LIRR Project - SQL Script Page

Database Structure

- Month (datetime, null)
- BranchLine (nvarchar(255), null)
- OTP (float, null)
- AMPeak (float, null)
- PMPeak (float, null)
- OffPeak (float, null)

	Month	BranchLine	OTP	AMPeak	PMPeak	OffPeak
1	2015-01-01 00:00:00.000	Babylon	0.899092	0.853618	0.815789	0.927
2	2015-01-01 00:00:00.000	Far Rockaway	0.941245	0.837321	0.912281	0.961
3	2015-01-01 00:00:00.000	Hempstead	0.940344	0.894737	0.912281	0.951
4	2015-01-01 00:00:00.000	Huntington	0.879638	0.814035	0.793233	0.903
5	2015-01-01 00:00:00.000	Long Beach	0.903586	0.850877	0.84689	0.922
6	2015-01-01 00:00:00.000	Montauk	0.882625	0.842105	0.808917	0.901
7	2015-01-01 00:00:00.000	Oyster Bay	0.926667	0.917293	0.815789	0.948
8	2015-01-01 00:00:00.000	Port Jefferson	0.861423	0.782895	0.849624	0.879
9	2015-01-01 00:00:00.000	Port Washington	0.91576	0.877193	0.800554	0.941
10	2015-01-01 00:00:00.000	Ronkonkoma	0.868255	0.78655	0.851974	0.889
11	2015-01-01 00:00:00.000	West Hempstead	0.938383	0.873684	0.885965	0.96
12	2015-01-02 00:00:00.000	Babylon	0.863895	0.745066	0.834586	0.901
13	2015-01-02 00:00:00.000	Far Rockaway	0.928304	0.799043	0.923977	0.951
14	2015-01-02 00:00:00.000	Hempstead	0.919631	0.889474	0.824561	0.939
15	2015-01-02 00:00:00.000	Huntington	0.833491	0.712281	0.770677	0.866
16	2015-01-02 00:00:00.000	Long Beach	0.900952	0.785088	0.856459	0.932
17	2015-01-02 00:00:00.000	Montauk	0.839196	0.717105	0.839744	0.86
18	2015-01-02 00:00:00.000	Oyster Bay	0.863905	0.796992	0.754386	0.9
19	2015-01-02 00:00:00.000	Port Jefferson	0.774421	0.664474	0.744361	0.804
20	2015-01-02 00:00:00.000	Port Washington	0.870504	0.80117	0.711911	0.912
21	2015-01-02 00:00:00.000	Ronkonkoma	0.842206	0.719298	0.865132	0.867
22	2015-01-02 00:00:00.000	West Hempstead	0.92623	0.905263	0.780702	0.962

Data Cleansing Tasks

-- 01. Find 'BranchLine' data where the numbers are abnormal and then delete

```
SELECT distinct [BranchLine], count (*)
FROM [LIRR].[dbo].[MTA_LIRR$]
GROUP BY BranchLine
```

	BranchLine	(No column name)
1	Hempstead	218
2	Atlantic	13
3	System Total	109
4	Port Jefferson	218
5	Montauk	218
6	Babylon	218
7	Huntington	218
8	West Hempstead	218
9	Grand Central Direct	2
10	Port Washington	218
11	Ronkonkoma	218
12	Long Beach	218
13	Oyster Bay	218
14	Far Rockaway	218

Findings showed that 'Atlantic', 'System Total' and 'Grand Central Direct' reflected abnormal counts thus implying they must be eliminated from the group

-- 02 Remove all records relating to BranchLine as 'Atlantic', 'System Total' or 'Grand Central Direct'

```
DELETE from MTA_LIRR$
WHERE
        BranchLine = 'Atlantic' or
        BranchLine like '%Total' or
        BranchLine = 'Grand Central Direct'
```

```
-- 03 Remove ‘OTP’ column as ‘AMPeak’, ‘PM Peak’ and ‘OffPeak’ represent the OTP Population

ALTER TABLE MTA_LIRR$
DROP COLUMN OTP

-- 04 Adding Spaces to ‘AMPeak’, ‘PMPeak’ and ‘OffPeak’

EXECUTE sp_rename '[dbo].[MTA_LIRR$].AMPeak', 'AM Peak', 'column'
EXECUTE sp_rename '[dbo].[MTA_LIRR$].PMPeak', 'PM Peak', 'column'
EXECUTE sp_rename '[dbo].[MTA_LIRR$].OffPeak', 'Off Peak', 'column'

-- 05 Change type of ‘Month’ as NVACHAR (255).  This was required so that the Date column can be added and constructed the correct way

 ALTER TABLE [dbo].[MTA_LIRR$]
 ALTER COLUMN [Month] NVArchar(255)

-- 06 Final SQL Statement to Upload to Power BI

SELECT -- Rearranging 'Month' so that the dates are arranged in the correct way
                CONCAT((YEAR([Month])),'-',(DAY([Month])),'-',(MONTH([Month]))) AS 'Date'
        ,[BranchLine]
    ,[PeakPeriod]
    ,[OTP]
        -- Adding column as an indicator if the OTP is equal to, or over 90%
        ,CASE
                WHEN [OTP] >=0.9 then 1
                ELSE 0
                END as OTP_Over90Perc
 FROM [LIRR].[dbo].[MTA_LIRR$]
 UNPIVOT
 ([OTP] for [PeakPeriod] in ([AM Peak], [PM Peak], [Off Peak])
 ) OTP
```

	Date	BranchLine	PeakPeriod	OTP	OTP_Over90Perc
1	2015-1-1	Babylon	AM Peak	0.853618	0
2	2015-1-1	Babylon	PM Peak	0.815789	0
3	2015-1-1	Babylon	Off Peak	0.927	1
4	2015-1-1	Far Rockaway	AM Peak	0.837321	0
5	2015-1-1	Far Rockaway	PM Peak	0.912281	1
6	2015-1-1	Far Rockaway	Off Peak	0.961	1
7	2015-1-1	Hempstead	AM Peak	0.894737	0
8	2015-1-1	Hempstead	PM Peak	0.912281	1
9	2015-1-1	Hempstead	Off Peak	0.951	1
10	2015-1-1	Huntington	AM Peak	0.814035	0
11	2015-1-1	Huntington	PM Peak	0.793233	0
12	2015-1-1	Huntington	Off Peak	0.903	1
13	2015-1-1	Long Beach	AM Peak	0.850877	0
14	2015-1-1	Long Beach	PM Peak	0.84689	0
15	2015-1-1	Long Beach	Off Peak	0.922	1
16	2015-1-1	Montauk	AM Peak	0.842105	0
17	2015-1-1	Montauk	PM Peak	0.808917	0
18	2015-1-1	Montauk	Off Peak	0.901	1
19	2015-1-1	Oyster Bay	AM Peak	0.917293	1
20	2015-1-1	Oyster Bay	PM Peak	0.815789	0
21	2015-1-1	Oyster Bay	Off Peak	0.948	1