

Assignment-4

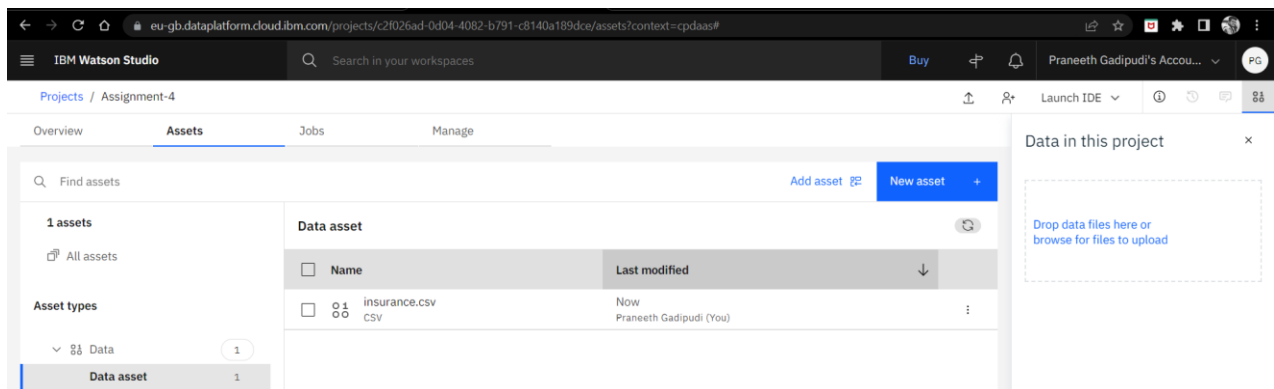
Name: G. Naga Sai Praneeth

Regd:19MIM10044

Clustering: Analysis of Medical Premium Charges for Insurance

Dataset: insurance.csv

Adding the dataset to the project

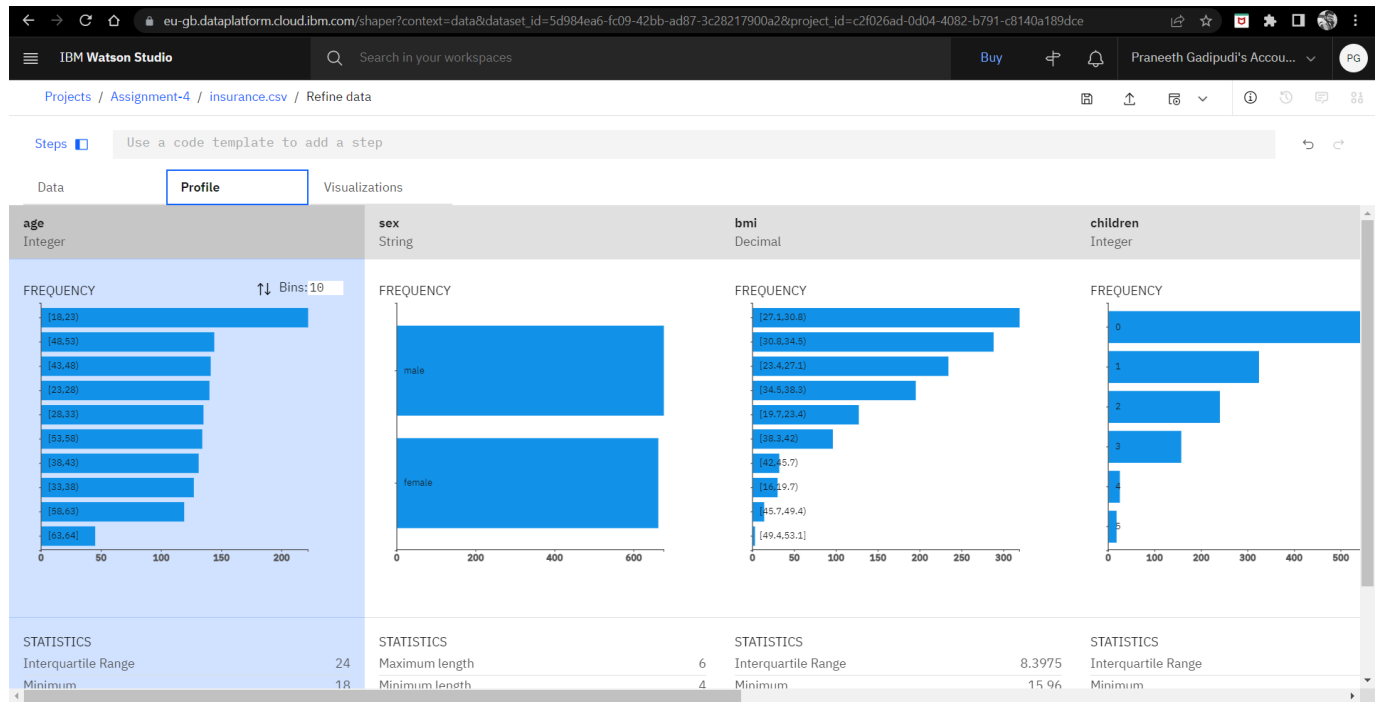


Data Refinery

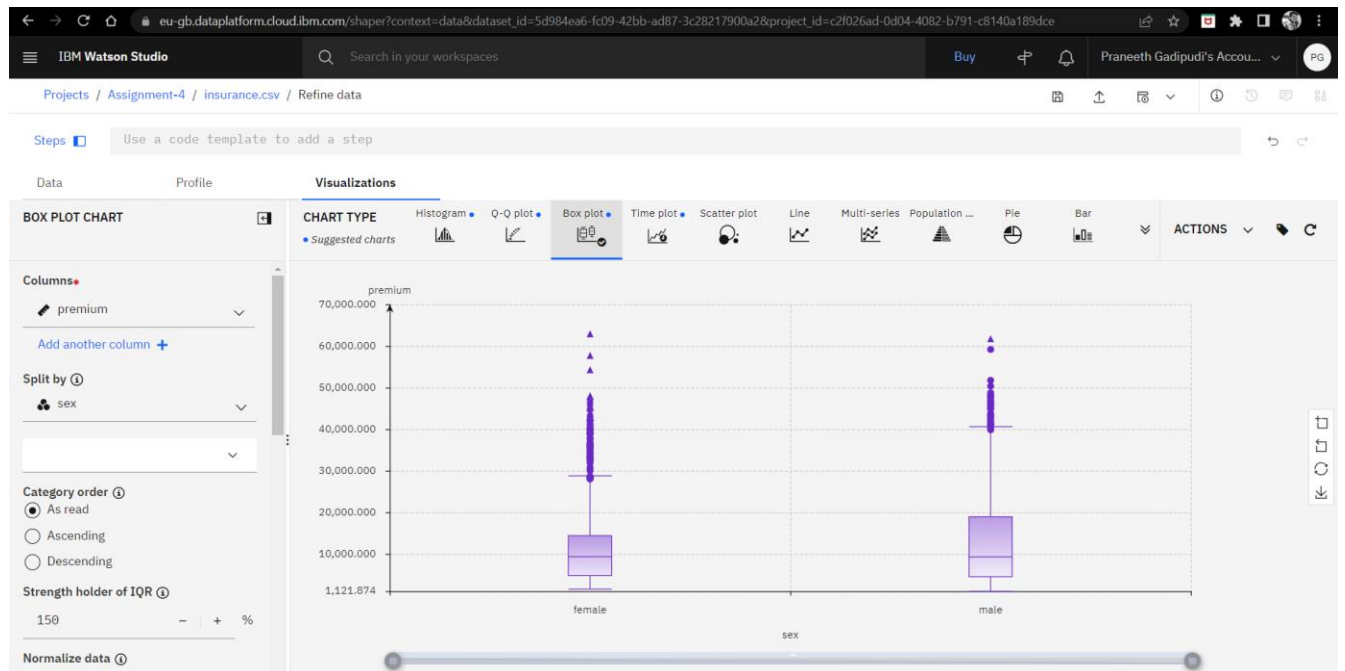
The screenshot shows the IBM Watson Studio interface with the 'Data Refinery' tab active. The 'insurance.csv' dataset is selected, and the 'Refine data' view is active. The table displays the first 50 rows of the dataset, including columns for age, sex, bmi, children, smoker, region, and premium. The 'Details' panel on the right shows the 'LOCATION' as 'Assignment-4' and the 'DATA REFINERY FLOW NAME' as 'insurance.csv_flow'. The 'STEPS' section shows 0 steps, and the 'DATA REFINERY FLOW OUTPUT' section shows the 'LOCATION' as 'Assignment-4'.

	age	sex	bmi	children	smoker	region	premium
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

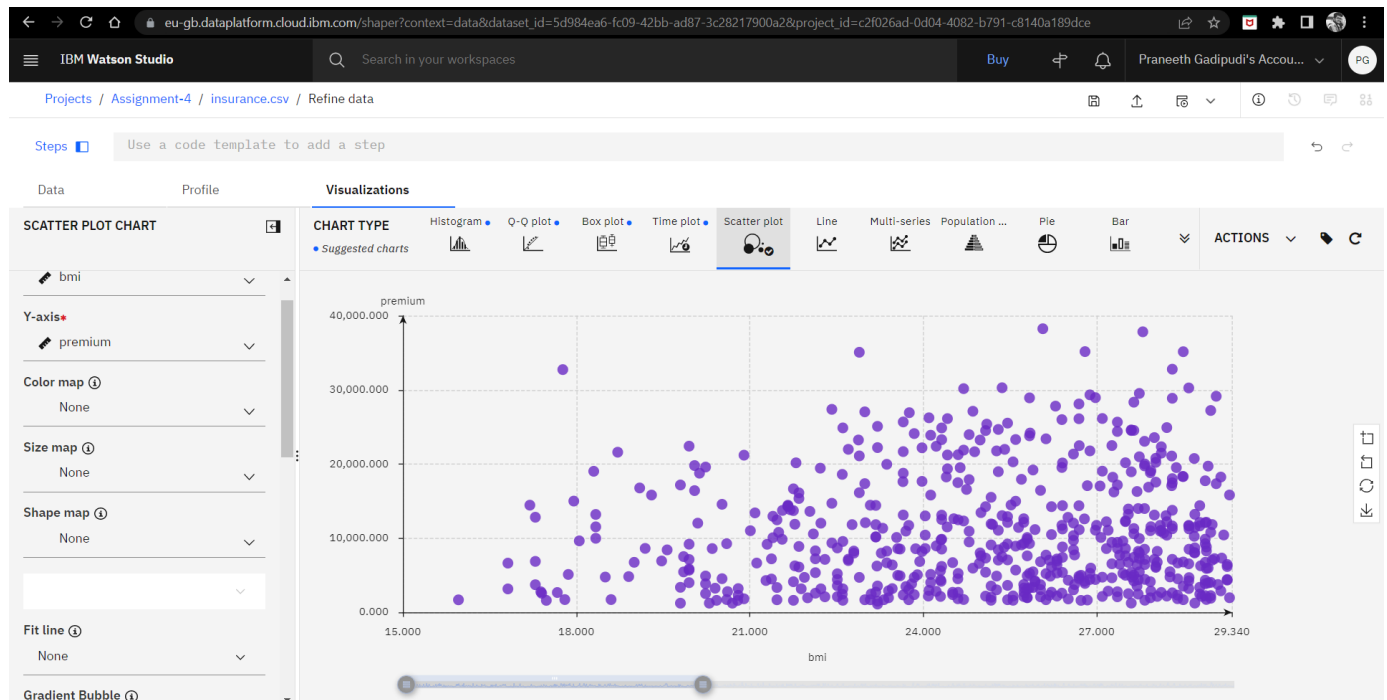
Profile:



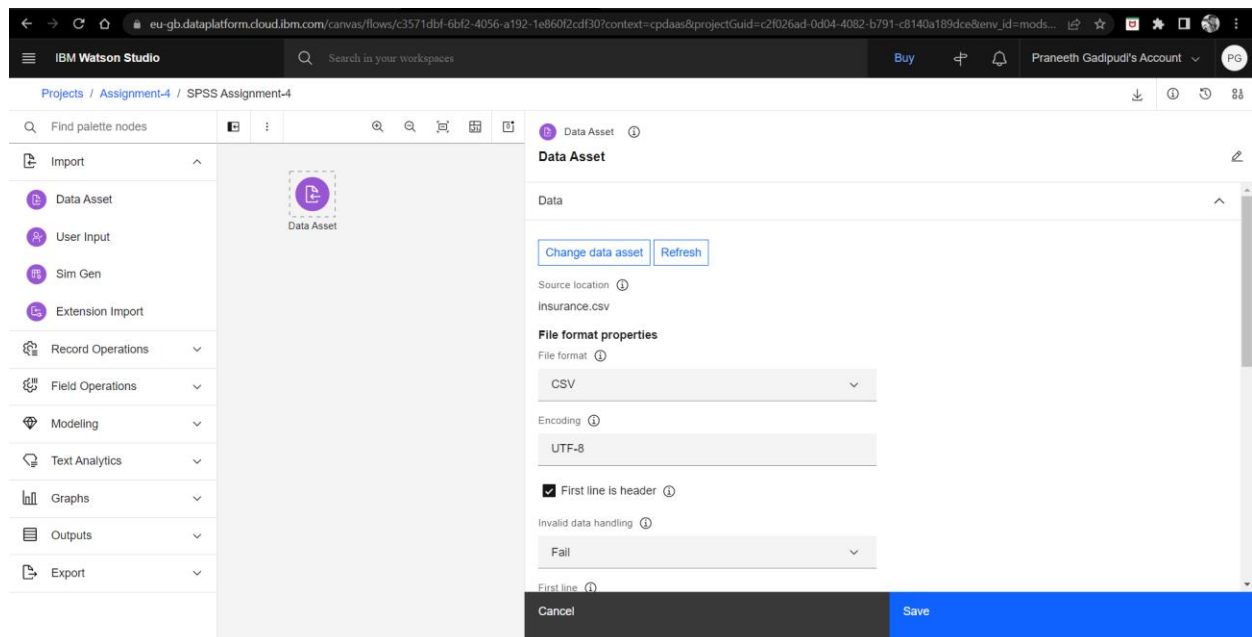
Visualization: Viewing Boxplot of Sex vs Premium



Scatterplot of BMI vs Premium:



Creating an SPSS Modeler and uploading the dataset:



Data Audit:

The screenshot shows the IBM Watson Studio interface. On the left, a sidebar lists various tools: Field Operations, Modeling, Text Analytics, Graphs, Outputs, Table, and Matrix. The main workspace displays a 'Data Asset' icon with a downward arrow pointing to a green '7 Fields' icon. A 'Run selection' button is visible above the workspace. On the right, a 'Messages' panel shows a 'Success' message: 'Run was successful'. Below it, a 'Clear all' link is present. Further right, an 'Outputs' panel shows '7 Fields' as the output, with a status of 'Just now'.

Data Audit output:

The screenshot shows the 'View Output: 7 Fields' window. It displays a table with 10 columns: Field, Graph, Measurement, Min, Max, Mean, Std. Dev, Skewness, Unique, and Valid. The table contains 6 rows of data for fields: age, sex, bmi, children, smoker, and region. Each row includes a small graph icon. A 'Compare' link is visible in the top right corner of the window.

	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

Type Node:

The screenshot shows the IBM Watson Studio interface. On the left, the 'Find palette nodes' search bar is active. The 'Type' node is selected in the 'Field Operations' section. The main canvas displays a workflow: 'Data Asset' (purple icon) points to 'Type' (blue icon), which points to '7 Fields' (green icon). The right-hand panel shows the 'Type' node configuration. It includes a 'Read values' button and a 'Clear values' button. Below these is a table with columns: Field, Measure, Role, Value mode, and Values. The table lists 7 fields: # age, abc sex, # bmi, # children, abc smoker, abc region, and # premium. The 'Role' column shows 'Input' for the first five fields and 'Target' for the last two. The 'Value mode' column shows 'Read' for all fields. The 'Values' column shows a dropdown arrow for each field. Below the table, there is a 'Default mode' section with two radio buttons: 'Read metadata' (selected) and 'Pass (do not scan)'. A 'Set unique fields to None' button is also present. At the bottom of the panel are 'Cancel' and 'Save' buttons.

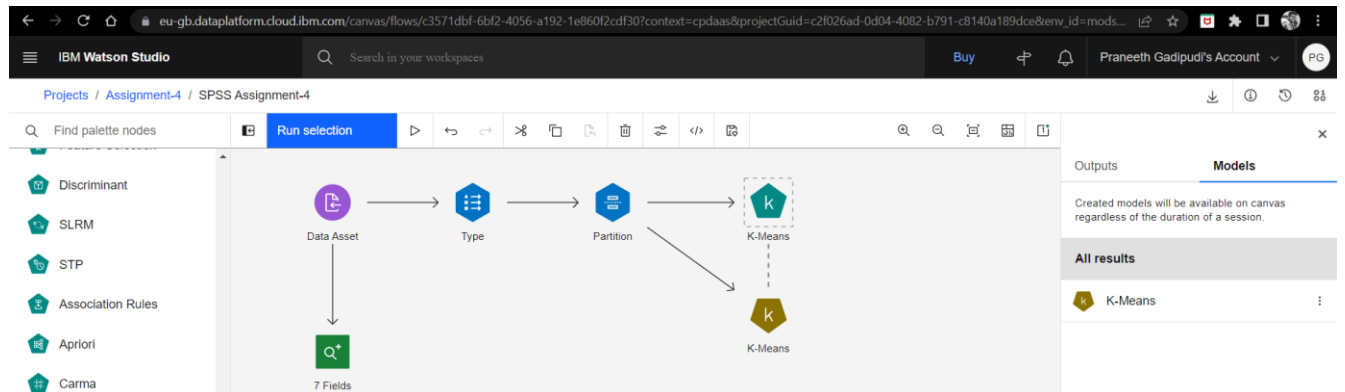
Field	Measure	Role	Value mode	Values
<input type="checkbox"/> # age	Continuous	Input	Read	
<input type="checkbox"/> abc sex	Categorical	Input	Read	
<input type="checkbox"/> # bmi	Continuous	Input	Read	
<input type="checkbox"/> # children	Continuous	Input	Read	
<input type="checkbox"/> abc smoker	Categorical	Input	Read	
<input type="checkbox"/> abc region	Categorical	Input	Read	
<input type="checkbox"/> # premium	Continuous	Target	Read	

Partition Node:

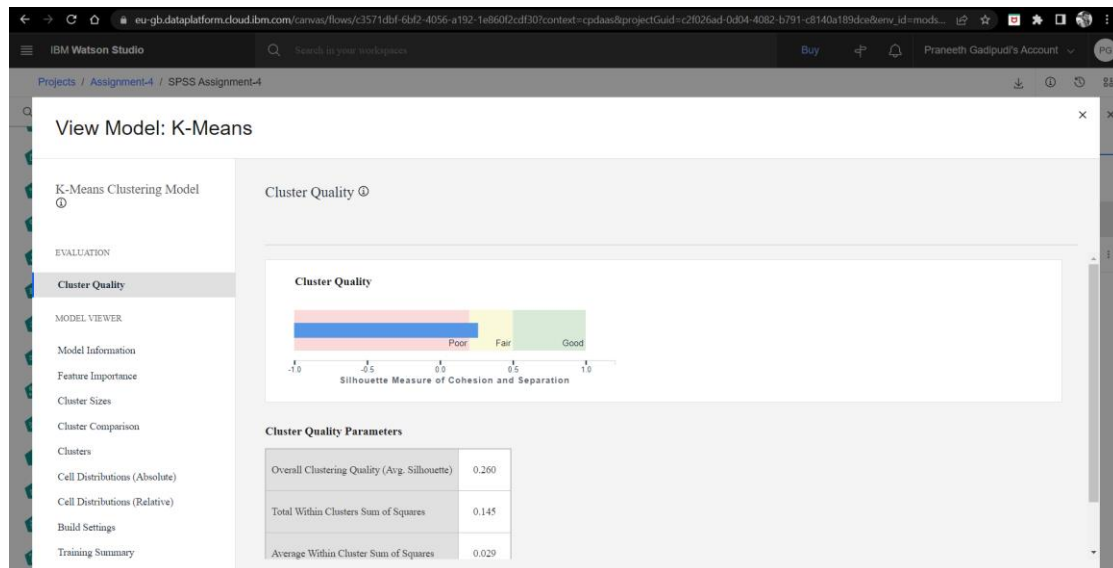
The screenshot shows the IBM Watson Studio interface. On the left, the 'Find palette nodes' search bar is active. The 'Partition' node is selected in the 'Field Operations' section. The main canvas displays a workflow: 'Data Asset' (purple icon) points to 'Type' (blue icon), which points to 'Partition' (blue icon). The right-hand panel shows the 'Partition' node configuration. It includes a 'Settings' section with a 'Derived Field Name' field containing 'Partition'. Below this is a 'Training Partition(%)' field with a value of 80. The 'Testing Partition(%)' field is highlighted with a blue border and contains a value of 20. There is a checkbox for 'Create validation partition' which is unchecked. A checkbox for 'Repeatable partition assignment' is checked. Below these is a 'Seed' field with a value of 1234567. At the bottom of the panel are 'Cancel' and 'Save' buttons.

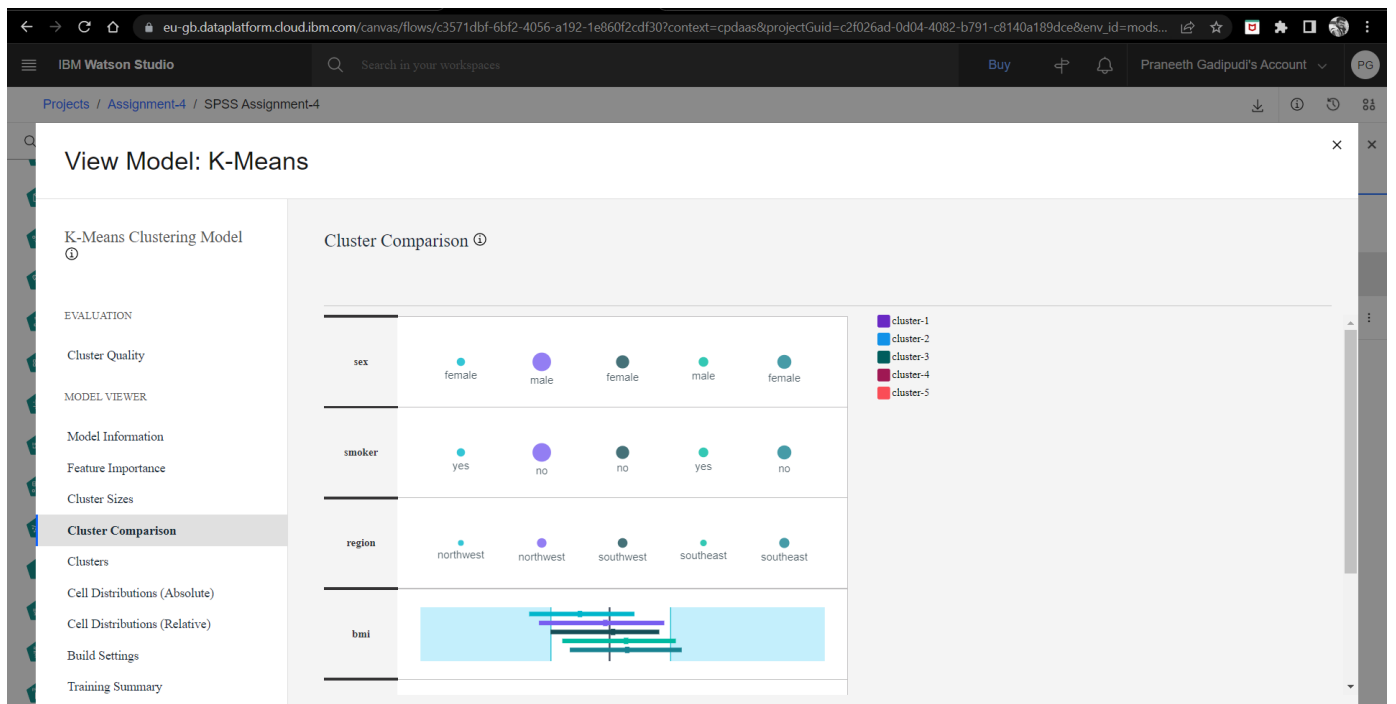
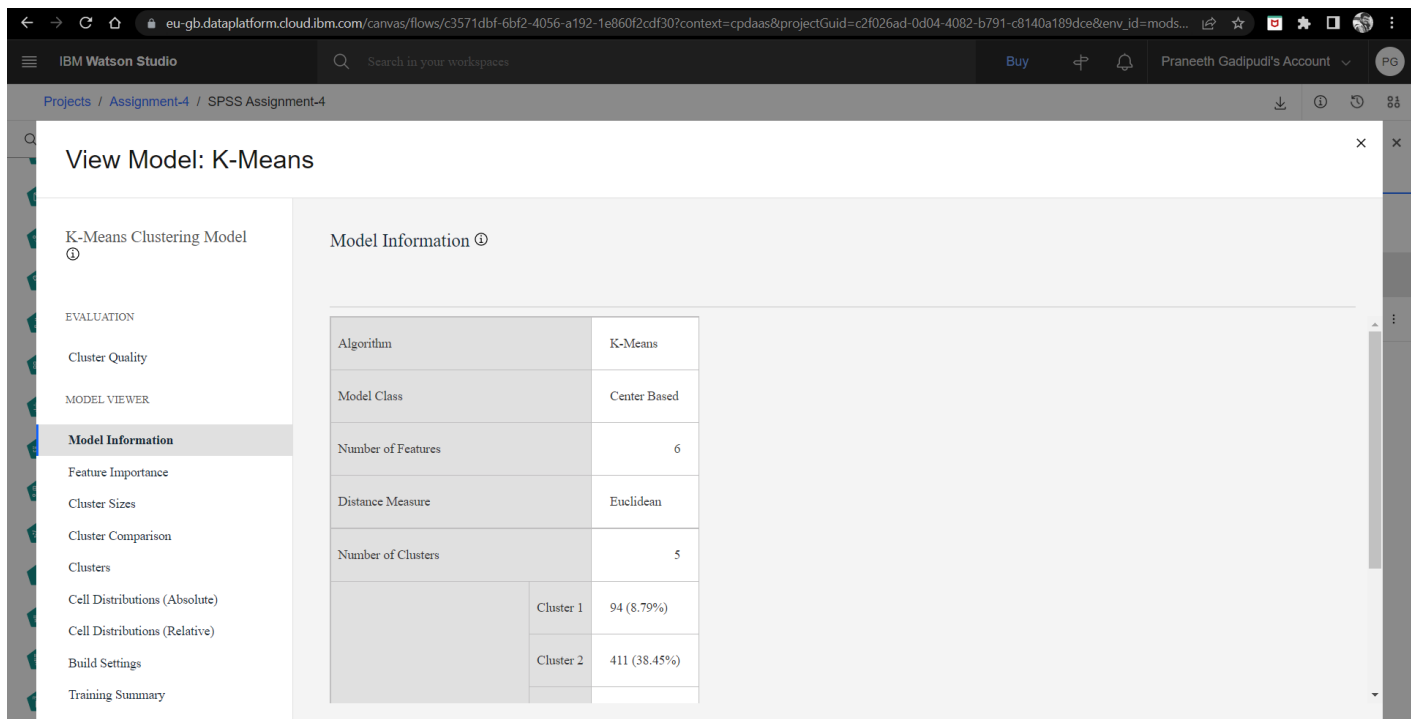
Field	Measure	Role	Value mode	Values
<input type="checkbox"/> # age	Continuous	Input	Read	
<input type="checkbox"/> abc sex	Categorical	Input	Read	
<input type="checkbox"/> # bmi	Continuous	Input	Read	
<input type="checkbox"/> # children	Continuous	Input	Read	
<input type="checkbox"/> abc smoker	Categorical	Input	Read	
<input type="checkbox"/> abc region	Categorical	Input	Read	
<input type="checkbox"/> # premium	Continuous	Target	Read	

K-Means Cluster Model:



K-means Output:





Plot Node:

IBM Watson Studio

Search in your workspaces

Buy

Praneeth Gadipudi's Account

Projects / Assignment-4 / SPSS Assignment-4

Find palette nodes

Run selection

Import

Record Operations

Field Operations

Modeling

Text Analytics

Graphs

Charts

Plot

Multiplot

Time plot

Distribution

Histogram

Collection

Data Asset

Type

Partition

K-Means

7 Fields

K-Means

Plot

3-D graph

X field

bmi

Y field

premium

Z field

age

Overlay

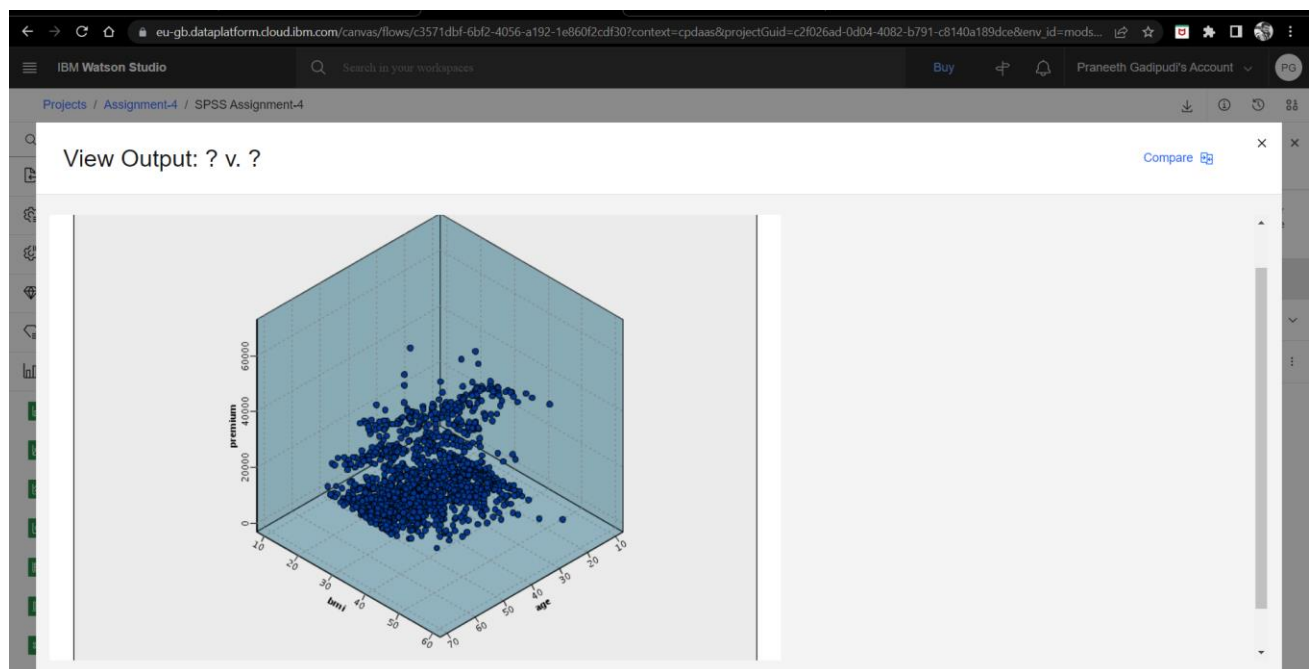
Options

Appearance

Cancel

Save

Output:



Assigning Color:

IBM Watson Studio

Search in your workspaces

Buy

Praneeth Gadipudi's Account

Projects / Assignment-4 / SPSS Assignment-4

Find palette nodes

Run selection

Import

Record Operations

Field Operations

Modeling

Text Analytics

Graphs

Charts

Plot

Multiplot

Time plot

Distribution

Histogram

Collection

Data Asset

Type

Partition

K-Means

7 Fields

K-Means

Plot

Overlay

Color

\$KM-K-Means

Size

Shape

Panel

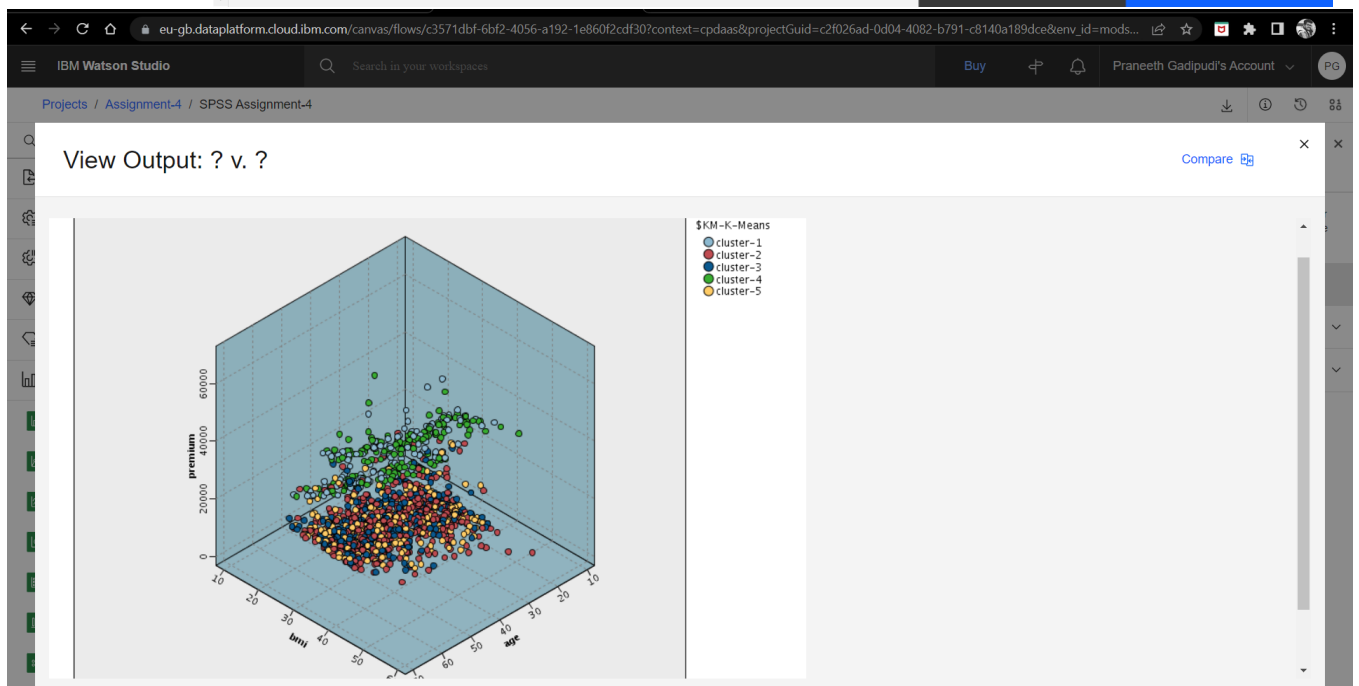
Transparency

Overlay type

None

Cancel

Save



Cluster =3:

