

Name: Sunny Kumar
Reg.no: 19BCE2637
University: VIT University, Vellore

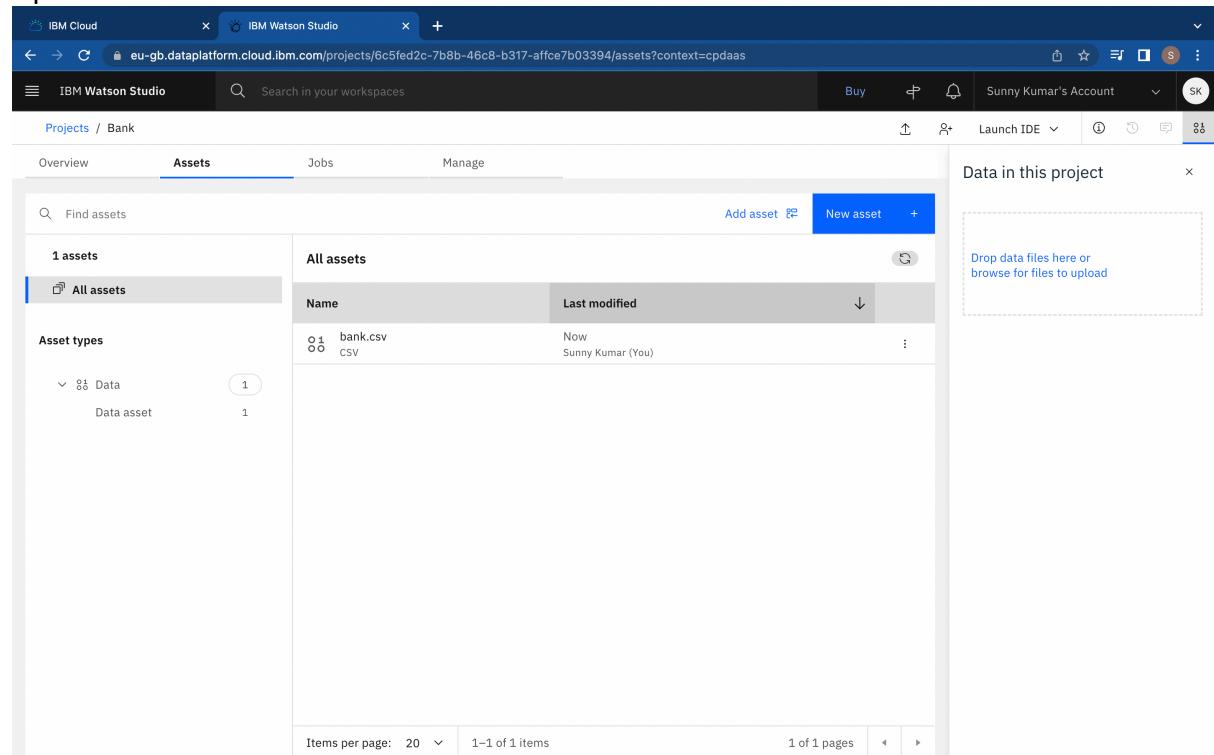
Assignment-3

Data set: Bank.csv

Link: <https://eu-gb.dataplatform.cloud.ibm.com/projects/6c5fed2c-7b8b-46c8-b317-affce7b03394/assets?context=cpdaas>

- Login to IBM Cloud.
- Go to service and Software and select Watson Studion.
- Select Watson Studio and lunch in IBM Cloud Pak for Data.
- Then create a project with suitable name.
- Then upload data asset.
- Add Data Asset to the flow

Upload Data



The screenshot shows the IBM Watson Studio interface within the IBM Cloud platform. The browser address bar indicates the URL: <https://eu-gb.dataplatform.cloud.ibm.com/projects/6c5fed2c-7b8b-46c8-b317-affce7b03394/assets?context=cpdaas>. The user is logged in as "Sunny Kumar's Account". The main navigation bar includes "Buy", "Watson Studio", "Jobs", "Manage", and "Launch IDE". The left sidebar shows "Projects / Bank" and "Overview". The central "Assets" tab is selected, displaying a table of assets. The table has columns "Name" and "Last modified". One asset is listed: "bank.csv" (CSV) last modified "Now" by "Sunny Kumar (You)". A "New asset" button is visible at the top right of the asset list. To the right of the asset list, there is a "Data in this project" section with a placeholder message: "Drop data files here or browse for files to upload". At the bottom of the page, there are pagination controls: "Items per page: 20", "1–1 of 1 items", "1 of 1 pages", and navigation arrows.

IBM Cloud IBM Watson Studio eu-gb.dataplatfrom.cloud.ibm.com/shaper?context=data&dataset_id=12f5e769-abc4-40d8-a061-e827c0c15b2c&project_id=6c5fed2c-7b8b-46c8-b317-affce7b...

IBM Watson Studio Search in your workspaces Buy Sunny Kumar's Account

Projects / Bank / bank.csv / Refine data

Use a code template to add a step

Data Profile Visualizations

	age	job	marital	education	default	balance	housing
1	59	admin.	married	secondary	no	2343	yes
2	56	admin.	married	secondary	no	45	no
3	41	technician	married	secondary	no	1270	yes
4	55	services	married	secondary	no	2476	yes
5	54	admin.	married	tertiary	no	184	no
6	42	management	single	tertiary	no	0	yes
7	56	management	married	tertiary	no	830	yes
8	60	retired	divorced	secondary	no	545	yes
9	37	technician	married	secondary	no	1	yes
10	28	services	single	secondary	no	5090	yes
11	38	admin.	single	secondary	no	100	yes
12	30	blue-collar	married	secondary	no	309	yes
13	29	management	married	tertiary	no	199	yes
14	46	blue-collar	single	tertiary	no	460	yes
15	31	technician	single	tertiary	no	703	yes
16	35	management	divorced	tertiary	no	3837	yes

SOURCE FILE: bank.csv SAMPLE SIZE: First 50 rows

Previewing the first 50 rows
Reading and processing data sample...

Details Help

LOCATION
Bank

DATA REFINERY FLOW NAME *bank.csv_flow*
Enter a description of the Data Refinery flow

STEPS
0

DATA REFINERY FLOW OUTPUT

Location
Bank/Data assets

Data set name

Descriptive analytics of Data - Profile

IBM Cloud IBM Watson Studio eu-gb.dataplatfrom.cloud.ibm.com/shaper?context=data&dataset_id=12f5e769-abc4-40d8-a061-e827c0c15b2c&project_id=6c5fed2c-7b8b-46c8-b317-affce7b...

IBM Watson Studio Search in your workspaces Buy Sunny Kumar's Account

Projects / Bank / bank.csv / Refine data

Search operations

CLEANSE

- Convert column type
- Convert column value to missing
- Extract date or time value
- Filter
- Remove column
- Remove duplicates
- Remove empty rows
- Replace missing values
- Replace substring
- Substitute
- Text

COMPUTE

- Calculate
- Math

ORGANIZE

Cancel Apply

Use a code template to add a step

Data **Profile** Visualizations

age Integer **job** String

FREQUENCY **BINS: 10**

FREQUENCY

Age Range	Frequency
[34,42]	2500
[26,34]	2200
[42,50]	1800
[50,58]	1400
[58,66]	1000
[18,26]	800
[66,74]	700
[74,82]	600
[82,90]	500
[90,95]	400

Information

Details Help

LOCATION
Bank

DATA REFINERY FLOW NAME *bank.csv_flow*
Enter a description of the Data Refinery flow

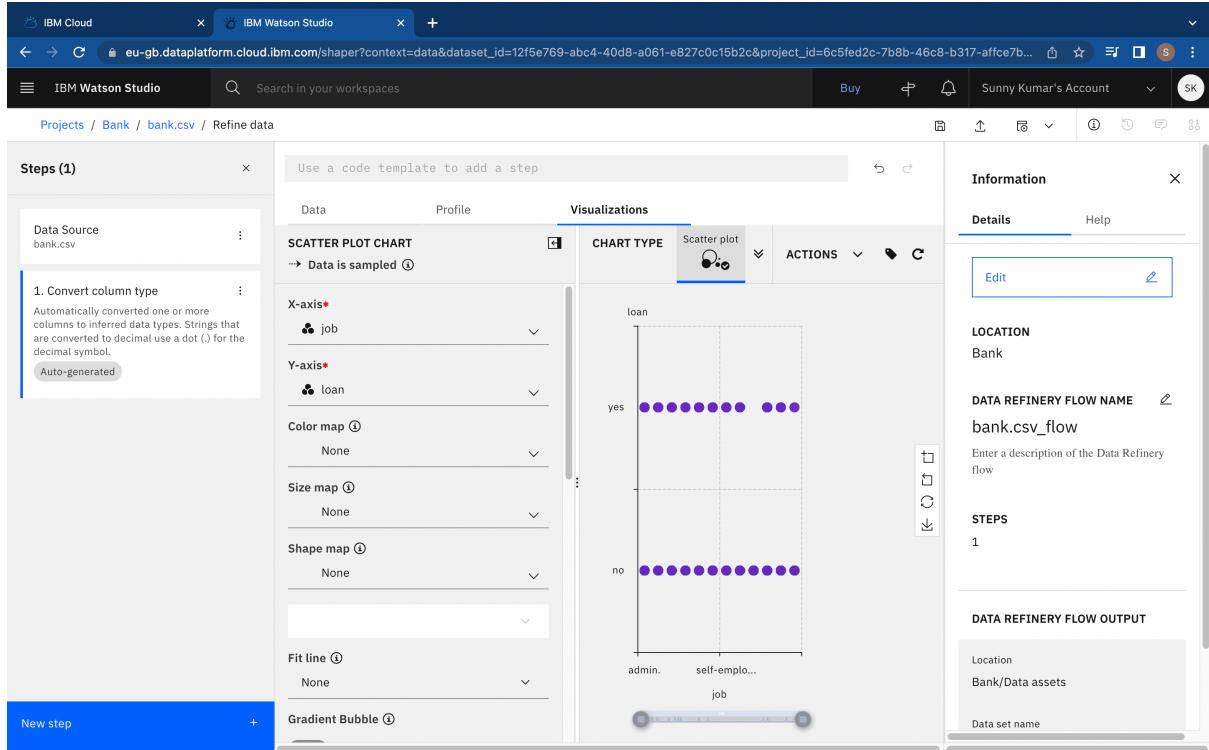
STEPS
1

DATA REFINERY FLOW OUTPUT

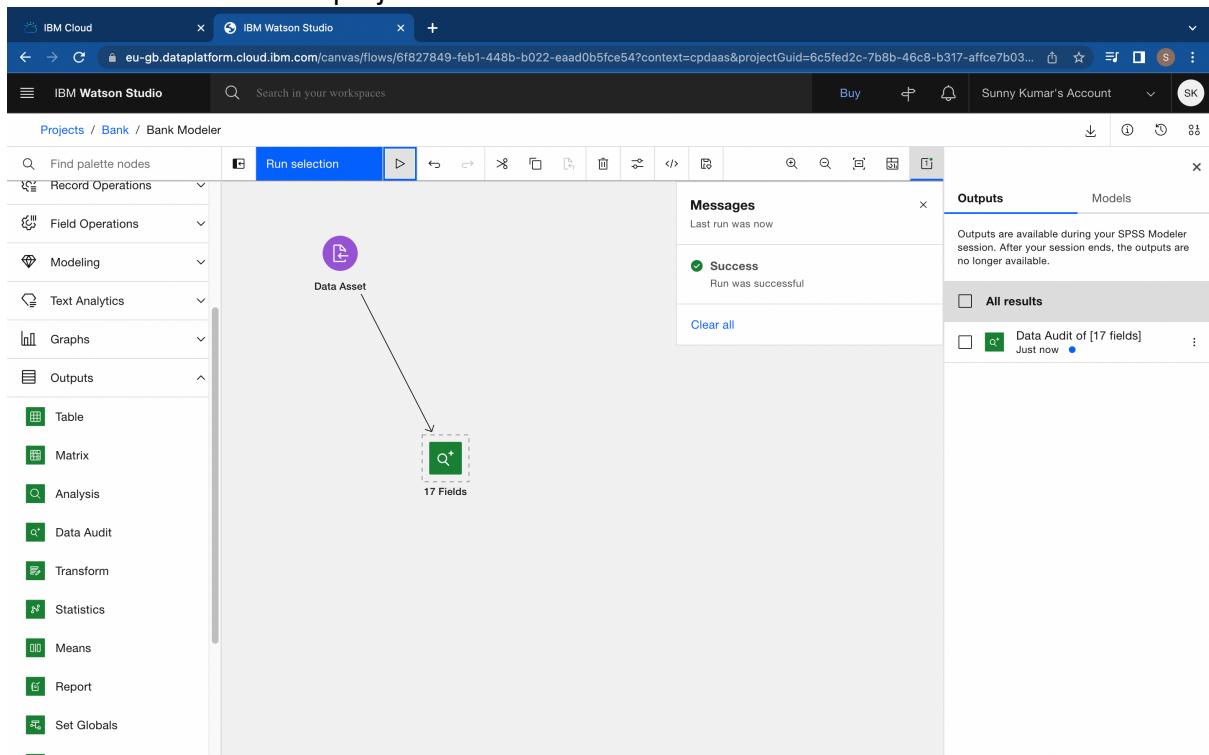
Location
Bank/Data assets

Data set name

Exploratory Analytic- Visualize



Add modeler flow to the project



Output

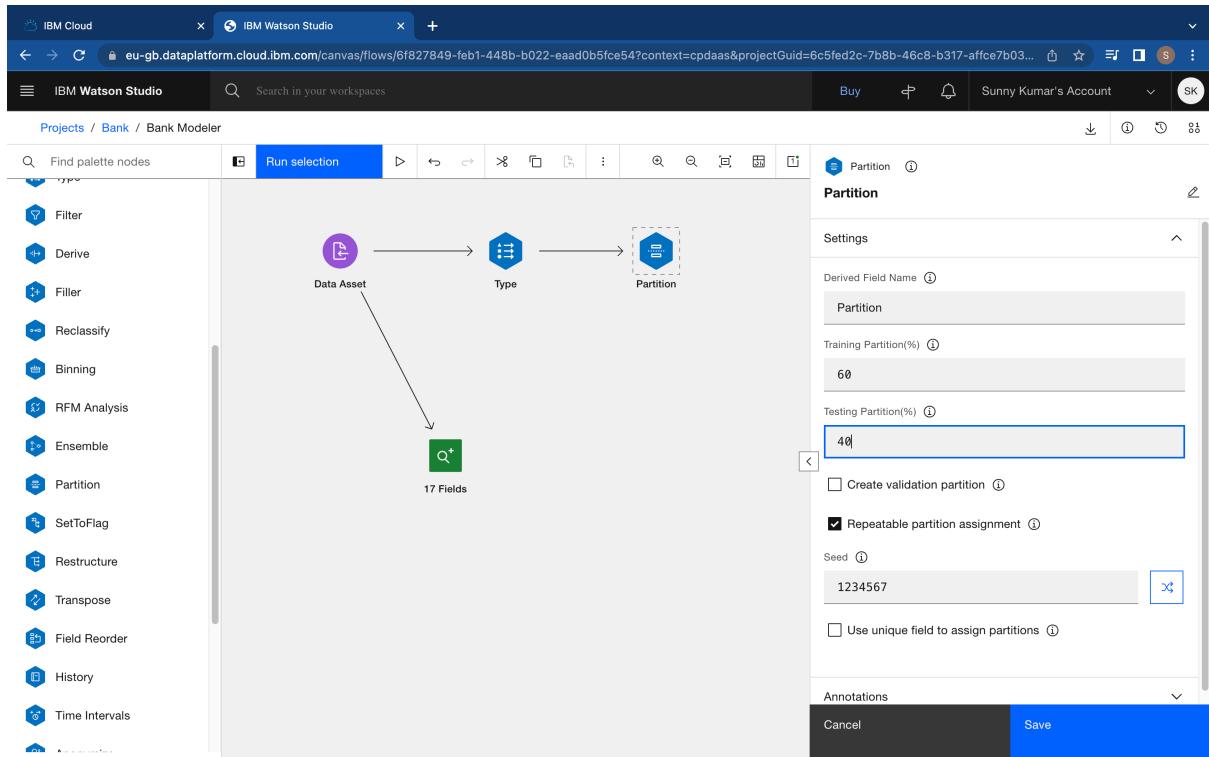
The screenshot shows a data audit report titled "View Output: Data Audit of [17 fields]". The report displays statistical information for each field, including histograms for continuous fields and bar charts for categorical fields. The fields listed are age, job, marital, education, default, balance, housing, and loan.

Field	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1 age	Continuous	18	95	41.232	11.913	0.863	--	11162
2 job	Categorical	--	--	--	--	--	12	11162
3 marital	Categorical	--	--	--	--	--	3	11162
4 education	Categorical	--	--	--	--	--	4	11162
5 default	Categorical	--	--	--	--	--	2	11162
6 balance	Continuous	-6847	81204	1528.539	3225.413	8.225	--	11162
7 housing	Categorical	--	--	--	--	--	2	11162
8 loan	Categorical	--	--	--	--	--	2	11162

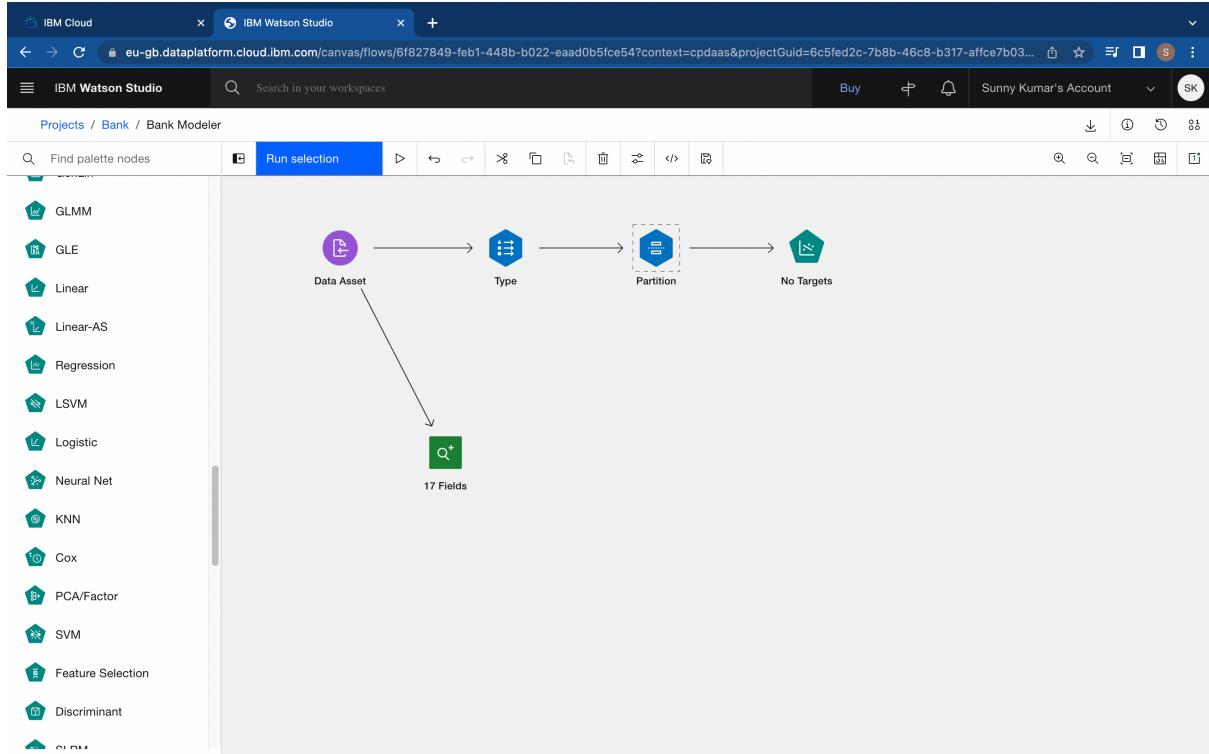
Now add Type Node

The screenshot shows a flow diagram in IBM Watson Studio. A "Data Asset" node is connected to a "Type" node, which then connects to a "17 Fields" node. The "17 Fields" node is highlighted with a dashed green border. The left sidebar shows various nodes under "Import" and "Field Operations". The right sidebar shows the "Outputs" tab with a list of results, including "Data Audit of [17 fi... Just now".

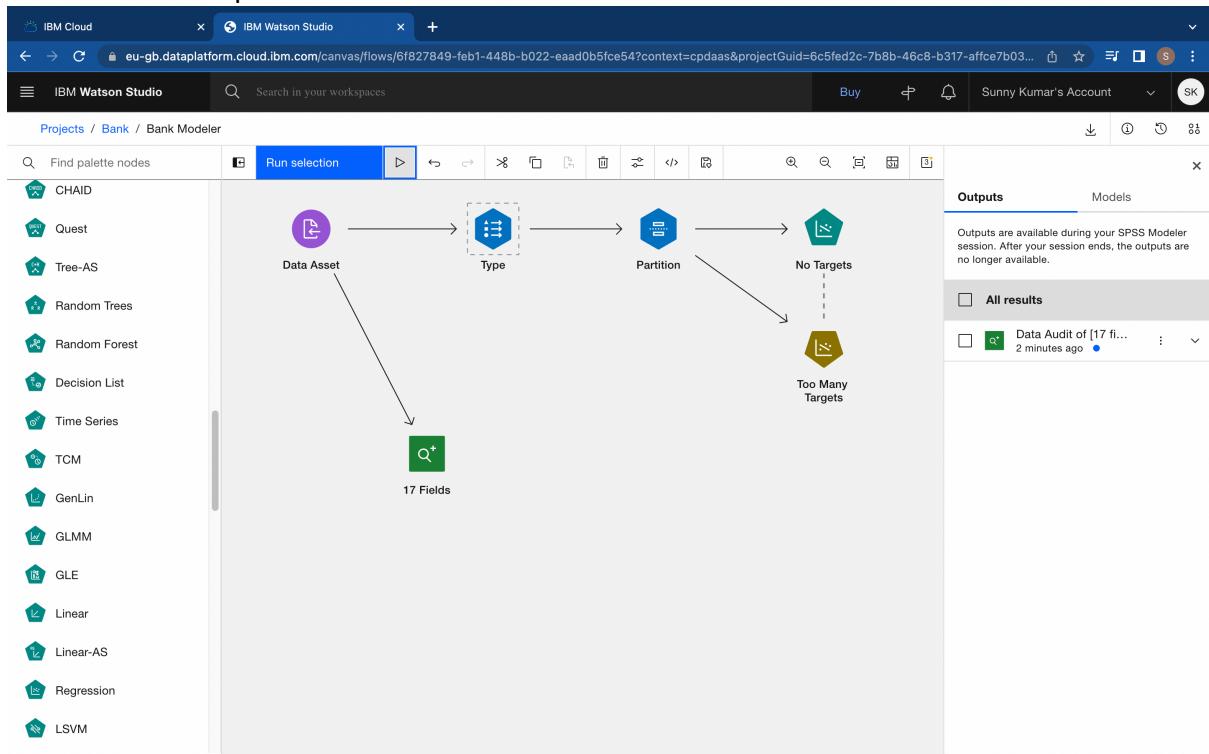
Add Partition Node from field operation and mark 60% 40% partition to Training and Testing fields.



Regression Node



Add Regression Node from Modelling pallet. Connect to Partition Node. It automatically identifies the Output Node



View the model

The screenshot shows the IBM Watson Studio interface with a project titled "Bank Modeler". On the left, a sidebar lists "Regression", "EVALUATION" (with "Model Summary" selected), "ANOVA", "Coefficients", "MODEL VIEWER", "Build Settings", and "Training Summary". The main panel displays the "Model Summary" table:

	1
R	0.005 {1}
R Square	0.000
Adjusted R Square	0.000
Std. Error of the Estimate	8.414

[1] Predictors: (Constant), %1; age