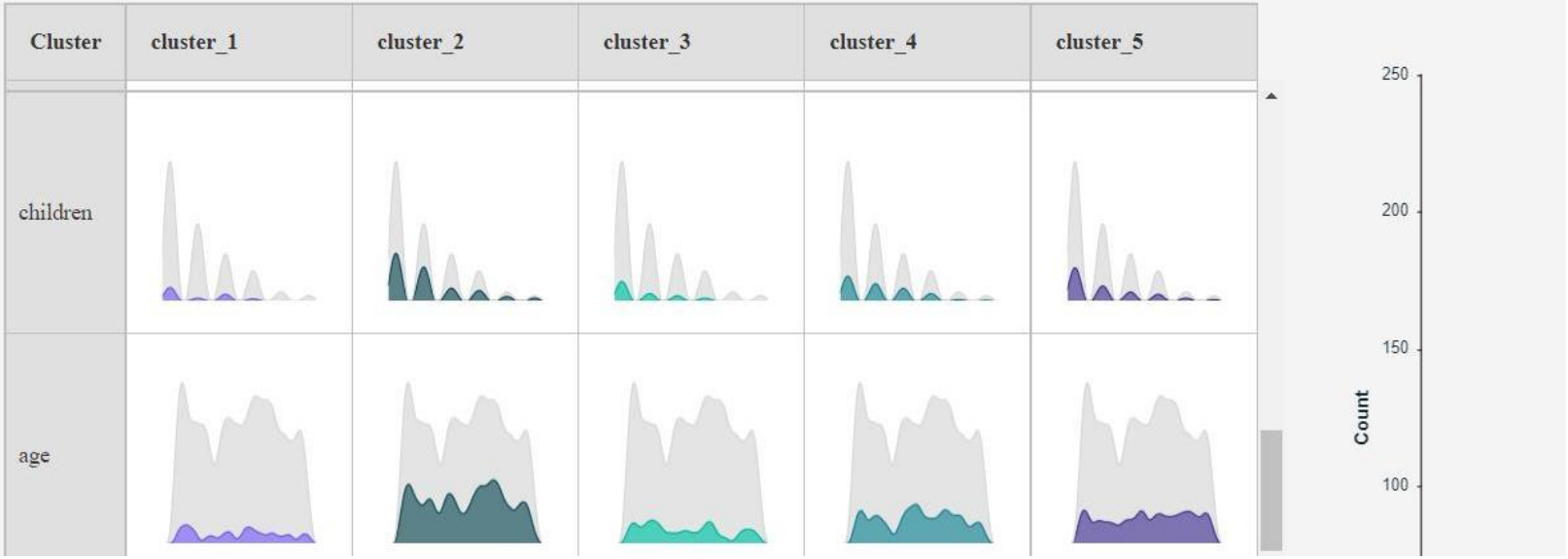
K-Means Clustering Model ⓘ

- Cluster Quality

MODEL VIEWER

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

Cell Distributions (Absolute) ⓘ



View Model: K-Means

K-Means Clustering Model ⓘ

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

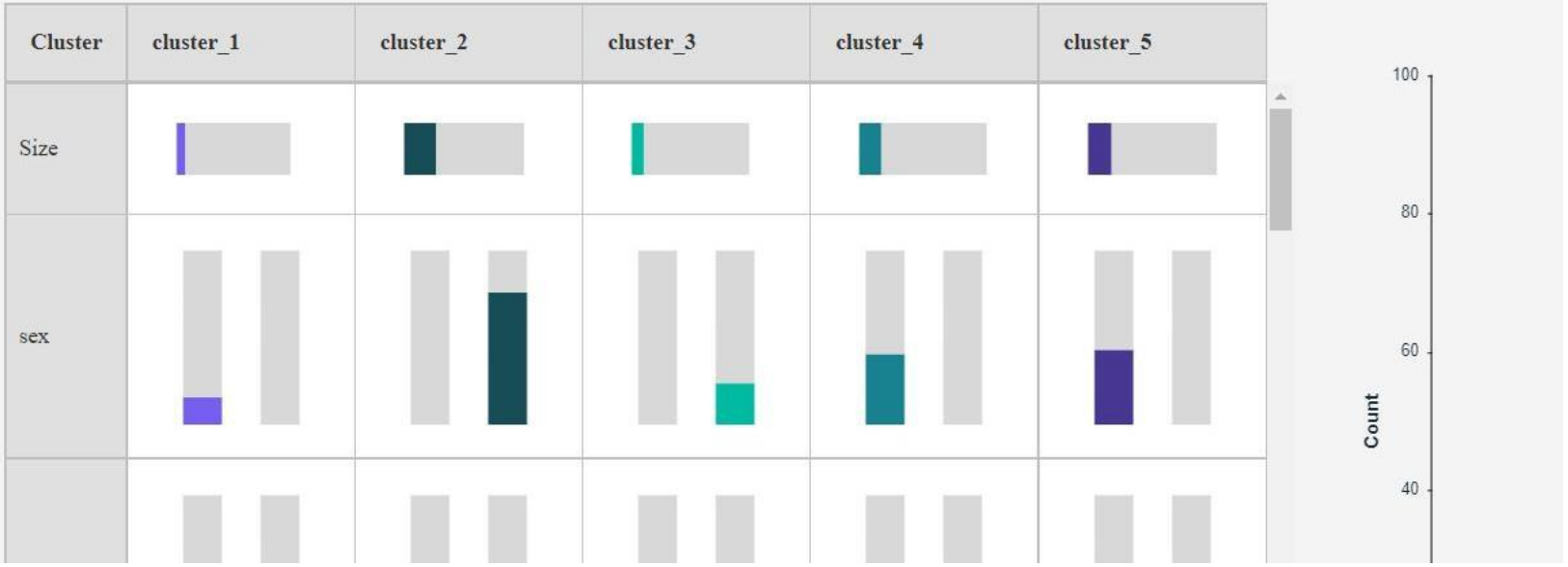
Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Cell Distributions (Relative) ⓘ



View Model: K-Means

K-Means Clustering Model ⓘ

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Cell Distributions (Relative) ⓘ

View Model: K-Means

K-Means Clustering Model ⓘ

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Cell Distributions (Relative) ⓘ

View Model: K-Means

K-Means Clustering Model ⓘ

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

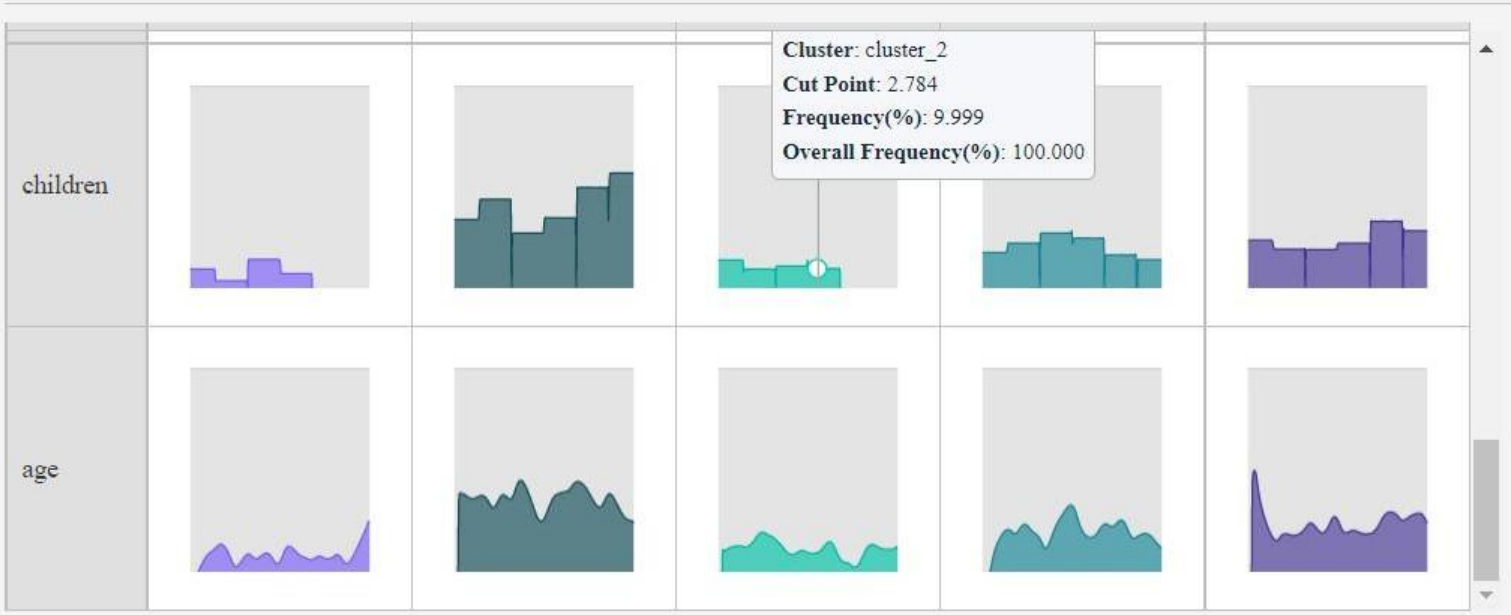
Clusters

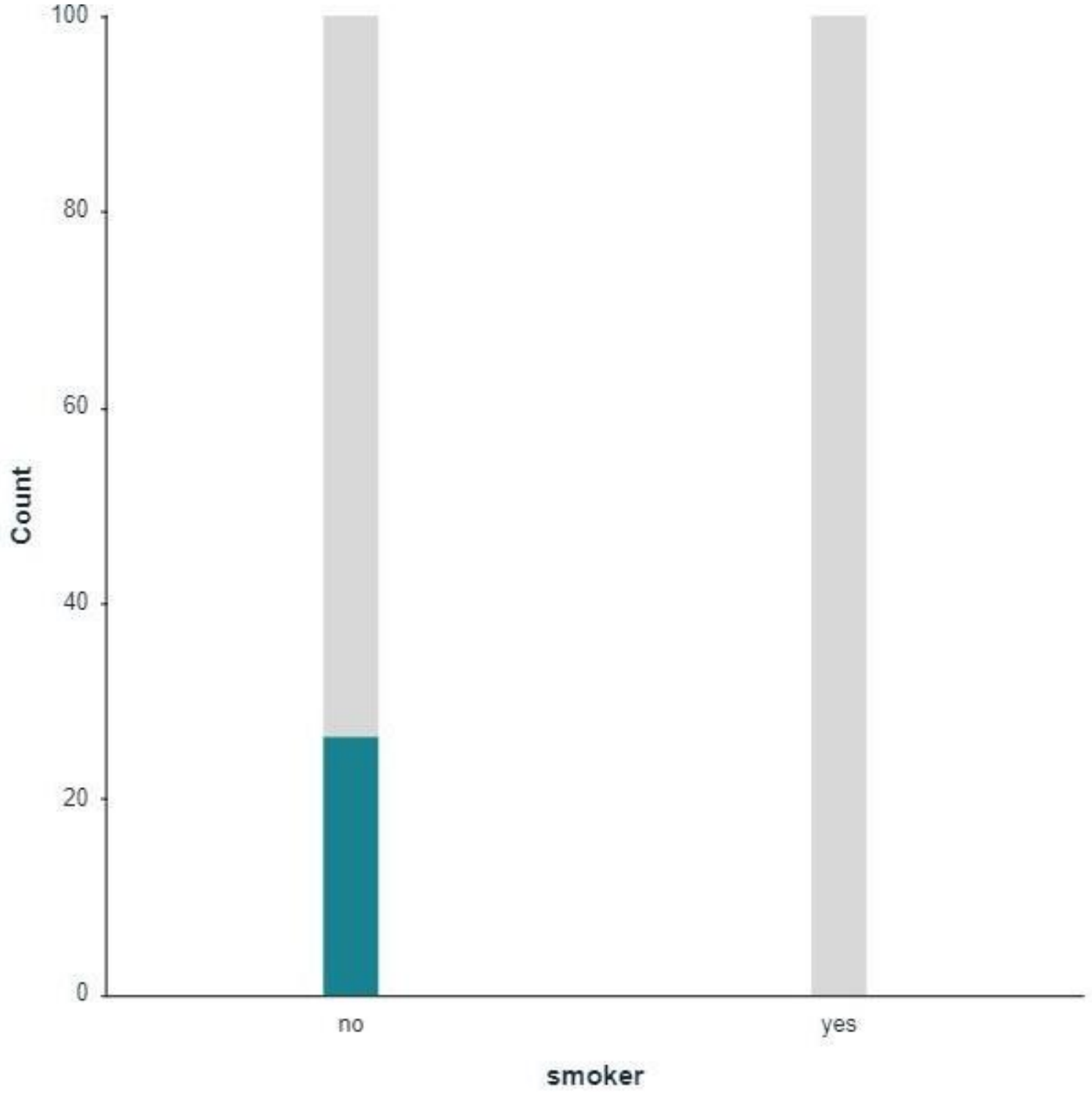
Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Cell Distributions (Relative) ⓘ




100
80
60
40
20
0
Count
no
yes
smoker

View Model: K-Means

K-Means Clustering Model ⓘ

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

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

View Model: K-Means

K-Means Clustering Model ⓘ

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

Cell Distributions (Absolute)

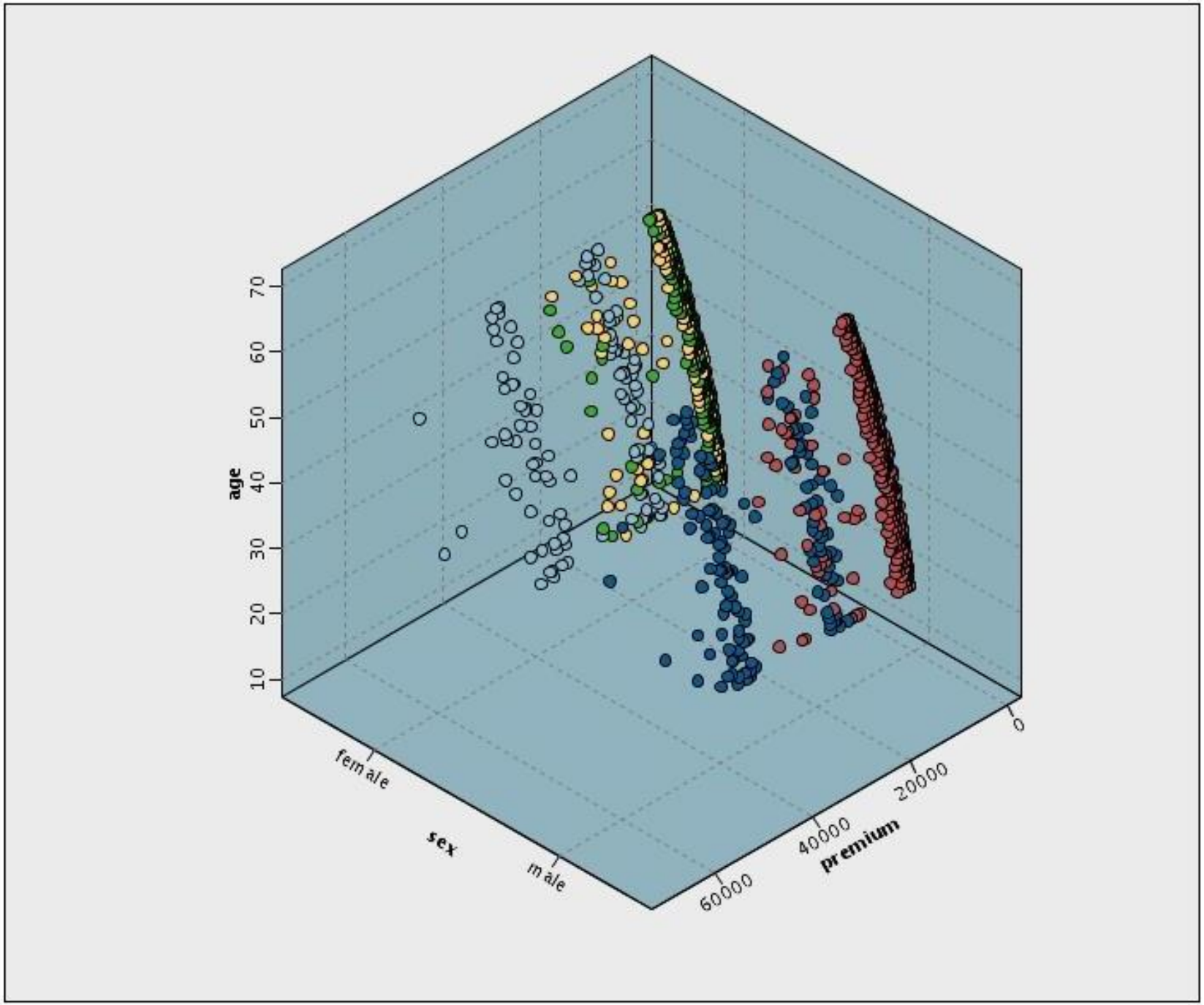
Cell Distributions (Relative)

Build Settings

Training Summary

Training Summary ⓘ

Algorithm	K-means
Model type	Clustering
Date built	Tue Dec 28 10:43:11 UTC 2021
Elapsed time for model build	0 hours, 0 mins, 0 secs


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