

19162121031

SMIT R PATEL

PRACTICAL 17

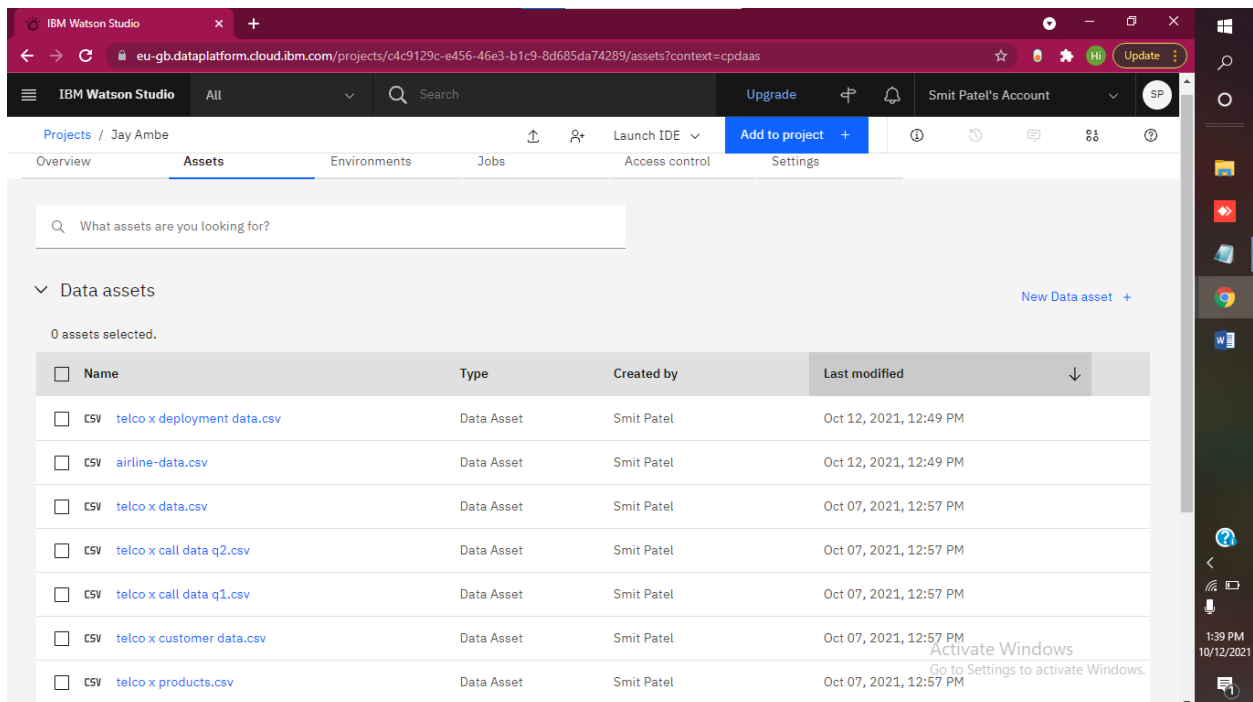
BIG DATA AND ANALYTICS

AIM- To understand the working of Data Refinery in Watson Studio.

Exercise: Acme airlines wants to analyse data with which they need to understand mean delay for some arrival and departure airports. They have a few data files in order to help understand the same. Use a tool in the IBM Cloud in order to get insights and obtain graphs to understand the results better.

Tasks:

1. Upload the data files in your Watson Studio Instance.

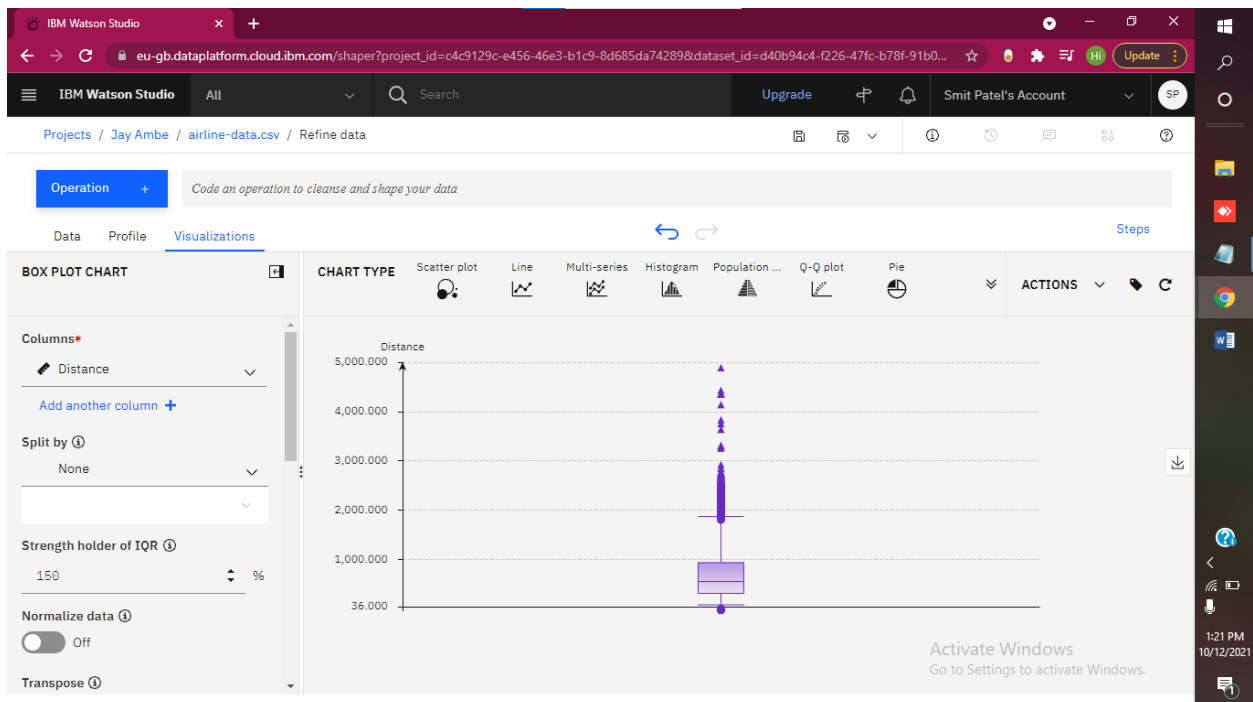
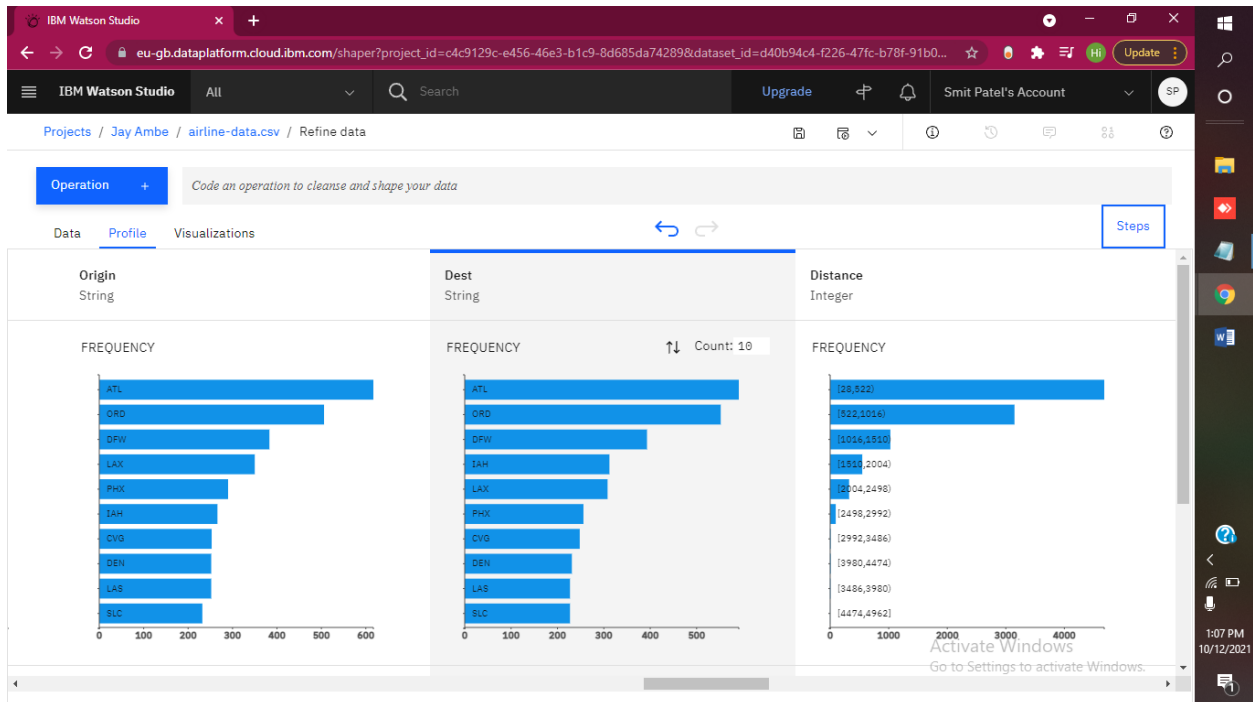


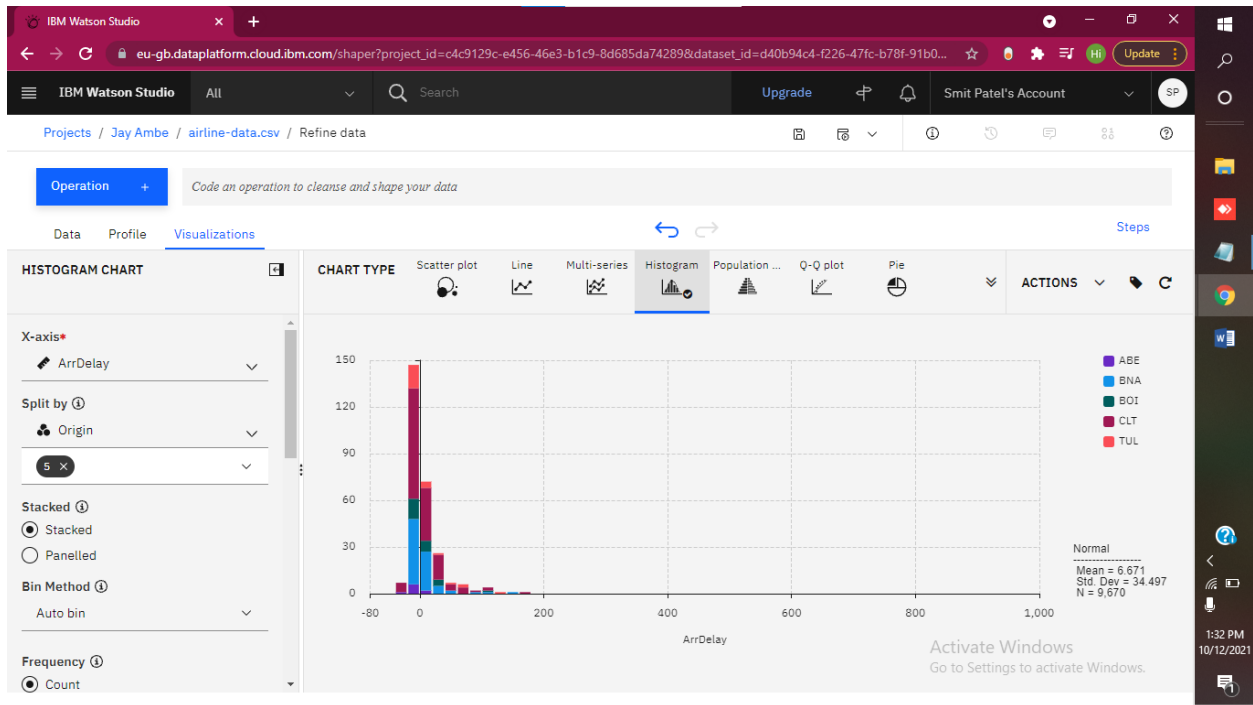
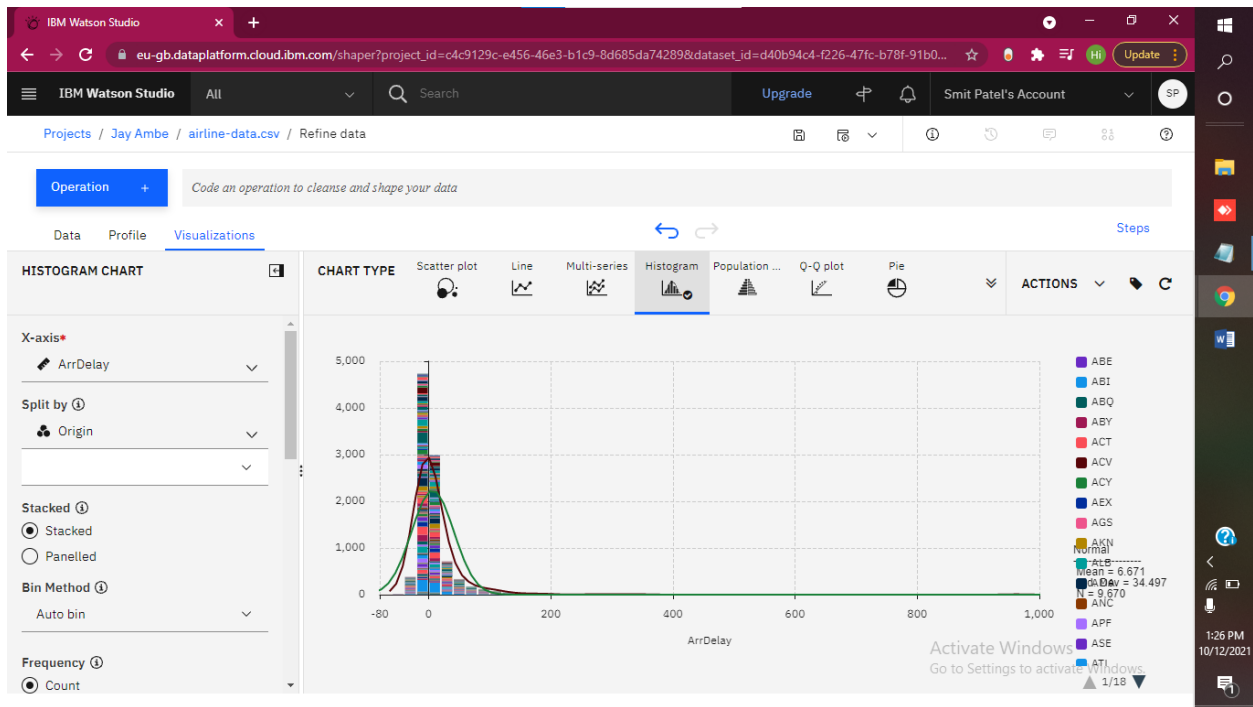
The screenshot displays the IBM Watson Studio web interface. The top navigation bar includes the IBM Watson Studio logo, a search bar, and user account information for Smit Patel. The main content area is titled 'Data assets' and shows a list of seven CSV files. The table columns are Name, Type, Created by, and Last modified. The files are all Data Assets created by Smit Patel, with the first two created on Oct 12, 2021, and the remaining five on Oct 07, 2021. A search bar at the top of the assets list asks 'What assets are you looking for?'. A 'New Data asset' button is visible in the top right corner of the assets list.

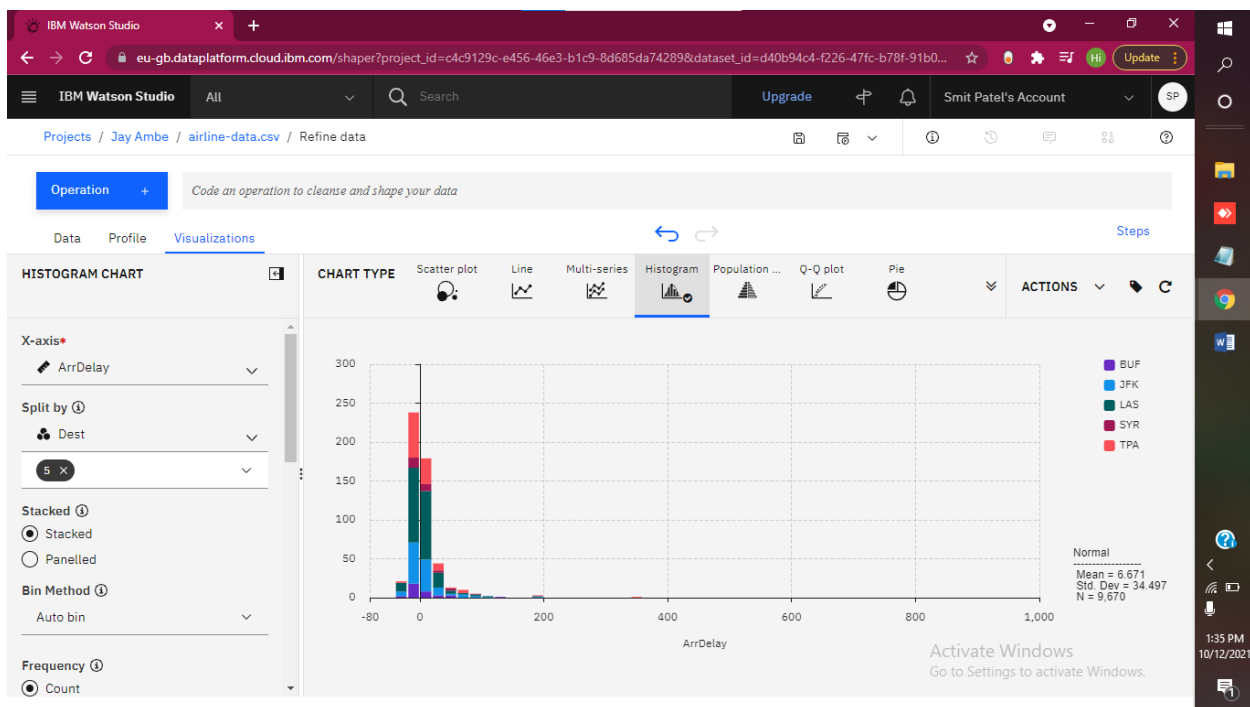
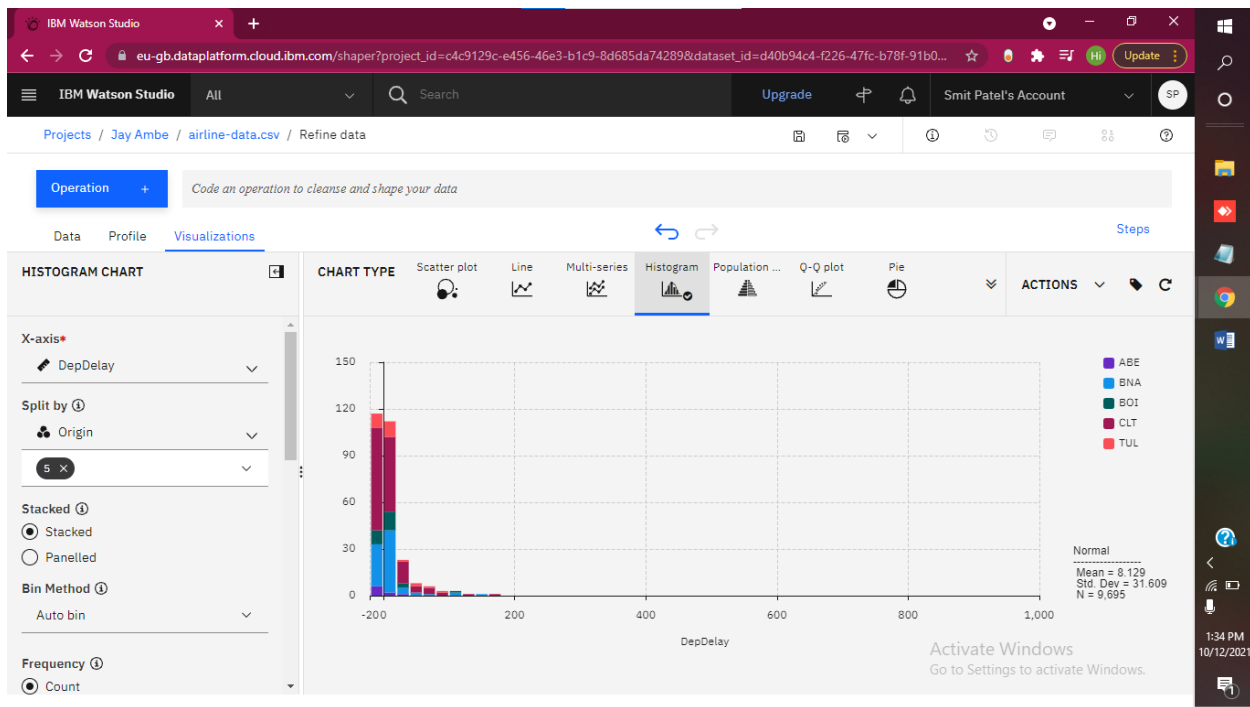
Name	Type	Created by	Last modified
telco x deployment data.csv	Data Asset	Smit Patel	Oct 12, 2021, 12:49 PM
airline-data.csv	Data Asset	Smit Patel	Oct 12, 2021, 12:49 PM
telco x data.csv	Data Asset	Smit Patel	Oct 07, 2021, 12:57 PM
telco x call data q2.csv	Data Asset	Smit Patel	Oct 07, 2021, 12:57 PM
telco x call data q1.csv	Data Asset	Smit Patel	Oct 07, 2021, 12:57 PM
telco x customer data.csv	Data Asset	Smit Patel	Oct 07, 2021, 12:57 PM
telco x products.csv	Data Asset	Smit Patel	Oct 07, 2021, 12:57 PM

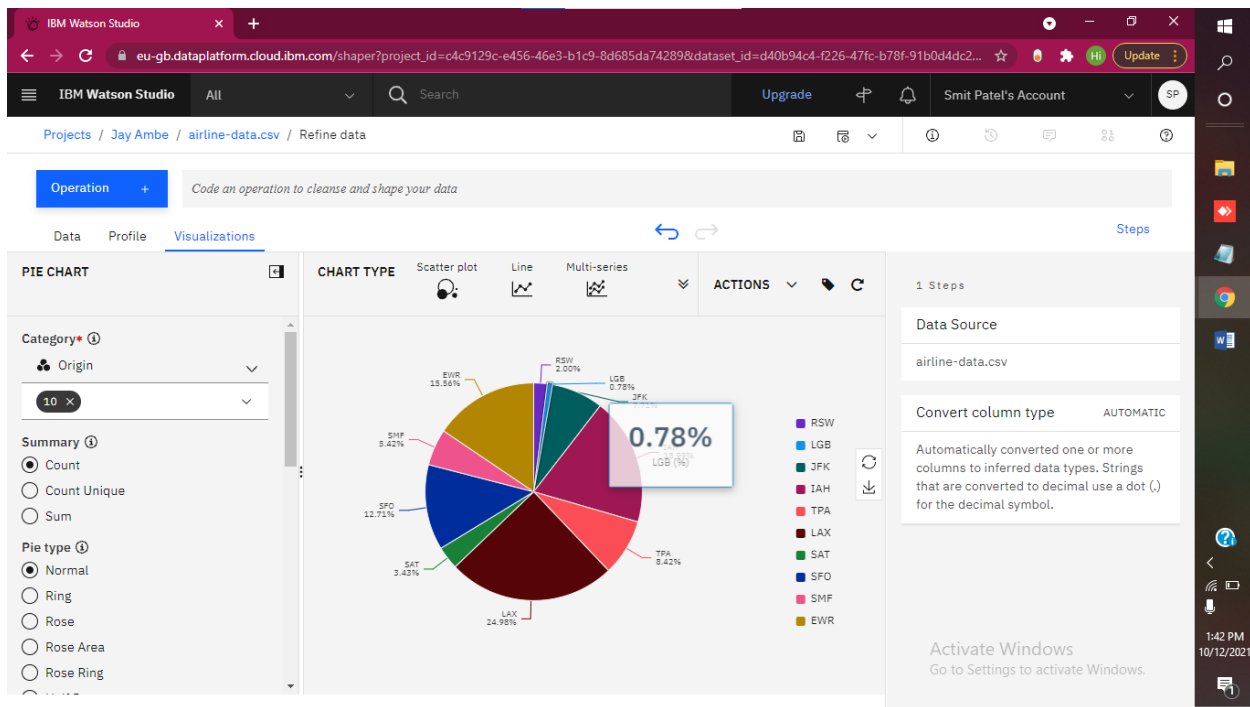
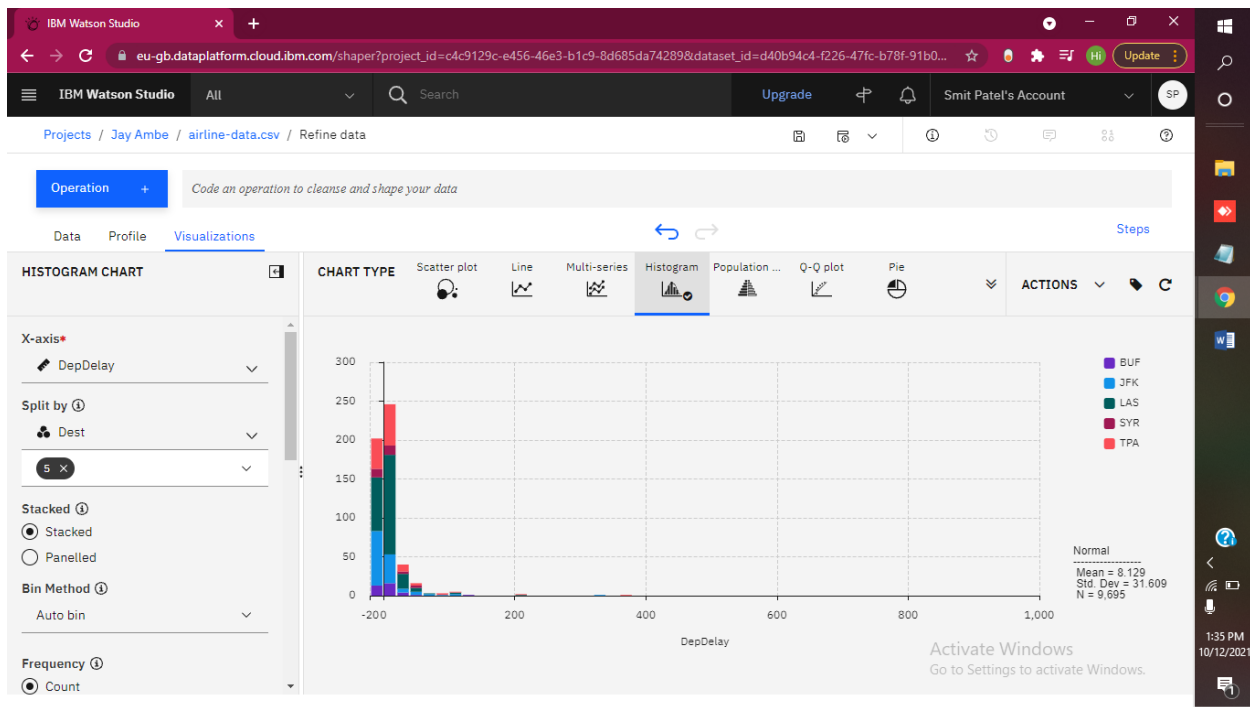
2. Understand basics of Data Refinery.

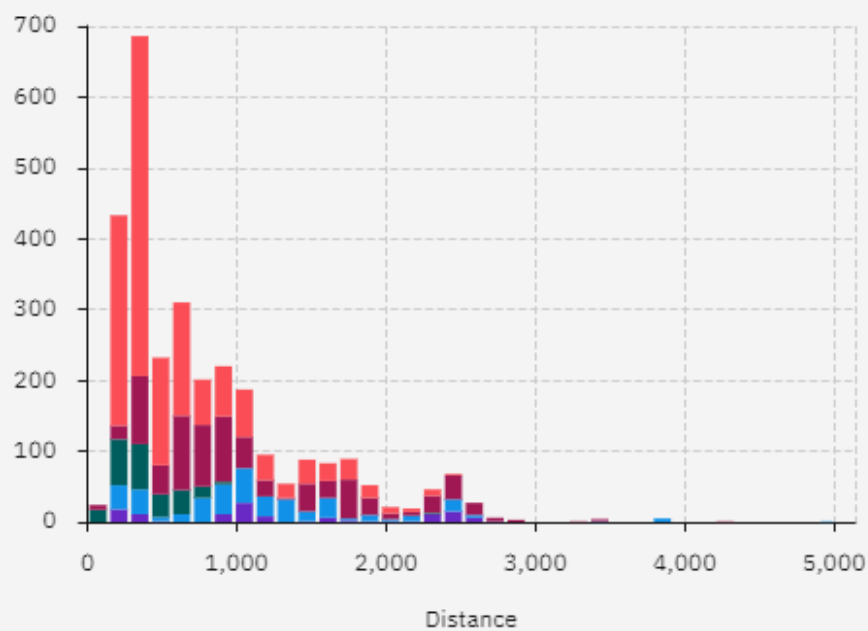
3. Create appropriate data refinery graphs to obtain the result.











B6
CO
DH
UA
WN

Log-normal

Scale = 529.6
Shape = 0.789
N = 9,852

Extra :-

IBM Watson Studio

eu-gb.dataplatform.cloud.ibm.com/shaper?project_id=c4c9129c-e456-46e3-b1c9-8d685da74289&dataset_id=f27518fc-d2e7-46b4-b2c9-0a02289c1...

IBM Watson Studio All Search Upgrade Smit Patel's Account

Projects / Jay Ambe / telco x deployment data.csv / Refine data

Operation + Code an operation to cleanse and shape your data

Data Profile Visualizations Steps

	customer_id String	data_known String	gender String	age Decimal	tariff String	dropped_ Decimal
1	N247880	yes	Female	31	CAT 50	0
2	N376450	yes	Female	42	CAT 50	6
3	N109880	yes	Female	62	CAT 100	9
4	N109885	no		NA		NA
5	N188070	yes	Male	24	CAT 200	0
6	N201040	yes	Female	20	CAT 200	2
7	N164070	yes	Female	31	Play 300	10
8	N201620	yes	Male	15	CAT 200	1
9	N408820	yes	Female	18	CAT 200	12
10	N178490	yes	Male	25	CAT 200	9
11	N348730	yes	Male	31	CAT 100	1
12	N173280	yes	Female	17	CAT 200	2

SOURCE FILE: telco x deployment data.csv SAMPLE SIZE: First 5678 rows

1 Steps

Data Source

telco x deployment data.csv

Convert column type AUTOMATIC

Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.

Activate Windows
Go to Settings to activate Windows.

1:51 PM
10/12/2021

IBM Watson Studio

eu-gb.dataplatform.cloud.ibm.com/shaper?project_id=c4c9129c-e456-46e3-b1c9-8d685da74289&dataset_id=f27518fc-d2e7-46b4-b2c9-0a02289c1...

IBM Watson Studio All Search Upgrade Smit Patel's Account

Projects / Jay Ambe / telco x deployment data.csv / Refine data

Operation + Code an operation to cleanse and shape your data

Data Profile Visualizations Steps

HISTOGRAM CHART

X-axis age

Split by gender

Stacked Stacked

Bin Method Auto bin

Frequency Count

CHART TYPE Scatter plot Line Multi-series

ACTIONS

1 Steps

Data Source

telco x deployment data.csv

Convert column type AUTOMATIC

Automatically converted one or more columns to inferred data types. Strings that are converted to decimal use a dot (.) for the decimal symbol.

Activate Windows
Go to Settings to activate Windows.

1:53 PM
10/12/2021

Log-normal
Scale = 27.94
Shape = 0.425
N = 5,668

