
DATA ANALYSIS AND VISUALIZATION OF NYC CITI BIKE TRIPS FOR BUSINESS INTELLIGENCE

MIS 636- DATA WAREHOUSING AND BUSINESS INTELLIGENCE
FALL -2020 FINAL PROJECT

Mahaswin Ramalingam Balaji



AGENDA

- Company Profile
- Why DW and BI
- BUS Matrix
- Opportunity Matrix
- Sample Data
- Logical Fact Table
- Steps to create Dimensional Modelling
- Fact Table
- Dimension Table
- High Level Data Model
- Slowly Changing Dimension
- OLAP Cube
- Tableau Visualization
- Dashboard Insights
- Conclusion





CITIBIKE COMPANY PROFILE

- Citibike is a public bike sharing program that allows users to ride rented bicycles from one bike station to another.
- Users can check bike availability at the station or on the mobile app and pick up a bike at one of the many stations around the city. They can end a ride by returning the bike to any station and sliding the bike firmly into an empty dock
- The density of bike share stations makes it convenient for riders to find and return bikes and this makes Citibike very successful.
- This project shows how Citi Bike, relates to the existing travel modes, and the characteristics of its users.
- Citibike is the largest bike-share program in the United States. It is most popular in New York City , and in its first two years has become a key element in the city's transportation network.

WHY DW AND BI

- Helps in maintaining historical records.
- Helps in analyzing the data to gain a better understanding of the business and to improve the business.
- Promotes better business intelligence for end-users
- Reduction in time to locate, access, and analyze information
- Effective Consolidation of disparate information sources
- Enables less time-to-market for products and services

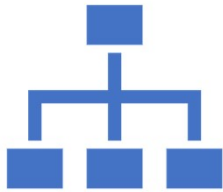
BUS MATRIX

- Bus matrix is used to describe dimension reuse across subjects
- Bus matrix takes business processes breaks them into objects that participate in those business processes
- It is a tool for designing and communicating the enterprise data warehouse bus architecture.
- The rows of the matrix are business processes, and the columns are dimensions.
- The cells of the matrix are then marked to indicate which dimensions apply to which processes.

BUS MATRIX

Business Process	Time	BikeStation	User_Info	Bicycle	Rental	Employee
Citibike Mobile App Sales transaction	x	x	x	x	x	
Sales transaction at Bike Station	x	x	x	x	x	x
Vendor Sales Transaction	x	x	x		x	
Return process	x	x	x	x		x

OPPORTUNITY MATRIX



The matrix consists of business processes versus the department or workgroups in the organization.



The matrix indicates which business functions/ departments are interested in which business processes in the organization.



The Opportunity or stakeholder matrix helps identify which business groups should be invited to the collaborative design sessions for each process-centric row.

OPPORTUNITY MATRIX

Business Process	Customer Service	Finance	Operations	Marketing	Logistics	Strategy Management
Citibike Mobile App Sales transaction	x	x	x	x	x	x
Sales transaction at Bike Station	x	x	x	x	x	x
Vendor Sales Transaction	x	x		x	x	
Trip Processing	x		x			
Return process	x		x		x	

SAMPLE DATA

201802-citibike-tripdata

tripduration	starttime	stoptime	start station id	start station name	start station latitude	start station longitude	end station id	end station name	end station latitude	end station longitude	bikeid	usertype	birth year	gender
771	2018-02-01 00:14:16.4120	2018-02-01 00:27:08.2290	72	W 52 St & 11 Ave	40.76727216	-73.99392888	379	W 31 St & 7 Ave	40.749156	-73.9916	14536	Subscriber	1952	1
264	2018-02-01 05:14:45.1790	2018-02-01 05:19:09.6860	72	W 52 St & 11 Ave	40.76727216	-73.99392888	478	11 Ave & W 41 St	40.76030096	-73.99884222	32820	Subscriber	1965	1
819	2018-02-01 06:48:55.2290	2018-02-01 07:02:35.0290	72	W 52 St & 11 Ave	40.76727216	-73.99392888	405	Washington St & Gansevoort St	40.739323	-74.008119	16131	Subscriber	1968	1
646	2018-02-01 07:12:50.1740	2018-02-01 07:23:36.5280	72	W 52 St & 11 Ave	40.76727216	-73.99392888	2006	Central Park S & 6 Ave	40.76590936	-73.97634151	20831	Subscriber	1990	2
1312	2018-02-01 07:46:48.8750	2018-02-01 08:08:41.5430	72	W 52 St & 11 Ave	40.76727216	-73.99392888	435	W 21 St & 6 Ave	40.74173969	-73.99415556	15899	Subscriber	1957	1
435	2018-02-01 07:50:45.4490	2018-02-01 07:58:00.8360	72	W 52 St & 11 Ave	40.76727216	-73.99392888	173	Broadway & W 49 St	40.76068327096590	-73.9845272898674	19749	Subscriber	1983	1
289	2018-02-01 07:51:07.7240	2018-02-01 07:55:57.3360	72	W 52 St & 11 Ave	40.76727216	-73.99392888	514	12 Ave & W 40 St	40.76087502	-74.00277668	17412	Subscriber	1988	2
1454	2018-02-01 07:53:43.0620	2018-02-01 08:17:57.3330	72	W 52 St & 11 Ave	40.76727216	-73.99392888	426	West St & Chambers St	40.71754834	-74.01322069	17548	Subscriber	1988	1
639	2018-02-01 08:16:56.5200	2018-02-01 08:27:35.7080	72	W 52 St & 11 Ave	40.76727216	-73.99392888	3173	Riverside Blvd & W 67 St	40.777507027548000	-73.98888587951660	18515	Subscriber	1988	1
530	2018-02-01 08:40:24.5540	2018-02-01 08:49:14.5940	72	W 52 St & 11 Ave	40.76727216	-73.99392888	458	11 Ave & W 27 St	40.751396	-74.005226	31050	Subscriber	1982	1
779	2018-02-01 08:53:27.2510	2018-02-01 09:06:27.0760	72	W 52 St & 11 Ave	40.76727216	-73.99392888	524	W 43 St & 6 Ave	40.75527307	-73.98316936	20522	Subscriber	1960	2
1167	2018-02-01 08:55:17.5330	2018-02-01 09:14:45.1670	72	W 52 St & 11 Ave	40.76727216	-73.99392888	496	E 16 St & 5 Ave	40.73726186	-73.99238967	32013	Subscriber	1959	1
1211	2018-02-01 08:59:40.0190	2018-02-01 09:19:51.4700	72	W 52 St & 11 Ave	40.76727216	-73.99392888	347	Greenwich St & W Houston St	40.728846	-74.008591	31524	Subscriber	1982	1
448	2018-02-01 09:10:39.4970	2018-02-01 09:18:07.9050	72	W 52 St & 11 Ave	40.76727216	-73.99392888	514	12 Ave & W 40 St	40.76087502	-74.00277668	30755	Subscriber	1990	1
529	2018-02-01 09:17:41.4010	2018-02-01 09:26:31.1250	72	W 52 St & 11 Ave	40.76727216	-73.99392888	447	8 Ave & W 52 St	40.76370739	-73.9851615	31657	Subscriber	1947	1
970	2018-02-01 09:19:46.6380	2018-02-01 09:35:57.1430	72	W 52 St & 11 Ave	40.76727216	-73.99392888	465	Broadway & W 41 St	40.75513557	-73.98658032	32613	Subscriber	1978	0
895	2018-02-01 09:36:52.1060	2018-02-01 09:51:47.3610	72	W 52 St & 11 Ave	40.76727216	-73.99392888	458	11 Ave & W 27 St	40.751396	-74.005226	31032	Subscriber	1984	2
916	2018-02-01 09:37:45.3410	2018-02-01 09:53:02.0790	72	W 52 St & 11 Ave	40.76727216	-73.99392888	405	Washington St & Gansevoort St	40.739323	-74.008119	31388	Subscriber	1961	1
653	2018-02-01 09:38:05.5610	2018-02-01 09:48:58.8190	72	W 52 St & 11 Ave	40.76727216	-73.99392888	440	E 45 St & 3 Ave	40.75255434	-73.97282625	31365	Subscriber	1988	1
387	2018-02-01 09:41:41.4220	2018-02-01 09:48:09.1970	72	W 52 St & 11 Ave	40.76727216	-73.99392888	495	W 47 St & 10 Ave	40.76269882	-73.99301222	30135	Subscriber	1956	2
711	2018-02-01 09:47:23.1880	2018-02-01 09:59:14.5040	72	W 52 St & 11 Ave	40.76727216	-73.99392888	458	11 Ave & W 27 St	40.751396	-74.005226	27978	Subscriber	1987	1
798	2018-02-01 09:47:56.7270	2018-02-01 10:01:15.4310	72	W 52 St & 11 Ave	40.76727216	-73.99392888	442	W 27 St & 7 Ave	40.746647	-73.993915	30538	Subscriber	1984	1
1101	2018-02-01 09:58:23.6830	2018-02-01 10:16:45.3810	72	W 52 St & 11 Ave	40.76727216	-73.99392888	3462	E 44 St & 2 Ave	40.75118387463280	-73.97138714790340	18685	Subscriber	1980	1
868	2018-02-01 10:02:01.8680	2018-02-01 10:16:30.2020	72	W 52 St & 11 Ave	40.76727216	-73.99392888	3472	W 15 St & 10 Ave	40.742753828659000	-74.00747358798980	20569	Subscriber	1984	2
2359	2018-02-01 10:36:18.8480	2018-02-01 11:15:38.0000	72	W 52 St & 11 Ave	40.76727216	-73.99392888	3435	Grand St & Elizabeth St	40.718822	-73.99596	32217	Subscriber	1959	2
706	2018-02-01 10:45:50.9110	2018-02-01 10:57:37.4510	72	W 52 St & 11 Ave	40.76727216	-73.99392888	359	E 47 St & Park Ave	40.75510267	-73.97498696	17693	Subscriber	1992	1
675	2018-02-01 11:58:54.6860	2018-02-01 12:10:10.5380	72	W 52 St & 11 Ave	40.76727216	-73.99392888	2006	Central Park S & 6 Ave	40.76590936	-73.97634151	26626	Subscriber	1991	1

COLUMNS IN THE DATA

- Trip Duration (seconds)
- Start Time and Date
- Stop Time and Date
- Start Station Name
- End Station Name
- Station ID
- Station Lat/Long
- Bike ID
- User Type (Customer = 24-hour pass or 3-day pass user; Subscriber = Annual Member)
- Gender (Zero=unknown; 1=male; 2=female)
- Year of Birth

STEPS TO CREATE DIMENSIONAL MODELLING

Select Business Process

OrgX – citibike Renting Process



Choose the Grain of the Business Process

OrgX – Each Rider, Trip



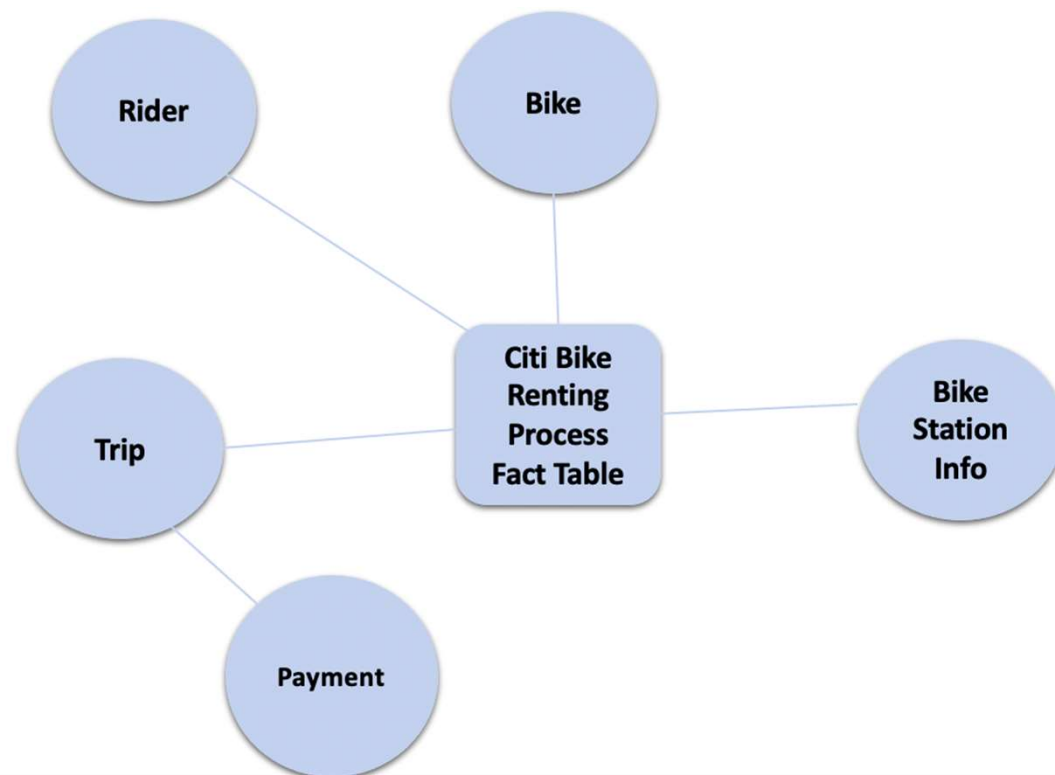
Choose Dimensions of the Business Process

OrgX – Rider, Trip, Bike, Bike Station Info

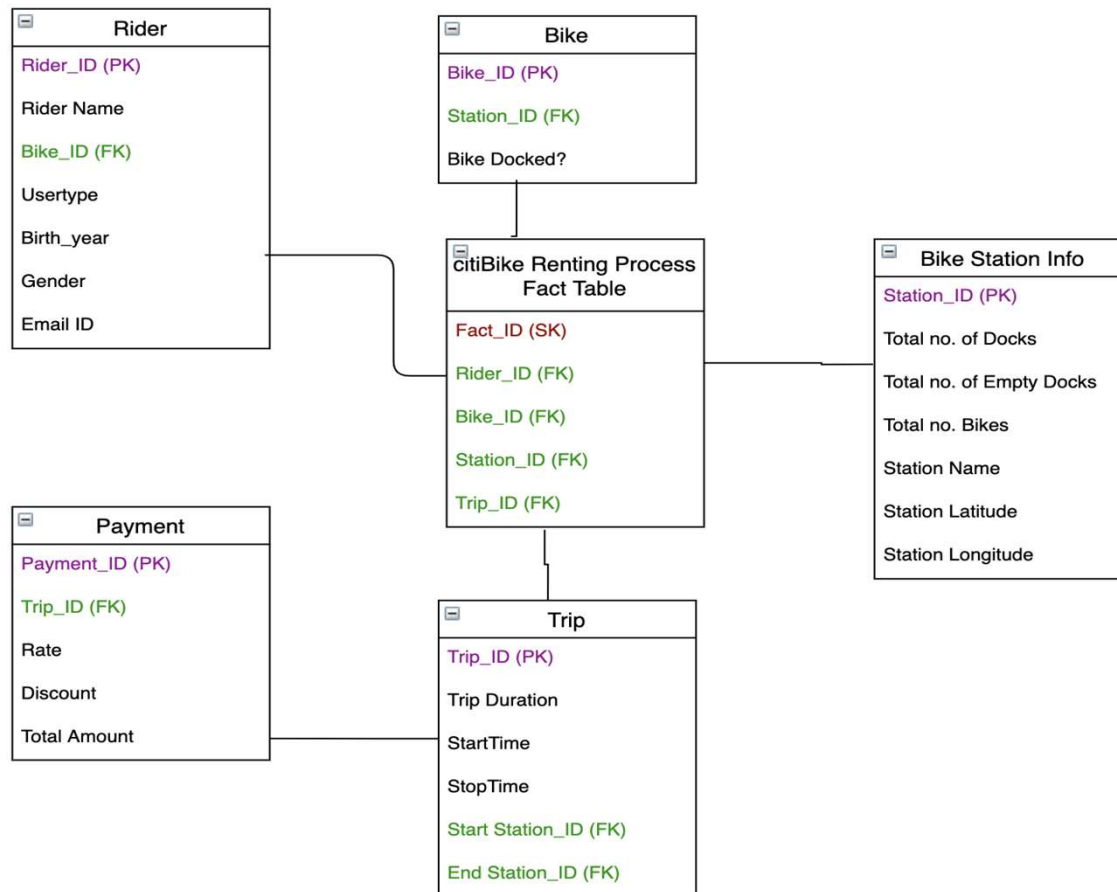


Choose the measured Facts

Orgx – # of bikes/docks available per station, total trip time, active riders renting at any given point in time, bike rented status at any given point in time



HIGH LEVEL DATA MODEL



LOGICAL FACT TABLE DIAGRAM

FACT TABLE

- A fact table consists of facts of a particular business process. Facts are also known as measurements or metrics. A fact table record captures a measurement or a metric.
- A fact table is used in the dimensional model in data warehouse design. A fact table is found at the center of a star schema or snowflake schema surrounded by dimension tables.

Fact_ID(SK)
Bike_ID(FK)
Rider_ID(FK)
Station_ID(FK)
Trip_ID(FK)

DIMENSION TABLE

- A dimension is a structure that categorizes facts and measure in order to answer business question
- A Dimension Table is a table in a star schema of a data warehouse. Data warehouses are built using dimensional data models which consist of fact and dimension tables
- Dimension tables are used to describe dimensions, they contain dimension keys, values and attributes

Dimension Tables

Rider
Trip
Bike
Bike Station Info

CONFORMED DIMENSION

- A conformed dimension provides the same meaning to every fact table in a data warehouse which it relates.
- A single conformed dimension table can be used against multiple fact tables in the same data mart, or a replicated dimension can be used in more than one data mart.
- So, for our data warehouse, following dimensions remain the same
 - Rider dimension
 - Bike dimension
 - Trip Dimension

SLOWLY CHANGING DIMENSION

Bike Station Info

Type 1

- In this scenario, the old record with the number of docks would be deleted and a new record with the new values would be inserted
- Updating Station Name

Type 2

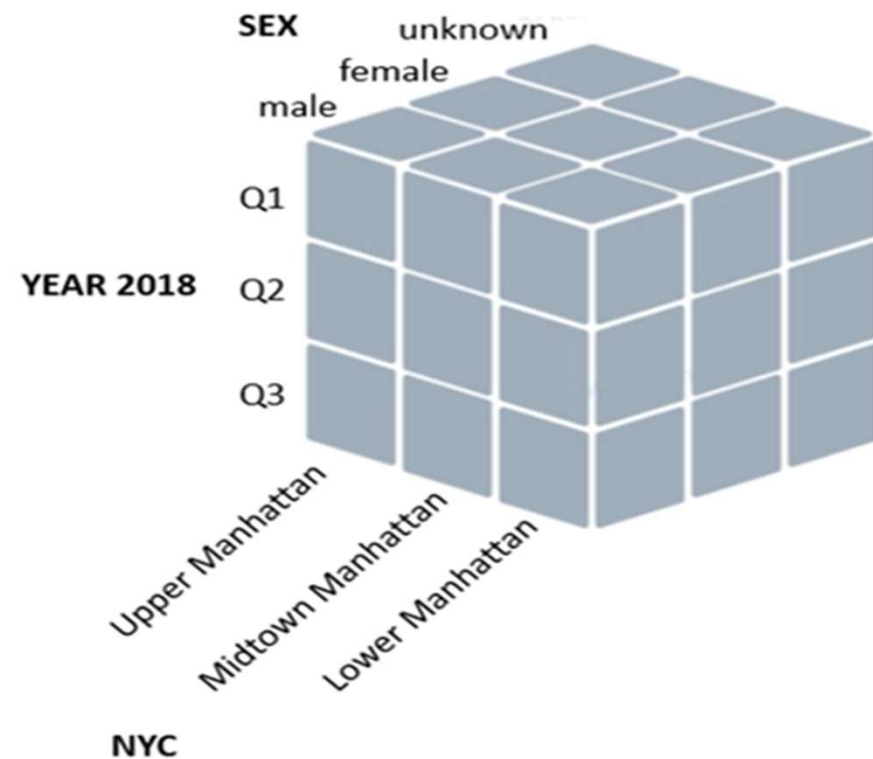
- In this scenario, a new row would be added with new Station name to maintain the history

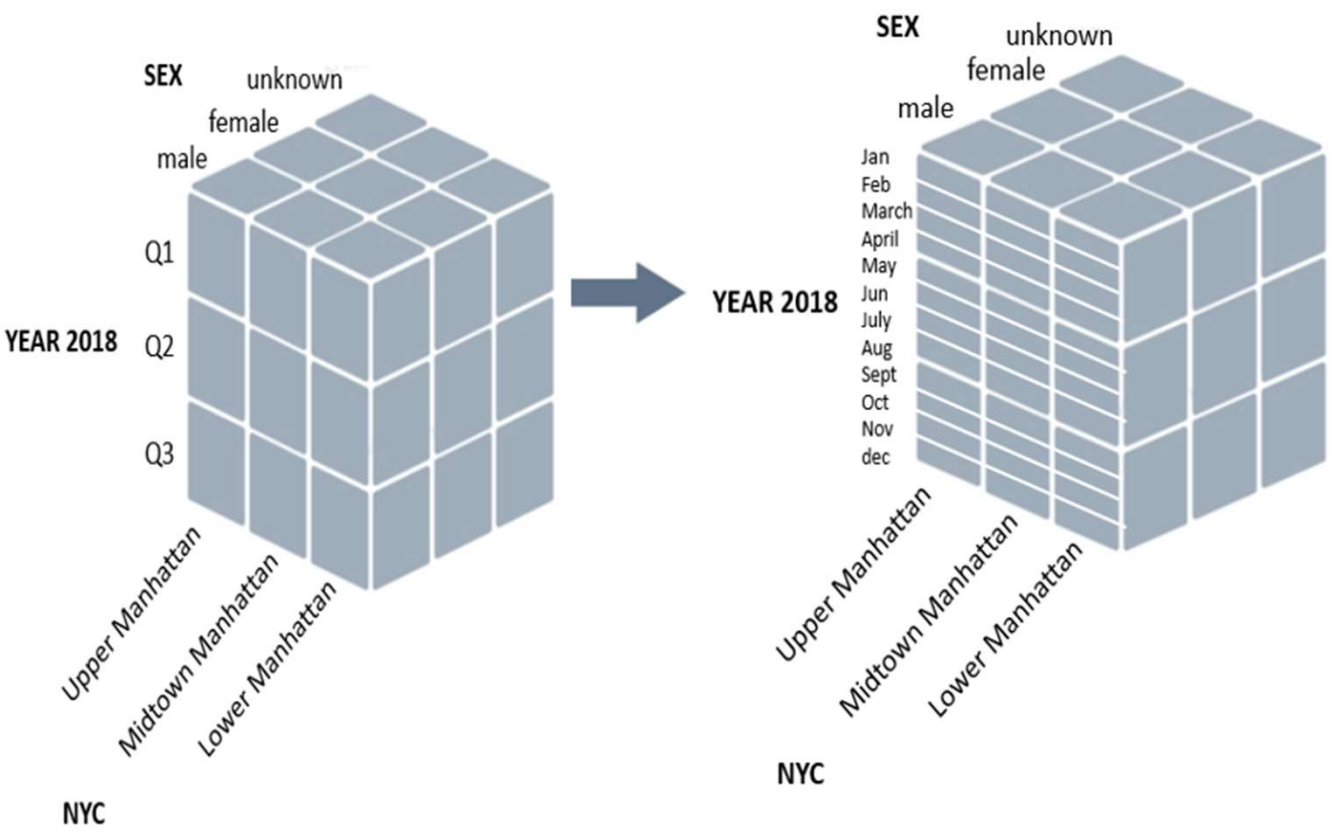
DIMENSIONAL TABLE DETAILED DIAGRAM: BIKE STATION INFO TABLE

Attribute Name	Attribute Description	Slowly changing dimensions policy	Sample Values
Station_ID	The Trip ID is the ID associated	Type-0: No Action	072
Total number of Docks	Total number of docks available at the station	Type-1: Overwrites	3
Total Number of Empty Docks	Number of Empty docks available for bike docking at the station	Type-0: No Action	379
Total No. Bikes	Number of bikes currently docked at the station	Type-0: No Action	W 52 St & 11 Ave
Station Name	Each station has specific name	Type-2: Add New Row	W 31 St & 7 Ave
Station Latitude	Each Station has specific latitude	Type-0: No Action	40.76727216
Station Longitude	Each Station has specific longitude	Type-0: No Action	-73.99392888

OLAP CUBE

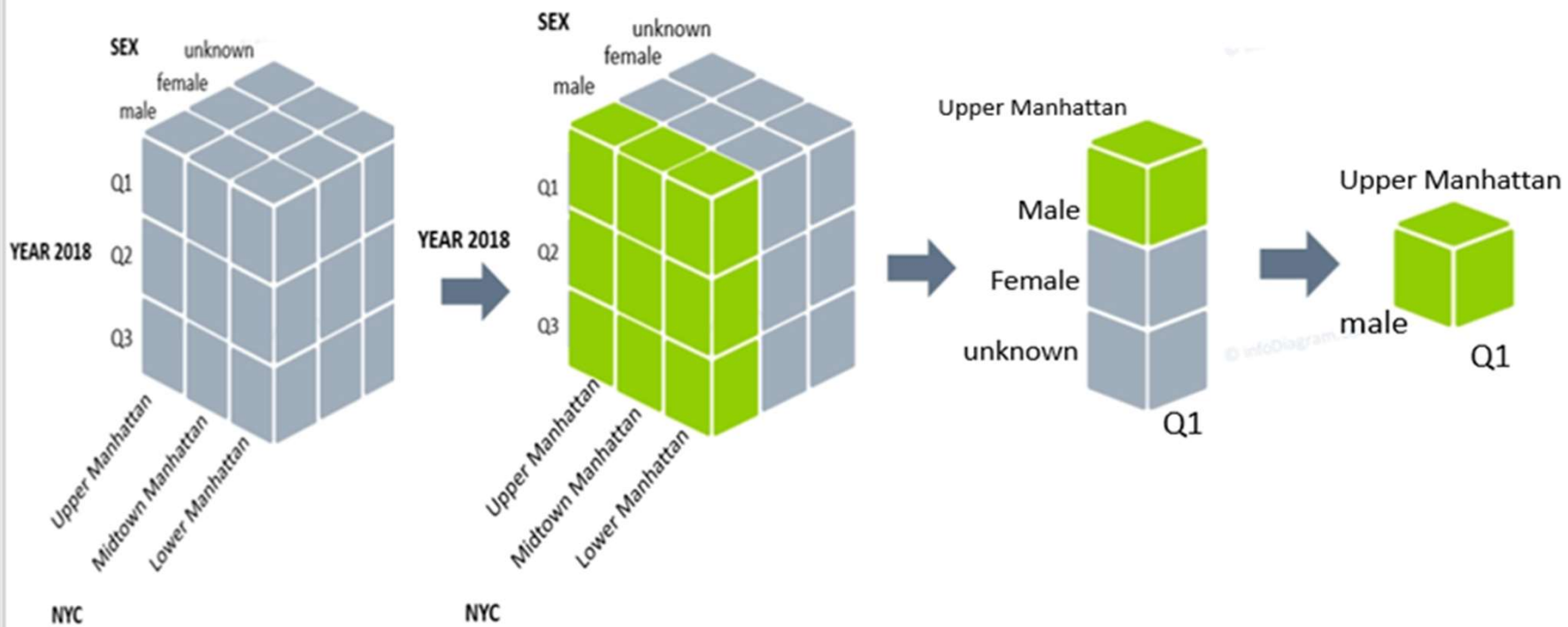
- It is a category of software that allows users to analyze information from multiple database systems at the same time.
- OLAP databases are divided into one or more cubes. The cubes are designed in such a way that creating and viewing reports become easy.
- It is a technology that enables analysts to extract and view business data from different points of view.
- Analysts frequently need to group, aggregate and join data. With OLAP data can be pre-calculated and pre-aggregated, making analysis faster.





DRILL DOWN

SLICING AND DICING



CITIBIKE BI DASHBOARD - [CLICK TO VIEW](#)

< Mahaswin Ramalingam Balaji - Profile

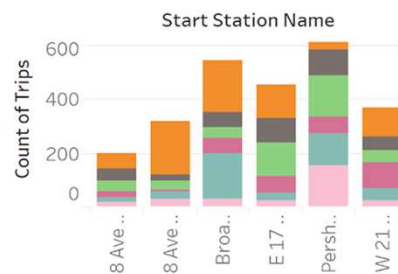
☆ Favorite

🔗 Explore (Beta)

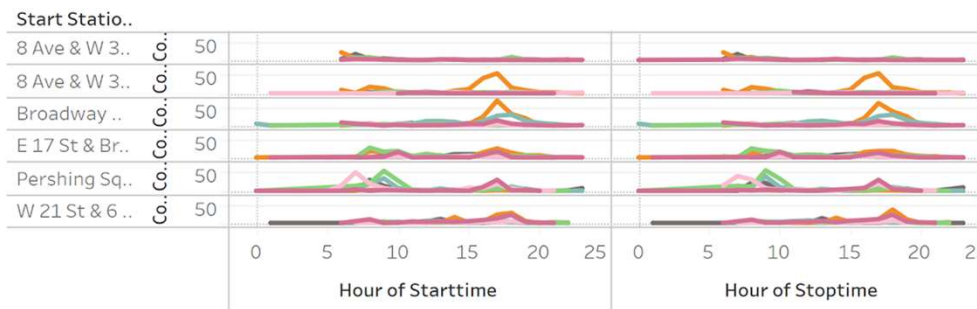
📄 Download

Data analysis & visualization of Citibike trips in NYC for Business Intelligence

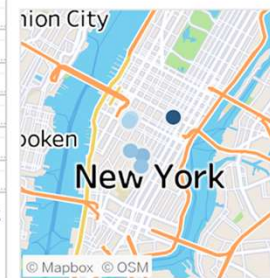
Popular start stations with end stations



Hour of start time and end time at top stations



Popular end stations in NYC



Count of 201802-citi...



Gender

- Not specified
- Male
- Female

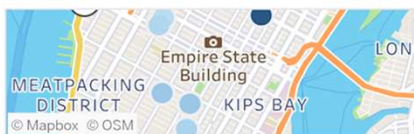
End Station Name

- Pershing Square ...
- W 21 St & 6 Ave
- Broadway & E 22 ...
- 8 Ave & W 31 St
- E 17 St & Broadw...
- 8 Ave & W 33 St

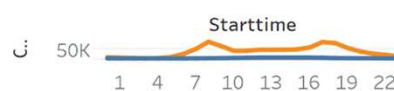
Usertype

- Customer
- Subscriber

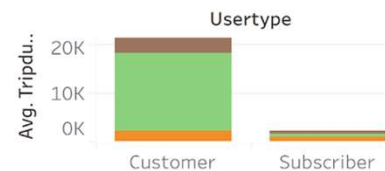
Popular start stations in NYC



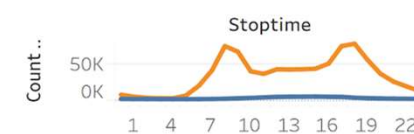
Start times for customers and subscribers



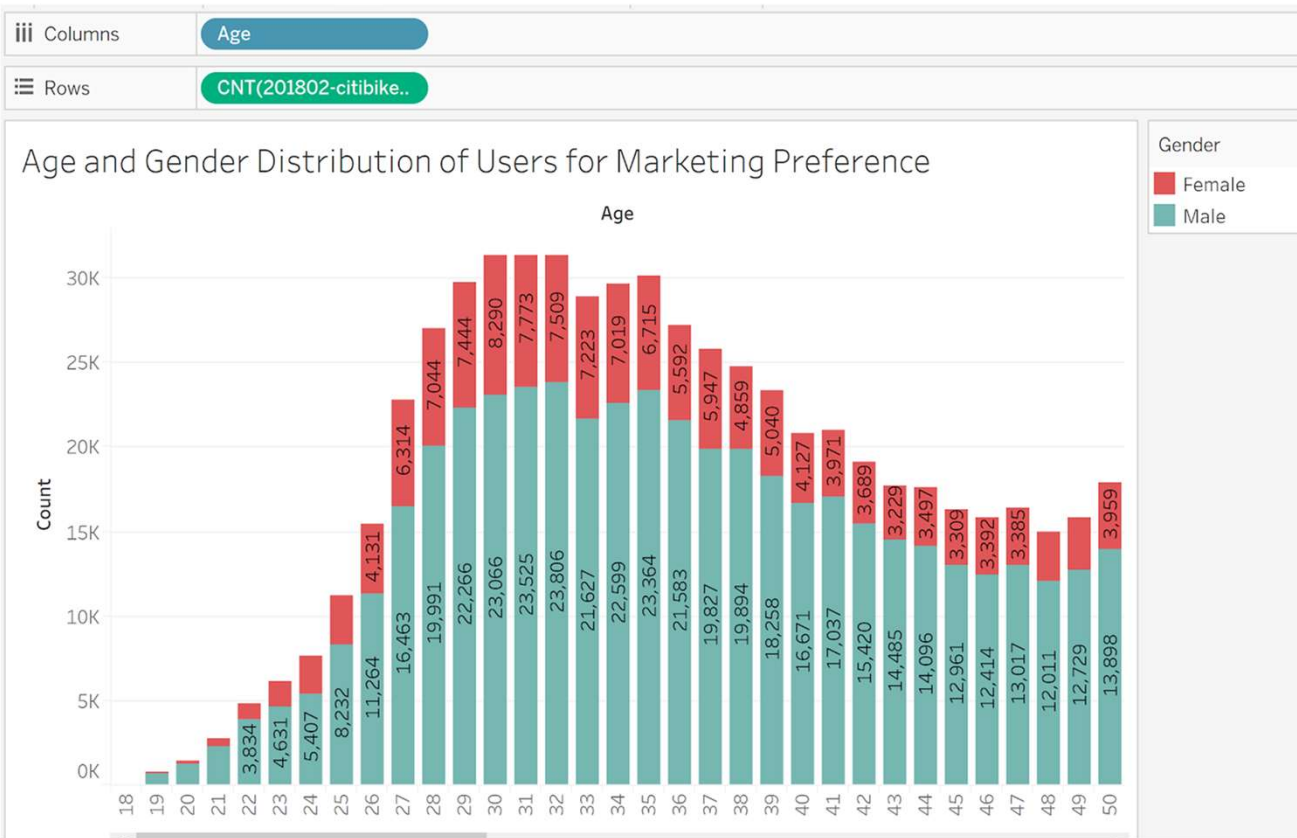
Average Trip Duration



End times for customers and subscribers



AGE AND GENDER DISTRIBUTION FOR MARKETING



Classification of Users based on Age and Gender

Male: Higher usage activity for ages 30-32

Female: Higher Usage activity for ages 29-33

DASHBOARD INSIGHTS

- Most Popular Start Station- Pershing Square
- Most Popular End Station- Pershing Square and Broadway at E 22nd street
- The hours of start time and end time at stations almost follow a similar pattern which can be observed from the dashboard that maximum number of starts and stops are around 8AM and between 5-6PM. This could be due to the fact that most office goers use the service
- Average Trip Duration of customers are much higher than that of subscribers and males outweigh female users

CONCLUSION

- The Citi might want to consider age when planning out things like bike infrastructure, much in the way that they do for crosswalks.
- As we can see that daily office workers use service more but it should be used in the odd times of the day too so Citi Bike could investigate adding more bikes to tourist heavy areas or any other way so that the non used service time is utilized.
- Citi Bike should consider adding promotional offers or student discounts to their subscribers to attract more youth in using Citi Bike.
- Citi Bike should introduce any new schemes where it would attract more female customers maybe by introducing female friendly bikes as we can see that males use the service more than females.

Thank You