

Part 2. Azure Synapse Analytics:

1 Resource Group

- Creating a Resource Group: **rg-dev-bikerental-datalake**

Indicates a development environment (**dev**) for managing bike rental data (**bikerental**) with a focus on analytics and data lake (**dataLake**). Provides clear insights into the resource group's purpose and strategic focus.

Microsoft Azure

Search resources, services, and doc

[Home](#) > [Resource groups](#) >

Create a resource group

Basics

Tags

Review + create

Resource group - A container that holds related resources for an Azure solution. The resource group can include all the resources for the solution, or only those resources that you want to manage as a group. You decide how you want to allocate resources to resource groups based on what makes the most sense for your organization. [Learn more](#)

Project details

Subscription * ⓘ

Azure for Students

Resource group * ⓘ

rg-dev-bikerental-datalake ✓

Resource details

Region * ⓘ

(US) East US

Review + create

< Previous

Next : Tags >

Microsoft Azure

Search resour

[Home](#) > [Resource groups](#) >

Create a resource group

Basics

Tags

Review + create

✓ Validation passed.

Basics

Subscription

Azure for Students

Resource group

rg-dev-bikerental-datalake

Region

East US

Tags

None

Create

< Previous

Next >

[Download a template for automation](#)

2 Create Synapse Analytics Workspace

- Creating a Synapse Analytics Workspace: **synapse-bikerental-datalake-ws**

Emphasizes the Synapse platform (**synapse**) followed by the bike rental project & a focus on a data lake (**bikerental** & **dataLake**) indicating it's workspace by (**ws**).

- Creating a Data Lake Storage Gen2 Account: **bikerentalxdlsgen2xacc**

Indicating the bike rental project (**bikerental**), utilization of Data Lake Storage Gen2 (**dlsgen2**), and separated as an account (**acc**) using 'x' as a separator.

Select Data Lake Storage Gen2 * ⓘ ☒ From subscription ☐ Manually via URL

Account name * ⓘ

File system name *

Create new

Data Lake Storage Gen2 account

Name * ✓

OK Cancel

Assign other users the appropriate [Synapse RBAC roles](#) using Synapse

- Creating a Synapse Analytics Workspace: **bikerental-dlsgen2-fs**

Reflects the bike rental project (**bikerental**), the use of Data Lake Storage Gen2 (**dlsgen2**), and signifies it as a file system (**fs**).

Select Data Lake Storage Gen2 * ⓘ ☒ From subscription ☐ Manually via URL

Account name * ⓘ

File system name *

Create new

Data Lake Storage Gen2 file system

Name * ✓

OK Cancel

Assign other users the appropriate [Synapse RBAC roles](#) using Synapse

Microsoft Azure

Search resources, services, and docs (G+/)

[Home](#) > [Azure Synapse Analytics](#) >

Create Synapse workspace

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all of your resources.

Subscription *

Azure for Students

Resource group *

rg-dev-bikerental-datalake

Create new

Managed resource group

Enter managed resource group name

Workspace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name *

synapse-bikerental-datalake-ws

✓

Region *

East US

✓

Select Data Lake Storage Gen2 *

☒ From subscription

☐ Manually via URL

Account name *

(New) bikerentalxdlsngen2xacc

Create new

File system name *

(New) bikerental-dlsgen2-fs

Create new

☒

Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively query it in the workspace.

We will automatically grant the workspace identity data access to the specified Data Lake Storage Gen2 account, using the [Storage Blob Data Contributor](#) role. To enable other users to use this storage account after you create your workspace, perform these tasks:

Assign other users to the **Contributor** role on workspace

Assign other users the appropriate [Synapse RBAC roles](#) using Synapse Studio

Assign yourself and other users to the **Storage Blob Data**

Review + create

< Previous

Next: Security >

- Setting a Password for the SQL Pools & Confirm it:

Microsoft Azure

Search resources, services, and docs (G+/)

[Home](#) > [Azure Synapse Analytics](#) >

Create Synapse workspace

✓ Validation succeeded

* Basics

* Security

Networking

Tags

Review + create

Product Details

Azure Synapse Analytics workspace
by Microsoft
[Terms of use](#) | [Privacy policy](#)

Serverless SQL est. cost/TB ⓘ
5.00 USD

Terms

By clicking Create, I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. For additional details see [Azure Marketplace Terms](#).

Basics

Subscription	Azure for Students
Resource group	rg-dev-bikerental-datalake
Region	East US
Workspace name	(new) synapse-bikerental-datalake-ws
Data Lake Storage Gen2 account	(new) https://bikerentalxdlsngen2xacc.dfs.core.windows.net
Data Lake Storage Gen2 file system	(new) bikerental-dlsngen2-fs
Managed resource group	None
Role assignments	The Storage Blob Data Contributor role will be assigned on the specified Data Lake Storage Gen2 account to both the workspace managed identity and the current user.

Security

Create

< Previous

Next >

[Download a template for automation](#)

- Review Synapse Resources Creation & their Costs:

• Verify Azure Synapse Analytics Resources Deployment:

Home > Microsoft.Azure.SynapseAnalytics-20231123172705 | Overview

Deployment

Search

Delete Cancel Redeploy Download Refresh

Your deployment is complete

Deployment name : Microsoft.Azure.SynapseAnalytics-20231123172705
Subscription : Azure for Students
Resource group : rg-dev-bikereental-datalake

Start time : 11/23/2023, 6:44:45 PM
Correlation ID : 92fbcd7d-a6b2-4164-b555-8cc93194ba0f

Deployment details

Resource	Type	Status	Operation details
StorageRoleDeploymentResource	Deployment	OK	Operation details
synapse-bikereental-datalake-ws/allowAll	Microsoft.Synapse/workspaces/1	OK	Operation details
synapse-bikereental-datalake-ws	Synapse workspace	OK	Operation details
synapse-bikereental-datalake-ws	Synapse workspace	OK	Operation details
bikereentalxdlsngen2xacc/default/bikereental-dlsngen2-fs	Microsoft.Storage/storageAccou	Created	Operation details
bikereentalxdlsngen2xacc	Storage account	OK	Operation details

Next steps

[Go to resource group](#)

3 Upload the Bike Rental Data to Bike Rental File System:

• Go to storage account: **bikereentalxdlsngen2xacc**

Home > Microsoft.Azure.SynapseAnalytics-20231123172705 | Overview > rg-dev-bikereental-datalake > bikereentalxdlsngen2xacc

Storage account

Search

Upload Open in Explorer Delete Move Refresh Open in mobile CLI / PS Feedback

Essentials

Resource group (move) : rg-dev-bikereental-datalake
Location : eastus
Primary/Secondary Location : Primary: East US, Secondary: West US
Subscription (move) : Azure for Students
Subscription ID : a3d42437-772d-4b14-a05c-8192e7a18547
Disk state : Primary: Available, Secondary: Available

Tags (edit) : Add tags

Properties Monitoring Capabilities (5) Recommendations (0) Tutorials Tools + SDKs

Data Lake Storage

Hierarchical namespace	Enabled
Default access tier	Hot
Blob anonymous access	Disabled
Blob soft delete	Disabled
Container soft delete	Disabled
Versioning	Disabled
Change feed	Disabled
NFS v3	Disabled
SFTP	Disabled

File service

Large file share	Disabled
Identity-based access	Not configured
Default share-level permissions	Disabled
Soft delete	Enabled (7 days)

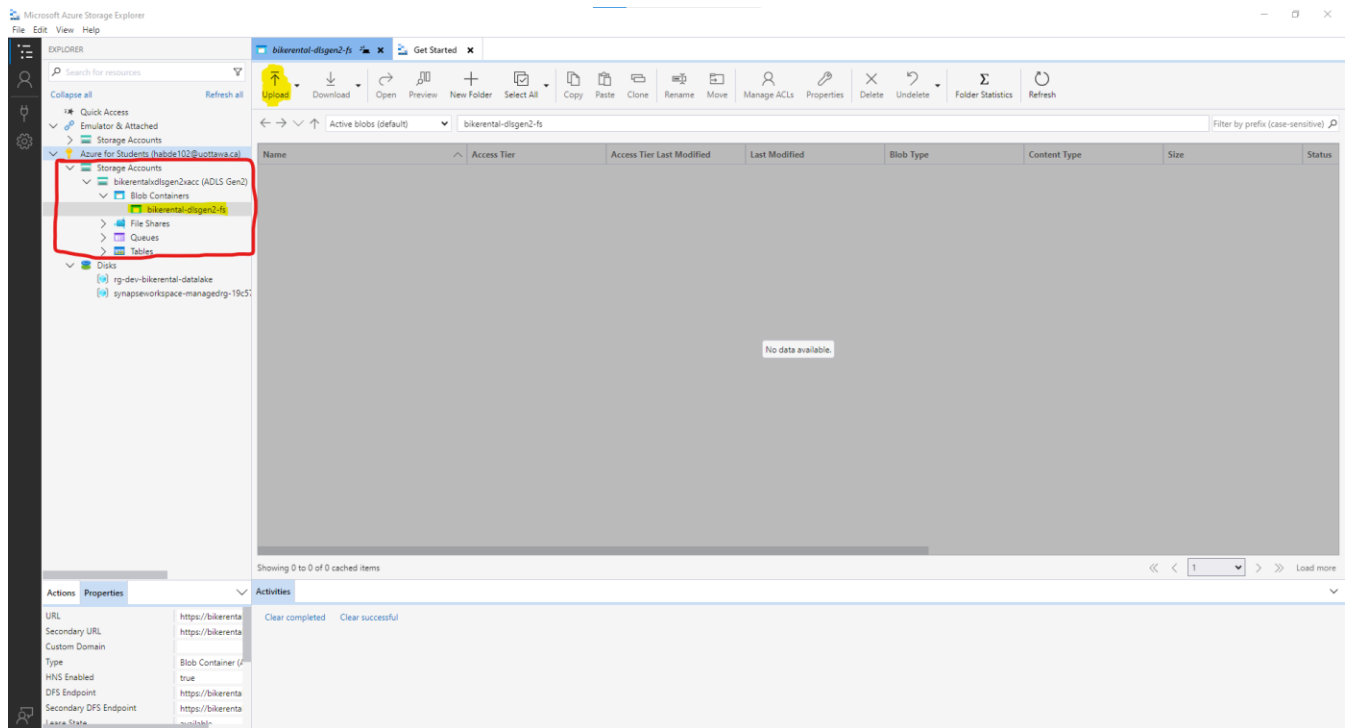
Security

Require secure transfer for REST API operations	Enabled
Storage account key access	Enabled
Minimum TLS version	Version 1.2
Infrastructure encryption	Disabled

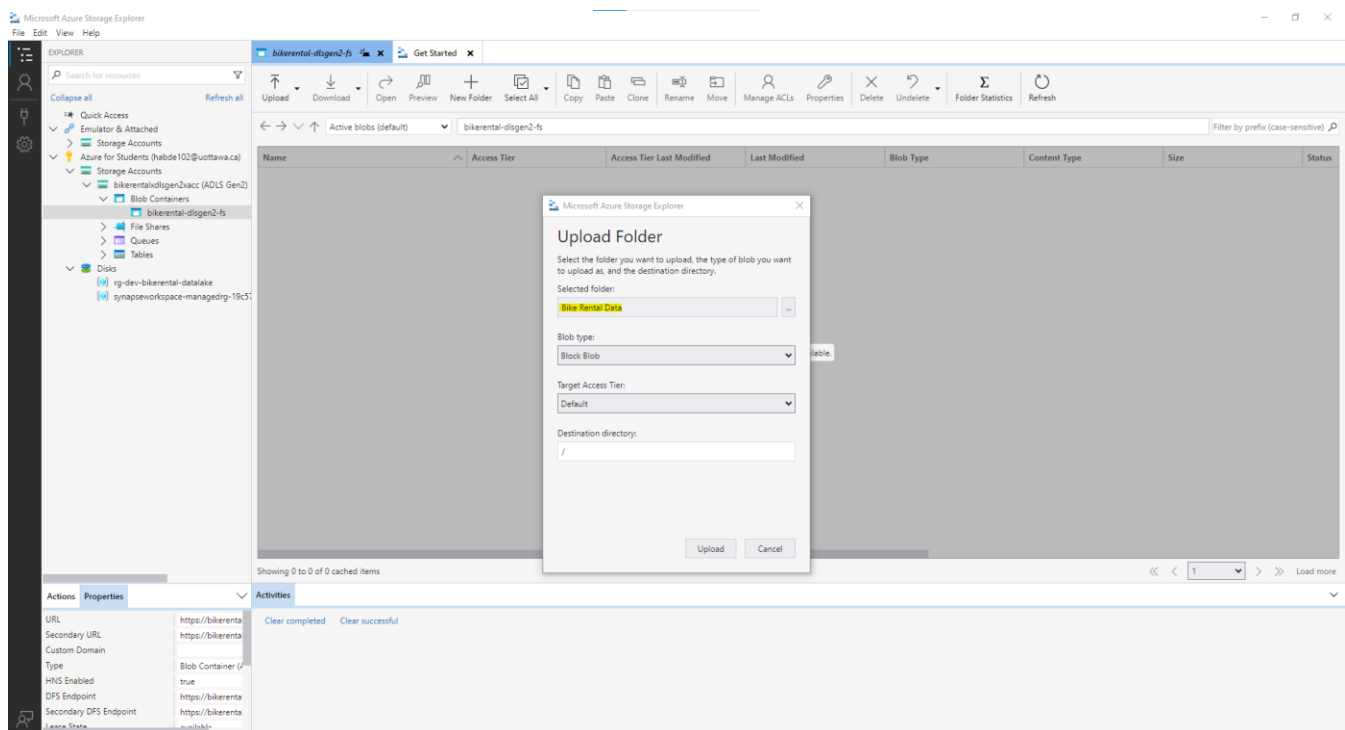
Networking

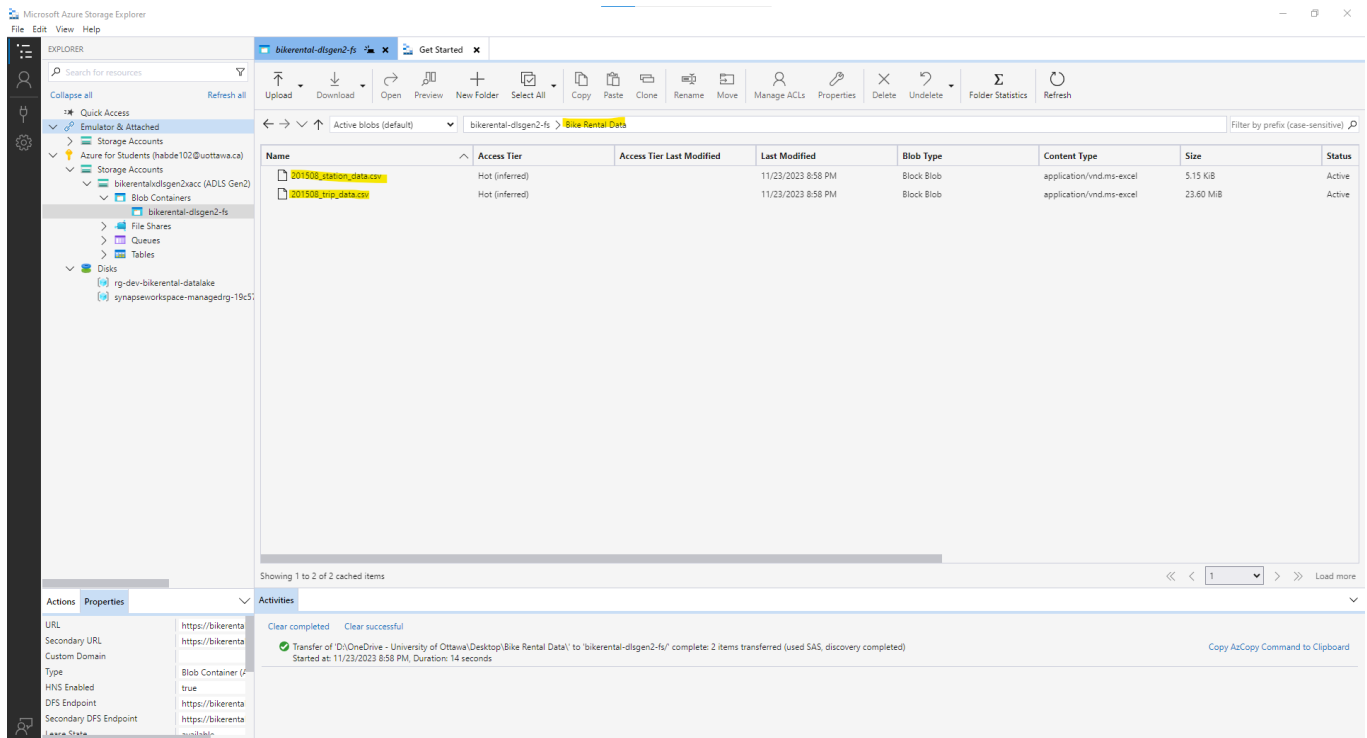
Allow access from	All networks
Number of private endpoint connections	0
Network routing	Microsoft network routing
Access for trusted Microsoft services	Yes
Endpoint type	Standard

- Access the Bike Rental File System (**bikerental-dlsgen2-fs**) through Azure Storage Explorer:



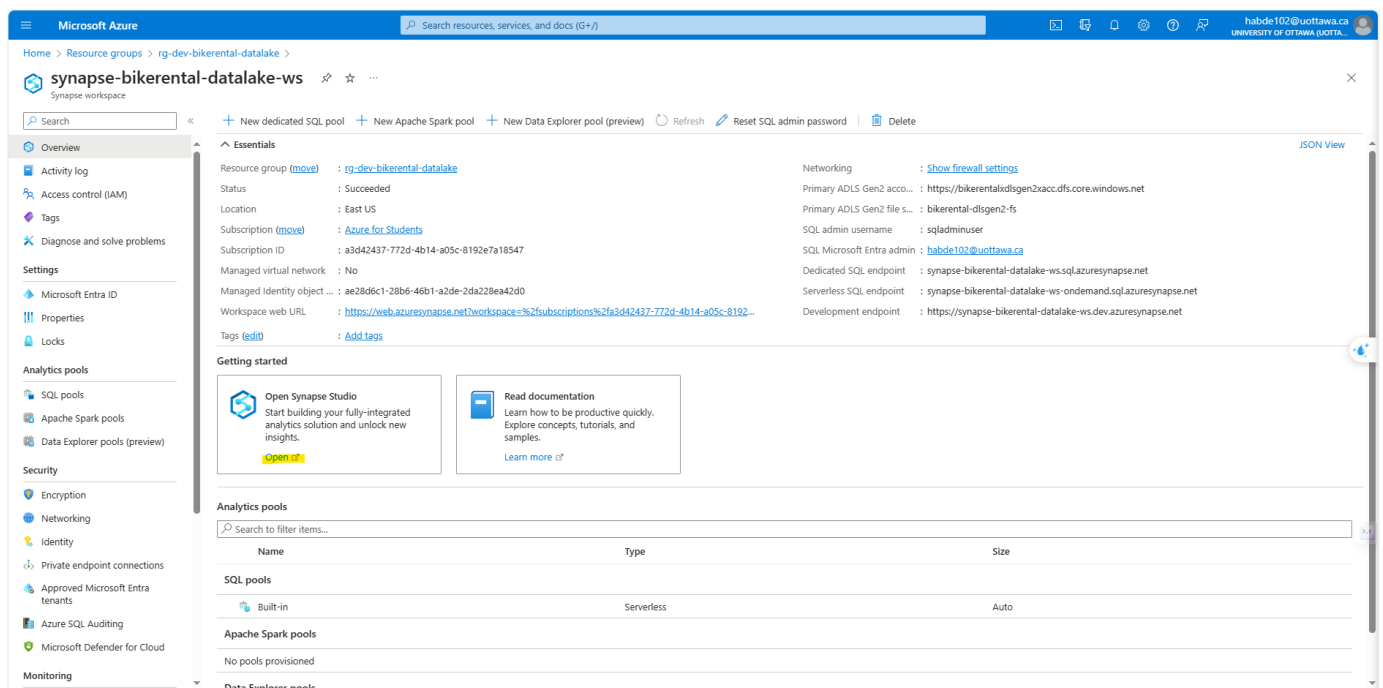
- Upload Bike Rental Data to the Bike Rental File System (**bikerental-dlsgen2-fs**) by Azure Storage Explorer:

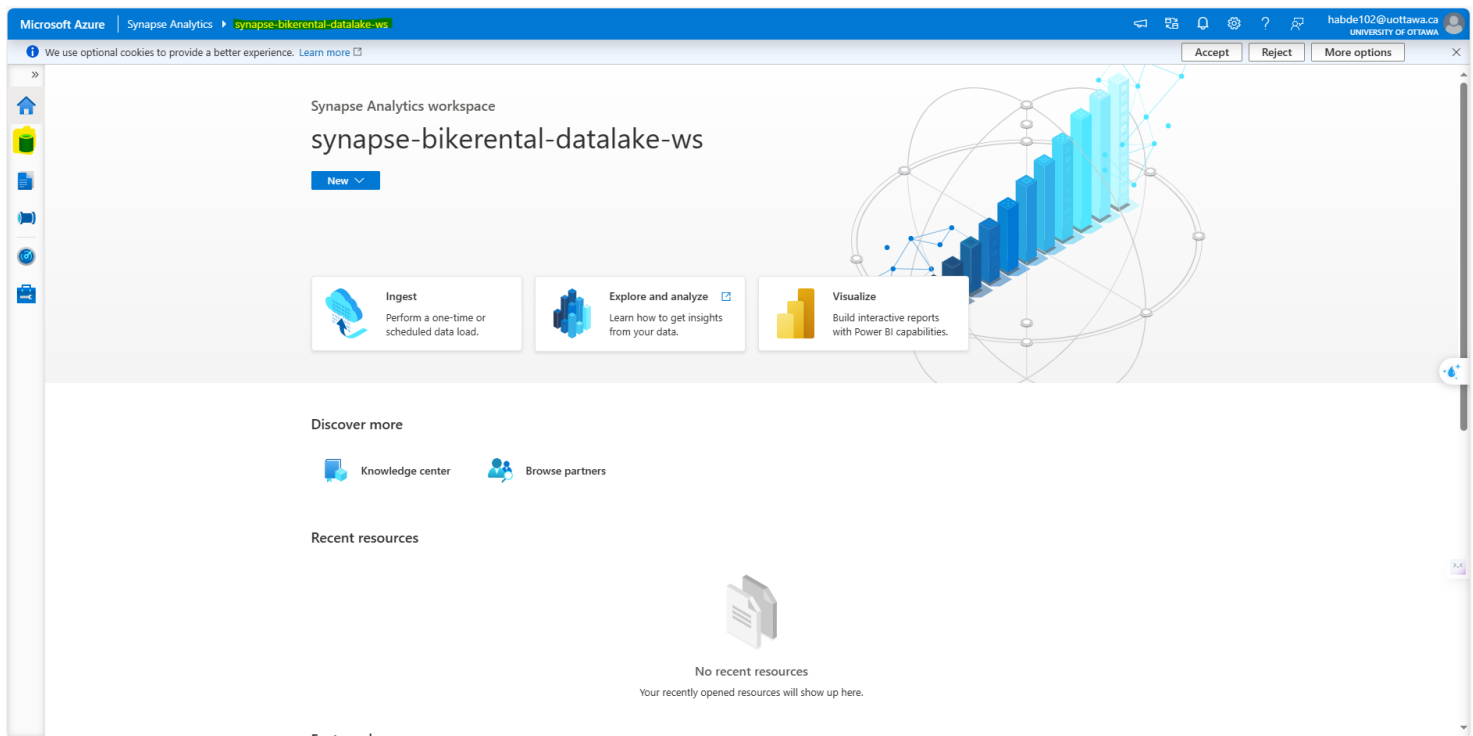




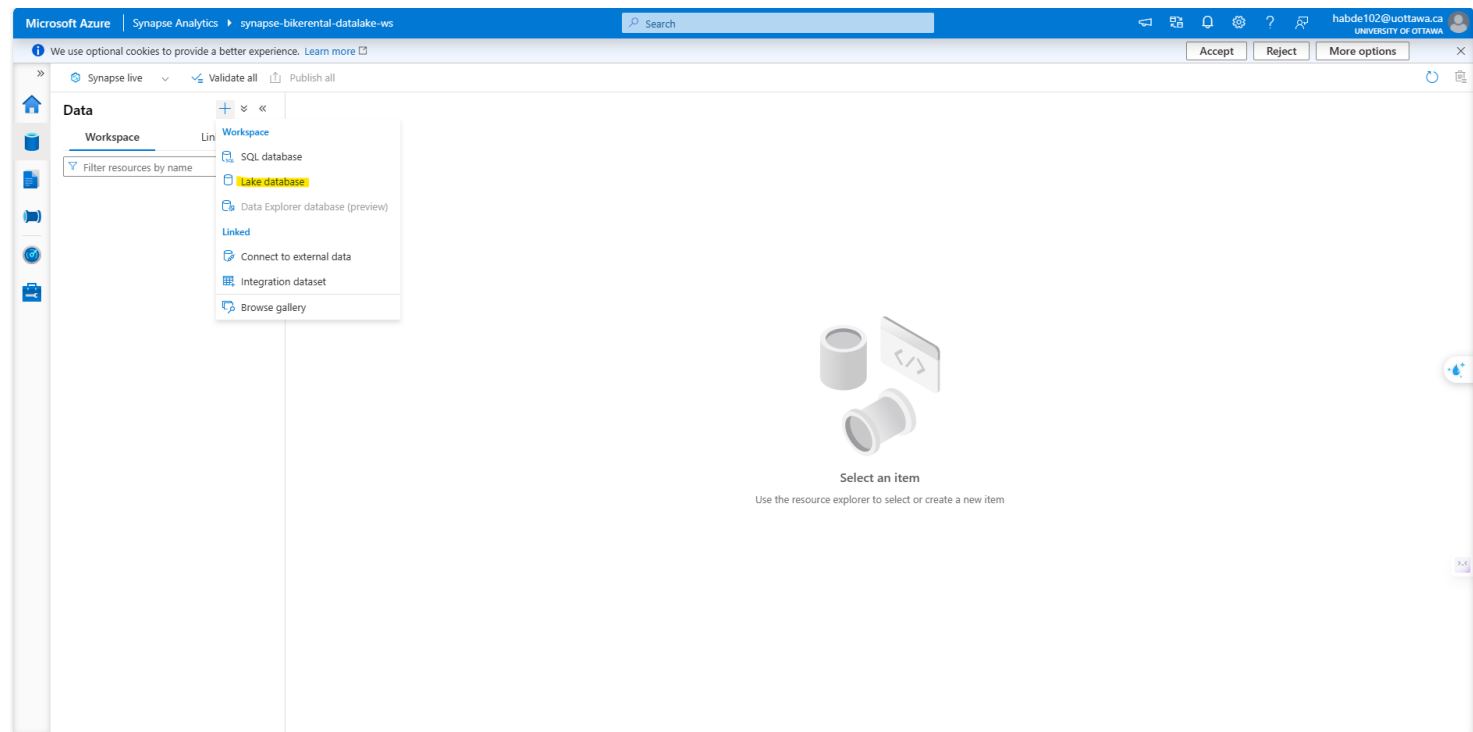
4 Create Lake Database:

- Open **Synapse Studio** of (synapse-bikerental1-datalake-ws) Synapse Analytics Workspace:





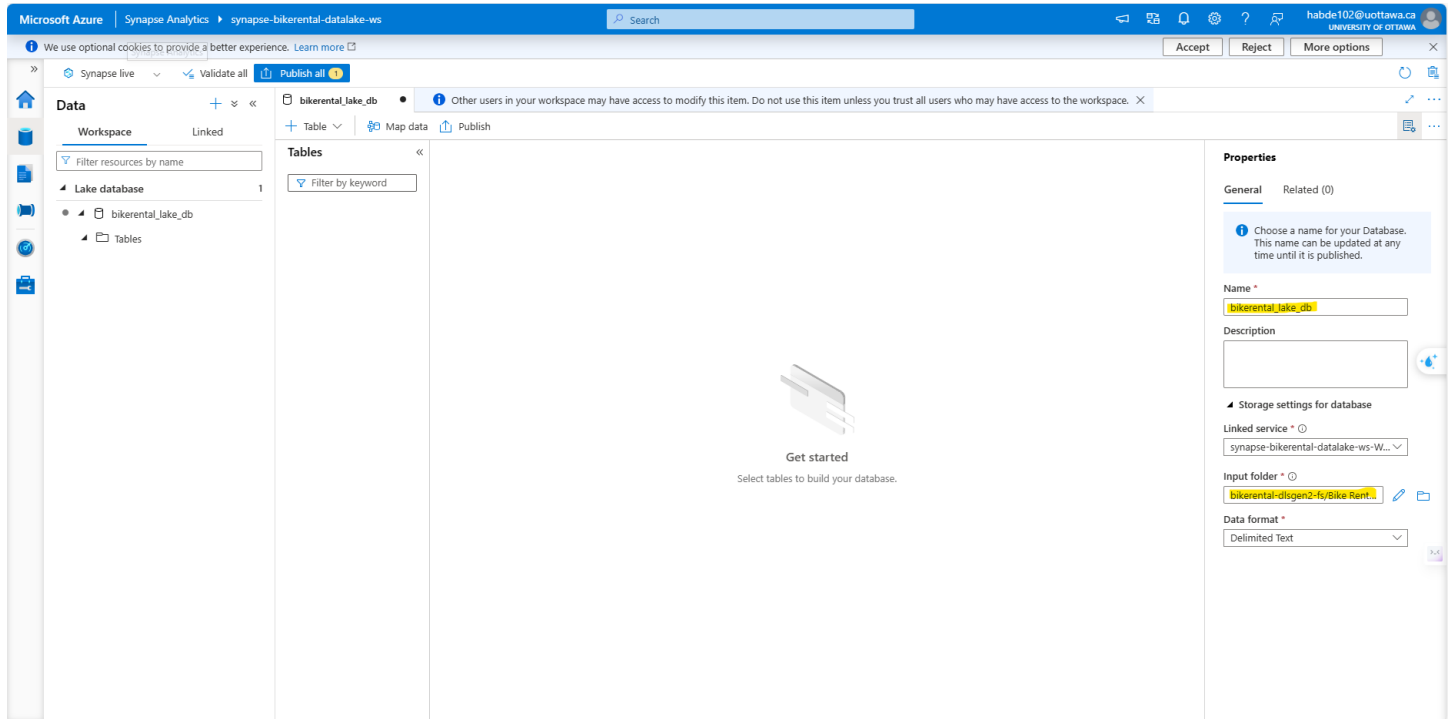
- Add Lake Database:



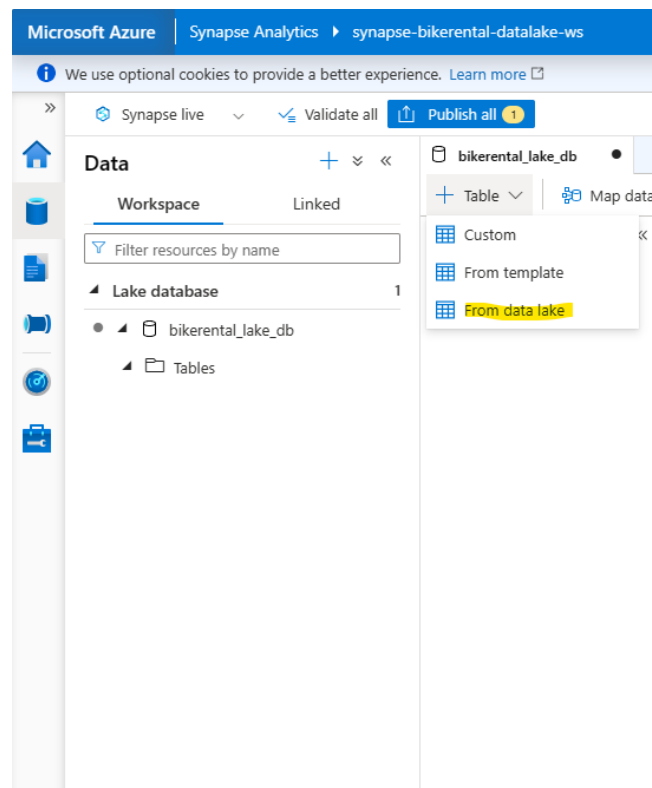
- Set Lake Database Name **bikerental_lake_db**

Connects the (**bikerental**) data with the Data Lake concept, representing it as a database (**db**).

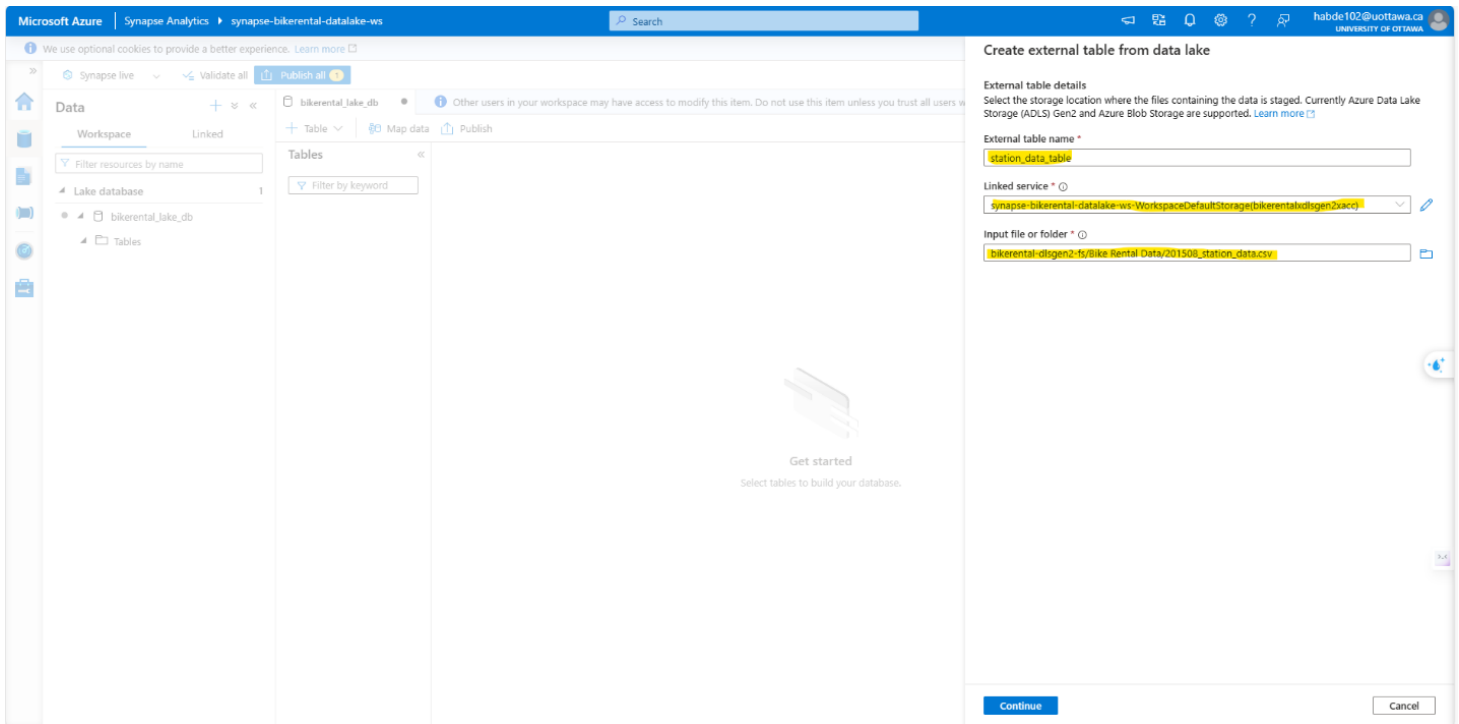
- Set **Input Folder Path** as the path to **Bike Rental Data** folder in **bikerental-dlsgen2-fs** file system.



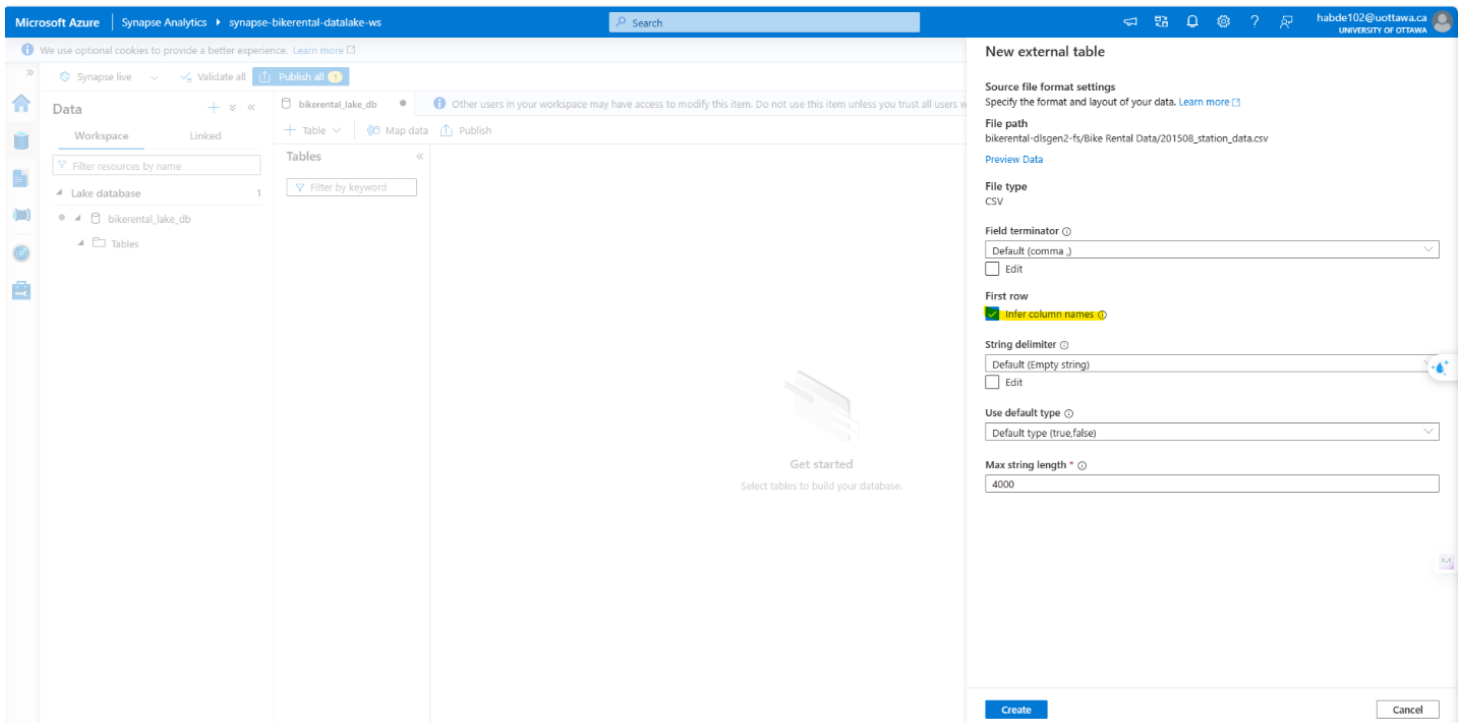
- Add **Data Tables**:



- Add Station Data Table **station_data_table** clearly identifies the table as related to station data within the project & Set it's input service account **bikerentalxdlsngen2xacc** to access **bikerental-dlsngen2-fs** file system so that import **201508_station_data.csv** file data



- Extract Column Names from the File:



- Correct **installation** Column Datatype from string to date:

The screenshot shows the Microsoft Azure Synapse Analytics interface. In the 'Data' section, the 'bikerental_lake_db1' database is selected. The 'Columns' tab is active, displaying a list of columns. The 'installation' column is highlighted in yellow, and its data type is set to 'date' with a format of 'MM/dd/yyyy'. The other columns are: station_id (int long), name (string), lat (double), long (double), dockcount (int long), and landmark (string).

Name	Keys	Description	Nullability	Data type	Format / Length
station_id	PK		Null	int long	
name	PK		Null	string	8000
lat	PK		Null	double	
long	PK		Null	double	
dockcount	PK		Null	int long	
landmark	PK		Null	string	8000
installation	PK		Null	date	MM/dd/yyyy

- Add Trip Data Table **trip_data_table** clearly identifies the table as related to trip data within the project & Set it's input service account **bikerentalxdlsngen2xacc** to access **bikerental-dlsngen2-fs** file system so that import **201508_trip_data.csv** file data

The screenshot shows the Microsoft Azure Synapse Analytics interface. In the 'Data' section, the 'bikerental_lake_db' database is selected. The 'Tables' tab is active, displaying a list of tables. The 'trip_data_table' is highlighted in yellow. The 'Columns' tab is also visible, showing the same columns as in the previous screenshot.

The 'Create external table from data lake' dialog is open, showing the following details:

- External table name:** trip_data_table
- Linked service:** synapse-bikerental-datalake-ws-WorkspaceDefaultStorage(bikerentalxdlsngen2xacc)
- Input file or folder:** bikerental-dlsngen2-fs/Bike Rental Data/201508_trip_data.csv

- Correct Columns Names & Datatypes:

Microsoft Azure | Synapse Analytics | synapse-bikereental-datalake-ws

Search

habde102@uottawa.ca UNIVERSITY OF OTTAWA

We use optional cookies to provide a better experience. [Learn more](#)

Accept Reject More options

Synapse live Validate all Publish all

Data Workspace Linked

Filter resources by name

Lake database 1

bikereental_lake_db

Tables

Filter by keyword

Others 2

General Columns Relationships

Filter by keyword + Column Clone Convert type Delete

Name	Keys	Description	Nullability	Data type	Format / Length
Trip ID	PK		Null	int long	
Duration	PK		Null	int long	
Start Date	PK		Null	abc string	8000
Start Terminal	PK		Null	int long	

Properties

General Related (0)

Choose a name for your Database. This name can be updated at any time until it is published.

Name * bikereental_lake_db

Description

Storage settings for database

Linked service * synapse-bikereental-datalake-ws-W...

Input folder * bikereental-dlsgen2-fs/Bike Rent...

Data format * Delimited Text

Microsoft Azure | Synapse Analytics | synapse-bikereental-datalake-ws

Search

habde102@uottawa.ca UNIVERSITY OF OTTAWA

We use optional cookies to provide a better experience. [Learn more](#)

Accept Reject More options

Synapse live Validate all Publish all

Data Workspace Linked

Filter resources by name

Lake database 2

bikereental_lake_db1

Tables

Filter by keyword

Others 2

General Columns Relationships

Filter by keyword + Column Clone Convert type Delete

Name	Keys	Description	Nullability	Data type	Format / Length
Trip ID	PK		Null	abc string	256
Duration	PK		Null	int long	
Start Date	PK		Null	timestamp	MM/dd/yyyy HH:mm:ss
Start Terminal	PK		Null	int long	
End Date	PK		Null	timestamp	MM/dd/yyyy HH:mm:ss
End Terminal	PK		Null	int long	
Bike	PK		Null	int long	
Subscriber type	PK		Null	abc string	8000
Zip Code	PK		Null	abc string	256

Properties

General Related (0)

Choose a name for your Database. This name can be updated at any time until it is published.

Name * bikereental_lake_db

Description

Storage settings for database

Linked service * synapse-bikereental-datalake-ws-W...

Input folder * bikereental-dlsgen2-fs/Bike Rent...

Data format * Delimited Text

- Publish **bikerental_lake_db** Database Creation it's Data tables (**station_data_table**, **Trip_data_table**) :

The screenshot shows the Microsoft Azure Synapse Analytics interface. On the left, the 'Data' pane shows the 'bikerental_lake_db' database with two tables: 'station_data_table' and 'trip_data_table'. The 'station_data_table' has columns: station_id, name, lat, long, dockcount, landmark, and installation. The 'trip_data_table' has columns: Trip_ID, Duration, Start_Date, Start_Terminal, End_Date, End_Terminal, Bike, Subscriber_Type, and Zip_Code. The 'Publish all' dialog is open on the right, showing a table with 'Pending changes (1)' for the 'bikerental_lake_db' database. The table has columns: NAME, CHANGE, and EXISTING. The row for 'bikerental_lake_db' shows a '(New)' change. At the bottom of the dialog are 'Publish' and 'Cancel' buttons.

5 Apply SQL Quieres on Bike Rental Database:

Create new SQL script for **bikerental_lake_db** database

The screenshot shows the Microsoft Azure Synapse Analytics interface. The 'Data' pane on the left shows the 'bikerental_lake_db' database with two tables: 'station_data_table' and 'trip_data_table'. The 'station_data_table' has columns: station_id (long), name (string), lat (double), long (double), dockcount (long), landmark (string), and installation (date). The 'trip_data_table' has columns: Trip_ID (string), Duration (long), Start_Date (timestamp), Start_Terminal (long), End_Date (timestamp), End_Terminal (long), Bike_Number (long), Subscriber_Type (string), and Zip_Code (string). A context menu is open over the 'bikerental_lake_db' database, showing options: 'New SQL script', 'New notebook', 'Open', 'Clone', and 'Delete'. The 'New SQL script' option is highlighted, and a sub-menu is open showing 'Empty script'.

a) Show Top 20 zip codes for bike up.

- Synapse Query Script:

```
SELECT TOP (20)
    [Zip_code],
    COUNT([Trip_ID]) AS [BikeUpsPickupCount]
FROM [bikerental_lake_db].[dbo].[trip_data_table]
GROUP BY [Zip_code]
ORDER BY [BikeUpsPickupCount] DESC;
```

The screenshot shows the Synapse Query Editor interface. At the top, there's a blue header bar with the text "bikerental-datalake-ws" and a search bar. Below the header, there's a "Publish all" button. The main area displays a tabbed interface with "Question A" through "Question E". The "Question D" tab is active, showing the following SQL query:

```
1 SELECT TOP (20)
2     [Zip_code],
3     COUNT([Trip_ID]) AS [BikeUpsPickupCount]
4 FROM [bikerental_lake_db].[dbo].[trip_data_table]
5 GROUP BY [Zip_code]
6 ORDER BY [BikeUpsPickupCount] DESC;
```

Below the query, there are buttons for "Run", "Undo", "Publish", and "Query plan". The "Connect to" dropdown is set to "Built-in", and the "Use database" dropdown is set to "bikerental_lake_db".

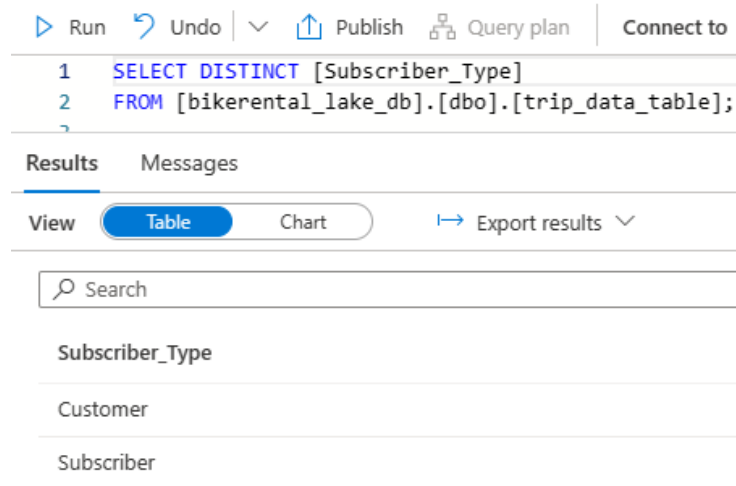
- Run Results:

The screenshot shows the Synapse Query Editor interface with the "Results" tab selected. The results are displayed in a table view with two columns: "Zip_code" and "BikeUpsPickupCount". The table contains 20 rows of data, sorted by "BikeUpsPickupCount" in descending order. The status bar at the bottom indicates "00:00:34 Query executed successfully."

Zip_code	BikeUpsPickupCount
94107	46622
94105	20311
94133	16246
94103	14959
94111	10960
94102	10150
nil	6649
94109	6413
95112	4829
94403	4199
94611	4088
94158	4071
94117	4070
94501	4034
94602	3850
94110	3703
94114	3575
95110	3528

- b) Show Monthly duration aggregate across the rental subscriber types, ordered in descending order of the busiest months (use a meaningful measure for the aggregate)

Since there's only two types of customers in the data as shown in the image below:



We will calculate the total duration of trips for each type of customers (Subscriber & Customer) within each month & order it by busiest month assuming that it's the one with largest number of trips made.

- Synapse Query Script:

```
SELECT
    FORMAT([Start_Date], 'MM/yyyy') AS [MonthYear],
    COUNT([Trip_ID]) AS [TripsCount],
    -- Add up all trips durations in each month to get a months trip total duration
    SUM([Duration]) AS [TripsTotalDuration],
    -- Since there's only two unique types of subscribers(Subscriber, Customer), we will add up durations of
    -- each of them across months
    -- to divide this durations over those types & define there portions from the whole total duration
    -- Subscriber Type Total Duration
    SUM(CASE WHEN [Subscriber_Type] = 'Subscriber' THEN [Duration] ELSE 0 END) AS [SubscriberTotalDuration],
    -- Customer Type Total Duration
    SUM(CASE WHEN [Subscriber_Type] = 'Customer' THEN [Duration] ELSE 0 END) AS [CustomerTotalDuration]
FROM [bikerental_lake_db].[dbo].[trip_data_table]
GROUP BY FORMAT([Start_Date], 'MM/yyyy')
-- Assume That the Busiest Month is the one with largest number of trips
-- order by busiest month then the month order itself incase multiple months were the as busy as others
ORDER BY [TripsCount] DESC, [MonthYear];
```


Question AQuestion BQuestion CQuestion DQuestion E

RunUndoPublishQuery planConnect toBuilt-inUse databasebikerental_lake_db

```
1 SELECT
2   FORMAT([Start_Date], 'MM/yyyy') AS [MonthYear],
3   COUNT([Trip_ID]) AS [TripsCount],
4   -- Add up all trips durations in each month to get a months trip total duration
5   SUM([Duration]) AS [TripsTotalDuration],
6   -- Since there's only two unique types of subscribers(Subscriber, Customer), we will add up durations of each of them across months
7   -- to divide this durations over those types & define there portions from the whole total duration
8   -- Subscriber Type Total Duration
9   SUM(CASE WHEN [Subscriber_Type] = 'Subscriber' THEN [Duration] ELSE 0 END) AS [SubscriberTotalDuration],
10  -- Customer Type Total Duration
11  SUM(CASE WHEN [Subscriber_Type] = 'Customer' THEN [Duration] ELSE 0 END) AS [CustomerTotalDuration]
12  FROM [bikerental_lake_db].[dbo].[trip_data_table]
13  GROUP BY FORMAT([Start_Date], 'MM/yyyy')
14  -- Assume That the Busiest Month is the one with largest number of trips
15  -- order by busiest month then the month order itself incase multiple months were the as busy as others
16  ORDER BY [TripsCount] DESC, [MonthYear];
17
```

• Run Results:

Question AQuestion BQuestion CQuestion DQuestion E

RunUndoPublishQuery planConnect toBuilt-inUse databasebikerental_lake_db

ResultsMessages

ViewTableChartExport results

Search

MonthYear	TripsCount	TripsTotalDuration	SubscriberTotalDuration	CustomerTotalDuration
10/2014	34220	33401099	17707291	15693808
07/2015	32476	33983062	15825478	18157584
06/2015	31907	34481927	16592087	17889840
08/2015	31904	31088866	15696847	15392019
09/2014	31682	33160021	16179559	16980462
03/2015	31626	29892301	16297208	13595093
04/2015	31363	28031940	16255508	11776432
05/2015	29540	31584633	16559867	15024766
01/2015	27840	25611358	14533223	11078135
02/2015	26401	25633016	13338619	12294397
11/2014	25516	22454934	12358940	10095994
12/2014	19677	41131402	13947859	27183543

00:00:30 Query executed successfully.

c) Show the top 5 busiest terminals for bike pickup.

- Synapse Query Script:

```
SELECT TOP (5)
    t.[Start_Terminal],
    s.[name] AS [StationName],
    -- Assume That the Busiest Terminal is the one with largest number of trips
    COUNT(t.[Trip_ID]) AS [PickupCount]
FROM [bikerental_lake_db].[dbo].[trip_data_table] t
    -- Join of station data Table done on start terminal as we are concerned about bike pickups
JOIN [bikerental_lake_db].[dbo].[station_data_table] s ON t.[Start_Terminal] = s.[station_id]
GROUP BY t.[Start_Terminal], s.[name]
ORDER BY [PickupCount] DESC; -- sort descending as we target terminals with most pickups
```

The screenshot shows the Synapse Query Editor interface. At the top, there are tabs for 'Question A', 'Question B', 'Question C', 'Question D', and 'Question E'. Below the tabs is a toolbar with buttons for 'Run', 'Undo', 'Publish', 'Query plan', 'Connect to', 'Use database', and a refresh icon. The 'Connect to' dropdown is set to 'Built-in' and 'Use database' is set to 'bikerental_lake_db'. The main area displays the SQL script from the previous block, with line numbers 1 through 11 on the left margin.

- Run Results:

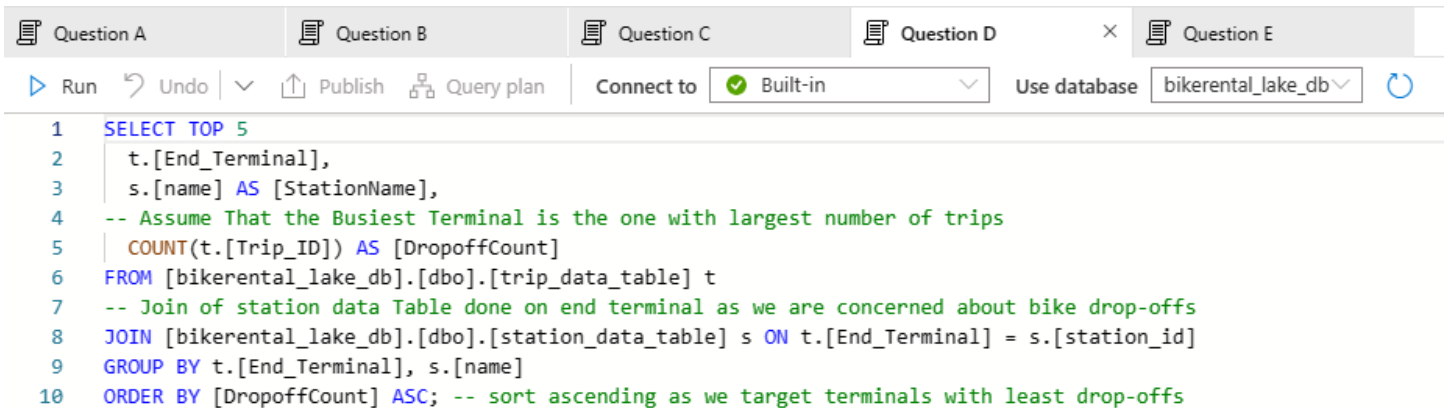
The screenshot shows the Synapse Query Editor interface with the 'Results' tab selected. The 'View' dropdown is set to 'Table'. The results are displayed in a table with three columns: 'Start_Terminal', 'StationName', and 'PickupCount'. The table contains five rows of data, representing the top 5 busiest terminals. A search bar is located at the top of the results area.

Start_Terminal	StationName	PickupCount
70	San Francisco Caltrain (Townsend...	26304
69	San Francisco Caltrain 2 (330 Tow...	21758
50	Harry Bridges Plaza (Ferry Buildin...	17255
55	Temporary Transbay Terminal (H...	14436
60	Embarcadero at Sansome	14158

d) Show the 5 terminals that has the least drop-offs.

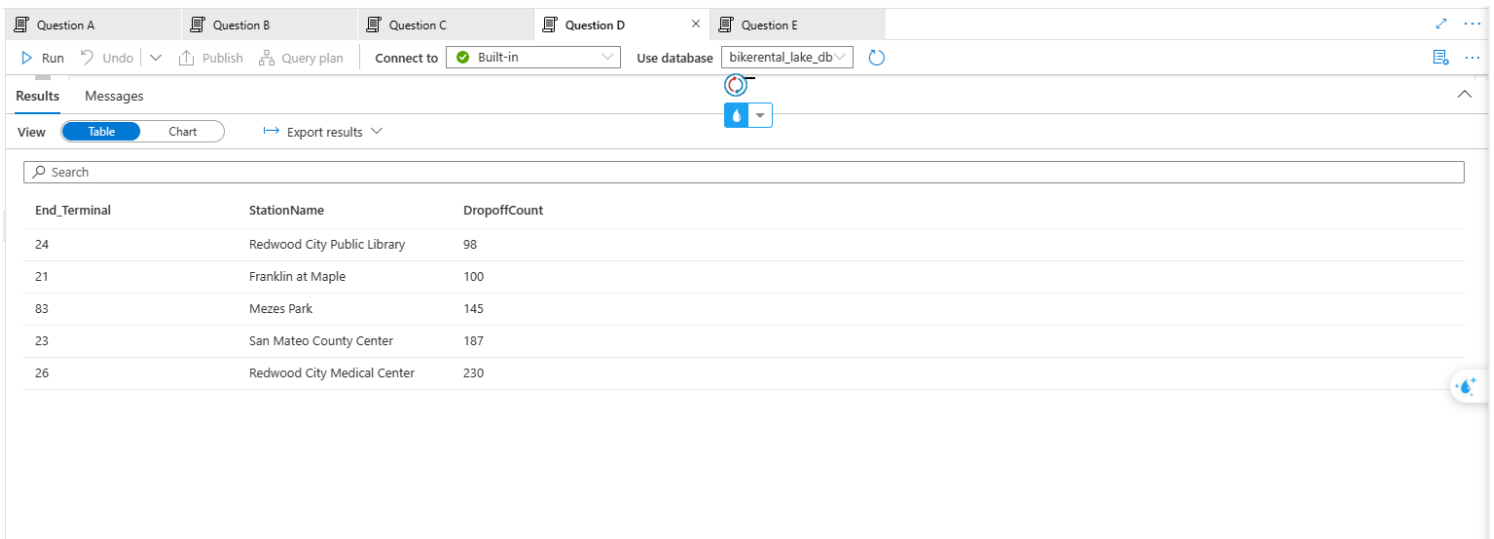
- Synapse Query Script:

```
SELECT TOP (5)
    t.[End_Terminal],
    s.[name] AS [StationName],
    -- Assume That the Busiest Terminal is the one with largest number of trips
    COUNT(t.[Trip_ID]) AS [DropoffCount]
FROM [bikerental_lake_db].[dbo].[trip_data_table] t
-- Join of station data Table done on end terminal as we are concerned about bike drop-offs
JOIN [bikerental_lake_db].[dbo].[station_data_table] s ON t.[End_Terminal] = s.[station_id]
GROUP BY t.[End_Terminal], s.[name]
ORDER BY [DropoffCount] ASC; -- sort ascending as we target terminals with least drop-offs
```



The screenshot shows the Synapse Query Editor interface. At the top, there are tabs for 'Question A', 'Question B', 'Question C', 'Question D', and 'Question E'. Below the tabs is a toolbar with buttons for 'Run', 'Undo', 'Publish', 'Query plan', 'Connect to' (set to 'Built-in'), 'Use database' (set to 'bikerental_lake_db'), and a refresh icon. The main area displays the SQL query script for 'Question E', which is the same script as shown in the previous block. The query is numbered from 1 to 10.

- Run Results:



The screenshot shows the Synapse Query Editor interface with the 'Results' tab selected. The 'Table' view is active, displaying the results of the query. The table has three columns: 'End_Terminal', 'StationName', and 'DropoffCount'. The results are sorted by 'DropoffCount' in ascending order, showing the 5 terminals with the least drop-offs.

End_Terminal	StationName	DropoffCount
24	Redwood City Public Library	98
21	Franklin at Maple	100
83	Mezes Park	145
23	San Mateo County Center	187
26	Redwood City Medical Center	230

e) Produce the monthly summary of bike rentals (format - month/year ex. 06/2020).

In this monthly summary we will try to include the aggregations for all columns in trip data table that their aggregations are applicable, logical & gives true insight:

- 1- Number of Trips Monthly
- 2- Sum of Monthly Trips Durations
- 3- Number of Subscribers & Customers that Made Trips Monthly
- 4- Sum of Subscribers & Customers Monthly Trips Durations
- 5- Sum of Numbers of Bikes Reserved Monthly

- **Synapse Query Script:**

```
SELECT
    FORMAT([Start_Date], 'MM/yyyy') AS [MonthYear],
    COUNT([Trip_ID]) AS [TripsCount],
    -- Add up all trips durations in each month to get a months trip total duration
    SUM([Duration]) AS [TripsTotalDuration],
    -- Count number of subscribers & customers across different trips through each month
    -- As there's only two unique types of subscribers(Subscriber, Customer)
    -- Subscriber Type Number of Trips
    COUNT(CASE WHEN [Subscriber_Type] = 'Subscriber' THEN [Trip_ID] END) AS [SubscriberTripsCount],
    -- Customer Type Number of Trips
    COUNT(CASE WHEN [Subscriber_Type] = 'Customer' THEN [Trip_ID] END) AS [CustomerTripsCount],
    -- Since there's only two unique types of subscribers(Subscriber, Customer), we will add up durations of
    each of them across months
    -- to divide this durations over those types & define their portions from the whole total duration
    -- Subscriber Type Total Duration
    SUM(CASE WHEN [Subscriber_Type] = 'Subscriber' THEN [Duration] ELSE 0 END) AS [SubscriberTotalDuration],
    -- Customer Type Total Duration
    SUM(CASE WHEN [Subscriber_Type] = 'Customer' THEN [Duration] ELSE 0 END) AS [CustomerTotalDuration],
    -- Total Number of Bikes Reserved Within each month
    SUM([Bike_Number]) AS [TotalBikesNumber]
FROM [bikerental_lake_db].[dbo].[trip_data_table]
GROUP BY FORMAT([Start_Date], 'MM/yyyy')
-- order by months order
ORDER BY [MonthYear] ASC;
```

```
Question A Question B Question C Question D Question E
Run Undo Publish Query plan Connect to Built-in Use database bikerental_lake_db
1 SELECT
2   FORMAT([Start_Date], 'MM/yyyy') AS [MonthYear],
3   COUNT([Trip_ID]) AS [TripsCount],
4   -- Add up all trips durations in each month to get a months trip total duration
5   SUM([Duration]) AS [TripsTotalDuration],
6   -- Count number of subscribers & customers across different trips through each month
7   -- As there's only two unqiue types of subscribers(Subscriber, Customer)
8   -- Subscriber Type Number of Trips
9   COUNT(CASE WHEN [Subscriber_Type] = 'Subscriber' THEN [Trip_ID] END) AS [SubscriberTripsCount],
10  -- Customer Type Number of Trips
11  COUNT(CASE WHEN [Subscriber_Type] = 'Customer' THEN [Trip_ID] END) AS [CustomerTripsCount],
12  -- Since there's only two unqiue types of subscribers(Subscriber, Customer), we will add up durations of each of them across months
13  -- to divide this durations over those types & define there portions from the whole total duration
14  -- Subscriber Type Total Duration
15  SUM(CASE WHEN [Subscriber_Type] = 'Subscriber' THEN [Duration] ELSE 0 END) AS [SubscriberTotalDuration],
16  -- Customer Type Total Duration
17  SUM(CASE WHEN [Subscriber_Type] = 'Customer' THEN [Duration] ELSE 0 END) AS [CustomerTotalDuration],
18  -- Total Number of Bikes Reserved Within each month
19  SUM([Bike_Number]) AS [TotalBikesNumber]
20 FROM [bikerental_lake_db].[dbo].[trip_data_table]
21 GROUP BY FORMAT([Start_Date], 'MM/yyyy')
22 -- order by months order
23 ORDER BY [MonthYear] ASC;
24
```

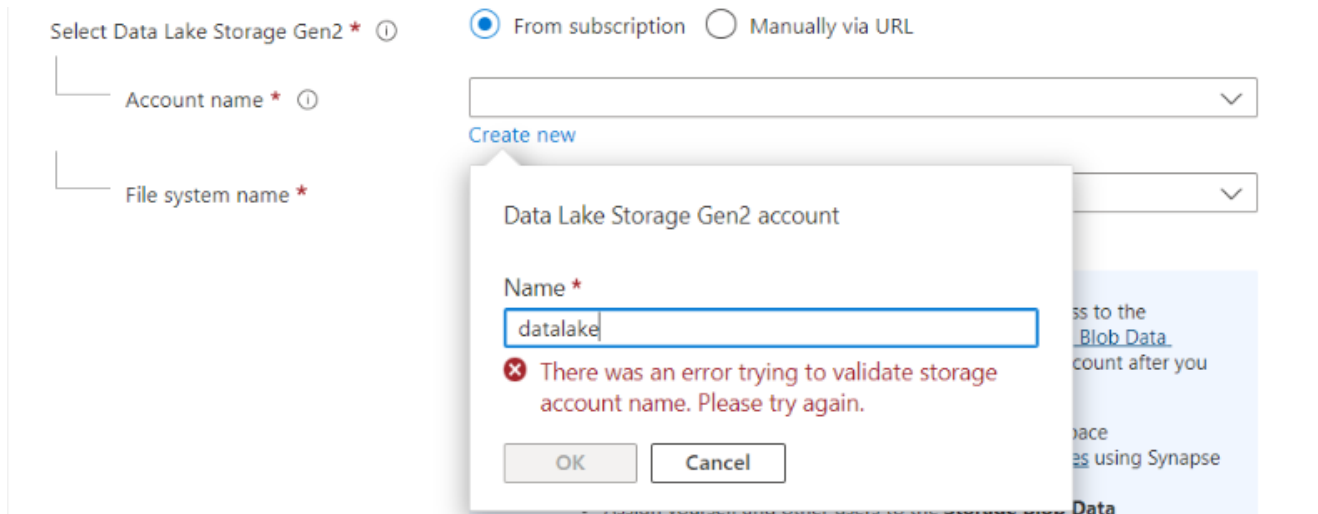
• Run Results:

MonthYear	TripsCount	TripsTotalDuration	SubscriberTripsCount	CustomerTripsCount	SubscriberTotalDu...	CustomerTotal...	TotalBikesNum...
01/2015	27840	25611358	25068	2772	14533223	11078135	11756577
02/2015	26401	25633016	23688	2713	13338619	12294397	11139951
03/2015	31626	29892301	27752	3874	16297208	13595093	13444028
04/2015	31363	28031940	28038	3325	16255508	11776432	13313721
05/2015	29540	31584633	25545	3995	16559867	15024766	12574550
06/2015	31907	34481927	27868	4039	16592087	17889840	13447867
07/2015	32476	33983062	27652	4824	15825478	18157584	13718375
08/2015	31904	31088866	27308	4596	15696847	15392019	13564645
09/2014	31682	33160021	27456	4226	16179559	16980462	13310244
10/2014	34220	33401099	29833	4387	17707291	15693808	14386280
11/2014	25516	22454934	22623	2893	12358940	10095994	10752350
12/2014	19677	41131402	17386	2291	13947859	27183543	8382802

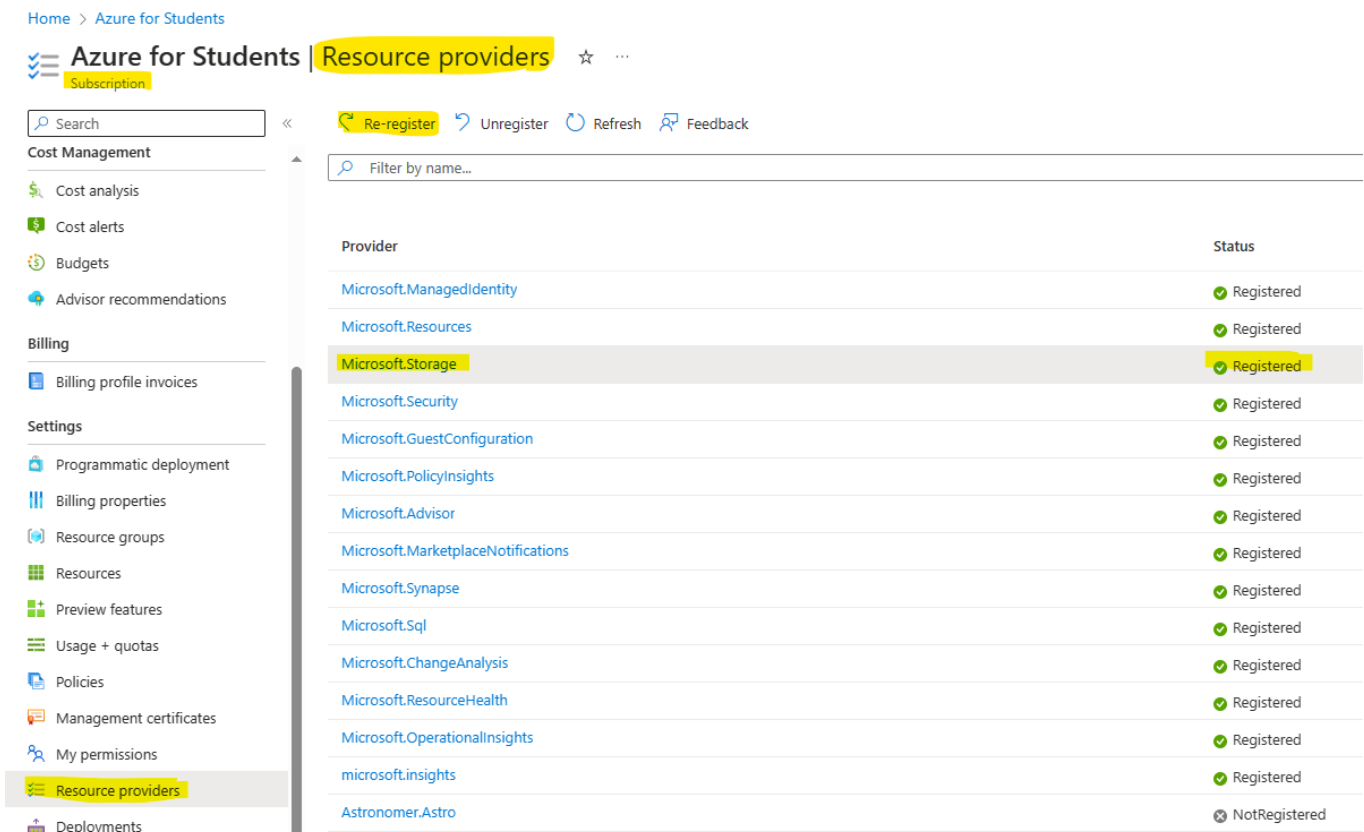
00:00:27 Query executed successfully.

6 Challenges On setting Synapse Analytics Environment:

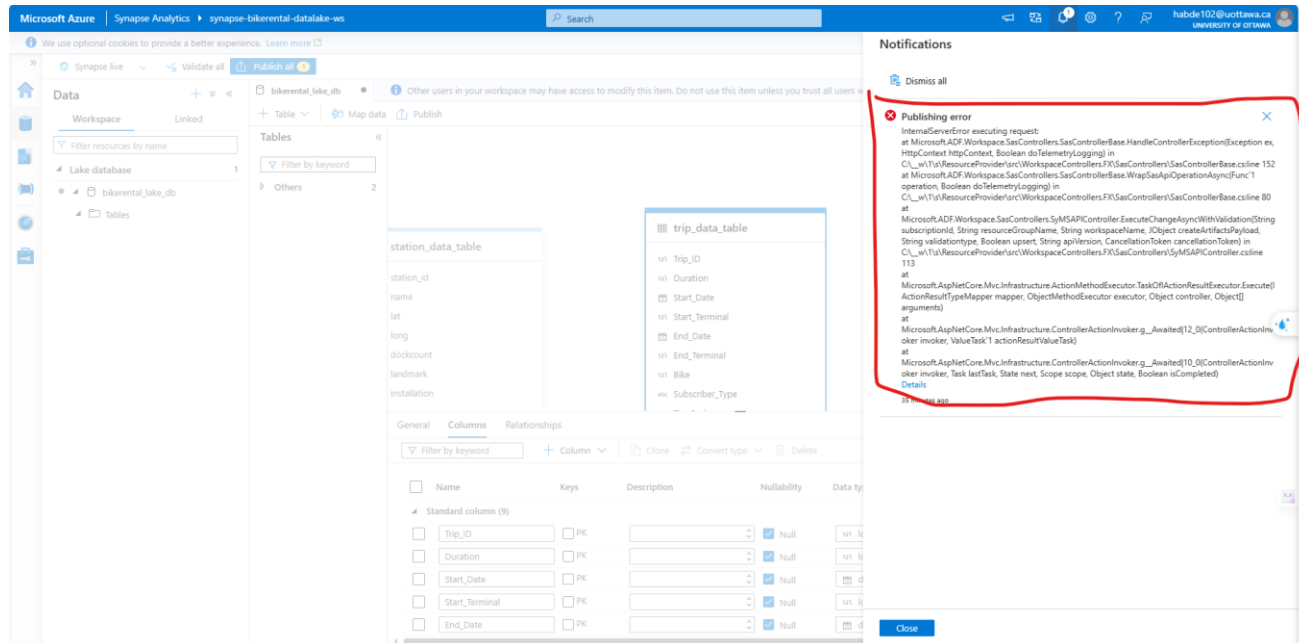
1. Couldn't create a data lake storage account Gen2 when creating an azure synapse resource



We have searched for the cause of this problem & we have found a solution for on the official Microsoft Q&A Section [1]. This problem required that we register Microsoft storage resource in our subscription through the resource provider.



2. Couldn't Publish synapse analytics lake databases changes of bikerental_lake_db Database Creation it's Data tables (station_data_table, Trip_data_table)



After searching for this problem we couldn't find a straight solution for it but after trying couple of times to publish changes it worked.

