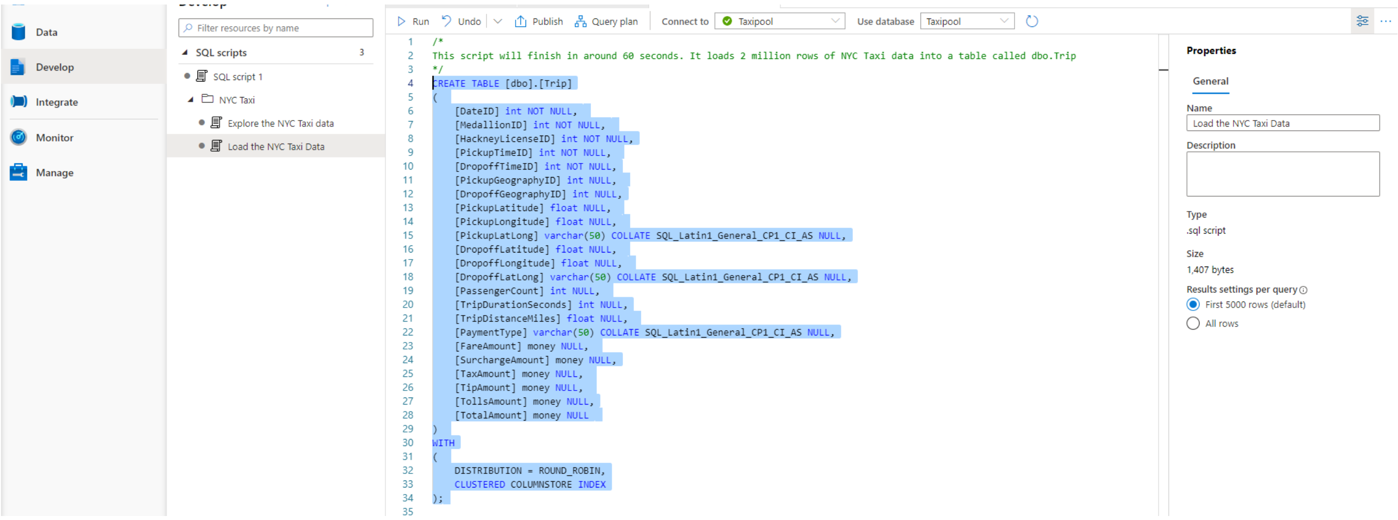
Demo: On working with dedicated SQL pool, and how we can create data, insert data or analyze data in SQL pool.

First I will create a table in SQL pool (Taxipool), next I will create folder called NYC Taxi, you can create a script in this folder or you can import a script from your desktop as well. I have imported it from my desktop.

This table is a Round robin distribution, basically Round robin means to evenly distribute data into different distributions of table.

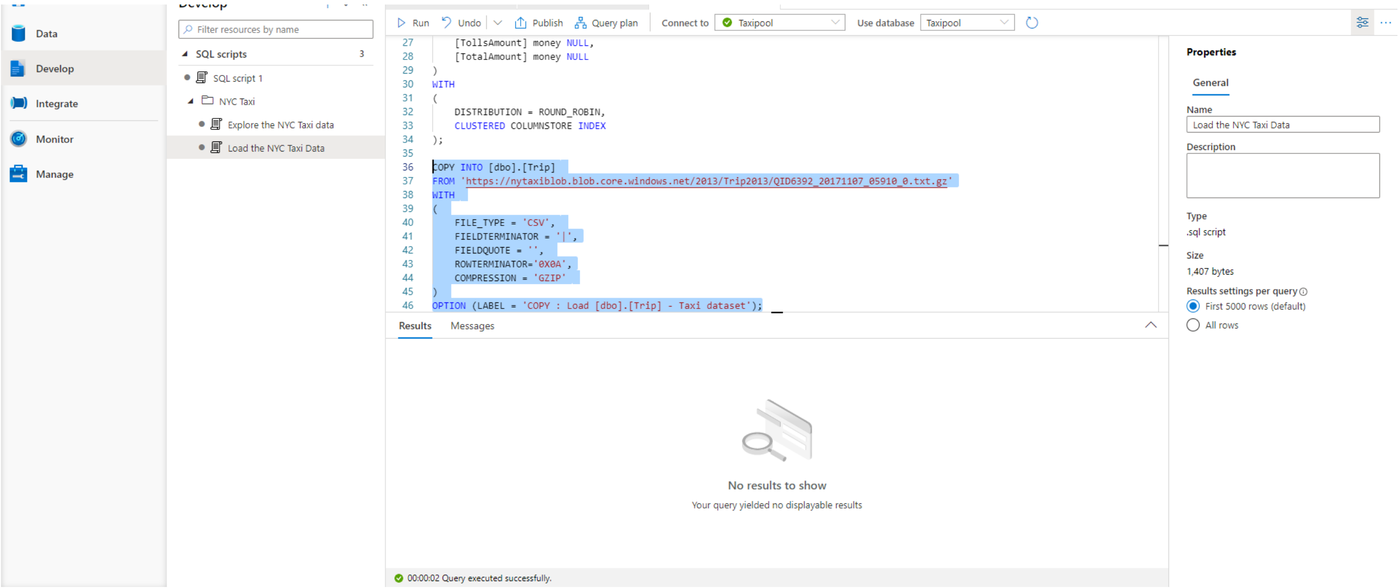


Connect to Taxipool Use database Taxipool

```sql
/*
This script will finish in around 60 seconds. It loads 2 million rows of NYC Taxi data into a table called dbo.Trip
*/
CREATE TABLE [dbo].[Trip]
(
    [DateID] int NOT NULL,
    [MedallionID] int NOT NULL,
    [HackneyLicenseID] int NOT NULL,
    [PickupTimeID] int NOT NULL,
    [DropoffTimeID] int NOT NULL,
    [PickupGeographyID] int NULL,
    [DropoffGeographyID] int NULL,
    [PickupLatitude] float NULL,
    [PickupLongitude] float NULL,
    [PickupLatLong] varchar(50) COLLATE SQL_Latin1_General_CP1_CI_AS NULL,
    [DropoffLatitude] float NULL,
    [DropoffLongitude] float NULL,
    [DropoffLatLong] varchar(50) COLLATE SQL_Latin1_General_CP1_CI_AS NULL,
    [PassengerCount] int NULL,
    [TripDurationSeconds] int NULL,
    [TripDistanceMiles] float NULL,
    [PaymentType] varchar(50) COLLATE SQL_Latin1_General_CP1_CI_AS NULL,
    [FareAmount] money NULL,
    [SurchargeAmount] money NULL,
    [TaxAmount] money NULL,
    [TipAmount] money NULL,
    [TollsAmount] money NULL,
    [TotalAmount] money NULL
)
WITH
(
    DISTRIBUTION = ROUND_ROBIN,
    CLUSTERED COLUMNSTORE INDEX
);
```

Name: Load the NYC Taxi Data
Description:
Type: .sql script
Size: 1,407 bytes
Results settings per query: ● First 5000 rows (default) ○ All rows

So basically here what we are doing, we are copying data into a Trip table from the open dataset. This script will insert a few million records into Trip table.



Connect to Taxipool Use database Taxipool

```sql
    [TollsAmount] money NULL,
    [TotalAmount] money NULL
)
WITH
(
    DISTRIBUTION = ROUND_ROBIN,
    CLUSTERED COLUMNSTORE INDEX
);

COPY INTO [dbo].[Trip]
FROM 'https://nytaxiblob.blob.core.windows.net/2013/Trip2013/QID6392_20171107_05910_0.txt.gz'
WITH
(
    FILE_TYPE = 'CSV',
    FIELDTERMINATOR = '|',
    FIELDQUOTE = '',
    ROWTERMINATOR='0X0A',
    COMPRESSION = 'GZIP'
)
OPTION (LABEL = 'COPY : Load [dbo].[Trip] - Taxi dataset');
```

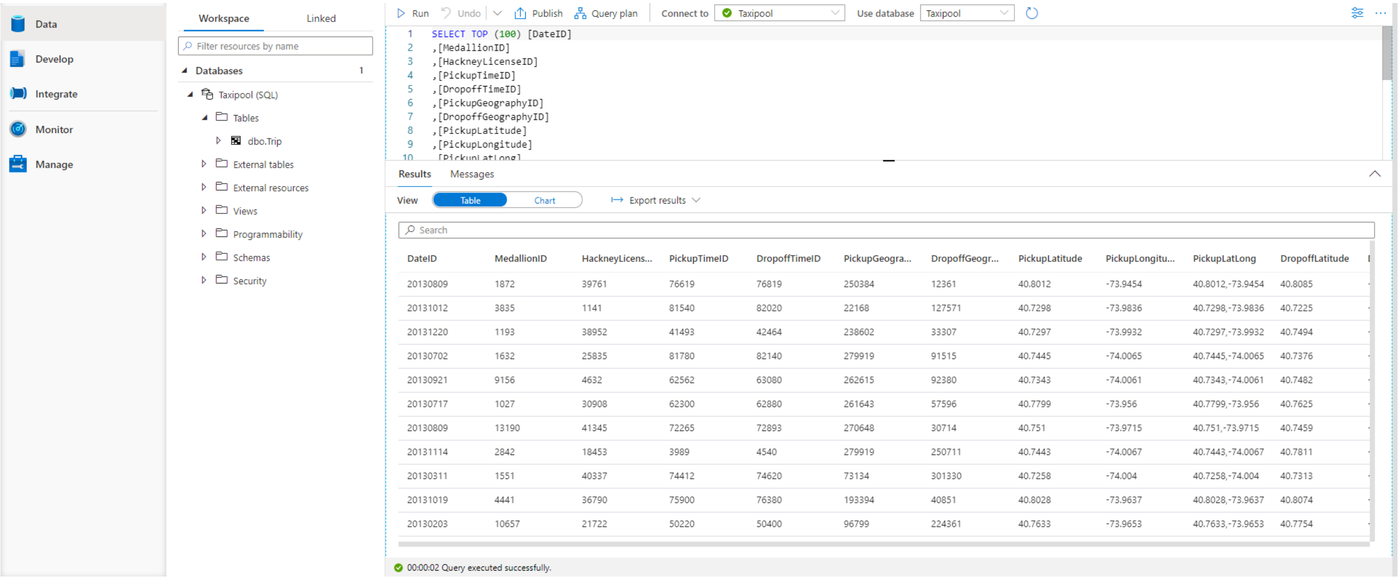
No results to show

Your query yielded no displayable results

00:00:02 Query executed successfully.

Name: Load the NYC Taxi Data
Description:
Type: .sql script
Size: 1,407 bytes
Results settings per query: ● First 5000 rows (default) ○ All rows

Now I should be able to see some data into Trip table here.

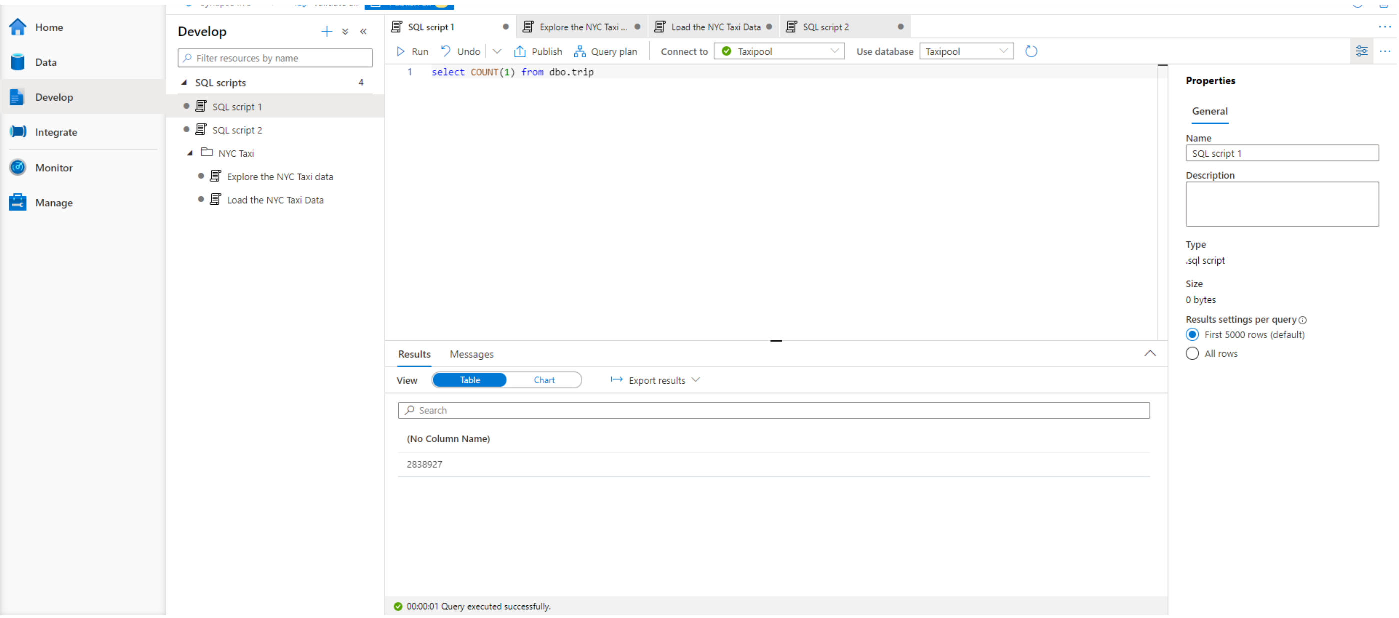


```sql
SELECT TOP (100) [DateID]
,[MedallionID]
,[HackneyLicenseID]
,[PickupTimeID]
,[DropoffTimeID]
,[PickupGeographyID]
,[DropoffGeographyID]
,[PickupLatitude]
,[PickupLongitude]
,[PickupLatLong]
```

DateID	MedallionID	HackneyLicens...	PickupTimeID	DropoffTimeID	PickupGeogra...	DropoffGeogr...	PickupLatitude	PickupLongitu...	PickupLatLong	DropoffLatitude
20130809	1872	39761	76619	76819	250384	12361	40.8012	-73.9454	40.8012,-73.9454	40.8085
20131012	3835	1141	81540	82020	22168	127571	40.7298	-73.9836	40.7298,-73.9836	40.7225
20131220	1193	38952	41493	42464	238602	33307	40.7297	-73.9932	40.7297,-73.9932	40.7494
20130702	1632	25835	81780	82140	279919	91515	40.7445	-74.0065	40.7445,-74.0065	40.7376
20130921	9156	4632	62562	63080	262615	92380	40.7343	-74.0061	40.7343,-74.0061	40.7482
20130717	1027	30908	62300	62880	261643	57596	40.7799	-73.956	40.7799,-73.956	40.7625
20130809	13190	41345	72265	72893	270648	30714	40.751	-73.9715	40.751,-73.9715	40.7459
20131114	2842	18453	3989	4540	279919	250711	40.7443	-74.0067	40.7443,-74.0067	40.7811
20130311	1551	40337	74412	74620	73134	301330	40.7258	-74.004	40.7258,-74.004	40.7313
20131019	4441	36790	75900	76380	193394	40851	40.8028	-73.9637	40.8028,-73.9637	40.8074
20130203	10657	21722	50220	50400	96799	224361	40.7633	-73.9653	40.7633,-73.9653	40.7754

00:00:02 Query executed successfully.

I have almost 2.8 million records in Trip table now.



```sql
select COUNT(1) from dbo.trip
```

(No Column Name)
2838927

00:00:01 Query executed successfully.

Name: SQL script 1

Description:

Type: .sql script

Size: 0 bytes

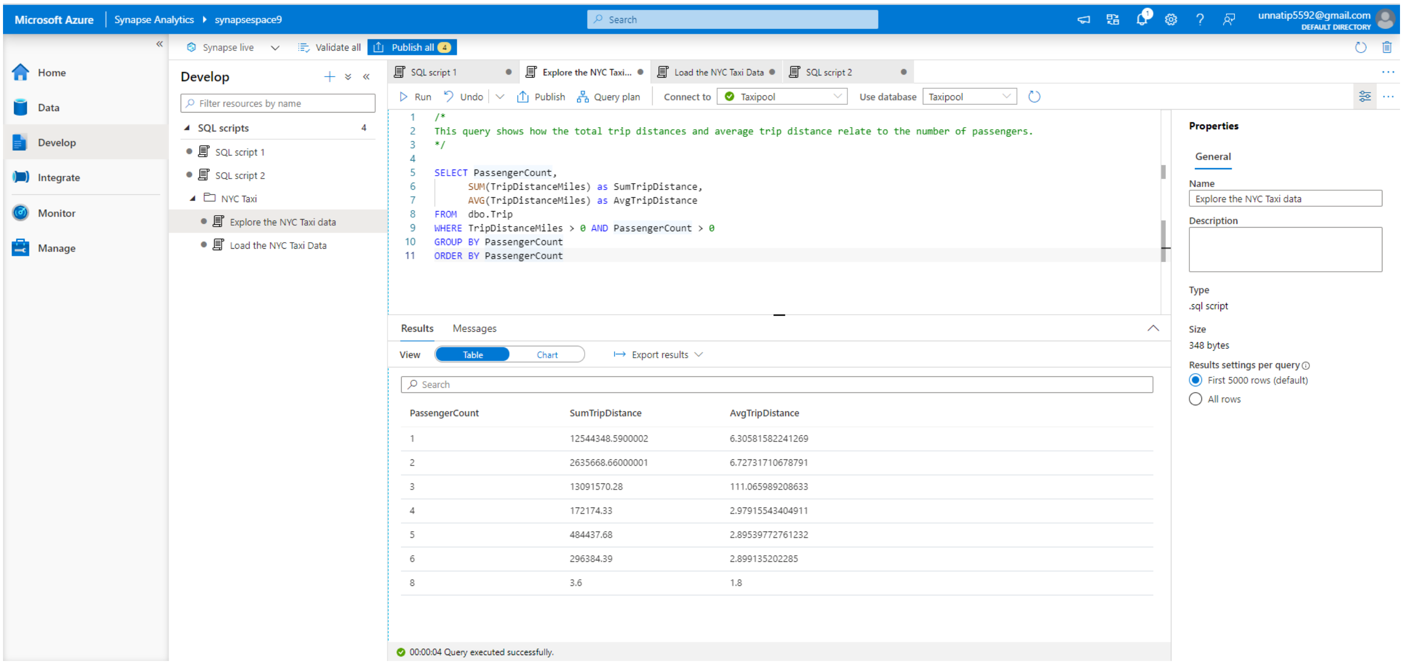
Results settings per query

● First 5000 rows (default)

○ All rows

let's explore some data from this table now in this query using a trip table, I'm taking some relevant data, data where the distance in miles is greater than Zero and and passenger count is greater than you.

This query will show the total trip distance and average trip distance related to number of passenger.



```sql
/*
This query shows how the total trip distances and average trip distance relate to the number of passengers.
*/

SELECT PassengerCount,
      SUM(TripDistanceMiles) as SumTripDistance,
      AVG(TripDistanceMiles) as AvgTripDistance
FROM  dbo.Trip
WHERE TripDistanceMiles > 0 AND PassengerCount > 0
GROUP BY PassengerCount
ORDER BY PassengerCount
```

PassengerCount	SumTripDistance	AvgTripDistance
1	12544348.5900002	6.30581582241269
2	2635668.66000001	6.72731710678791
3	13091570.28	111.065989208633
4	172174.33	2.97915543404911
5	484437.68	2.89539772761232
6	296384.39	2.899135202285
8	3.6	1.8

00:00:04 Query executed successfully.

Properties

General

Name: Explore the NYC Taxi data

Description

Type: .sql script

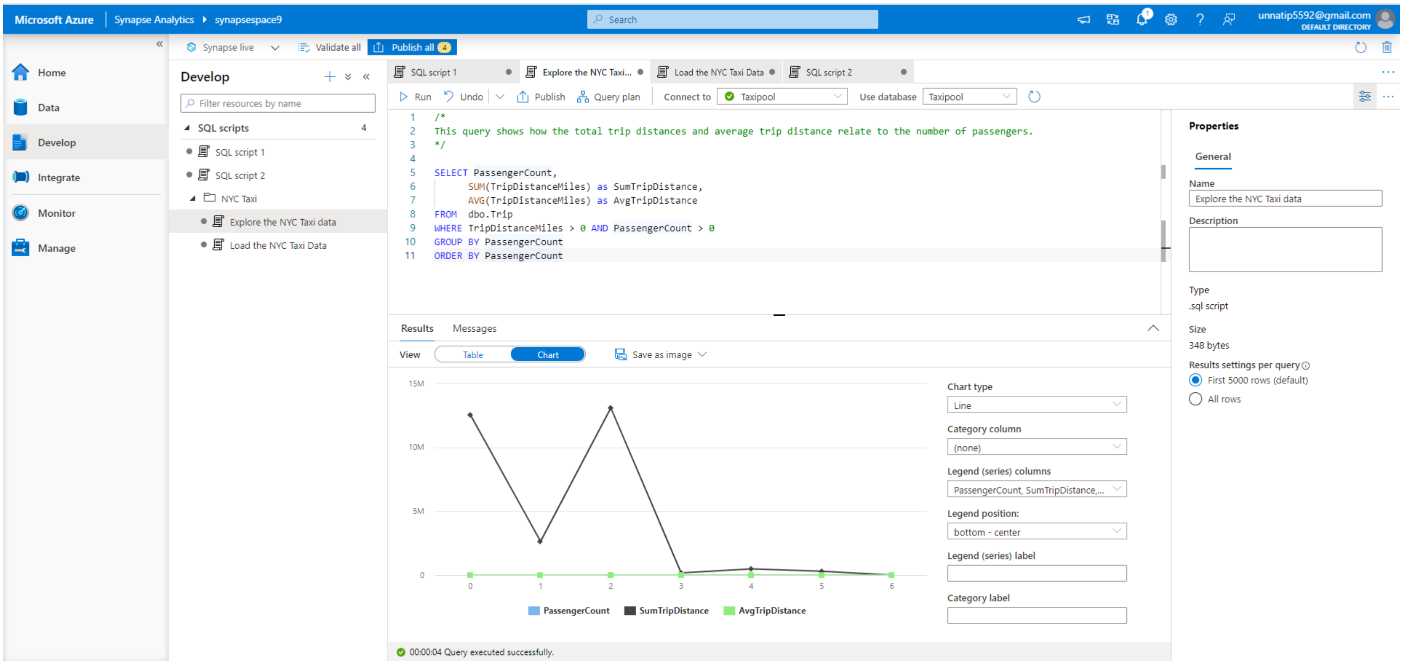
Size: 348 bytes

Results settings per query

● First 5000 rows (default)

○ All rows

You can also convert this query result in a chart as well.



```sql
/*
This query shows how the total trip distances and average trip distance relate to the number of passengers.
*/

SELECT PassengerCount,
      SUM(TripDistanceMiles) as SumTripDistance,
      AVG(TripDistanceMiles) as AvgTripDistance
FROM  dbo.Trip
WHERE TripDistanceMiles > 0 AND PassengerCount > 0
GROUP BY PassengerCount
ORDER BY PassengerCount
```

Chart type: Line

Category column: (none)

Legend (series) columns: PassengerCount, SumTripDistance,...

Legend position: bottom - center

Legend (series) label

Category label

00:00:04 Query executed successfully.

Properties

General

Name: Explore the NYC Taxi data

Description

Type: .sql script

Size: 348 bytes

Results settings per query

● First 5000 rows (default)

○ All rows