Section 1: Theory

1. **What is a Data Moat? Why is it important to have one?**
   The data moat describes the competitive advantage that a company holds because of combining proprietary data sources with others, and building or improving services using automated decision making, fueled by AI and data. It is a modern extension of traditional economic moats, such as economies of scale, high switching costs, network effects, and intellectual property. It's important to have one because a strong data moat inhibits competitors from stealing market share.

2. **What is the difference between OLAP and OLTP Databases? Why would you choose one over the other?**
   OLAP (On-line Analytical Processing) Database is characterized by relatively low volume of transactions and it is widely used by Data Mining techniques. Aggregated and historical data in OLAP Database are stored in multi-dimensional schema. Records are usually short and wide.
   OLTP (On-line Transaction Processing) Database is characterized by a large number of short on-line transactions and main emphasis is putting on very fast query processing. Detailed and current data in transactional database are modeled using an entity-relationship model. Records are usually narrow and long.
   We usually choose one over the other because they serve different purposes.

3. **What are the 3 different roles in a modern data team? Which problems do each of them solve? How do they compare with each other?**
   1) Data Engineer; 2) Data Analyst; 3) Data Scientist
   Data engineer (ETL engineer) is responsible for moving and propagating access to data. Their chief mandate is piping the data to the right places in the most efficient way.
   Data analyst focuses on answering business questions by using data. This person knows SQL and may be comfortable in a few other languages such as Python, or R, and he/she effectively serves as the bridge between data and business insights.
   Data scientist builds predictive models and automated classifications of the existing data to help guide future decisions and predict outcomes.

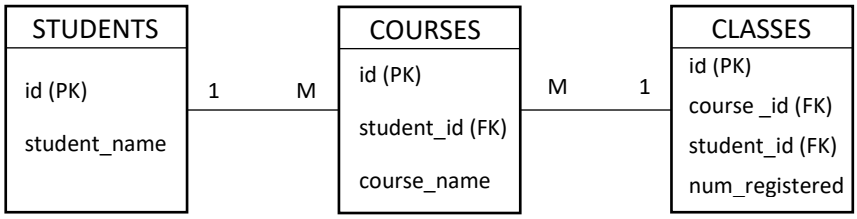
4. **What is the difference between the WHERE and HAVING clauses?**
   The WHERE clause is used to filter rows before the grouping is performed, whereas the HAVING clause is used to filter rows after the grouping is performed. The HAVING clause often includes the result of aggregate functions and is used with GROUP BY.

5. **How would you define the relationship between employees and offices in the Entity Relationship (ER) model? Please provide an explanation why using real world examples.**
   Normally, 1 employee can work in only 1 office, so this relationship is one-to-one (1:1) for employees. 1 office usually has many employees, so this relationship is one-to-many (1:M) for offices. However, there are also some special cases. For example, in a multi-national enterprise, 1 employee can also work in many offices and 1 office can hold many employees, so the relationship between employees and offices can also be many-to-many (M:M).

Section 2: Database Design

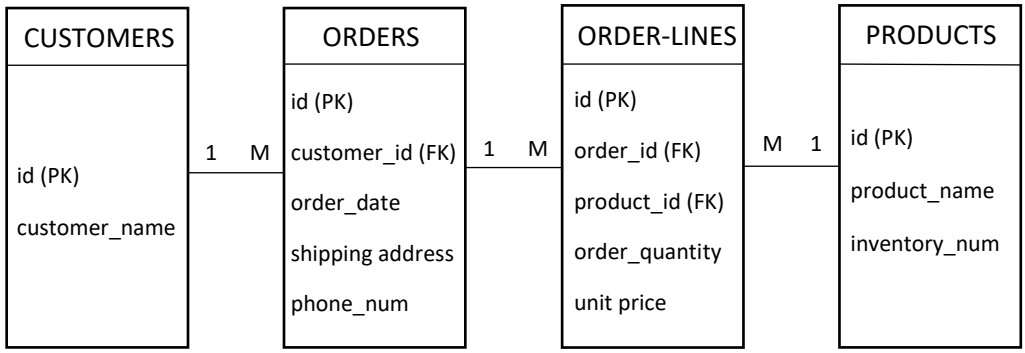
1. **You are asked to model the many to many relationship between students and classes in a relational database.**
    - **What changes do you need to make to support this relationship?**
      Add “COURSES” entity as the bridge between “STUDENTS” and “CLASSES”.

    - **Please create an ER diagram to show how these entities will relate to each after your changes.**
      STUDENTS M:M CLASSES

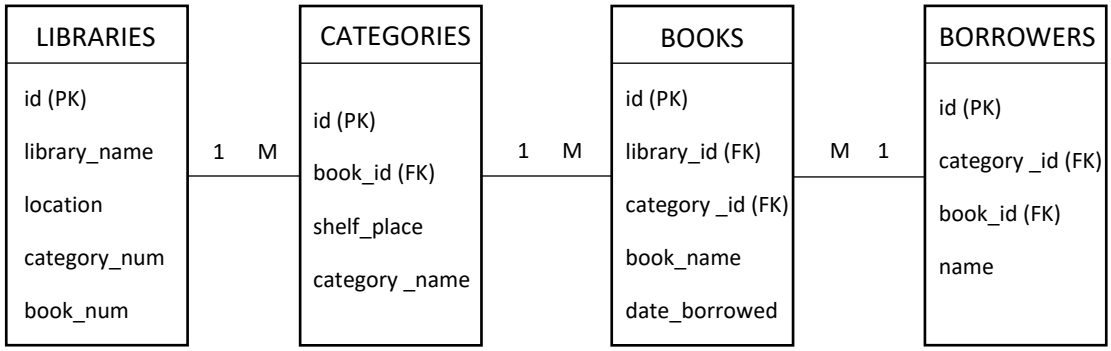


2. **You are asked to model the many to many relationship between customers and products in a relational database.**
    - **What changes do you need to make to support this relationship?**
      Add “ORDERS” and “ORDER-LINES” as the bridge between “CUSTOMERS” and “PRODUCTS”.

    - **Please create an ER diagram to show how these entities will relate to each after your changes.**
      CUSTOMERS M:M PRODUCTS



3. Design an ER diagram for a library reservation system for a family of libraries based on the given characteristics.
    - This system is for multiple libraries
    - This system is for multiple borrowers
    - There are multiple types of content that can be borrowed
    - Borrowers can borrow multiple items at the same time
    - Borrowers can borrow multiple types of content



Section 3: Data Analysis with SQL

In this section, you're going to use the data set **new_york_citibike** (under bigquery-public-data) in Google BigQuery to answer some business questions using SQL. Take some time to familiarize yourself with the data set before answering your questions.

Your output will be a 1-page report, which diagnoses the problems you see, provides a few potential resolutions, and recommends one solution with a justification of why. The report must fit on one page.

You will also submit an appendix, which includes all the SQL you ran to get to your answer and any tables, maps, or charts you think are helpful to make your point. Please add a comment on top of each figure in your appendix to explain what insight it is providing.

Business Questions

You've been told by customer support that customers frequently complain about bike stations being empty. You need to analyze the data in your data set to understand this problem and make suggestions about how to address it. Some items to consider are below. Please note that the questions below are just a guiding point for your analysis. You don't need to explicitly answer them all:

- Can you find any traces of empty stations?
  - If yes, how big is this problem?
- What are the most popular stations in the network?
  - When does their usage peak?
- What are the most popular trips in the network?
- Are there differences in the types of rides that people take?
- Is there a pattern in the types of stations that are empty?

Report

Problem Diagnosis

1. The scarcity of bike distribution are lying in the Upper East Side of Manhattan and Brooklyn, which are the residential area. *(See Figure 2)* Moreover, the most frequent users of citibikes in these empty stations are subscribers. *(See Figure 12)* Therefore, we can assume that the scarcity of bike distribution in these empty stations will most probably affect the workers to get to work in the Lower Manhattan and Midtown Manhattan areas, which are financial district and central business district (CBD).

2. The most popular stations are located in Midtown Manhattan area *(See Figure 4 &5)*, and the most popular trips are occurred in Midtown Manhattan and Lower Manhattan areas. *(See Figure 8 &9)* The most frequent users of popular stations and trips are mostly customers *(See Figure 10 & 11)*, and the usage peak are during Friday and weekends.

Potential Resolutions

1. As there are clear patterns in subscribers (annual member —- workers/residents) and customers (24-hour pass or 7-day pass user —- tourists) —- subscribers usually use the bikes during the rush hours on weekdays, whereas customers' demand of bikes is usually during non-rush hours and weekends, we can distribute more bikes in Upper East Side and Brooklyn areas during the morning and evening rush hours on weekdays, and place more bikes in the Midtown Manhattan and Lower Manhattan areas during the rest of the hours on weekdays and weekends.

2. Encourage the users to download citibike app and use the function of "find the nearest bike or dock" by giving out incentives, like point system —- redeem points for gift cards, merchandises, or even cash back).

One Recommendation

New York City is a highly populated city with a large number of residents and tourists. In order to satisfy both groups of people, so peak shift distribution strategy is the most ideal way to solve the problem of empty stations. Apart from this, the regular maintenance should also be paid attention to. Besides, as subway stations in NYC are not too far from each other, people can actually be encouraged to find the nearest available bikes by themselves. *(See Figure 6)* Large stations can contain more docks and bikes available for rental.

Appendix

1. Empty stations:

[Input]

```
SELECT station_id, name, ST_GeogPoint(longitude, latitude)  AS WKT, capacity, num_bikes_available, num_docks_available

FROM `bigquery-public-data.new_york_citibike.citibike_stations`

WHERE num_bikes_available = 0
```

[Output]

Comment: There are 11 empty stations. Among all, E 81 St & 3 Ave has the largest capacity.

Row	station_id	name	WKT	capacity	num_bikes_available	num_docks_available
1	3903	Jefferson Ave & Evergreen Ave	POINT(-73.91458 40.68989)	17	0	17
2	3913	Sands St Gate	POINT(-73.979827 40.699569)	19	0	17
3	3196	Riverview Park	POINT(-74.0439909 40.7443187)	18	0	18
4	3755	DeKalb Ave & Franklin Ave	POINT(-73.9574617892504 40.6906480708294)	19	0	18
5	3833	Madison St & Evergreen Ave	POINT(-73.91693 40.69122)	18	0	18
6	3557	40 Ave & 9 St	POINT(-73.945133 40.75742)	19	0	19
7	3878	Cypress Ave & George St	POINT(-73.90271 40.69812)	19	0	19
8	3884	Centre St & Seneca Ave	POINT(-73.9022 40.69969)	23	0	23
9	3156	E 72 St & York Ave	POINT(-73.95348296 40.76663814)	37	0	37
10	3286	E 89 St & 3 Ave	POINT(-73.9521667 40.7806284)	39	0	39
11	3146	E 81 St & 3 Ave	POINT(-73.9567526 40.77573034)	47	0	47

Figure 1: Table of Empty Stations

[Bigquery Geo Viz]

Comment: Most empty stations are located in Upper East Side of Manhattan and Brooklyn areas, which are mostly residential areas. Besides, stations in Upper East Side have more capacity. (The size of red dots stands for capacity)

Figure 2: Map of Empty Stations

2. Most Popular Stations:

[Input]

```
SELECT starttime, stoptime, start_station_name, end_station_name, num_bikes_available, COUNT(*) AS num_trips,

ST_GeogPoint(start_station_longitude, start_station_latitude)  AS start_WKT,

ST_GeogPoint(end_station_longitude, end_station_latitude)  AS end_WKT,

EXTRACT (DAYOFWEEK FROM starttime) AS day,

EXTRACT (HOUR FROM starttime) AS hour,

FROM `bigquery-public-data.new_york_citibike.citibike_trips` AS t

LEFT JOIN `bigquery-public-data.new_york_citibike.citibike_stations` AS s

ON s.station_id = t.start_station_id

WHERE num_bikes_available > 40

GROUP BY starttime, stoptime, start_station_name, end_station_name, num_bikes_available, start_station_longitude, start_station_latitude, end_station_longitude, end_station_latitude

ORDER BY num_trips DESC

LIMIT 10
```

[Output]

Comment: The most popular start stations are 8 Ave & W 33 St, Perishing Square North, Broadway & W 55 St, and Grand Army Plaze & Central Park S, and the most popular end stations are Broadway & W 49 St, Grand Army Plaza & Central Park S, 12 Ave & W 40 St, 9 Ave & W 22 St, etc. The usage peak is mostly during Friday and weekends.

Row	starttime	stoptime	start_station_name	end_station_name	num_bikes_available	num_trips	start_WKT	end_WKT	day	hour
1	2015-06-26T21:35:00	2015-06-26T21:48:00	8 Ave & W 33 St	Broadway & W 49 St	42	6	POINT(-73.993934 40.751551)	POINT(-73.98442659 40.76064679)	6	21
2	2015-06-28T12:30:00	2015-06-28T12:51:00	Pershing Square North	Grand Army Plaza & Central Park S	42	5	POINT(-73.977706 40.751873)	POINT(-73.97371465 40.7643971)	1	12
3	2015-06-30T09:10:00	2015-06-30T09:28:00	Broadway & W 55 St	12 Ave & W 40 St	48	5	POINT(-73.98192338 40.7652654)	POINT(-74.00277668 40.76087502)	3	9
4	2015-06-21T17:44:00	2015-06-21T17:53:00	8 Ave & W 33 St	9 Ave & W 22 St	42	5	POINT(-73.993934 40.751551)	POINT(-74.00197139 40.7454973)	1	17
5	2015-06-27T12:29:00	2015-06-27T13:10:00	Grand Army Plaza & Central Park S	Grand Army Plaza & Central Park S	44	4	POINT(-73.97371465 40.7643971)	POINT(-73.97371465 40.7643971)	7	12
6	2015-06-28T13:58:00	2015-06-28T14:20:00	Grand Army Plaza & Central Park S	Broadway & W 32 St	44	4	POINT(-73.97371465 40.7643971)	POINT(-73.98808416 40.74854862)	1	13
7	2015-06-12T16:09:00	2015-06-12T16:27:00	Grand Army Plaza & Central Park S	E 45 St & 3 Ave	44	4	POINT(-73.97371465 40.7643971)	POINT(-73.97282625 40.75255434)	6	16
8	2015-06-20T13:21:00	2015-06-20T13:37:00	Broadway & W 55 St	W 45 St & 6 Ave	48	4	POINT(-73.98192338 40.7652654)	POINT(-73.98291153 40.7568001)	7	13
9	2015-06-28T12:59:00	2015-06-28T13:54:00	Grand Army Plaza & Central Park S	Grand Army Plaza & Central Park S	44	4	POINT(-73.97371465 40.7643971)	POINT(-73.97371465 40.7643971)	1	12
10	2015-06-12T21:51:00	2015-06-12T22:05:00	8 Ave & W 33 St	Broadway & W 60 St	42	4	POINT(-73.993934 40.751551)	POINT(-73.98191841 40.76915505)	6	21

Figure 3: Table of Popular Stations

[Bigquery Geo Viz]

Comment: The most popular start and end stations are all lying in the Midtown Manhattan area. (The size of red dots stands for number of bikes available)

Figure 4: Map of Start Stations of Popular Stations

Figure 5: Map of End Stations of Popular Stations

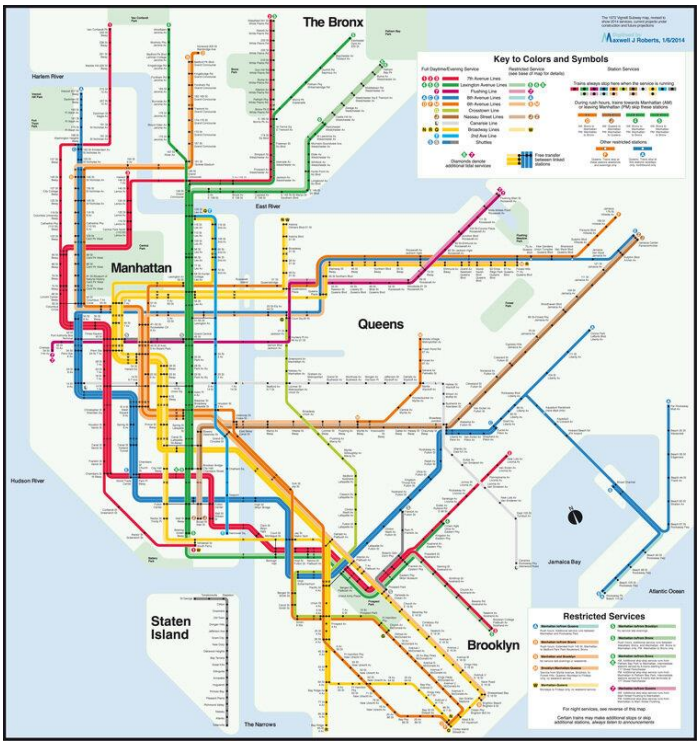


Figure 6: New York Subway Lines

Comment: The busiest part of NYC subway lines is mostly distributed in the Lower Manhattan and Midtown Manhattan, which are financial district and central business district (CBD).

3. Most Popular Trips:

[Input]

```
SELECT starttime, stoptime, start_station_name, end_station_name, count(*) as num_trips,
avg(tripduration)/60 as avg_trip_duration,

ST_GeogPoint(start_station_longitude, start_station_latitude)  AS start_WKT,

ST_GeogPoint(end_station_longitude, end_station_latitude)  AS end_WKT,
```

EXTRACT (DAYOFWEEK FROM starttime) AS day,

EXTRACT (HOUR FROM starttime) AS hour,

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

GROUP BY starttime, stoptime, start_station_name, end_station_name, start_station_longitude, start_station_latitude, end_station_longitude, end_station_latitude

HAVING num_trips > 4

ORDER BY num_trips DESC

LIMIT 10

[Output]

Comment: The most popular trip is where the red oval is highlighted in the table below.

Row	starttime	stoptime	start_station_name	end_station_name	num_trips	avg_trip_duration	start_WKT	end_WKT	day	hour
1	null	null			5828994	null	null	null	null	null
2	2015-06-05T15:53:00	2015-06-05T16:02:0	Barrow St & Hudson St	Sullivan St & Washington Sq	7	8.945238095238096	POINT(-74.00674436 40.73172428)	POINT(-73.99906065 40.73047747)	6	15
3	2015-06-22T12:06:00	2015-06-22T12:35:00	Broad St & Bridge St	Christopher St & Greenwich St	6	28.877777777777776	POINT(-74.01167797 40.70365182)	POINT(-74.00711384 40.73291553)	2	12
4	2015-06-08T21:30:00	2015-06-08T21:54:00	E 7 St & Avenue A	Peck Slip & Front St	6	24.17777777777778	POINT(-73.98379855 40.72621788)	POINT(-74.00167 40.707873)	2	21
5	2015-06-20T18:52:00	2015-06-20T19:15:00	Lafayette St & E 8 St	E 20 St & Park Ave	6	22.680555555555554	POINT(-73.9907647 40.73028666)	POINT(-73.98751968 40.73827428)	7	18
6	2015-06-26T21:35:00	2015-06-26T21:48:00	8 Ave & W 33 St	Broadway & W 49 St	6	13.213888888888887	POINT(-73.993934 40.751551)	POINT(-73.98442659 40.76064679)	6	21
7	2015-06-07T15:27:00	2015-06-07T15:36:00	Vesey Pl & River Terrace	Little West St & 1 Pl	6	9.183333333333334	POINT(-74.01658354 40.7153379)	POINT(-74.01677685 40.70569254)	1	15
8	2015-06-20T16:47:00	2015-06-20T17:03:00	11 Ave & W 27 St	W 4 St & 7 Ave S	6	15.96111111111111	POINT(-74.005226 40.751396)	POINT(-74.00293877 40.73401143)	7	16
9	2015-06-21T18:40:00	2015-06-21T18:57:00	Central Park S & 6 Ave	Grand Army Plaza & Central Park S	6	17.269444444444442	POINT(-73.97634151 40.76590936)	POINT(-73.97371465 40.7643971)	1	18
10	2015-06-21T13:37:00	2015-06-21T13:50:00	Cumberland St & Lafayette Ave	DeKalb Ave & Skillman St	6	13.197222222222223	POINT(-73.97265183 40.68753406)	POINT(-73.95643107 40.6906495)	1	13

Figure 7: Table of Popular Trips

[Bigquery Geo Viz]

Comment: The most popular trip networks are lying in the Lower Manhattan and Midtown Manhattan, as they are the busiest area in NYC – financial district and CBD. (The size of red dots stands for number of trips)

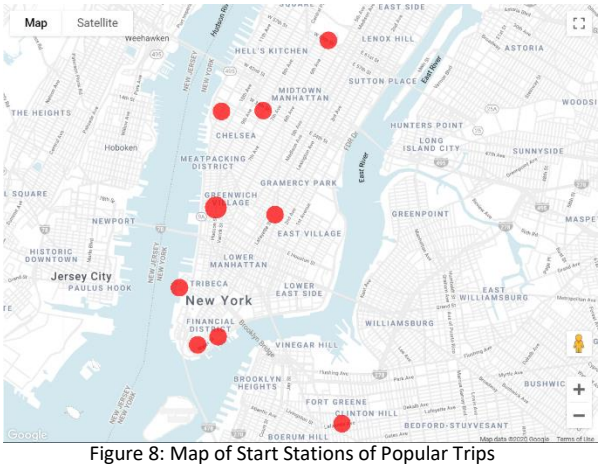


Figure 8: Map of Start Stations of Popular Trips

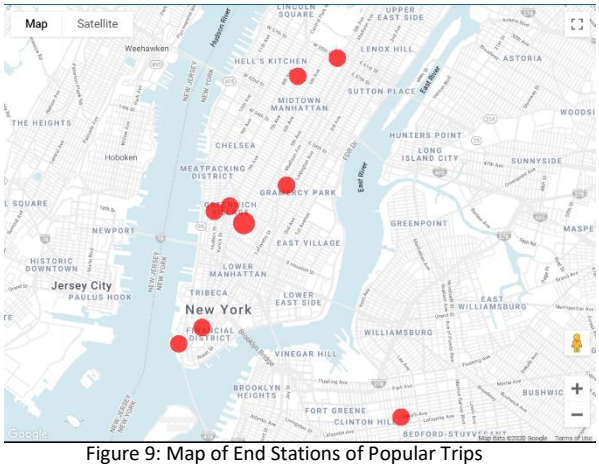


Figure 9: Map of End Stations of Popular Trips

4. Differences in the User Types

1) Most Popular Stations

[Input]

```
SELECT starttime, stoptime, start_station_name, end_station_name, usertype, num_bikes_available,
COUNT(*) AS num_trips

FROM `bigquery-public-data.new_york_citibike.citibike_trips` AS t

LEFT JOIN `bigquery-public-data.new_york_citibike.citibike_stations` AS s

ON s.station_id = t.start_station_id

WHERE num_bikes_available > 40

GROUP BY starttime, stoptime, start_station_name, end_station_name, usertype, num_bikes_available

ORDER BY num_trips DESC

LIMIT 10
```

[Output]

Comment: In the most popular stations, most frequent user type is customer.

Row	starttime	stoptime	start_station_name	end_station_name	usertype	num_bikes_available	num_trips
1	2015-06-26T21:35:00	2015-06-26T21:48:00	8 Ave & W 33 St	Broadway & W 49 St	Customer	42	6
2	2015-06-21T17:44:00	2015-06-21T17:53:00	8 Ave & W 33 St	9 Ave & W 22 St	Customer	42	5
3	2015-06-28T12:30:00	2015-06-28T12:51:00	Pershing Square North	Grand Army Plaza & Central Park S	Customer	42	5
4	2015-06-30T09:10:00	2015-06-30T09:28:00	Broadway & W 55 St	12 Ave & W 40 St	Customer	48	5
5	2015-06-28T12:59:00	2015-06-28T13:54:00	Grand Army Plaza & Central Park S	Grand Army Plaza & Central Park S	Customer	44	4
6	2015-06-28T13:58:00	2015-06-28T14:20:00	Grand Army Plaza & Central Park S	Broadway & W 32 St	Customer	44	4
7	2015-06-27T12:29:00	2015-06-27T13:10:00	Grand Army Plaza & Central Park S	Grand Army Plaza & Central Park S	Customer	44	4
8	2015-06-20T13:21:00	2015-06-20T13:37:00	Broadway & W 55 St	W 45 St & 6 Ave	Customer	48	4
9	2015-06-12T16:09:00	2015-06-12T16:27:00	Grand Army Plaza & Central Park S	E 45 St & 3 Ave	Customer	44	4
10	2015-06-25T12:50:00	2015-06-25T14:15:00	Grand Army Plaza & Central Park S	Grand Army Plaza & Central Park S	Customer	44	4

Figure 10: Table of Usertype of Popular Stations

2) Most Popular Trips

[Input]

```
SELECT starttime, stoptime, start_station_name, end_station_name, usertype, count(*) as num_trips,
avg(tripduration)/60 as avg_trip_duration,

FROM `bigquery-public-data.new_york_citibike.citibike_trips`

GROUP BY starttime, stoptime, start_station_name, end_station_name, usertype

HAVING num_trips > 4

ORDER BY num_trips DESC

LIMIT 10
```

[Output]

Comment: In the most popular trips, most frequent user type is still customer.

Row	starttime	stoptime	start_station_name	end_station_name	usertype	num_trips	avg_trip_duration
1	*null*	*null*				5828994	*null*
2	2015-06-26T21:35:00	2015-06-26T21:48:00	8 Ave & W 33 St	Broadway & W 49 St	Customer	6	13.213888888888887
3	2015-06-07T20:21:00	2015-06-07T20:30:00	Front St & Maiden Ln	Liberty St & Broadway	Customer	6	8.980555555555556
4	2015-06-21T18:40:00	2015-06-21T18:57:00	Central Park S & 6 Ave	Grand Army Plaza & Central Park S	Customer	6	17.269444444444442
5	2015-06-07T19:17:00	2015-06-07T19:24:00	Front St & Maiden Ln	Bus Slip & State St	Customer	6	6.730555555555555
6	2015-06-21T13:37:00	2015-06-21T13:50:00	Cumberland St & Lafayette Ave	DeKalb Ave & Skillman St	Customer	6	13.197222222222223
7	2015-06-30T12:11:00	2015-06-30T12:34:00	Water - Whitehall Plaza	Centre St & Chambers St	Customer	6	23.502777777777776
8	2015-06-20T18:52:00	2015-06-20T19:15:00	Lafayette St & E 8 St	E 20 St & Park Ave	Customer	6	22.680555555555557
9	2015-06-08T21:30:00	2015-06-08T21:54:00	E 7 St & Avenue A	Peck Slip & Front St	Customer	6	24.17777777777778
10	2015-06-22T12:06:00	2015-06-22T12:35:00	Broad St & Bridge St	Christopher St & Greenwich St	Subscriber	6	28.87777777777778

Figure 11: Table of Usertype of Popular Trips

5. Pattern in the Empty Stations

[Input]

```
SELECT station_id, start_station_id, end_station_id, name, usertype, capacity, num_bikes_available,
num_docks_available

FROM `bigquery-public-data.new_york_citibike.citibike_trips` AS t

LEFT JOIN `bigquery-public-data.new_york_citibike.citibike_stations` AS s

ON s.station_id = t.start_station_id OR s.station_id = t.end_station_id

WHERE num_bikes_available = 0

ORDER BY num_docks_available, capacity ASC

LIMIT 10
```

[Output]

Comment: In empty stations, most frequent user type is subscriber.

Row	station_id	start_station_id	end_station_id	name	usertype	capacity	num_bikes_available	num_docks_available
1	3146	3431	3146	E 81 St & 3 Ave	Subscriber	47	0	47
2	3146	3135	3146	E 81 St & 3 Ave	Subscriber	47	0	47
3	3146	3146	3155	E 81 St & 3 Ave	Subscriber	47	0	47
4	3146	3134	3146	E 81 St & 3 Ave	Subscriber	47	0	47
5	3146	3151	3146	E 81 St & 3 Ave	Subscriber	47	0	47
6	3146	3369	3146	E 81 St & 3 Ave	Subscriber	47	0	47
7	3146	3146	3140	E 81 St & 3 Ave	Subscriber	47	0	47
8	3146	3146	3165	E 81 St & 3 Ave	Customer	47	0	47
9	3146	3283	3146	E 81 St & 3 Ave	Customer	47	0	47
10	3146	3146	3167	E 81 St & 3 Ave	Customer	47	0	47

Figure 12: Table of Usertype of Empty Stations

Section 4: Data Visualization on Top of SQL

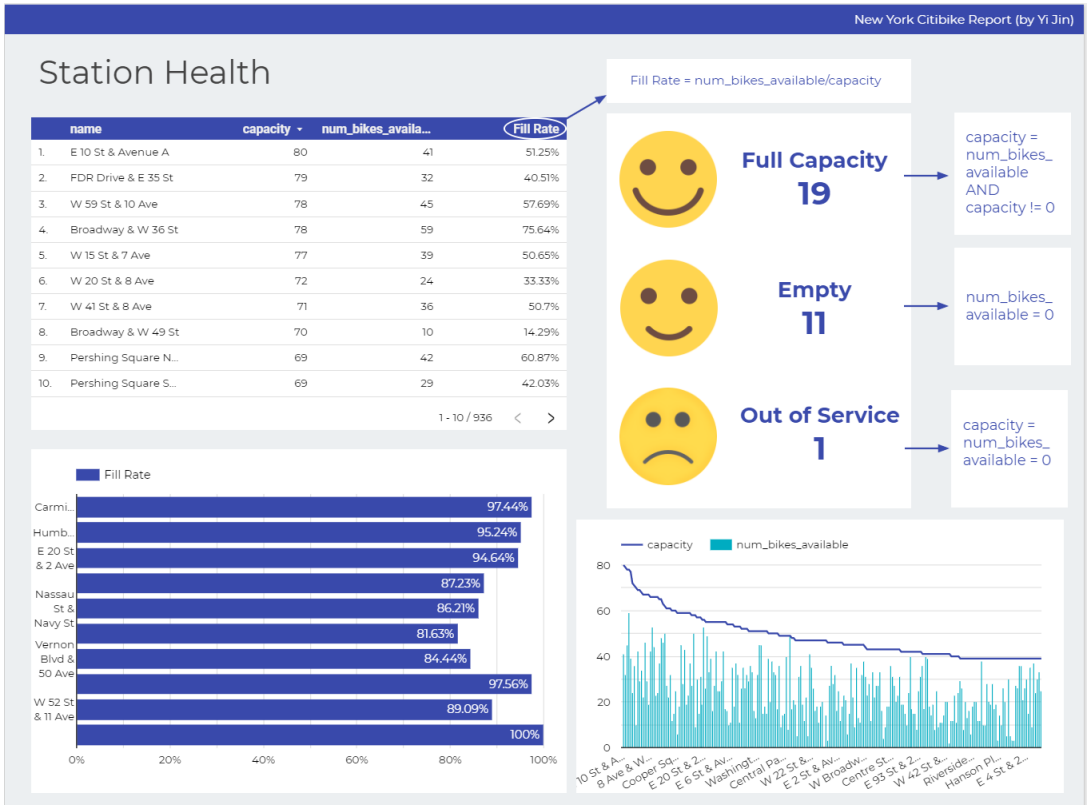
Build an operational dashboard to answer the following business questions:

Station Health

- How many stations are at capacity, empty, or out of service?
- What is the fill rate (bikes available/capacity) for each station?
- What is the most popular station to start rides for all time?
- What is the most popular station to end rides for all time?
- What are the top 3 most popular trips (start and end station combination) for all time?
- Which hours of the day does usage peak on weekdays?
- Which hours of day does usage peak on weekends?

System Health

- How many trips are there per day?
- What is the average trip duration?
- What was the shortest trip?
- What was the longest trip?
- How many total hours of usage does each bike have?



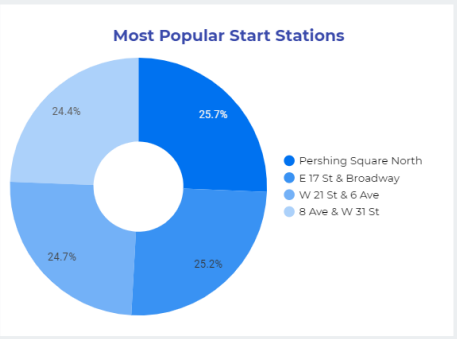
Station Health

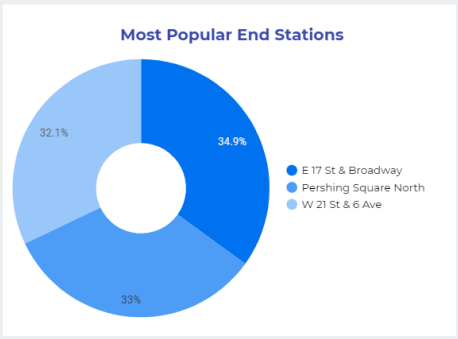
	start_station_name	end_station_name	num_trips
1.	Central Park S & 6 Ave	Central Park S & 6 Ave	
2.	Grand Army Plaza & Central Park S	Grand Army Plaza & Central Park S	
3.	Centre St & Chambers St	Centre St & Chambers St	
4.	Broadway & W 60 St	Broadway & W 60 St	
5.	12 Ave & W 40 St	West St & Chambers St	
6.	W 21 St & 6 Ave	9 Ave & W 22 St	
7.	W 21 St & 6 Ave	W 22 St & 10 Ave	
8.	West St & Chambers St	12 Ave & W 40 St	
9.	West St & Chambers St	West St & Chambers St	
10.	12 Ave & W 40 St	12 Ave & W 40 St	

1 - 10 / 10

The most popular station to start rides for all time is **Perishing Sqaure North**, and the most popular station to end rides for all time is **E17 St & Broadway**.

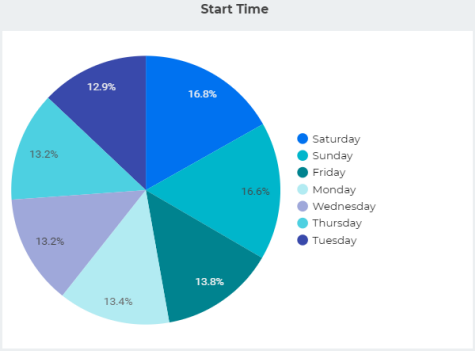
The top 3 most popular trips for all time are **Central Park**, **Grand Army Plaza**, **and Chambers St**.

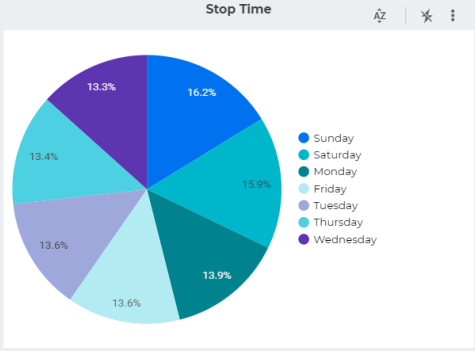


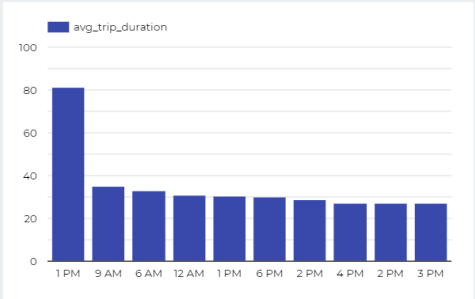


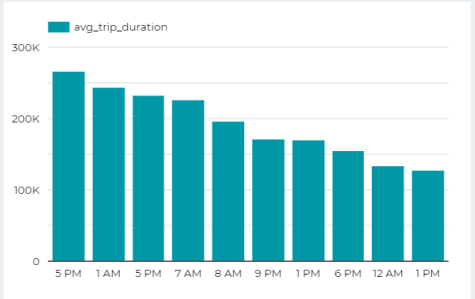
Station Health

The usage peaks mostly during the weekends. For the start time, usage peaks at 1 PM, 9 AM, and 6 AM; for stop time, it usually peaks at 5 PM.









System Health

	bikeid	total_usage_hours
1.	30503	5,660.78
2.	16533	5,531.67
3.	16777	5,529.75
4.	18504	5,169.15
5.	15336	4,876.96
6.	30875	4,663.82
7.	19957	4,552
8.	18251	4,202.84
9.	17655	4,105.26
10.	21523	4,047.47
11.	16355	4,036.93
12.	19422	3,988.33
13.	20999	3,965.95
14.	16734	3,755.41
15.	17607	3,753.77
16.	19903	3,746.31
17.	30446	3,723.62
18.	17094	3,588.9
19.	24939	3,495.38
20.	18565	3,354.79
21.	33252	3,340.88
22.	27076	3,338.31
23.	18275	3,291.39

1 - 100 / 17683

number of trips per day
114,220.38

average trip duration (in min)
16.04

shortest trip (in min)
1.0

longest trip (in min)
325,167.48