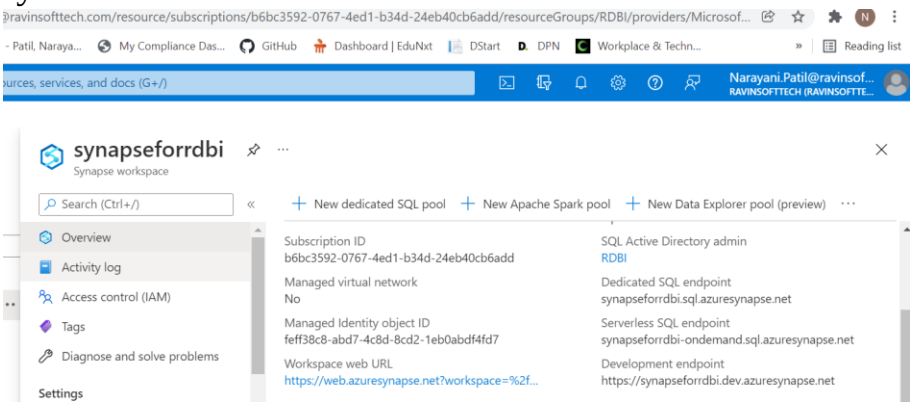
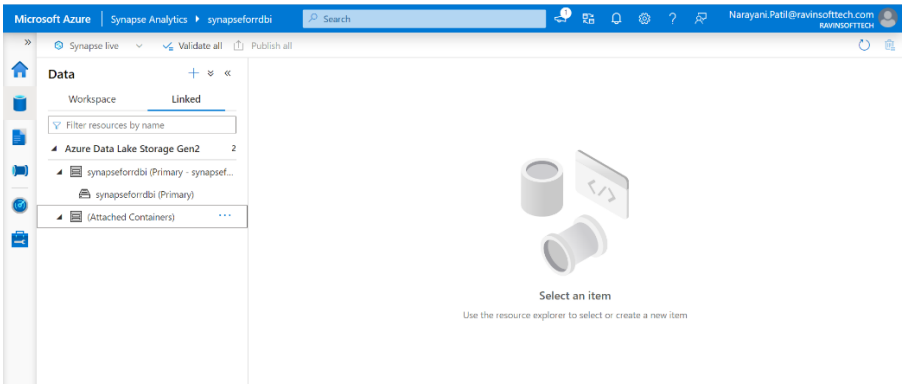
3.2 Data Analysis on Cloud

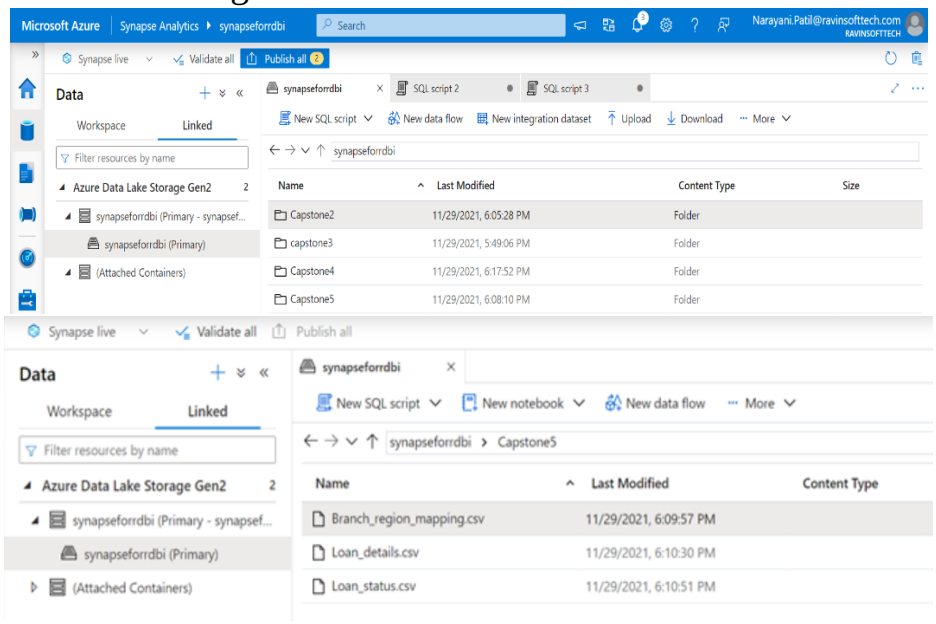
- Move the Data Set to Azure Synapse Storage Gen2

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

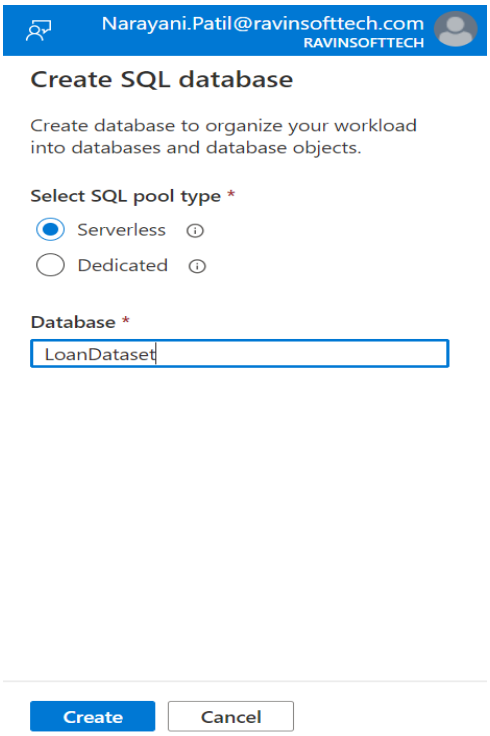




Next step is, we created folder named Capstone5 to upload datafiles from local system to azure data storage.

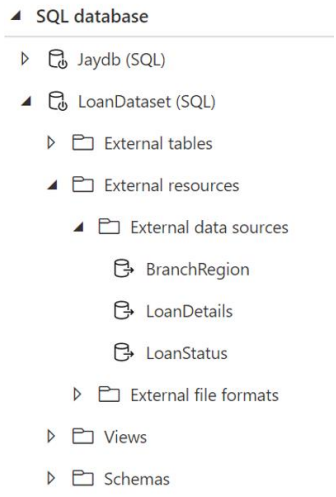


- Create a serverless SQL pool to query the data from Storage gen1



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')
```

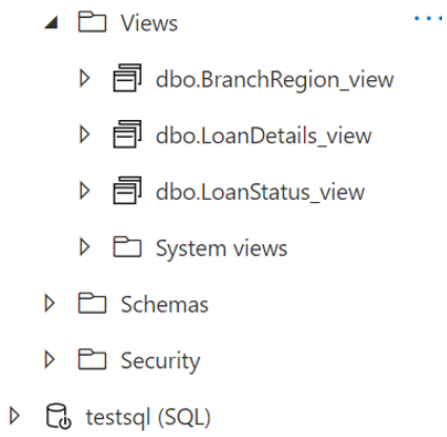


Next step is after creating external data sources, we created three views for each data source.

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

2. 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

3. 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.

```
select * from dbo.BranchRegion_view
```

Results | Messages

View: Table | Chart — Export results

Search

C1	C2
1	East
2	East
3	East
5	East
7	East
8	East
9	East

SQL script 2

Run | Undo | Publish | Query plan | Connect to: Built-in

```
select * from dbo.LoanDetails_view
```

Results | Messages

View: Table | Chart — Export results

Search

C1	C2	C3	C4	C5
1	36439	65850	56.19	64
2	48749	69303	72.15	67
3	55348	66340	85	2
4	48849	64133	77.96	217
5	40394	59386	70.72	74
6	51803	67466	79.3	162

00:00:09 Query executed successfully.

SQL script 2

Run | Undo | Publish | Query plan | Connect to: Built-in

```
select * from dbo.LoanStatus_view
```

Results | Messages

View: Table | Chart — Export results

Search

C1	C2
1	0
2	0
3	1
4	0
5	0
6	0
7	0

00:00:03 Query executed successfully.

Notifications

Dismiss all

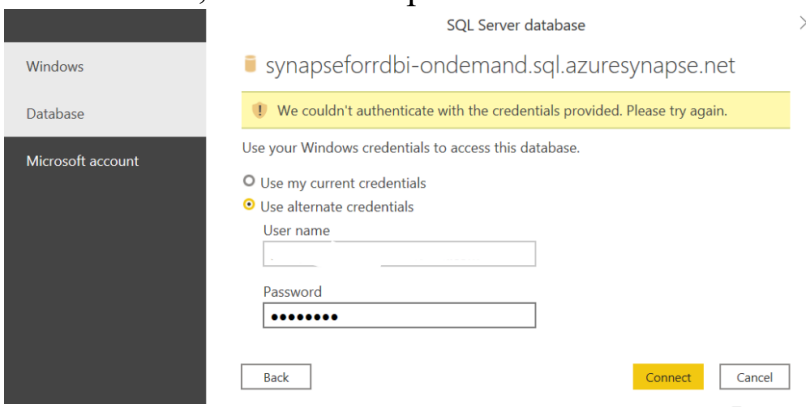
Publishing completed

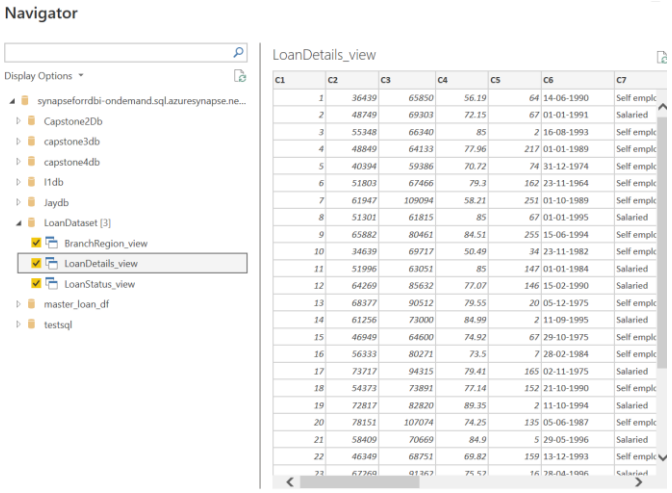
Successfully published

a few seconds ago

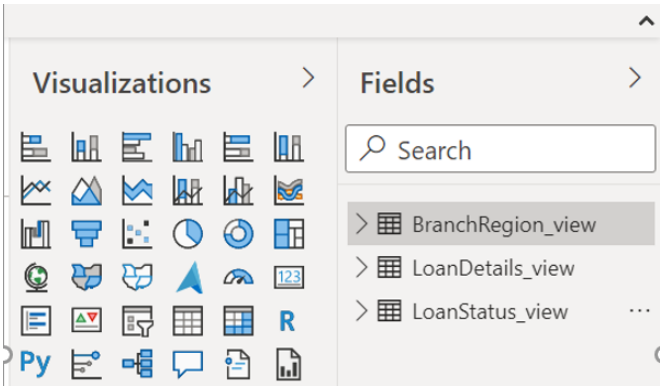
- Create a Linked service to PowerBI

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

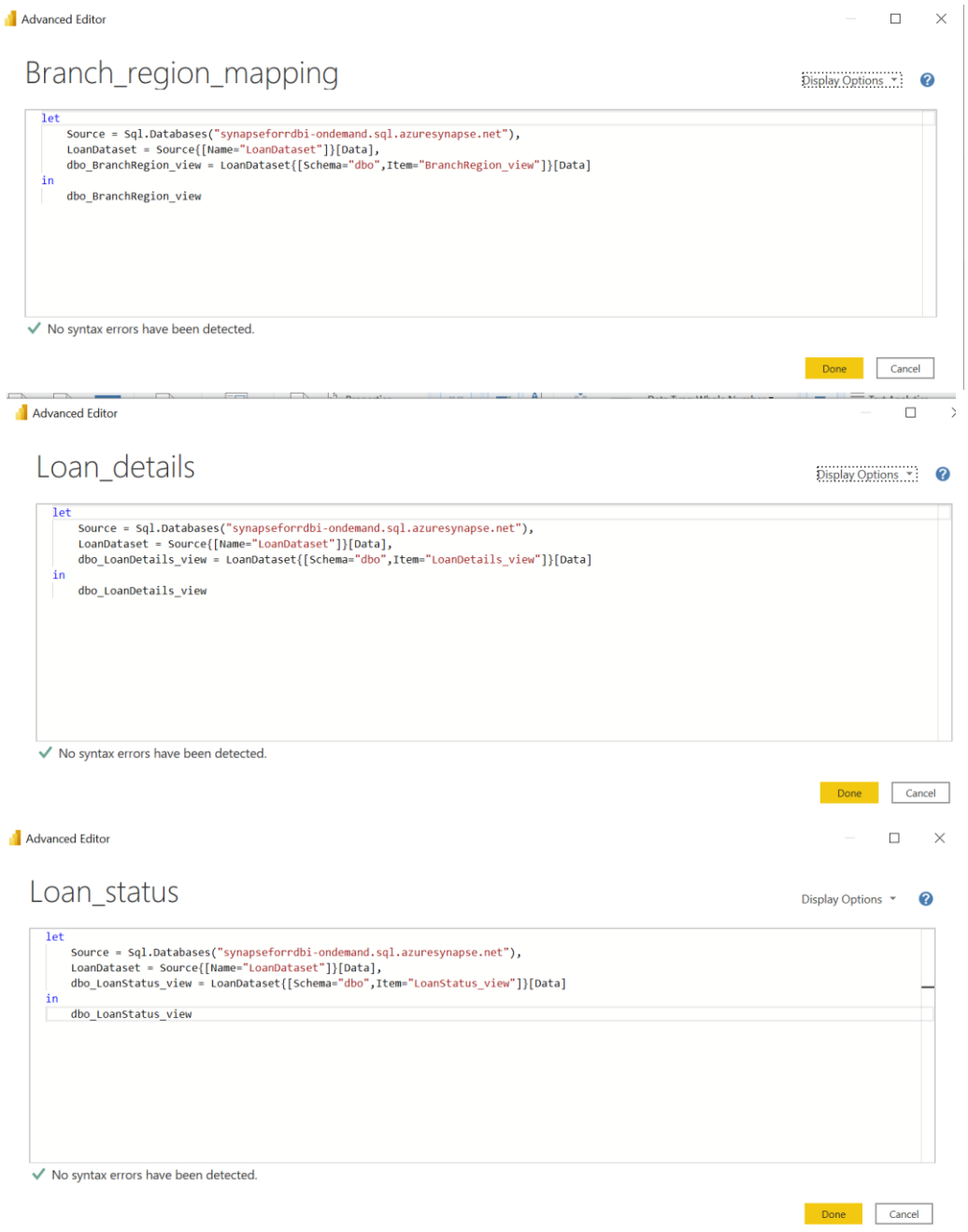




After loading dataset in Local system Power Bi from Azure Storage, will be able to see data fields.



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



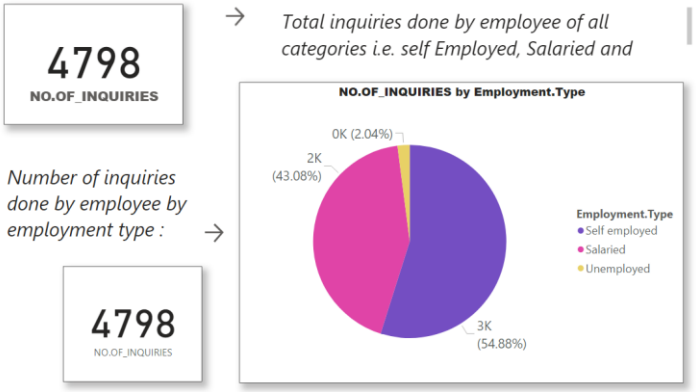
- Perform various analytics on PowerBI

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?

- What is the average asset cost for each employment type?

→

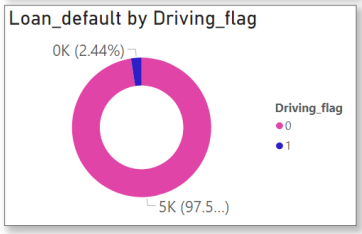
Employment.Type	Average of asset_cost
Self employed	76536.20
Salaried	74384.49
	82966.04
Total	**75842.18**

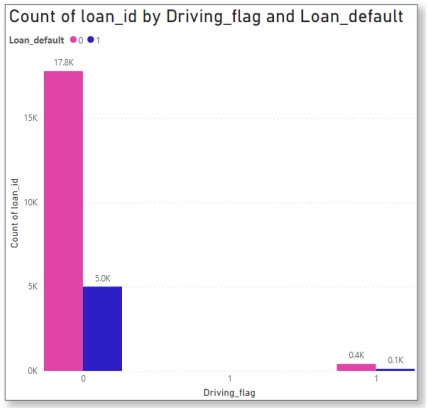
Total average asset cost is : 75.84K
Average of asset_cost

- What is the average loan default for each driving flag?
  - What is the average loan default for each driving flag?

→

Driving_flag	0	1	Total
0	17774	5001	**22775**
1	415	125	**540**
Total	**18189**	**5126**	**23315**

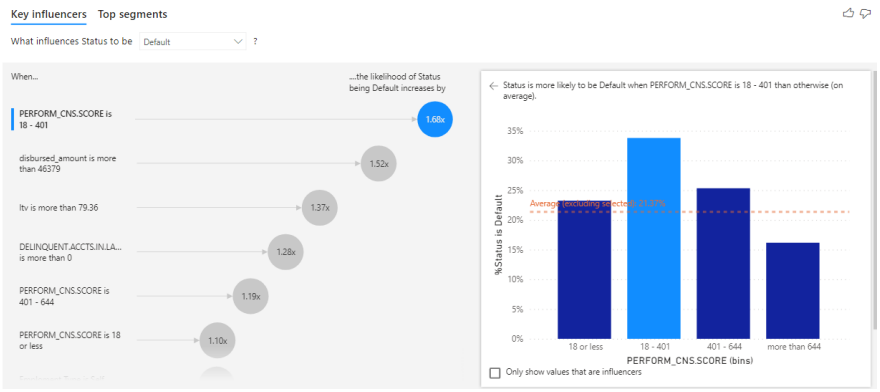




- Display to Key Influencer Visual for the appropriate columns and indicate your inferences.

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

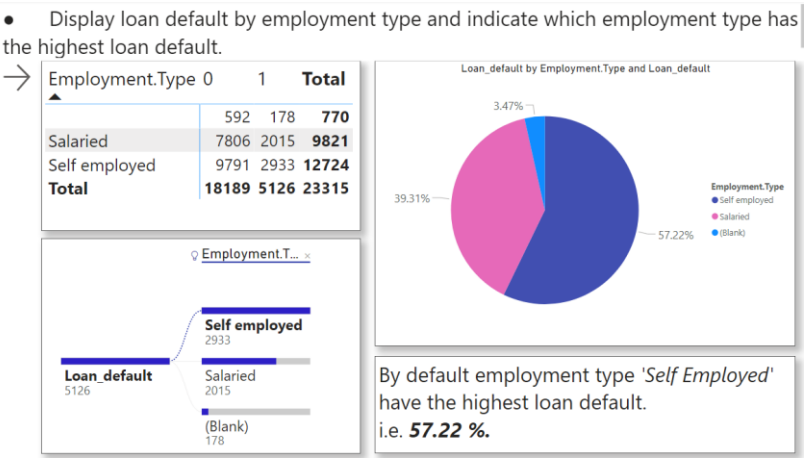
→ It is showing the top contributors to the data metric.



- Display to Key Influencer Visual for the appropriate columns and indicate inferences.

→ It is showing the top contributors i.e. here, ***Employment Type*** to the data metric.

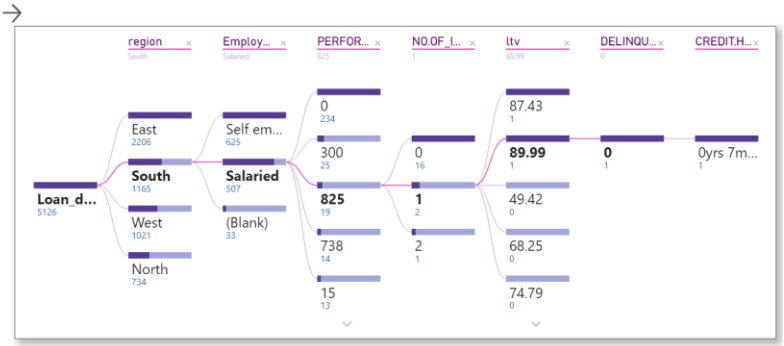
- Display loan default by employment type and indicate which employment type has the highest loan default.



- Display a decomposition tree for the data.

- Display a decomposition tree for the data.

→



Decomposition tree for 'Loan Data', helping us to ***understand and visualize across multiple dimensions***.