

DATA ANALYTICS

ASSIGNMENT – 4

CLUSTERING MEDICAL PREMIUM CHARGES

NAME : Harendhar K

Reg no : 19BEE0140

Campus: 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

1. Convert column type ⓘ

Details Help

Edit ↗

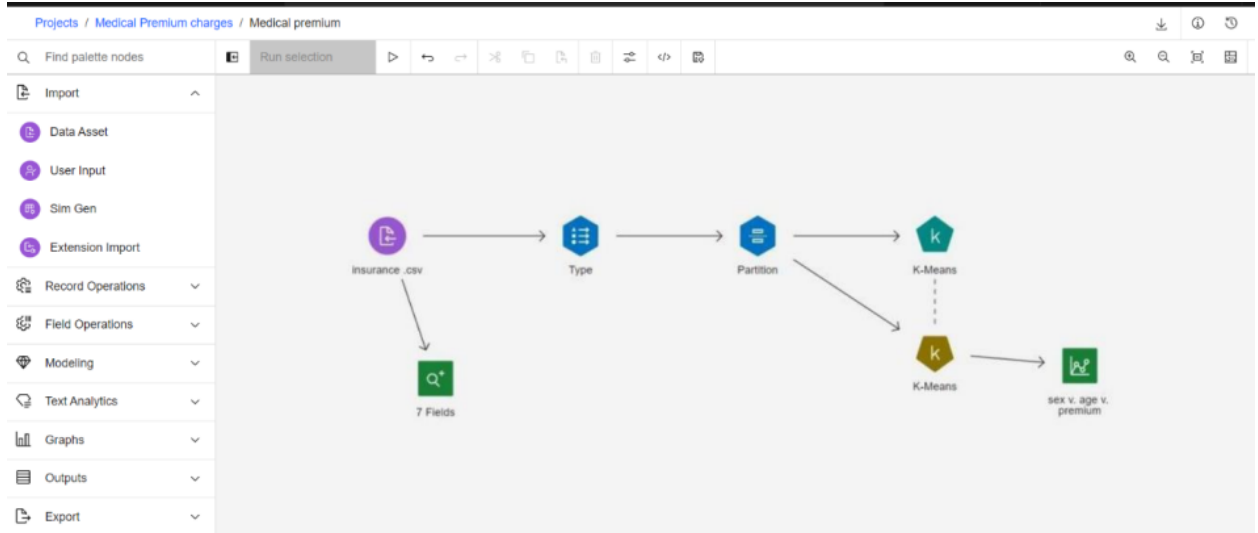
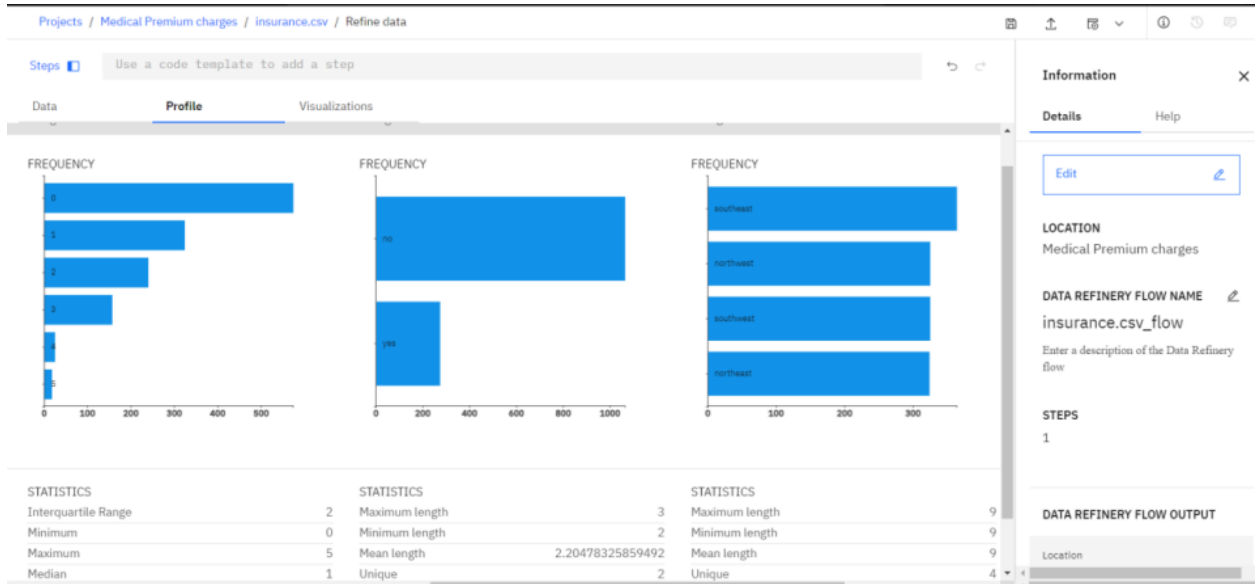
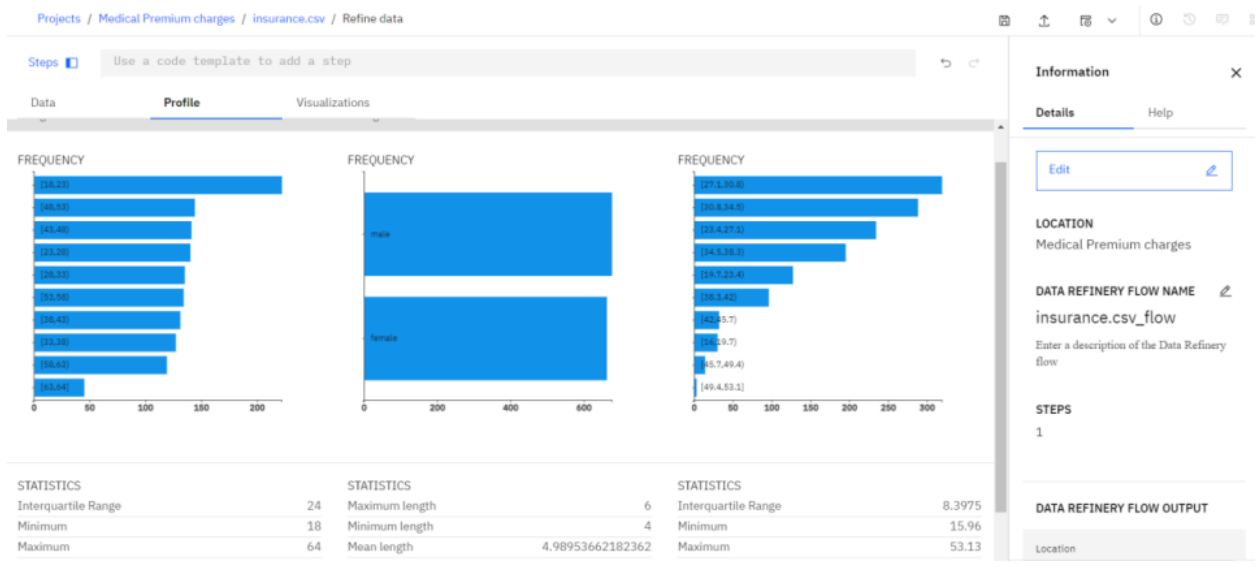
LOCATION
Medical Premium charges

DATA REFINERY FLOW NAME ↗
insurance.csv_flow
Enter a description of the Data Refinery flow







STEPS
1

DATA REFINERY FLOW OUTPUT

| | 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 |





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]

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 Size

Cluster Comparison

Clusters

Cell Distributions (Absolute)

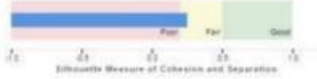
Cell Distributions (Relative)

Build Settings

Training Summary

Cluster Quality

Cluster Quality



Cluster Quality Parameters

| | |
|--|-------|
| Overall Clustering Quality (Avg. Silhouette) | 0.349 |
| Total Within Cluster Sum of Squares | 0.132 |
| Average Within Cluster Sum of Squares | 0.028 |
| Average SSB (Between, ss) | 0.071 |

View Model: K-Means

K-Means Clustering Model

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Size

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 | 81 (8.7%) |
| | Cluster 2 | 368 (39.57%) |
| | Cluster 3 | 112 (12.07%) |
| | Cluster 4 | 190 (20.41%) |
| | Cluster 5 | 180 (19.33%) |
| Ratio of sizes (Largest to smallest) | | 4.543 |

View Model: K-Means

K-Means Clustering Model

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparisons

Clusters

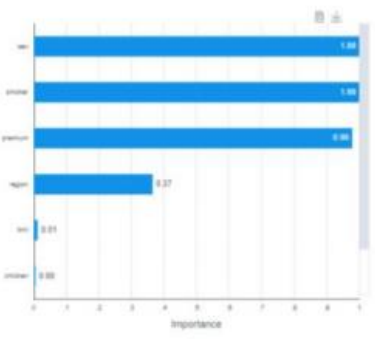
Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Feature Importance



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

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



View Model: K-Means

K-Means Clustering Model

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

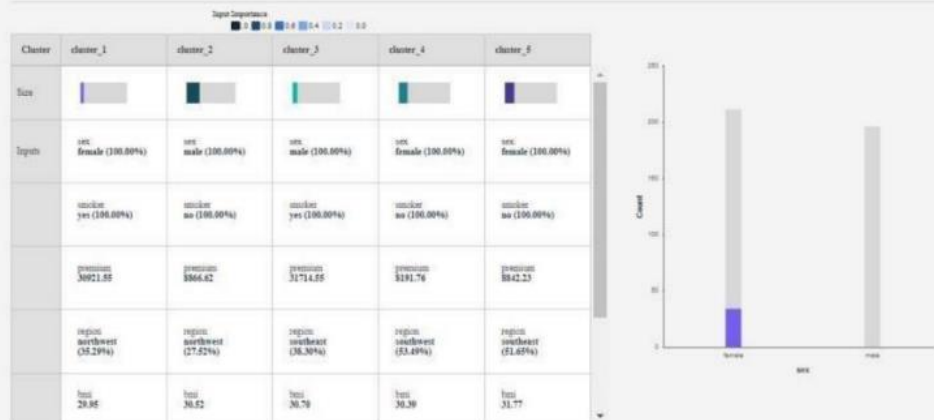
Cell Distributions (Absolute)

Cell Distributions (Relative)

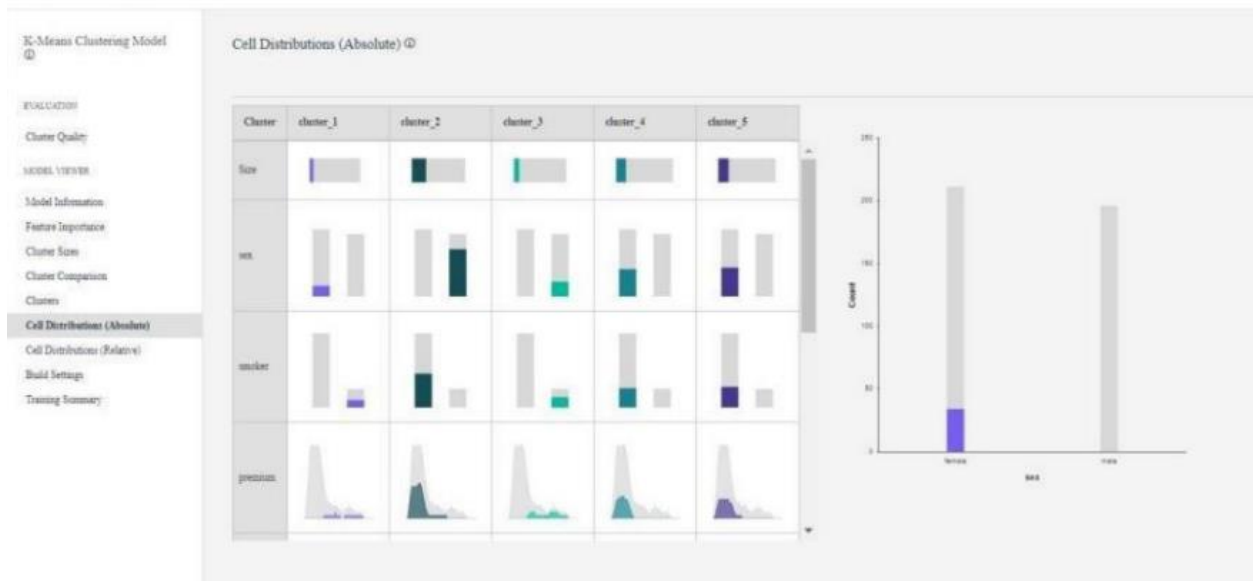
Build Settings

Training Summary

Clusters



View Model: K-Means



View Model: K-Means



