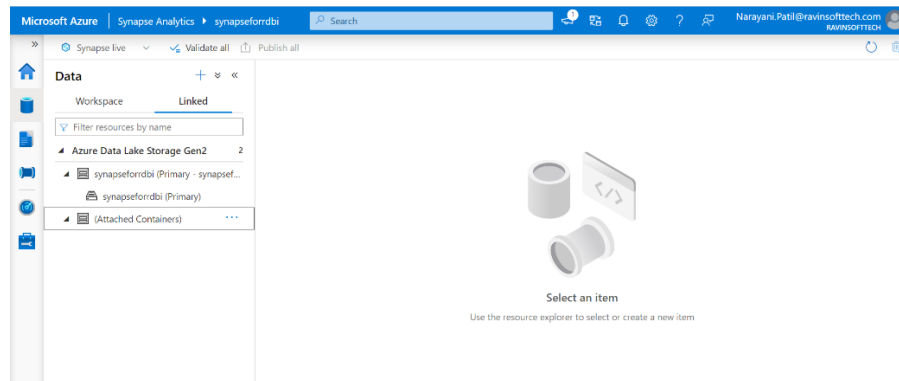
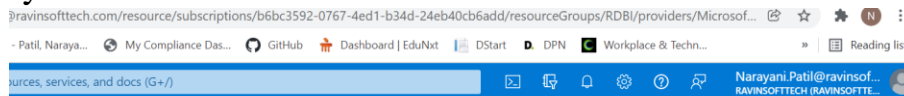


3.2 Data Analysis on Cloud

- Move the Data Set to Azure Synapse Storage Gen1

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.

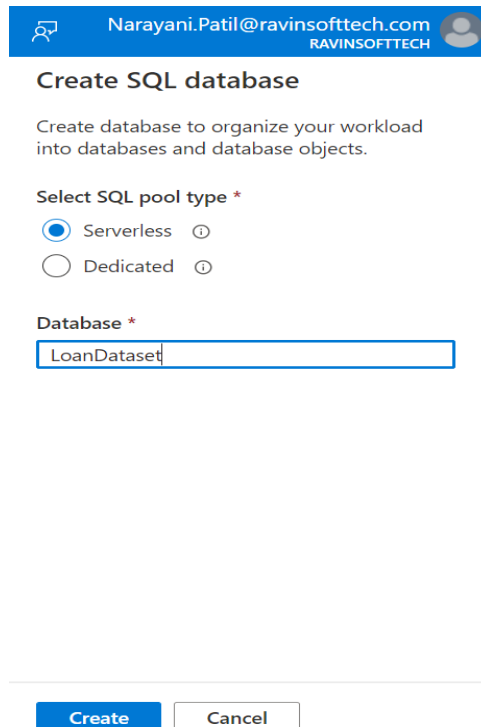
The screenshot displays the Microsoft Azure Synapse Analytics interface. The top navigation bar shows 'Microsoft Azure | Synapse Analytics | synapseforrdbi'. The left sidebar contains a 'Data' section with 'Workspace' and 'Linked' tabs. The 'Linked' tab is active, showing a list of resources under 'Azure Data Lake Storage Gen2'. The main pane shows a table of folders within the 'synapseforrdbi' workspace:

Name	Last Modified	Content Type	Size
Capstone2	11/29/2021, 6:05:28 PM	Folder	
capstone3	11/29/2021, 5:49:06 PM	Folder	
Capstone4	11/29/2021, 6:17:52 PM	Folder	
Capstone5	11/29/2021, 6:08:10 PM	Folder	

Below this, the 'Capstone5' folder is expanded, showing a list of CSV files:

Name	Last Modified	Content Type
Branch_region_mapping.csv	11/29/2021, 6:09:57 PM	
Loan_details.csv	11/29/2021, 6:10:30 PM	
Loan_status.csv	11/29/2021, 6:10:51 PM	

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



Narayani.Patil@ravinsofttech.com
RAVINSOFTTECH

Create SQL database

Create database to organize your workload into databases and database objects.

Select SQL pool type *

☒ Serverless ⓘ

☐ Dedicated ⓘ

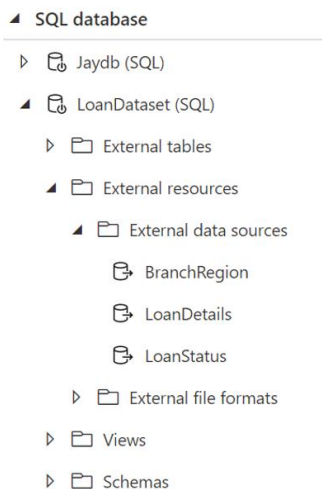
Database *

LoanDataset

Create Cancel

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

1. CREATE EXTERNAL DATA SOURCE BranchRegion
WITH (LOCATION = 'https://synapse2811.blob.core.windows.net/Capstone5')
1. CREATE EXTERNAL DATA SOURCE LoanDetails
WITH (LOCATION = 'https://synapse2811.blob.core.windows.net/Capstone5')
1. CREATE EXTERNAL DATA SOURCE LoanStatus
WITH (LOCATION = 'https://synapse2811.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

- Views
 - dbo.BranchRegion_view
 - dbo.LoanDetails_view
 - dbo.LoanStatus_view
 - System views
 - Schemas
 - Security
 - testsql (SQL)

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

1

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

1 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.

SQL script 2

Run Undo Publish Query plan Connect to Built-in

1 select * from dbo.LoanDetails_view

Results

Messages

View

Table

Chart


Export results

Search

00:00:09 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.

SQL Server database

synapseforrdbi-ondemand.sql.azuresynapse.net

We couldn't authenticate with the credentials provided. Please try again.

Use your Windows credentials to access this database.

☐ Use my current credentials
☒ Use alternate credentials

User name

Password

Back Connect Cancel

Navigator

Display Options

- synapseforrdbi-ondemand.sql.azuresynapse.net
 - Capstone2Db
 - capstone3db
 - capstone4db
 - l1db
 - Jaydb
 - LoanDataset (3)
 - BranchRegion_view
 - LoanDetails_view**
 - LoanStatus_view
 - master_loan_df
 - testsql

LoanDetails_view

	C1	C2	C3	C4	C5	C6	C7
1	36439	65850	56.19	64	14-06-1990	Self emplc	
2	48749	69303	72.15	67	01-01-1991	Salaried	
3	55348	66340	85	2	16-08-1993	Self emplc	
4	48849	64133	77.96	217	01-01-1989	Self emplc	
5	40394	59386	70.72	74	31-12-1974	Self emplc	
6	51803	67466	79.3	162	23-11-1964	Self emplc	
7	61947	109094	58.21	251	01-10-1989	Self emplc	
8	51301	61815	85	67	01-01-1995	Salaried	
9	65882	80461	84.51	255	15-06-1994	Self emplc	
10	34639	69717	50.49	34	23-11-1982	Self emplc	
11	51996	63051	85	147	01-01-1984	Salaried	
12	64269	85632	77.07	146	15-02-1990	Salaried	
13	68377	90512	79.55	20	05-12-1975	Self emplc	
14	61256	73000	84.99	2	11-09-1995	Salaried	
15	46949	64600	74.92	67	29-10-1975	Self emplc	
16	56333	80271	73.5	7	28-02-1984	Self emplc	
17	73717	94315	79.41	165	02-11-1975	Salaried	
18	54373	73891	77.14	152	21-10-1990	Self emplc	
19	72817	82820	89.35	2	11-10-1994	Salaried	
20	78151	107074	74.25	135	05-06-1987	Self emplc	
21	58409	70669	84.9	5	29-05-1996	Salaried	
22	46349	68751	69.82	159	13-12-1993	Self emplc	
23	67760	91363	75.53	161	28-04-1996	Salaried	

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

Visualizations

Fields

Search

- BranchRegion_view
- LoanDetails_view**
- LoanStatus_view

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

The image displays three screenshots of the 'Advanced Editor' window in Azure Synapse Studio, each showing a different SQL query. The window title is 'Advanced Editor' and it includes standard window controls (minimize, maximize, close) and a 'Display Options' dropdown menu.

Branch_region_mapping

```
let
Source = Sql.Databases("synapseforrdbi-ondemand.sql.azuresynapse.net"),
LoanDataset = Source[Name="LoanDataset"]{Data},
dbo_BranchRegion_view = LoanDataset[Schema="dbo",Item="BranchRegion_view"]{Data}
in
dbo_BranchRegion_view
```

✓ No syntax errors have been detected.

Loan_details

```
let
Source = Sql.Databases("synapseforrdbi-ondemand.sql.azuresynapse.net"),
LoanDataset = Source[Name="LoanDataset"]{Data},
dbo_LoanDetails_view = LoanDataset[Schema="dbo",Item="LoanDetails_view"]{Data}
in
dbo_LoanDetails_view
```

✓ No syntax errors have been detected.

Loan_status

```
let
Source = Sql.Databases("synapseforrdbi-ondemand.sql.azuresynapse.net"),
LoanDataset = Source[Name="LoanDataset"]{Data},
dbo_LoanStatus_view = LoanDataset[Schema="dbo",Item="LoanStatus_view"]{Data}
in
dbo_LoanStatus_view
```

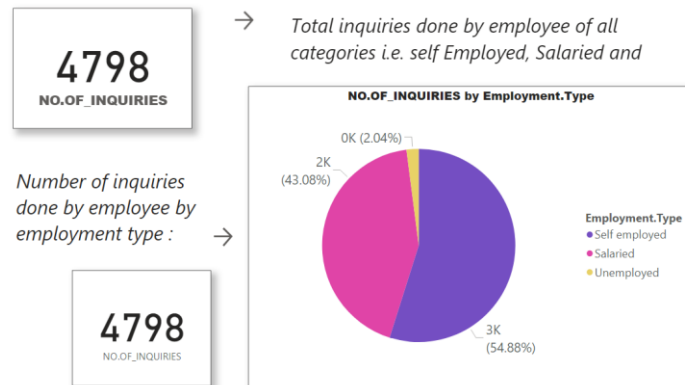
✓ No syntax errors have been detected.

- Perform various analytics on PowerBI

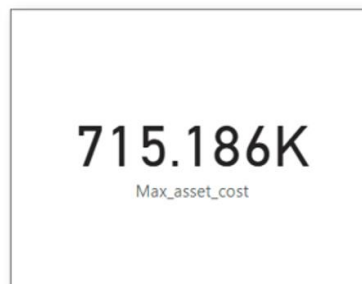
After loading dataset successfully, various analytics performed as given below.

- 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 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

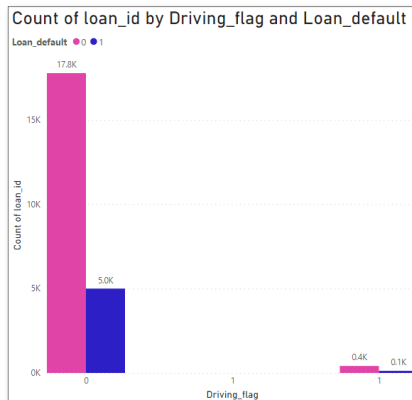
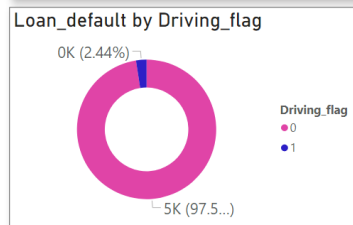
Average of asset_cost

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

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

→

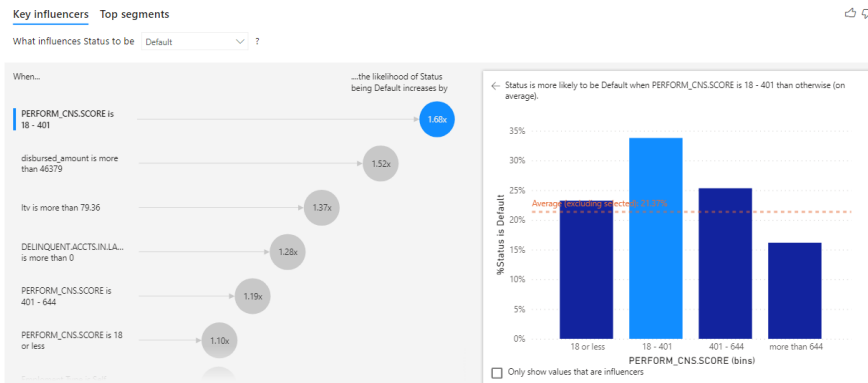
Driving_flag	0	1	Total
0	17774	5001	22775
1	415	125	540
Total	18189	5126	23315



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

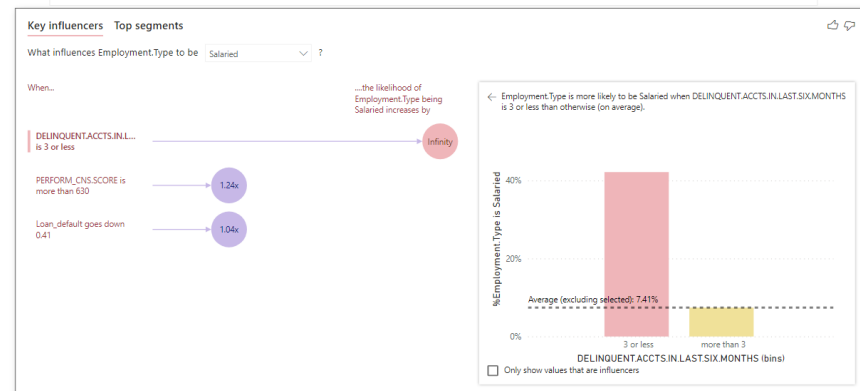
- Display to Key Influencer Visual for the appropriate columns and indicate your inferences.

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



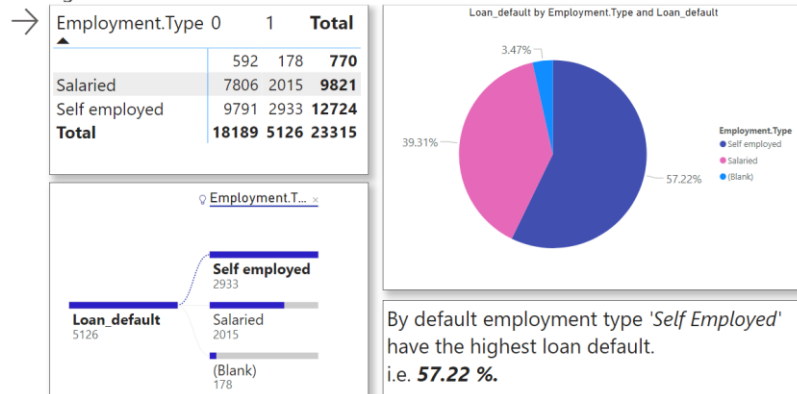
- Display to Key Influencer Visual for the appropriate columns and indicate inferences.

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



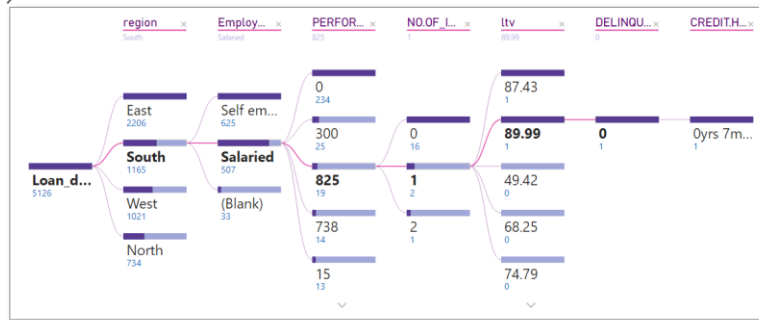
- Display loan default by employment type and indicate which employment type has the highest loan default.

- Display 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.***