

DATA ANALYTICS

ASSIGNMENT – 4

NAME: ESWAR S S

REG NO: 19BEE0133

CAMPUS: VIT VELLORE

My projects / Medical Premium charges / insurance.csv

Preview

Activities

Schema: 7 Columns

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

Last refresh: just now ↻

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

Sachin K S

Usage

Created on Apr 28, 2022, 10:58 PM

Size

55.628 KB

Projects / Medical Premium charges / insurance.csv / Refine data

Steps (1) x

Use a code template to add a step ↻

Data Source

insurance.csv

1. Convert column type ⓘ

Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.

Auto-generated

New step +

Data

Profile

Visualizations

	age Integer	sex String	bmi Decimal	children Integer	smoker String
1	19	female	27.9	0	yes
2	18	male	33.77	1	no
3	28	male	33	3	no
4	33	male	22.705	0	no
5	32	male	28.88	0	no
6	31	female	25.74	0	no
7	46	female	33.44	1	no
8	37	female	27.74	3	no
9	37	male	29.83	2	no
10	60	female	25.84	0	no
11	25	male	26.22	0	no
12	62	female	26.29	0	yes
13	23	male	34.4	0	no
14	56	female	39.82	0	no

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 1

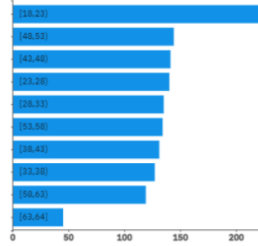
Use a code template to add a step

Data

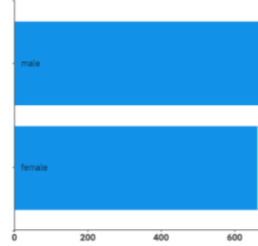
Profile

Visualizations

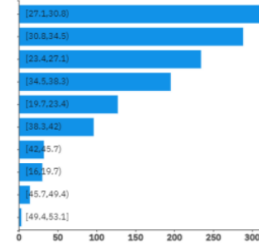
FREQUENCY



FREQUENCY



FREQUENCY



STATISTICS

Interquartile Range	24
Minimum	18
Maximum	64
Median	39

STATISTICS

Maximum length	6
Minimum length	4
Mean length	4.98953662182362
Unique	2

STATISTICS

Interquartile Range	8.3975
Minimum	15.96
Maximum	53.13
Median	30.4

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

Steps 1

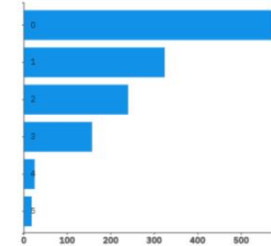
Use a code template to add a step

Data

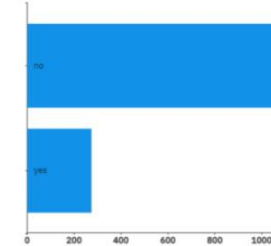
Profile

Visualizations

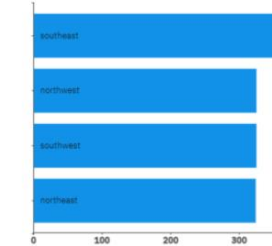
FREQUENCY



FREQUENCY



FREQUENCY



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

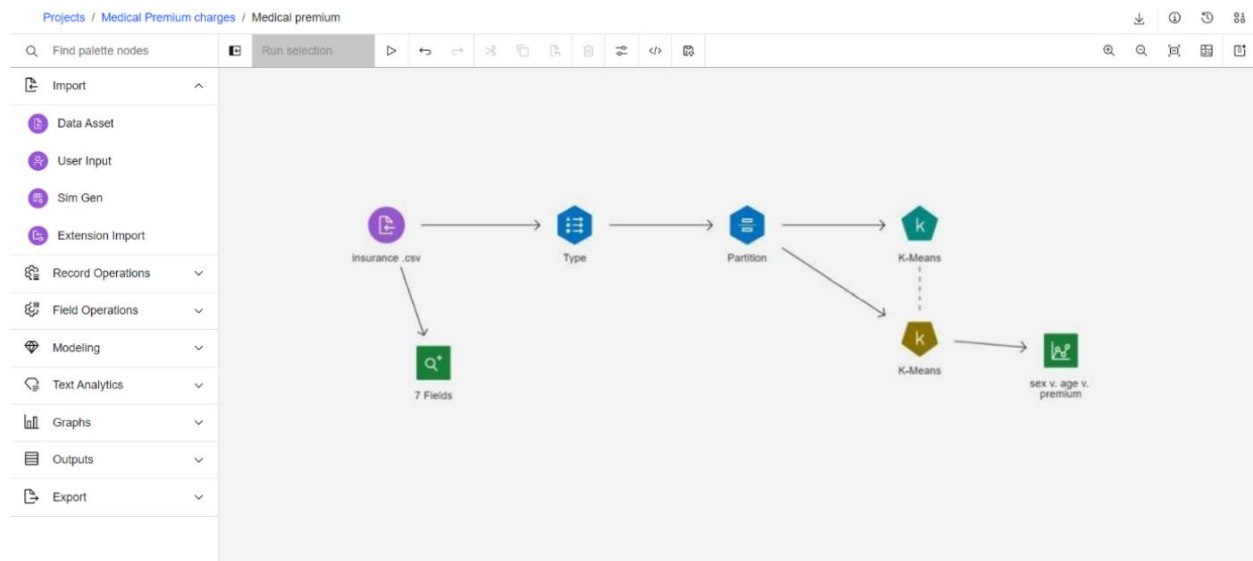
Enter a description of the Data Refinery flow

STEPS

1

DATA REFINERY FLOW OUTPUT



Location



View Output: Data Audit of [7 fields]

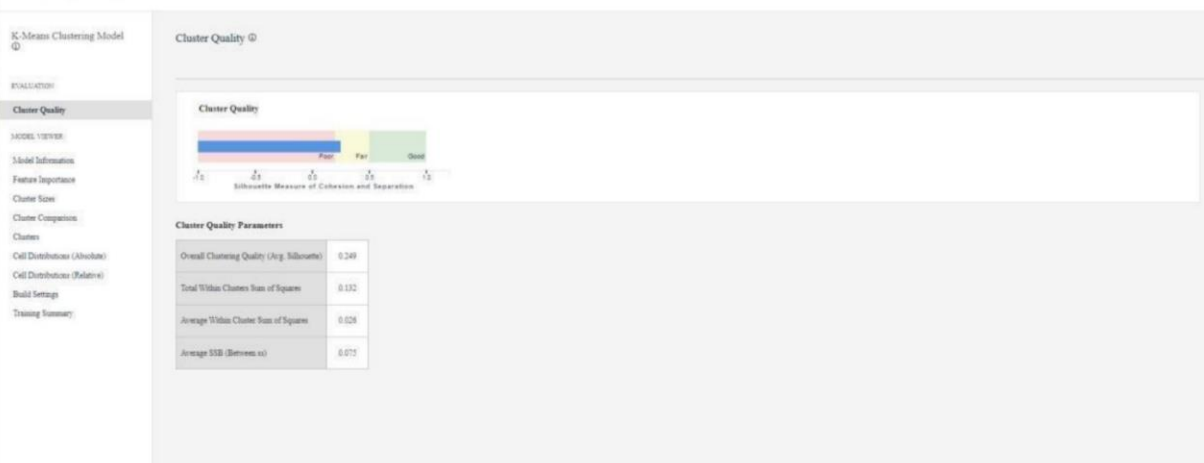
	Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	age		Continuous	18	64	39.207	14.050	0.056	--	1338
2	sex		Categorical	--	--	--	--	--	2	1338
3	bmi		Continuous	15.960	53.130	30.663	6.098	0.284	--	1338
4	children		Continuous	0	5	1.095	1.205	0.938	--	1338
5	smoker		Categorical	--	--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

View Output: Data Audit of [7 fields]

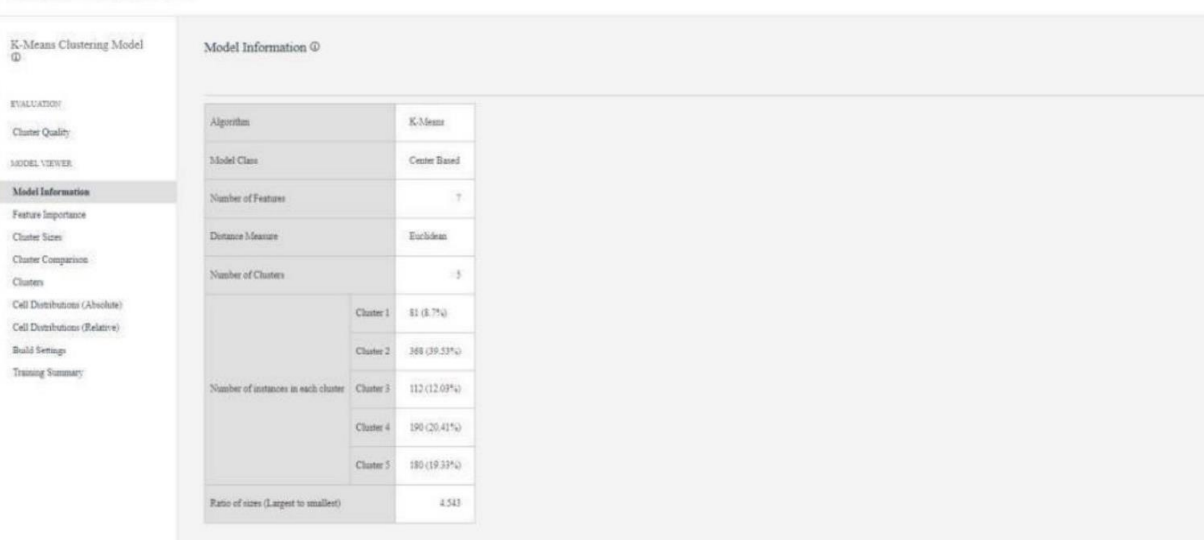
6	region		Categorical	--	--	--	--	--	4	1338
7	premium		Continuous	1121.874	63770.428	13270.422	12110.011	1.516	--	1338

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

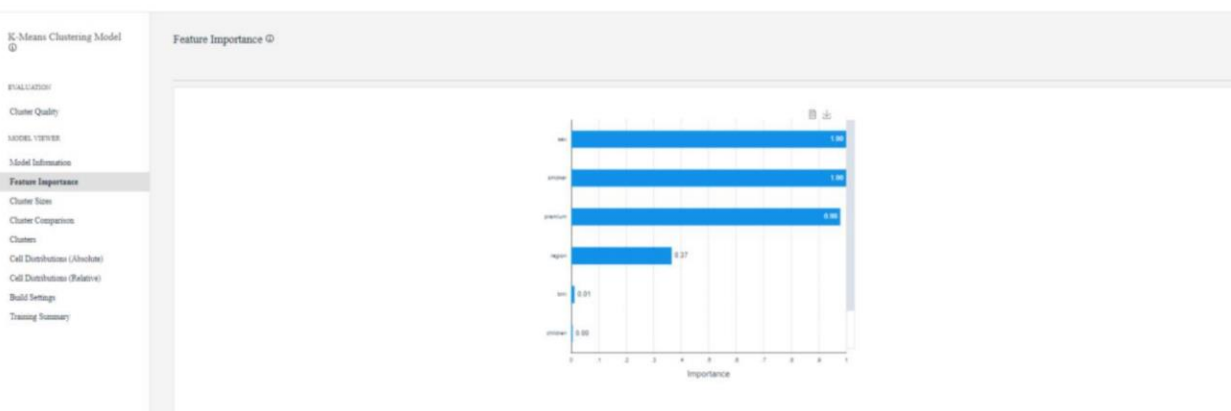
View Model: K-Means



View Model: K-Means



View Model: K-Means



View Model: K-Means

K-Means Clustering Model
Q

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 Q



View Model: K-Means

K-Means Clustering Model
Q

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

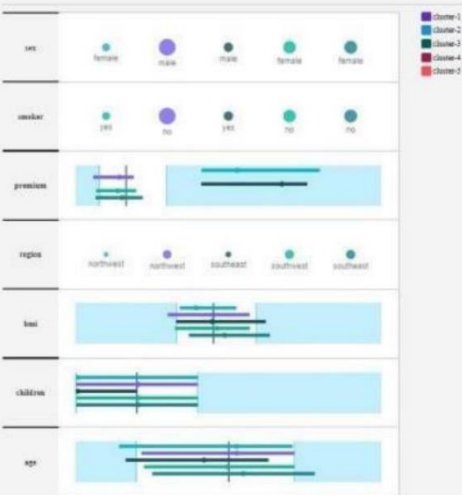
Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Cluster Comparison Q



View Model: K-Means

K-Means Clustering Model
Q

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

Cell Distributions (Absolute)

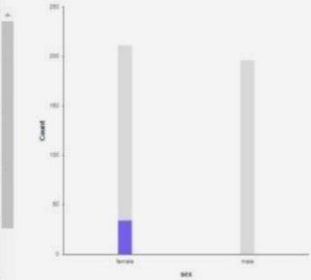
Cell Distributions (Relative)

Build Settings

Training Summary

Clusters Q

Cluster	Signal Importance				
	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5
Size	1	1	1	1	1
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%)
	premium: 30921.88	premium: 8866.42	premium: 31714.88	premium: 8181.78	premium: 8842.23
	region: northwest (35.29%)	region: northwest (27.62%)	region: southeast (38.30%)	region: southeast (53.69%)	region: southeast (51.65%)
	loss: 29.96	loss: 30.52	loss: 30.70	loss: 30.39	loss: 31.77



View Model: K-Means

K-Means Clustering Model
🔍

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparisons

Clusters

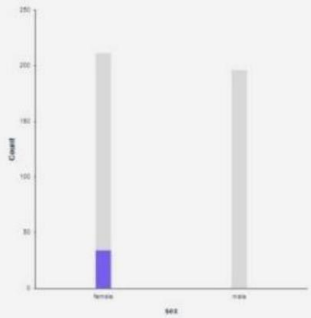
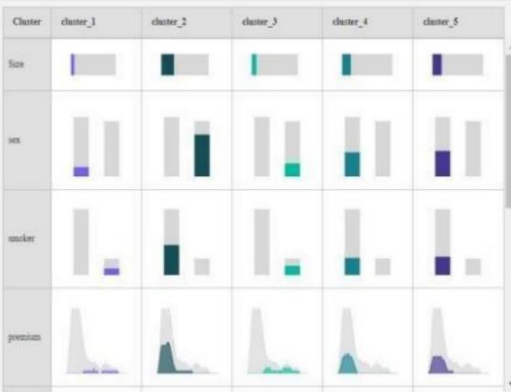
Cell Distributions (Absolute)

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 Comparisons

Clusters

Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Cell Distributions (Relative) 🔍

