19BCE2647

Aadarshbhushan Singh

VIT Vellore

Smartinternz Assignment 03

Data Analytics

23rd April 2022

Dataset Name: Diabetes.csv

Link to the dashboard:

https://eu-gb.dataplatform.cloud.ibm.com/canvas/flows/40078de2-3f00-4466-a866-be2e527d2ec5?context=cpdaas&project_id=6b7b24a3-cebb-4b29-b99c-77050327294e

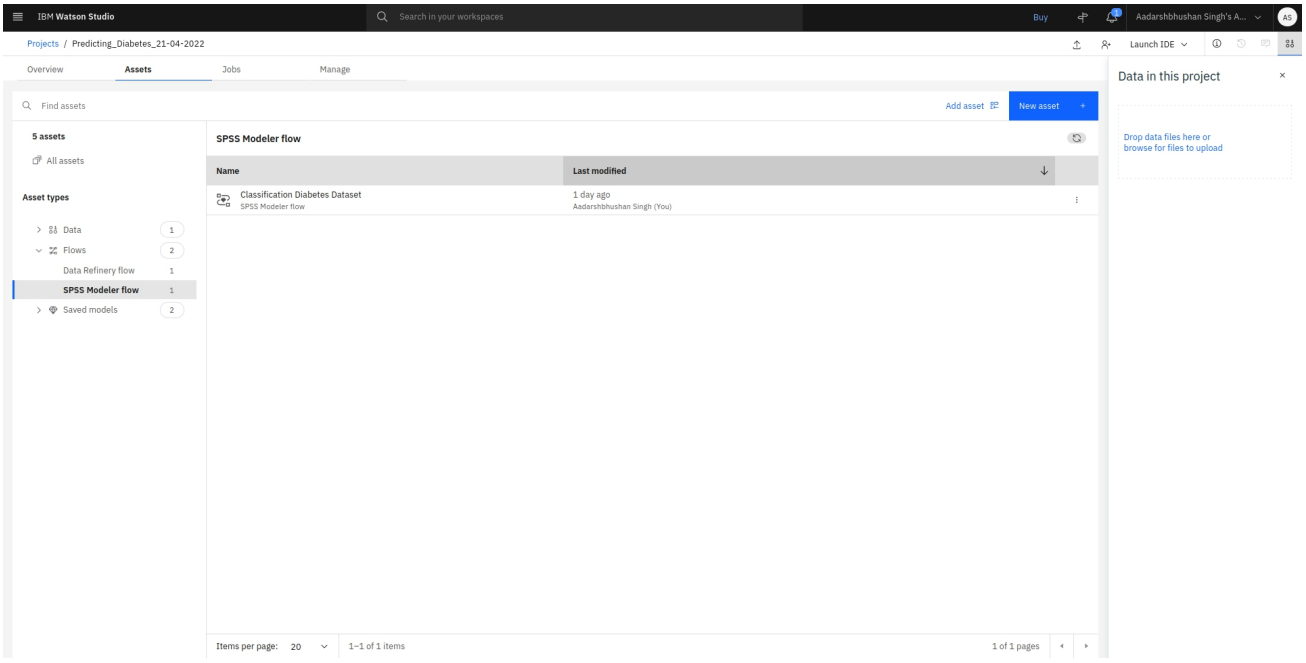


Fig: Create New Project

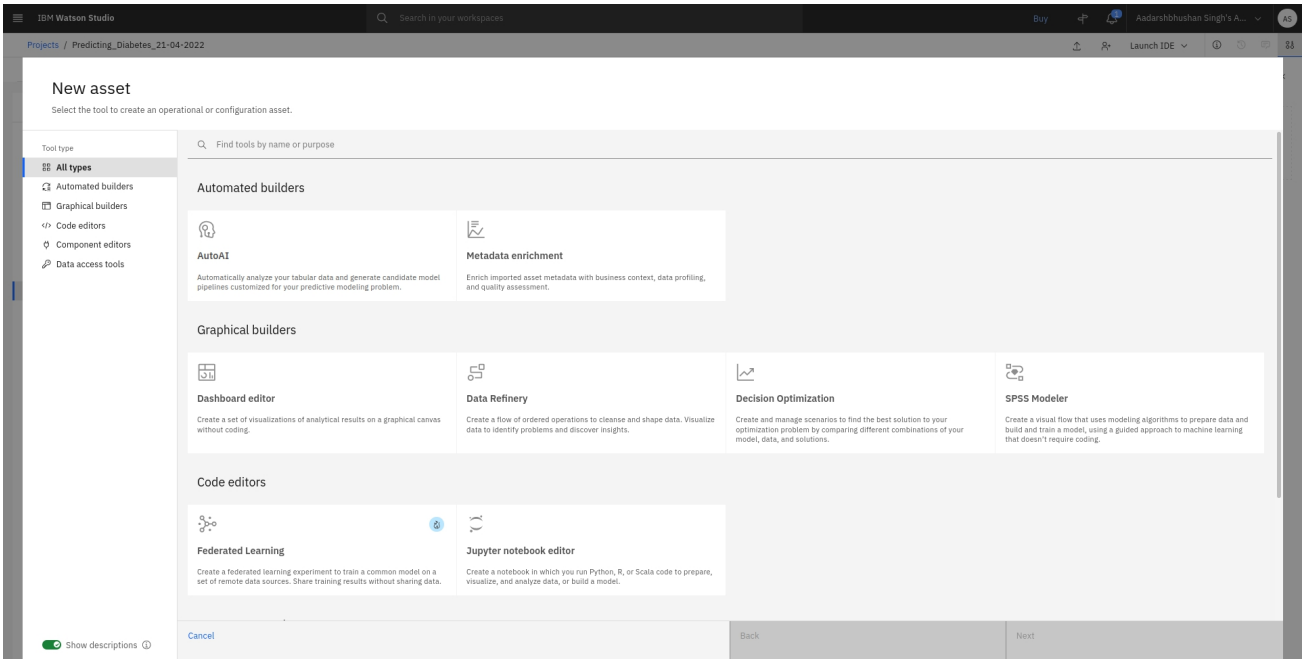


Fig: Go to Data Refinary

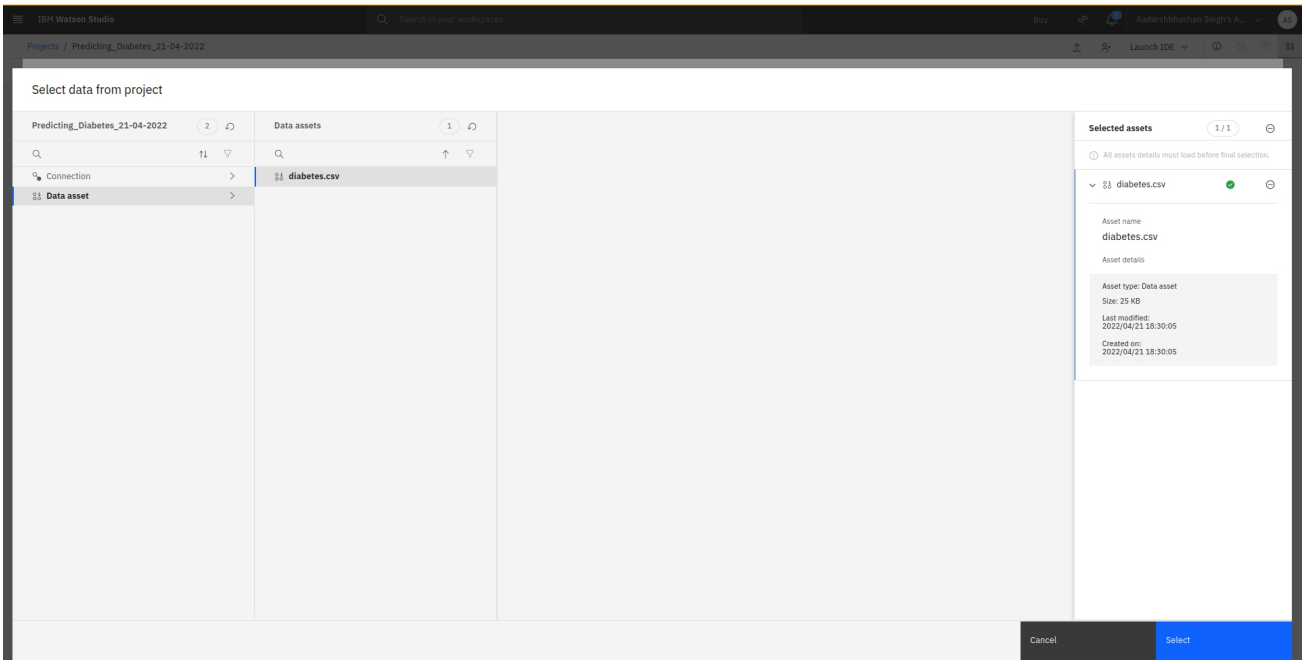


Fig: Select diabetes.csv data

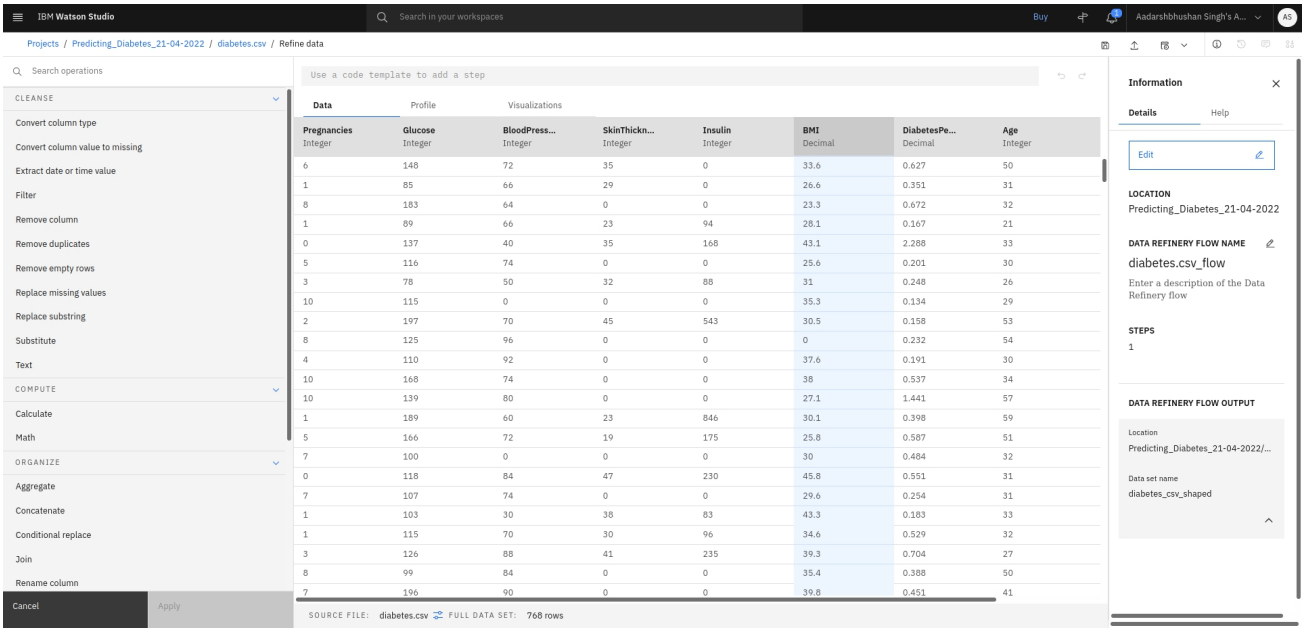


Fig: Date Refining

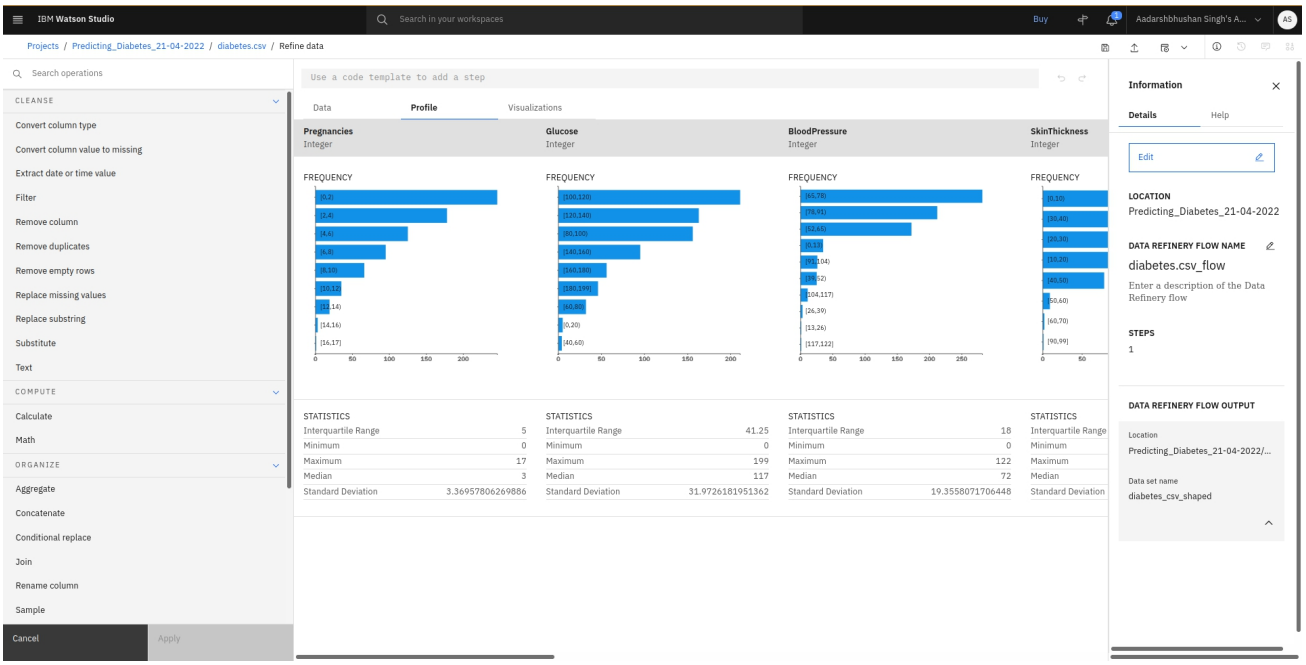


Fig: Reading profile of data through data refining

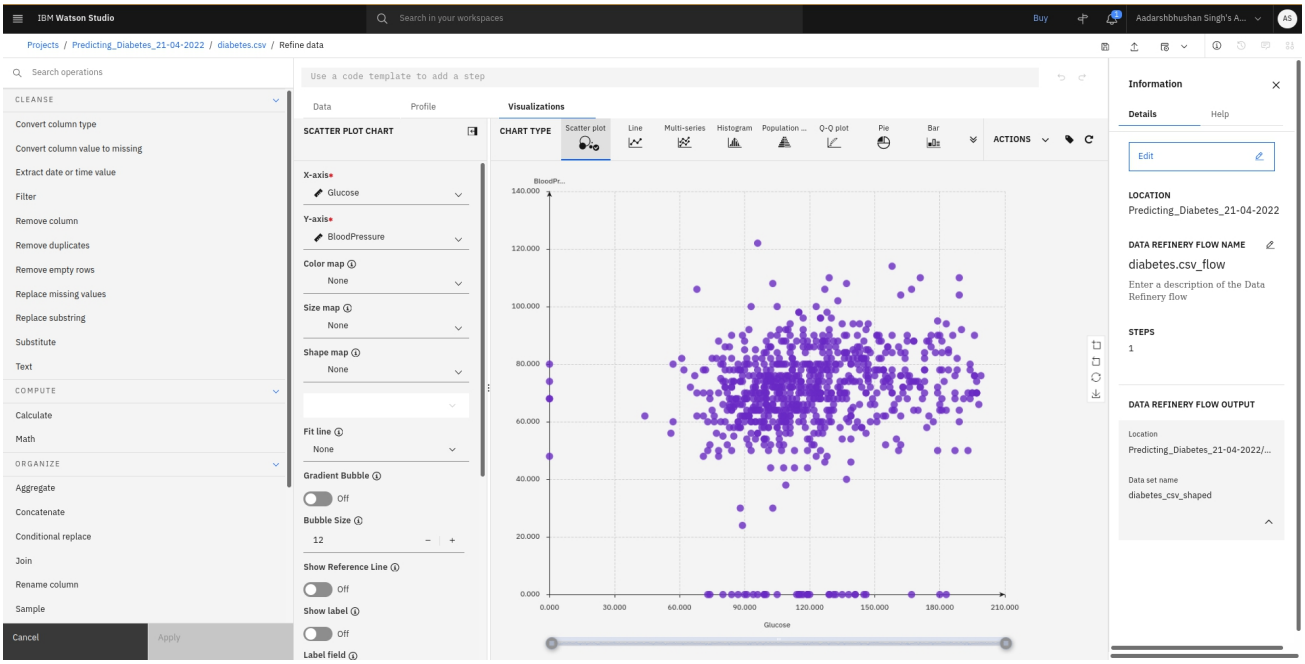


Fig: Visualising Glucose and Blood Pressure Data

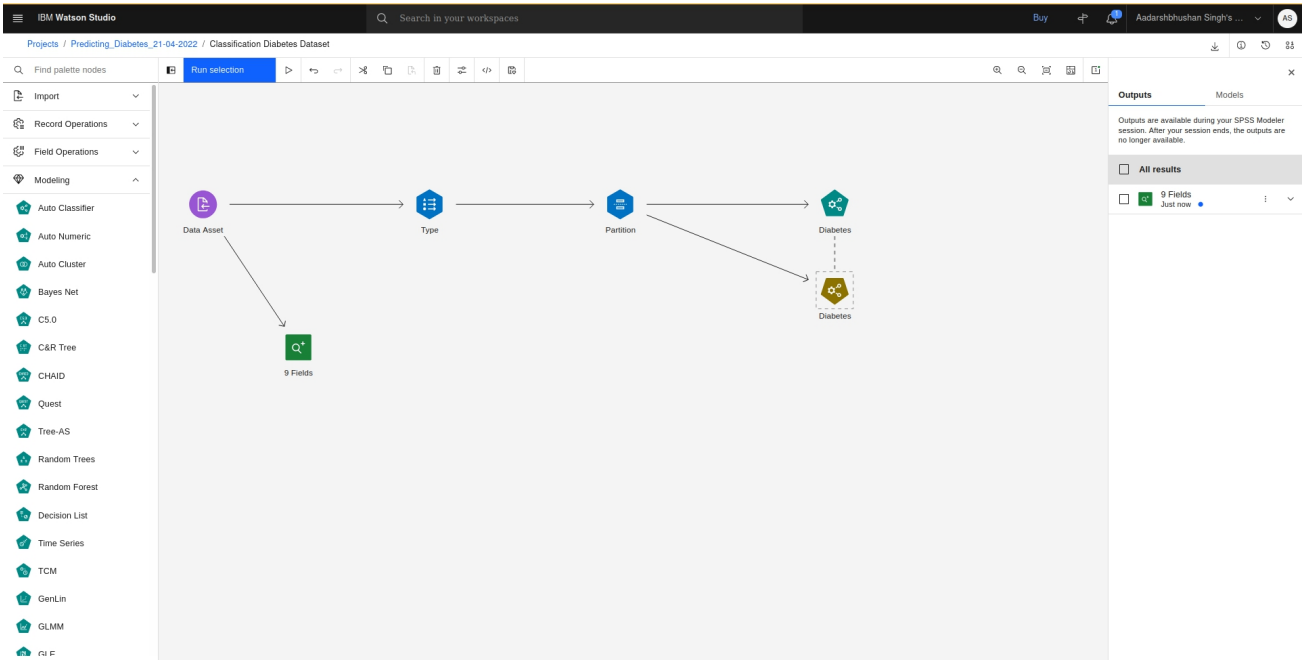


Fig: SPSS Modeler of diabetes

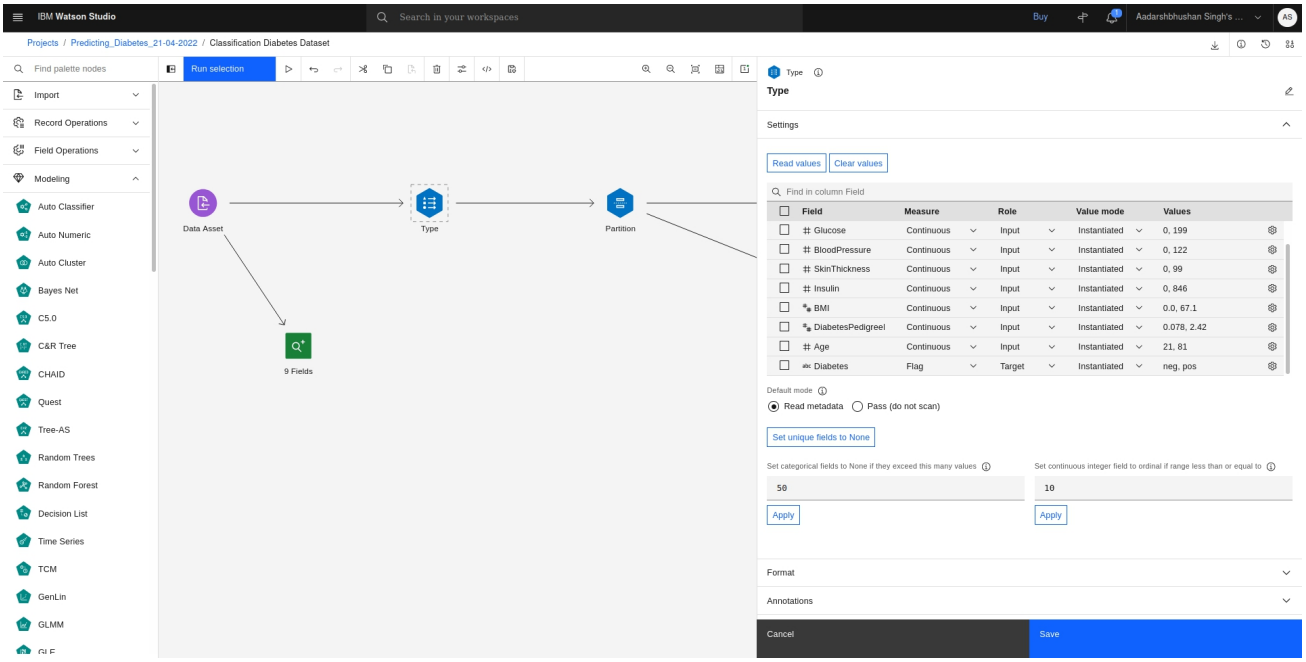
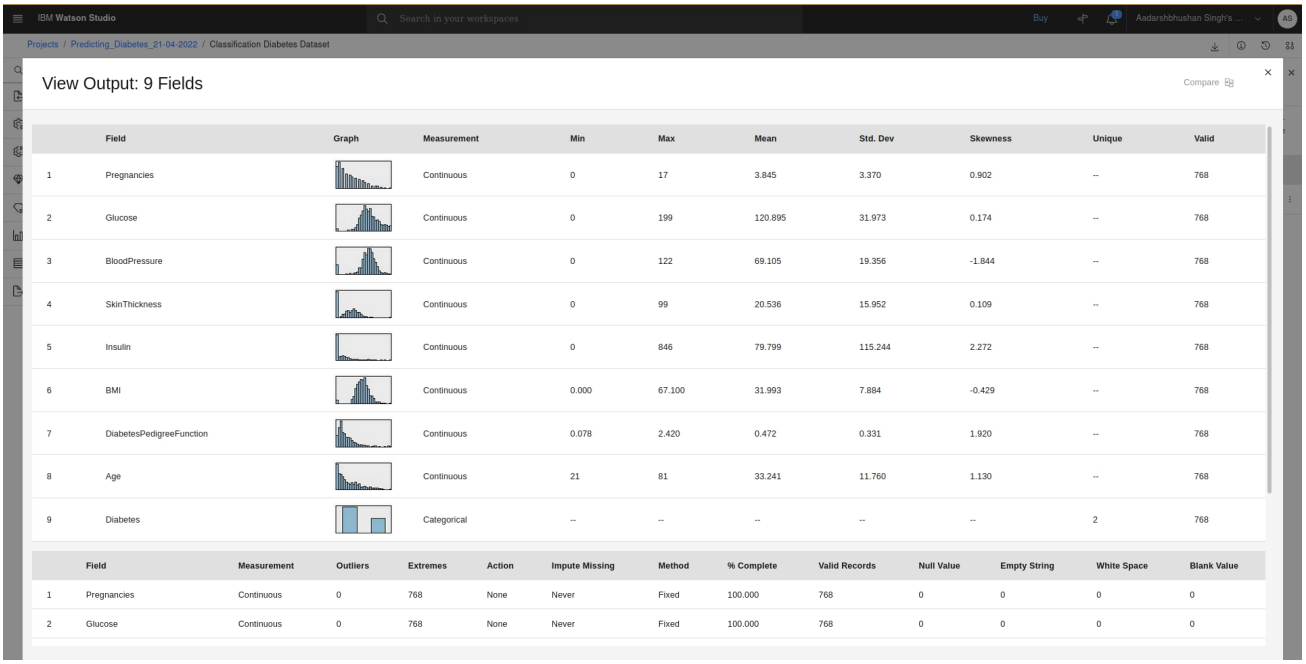


Fig: Setting the types of attribute



	Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	Pregnancies		Continuous	0	17	3.845	3.370	0.902	--	768
2	Glucose		Continuous	0	199	120.895	31.973	0.174	--	768
3	BloodPressure		Continuous	0	122	69.105	19.356	-1.844	--	768
4	SkinThickness		Continuous	0	99	20.536	15.952	0.109	--	768
5	Insulin		Continuous	0	846	79.799	115.244	2.272	--	768
6	BMI		Continuous	0.000	67.100	31.993	7.884	-0.429	--	768
7	DiabetesPedigreeFunction		Continuous	0.078	2.420	0.472	0.331	1.920	--	768
8	Age		Continuous	21	81	33.241	11.760	1.130	--	768
9	Diabetes		Categorical	--	--	--	--	--	2	768

	Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
1	Pregnancies	Continuous	0	768	None	Never	Fixed	100.000	768	0	0	0	0
2	Glucose	Continuous	0	768	None	Never	Fixed	100.000	768	0	0	0	0

Fig: View Output of fields

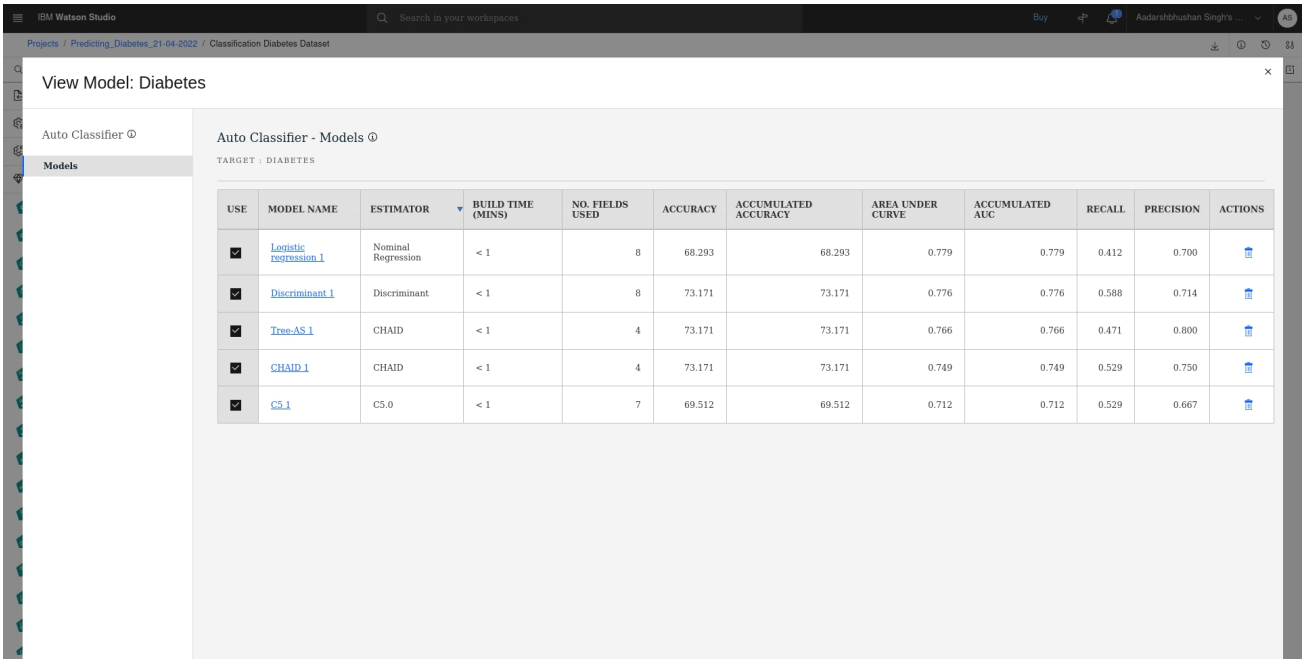


Fig: Result of Auto Classifier

Dataset Name: insurance.csv

Link to the dashboard:

https://eu-gb.dataplatform.cloud.ibm.com/canvas/flows/3162bb54-7f3c-4fce-a6a8-da074a9ebc93?context=cpdaas&projectGuid=6aa29757-cd4e-4b19-9b87-d6e390c25342&env_id=mods-6aa29757-cd4e-4b19-9b87-d6e390c25342

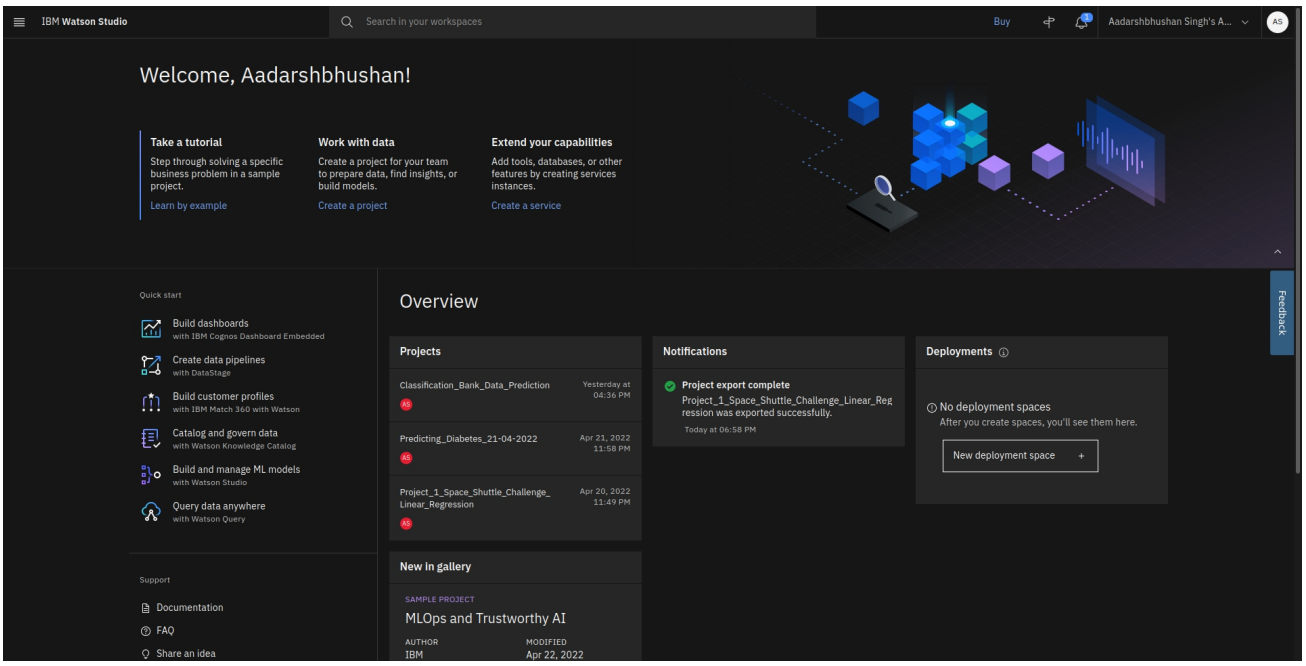


Fig: Create New Project

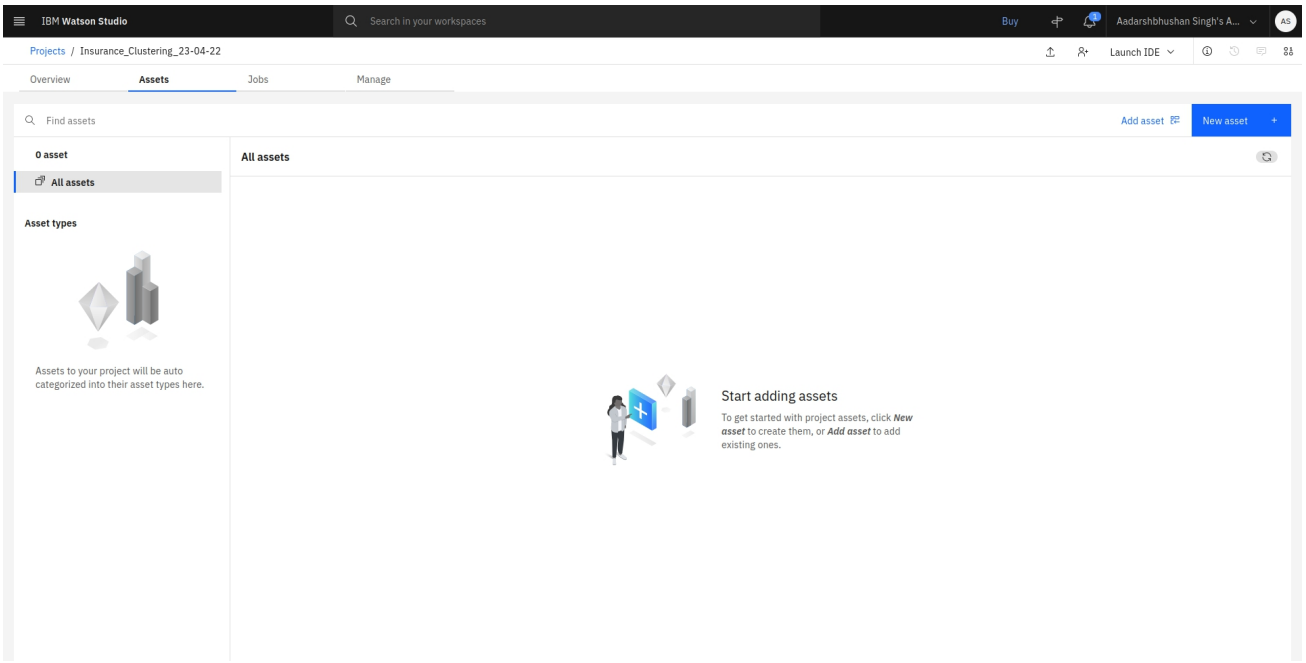


FIg: Go to Asset and Click on new asset

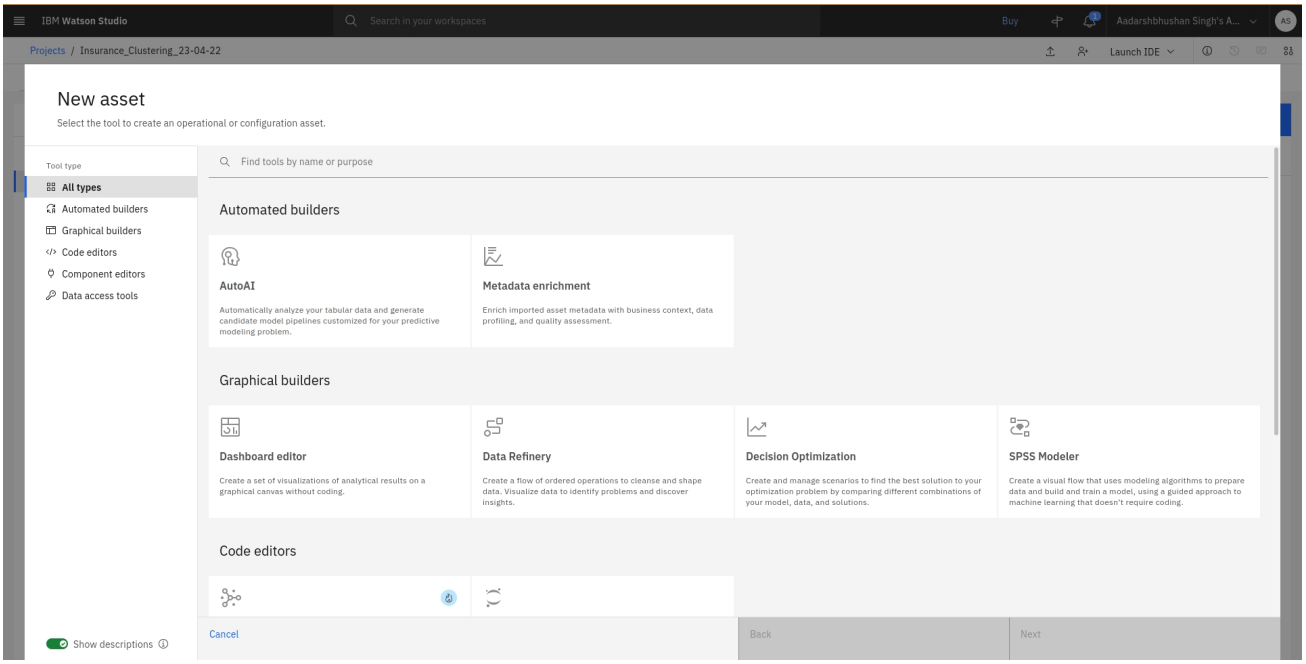


FIg: Data Refinary

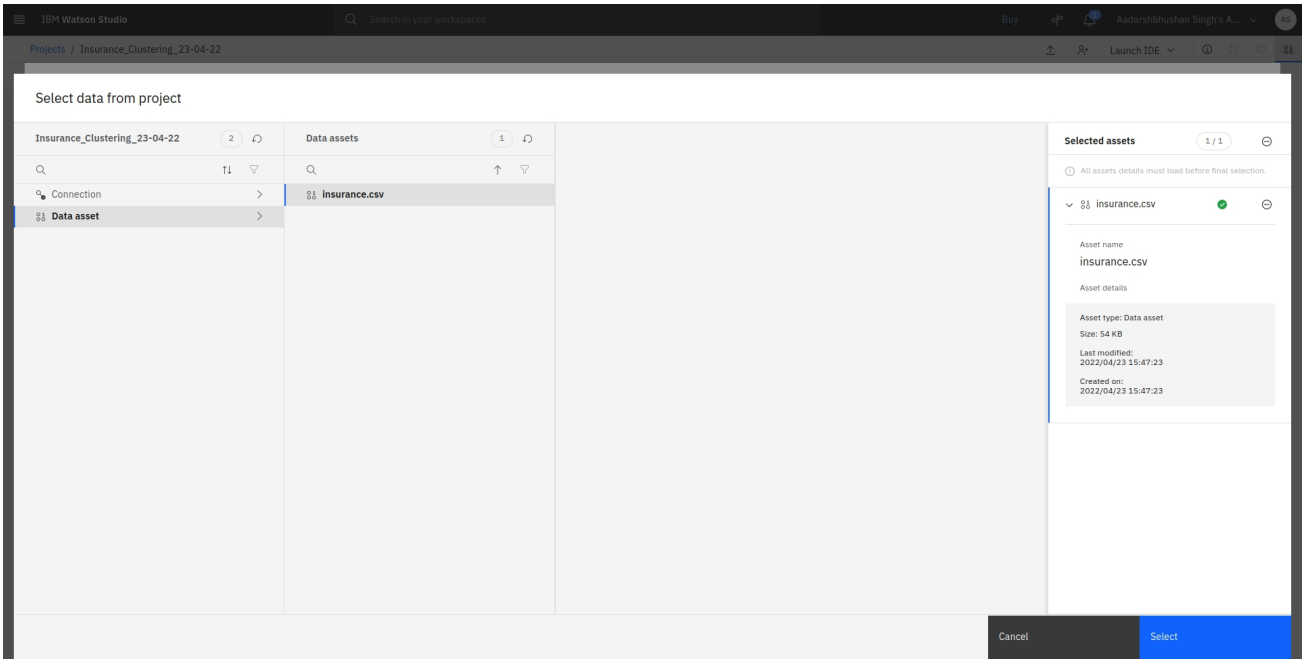


Fig: Select the dataset

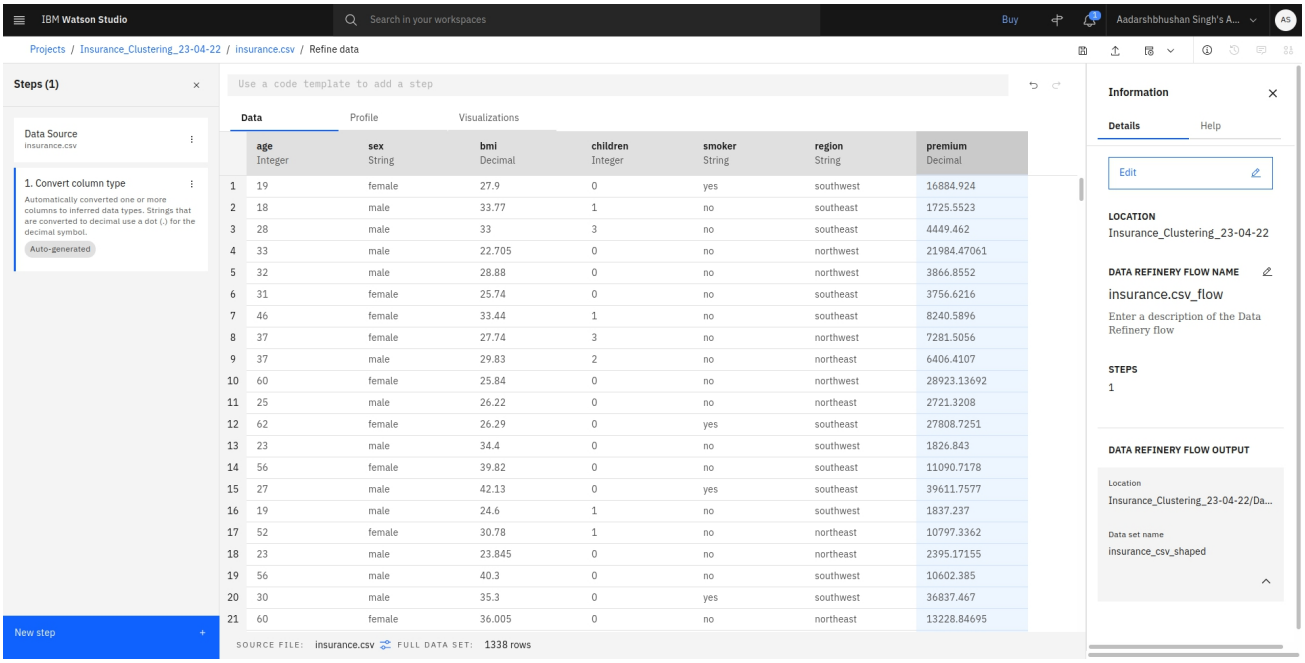


Fig: Data Refining

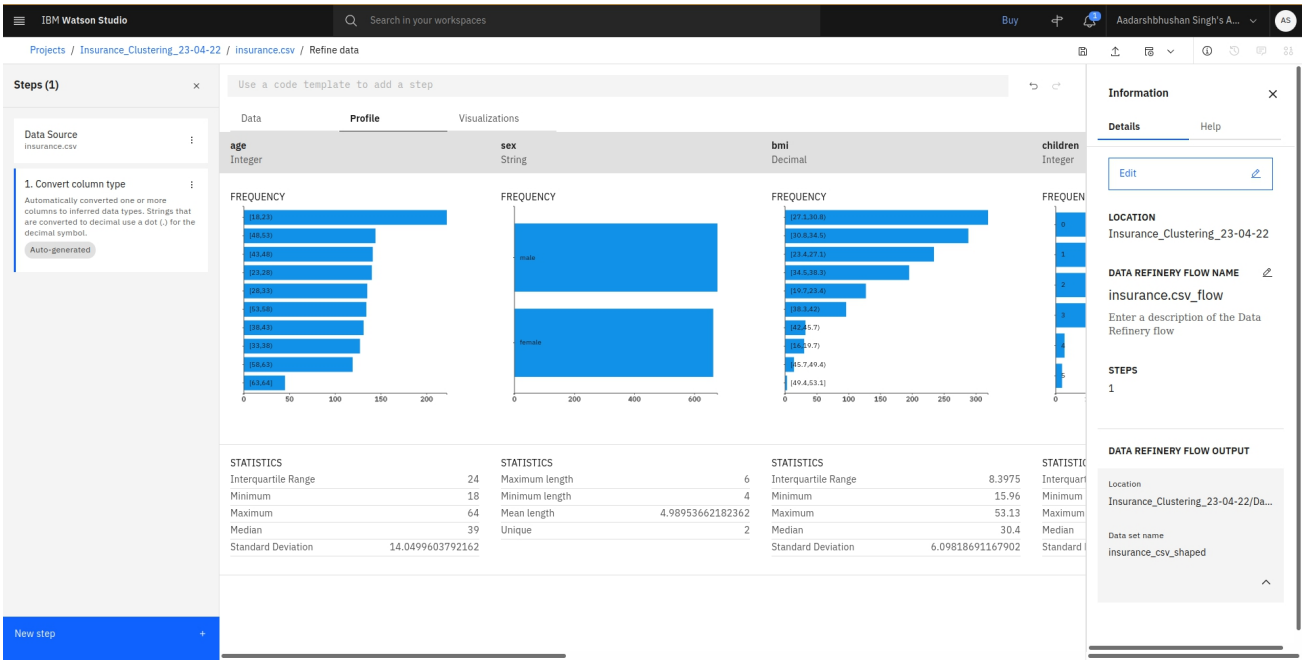


Fig: Observing data through data profile

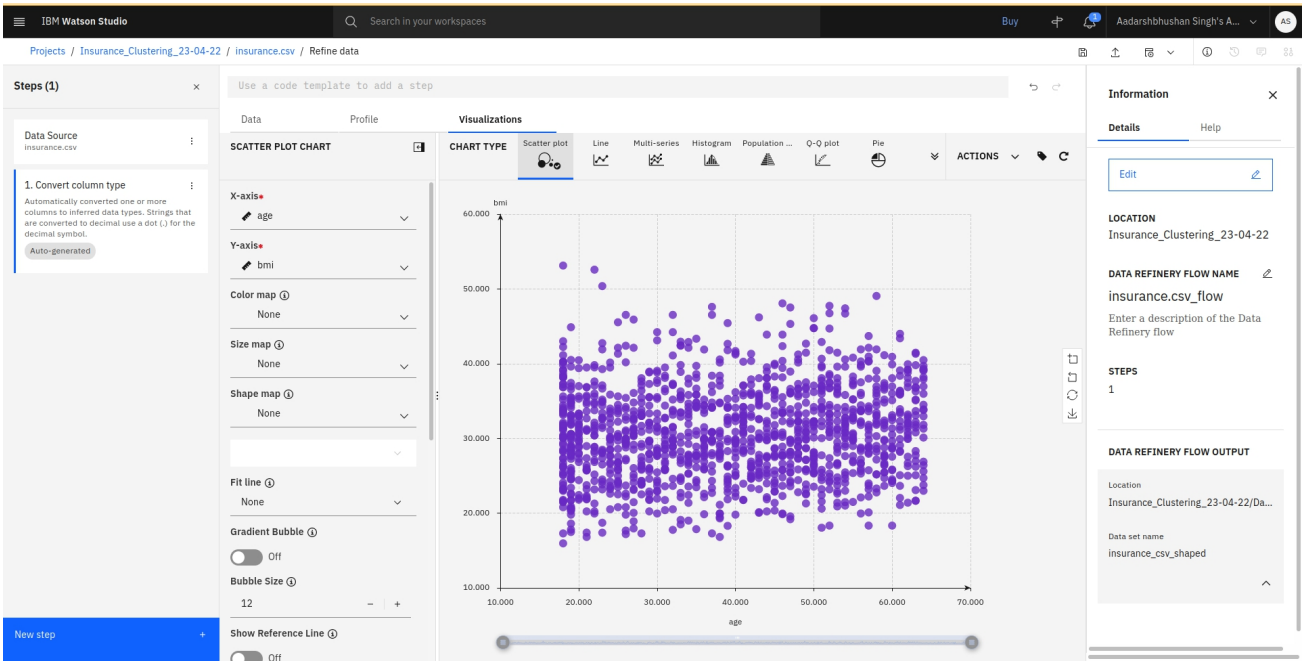


Fig: Observing bmi vs age graph

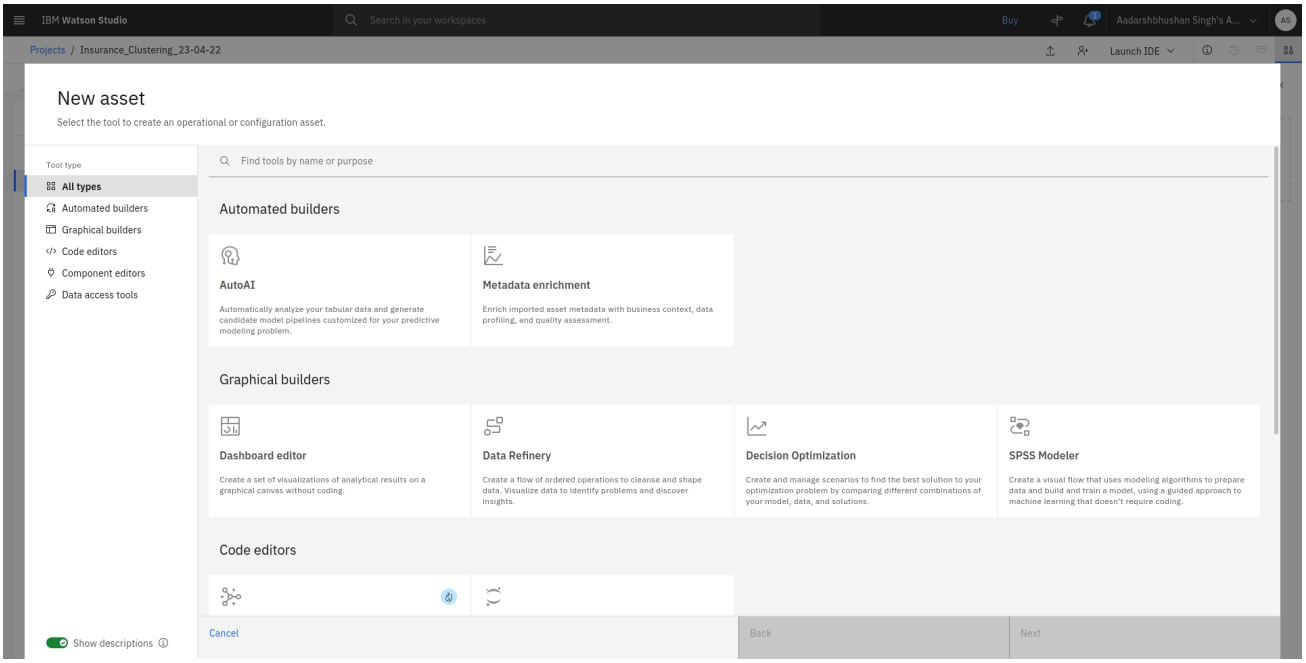


Fig: Creating new asset and selecting SPSS Modeler

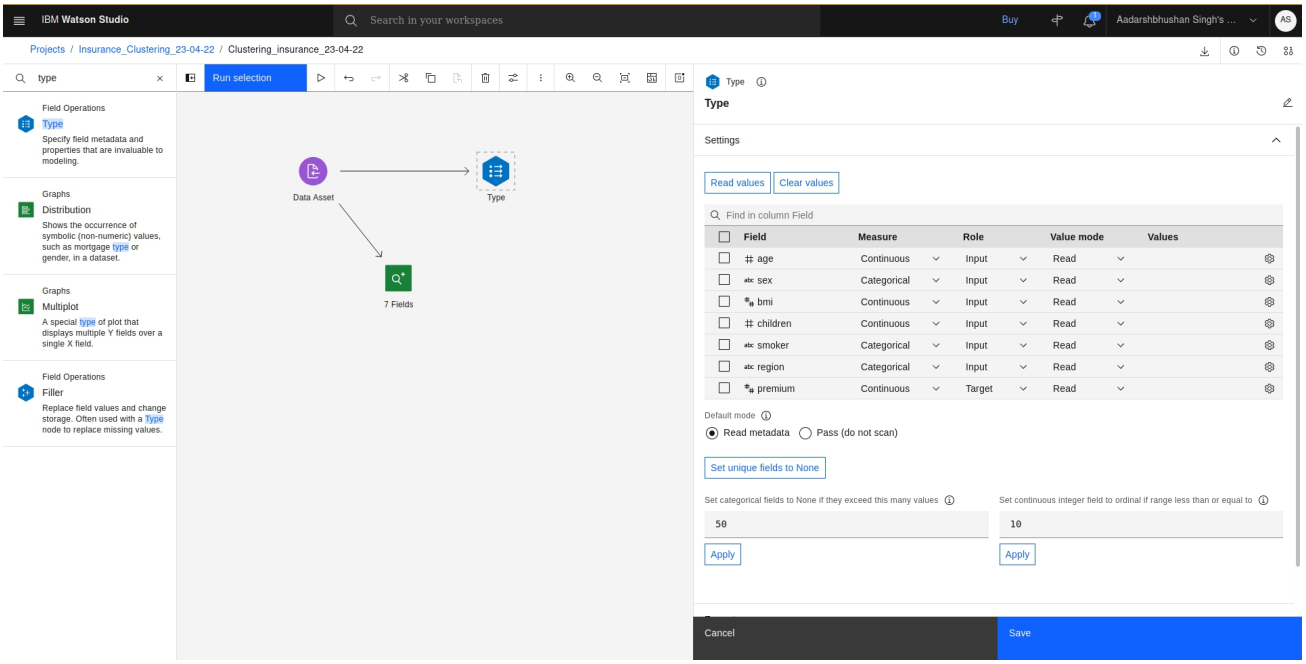


Fig: Setting up target and input attribute

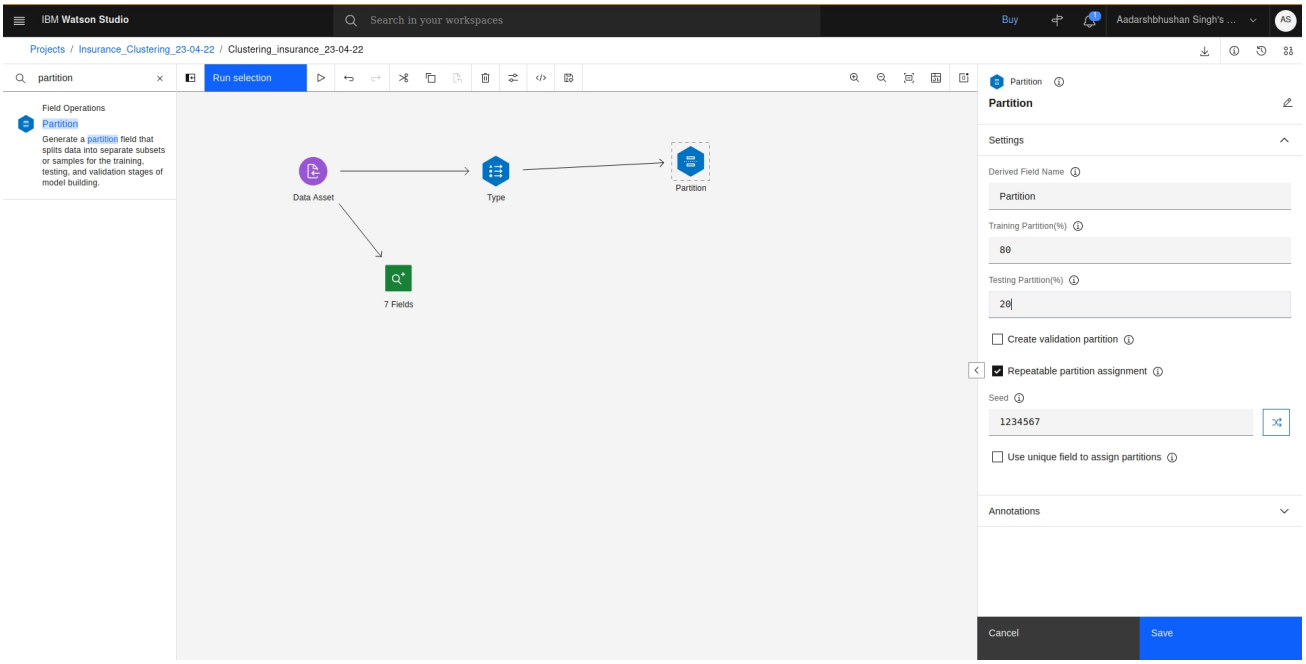


Fig: Creating partition

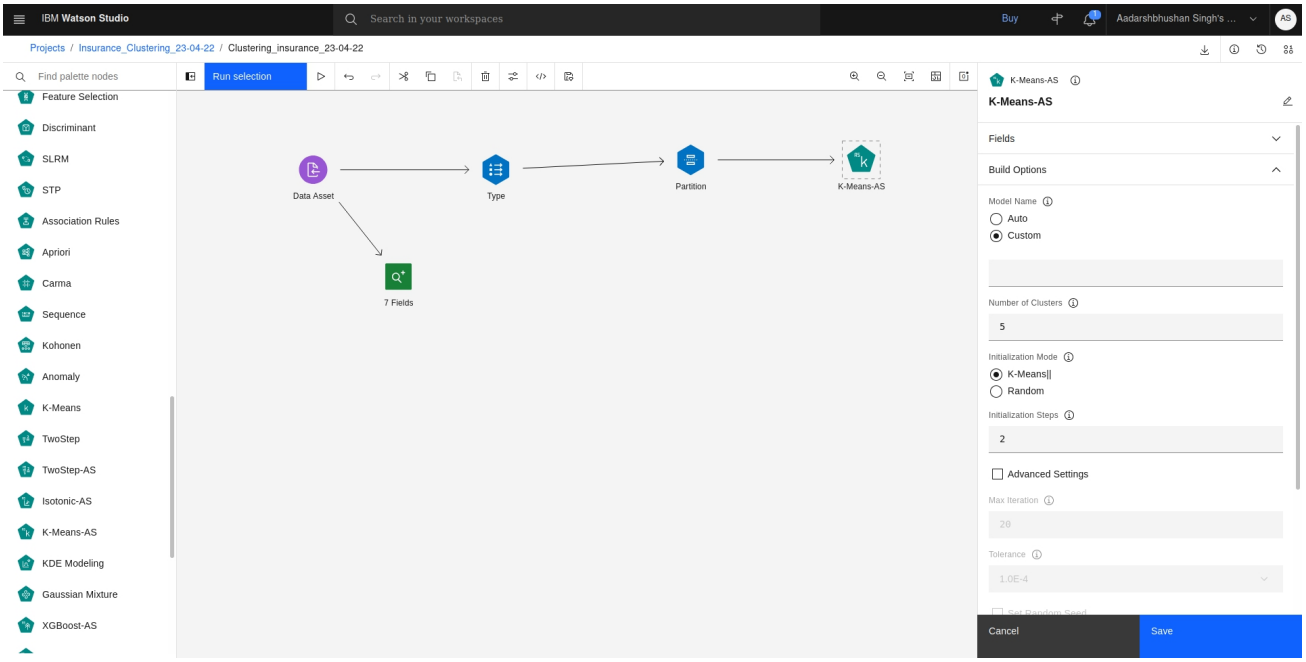
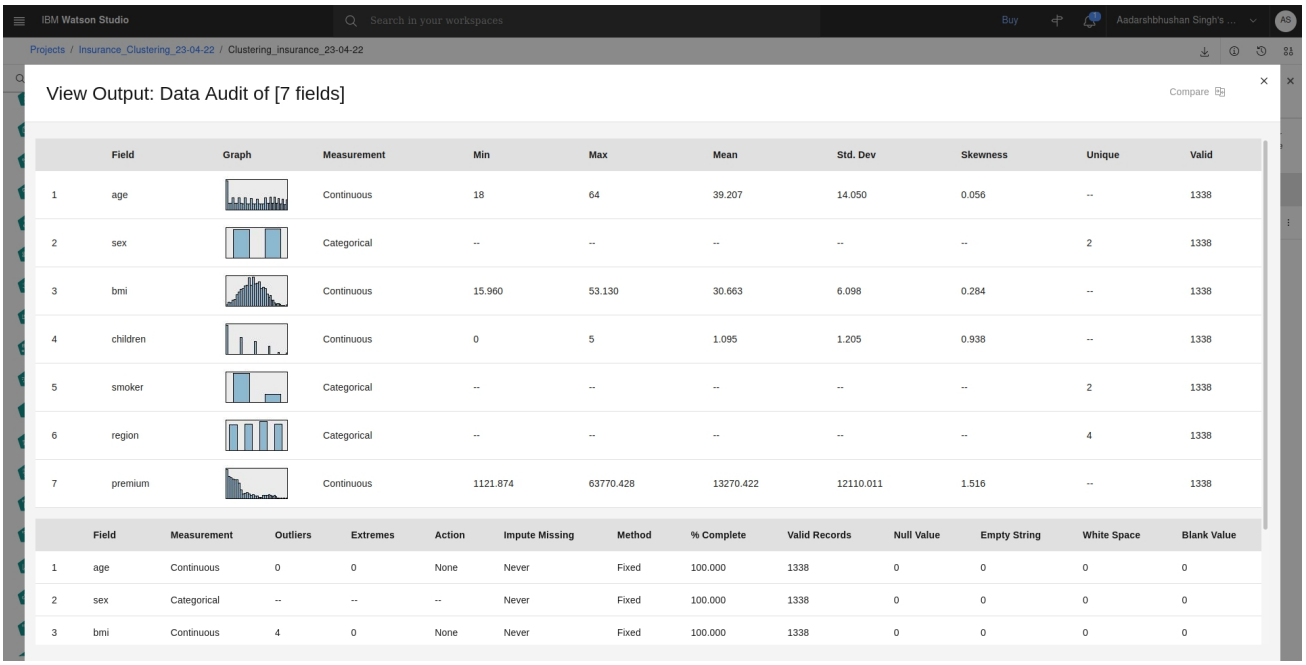


Fig: Selecting Number of clusters



	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

	Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
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

FIg: Data Audit output

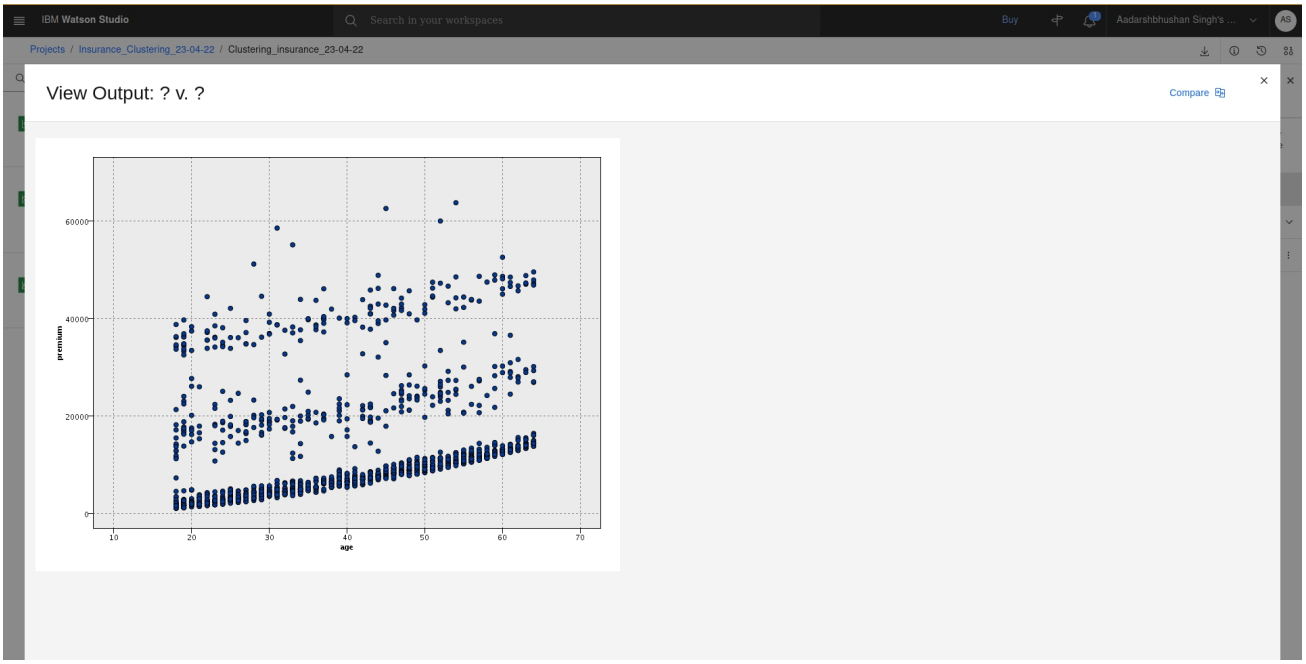


Fig: Ploting of Premium and age graph after clustering

Dataset Name: Mall_Customers.csv

Link to the dashboard:

https://eu-gb.dataplatform.cloud.ibm.com/canvas/flows/8e0077ad-828d-4e1d-9495-09ad62a41229?context=cpdaas&projectGuid=ccdac624-e9ea-47b8-b33f-348063a32f87&env_id=mods-ccdac624-e9ea-47b8-b33f-348063a32f87

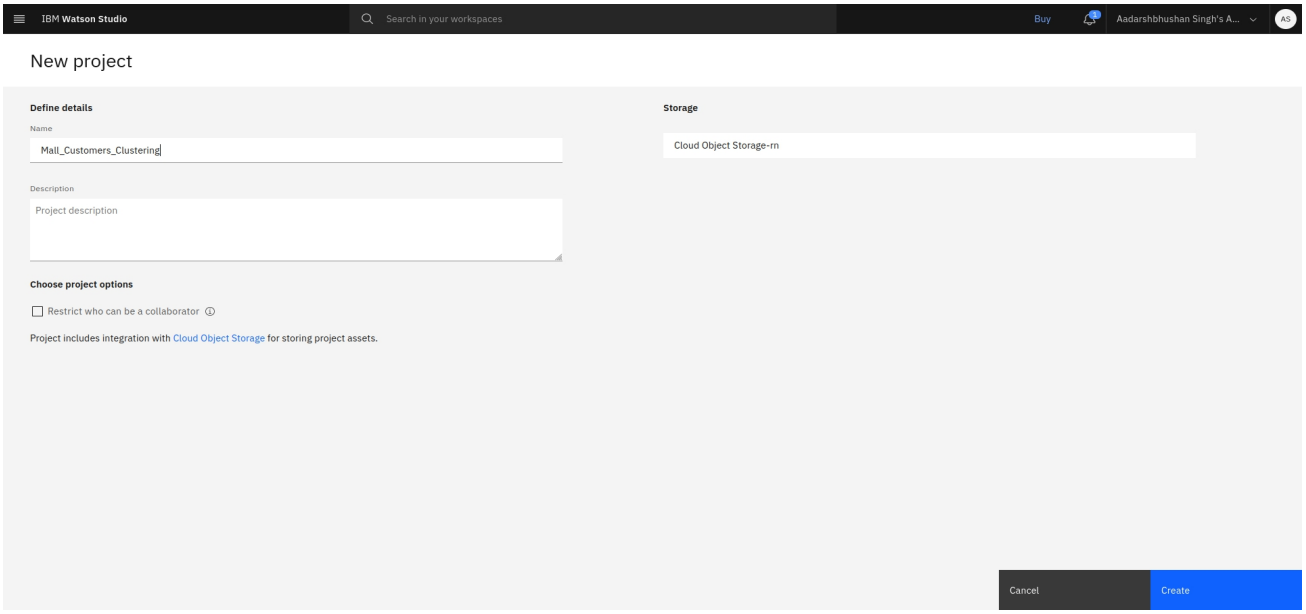


FIg: Creating new project

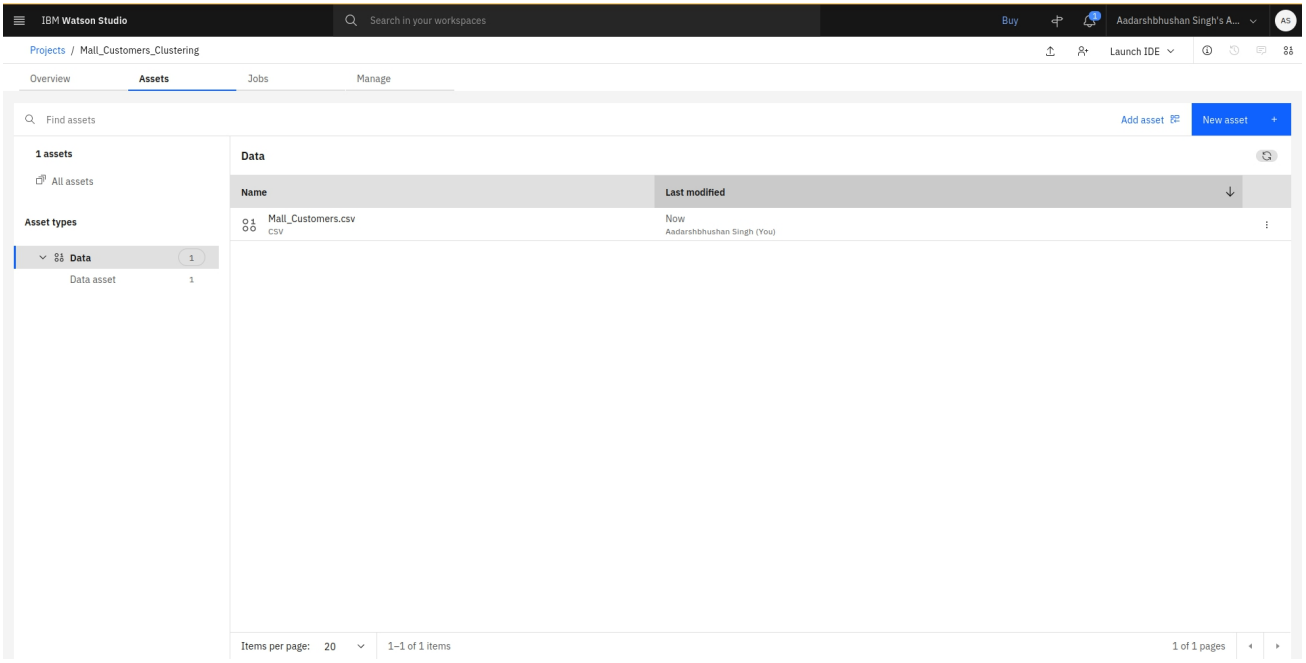


Fig: Adding dataset Mall Customers

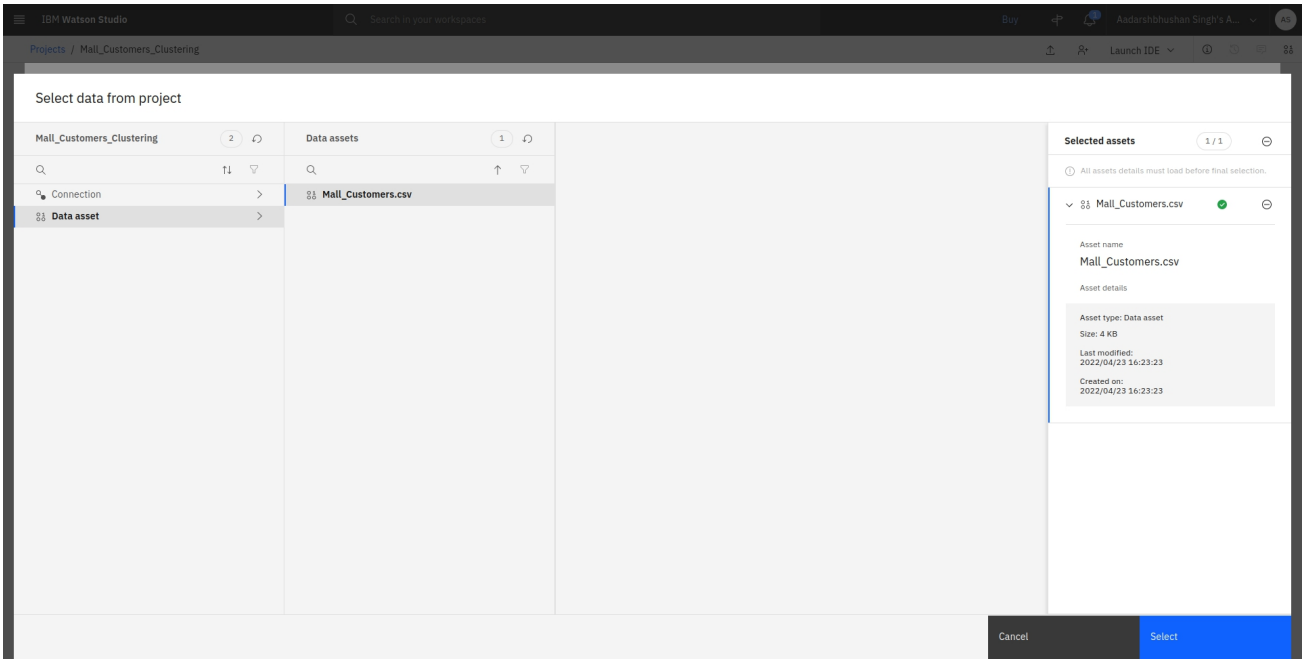


Fig: Select dataset for data refinery

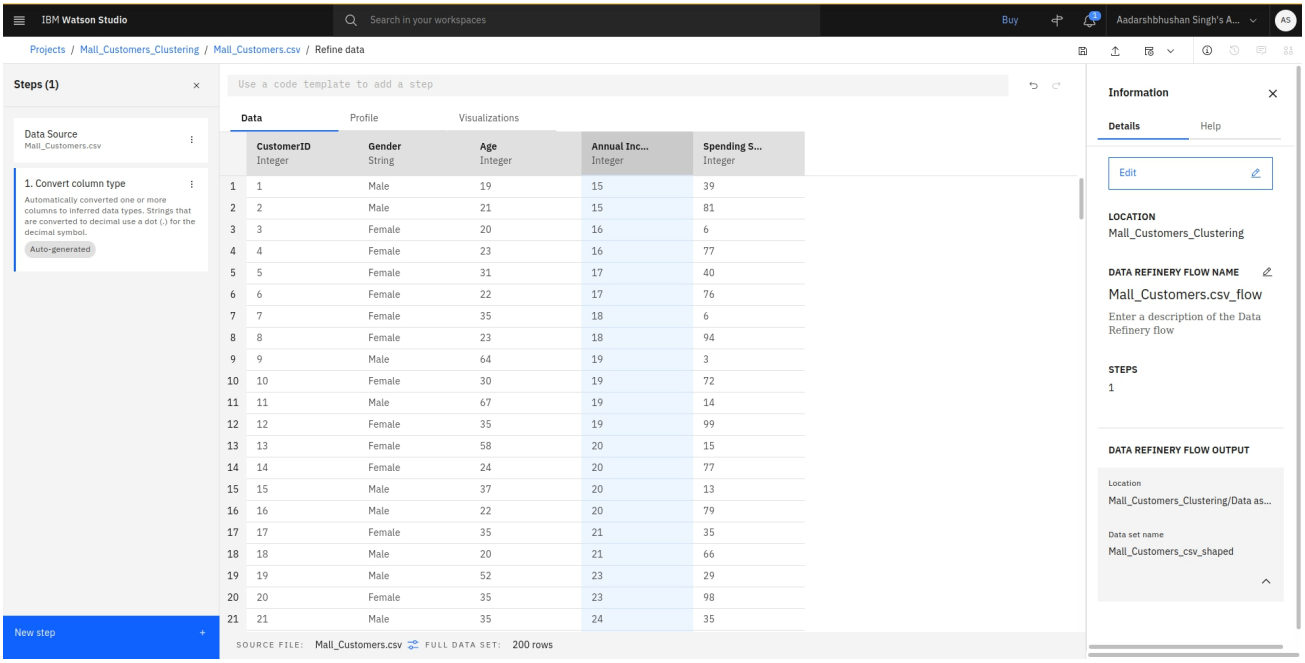


Fig: Data Refining

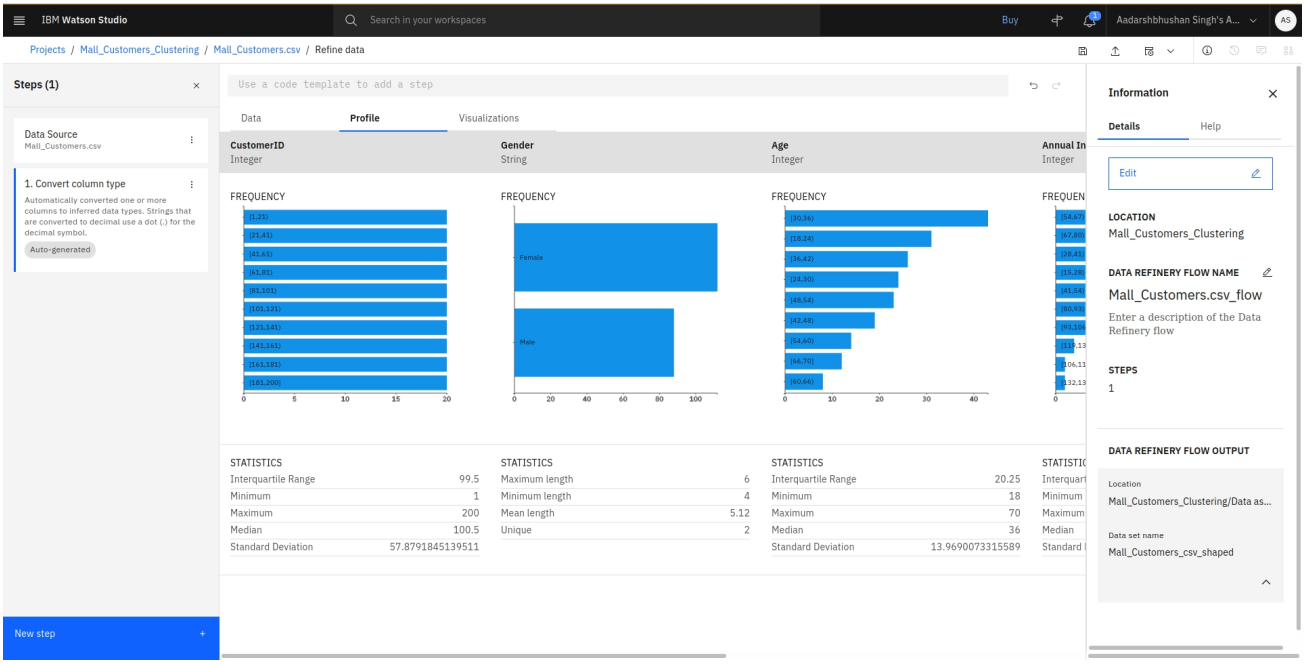


Fig: Understanding data through profile

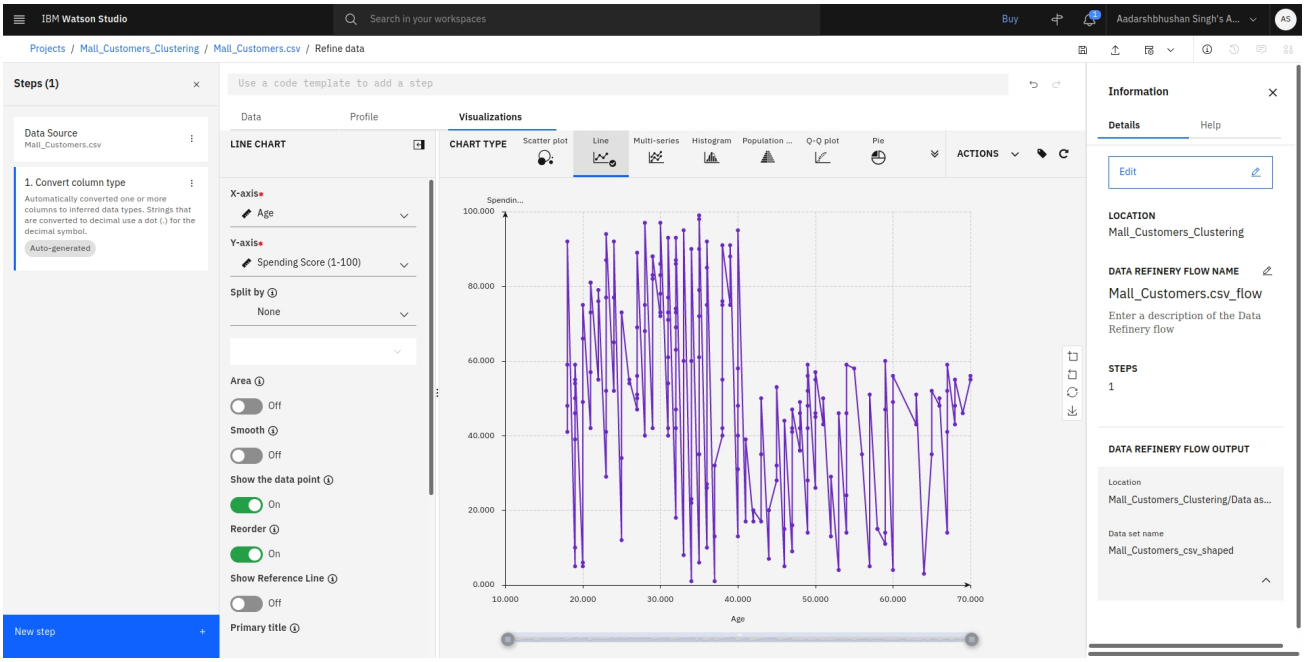


Fig: Visualising data through graph: Age vs Spending score

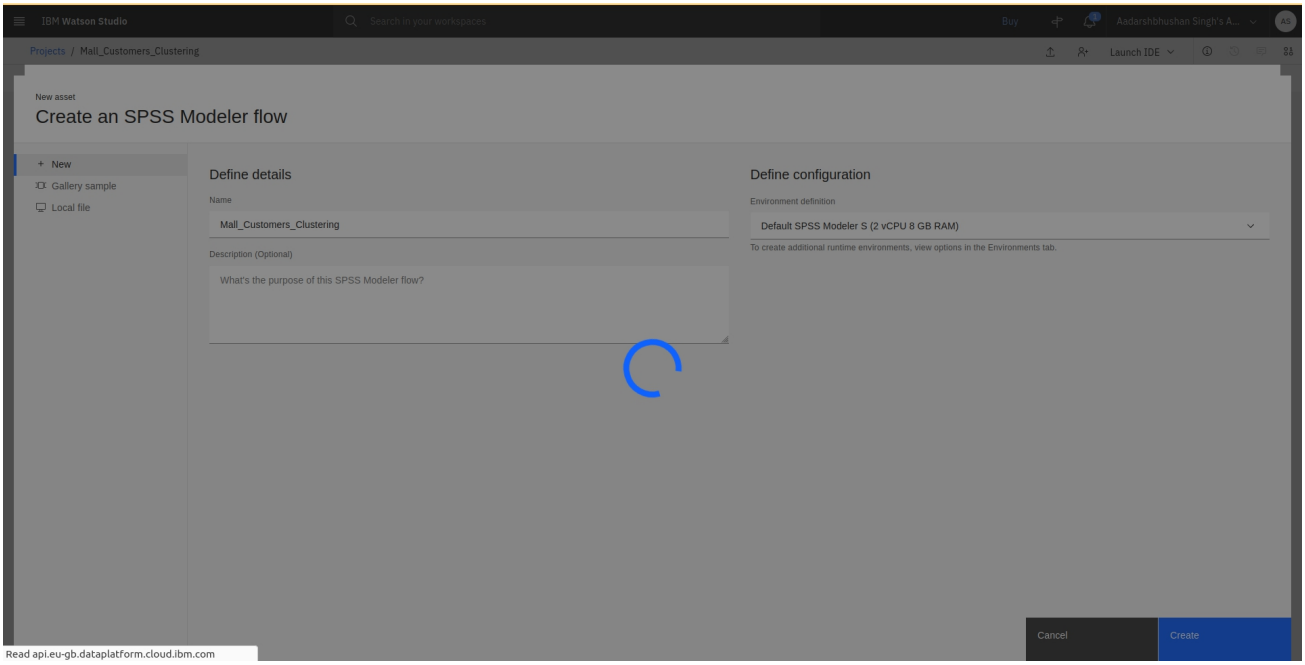


Fig: Creating SPSS Modeler

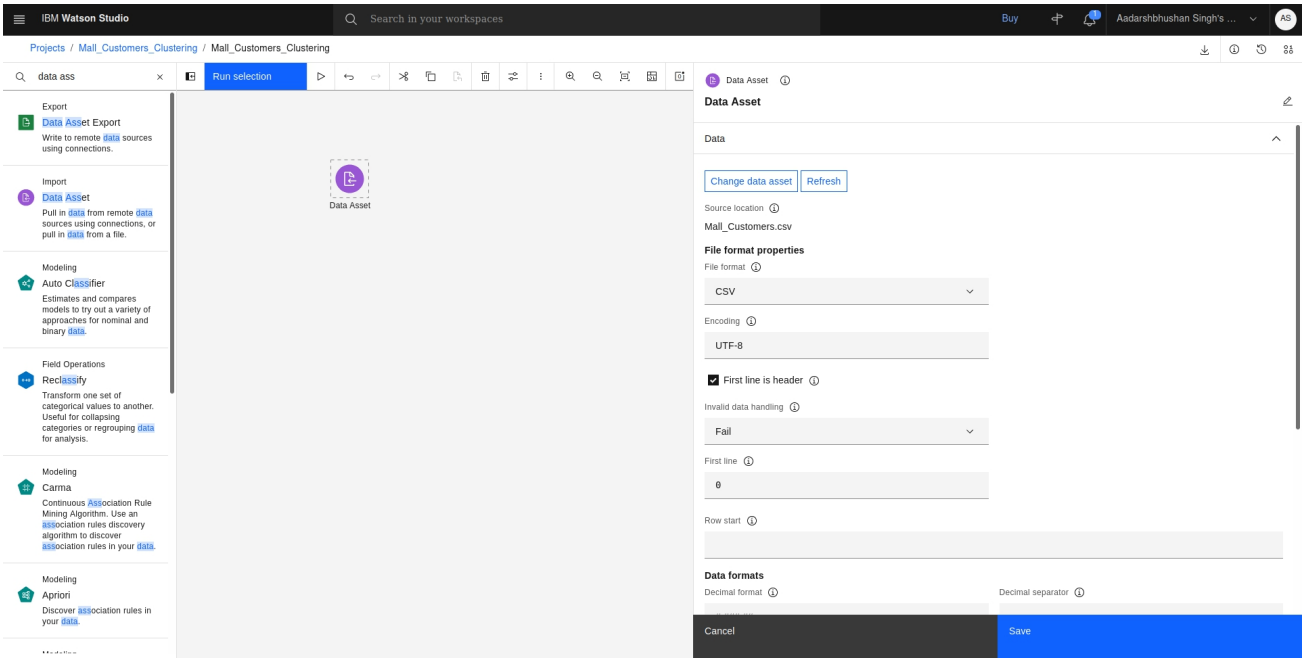


Fig: Selecting Data Asset

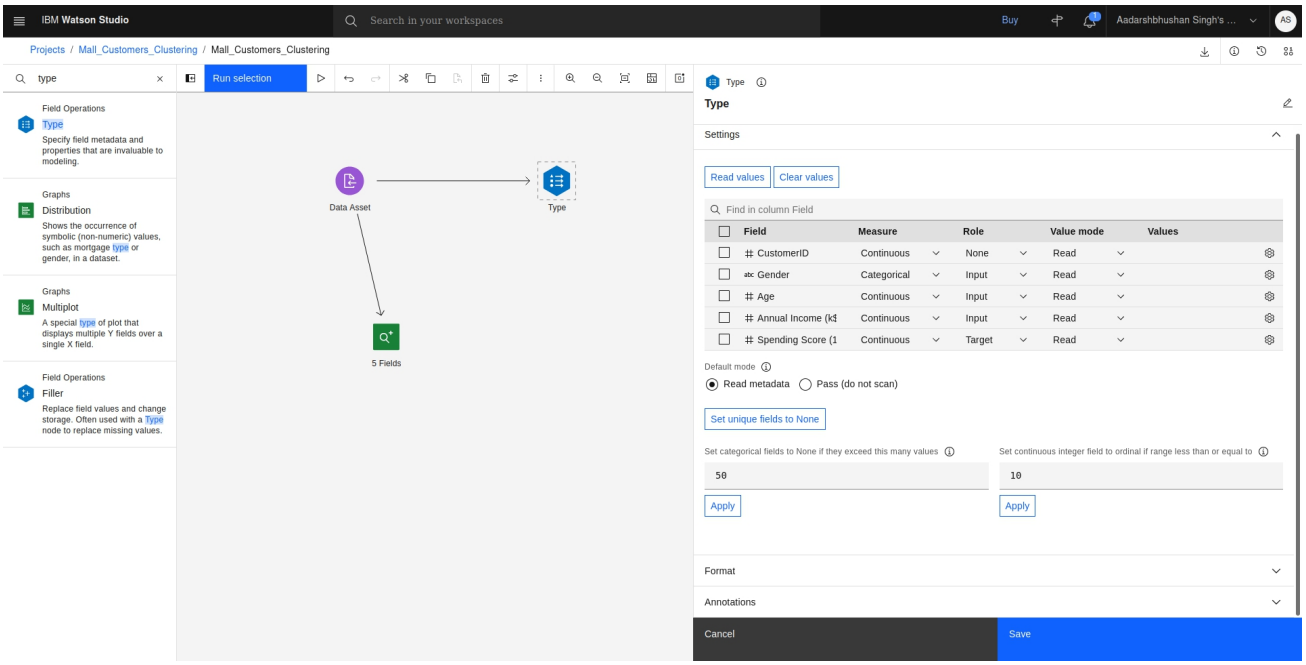


Fig: Fixing the target and input attribute

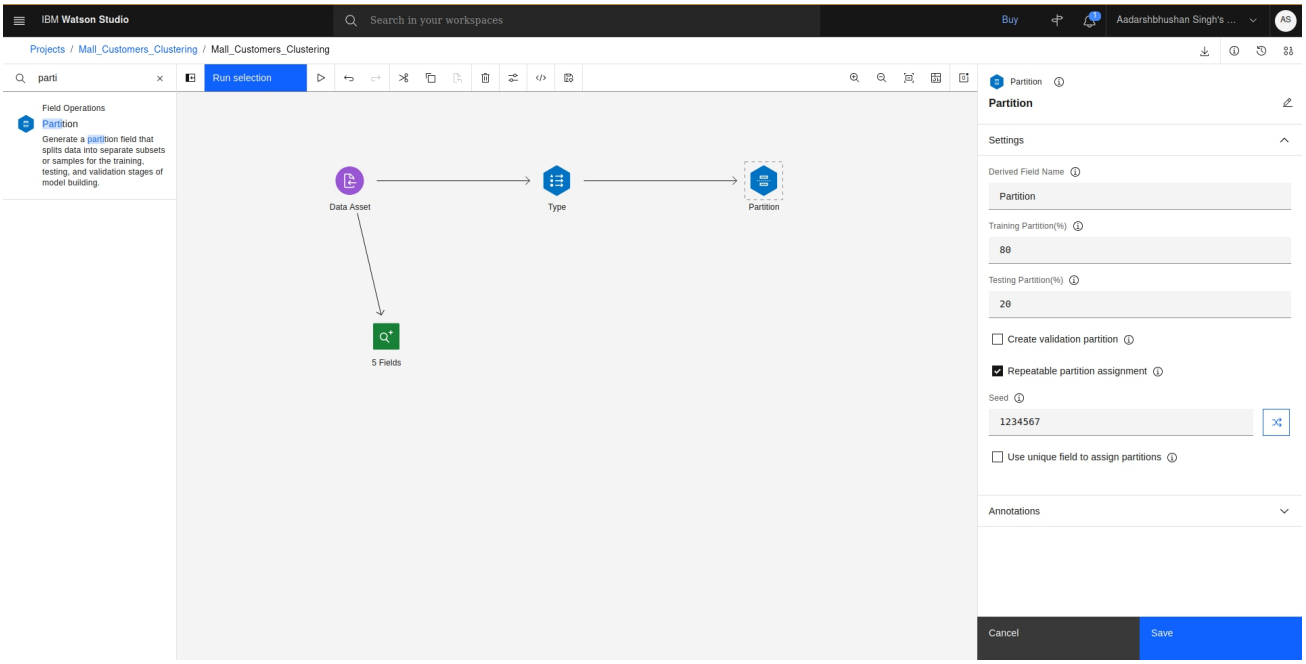
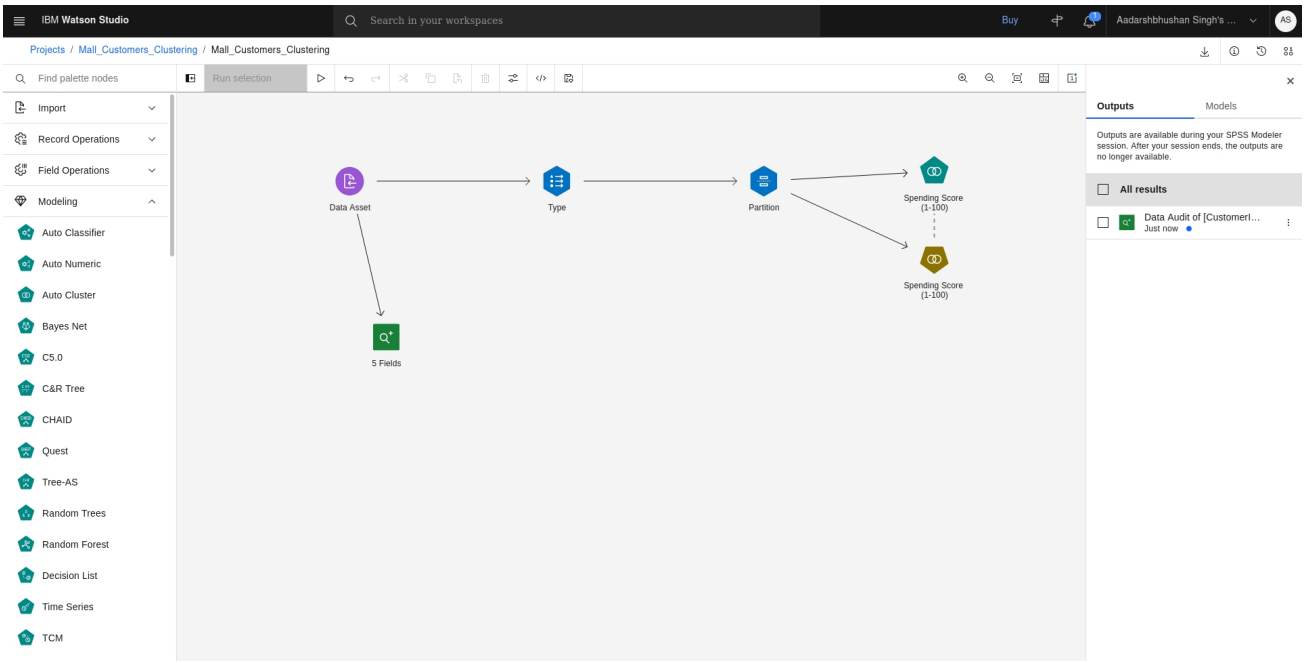
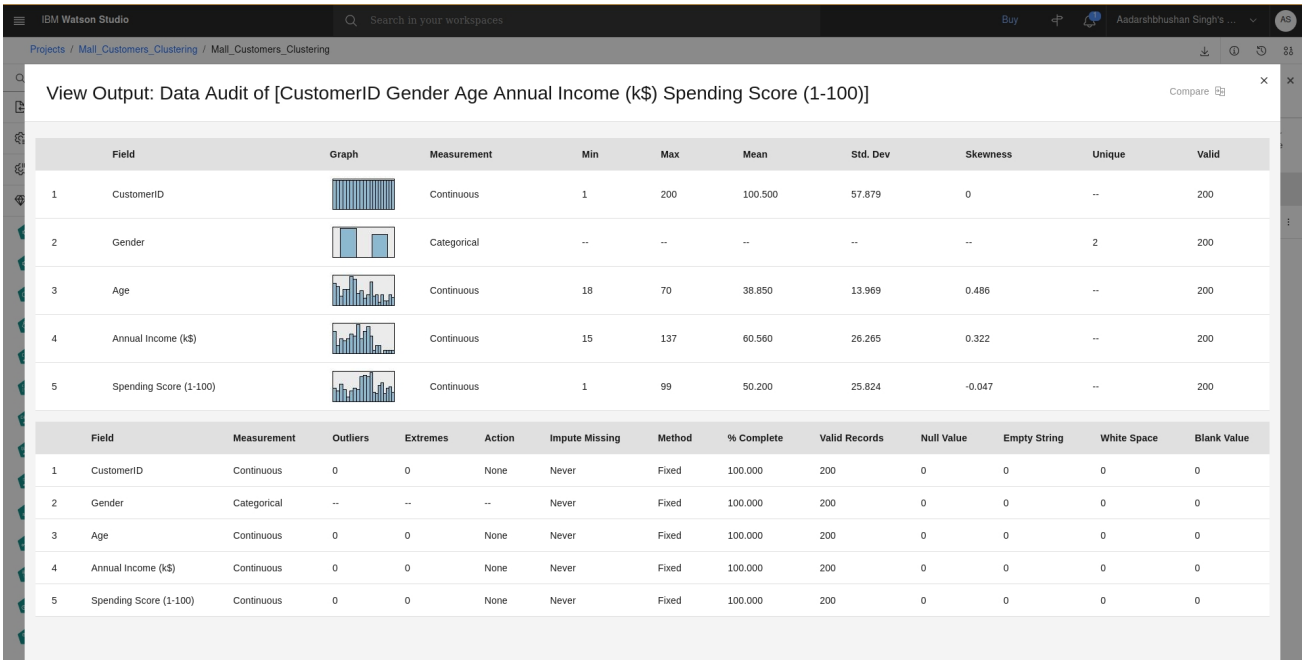


Fig: Setting the testing and training attribute



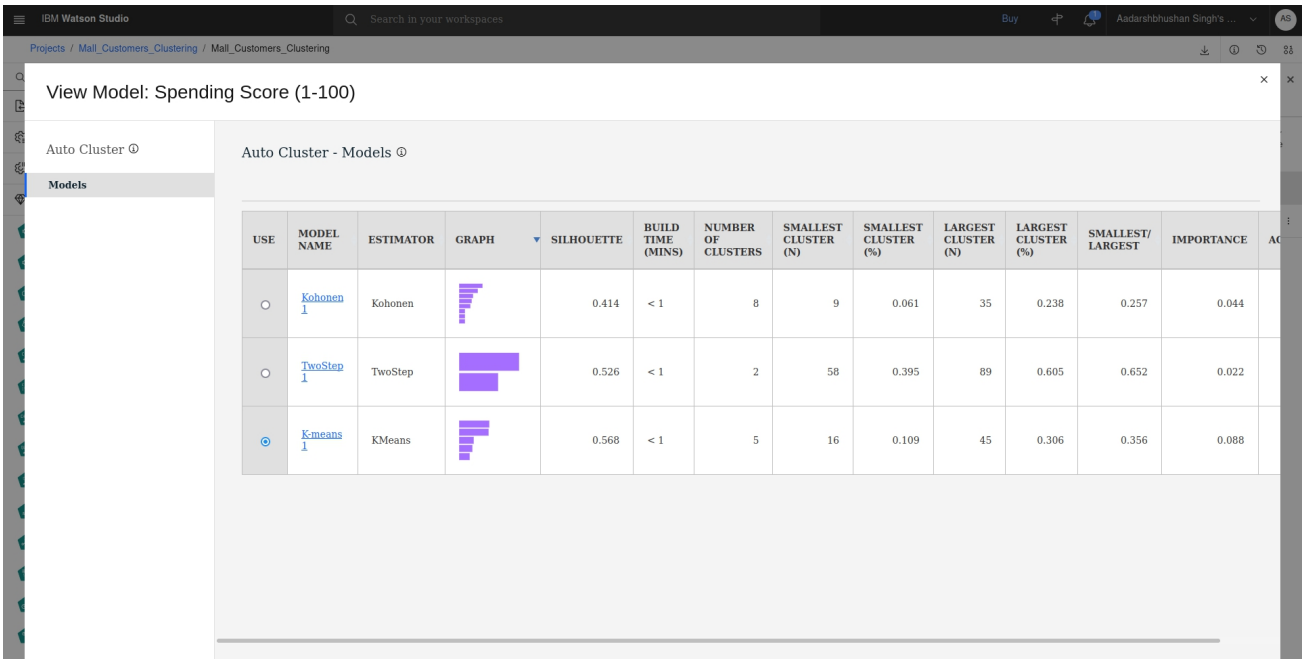
Final SPSS Modeler



	Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	CustomerID		Continuous	1	200	100.500	57.879	0	--	200
2	Gender		Categorical	--	--	--	--	--	2	200
3	Age		Continuous	18	70	38.850	13.969	0.486	--	200
4	Annual Income (k$)		Continuous	15	137	60.560	26.265	0.322	--	200
5	Spending Score (1-100)		Continuous	1	99	50.200	25.824	-0.047	--	200

	Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space	Blank Value
1	CustomerID	Continuous	0	0	None	Never	Fixed	100.000	200	0	0	0	0
2	Gender	Categorical	--	--	--	Never	Fixed	100.000	200	0	0	0	0
3	Age	Continuous	0	0	None	Never	Fixed	100.000	200	0	0	0	0
4	Annual Income (k$)	Continuous	0	0	None	Never	Fixed	100.000	200	0	0	0	0
5	Spending Score (1-100)	Continuous	0	0	None	Never	Fixed	100.000	200	0	0	0	0

Fig: Data Audit Output



USE	MODEL NAME	ESTIMATOR	GRAPH	SILHOUETTE	BUILD TIME (MINS)	NUMBER OF CLUSTERS	SMALLEST CLUSTER (N)	SMALLEST CLUSTER (%)	LARGEST CLUSTER (N)	LARGEST CLUSTER (%)	SMALLEST/ LARGEST	IMPORTANCE
○	Kohonen 1	Kohonen		0.414	< 1	8	9	0.061	35	0.238	0.257	0.044
○	TwoStep 1	TwoStep		0.526	< 1	2	58	0.395	89	0.605	0.652	0.022
◉	K-means 1	KMeans		0.568	< 1	5	16	0.109	45	0.306	0.356	0.088

Fig: Final Clustering

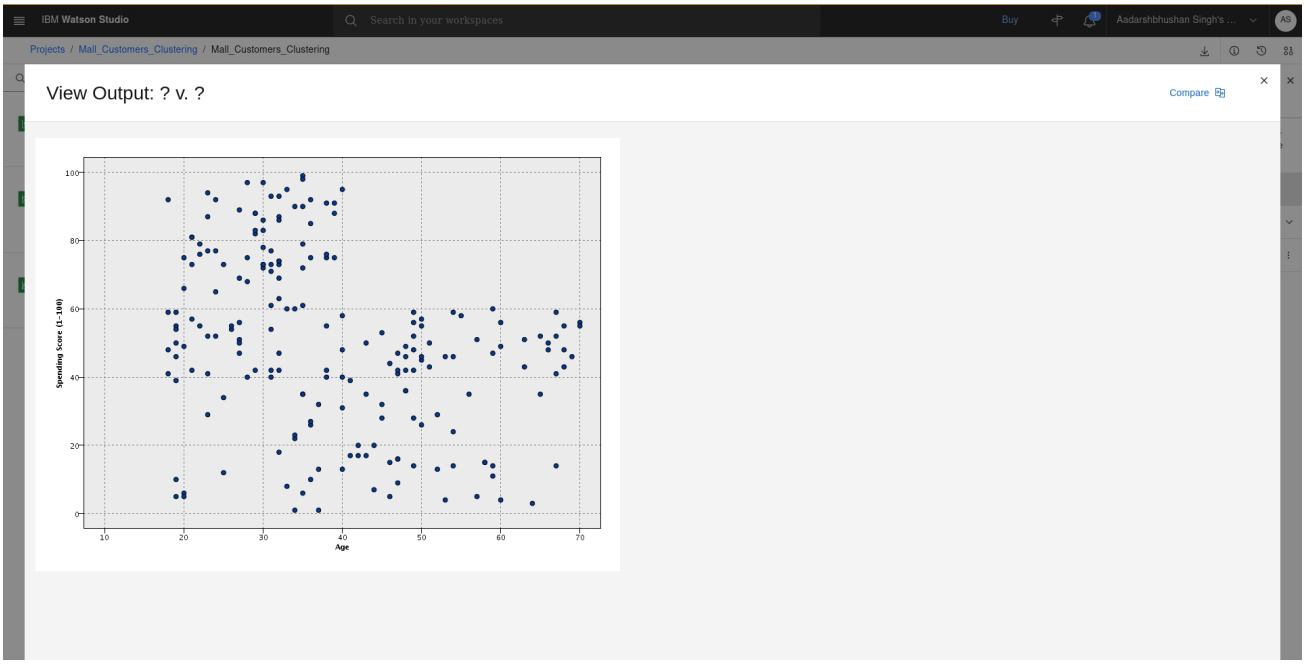


Fig: Data plotting of Spending score vs age