

LAB 1: CONNECTING TO AN ONLINE DATA SOURCE

**BY
JORDAN GOLDMEIER**

LAB: CONNECTING TO AN ONLINE DATA SOURCE

MODULE OUTCOMES

- ▶ Connecting to a remote data service
- ▶ Using Power Query to transform the data
- ▶ Using Power BI Visualization to quickly create an interactive visualization

DESCRIPTION

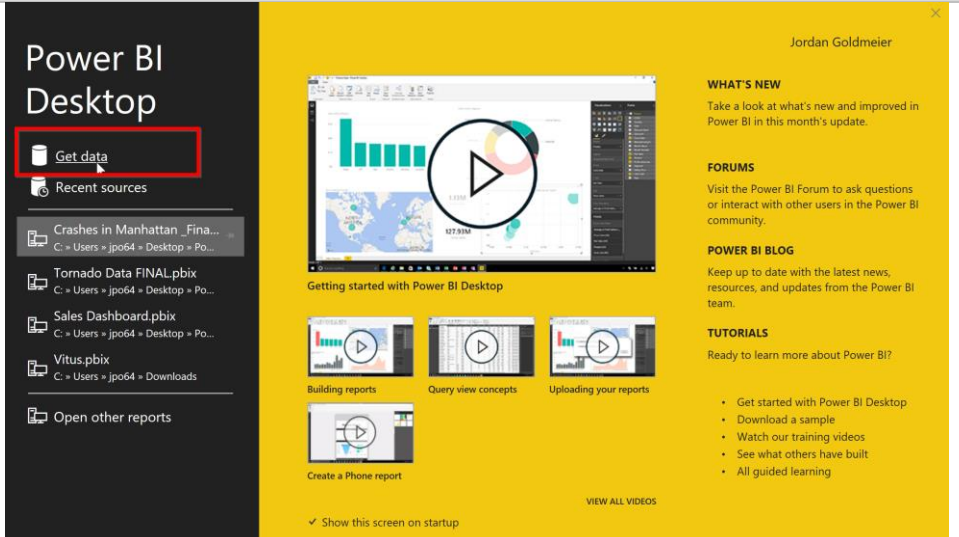
This lab is split into three parts:

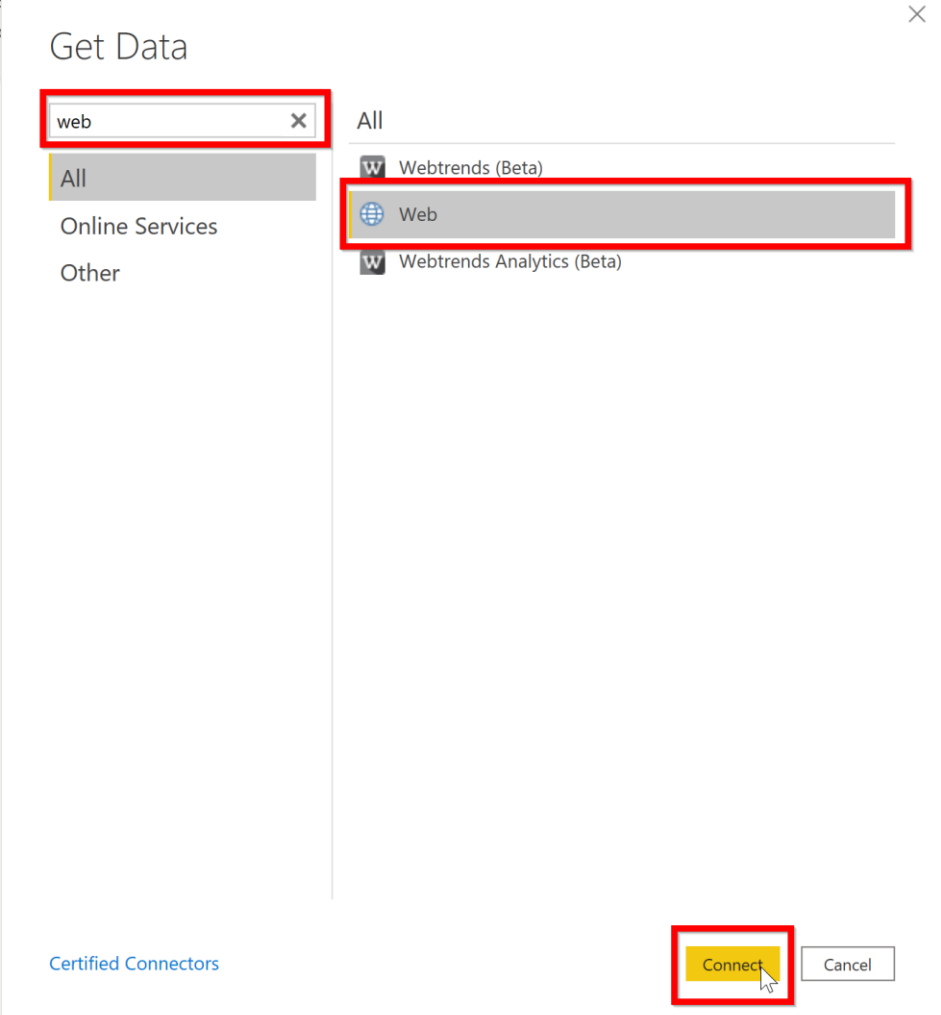
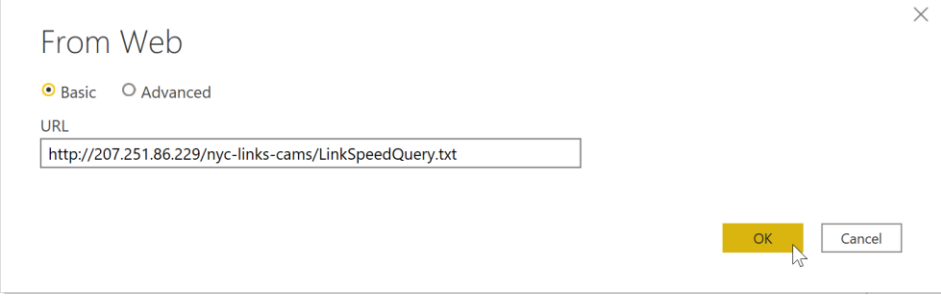
- ▶ Part 1: Connecting to the data source
- ▶ Part 2: Wrangling with Power Query
- ▶ Part 3: Creating the report

Links:

<https://data.cityofnewyork.us/> - <http://207.251.86.229/nyc-links-cams/LinkSpeedQuery.txt>

PART 1: CONNECTING TO THE DATA SOURCE

Click Steps	Screen Shots
<ol style="list-style-type: none">1. Open Power BI2. Select Get data on the left side of the screen	

Click Steps	Screen Shots
<ol style="list-style-type: none"> 3. Type "web" in the Get Data search box to search of the Web Connector 4. Click on the Web connector on the right side of the screen when it pops up. 5. Press Connect once selected. 	 <p>The screenshot shows the 'Get Data' window. A search box at the top left contains the text 'web'. Below the search box, there are three categories: 'All', 'Online Services', and 'Other'. Under the 'All' category, three connectors are listed: 'Webtrends (Beta)', 'Web', and 'Webtrends Analytics (Beta)'. The 'Web' connector is highlighted with a red box. At the bottom right, there is a yellow 'Connect' button, which is also highlighted with a red box, and a 'Cancel' button next to it.</p>
<ol style="list-style-type: none"> 6. We are going to connect to the NYC speed camera database. Use the following connection link: http://207.251.86.229/nyc-links-cams/LinkSpeedQuery.txt 7. Click OK 	 <p>The screenshot shows the 'From Web' window. At the top, there are two tabs: 'Basic' (selected) and 'Advanced'. Below the tabs, there is a 'URL' label and a text box containing the URL: 'http://207.251.86.229/nyc-links-cams/LinkSpeedQuery.txt'. At the bottom right, there is a yellow 'OK' button, which is highlighted with a red box, and a 'Cancel' button next to it.</p>

Click Steps

- Once the data is loaded, click Transform Data.

Note: this is real-time data, so your results might differ.

Screen Shots

<http://207.251.86.229/nyc-links-cams/LinkSpeedQuery.txt>

File Origin: 1252: Western European (Windows) | Delimiter: Tab | Data Type Detection: Based on first 200 rows

id	Speed	TravelTime	Status	DataAsOf	linkid	linkPoints	Enc
1	23.61	228	0	3/5/2020 6:24:12 PM	4616337	40.74047,-74.009251 40.74137,-74.00893 40.7431706,-...	jbtfwFx ubMsD_AglcA
106	26.09	86	0	3/5/2020 6:24:12 PM	4616323	40.77158,-73.994441 40.7713004,-73.99455 40.77085,-...	kezwFF sbMv@TxAVn
107	0	0	-101	1/1/1978 12:00:00 AM	4616279	40.64289,-74.21316 40.64275,-74.21222 40.64281,-74...	aaawFdwJcMzYDgMc
108	0	0	-101	1/1/1978 12:00:00 AM	4616280	40.62464,-74.17757 40.62345,-74.17679 40.62242,-74...	_oJvFxxvcMIF CIEk@b
110	47.84	159	0	3/5/2020 6:23:12 PM	4616281	40.5256,-74.23039 40.52551,-74.23214 40.52539,-74.2...	_dljF badMP IV'GXh
119	23.61	298	0	3/5/2020 6:24:03 PM	4456502	40.70631,-74.01501 40.705380,-74.01528 40.70496,-74...	mmmwFx'wbMxDt@r
122	0	0	-101	1/1/1978 12:00:00 AM	4616344	40.7073904,-74.01559 40.7078006,-74.01549 40.70853...	etmwFidwbMqASqCu
123	0	0	-101	1/1/1978 12:00:00 AM	4616345	40.70738,-74.01558 40.70669,-74.0159 40.706371,-74...	mlmwFr~vbMUKw@A
124	13.04	535	0	3/5/2020 6:24:03 PM	4456501	40.68036,-74.00441001 40.6822,-74.0057201 40.6837,-...	gkhwFp~tbMoldGkHn
126	16.15	462	0	3/5/2020 6:24:05 PM	4616247	40.8271606,-73.84993 40.82771,-73.84671 40.8284105...	w'exF~xvaMmBaskCS
129	52.19	88	0	3/5/2020 6:24:05 PM	4616246	40.8240706,-73.874311 40.8247,-73.86959 40.8251906...	mmxdFq aM Bm\ aBw
137	24.23	182	0	3/5/2020 6:24:05 PM	4616260	40.8242005,-73.874361 40.8249804,-73.868411 40.825...	gndxFvq aM Ced@Sd
140	26.09	108	0	3/5/2020 6:24:03 PM	4456479	40.79789,-73.91988 40.79771,-73.92004 40.79758,-73...	yl_xFndbMb@XkbT
141	44.11	158	0	3/5/2020 6:24:03 PM	4456478	40.772251,-73.918991 40.77391,-73.9222 40.77474,-73...	qizwFhndbMkillMeD'D
142	49.08	151	0	3/5/2020 6:24:05 PM	4616261	40.83037,-73.85062 40.82996,-73.849251 40.8294605...	ytexJ VaMaPqGgBgH
145	10.56	241	0	3/5/2020 6:24:12 PM	4616342	40.7081105,-73.99944 40.7084705,-73.99884 40.70868...	uxmwFn_tbMgAwBI@
148	18.01	827	0	3/5/2020 6:24:12 PM	4616257	40.69153,-73.99911 40.6922605,-73.99937 40.6928906...	aqjwF sbMqCr@ B?y
149	18.64	385	0	3/5/2020 6:24:12 PM	4616339	40.6916,-73.99915 40.69236,-73.99939 40.69272,-73.9...	oqjwFt sbMwCn@GAB
150	6.83	1349	0	3/5/2020 6:24:12 PM	4616340	40.7016405,-73.991121 40.70158,-73.99033 40.701440...	gplwFnrkMJ CZIEh@
153	21.12	334	0	3/5/2020 6:24:12 PM	4616235	40.69153,-73.99911 40.6922605,-73.99937 40.6928906...	sowFx_jbMwDgHuHg

Preview downloaded on Thursday, March 5, 2020

Load Transform Data Cancel

- Your data should now be loaded into Power Query. Your screen will look similar to the one to the right.

Untitled - Power Query Editor

Home | View | Tools | Help

Queries | LinkSpeedQuery

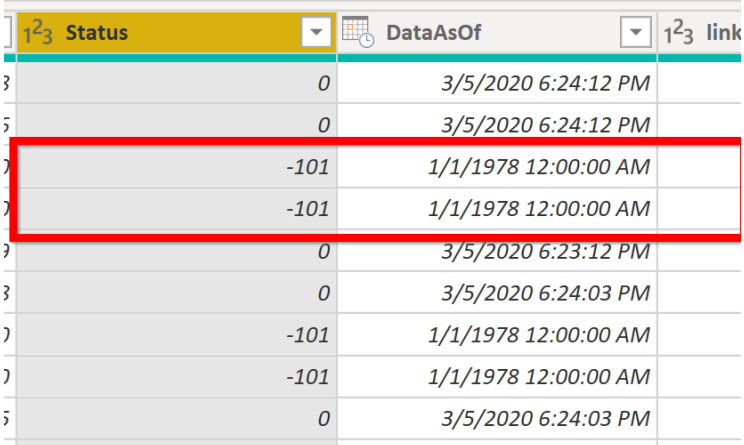
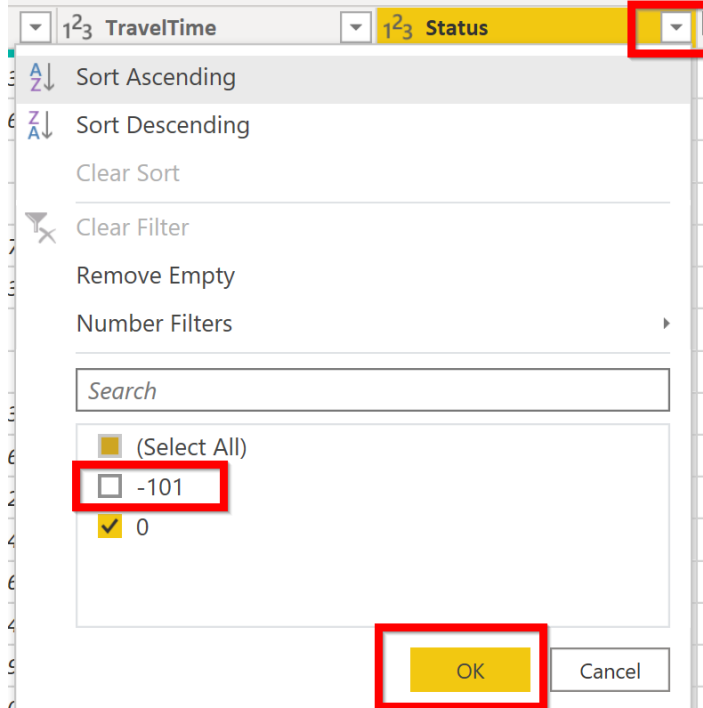
Table: TransformColumns(Promoted Headers), ({"id", Int64.Type), ("Speed", Type number), ("TravelTime", Int64.Type), ("Status", Int64.Type), ("DataAsOf", Int64.Type), ("linkid", Int64.Type), ("linkPoints", Int64.Type), ("Enc", Int64.Type)

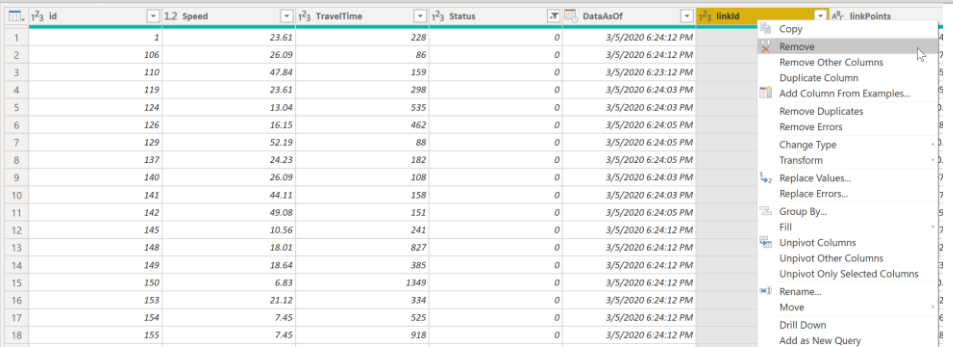
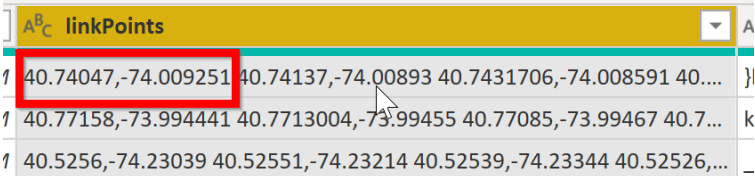
id	Speed	TravelTime	Status	DataAsOf	linkid	linkPoints	Enc
1	23.61	228	0	3/5/2020 6:24:12 PM	4616337	40.74047,-74.009251 40.74137,-74.00893 40.7431706,-...	jbtfwFx ubMsD_AglcA
2	26.09	86	0	3/5/2020 6:24:12 PM	4616323	40.77158,-73.994441 40.7713004,-73.99455 40.77085,-...	kezwFF sbMv@TxAVn
3	0	0	-101	1/1/1978 12:00:00 AM	4616279	40.64289,-74.21316 40.64275,-74.21222 40.64281,-74...	aaawFdwJcMzYDgMc
4	0	0	-101	1/1/1978 12:00:00 AM	4616280	40.62464,-74.17757 40.62345,-74.17679 40.62242,-74...	_oJvFxxvcMIF CIEk@b
5	47.84	159	0	3/5/2020 6:23:12 PM	4616281	40.5256,-74.23039 40.52551,-74.23214 40.52539,-74.2...	_dljF badMP IV'GXh
6	23.61	298	0	3/5/2020 6:24:03 PM	4456502	40.70631,-74.01501 40.705380,-74.01528 40.70496,-74...	mmmwFx'wbMxDt@r
7	0	0	-101	1/1/1978 12:00:00 AM	4616344	40.7073904,-74.01559 40.7078006,-74.01549 40.70853...	etmwFidwbMqASqCu
8	0	0	-101	1/1/1978 12:00:00 AM	4616345	40.70738,-74.01558 40.70669,-74.0159 40.706371,-74...	mlmwFr~vbMUKw@A
9	13.04	535	0	3/5/2020 6:24:03 PM	4456501	40.68036,-74.00441001 40.6822,-74.0057201 40.6837,-...	gkhwFp~tbMoldGkHn
10	16.15	462	0	3/5/2020 6:24:05 PM	4616247	40.8271606,-73.84993 40.82771,-73.84671 40.8284105...	w'exF~xvaMmBaskCS
11	52.19	88	0	3/5/2020 6:24:05 PM	4616246	40.8240706,-73.874311 40.8247,-73.86959 40.8251906...	mmxdFq aM Bm\ aBw
12	24.23	182	0	3/5/2020 6:24:05 PM	4616260	40.8242005,-73.874361 40.8249804,-73.868411 40.825...	gndxFvq aM Ced@Sd
13	26.09	108	0	3/5/2020 6:24:03 PM	4456479	40.79789,-73.91988 40.79771,-73.92004 40.79758,-73...	yl_xFndbMb@XkbT
14	44.11	158	0	3/5/2020 6:24:03 PM	4456478	40.772251,-73.918991 40.77391,-73.9222 40.77474,-73...	qizwFhndbMkillMeD'D
15	49.08	151	0	3/5/2020 6:24:05 PM	4616261	40.83037,-73.85062 40.82996,-73.849251 40.8294605...	ytexJ VaMaPqGgBgH
16	10.56	241	0	3/5/2020 6:24:12 PM	4616342	40.7081105,-73.99944 40.7084705,-73.99884 40.70868...	uxmwFn_tbMgAwBI@
17	18.01	827	0	3/5/2020 6:24:12 PM	4616257	40.69153,-73.99911 40.6922605,-73.99937 40.6928906...	aqjwF sbMqCr@ B?y
18	18.64	385	0	3/5/2020 6:24:12 PM	4616339	40.6916,-73.99915 40.69236,-73.99939 40.69272,-73.9...	oqjwFt sbMwCn@GAB
19	6.83	1349	0	3/5/2020 6:24:12 PM	4616340	40.7016405,-73.991121 40.70158,-73.99033 40.701440...	gplwFnrkMJ CZIEh@
20	21.12	334	0	3/5/2020 6:24:12 PM	4616235	40.69153,-73.99911 40.6922605,-73.99937 40.6928906...	sowFx_jbMwDgHuHg

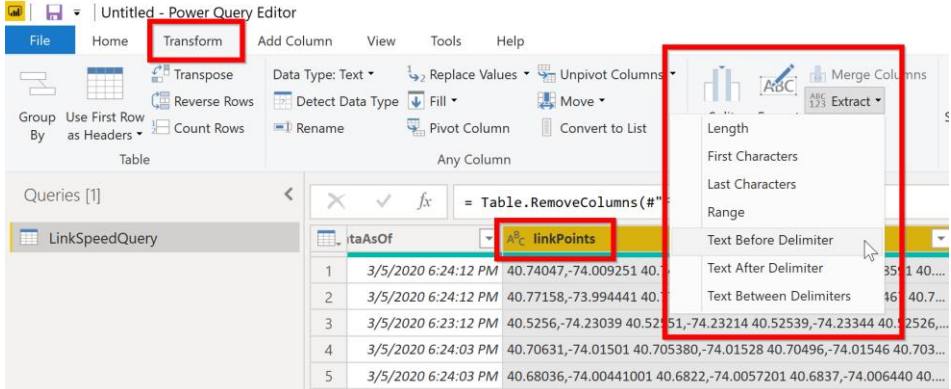
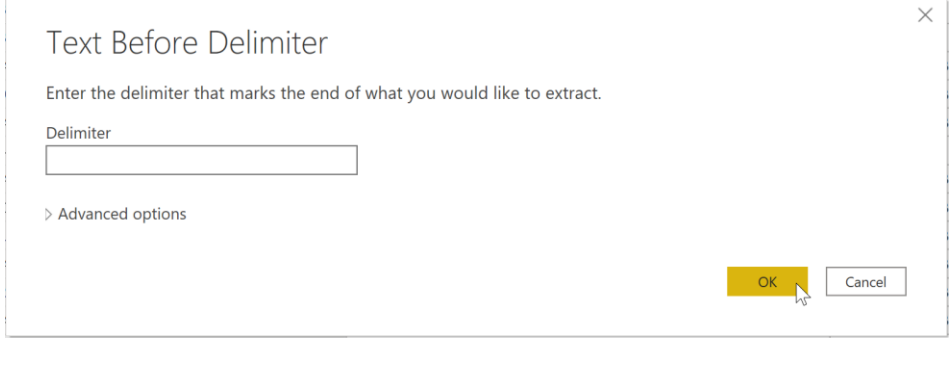
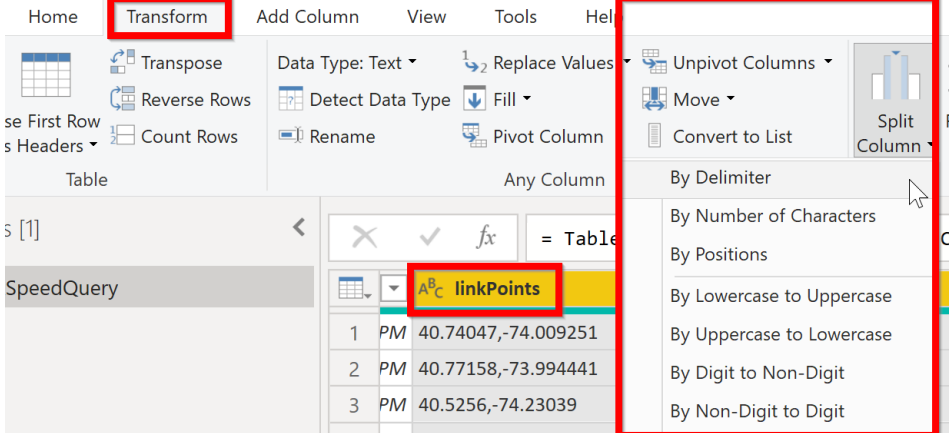
13 COLUMNS, 153 ROWS | Column profiling based on top 1000 rows

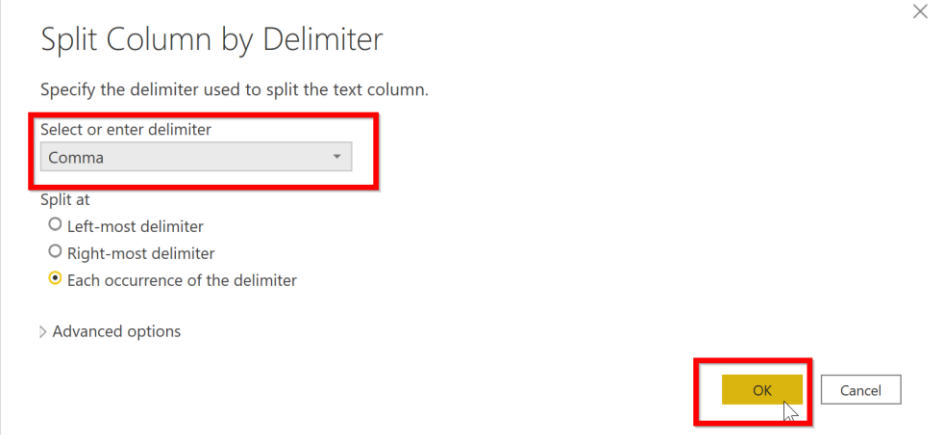
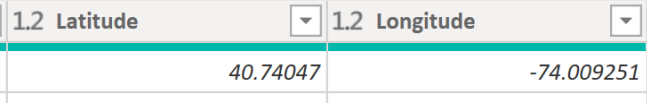
PREVIEW DOWNLOADED ON THURSDAY, MARCH 5, 2020

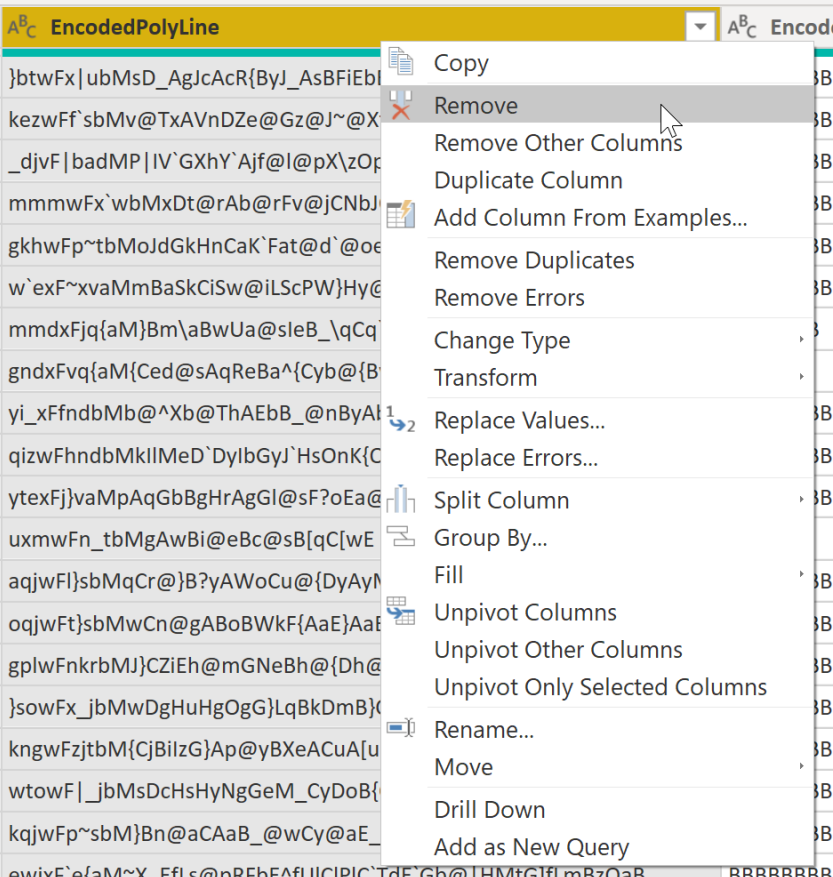
PART 2: WRANGLING WITH POWER QUERY

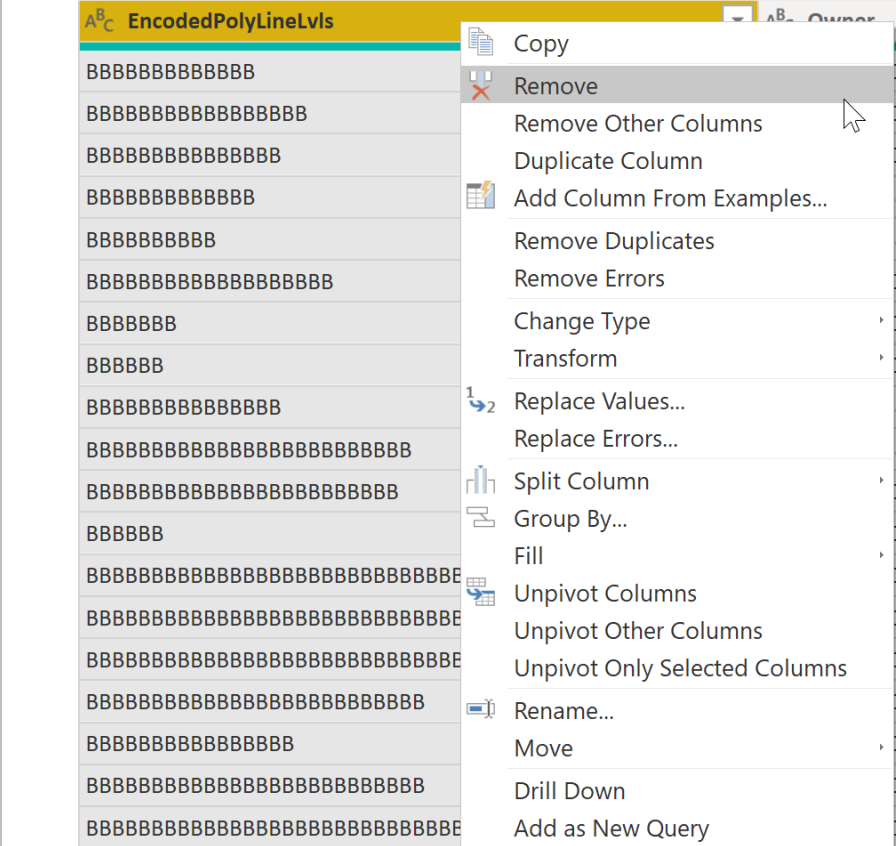
Click Steps	Screen Shots
<p>10. We'll start by using Power Query to clean up the dataset before we do analysis on it.</p> <p>11. Notice that Status contains some nonsense data.</p>	
<p>12. Let's filter out the negatives to removed the dirty data.</p> <p>Click on the dropdown to the right of status.</p> <p>13. Deselect anything that isn't a 0.</p> <p>14. Click OK.</p>	

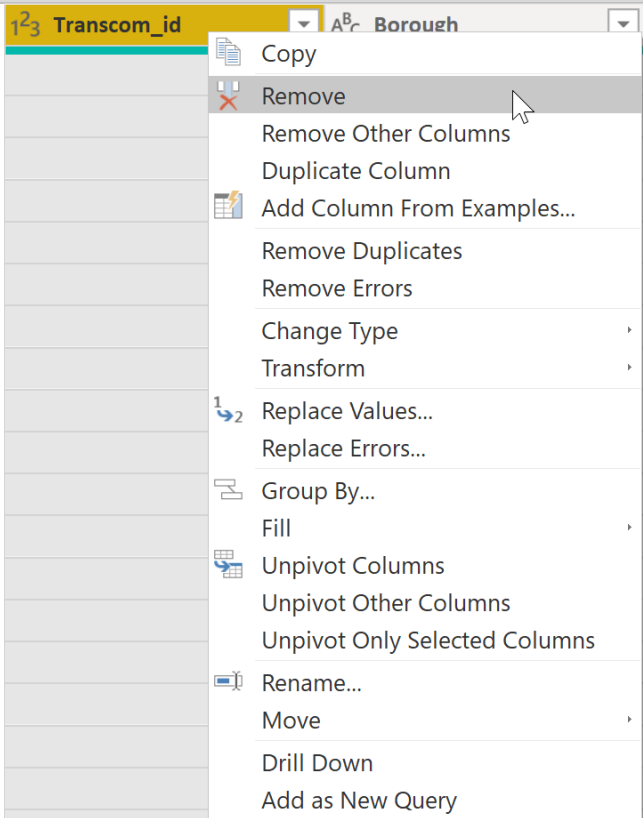
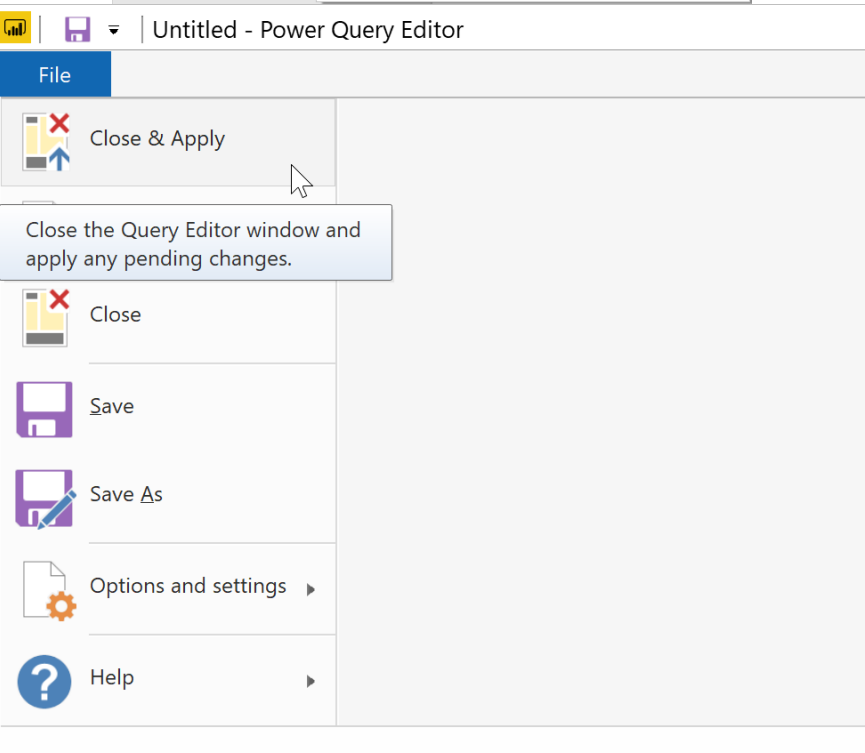
Click Steps	Screen Shots
<p>15. Let's remove the LinkID column. Right click on the LinkID column header.</p> <p>16. Click Remove.</p>	
<p>17. Now we need to create longitude and latitude columns. The linkPoints column contains multiple longitude and latitude points. This reflects the entire physical footprint (point by point) of the camera sensor but we don't really need all of that information. We're only interested in the first two points in the red highlight.</p> <p>Notice that each longitude and latitude point uses a comma to refer to one point. Multiple points are separated by a space.</p> <p>Again, we're only interested in the first set of pairwise points.</p>	

Click Steps	Screen Shots
<p>18. Let's extract only those points first. Click the linkPoints header field.</p> <p>19. Click the Transform tab.</p> <p>20. Click Extract.</p> <p>21. Click Text Before Delimiter.</p>	
<p>22. Type in a space [hit the space bar key] in the delimiter textbox. You probably can't see it in the image, but there is a space in the textbox.</p> <p>23. Click OK. Notice that the other geo points have been removed.</p>	
<p>24. Now we have to separate the latitude and longitude points. To make the separation, click the on the linkPoints column.</p> <p>25. Click Transform</p> <p>26. Click the Split Column button.</p> <p>27. Click By Delimeter.</p>	

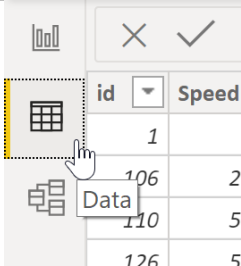
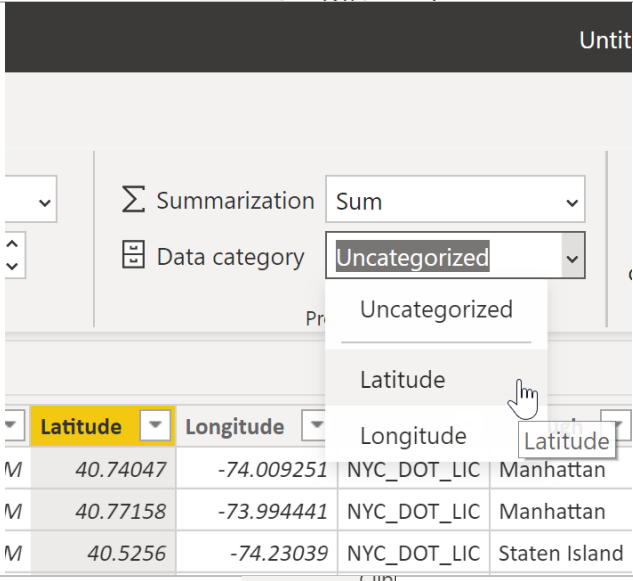
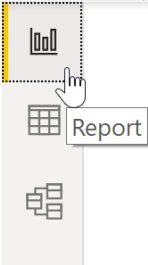
Click Steps	Screen Shots				
<p>28. From the dropdown select Comma.</p> <p>29. Click OK.</p>					
<p>30. The new columns will be named LinkPoints.1 and LinkPoints.2. Rename them to latitude and longitude respectively.</p>	 <table border="1"> <thead> <tr> <th>1.2 Latitude</th><th>1.2 Longitude</th></tr> </thead> <tbody> <tr> <td>40.74047</td><td>-74.009251</td></tr> </tbody> </table>	1.2 Latitude	1.2 Longitude	40.74047	-74.009251
1.2 Latitude	1.2 Longitude				
40.74047	-74.009251				

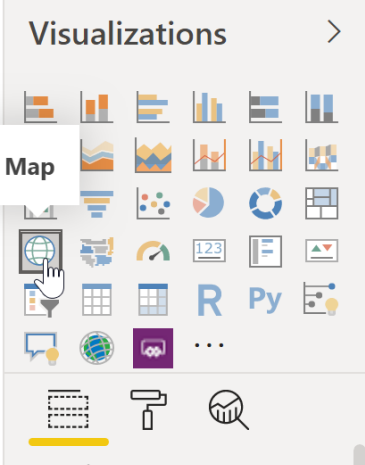
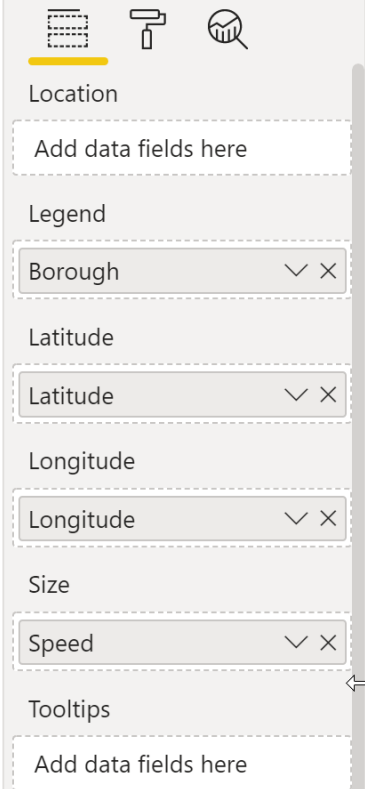
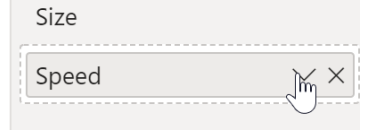
Click Steps	Screen Shots
<p>31. Let's remove more columns we don't need. Right click on the EncodedPolyLine column header. Click Remove.</p>	

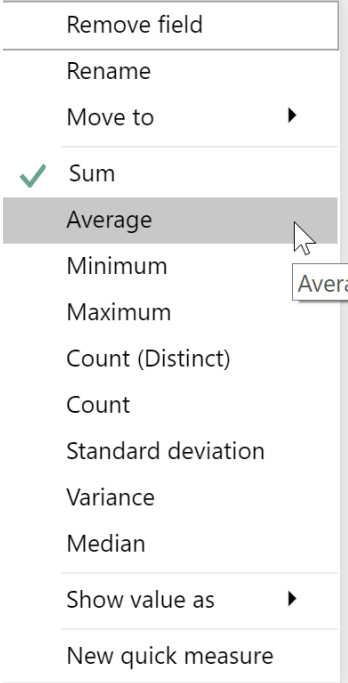
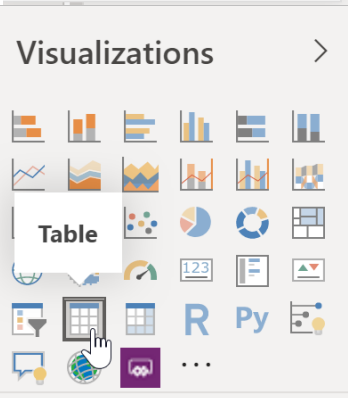

Click Steps	Screen Shots
32. Let's remove more columns we don't need. Right click on the EncodedPolyLineLvis column header. Click Remove.	

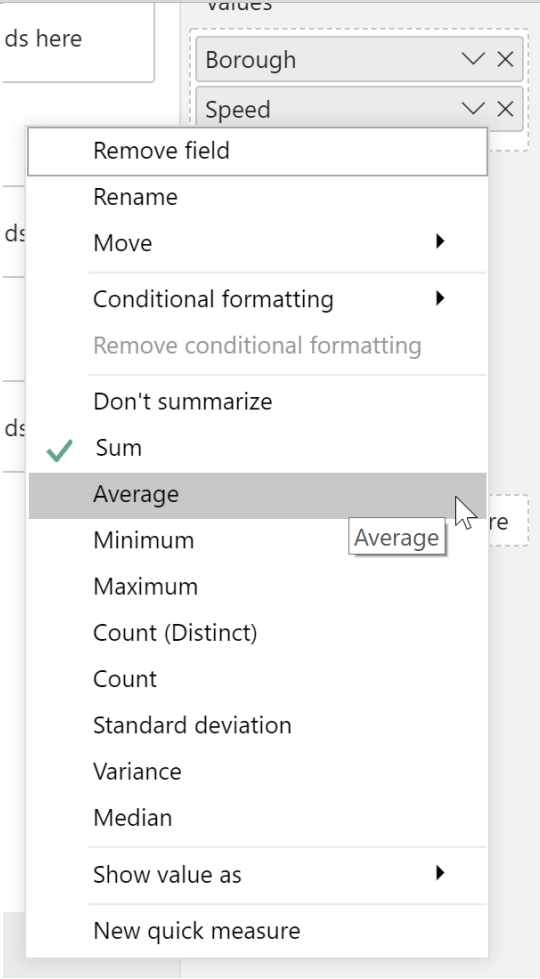
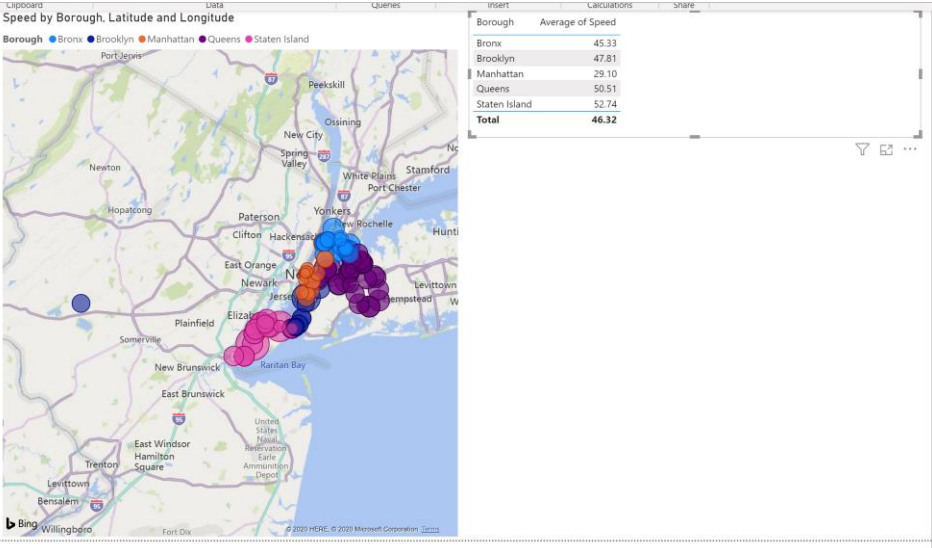
Click Steps	Screen Shots
<p>33. Remove Transcom_id. We don't need it. Right click the column header. Select Remove.</p>	 <p>The screenshot shows the Power Query Editor interface. A column named 'Transcom_id' is selected. A right-click context menu is open, displaying various options. The 'Remove' option, indicated by a red 'X' icon, is highlighted by the mouse cursor. Other visible options include 'Copy', 'Remove Other Columns', 'Duplicate Column', 'Add Column From Examples...', 'Remove Duplicates', 'Remove Errors', 'Change Type', 'Transform', 'Replace Values...', 'Replace Errors...', 'Group By...', 'Fill', 'Unpivot Columns', 'Unpivot Other Columns', 'Unpivot Only Selected Columns', 'Rename...', 'Move', 'Drill Down', and 'Add as New Query'.</p>
<p>34. We're finished. Now Click Close & Apply.</p>	 <p>The screenshot shows the 'File' menu in the Power Query Editor. The 'Close & Apply' option, represented by a document icon with a red 'X' and a blue arrow, is selected. A tooltip is displayed over this option, stating: 'Close the Query Editor window and apply any pending changes.' Other menu items visible include 'Close', 'Save', 'Save As', 'Options and settings', and 'Help'.</p>

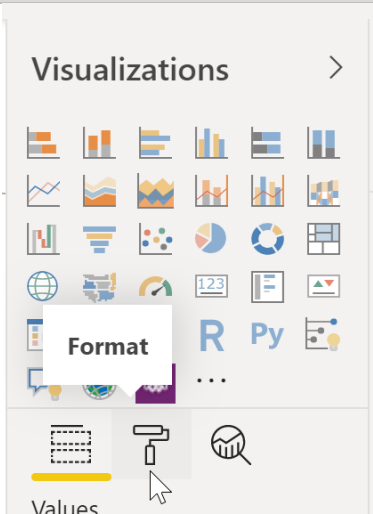
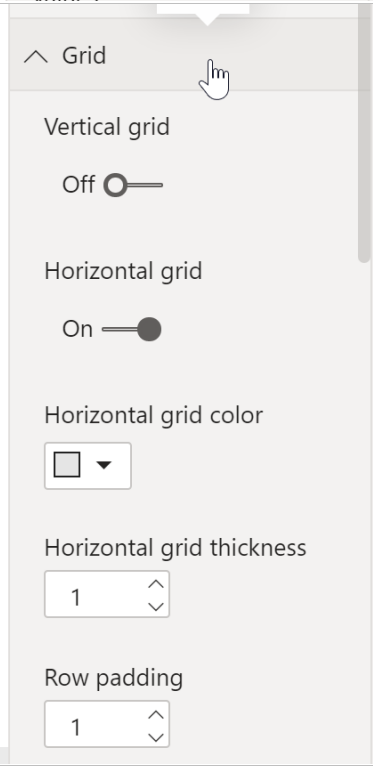
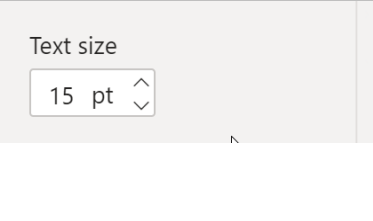
PART 3: CREATING THE REPORT

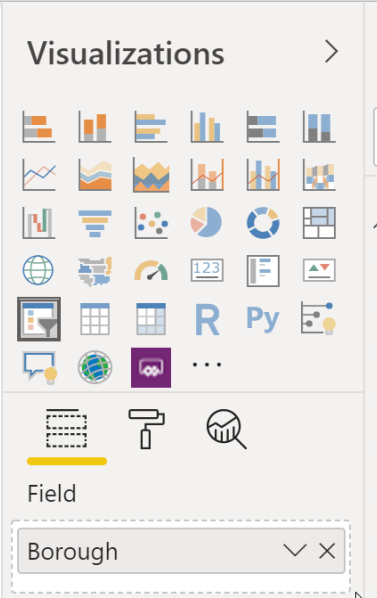
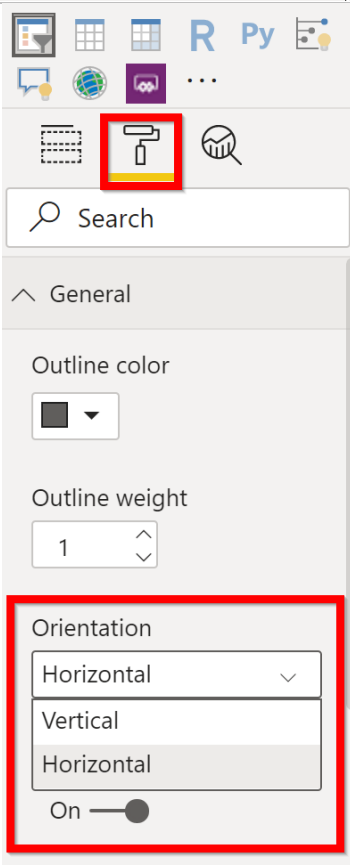
Click Steps	Screen Shots
1. When Power Query closes click on the Data icon to go to the data view.	
2. Select the latitude column header. Select the Data Category dropdown latitude. This will tell Power BI to treat it as a latitude data point. 3. Click on the Longitude header and do the same (but click the Longitude item).	
4. Click the Report icon to move to the Report view.	

Click Steps	Screen Shots
<p>5. Click the Map button to insert a map in the upper left.</p>	
<p>6. Place the Latitude column in the Latitude field well.</p> <p>7. Place the Longitude column in the Longitude field well.</p> <p>8. Place the Borough column into the Legend field well.</p> <p>9. Place the Speed column into the Size field well.</p>	
<p>10. Click the down arrow next to the Speed field in the Size field well.</p>	

Click Steps	Screen Shots
<p>11. Select Average.</p>	 <p>A screenshot of the Power BI context menu. The menu is open, showing various aggregation options. The 'Average' option is highlighted with a mouse cursor. Other options include 'Remove field', 'Rename', 'Move to', 'Sum' (with a green checkmark), 'Minimum', 'Maximum', 'Count (Distinct)', 'Count', 'Standard deviation', 'Variance', 'Median', 'Show value as', and 'New quick measure'.</p>
<p>12. Now let's create a table to show the average speed.</p> <p>13. Click the Table button to insert a new table.</p>	 <p>A screenshot of the Power BI Visualizations pane. The 'Table' visualization is selected, indicated by a white box and a hand cursor. The pane shows various other visualization options like bar charts, line charts, pie charts, and maps.</p>
<p>14. Add Borough and Speed to the Values field well.</p>	 <p>A screenshot of the Power BI Values field well. The field well is titled 'Values' and contains two items: 'Borough' and 'Speed'. Each item has a dropdown arrow and a close button (X) to its right.</p>

Click Steps	Screen Shots														
<p>15. Click on the down arrow next to speed and select Average.</p>															
<p>16. Resize the table so that it appears like the one in the picture.</p>	 <table border="1"> <thead> <tr> <th>Borough</th> <th>Average of Speed</th> </tr> </thead> <tbody> <tr> <td>Bronx</td> <td>45.33</td> </tr> <tr> <td>Brooklyn</td> <td>47.81</td> </tr> <tr> <td>Manhattan</td> <td>29.10</td> </tr> <tr> <td>Queens</td> <td>50.51</td> </tr> <tr> <td>Staten Island</td> <td>52.74</td> </tr> <tr> <td>Total</td> <td>46.32</td> </tr> </tbody> </table>	Borough	Average of Speed	Bronx	45.33	Brooklyn	47.81	Manhattan	29.10	Queens	50.51	Staten Island	52.74	Total	46.32
Borough	Average of Speed														
Bronx	45.33														
Brooklyn	47.81														
Manhattan	29.10														
Queens	50.51														
Staten Island	52.74														
Total	46.32														

Click Steps	Screen Shots
<p>Let's format the table. With the table selected, Click the Format paintbrush button.</p>	
<p>17. Click the Grid dropdown. And then scroll down.</p>	
<p>18. Change the Text size to 15pt. You might need to resize the final table to make it look how you'd like it.</p>	

Click Steps	Screen Shots
<p>19. Let's add a slicer. Click the Slicer icon.</p> <p>20. Add Borough to the field well. This will add a slicer checklist. Let's turn it into a the chicklet type of slicer.</p>	
<p>21. With the slicer selected click the Format button</p> <p>22. Press the General button to open the options.</p> <p>23. From the Orientation dropdown select Horizontal.</p>	

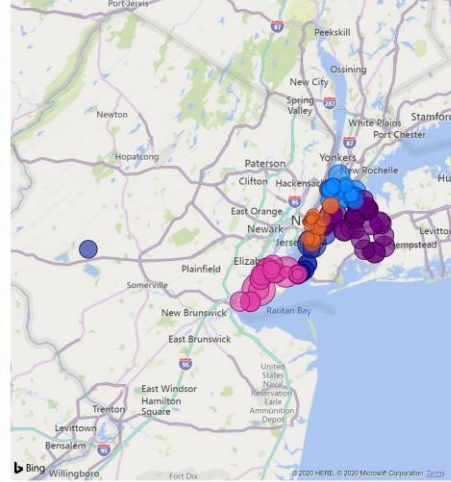
Click Steps

24. If you did everything right, your report should look like the screenshot.

Screen Shots

Speed by Borough, Latitude and Longitude

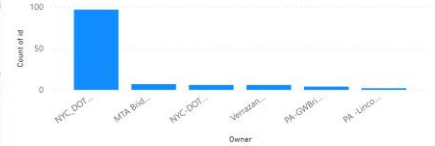
Borough ● Bronx ● Brooklyn ● Manhattan ● Queens ● Staten Island



Borough Average of Speed

Borough	Average of Speed
Bronx	45.33
Brooklyn	47.81
Manhattan	29.10
Queens	50.51
Staten Island	52.74
Total	46.32

Count of id by Owner



Borough

Borough	Queens
Staten Island	
Manhattan	