Data assets

New Data asset +

0 assets selected.

Name	Type	Created by	Last modified
CSV insurance .csv	Data Asset	Krish gupta	Dec 22, 2021, 11:30 PM

Models

Watson Machine Learning models

Name	Type	Software specification	Last modified
Medical Premium Charges	spss-modeler_18.2	spss-modeler_18.2	Dec 23, 2021

Modeler flows

New Modeler flow +

Name	Type	Created by	Last modified
Medical Premium	SPSS Modeler	Krish gupta	Dec 23, 2021, 12:15 AM

Models

Watson Machine Learning models

Name	Type	Software specification	Last modified
Medical Premium Charges	spss-modeler_18.2	spss-modeler_18.2	Dec 23, 2021

Modeler flows

New Modeler flow +

Name	Type	Created by	Last modified
Medical Premium	SPSS Modeler	Krish gupta	Dec 23, 2021, 12:15 AM

Data Refinery flows

New Data Refinery flow +

Name	Type	Created by	Last modified
insurance .csv_flow	Data Refinery flow	Krish gupta	Dec 22, 2021, 11:35 PM

Schema: 7 Columns

Preview: First 1000 rows

Last refresh: 13 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
56	female	39.82	0	no	southeast	11090.7178

Medical Premium Charges

Promote to deployment space

Overview | Activities

Input Schema

Input

Column	Type
age	"integer"
bmi	"double"
children	"integer"
premium	"double"
region	"string"
sex	"string"
smoker	"string"

Output

Column	Type
Partition	"string"
age	"integer"
bmi	"double"
children	"integer"
premium	"double"
region	"string"
sex	"string"
smoker	"string"

Medical Premium Charges

Last modified at Dec 28, 2021 8:27 PM

Description
No description provided.

Created
Dec 23, 2021 12:15 AM

Type
spss-modeler_18.2

Model ID
d49e0763-2606-4f84-9685-65...

Software specification
spss-modeler_18.2

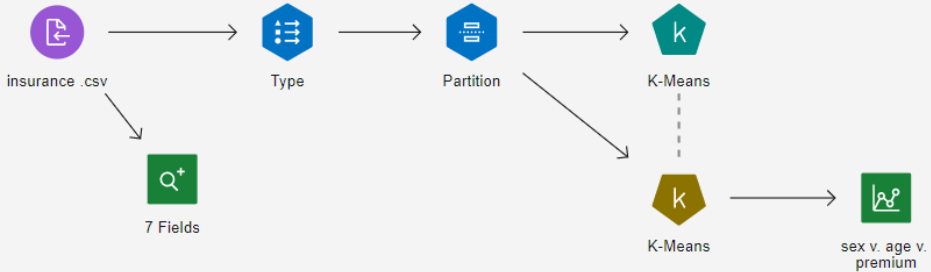
Hybrid pipeline software specifications

Tags
Add tags to make assets easier to find.

Find palette nodes

Run

- Import
- Record Operations
- Field Operations
- Modeling
- Text Analytics
- Graphs
- Outputs
- Export



Type

Type

Read Values | Clear All Values

Find in column Field

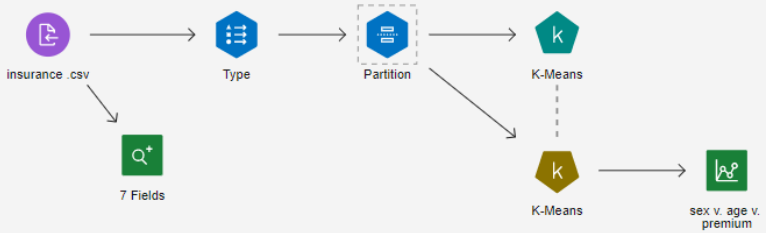
	Field	Measure	Role	Value mode	Values	Check
☐	age	Continuous	Input	Instantiated	18, 64	None
☐	sex	Flag	Input	Instantiated	female, male	None
☐	bmi	Continuous	Input	Instantiated	15.96, 53.13	None
☐	children	Continuous	Input	Instantiated	0, 5	None
☐	smoker	Flag	Input	Instantiated	no, yes	None
☐	region	Nominal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

Cancel | Save

Find palette nodes

Run

- Import
- Record Operations
- Field Operations
- Modeling
- Text Analytics
- Graphs
- Outputs
- Export



Partition

Partition

Settings

Derived Field Name

Partition

Training Partition(%)

70

Testing Partition(%)

30

Create validation partition

Repeatable partition assignment

Seed

Cancel | Save

View Output: Data Audit of [7 fields]

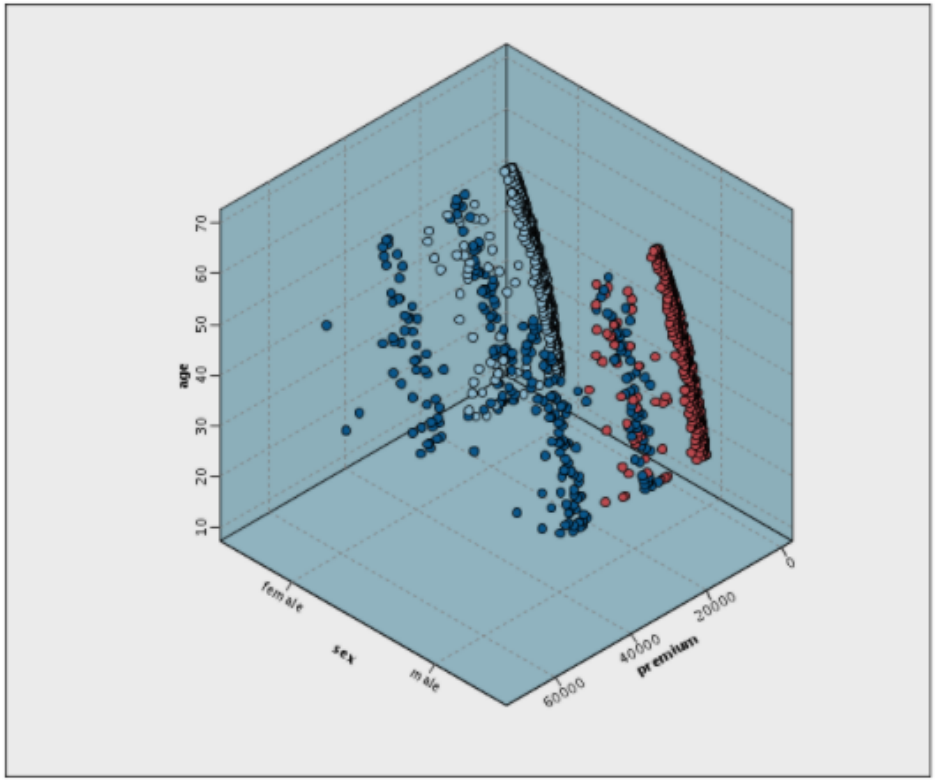
	Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	age		Continuous	18	64	39.207	14.050	0.056	--	1338
2	sex		Categorical	--	--	--	--	--	2	1338
3	bmi		Continuous	15.960	53.130	30.663	6.098	0.284	--	1338
4	children		Continuous	0	5	1.095	1.205	0.938	--	1338
5	smoker		Categorical	--	--	--	--	--	2	1338
6	region		Categorical	--	--	--	--	--	4	1338
7	premium		Continuous	1121.874	63770.428	13270.422	12110.011	1.516	--	1338

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	White Space	Bl
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

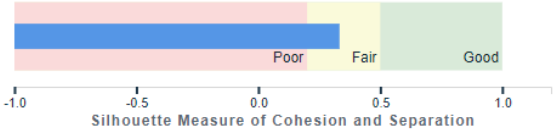
Feature Importance

Cluster Sizes

Cluster Comparison

Cluster Quality ⓘ

Cluster Quality



Cluster Quality Parameters

Cluster Quality ⓘ

Cluster Quality Parameters

Overall Clustering Quality (Avg. Silhouette)	0.331
Total Within Clusters Sum of Squares	0.143
Average Within Cluster Sum of Squares	0.048
Average SSB (Between ss)	0.064

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		3
Number of instances in each cluster	Cluster 1	370 (39.74%)
	Cluster 2	368 (39.53%)
	Cluster 3	193 (20.73%)
Ratio of sizes (Largest to smallest)		1.917

View Model: K-Means

K-Means Clustering Model

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

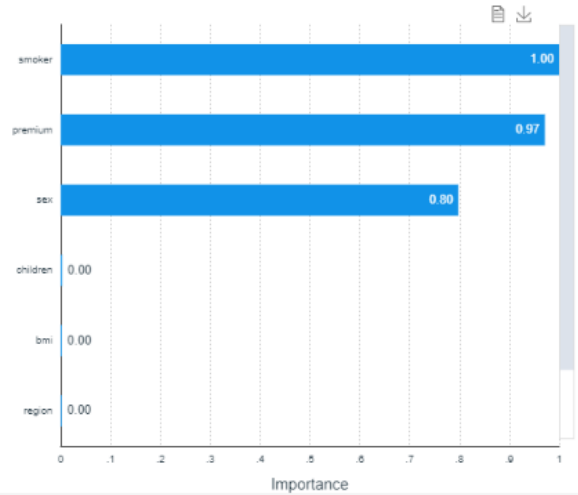
Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Feature Importance



View Model: K-Means

K-Means Clustering Model

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

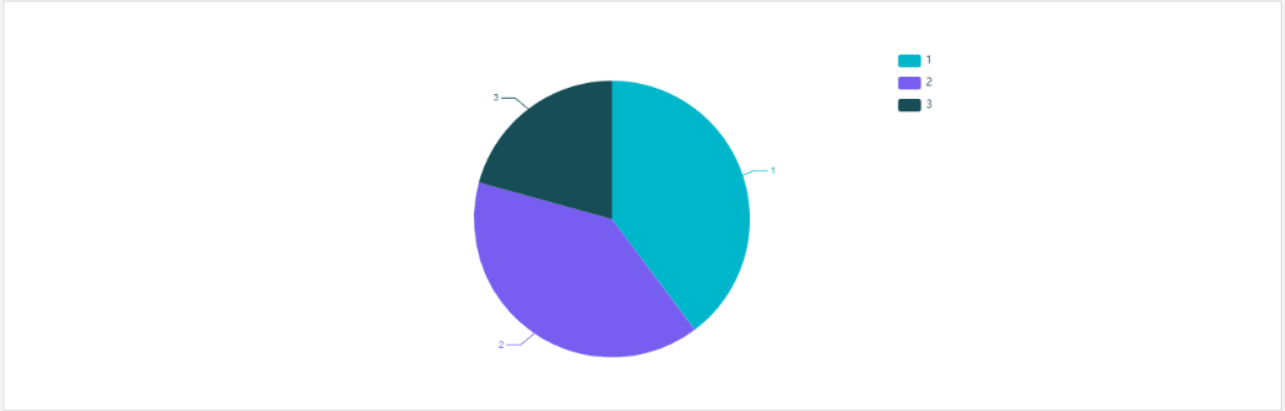
Cell Distributions (Absolute)

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

View Model: K-Means

K-Means Clustering Model

EVALUATION

Cluster Quality

MODEL VIEWER

Model Information

Feature Importance

Cluster Sizes

Cluster Comparison

Clusters

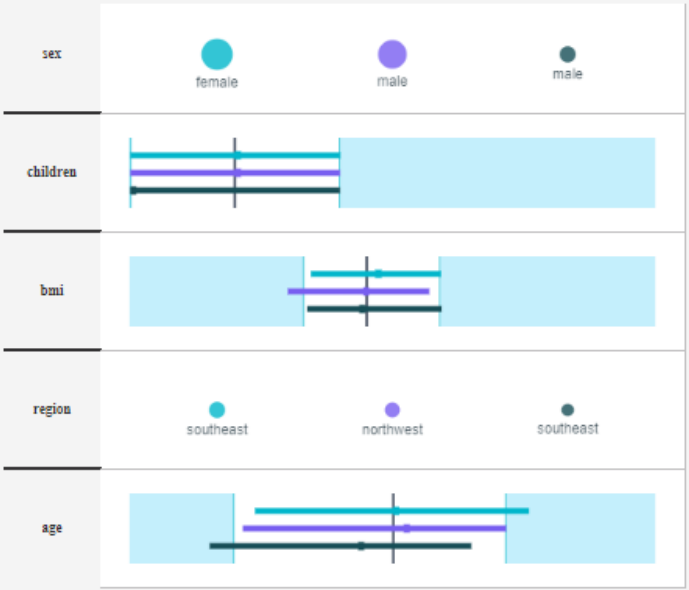
Cell Distributions (Absolute)

Cell Distributions (Relative)

Build Settings

Training Summary

Cluster Comparison



View Model: K-Means

K-Means Clustering Model ⓘ

EVALUATION

Cluster Quality

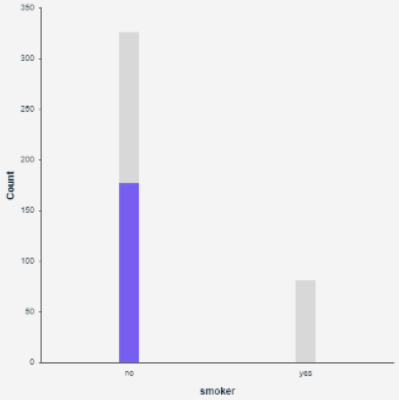
MODEL VIEWER

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

Clusters ⓘ

Input Importance
1.0 0.8 0.6 0.4 0.2 0.0

Cluster	cluster_1	cluster_2	cluster_3
Size			
Inputs	smoker **no (100.00%)**	smoker **no (100.00%)**	smoker **yes (100.00%)**
	premium **8526.18**	premium **8866.62**	premium **31381.69**
	sex **female (100.00%)**	sex **male (100.00%)**	sex **male (58.02%)**
	children **1.11**	children **1.11**	children **.83**



View Model: K-Means

K-Means Clustering Model ⓘ

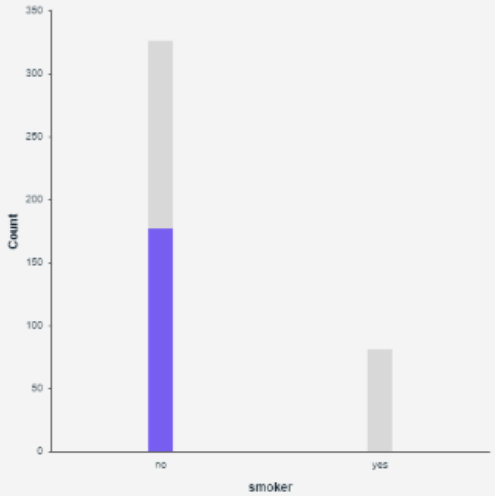
EVALUATION

- 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

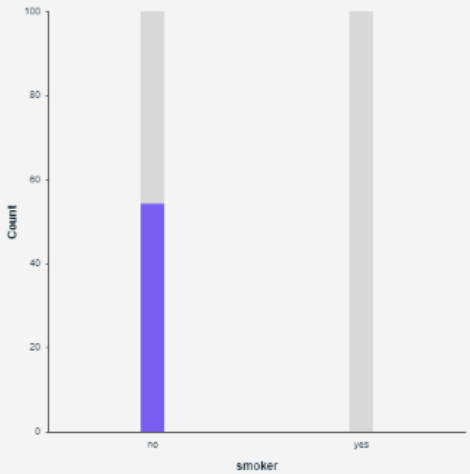
- Cluster Quality

MODEL VIEWER

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

Cell Distributions (Relative) ⓘ

Cluster	cluster_1	cluster_2	cluster_3
Size			
smoker			
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

Build Settings

Use partitioned data	true
Calculate raw propensity scores	false
Calculate adjusted propensity scores	false
Number of clusters	3
Generate distance field	false
Cluster label	String
Label prefix	cluster
Optimize	Memory
Mode	Simple

View Model: K-Means

K-Means Clustering Model

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 15:05:27 UTC 2021
Elapsed time for model build	0 hours, 0 mins, 0 secs

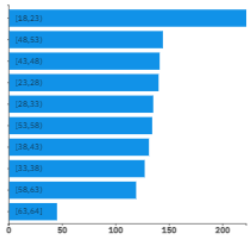
	age Integer	sex String	bmi Decimal	children Integer	smoker String	region String	premium Decimal
1	19	female	27.9	0	yes	southwest	16884.924
2	18	male	33.77	1	no	southeast	1725.5523
3	28	male	33	3	no	southeast	4449.462
4	33	male	22.705	0	no	northwest	21984.47061
5	32	male	28.88	0	no	northwest	3866.8552
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

Steps | Use a code template to add a step

Data | **Profile** | Visualizations

age
Integer

FREQUENCY

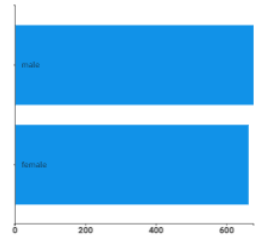


STATISTICS

Interquartile Range	24
Minimum	18
Maximum	64
Median	39
Standard Deviation	14.0499603792162

sex
String

FREQUENCY

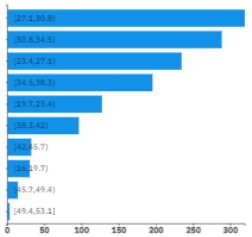


STATISTICS

Maximum length	6
Minimum length	4
Mean length	4.98953662182362
Unique	2

bmi
Decimal

FREQUENCY

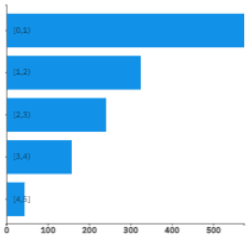


STATISTICS

Interquartile Range	8.3975
Minimum	15.96
Maximum	53.13
Median	30.4
Standard Deviation	6.09818691167902

children
Integer

FREQUENCY

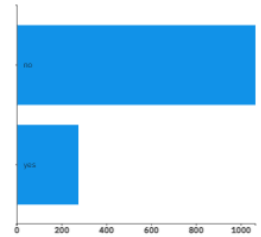


STATISTICS

Interquartile Range	2
Minimum	0
Maximum	5
Median	1
Standard Deviation	1.20549273978191

smoker
String

FREQUENCY

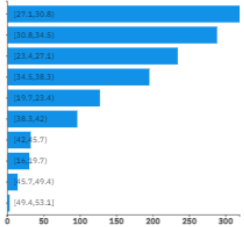


STATISTICS

Maximum length	3
Minimum length	2
Mean length	2.20478325859492
Unique	2

bmi
Decimal

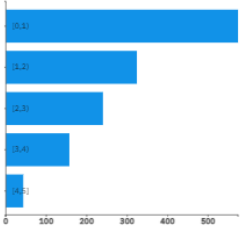
FREQUENCY



STATISTICS	
Interquartile Range	8.3975
Minimum	15.96
Maximum	53.13
Median	30.4
Standard Deviation	6.09818691167902

children
Integer

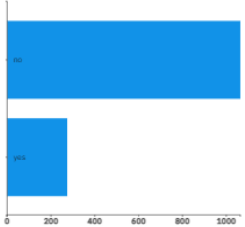
FREQUENCY



STATISTICS	
Interquartile Range	2
Minimum	0
Maximum	5
Median	1
Standard Deviation	1.20549273978191

smoker
String

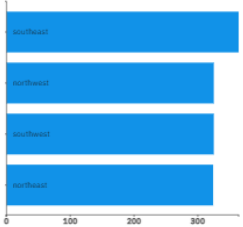
FREQUENCY



STATISTICS	
Maximum length	3
Minimum length	2
Mean length	2.20478325859492
Unique	2

region
String

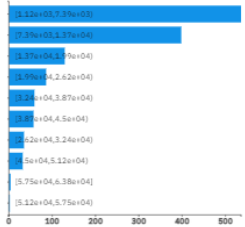
FREQUENCY



STATISTICS	
Maximum length	9
Minimum length	9
Mean length	9
Unique	4

premium
Decimal

FREQUENCY



STATISTICS	
Interquartile Range	11899.625365
Minimum	1121.8739
Maximum	63770.42801
Median	9382.033
Standard Deviation	12110.011236694

Steps | Use a code template to add a step

Data | Profile | Visualizations

SCATTER PLOT CHART

CHART TYPE: Scatter plot | Line | Multi-series | Histogram | Population ... | Q-Q plot | Pie | Bar | Parallel | Relationship | Box plot | Heat map | t-SNE | Word cloud | Map | Sunburst

ACTIONS

X-axis* age

Y-axis* region

Color map: None

Size map: None

Shape map: None

Fit line: None

Gradient Bubble: Off

Bubble Size: 12