NYC TRANSPORTATION

EXCEL

SUMMARY

- Data Source: NYC Open Data (data.cityofnewyork.us)
- Conclusions:
  - Mechanical failures and heavy traffic are the most common reasons for bus breakdowns and delays.
  - The NYC boroughs with the highest average bus delay times represent the most populous boroughs with the largest populations of people with vehicles.
  - There is a strong correlation between the day of the week and the number of bus breakdowns and delays.
- Challenges: The “bus company name” data was non-standardized and required some guess work due to lack of domain knowledge and access to additional information to better reconcile the information presented.
- With more time, I would explore further into the data to determine the number of breakdowns per bus company, the most frequent breakdown location of each bus company and the total number of trips made by each bus company.

QUESTIONS PRESENTED

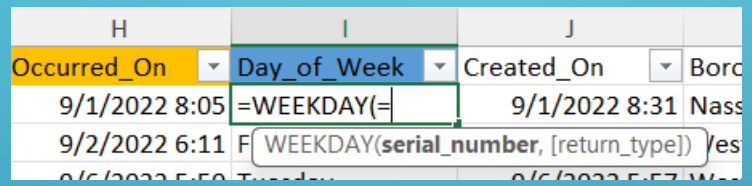
- What are the most common reasons for bus delays and breakdowns?
- How do bus delay times vary by bus company and borough?
- Is there a correlation between specific days of the week and the frequency of bus breakdowns or delays?

DATA PREPARATION

- Techniques used:
  - Data extraction function
  - Text to columns

DATA EXTRACTION – WEEKDAY FUNCTION

Occurred_On
9/1/2022 8:05
9/2/2022 6:11
9/6/2022 5:50
9/6/2022 8:12
9/7/2022 7:27
9/7/2022 7:40
9/8/2022 6:20
3/30/2020 6:30
9/4/2019 6:27
9/4/2019 6:37
9/16/2019 7:34
10/15/2019 13:14
10/31/2019 8:05
11/22/2019 14:35
12/6/2019 6:40
12/6/2019 6:50
12/6/2019 6:50
12/20/2019 6:30
10/27/2021 7:43

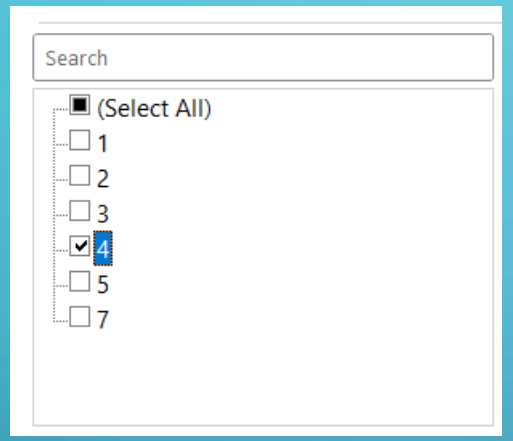


The weekday function returns a number between 1 and 7 to represent the corresponding day of the week for a particular date.

Day_of_Week
4
5
2
2
3
3
4
1
3
3
1
2
4
5
5
5
5
5
3

DATA EXTRACTION – WEEKDAY FUNCTION

Day_of_Week
4
5
2
2
3
3
4
1
3
3
1
2
4
5
5
5
5
5
3



The data is then filtered for each individual number at which point the corresponding day of the week is manually entered and dragged downwards (double-click) for all relevant cells.

Day_of_Week
4
4
4
4
4
4
4
4
4
4
4
4
4
4
4
4
4
4
4
4
4

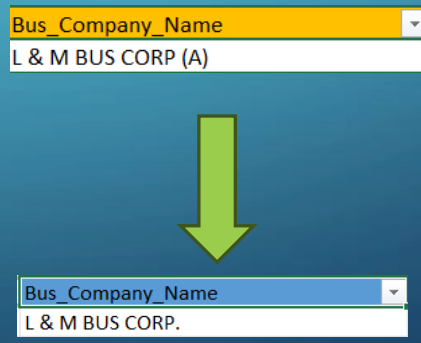
Day_of_Week
Thursday
Thursday
Thursday
Thursday
Thursday
Thursday
Thursday
4
4
4

NON-STANDARDIZED INFORMATION

Non-standardized Standardized

Bus_Company_Name	Bus_Company_Name
L & M BUS CORP (A)	L & M BUS CORP.
CONSOLIDATED BUS TRANSIT, INC.	CONSOLIDATED BUS TRANS. INC.
BORO TRANSIT, INC.	BORO TRANSIT, INC.
LEESEL TRANSPORTATION CORP (B2192)	LEESEL TRANSPORTATION COR
VAN TRANS LLC (B2192)	VAN TRANS LLC (B2192)
GRANDPA`S BUS CO., INC.	GRANDPA`S BUS CO., INC.
RELIANT TRANSPORTATION, INC (B2321)	RELIANT TRANSPORTATION, INC (B2321)
RELIANT TRANSPORTATION, INC (B2321)	RELIANT TRANSPORTATION, INC (B2321)
RELIANT TRANSPORTATION, INC (B2321)	RELIANT TRANSPORTATION, INC (B2321)
G.V.C. LTD. (B2192)	GVC LTD
L & M BUS CORP.	L & M BUS CORP.
DON THOMAS BUSES, INC.	Don Thomas Buses
PIONEER TRANSPORTATION CORP	PIONEER TRANSPORTATION CORP
VAN TRANS LLC (B2192)	VAN TRANS LLC (B2192)
G.V.C. LTD. (B2192)	GVC LTD
RELIANT TRANSPORTATION, INC (B2321)	RELIANT TRANSPORTATION, INC (B2321)
SNT BUS INC	SNT BUS INC
BOBBY`S BUS CO. INC.	BOBBY`S BUS CO. INC.
GVC LTD	GVC LTD
L & M BUS CORP.	L & M BUS CORP.
VAN TRANS LLC (B2192)	VAN TRANS LLC (B2192)
RELIANT TRANSPORTATION, INC (B2321)	RELIANT TRANSPORTATION, INC (B2321)

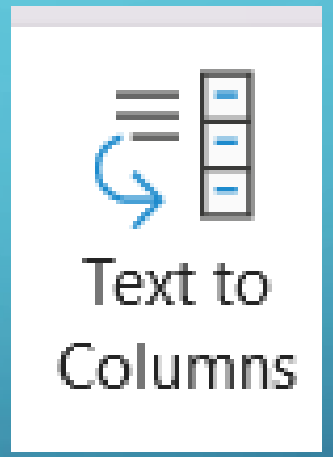
The bus company name data contained many inconsistencies. As a work around, I reconciled an apparent duplicate by re-naming the company the most generic name provided for all similarly named bus companies contained within the data set. I used the filter function to group the similar names before dragging the correction downwards for ease of correction across the thousands of different cells.

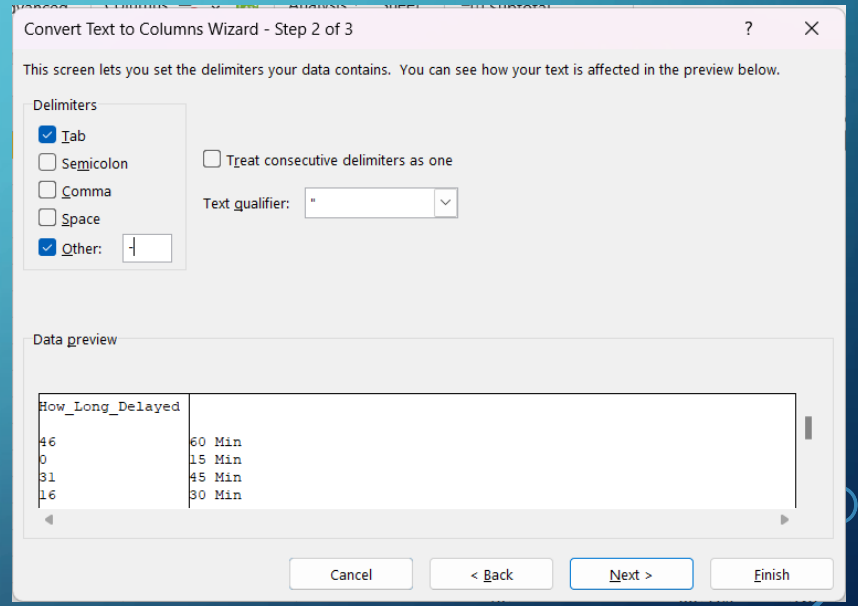


TEXT-TO-COLUMNS

How_Long_Delaye
46-60 Min
31-45 Min
16-30 Min
16-30 Min
61-90 Min
46-60 Min
31-45 Min
31-45 Min
31-45 Min
31-45 Min
31-45 Min
31-45 Min
31-45 Min
16-30 Min
46-60 Min
16-30 Min

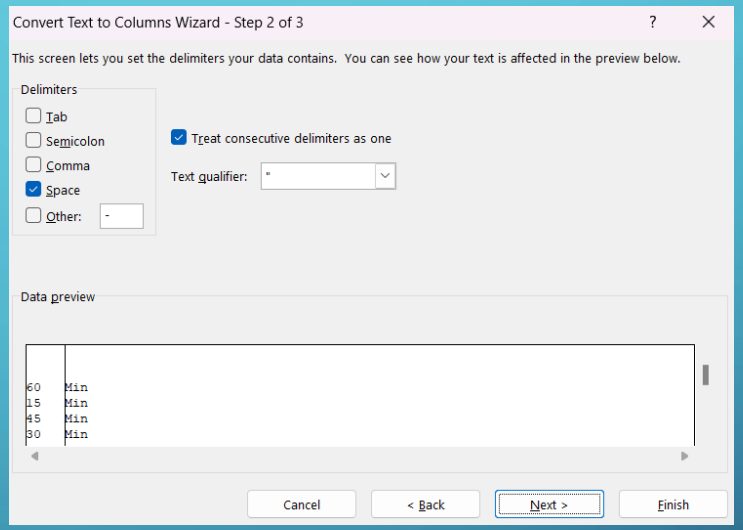
This information was split into multiple columns to allow for the averaging of delay times at the low end versus the high end to allow for ease of comparability in average bus delay times between bus companies and boroughs.





TEXT-TO-COLUMNS

How_Long_Delaye	
46	60 Min
31	45 Min
16	30 Min
16	30 Min
61	90 Min
46	60 Min
31	45 Min
31	45 Min
31	45 Min
31	45 Min
31	45 Min
31	45 Min
16	30 Min
46	60 Min
16	30 Min



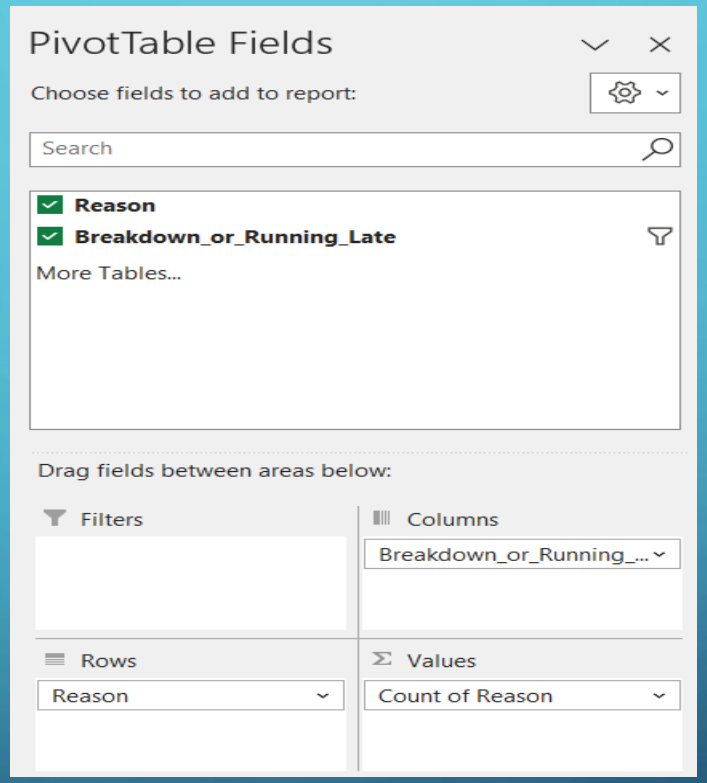
46	60
31	45
16	30
16	30
61	90
46	60
31	45
31	45
31	45
31	45
31	45
31	45
31	45
16	30
46	60
16	30

QUESTIONS PRESENTED

QUESTION PRESENTED #1

What are the most common reasons for bus delays and breakdowns?

Reason	Breakdown_or_Running_Late
Mechanical Problem	Breakdown
Heavy Traffic	Running Late
Heavy Traffic	Running Late
Heavy Traffic	Running Late
Heavy Traffic	Running Late
Heavy Traffic	Running Late
Heavy Traffic	Running Late
Heavy Traffic	Running Late
Mechanical Problem	Running Late
Heavy Traffic	Running Late
Heavy Traffic	Running Late
Heavy Traffic	Running Late
Heavy Traffic	Running Late
Heavy Traffic	Running Late

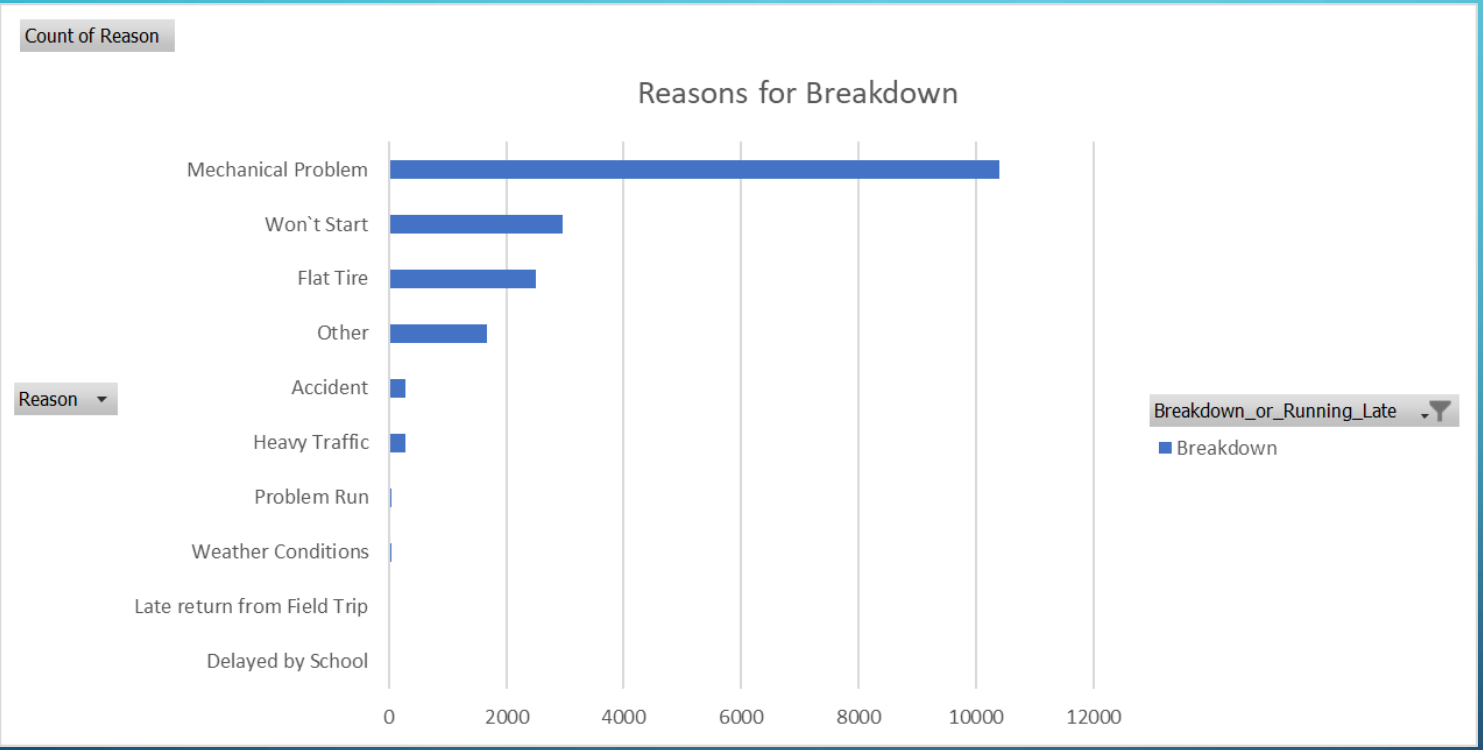


Count of Reason	Column Labels	
Row Labels	Breakdown	Grand Total
Delayed by School	12	12
Late return from Field Trip	16	16
Weather Conditions	41	41
Problem Run	48	48
Heavy Traffic	279	279
Accident	290	290
Other	1674	1674
Flat Tire	2499	2499
Won`t Start	2970	2970
Mechanical Problem	10399	10399
Grand Total	**18228**	**18228**

Count of Reason	Column Labels	
Row Labels	Running Late	Grand Total
Delayed by School	1737	1737
Won`t Start	1932	1932
Accident	2101	2101
Late return from Field Trip	2255	2255
Flat Tire	2401	2401
Weather Conditions	4318	4318
Problem Run	10193	10193
Mechanical Problem	13695	13695
Other	49135	49135
Heavy Traffic	176196	176196
Grand Total	**263963**	**263963**

I ran a pivot table on the “reason” and “count of reason” to best compare the type of bus delay (breakdown versus running late) and its number of occurrences across the entire data set.

QUESTION PRESENTED #1



I selected a bar chart to visualize the data given its categorical nature.

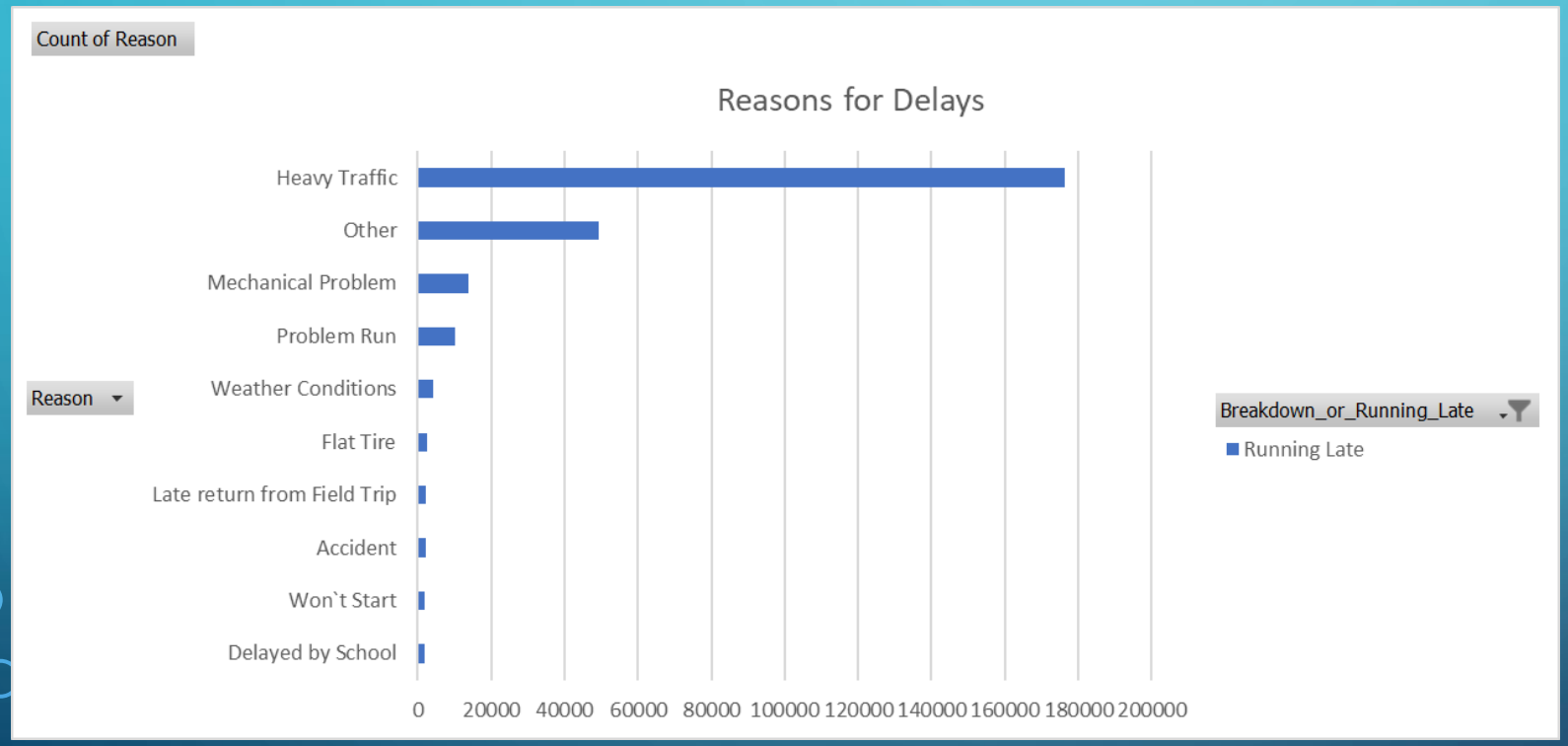
Conclusion:

Most bus breakdowns are due to mechanical problems.

Recommendation:

Establish a regular preventative maintenance schedule for the bus fleet to combat the high rate of mechanical problems.

QUESTION PRESENTED #1



Conclusion:

Heavy traffic is the overwhelming reason for bus delays.

Recommendations:

- Have buses execute their runs to pick-up children outside of rush hour.
- Investigate new route options that may circumvent the current heavy traffic that many buses encounter.

QUESTION PRESENTED #2

How do bus delay times vary by bus company and borough?

Bus_Company_Name	Short Delay	Long Delay
L & M BUS CORP.	46	60
BORO TRANSIT, INC.	31	45
GVC LTD	16	30
PIONEER TRANSPORTATION CORP	16	30
PRIDE TRANSPORTATION (SCH AGE)	61	90
CONSOLIDATED BUS TRANS. INC.	46	60
RELIANT TRANSPORTATION, INC (B2321)	31	45
RELIANT TRANSPORTATION, INC (B2321)	31	45
Don Thomas Buses	31	45
LEESEL TRANSPORTATION COR	31	45
GVC LTD	31	45

Row Labels	Average of Short_Delay_Estimate	Average of Long_Delay_Estimate
MAR-CAN TRANSPORT CO. INC (B2192)	45	66
LORINDA ENTERPRISES, LTD.	45	63
MJT BUS COMPANY, INC	45	60
FORTUNA BUS COMPANY	46	60
IC BUS INC.	46	60
LOGAN BUS COMPANY INC.	46	66
LITTLE RICHIE BUS SERVICE	47	67
PRIDE TRANSPORTATION (SCH AGE)	49	70
LITTLE LINDA BUS CO.,INC.	52	75
SELBY TRANSPORTATION	60	88
Grand Total	**48**	**69**

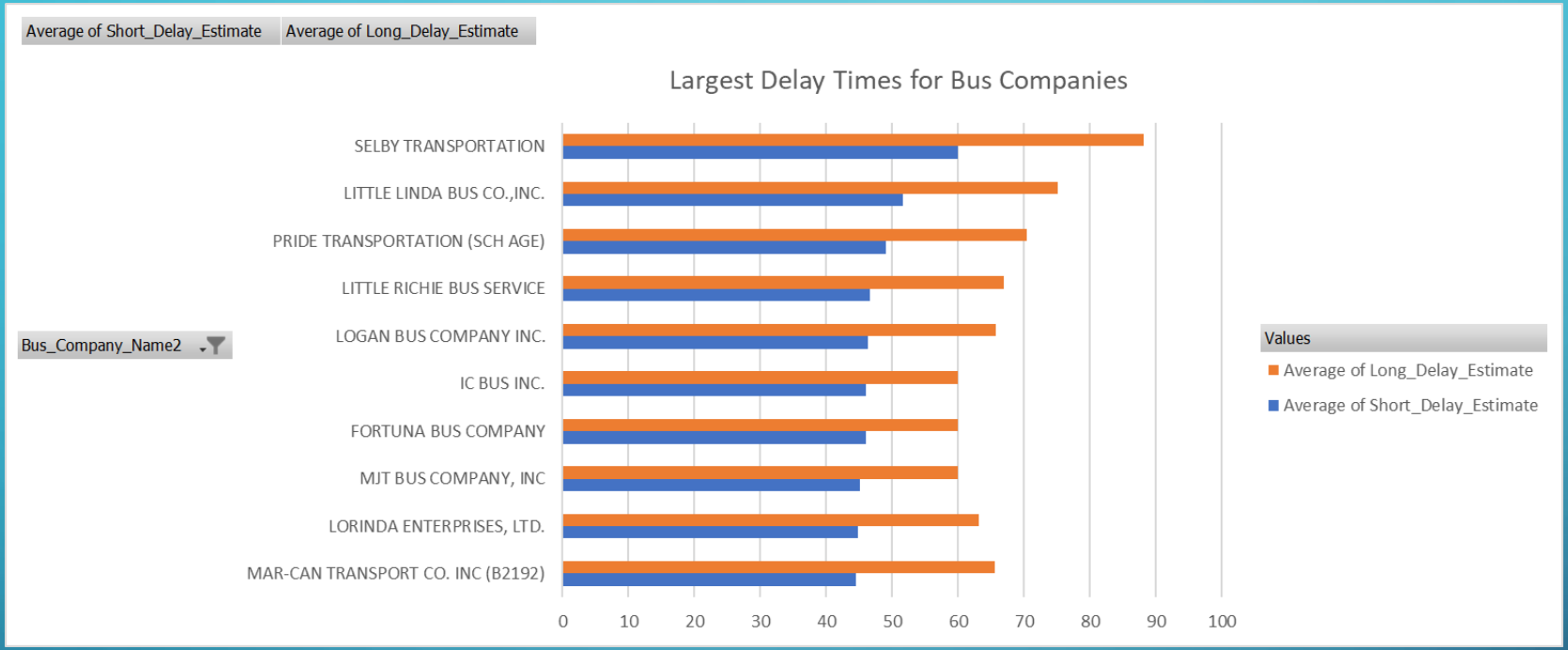
Row Labels	Average of Short_Delay_Estimate	Average of Long_Delay_Estimate
Connecticut	12	26
Bronx	20	35
All Boroughs	21	36
Nassau County	26	41
Brooklyn	27	43
New Jersey	29	45
Westchester	30	48
Staten Island	33	47
Queens	33	50
Rockland County	33	51
Manhattan	37	53
Grand Total	**29**	**45**

Connecticut
Bronx
All Boroughs
Nassau County
Brooklyn
New Jersey
Westchester
Staten Island
Queens
Rockland County
Manhattan

I ran a pivot table on borough and bus company name against the average short/long average bus delay times.

I limited the bus company output to the top 10 bus companies (by highest average delay times).

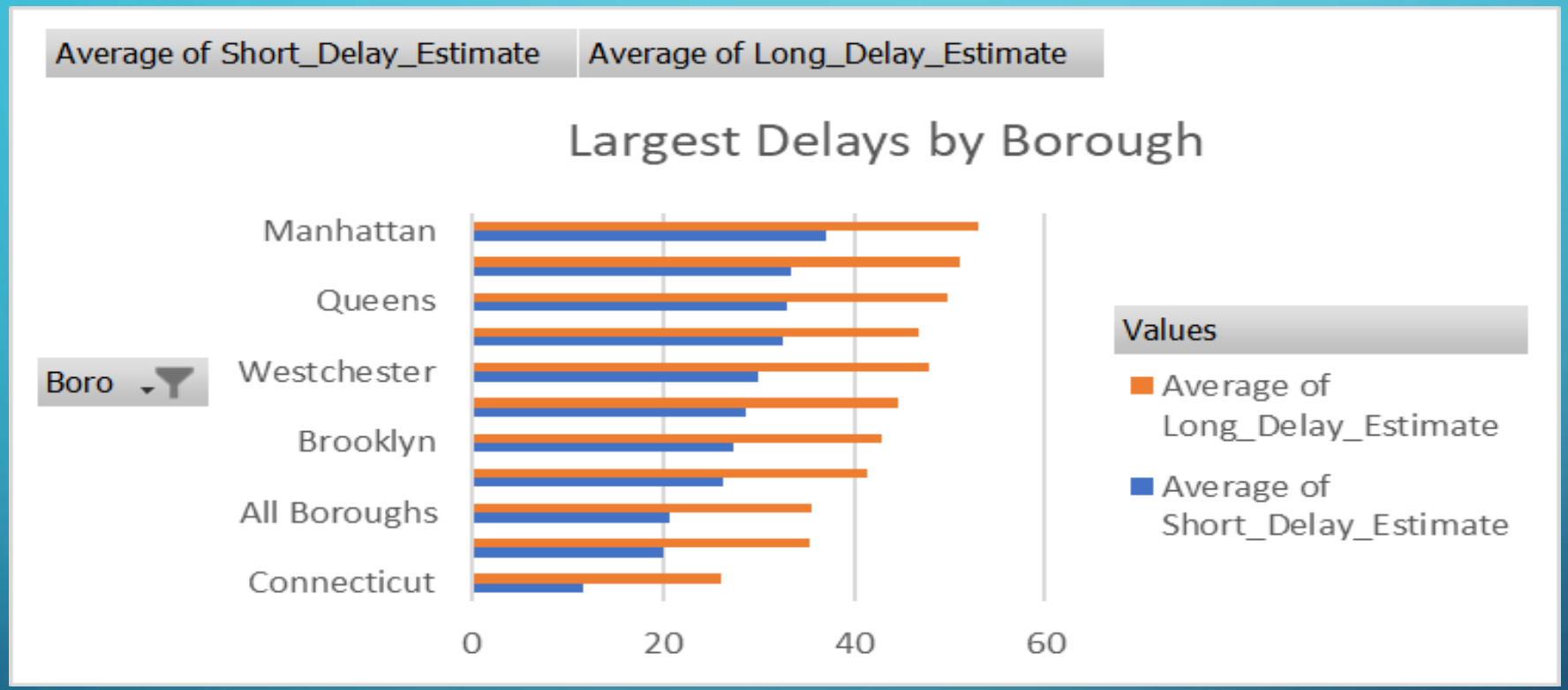
QUESTION PRESENTED #2



Conclusion:

These high average bus delay times could be the result of mechanical problems which may be alleviated with regular preventative maintenance; however, additional exploratory analysis of the data may support the cancelation of some or all contracts of the bus companies identified above as well as others.

QUESTION PRESENTED #2



Conclusion:

The boroughs with the highest average bus delay times have the highest concentration of people and vehicles which accounts for longer delay times.

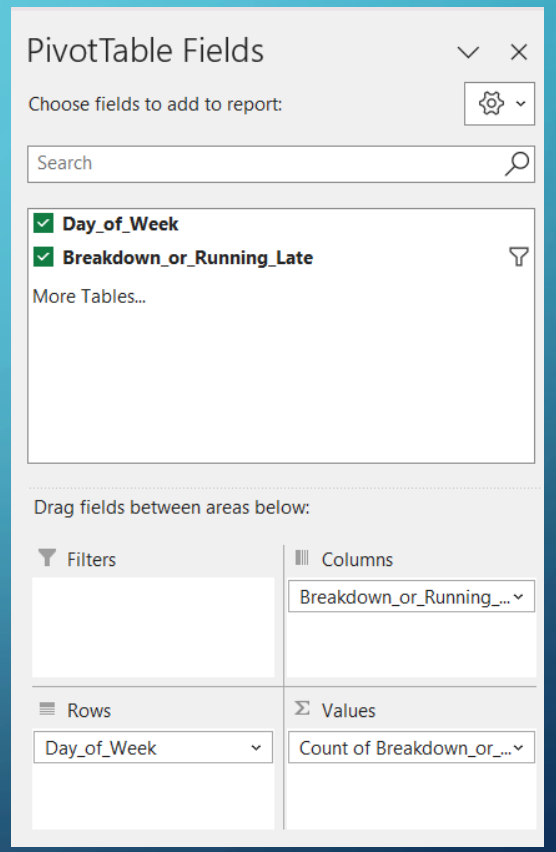
QUESTION PRESENTED #3

Is there a correlation between specific days of the week and the frequency of bus breakdowns or delays?

Count of Breakdown_or_Running_Late	Column Labels	
Row Labels	Breakdown	Grand Total
Monday	4146	4146
Tuesday	3619	3619
Wednesday	3573	3573
Thursday	3572	3572
Friday	3318	3318
Grand Total	**18228**	**18228**

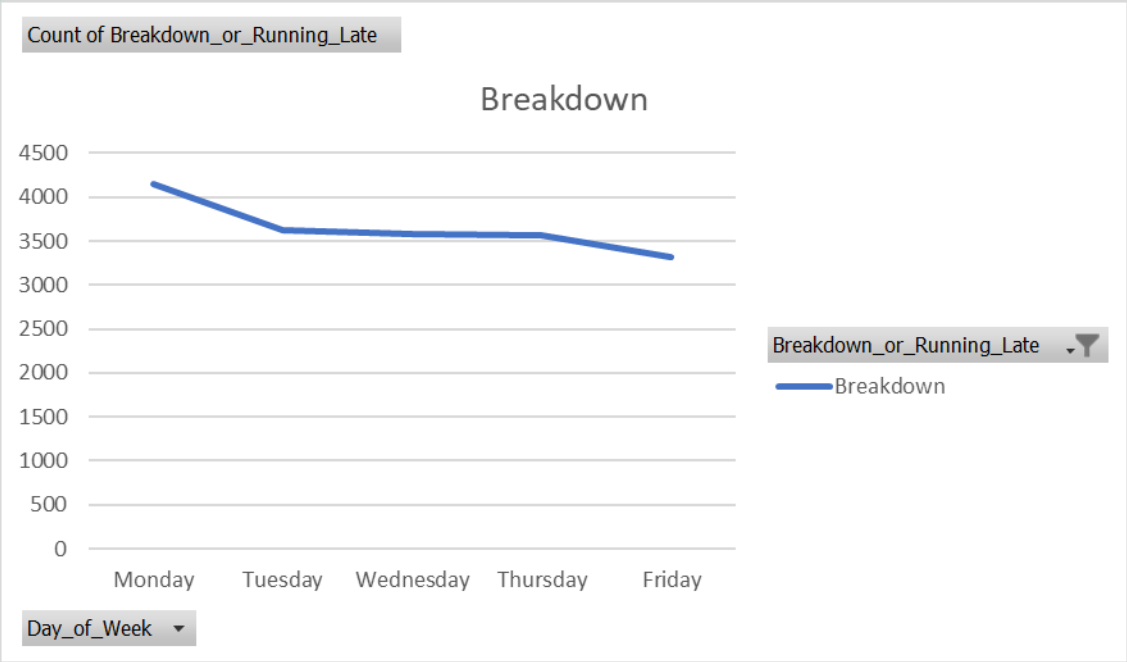
Count of Breakdown_or_Running_Late	Column Labels	
Row Labels	Running Late	Grand Total
Monday	54667	54667
Tuesday	52332	52332
Wednesday	53589	53589
Thursday	53862	53862
Friday	49512	49512
Grand Total	**263962**	**263962**

Day_of_Week	Breakdown_or_Running_Late
Thursday	Breakdown
Thursday	Running Late
Friday	Running Late
Tuesday	Running Late
Tuesday	Running Late
Wednesday	Running Late
Wednesday	Running Late
Thursday	Running Late
Monday	Running Late
Wednesday	Running Late
Wednesday	Running Late
Monday	Running Late
Tuesday	Running Late
Thursday	Running Late
Friday	Running Late
Friday	Running Late
Friday	Running Late
Friday	Running Late
Friday	Running Late
Wednesday	Running Late



I ran a pivot on day-of-week, delay type and count-of-delay type for ease of comparison.

QUESTION PRESENTED #3



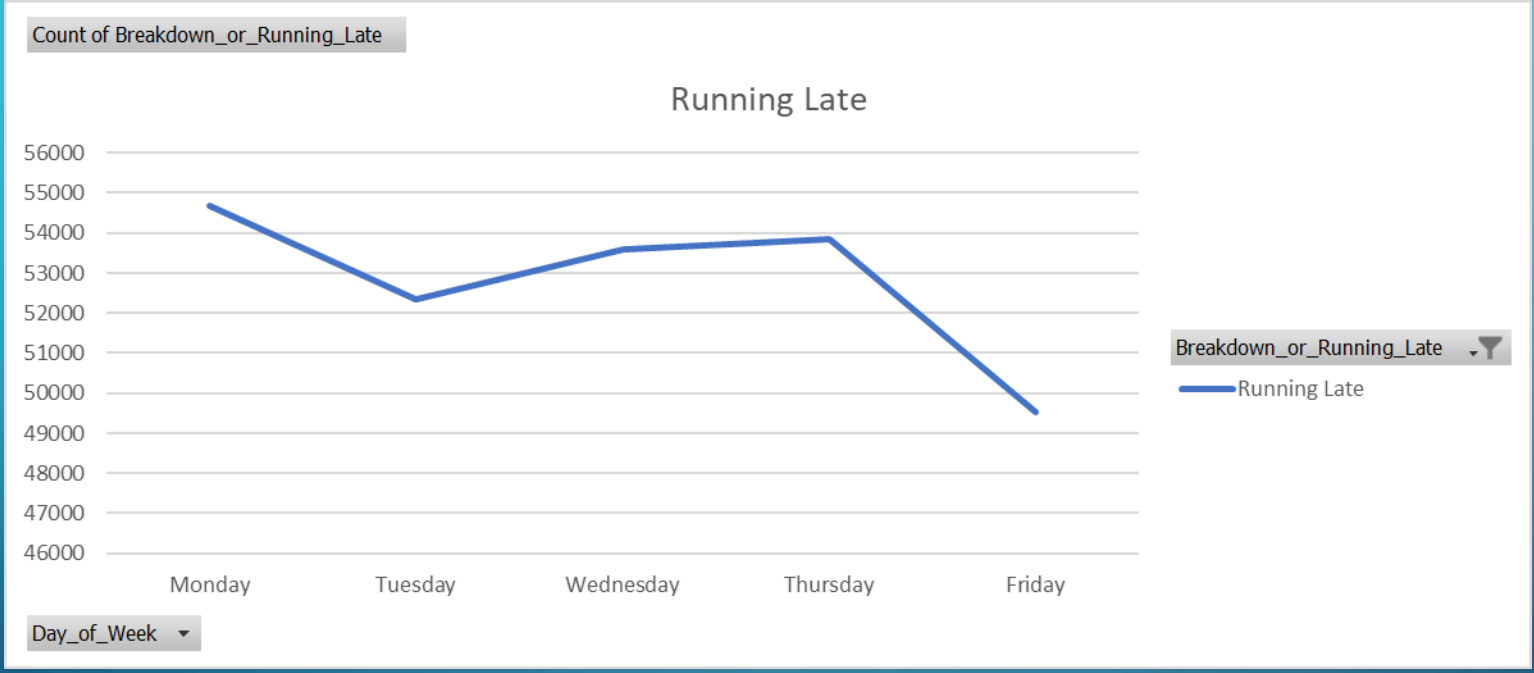
Conclusion:

The greatest number of bus breakdowns are realized on Mondays, followed by a gradual decrease in the number of bus breakdowns throughout the remainder of the week.

Recommendation:

Schedule weekend bus maintenance to alleviate weekday bus breakdowns.

QUESTION PRESENTED #3



Conclusion:

Most bus delays happen on Mondays, followed by a gradual decrease in bus delays during the remainder of the week.

Since most bus delays are caused by heavy traffic, we can assume that the gradual decrease in bus delays throughout the week is due to less people driving on the roads later in the week.