

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

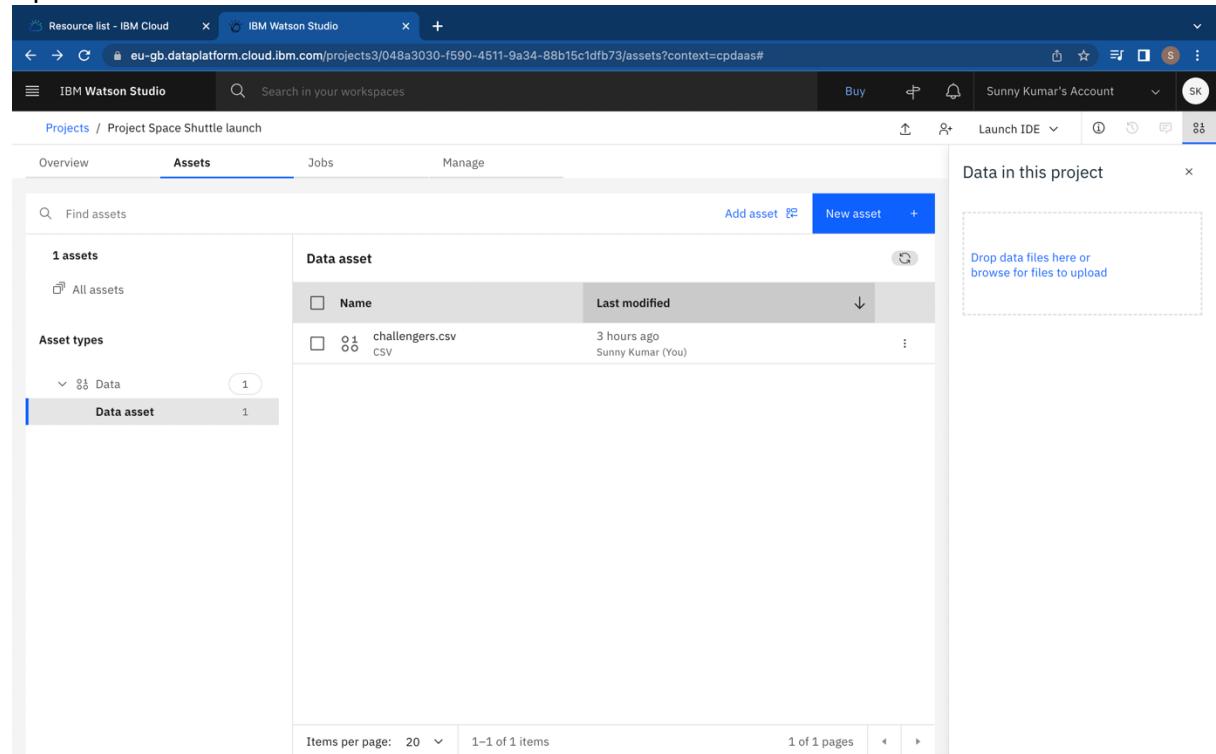
Assignment-2

Data set: challengers.csv

Link: <https://eu-gb.dataplayer.cloud.ibm.com/projects/048a3030-f590-4511-9a34-88b15c1dfb73/assets?context=cpdaas>

- 1) Login to IBM Cloud.
- 2) Go to service and software and select Watson Studio.
- 3) Select Watson Studio and lunch in IBM Cloud Pak for Data.
- 4) Then create a project with suitable name.
- 5) Then upload data asset

Upload Data



The screenshot shows the IBM Watson Studio interface with the 'Assets' tab selected. A single data asset named 'challengers.csv' is listed, which is a CSV file uploaded 3 hours ago by Sunny Kumar (You). To the right, there is a dashed box labeled 'Data in this project' with the instruction 'Drop data files here or browse for files to upload'.

Name	Last modified
challengers.csv	3 hours ago

Profile of data

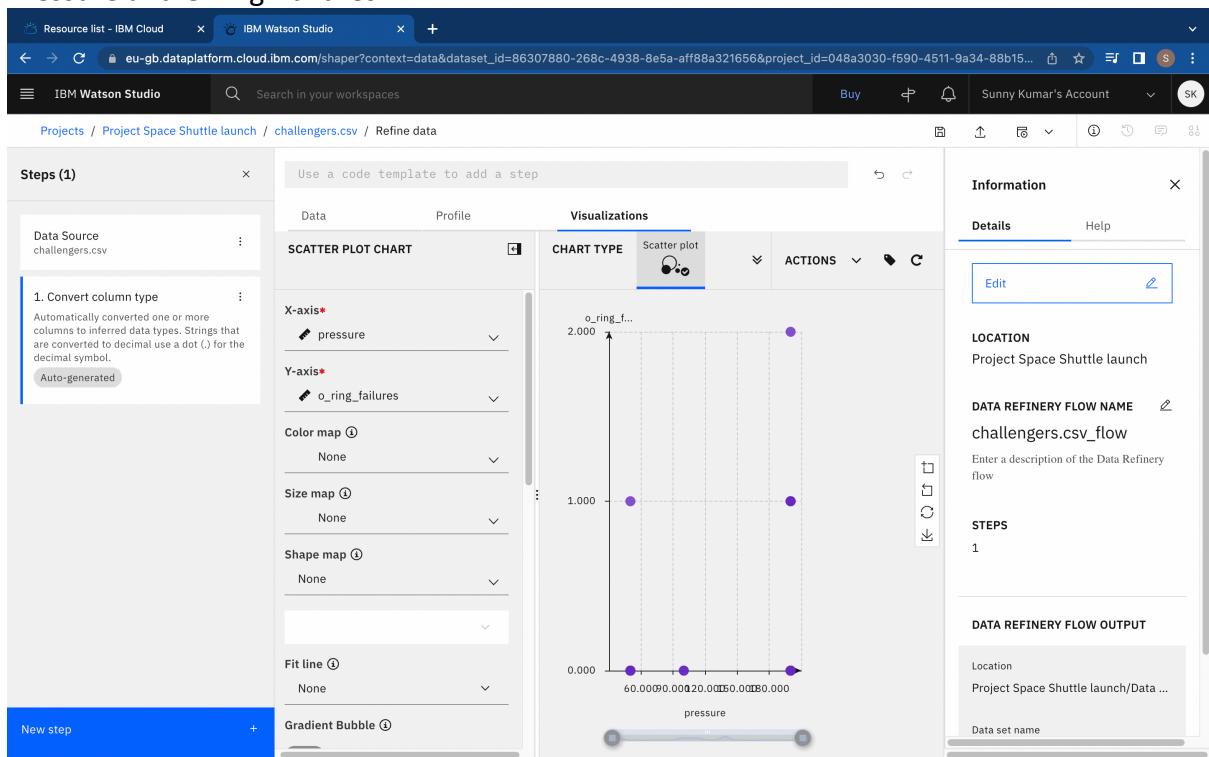
The screenshot shows the 'Profile' step in IBM Watson Studio. The left sidebar lists various data refinement steps under 'CLEANSE', 'COMPUTE', and 'ORGANIZE'. The central area displays two histograms: one for 'o_ring_ct' (frequency vs value) and one for 'o_ring_failures' (frequency vs value). Below the histograms are tables of statistical measures for each column. The right panel contains an 'Information' box with details like 'DATA REFINERY FLOW NAME: challengers.csv_flow' and 'STEPS: 1'.

Visualizations:

Co-relation of Temperature and O-ring-failures

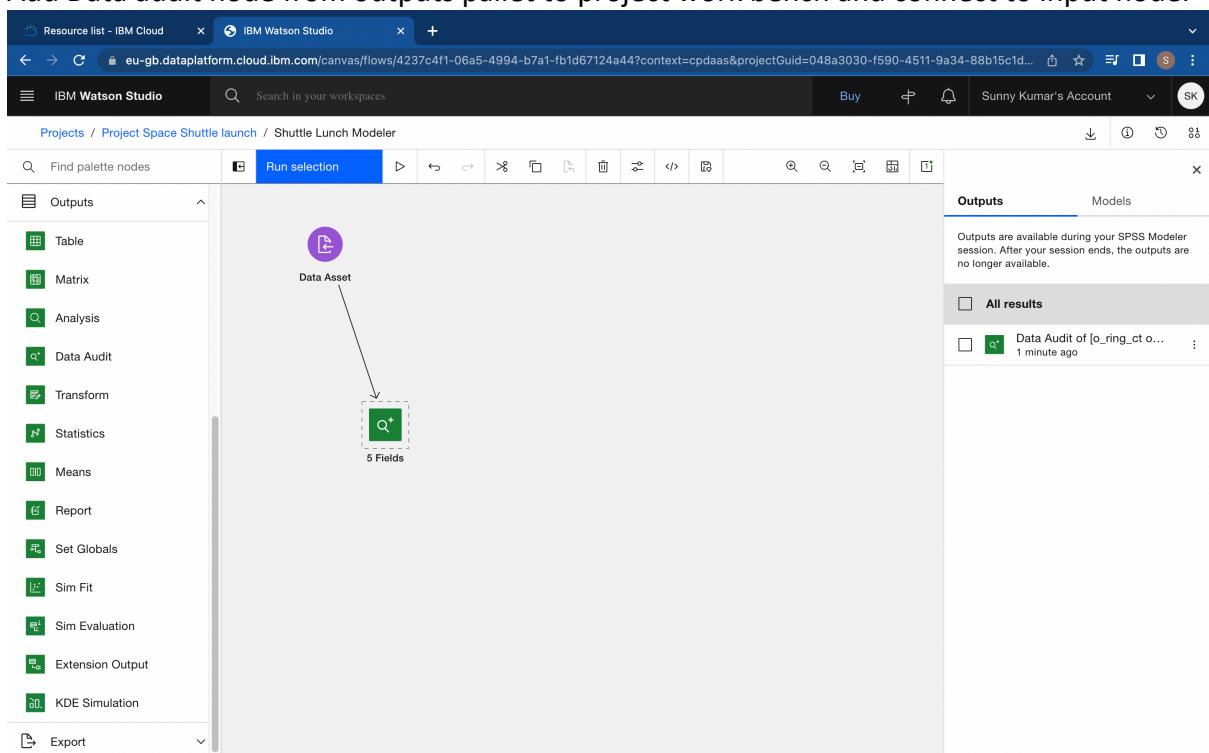
The screenshot shows a 'SCATTER PLOT CHART' visualization. The X-axis is labeled 'temperature' and the Y-axis is labeled 'o_ring_failures'. The plot shows several data points, mostly clustered at a Y-value of 0.000, with a few outliers at 2.000. The right side of the screen displays an 'Information' box with details like 'DATA REFINERY FLOW NAME: challengers.csv_flow' and 'STEPS: 1'.

Pressure and O-ring- failures



Add modeler flow to the project:

Add Data audit node from outputs pallet to project work bench and connect to input node.



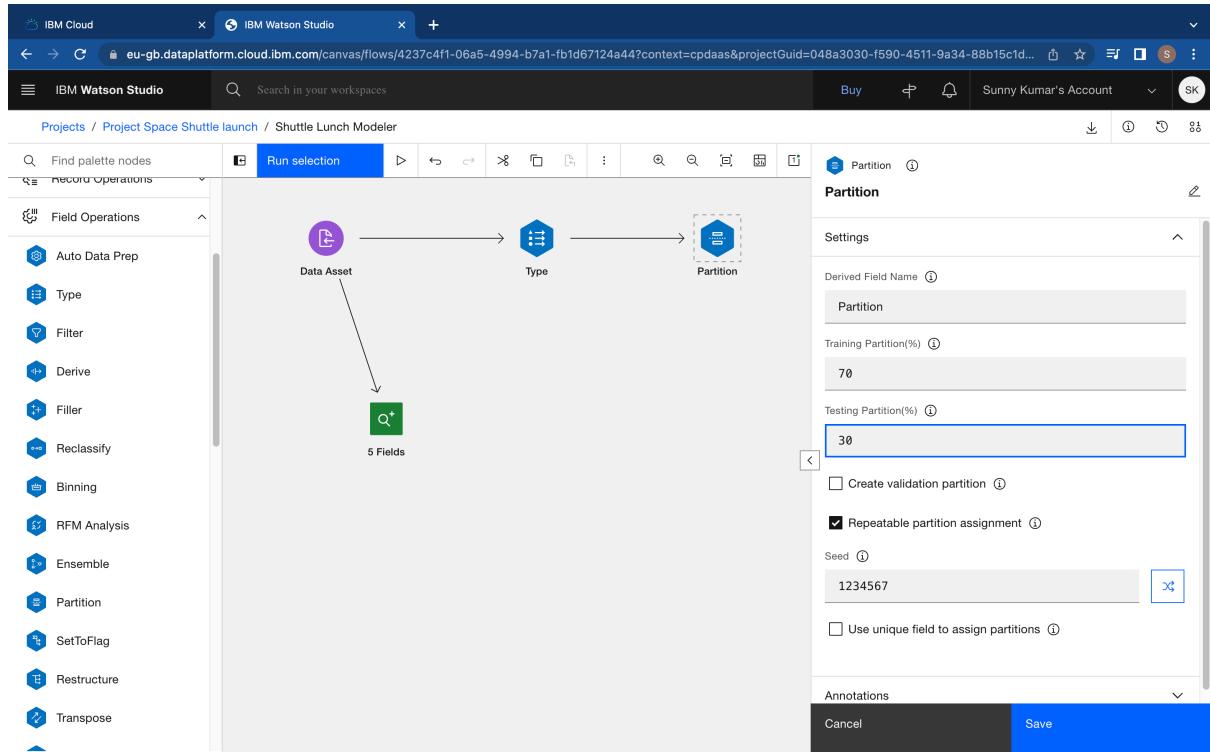
Output of Data Audit Node:

Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1 o_ring_ct		Continuous	6	6	6	0	--	--	23
2 o_ring_failures		Continuous	0	2	0.304	0.559	1.735	--	23
3 temperature		Continuous	53	81	69.565	7.057	-0.654	--	23
4 pressure		Continuous	50	200	152.174	68.221	-0.791	--	23
5 launch_id		Continuous	1	23	12	6.782	0	--	23

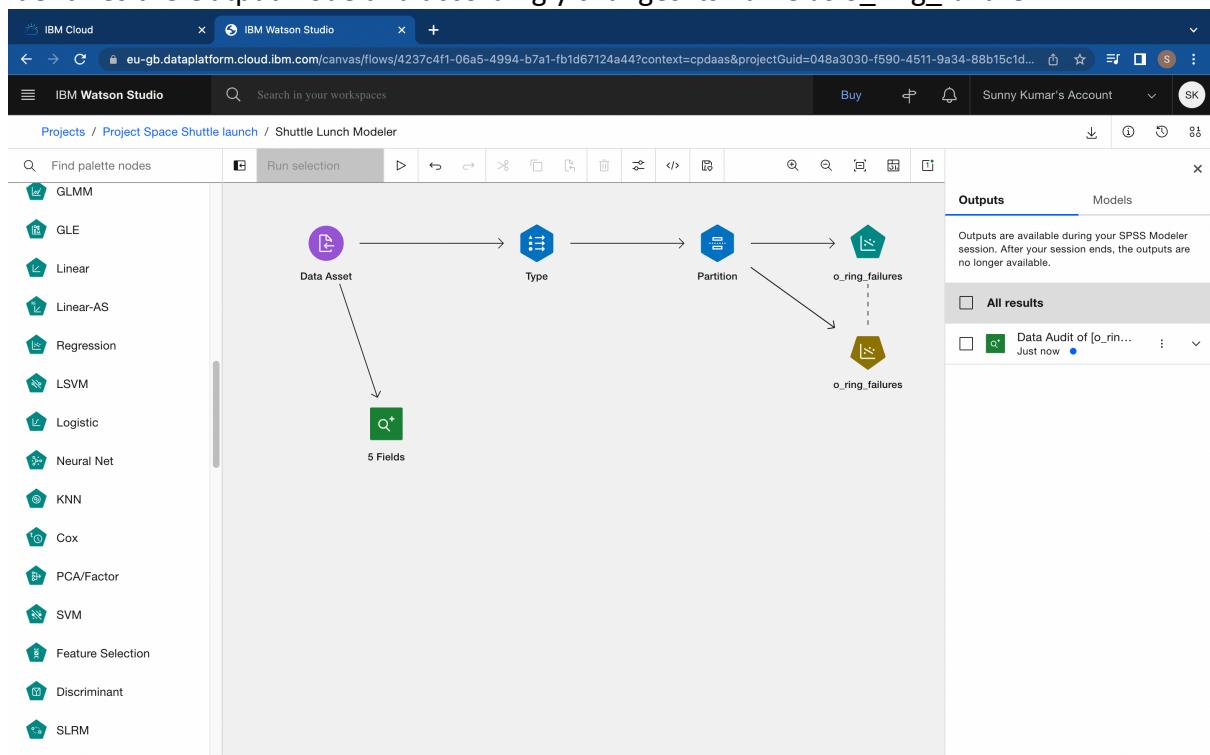
Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Error
1 o_ring_ct	Continuous	0	0	None	Never	Fixed	100.000	23	0	0
2 o_ring_failures	Continuous	1	0	None	Never	Fixed	100.000	23	0	0

Now add Type Node from field Operations connect to Data Asset in the work bench

Add Partition Node from filed operation and mark 70% 30% partition to Training and Testing fileds.



Add Regression Node from Modelling pallet. Connect to Partition Node. It automatically identifies the Output Node and accordingly changes its name as o_ring_failure.



View the Model

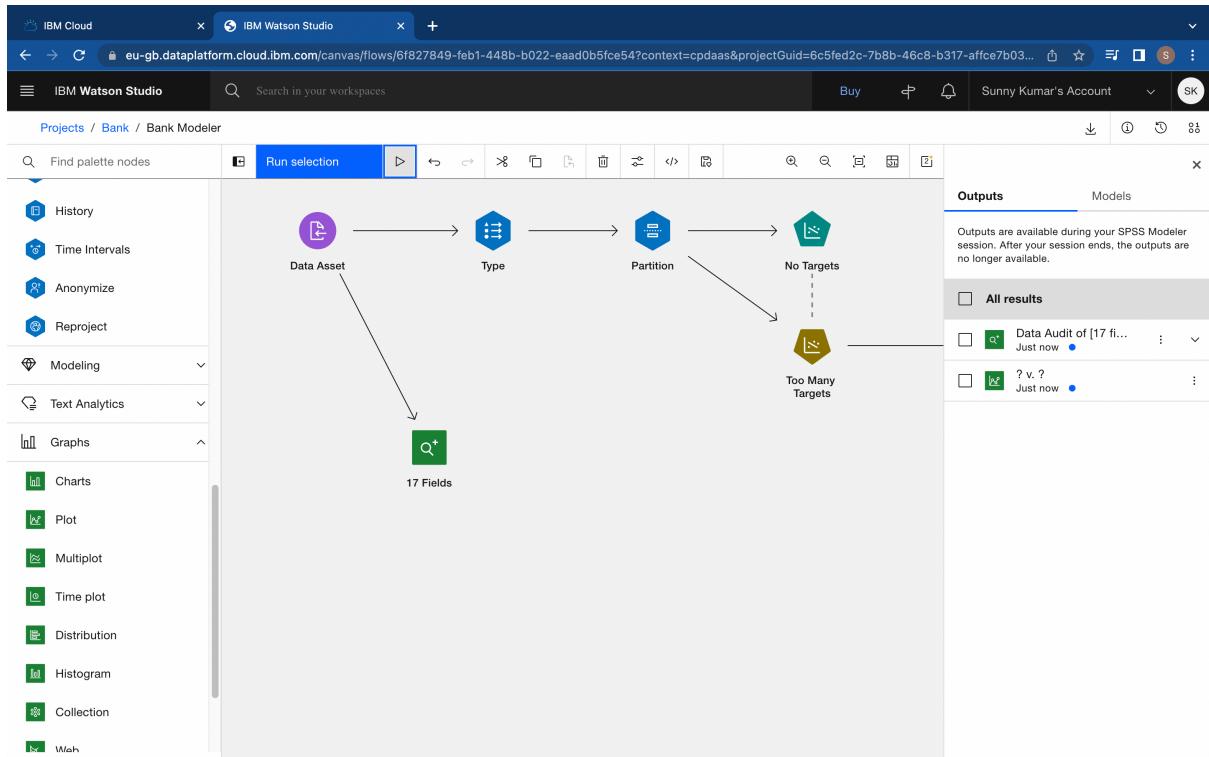
The screenshot shows the 'View Model: o_ring_failures' page in IBM Watson Studio. The left sidebar has sections for Regression, EVALUATION (Model Summary selected), ANOVA, Coefficients, MODEL VIEWER, Build Settings, and Training Summary. The main area displays the 'Model Summary' table:

	I
R	0.766 [1]
R Square	0.587
Adjusted R Square	0.561
Std. Error of the Estimate	0.402

[1] Predictors: (Constant), [%1:, temperature]

Plot Node Properties

The screenshot shows the 'Plot Node Properties' page in IBM Watson Studio. The left sidebar lists various node types: Record Operations, Field Operations, Modeling, Text Analytics, Graphs, Charts, Plot, Multiplot, Time plot, Distribution, Histogram, Collection, Web, Evaluation, and Outputs. The central area shows a flow diagram starting with a 'Data Asset' node, followed by a 'Type' node, a 'Partition' node, and an 'o_ring_failures' node. A dashed arrow points from the 'o_ring_failures' node to a 'Plot' node, which is connected to a '5 Fields' node.



Regression plot output

