

NAME:S.S.NITHISH SOORYA

REG NO:19BEC0212

CAMPUS : VELLORE

DATA ANALYTICS

ASSIGNMENT - 4

CLUSTERING MEDICAL PREMIUM CHARGES

IBM Watson Studio

Search in your workspaces

Buy

NITHISH SOORYA S S's Ac...

NS

My projects / Medical premium Charges / insurance.csv

Information

Preview

Activities

Schema: 7 Columns

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

Last refresh: 12 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

Information

×

Data asset

insurance.csv

Description

No description available for this asset

Tags

No tags available for this asset

Creator

NITHISH SOORYA S S

Usage

Created on Apr 25, 2022, 10:00 PM

Size

55.628 KB

IBM Watson Studio

Search in your workspaces

Buy

NITHISH SOORYA S S's Ac...

NS

Projects / Medical premium Charges / insurance.csv_flow

Information

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
14	56	female	39.82	0	no	southeast	11090.7178
	27	male	42.13	0	yes	southeast	39611.7577

SOURCE FILE: insurance.csv

FULL DATA SET: 1338 rows

Information

×

Details

Help

Edit

LOCATION

Medical premium Charges

DATA REFINERY FLOW NAME

insurance.csv_flow

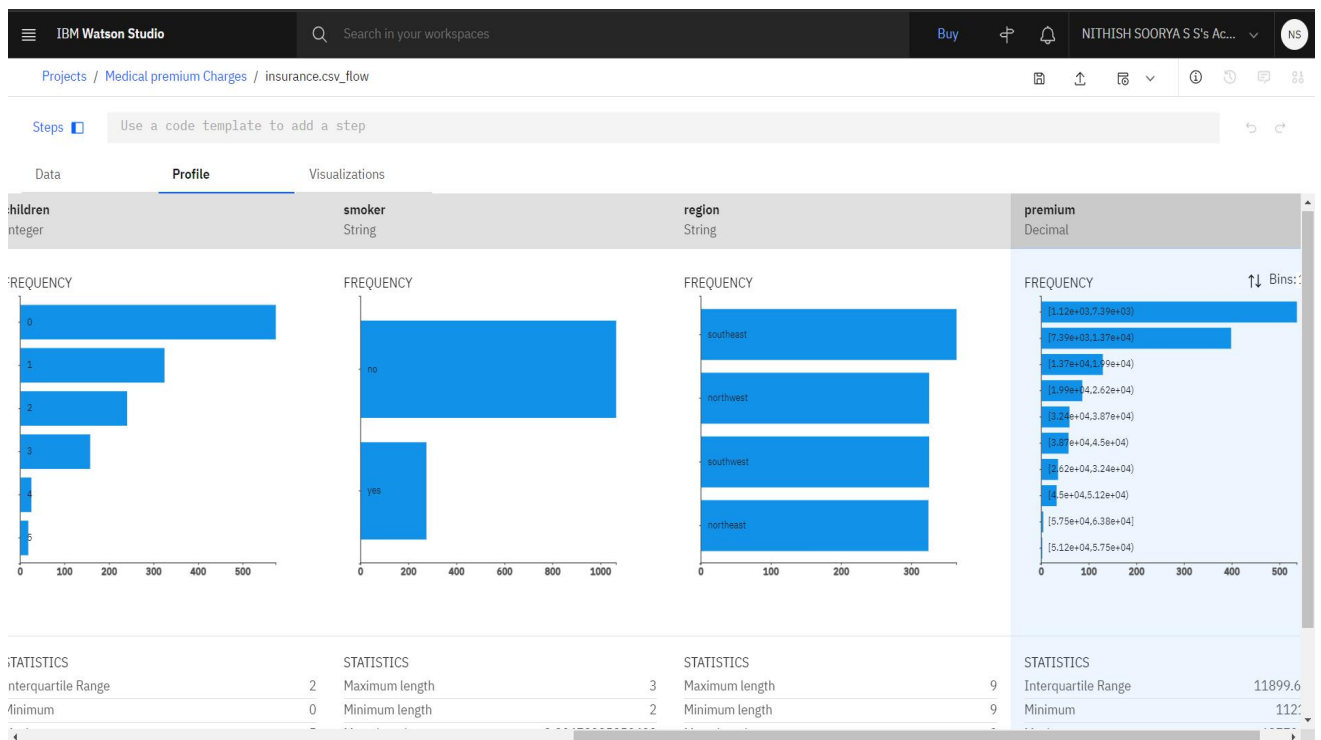
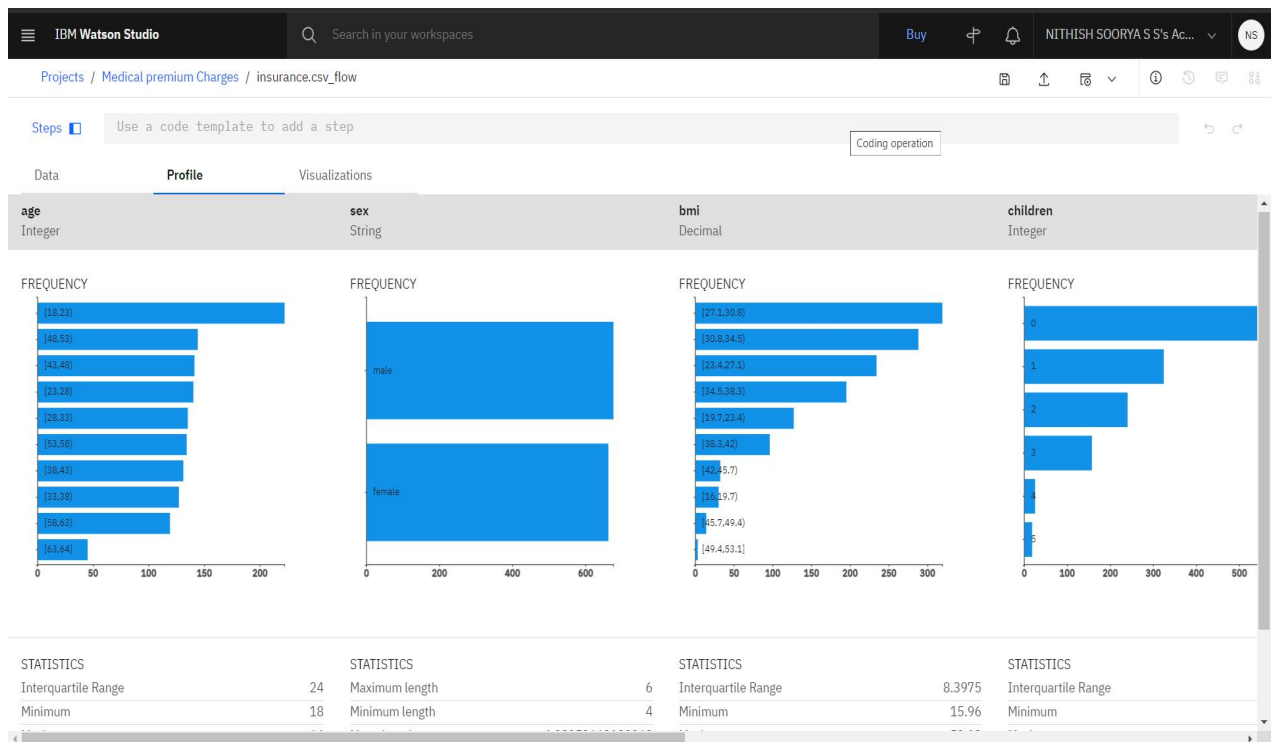
Enter a description of the Data Refinery flow

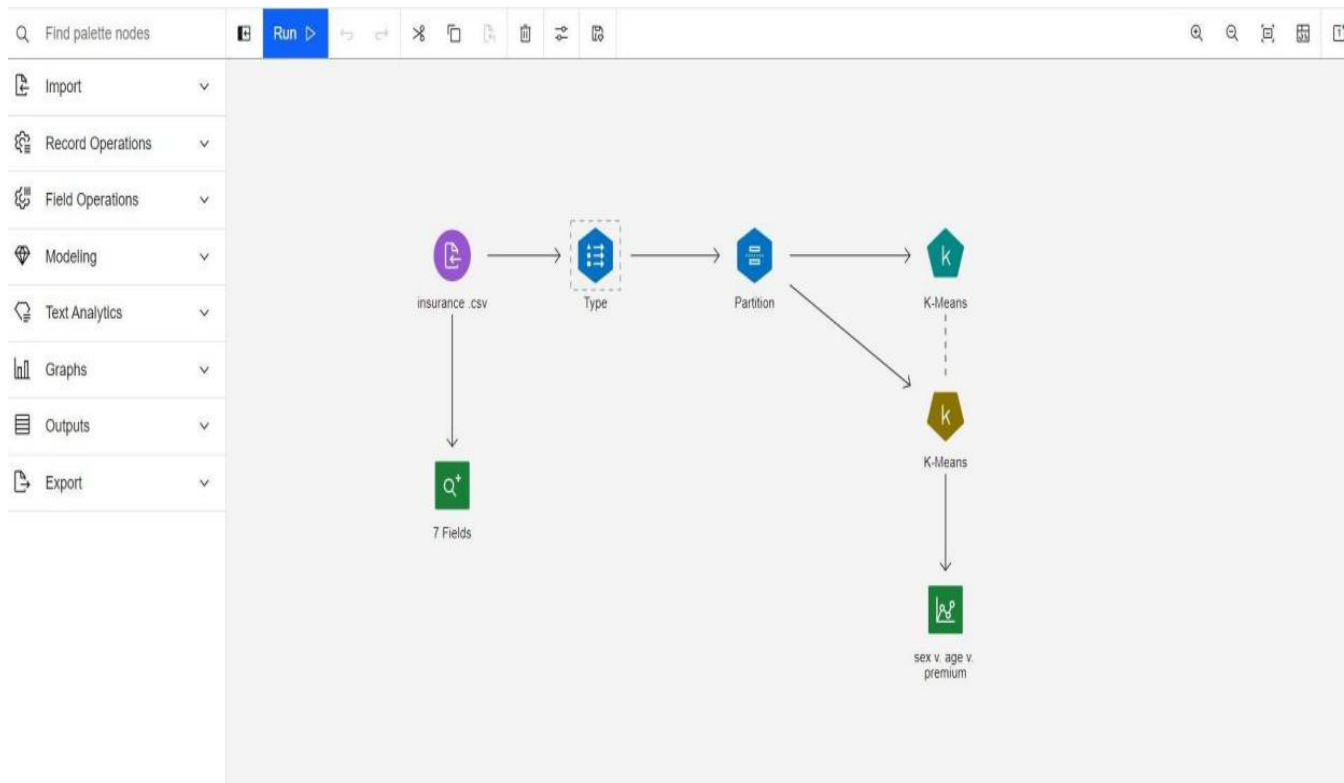
STEPS

1

DATA REFINERY FLOW OUTPUT

Location







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



Silhouette Measure of Cohesion and Separation

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

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

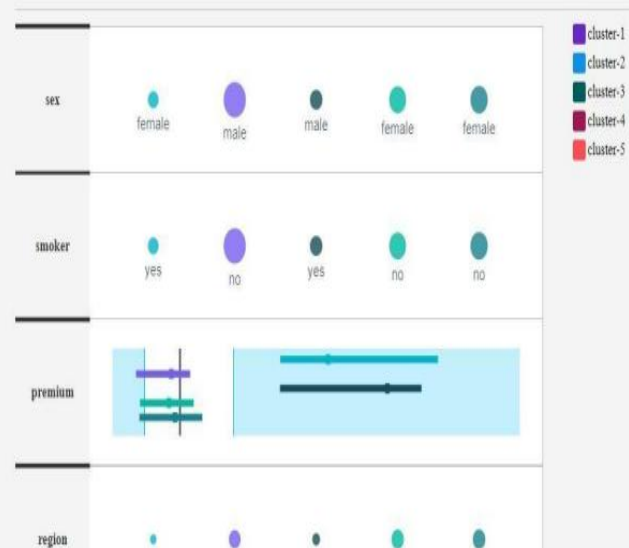
Clusters

Cell Distributions (Absolute)

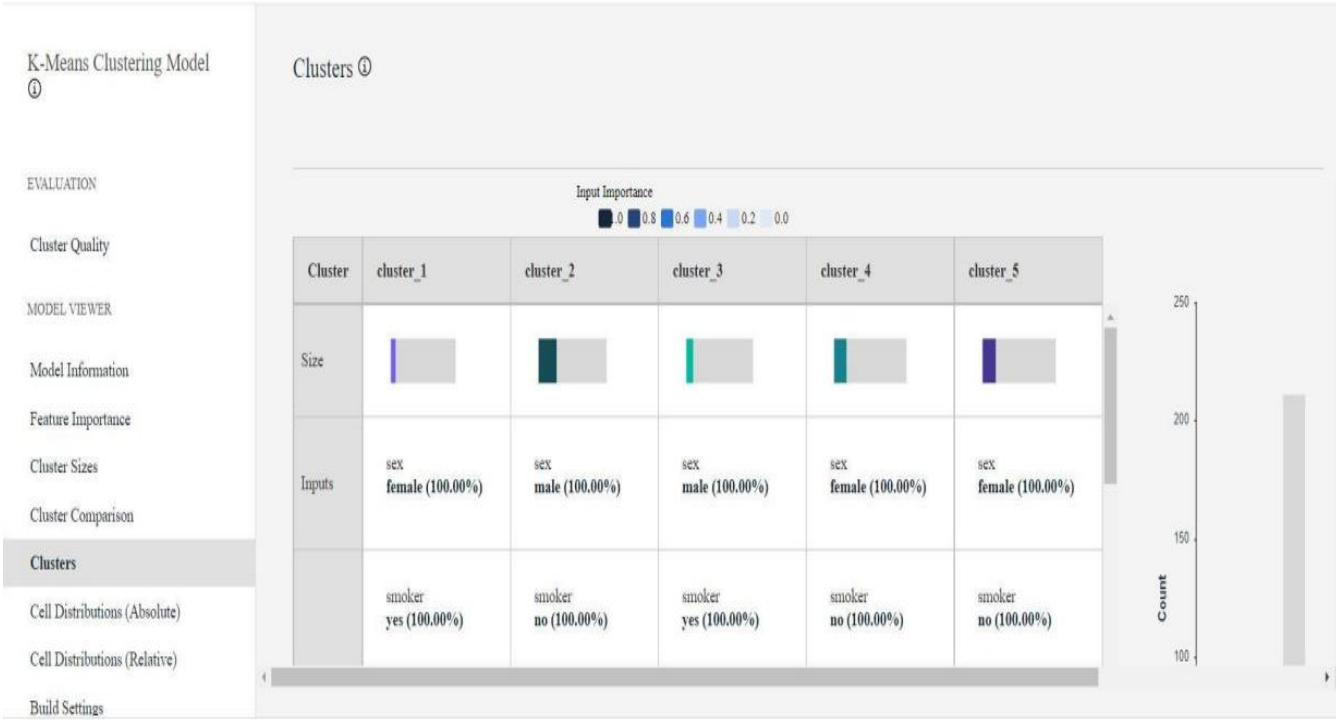
Cell Distributions (Relative)

Build Settings

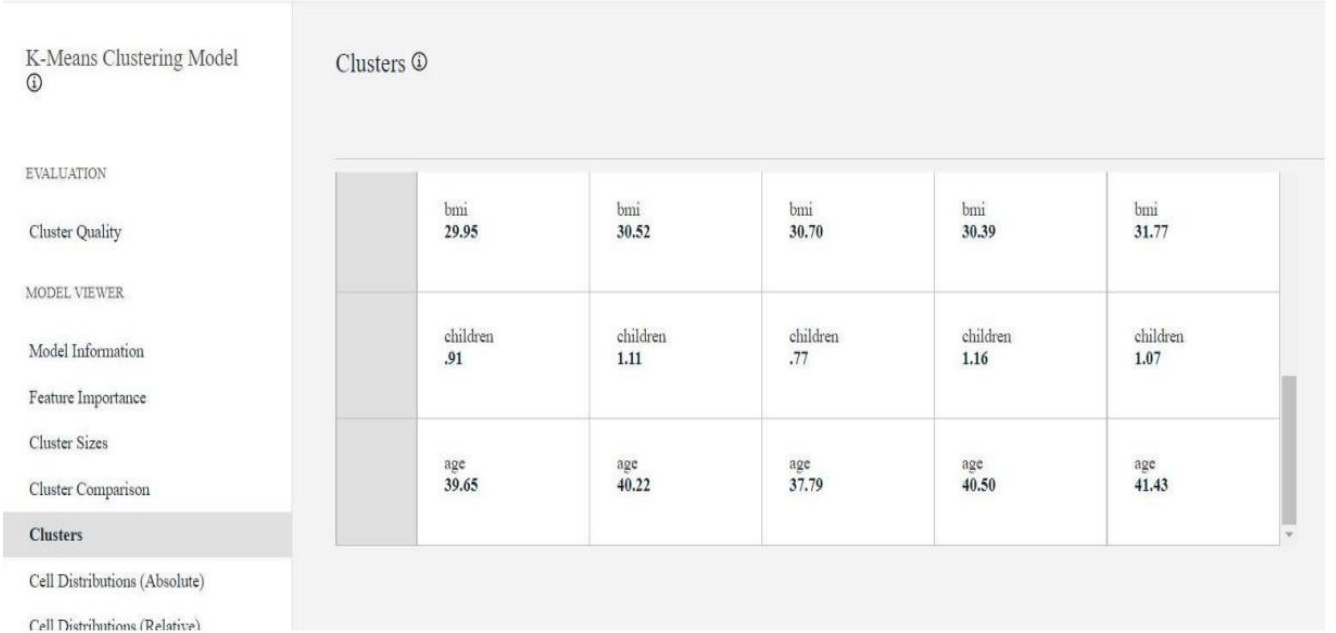
Cluster Comparison ⓘ



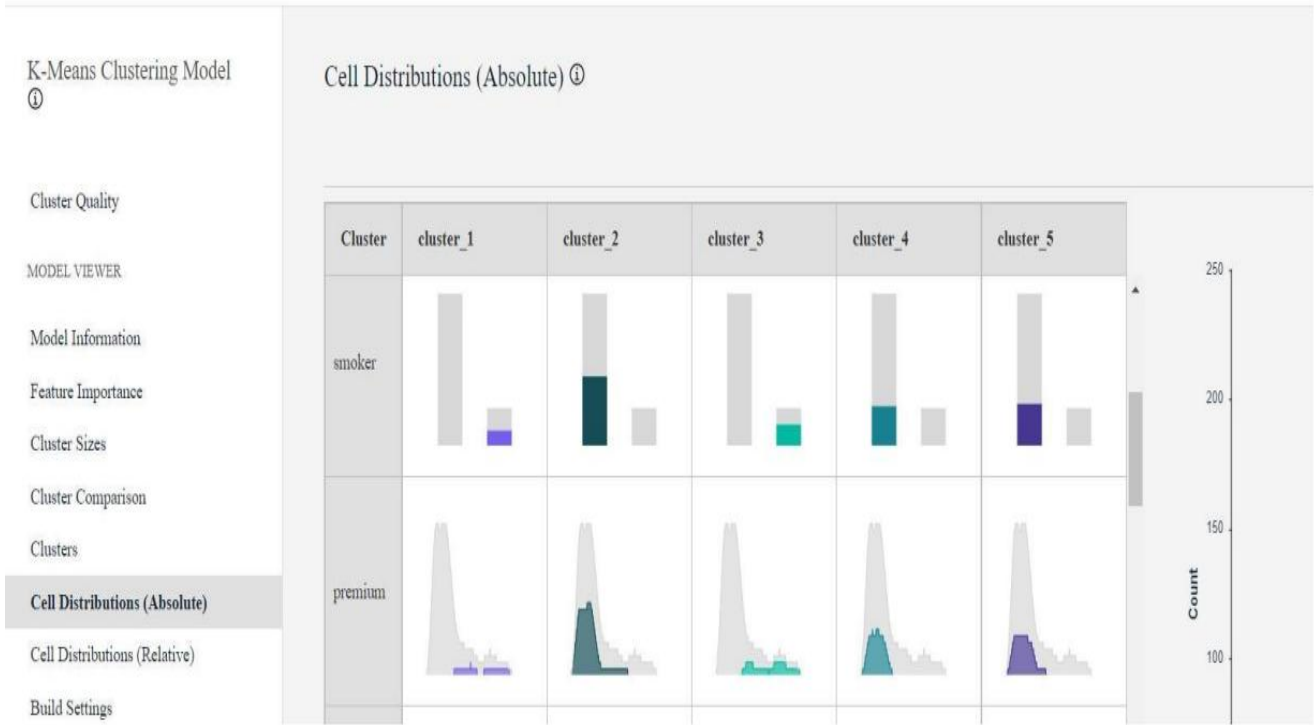
View Model: K-Means



View Model: K-Means



View Model: K-Means



View Model: K-Means



View Model: K-Means



View Model: K-Means

K-Means Clustering Model ⓘ	Training Summary ⓘ
Cluster Quality	
MODEL VIEWER	
Model Information	
Feature Importance	
Cluster Sizes	
Cluster Comparison	
Clusters	
Cell Distributions (Absolute)	
Cell Distributions (Relative)	
Build Settings	
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

