3.2 Data Analysis on Cloud

* Move the Data Set to Azure Synapse Storage Gen2

To move our Loan Dataset to Azure Storage, we used synapsefordbi which was given already.

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

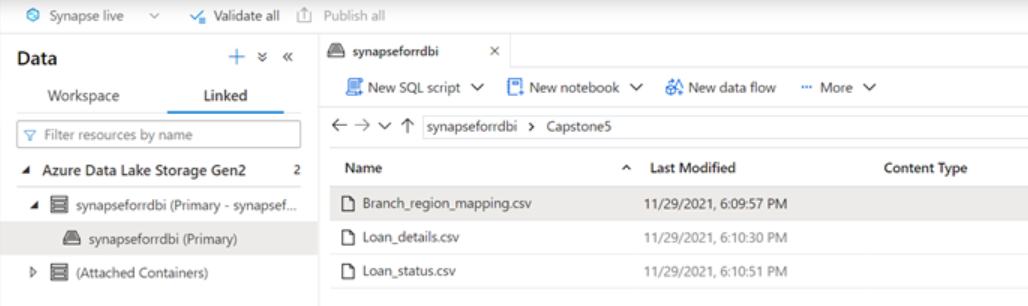
Description automatically generated

Graphical user interface, application

Description automatically generated

Next step is, we created folder named Capstone5 to upload datafiles from local system to azure data storage.

Graphical user interface, text, application, email

Description automatically generated

* Create a serverless SQL pool to query the data from Storage gen1

Graphical user interface, text, application, email

Description automatically generated

After creating serverless SQL pool, we create three external data sources.

1. CREATE EXTERNAL DATA SOURCE BranchRegion

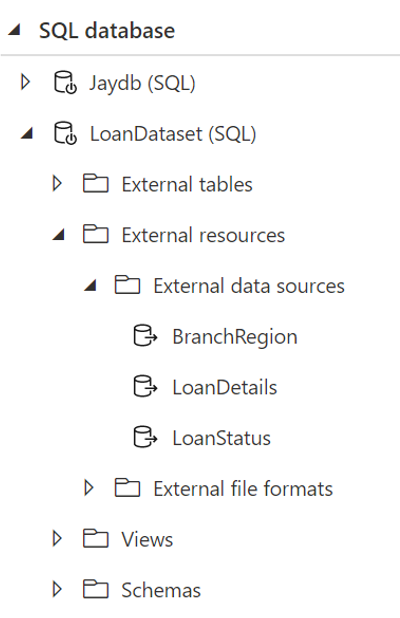
WITH ( LOCATION = 'https://synapase2811.blob.core.windows.net/Capstone5')

1. CREATE EXTERNAL DATA SOURCE LoanDetails

WITH ( LOCATION = 'https://synapase2811.blob.core.windows.net/Capstone5')

1. CREATE EXTERNAL DATA SOURCE LoanStatus

WITH ( LOCATION = 'https://synapase2811.blob.core.windows.net/Capstone5')



Next step is after creating external data sources, we created three views for each data source.

1. CREATE VIEW BranchRegion\_view AS

SELECT

C1 as branch\_id,

C2 as region

FROM OPENROWSET

(BULK 'https://synapseforrdbi.dfs.core.windows.net/synapseforrdbi/Capstone5/Branch\_region\_mapping.csv',

format = 'csv', parser\_version = '2.0', firstrow = 2 ) as rows

1. CREATE VIEW LoanDetails\_view AS

SELECT

C1 as Loan\_id,

C2 as disbursed\_amount,

C3 as asset\_cost,

C4 as ltv,  C5 as branch\_id,

C6 as [Date.of.Birth],

C7 as [Employment.Type],

C8 as DisbursalDate,

C9 as MobileNo\_Avl\_Flag,

C10 as Aadhar\_flag,

C11 as PAN\_flag,

C12 as VoterID\_flag,

C13 as Driving\_flag,

C14 as Passport\_flag,

C15 as [PERFORM\_CNS.SCORE],

C16 as [DELINQUENT.ACCTS.IN.LAST.SIX.MONTHS],

C17 as [CREDIT.HISTORY.LENGTH],

C18 as [NO.OF\_INQUIRIES]

FROM OPENROWSET

(BULK 'https://synapseforrdbi.dfs.core.windows.net/synapseforrdbi/Capstone5/Loan\_details.csv',

format = 'csv', parser\_version = '2.0', firstrow = 2 ) as rows

1. CREATE VIEW LoanStatus\_view AS

SELECT

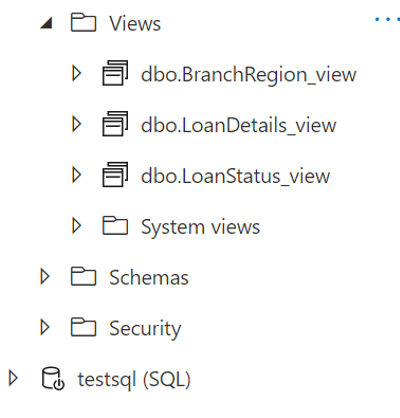
C1 as Loan\_default,

C2 as loan\_id

FROM OPENROWSET

(BULK 'https://synapseforrdbi.dfs.core.windows.net/synapseforrdbi/Capstone5/Loan\_status.csv',

format = 'csv', parser\_version = '2.0', firstrow = 2 ) as rows

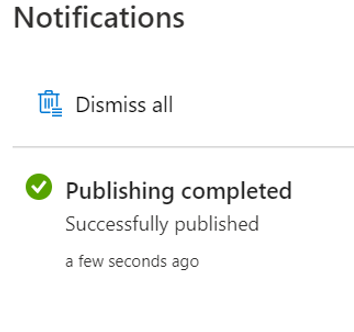


After creating views, we queried our dataset to view data.

Graphical user interface, text

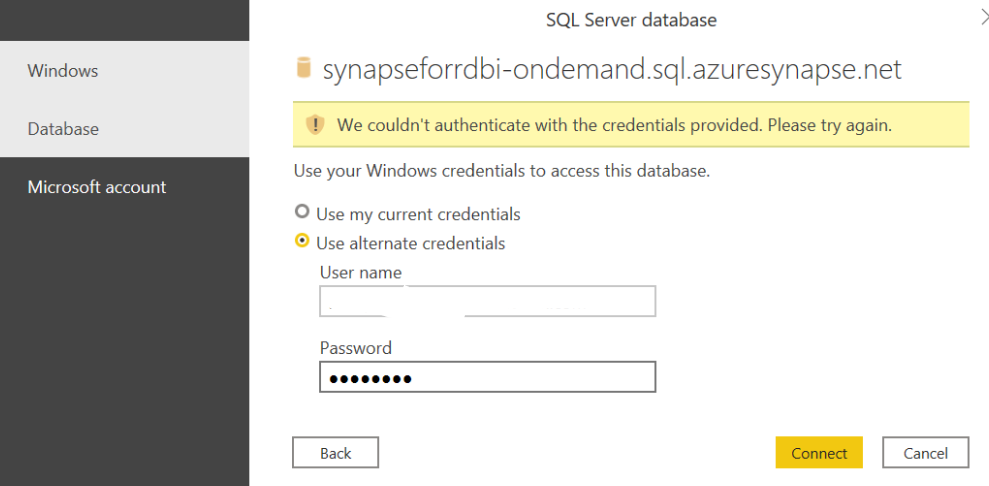
Description automatically generated with medium confidenceTable

Description automatically generatedGraphical user interface, text

Description automatically generated 

* Create a Linked service to PowerBI

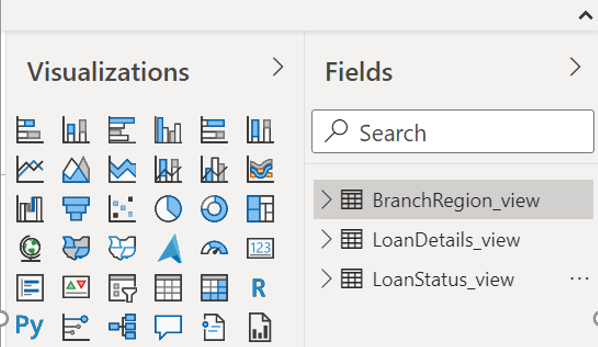
To link azure with PowerBI, in the desktop Power Bi “Get Data” from azure.



Graphical user interface, table

Description automatically generated

After loading dataset in Local system Power Bi from Azure Storage, will be able to see data fields.



Below queries shows, how dataset is connected to azure synapse.

Graphical user interface, application

Description automatically generatedGraphical user interface, application, Word

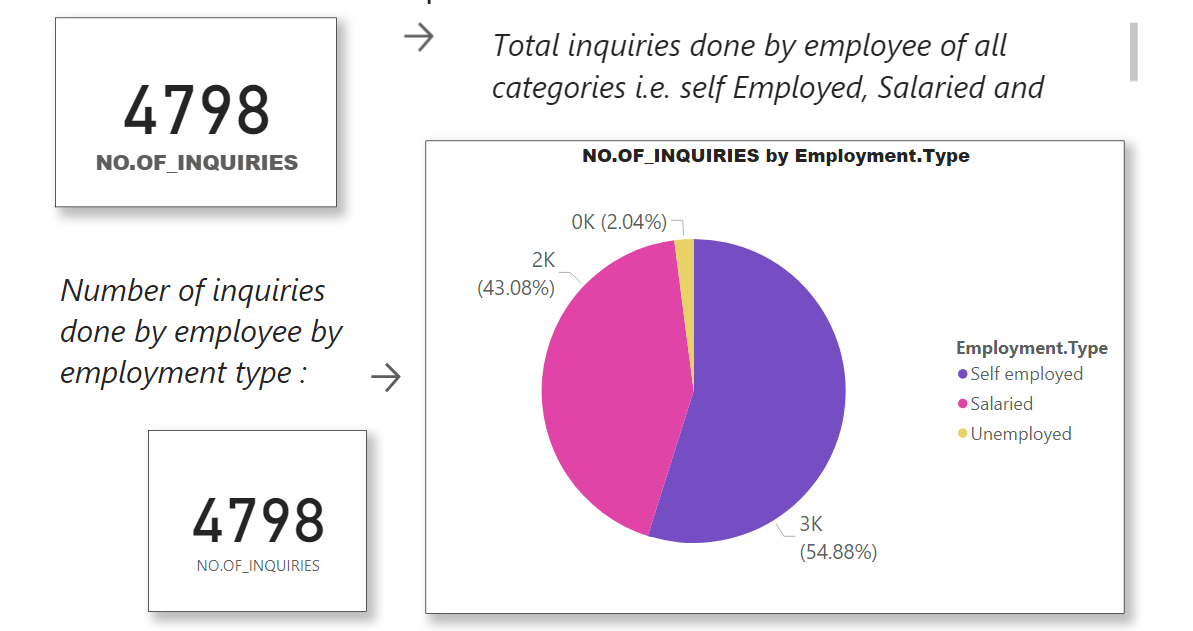
Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

* Perform various analytics on PowerBI

After loading dataset successfully, various analytics performed as given below.

* Ensure you have sufficient privileges on Synapse to access the serverless sql pool.
* Perform the tasks mentioned in Task 2.3
  + What were the total enquiries done?



* + What was the maximum asset cost?



* + What is the average asset cost for each employment type?

Table

Description automatically generated with medium confidence

* + What is the average loan default for each driving flag?

Chart

Description automatically generated

* + Display to Key Influencer Visual for the appropriate columns and indicate your inferences.

Chart, bubble chart

Description automatically generatedGraphical user interface, application

Description automatically generated

* + Display loan default by employment type and indicate which employment type has the highest loan default.

Graphical user interface, chart, application, pie chart

Description automatically generated

* + Display a decomposition tree for the data.

A picture containing graphical user interface

Description automatically generated