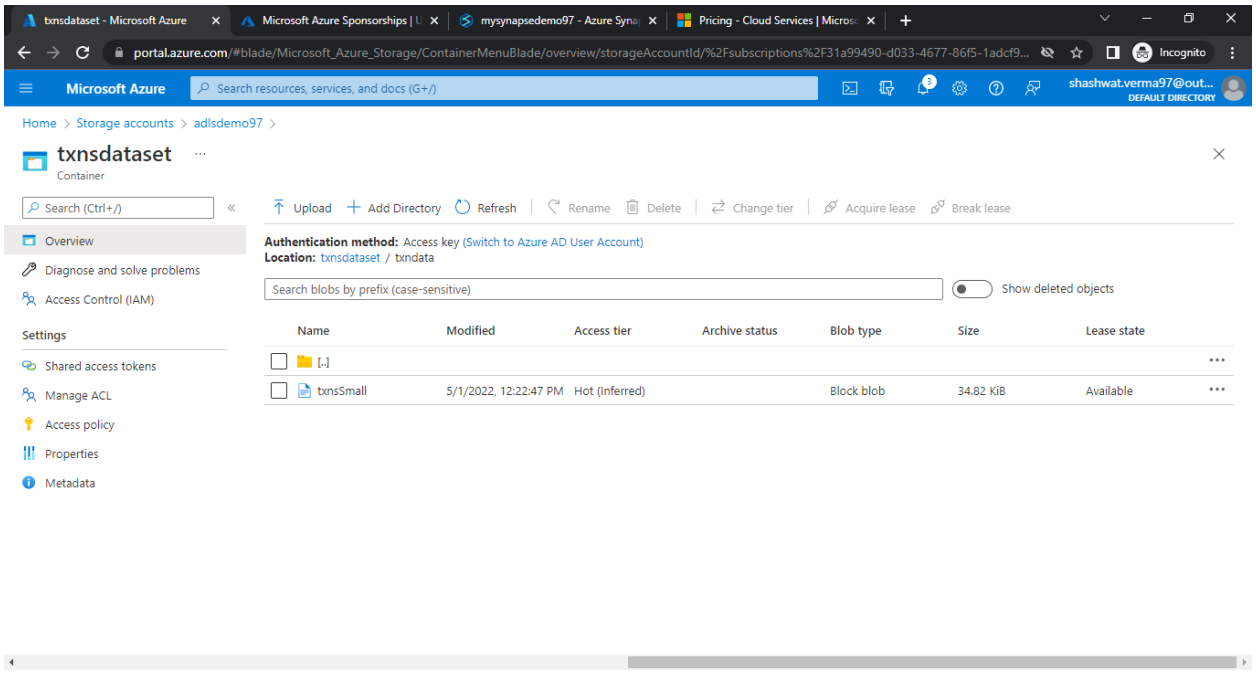
Assignment 1: Azure Synapse Analytics

In the 'adlsdemo97' storage account we created a container named 'txnsdataset'. Inside txns data set we created a directory 'txndata' where we uploaded 'txnsSmall' file as shown in screenshot below:



Source Code:

```
-- First we need to create a database in the serverless pool
CREATE DATABASE [walmart]

-- Ensure to switch the context to the new database first
USE [walmart]

-- Here we are creating a database master key. This key will be
used to protect the Shared Access Signature which is specified
in the next step
CREATE MASTER KEY ENCRYPTION BY PASSWORD = 'P@ssw0rd@123';
```

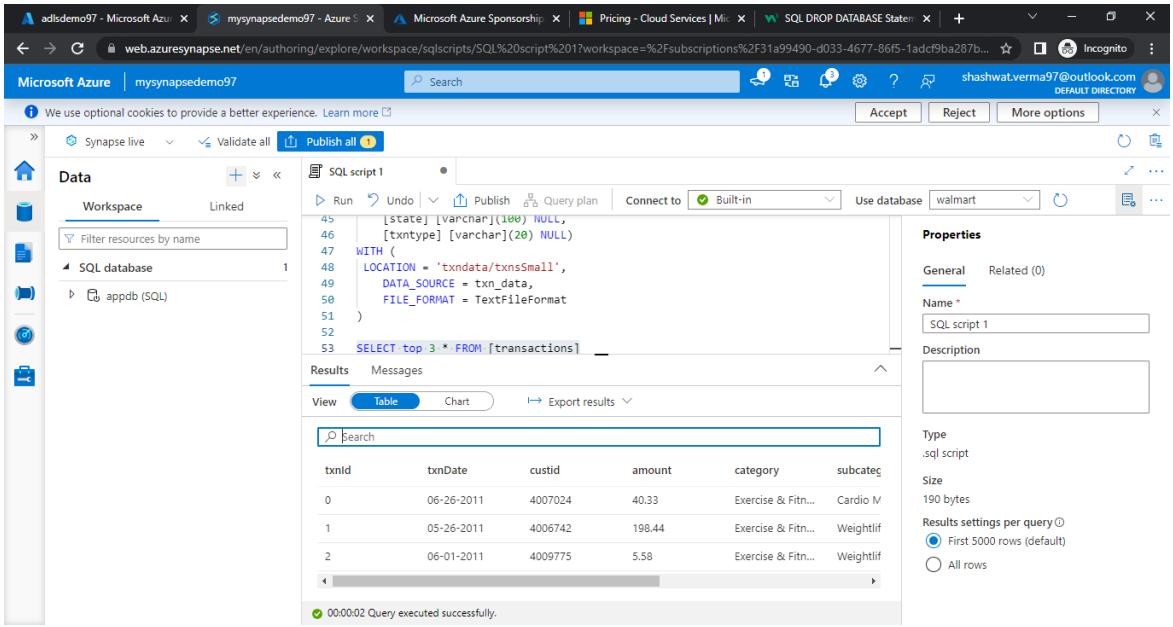
```
-- Here we are using the Shared Access Signature to authorize
the use of the Azure Data Lake Storage account
CREATE DATABASE SCOPED CREDENTIAL SasToken
WITH IDENTITY='SHARED ACCESS SIGNATURE'
, SECRET =
'?sv=2020-08-04&ss=bfqt&srt=sco&sp=rwdlacupx&se=2022-05-31T15:49
:55Z&st=2022-05-01T07:49:55Z&spr=https&sig=a%2FTrdAsJ0kGY1dFAqi3
yXoA0soWRUDYcdtUfpjMEpdU%3D';

-- This defines the source of the data.
CREATE EXTERNAL DATA SOURCE txn_data
WITH (    LOCATION   =
'https://adlsdemo97.dfs.core.windows.net/txnsdataset',
          CREDENTIAL = SasToken
)


/* This creates an External File Format object that defines the
external data that can be
present in Hadoop, Azure Blob storage or Azure Data Lake Store

Here with FIRST_ROW, we are saying please skip the first row
because this contains header information
Ref:
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-ext
ernal-file-format-transact-sql?view=sql-server-ver15&tabs=delimi
ted
*/
CREATE EXTERNAL FILE FORMAT TextFileFormat WITH (
      FORMAT_TYPE = DELIMITEDTEXT,
    FORMAT_OPTIONS (
        FIELD_TERMINATOR = ',',
        FIRST_ROW = 1))
```

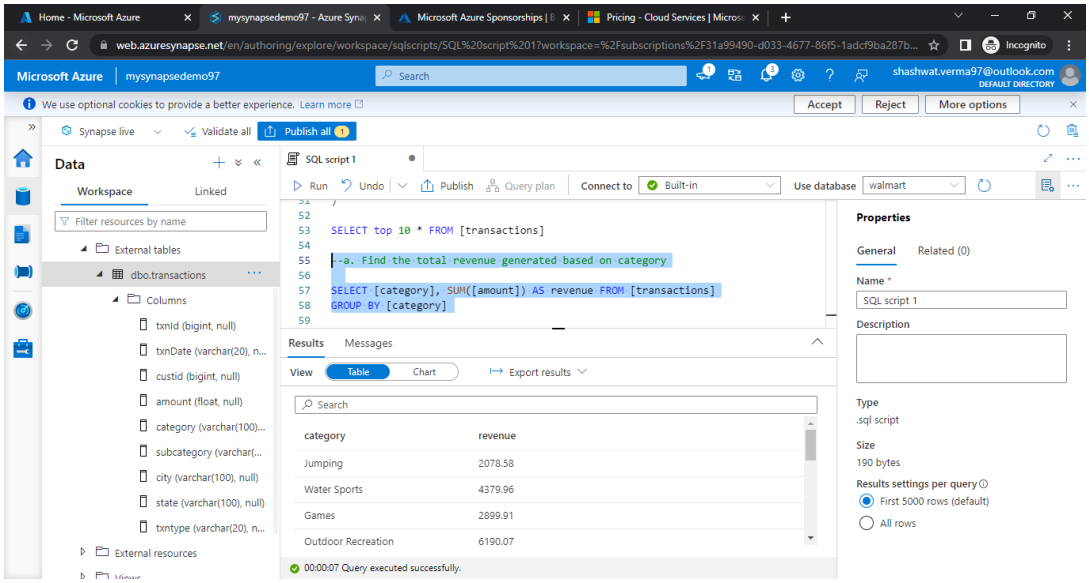
```
-- Here we define the external table
CREATE EXTERNAL TABLE [transactions]
(
     [txnId] [bigint] NULL,
     [txnDate] [varchar](20) NULL,
     [custid] [bigint] NULL,
     [amount] [float] NULL,
     [category] [varchar](100) NULL,
     [subcategory] [varchar](200) NULL,
     [city] [varchar](100) NULL,
     [state] [varchar](100) NULL,
     [txntype] [varchar](20) NULL)
WITH (
 LOCATION = 'txndata/txnsSmall',
     DATA_SOURCE = txn_data,
     FILE_FORMAT = TextFileFormat
)
--To display top 3 rows for confirmation of successful table
creation:
SELECT top 10 * FROM [transactions]
```



Problems:

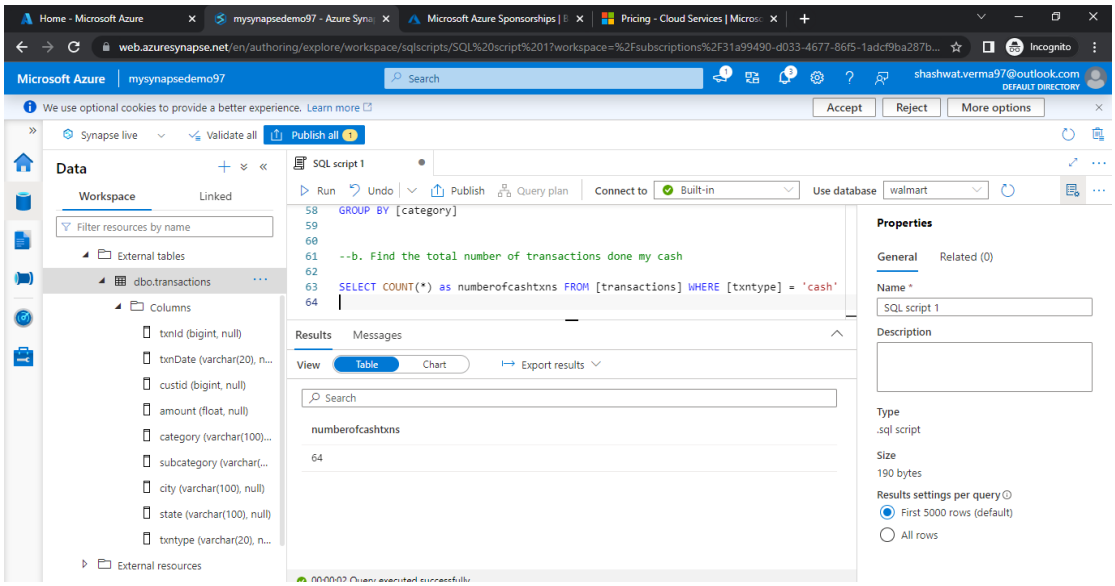
```
--a. Find the total revenue generated based on category

SELECT [category], SUM([amount]) AS revenue FROM [transactions]
GROUP BY [category]
```



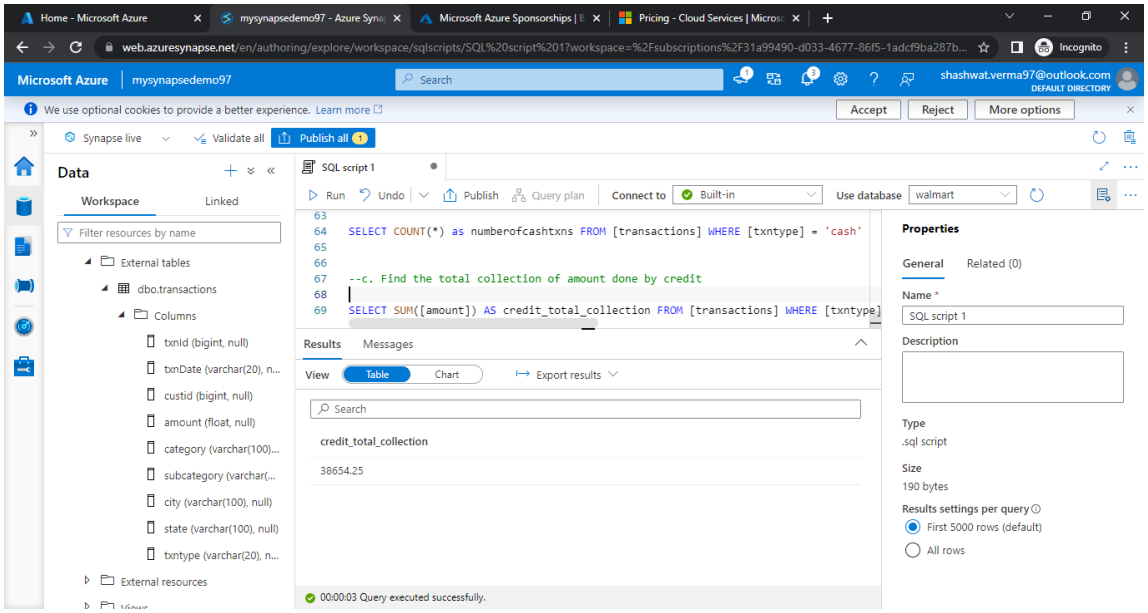

```
--b. Find the total number of transactions done my cash

SELECT COUNT(*) as numberofcashtxns FROM [transactions] WHERE
[txntype] = 'cash'
```



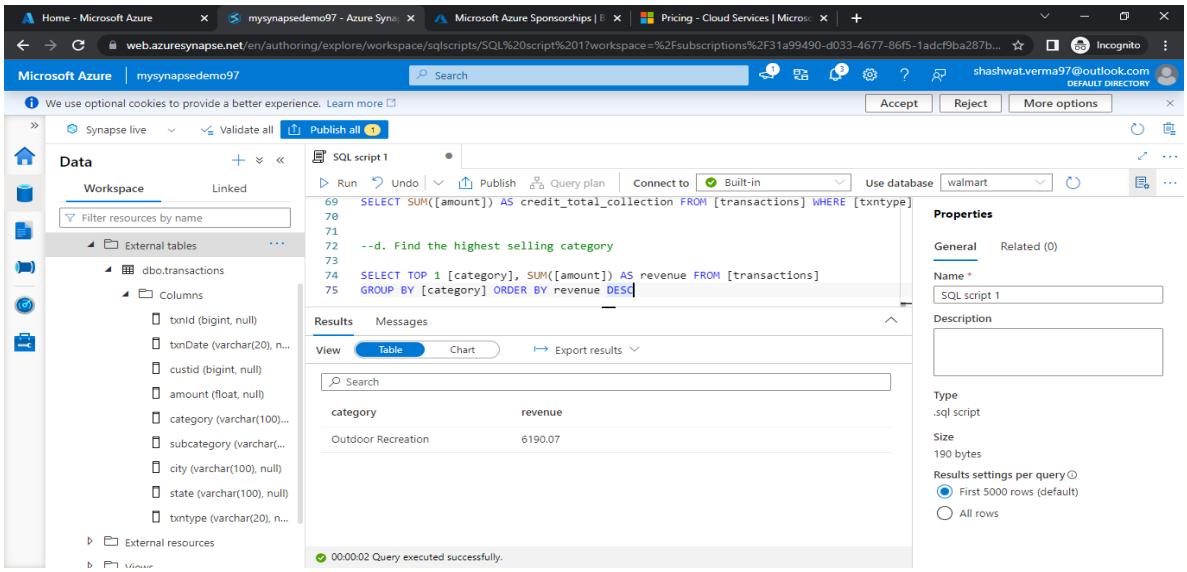
```
--c. Find the total collection of amount done by credit

SELECT SUM([amount]) AS credit_total_collection FROM
[transactions] WHERE [txntype] = 'credit'
```




```
--d. Find the highest selling category

SELECT TOP 1 [category], SUM([amount]) AS revenue FROM
[transactions] GROUP BY [category] ORDER BY revenue DESC
```



```
--e. Find the lowest selling category

SELECT TOP 1 [category], SUM([amount]) AS revenue FROM
[transactions] GROUP BY [category] ORDER BY revenue
```