Cyclistic - Data Analysis Case Study

Ask: A clear statement of the business task

1. Total number of bicycles?
2. Total number of stations?
3. Total types of cyclists?
4. Total types of bicycles?
5. Total rides per day for each cyclist type?
    a. Comparison between riders on weekdays and weekend days?
6. Total number of riders per cyclist type?
7. Total rides per month?
    a. Comparison between frequency of riders for each season?
8. Average ride duration per cyclist type?
9. What is the usage of each bicycle type?
10. Rides per hour for each cyclist type?
11. Geographical density for each cyclist type?

Prepare: A description of all data sources used

The data source is available on https://divvy-tripdata.s3.amazonaws.com/index.html

- The sample dataset for EXCEL project was Jan 2021.
- The sample dataset for SQL project was February 2021.
- The Dataset used for TABLEAU was FY-2021

Process: Documentation of any cleaning or manipulation of data

EXCEL: Sample Cyclist DA.xlsx

- Removed ride_id column.
- Created start_date and end_date column.
- Created start_day column and used, =TEXT(D2,"dddd") , to identify the day.
- Created start_month column and used, =TEXT(D2,"mmmm") , to identify the month.
- Created start_time column and used, =TEXT(B2,"h")
- Created trip_duration column and used, =C2-B2 , and changed format to , Custom – [h]:mm
- Created a new sheet – result, to record my findings.

SQL:

- Created start_date, end_date, start_time, end_time, columns:
  - # To create start_time and end_time columns.
  - ALTER TABLE february_2021 ADD start_time TIME;
    UPDATE february_2021 SET start_time = started_at;
  - ALTER TABLE february_2021 ADD end_time TIME;
    UPDATE february_2021 SET end_time = ended_at;

    # To update the datatype from DATETIME to DATE.
  - ALTER TABLE february_2021
    MODIFY COLUMN started_at DATE;
  - ALTER TABLE february_2021
    MODIFY COLUMN ended_at DATE;

    # To rename started_at and ended_at columns.
  - ALTER TABLE february_2021
    RENAME COLUMN started_at TO start_date;
  - ALTER TABLE february_2021
    RENAME COLUMN ended_at TO end_date;
- Created start_day column.
  - # To create day column with numeric values
  - ALTER TABLE february_2021 ADD start_day INT;
    UPDATE february_2021 SET start_day = WEEKDAY(start_date);
    # To create day column with string values
  - ALTER TABLE february_2021 ADD start_day INT;
    UPDATE february_2021 SET start_day = DAYNAME(start_date);
- Created start_month column.
  - # To create a column for start_month
  - ALTER TABLE february_2021 ADD start_month INT;
    UPDATE february_2021 SET start_month = MONTH(start_date);

- Created trip_duration column.
  # To create trip_duration column
    - ALTER TABLE february_2021 ADD trip_duration TIME;
      UPDATE february_2021 SET trip_duration = TIMEDIFF(end_time, start_time);
  # To drop rows with 0 min in trip_duration
    - DELETE FROM february_2021 WHERE trip_duration = 0;

Analyze: A summary of your analysis

EXCEL: Sample dataset

Q1 - Total number of bicycles?
NA

Q2 - Total number of stations?
Count the unique value in start_station_id column , =COUNTA(UNIQUE(raw_data!L2:L103771))

Q2 - Total number of stations?
758

Q3 - Total types of cyclists?
There are two types of cyclists,

1. Casual
2. Member

To display list of (unique values) types of cyclist , =UNIQUE(raw_data!J2:J103771,FALSE,FALSE)
To count the types of cyclist , =COUNTA(UNIQUE(raw_data!J2:J103771,FALSE,FALSE))

Q3 - Total types of cyclists?
casual
member
2

Q4 - Total types of bicycles?
There are three type of bicycles,

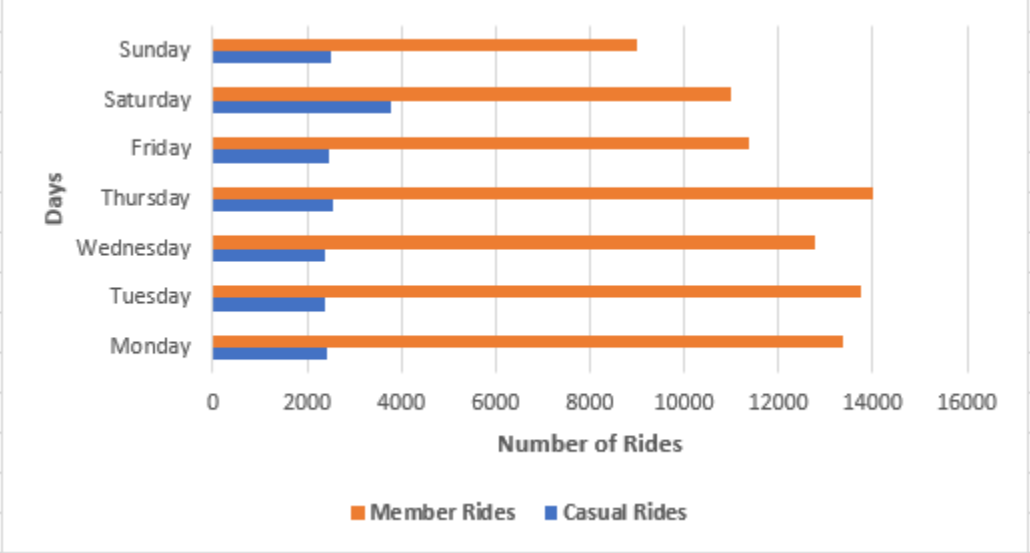
1. Electric bike
2. Classic bike
3. Docked bike

To display list of (unique values) types of bicycles , =UNIQUE(raw_data!A2:A103771,FALSE,FALSE)
To count the types of bicycles , =COUNTA(UNIQUE(raw_data!A2:A103771,FALSE,FALSE))

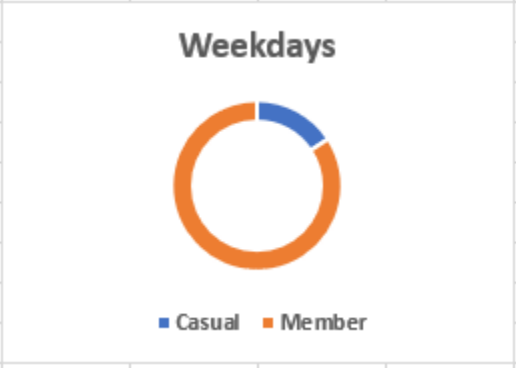
Q4 - Total types of bicycles?
electric_bike
classic_bike
docked_bike
3

Q5 - Total rides per day for each cyclist type?
To find total riders per day , =COUNTA(FILTER(raw_data!J2:J103771, raw_data!E2:E103771="Monday"))
To count total number of casual riders on Monday , =COUNTA(FILTER(raw_data!J2:J103771,(raw_data!E2:E103771="Monday")*(raw_data!J2:J103771="casual")))
To count total number of member riders on Monday , =COUNTA(FILTER(raw_data!J2:J103771,(raw_data!E2:E103771="Monday")*(raw_data!J2:J103771="member")))

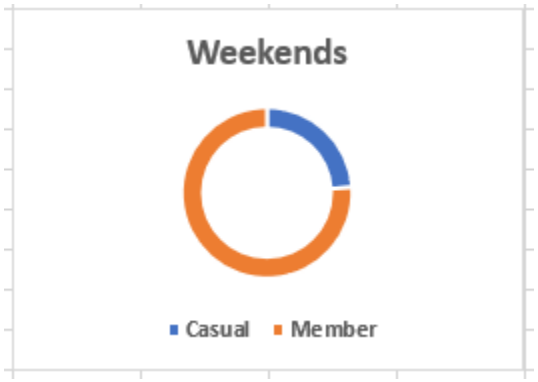
Q5 - Total rides per day for each cyclist type?			
	Casual Rides	Member Rides	Total Rides
Monday	2429	13374	15803
Tuesday	2394	13755	16149
Wednesday	2389	12785	15174
Thursday	2542	14011	16553
Friday	2459	11352	13811
Saturday	3791	10979	14770
Sunday	2515	8994	11509



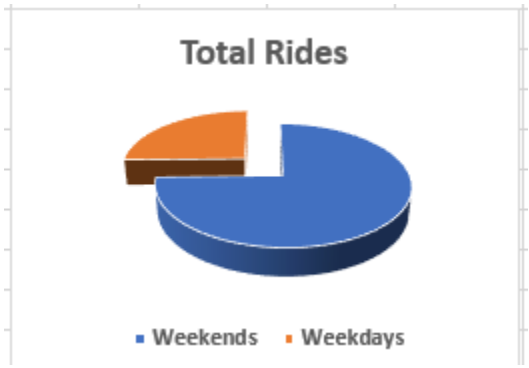
	Casual	Member
Weekdays	12213	65277



	Casual	Member
Weekends	6306	19973



	Total Rides
Weekends	77490
Weekdays	26279



Q.6 - Total number of riders per cyclist type?
To find total casual riders, =COUNTA(FILTER(raw_data!J2:J103771, raw_data!J2:J103771="casual")) .
To find total member riders, =COUNTA(FILTER(raw_data!J2:J103771, raw_data!J2:J103771="member"))

Casual_riders	18520
Member_rider	85250

Q.7 - Total rides per month? Comparison between frequency of riders for each season?

Q.8 - Average ride duration per cyclist type?
To find average ride duration casual riders, =AVERAGE(FILTER(raw_data!I2:I103771, raw_data!J2:J103771="member"))
To find average ride duration member riders, =AVERAGE(FILTER(raw_data!I2:I103771, raw_data!J2:J103771="casual"))
Format both cells in [h]:mm:ss

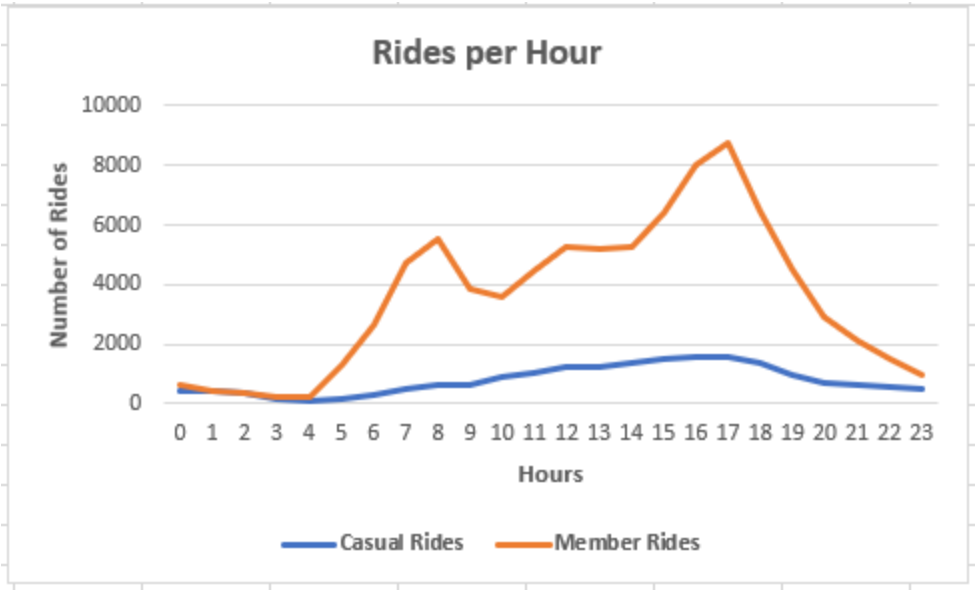
	Member Rides	Casual Rides
Average Ride duration	0:11:59	0:30:23

Q.9 - What is the usage of each bicycle type?
To find total number of rides per bike type,
=COUNTA(FILTER(raw_data!A2:A103771, raw_data!A2:A103771="classic_bike"))
=COUNTA(FILTER(raw_data!A3:A103772, raw_data!A3:A103772="electric_bike"))
=COUNTA(FILTER(raw_data!A4:A103773, raw_data!A4:A103773="docked_bike"))
To find total hours of utilization per bike type,
=SUM(FILTER(raw_data!I2:I103771, raw_data!A2:A103771="classic_bike"))
=SUM(FILTER(raw_data!I2:I103771, raw_data!A2:A103771="electric_bike"))
=SUM(FILTER(raw_data!I2:I103771, raw_data!A2:A103771="docked_bike"))

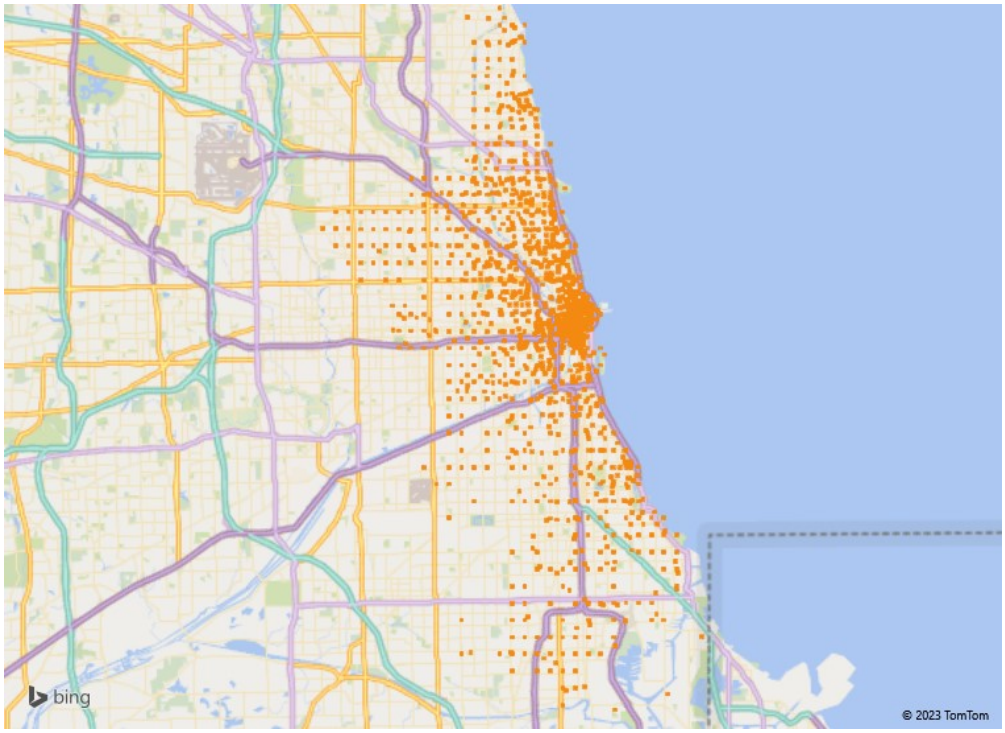
Bikes	Total Rides	Total Time
classic_bike	55067	14154:57:38
electric_bike	47741	8672:49:16
docked_bike	961	3572:27:17

Q.10 - Rides per hour for each cyclist type?
To find number of casual riders per hour,
=COUNTA(FILTER(raw_data!K2:K103771,(raw_data!H2:H103771="0")*(raw_data!K2:K103771="casual")))
To find number of member riders per hour,
=COUNTA(FILTER(raw_data!K2:K103771,(raw_data!H2:H103771="0")*(raw_data!K2:K103771="member")))

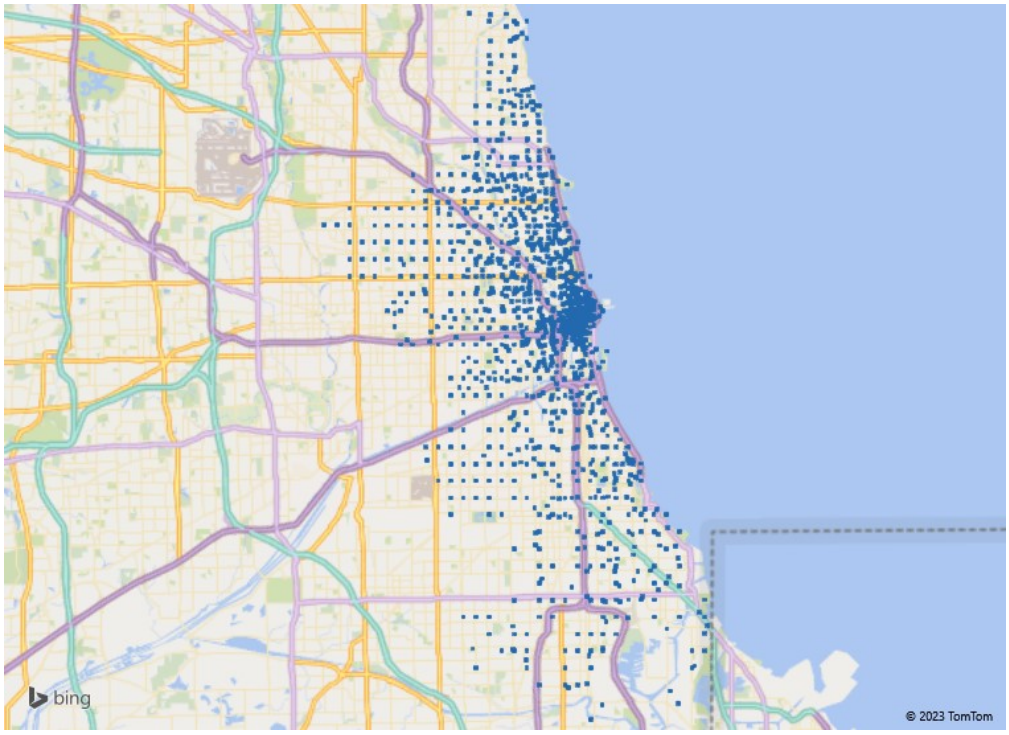
	Casual Rides	Member Rides
0	434	648
1	402	456
2	329	334
3	142	195
4	98	254
5	160	1288
6	279	2610
7	514	4738
8	603	5517
9	619	3839
10	877	3577
11	1052	4453
12	1253	5229
13	1239	5211
14	1379	5230
15	1484	6404
16	1548	8022
17	1536	8743
18	1334	6482
19	944	4539
20	670	2880
21	628	2121
22	536	1498
23	460	982



Q.11 - Geographical density for each cyclist type?
To find casual riders locations, =FILTER(raw_data!P2:P103771,(raw_data!K2:K103771="casual")) and =FILTER(raw_data!Q2:Q103771,(raw_data!K2:K103771="casual"))
To find member riders locations, =FILTER(raw_data!P2:P103771,(raw_data!K2:K103771="member")) and =FILTER(raw_data!Q2:Q103771,(raw_data!K2:K103771="member"))
Member Riders:



Casual Riders:


bing
© 2023 TomTom

SQL: Sample dataset

Q1. Total number of bicycles?

NA

Q2. Total number of stations?

To find riders on weekends.

- SELECT (DISTINCT start_station_name) FROM february_2021;

Q3. Total types of cyclists?

- SELECT count(DISTINCT member_casual) FROM february_2021;

Q4. Total types of bicycles?

- SELECT DISTINCT rideable_type FROM february_2021;

Q5. Total rides per day for each cyclist type?

Total rides per day

- SELECT distinct start_day, count(start_day) FROM february_2021 GROUP BY start_day ORDER BY start_day;

Total rides per day for each cyclist type

- SELECT distinct start_day, COUNT(start_day) FROM february_2021 WHERE member_casual like "casual" GROUP BY start_day ORDER BY start_day;
- SELECT distinct start_day, COUNT(start_day) FROM february_2021 WHERE member_casual like "member" GROUP BY start_day ORDER BY start_day;

Q5.1. Difference between riders on weekdays and weekend days?

To find riders on weekdays

- SELECT count(start_day) FROM february_2021 WHERE start_day between 0 and 4;

To find riders on weekends.

- SELECT count(start_day) FROM february_2021 WHERE start_day between 5 and 6;

Q6. Total number of riders per cyclist type?

Total rides per cyclist type

- SELECT distinct member_casual, count(start_day) FROM february_2021 GROUP BY member_casual;

Q7. Total rides per month for each cyclist type?

Total rides per cyclist type every month

- SELECT distinct start_month, count(member_casual) FROM february_2021 GROUP BY start_month;

Q7.1. Total riders for each season?

Total riders for each season

- SELECT count(start_month) as Spring FROM february_2021 WHERE start_month between 3 and 5 GROUP BY member_casual;
- SELECT count(start_month) as Summer FROM february_2021 WHERE start_month between 6 and 8 GROUP BY member_casual;
- SELECT count(start_month) as Fall FROM february_2021 WHERE start_month between 9 and 11 GROUP BY member_casual;
- SELECT count(start_month) as Winter FROM february_2021 WHERE start_month between 12 and 2 GROUP BY member_casual;

Q8. Average ride duration per cyclist type?

To find average ride duration per cyclist type

- SELECT distinct member_casual, avg(trip_duration) FROM february_2021 GROUP BY member_casual;

Q9. What is the usage of each bicycle type?

Total usage of each bicycle type

- SELECT distinct rideable_type, count(rideable_type) FROM february_2021 GROUP BY rideable_type;

Q10. Rides per hour for each cyclist type?

Total rides per hour for each cyclist type

- SELECT distinct HOUR(start_time) as start_hour, count(HOUR(start_time)) as total FROM february_2021 GROUP BY HOUR(start_time);

- There was more participation of casual riders on weekends.
- Casual riders tend to ride more in warmer months of Chicago, namely July and August.
- All riders tend to ride for a longer time in the cold months of Chicago.
- Casual riders spent on average a lot longer time per ride.
- Member riders did not prefer the docked bikes.
- There is a surge of riders in the evening around 5 PM.

Share: Supporting visualizations and key findings

#GrowWithGoogle

TABLEAU: https://public.tableau.com/shared/KNQM4WSNY?:display_count=n&:origin=viz_share_link

EXCEL: Sample Cyclist DA.xlsx

SLIDES: **PPT - Cyclistic Case Study.pdf**

CREDLY : https://www.credly.com/users/tushar-jawale.0b5d4935

Act: Your top three recommendations based on your analysis

- Introducing plans that may be more appealing to casual riders in summer months.The marketing should be done in winter.
- In order to gain more members the pricing structure can be altered in order to make single use more costly to casual riders and also lowering the long term membership rate.
- Different membership rates specifically for weekend users and summer riders can also help to improve total memberships.