CASE STUDY

How Does a Bike-Share Navigate Speedy Success?

INTRODUCTION

Cyclistic is a bike-sharing company in Chicago. The director of marketing believes the company's future success depends on increasing the number of annual memberships. The marketing team is aiming to understand the differences in bike use between casual riders and members.

Ask

How do annual members and casual riders use Cyclistic bikes differently?

Business Goal: Designing new marketing strategies aimed at converting casual riders into annual members.

Problem: Casual riders are not likely to upgrade their membership to annual memberships

Insight Advantages: Get more casual riders to upgrade their membership and possibly get more members by finding patterns in bike usage metrics.

Business Task: Finding similarities and differences between annual and casual members when using Cyclistic bikes.

Key stakeholders

• Cyclistic executive team

• Lily Moreno: director of marketing and manager

• Cyclistic marketing analytics team

PREPARE

Data Location: Downloaded onto the local desktop.

Data Organization: zip files dated by year and month. The file size is shown as well.

Issues with bias or Credibility:

- RELIABILITY- Data was collected on different days each month. There is data duplicated. There are missing data entries.
- ORIGINALITY-collected by the company itself
- COMPREHENSIBILITY-inadequately researched
- CURRENCY- Use without commercialization prohibited
- CITATION- https://divvy-tripdata.s3.amazonaws.com/index.html

Licensing: The data was made available through the license on this link https://ride.divvybikes.com/data-license-agreement. It belongs solely to the City of Chicago.

Privacy: Public use on condition of Data Anonymization. Forbidden use of Personally Identifiable Information (PII)

Security: Passwords and cloud storage

Accessibility: (Open-source data) free access, usage, and sharing of data.

STEPS IN DATA INTEGRITY VERIFICATION

1. Checking Metadata and file information.
2. Importing data into databases or spreadsheet tools like Excel or Google Sheets.
3. Checking data comptonization like replication, transfer, or manipulation.
4. Continuously saving to prevent loss of progress.
5. Prevent unauthorized access.
6. Checking for malware using dedicated antiviruses.
7. Updating software regularly to fix errors and patch updates.
8. Hypothesis testing.

The question will be answered once I finished with the verification and data cleaning. I might start getting other ideas that may address other business problems.

PROBLEMS

1. Too much data and inconsistencies in some naming processes.
2. Using data just from the previous year might not let me be able to distinguish other problems or underlying issues.
3. Mismatched expectations.

4. Incomplete or misaligned data.
5. Sample size should be representative of the whole population for the whole year. The issue might be some users only utilize the bikes for several months and then stop completely.
6. Casual users might move cities.
7. Weather and other issues might result in some routes not being taken or people opting out of bike rides in some seasons.

Deliverable

The data source used is a new database comprised of data from excel.

PROCESS

<u>Tools</u>

Excel workbooks.

<u>About Ensuring Data Integrity</u>

Yes, I checked for consistency when data was being transferred from Excel files into the Database. I checked the column names and also the tables and also row width.

<u>Steps taken to ensure Data is Clean</u>

1. Removing leading, trailing and repeated spaces in data using the TRIM() FUNCTION in Excel.
2. Using MID FUNCTION to find out if certain columns such as ride_id have the equal number of characters.

<u>Data is clean and ready to analyze when:</u>

1. There is no missing data or blank cells.
2. Full documentation of data cleaning effort.
3. There is no duplicated data.

NB: Documentation attached of the cleaning process.

Deliverable

Documentation of any cleaning or manipulation of data

ANALYZE

1. Combine all workbooks for different seasons into a new workbook called Divvybikes.xls.
1. Save the file and duplicate the procedure for the rest of the workbooks.

Most of the workbooks were too large to be merged using Excel's PowerQuery so SQLITE was used instead to merge them into one file.

1 Rename files to the CSV files that can be read by the database for later querying ease eg 2021-10_divvybike-data.csv was renamed to oct2021.csv.

2 Create New Database and save it as divvybikes.db

3 Create a new table "divvybikes" from the separate and renamed CSV files.

4 Set datatypes for each column to Not Null and the Ride ID to be Unique.

Excel for cleaning and SQLITE for combining processed EXCEL files into one document. Then export it into a workbook.

Ensuring data's integrity

Yes, I checked for consistency when data was being transferred from Excel files into the Database. I checked the column names and also the tables and also row width.

Steps taken to ensure that your data is clean

Steps like

Sort by Ascending Order each column and then Filtered by blanks for each column.

Deleted rows that had incomplete data in said view after Filtering.

Removing leading, trailing, and repeated spaces in data using the TRIM() FUNCTION in database.

SQL Queries

1. DELETE FROM divvybikes WHERE start_station_name IS NULL OR trim(start_station_name) = '';
2. DELETE FROM divvybikes WHERE ride_length IS NULL OR trim(ride_length) = '';
3. DELETE FROM divvybikes WHERE member_casual IS NULL OR trim(member_casual) = '';

Table rows were reduced from approximately 10000000 rows to 4597721.

Export into Excel.

EXCEL functions

For December 2021 WINTER excel workbook:

1. Check and Remove blanks again if they are there.
2. Remove duplicates as well.
3. I calculate

>   Maximum length of ride: =MAX(E1:E176372) = 2 hours 40 minutes
>   Mode of the day of week:5 which is Thursday (Most frequently occurring number ie day in this case)=MODE(F1:F176372)
>   Median ride length: 9 minutes
>   >   Formula: =MEDIAN(E1:E176372)
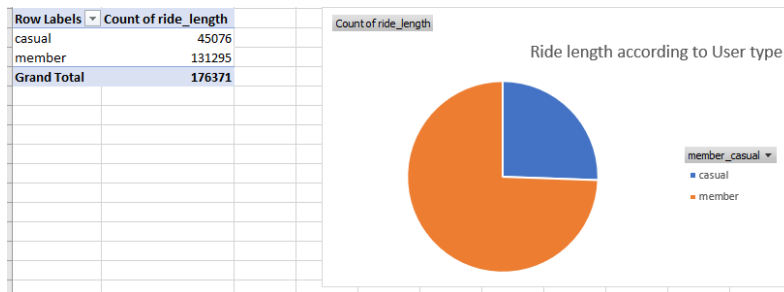>   Mode of the length of ride: =MODE.SNGL(E1:E176372) =6minutes
>   *Mean ride length = AVERAGE(e1:E176372)= 0:14:14

NB: 1 is Sunday and 7 is Saturday.

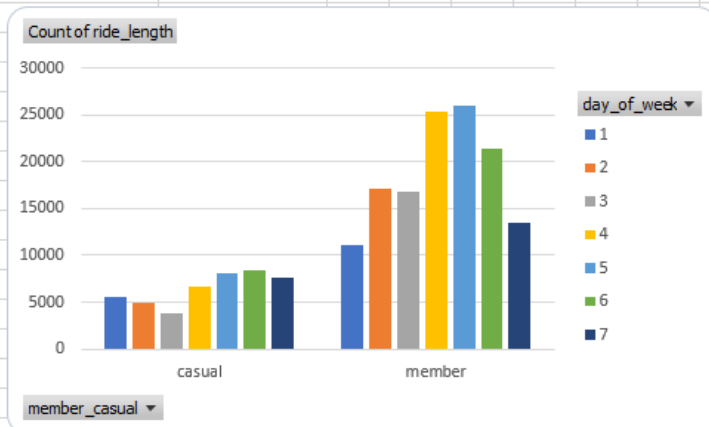## PIVOT TABLE for December 2021(WINTER)
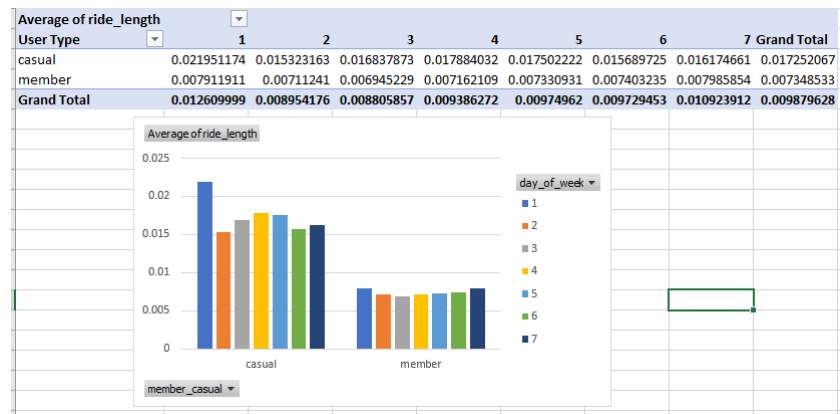
Total Ride Length according to Type of Subscription

Row Labels	Count of ride_length
casual	45076
member	131295
Grand Total	176371

Total Ride Length Per Day per User Type

Count of ride_length									
User Type		1	2	3	4	5	6	7	Grand Total
casual		5552	4927	3898	6652	8125	8358	7564	45076
member		11039	17038	16827	25415	26043	21415	13518	131295
Grand Total		16591	21965	20725	32067	34168	29773	21082	176371

Average Ride Length for users by day of week

Average of ride_length									
User Type		1	2	3	4	5	6	7	Grand Total
casual		0.021951174	0.015323163	0.016837873	0.017884032	0.017502222	0.015689725	0.016174661	0.017252067
member		0.007911911	0.00711241	0.006945229	0.007162109	0.007330931	0.007403235	0.007985854	0.007348533
Grand Total		0.012609999	0.008954176	0.008805857	0.009386272	0.00974962	0.009729453	0.010923912	0.009879628

MARCH 2022 spring

EXCEL functions

For March 2022 excel workbook:

1. I calculate

    Maximum length of ride: =MAX(e1:E215984)=20:34:04

    Mode of the day of week: which is =MODE(F1:F215984)=4 which is Wednesday
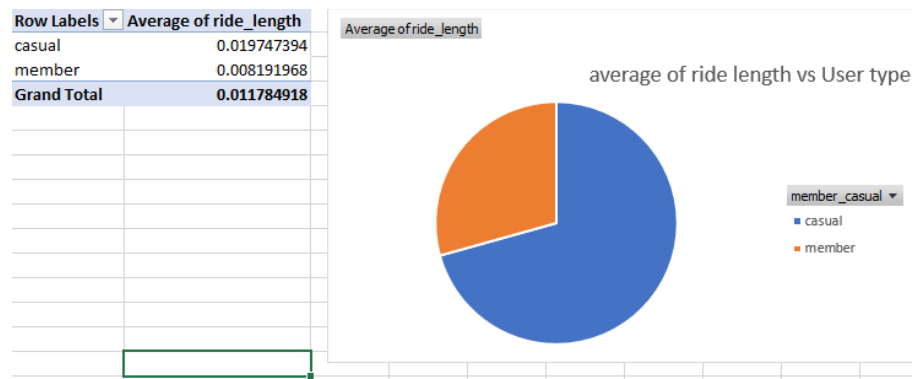
    Median ride length: 0.006782407 minutes

        Formula: =MEDIAN(e1:E215984)

    Mode of the length of ride: =MODE.SNGL(e1:E215984) = 05:12 minutes

    *Mean ride length = AVERAGE (e1:E215984) =0:16:58

PIVOT TABLE for March 2022(spring)

Average ride length vs User Type

Row Labels	Average of ride_length
casual	0.019747394
member	0.008191968
Grand Total	0.011784918

average ride_length for users by day_of_week

Average of ride_length	Column Labels							
Row Labels	1	2	3	4	5	6	7	Grand Total
casual	0.022039604	0.021716552	0.014933202	0.018763606	0.015996064	0.014629842	0.024065935	0.019747394
member	0.009434099	0.008637012	0.007431917	0.008108692	0.007405296	0.007571806	0.009591062	0.008191968
Grand Total	0.014965134	0.012867575	0.009019125	0.01106454	0.0096175	0.009297267	0.016035206	0.011784918

average ride_length for users by day_of_week

number of rides for users by day_of_week

JUNE 2022 SUMMER

EXCEL functions

For June 2022 excel workbook:

1. I calculate

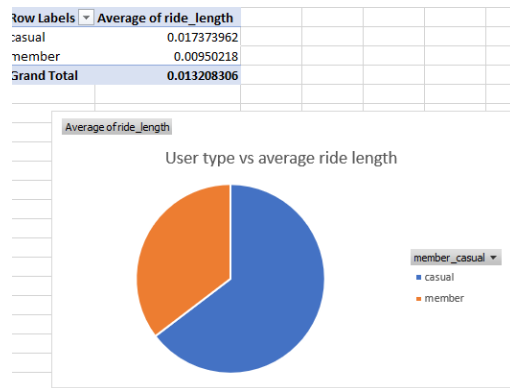   Maximum length of ride: =MAX(E1:E620351) = 15:12:52
    Mode of the day of week: (Most frequently occurring number ie day in this case)=MODE(f1:f620351)=5 WHICH IS thursday
    Median ride length Formula: =MEDIAN(E1:E620351)= 0:12:08
    Mode of the length of ride: =MODE.SNGL(e1:E620351) =0:06:38
    *Mean ride length = AVERAGE(e1:E620351)= 0:19:01

NB: 1 is Sunday and 7 is Saturday.

Row Labels ▼	Average of ride_length			
casual	0.017373962			
member	0.00950218			
Grand Total	0.013208306			

Average of ride_length

User type vs average ride length

member_casual ▼
- casual
- member

SEPTEMBER 2022 FALL

EXCEL functions

For September 2022 excel workbook:

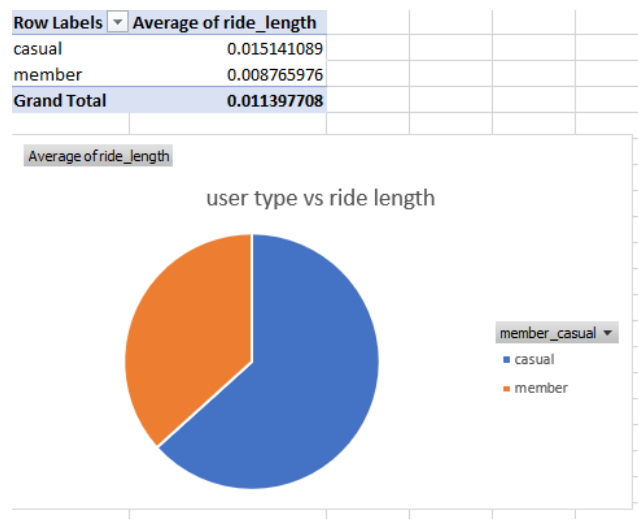Maximum length of ride: =MAX(E1:E535146) = 0:55:00

*Mode of the day of week: (Most frequently occurring number ie day in this case)=MODE(F1:F535146) =6 which is Friday
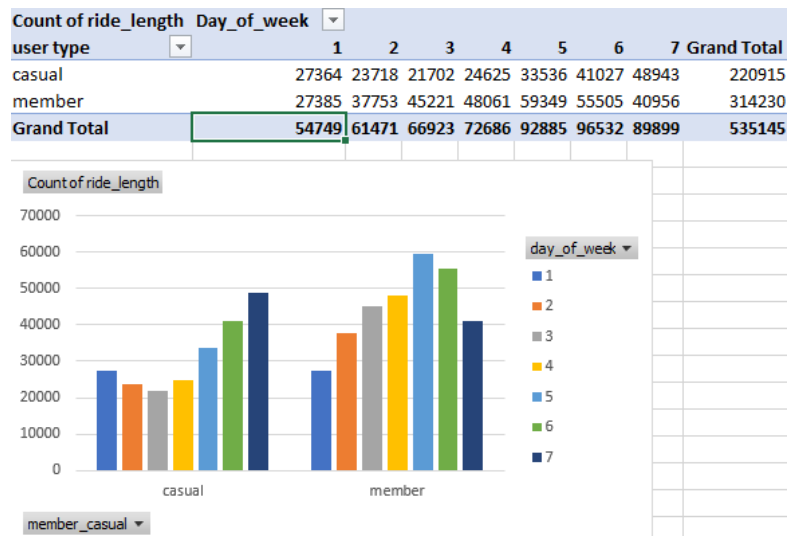Median RIDE length Formula: =MEDIAN(E1:E535146)= 0:10:00
Mode of the length of ride: =MODE(E1:E535146) =0:06:00
*Mean ride length =AVERAGE(e1:E535146)= 0:16:25
NB: 1 is Sunday and 7 is Saturday.

Row Labels ▼	Average of ride_length			
casual	0.015141089			
member	0.008765976			
Grand Total	0.011397708			

Average of ride_length

user type vs ride length

member_casual ▼
- casual
- member

Count of ride_length	Day_of_week							
user type	1	2	3	4	5	6	7	Grand Total
casual	27364	23718	21702	24625	33536	41027	48943	220915
member	27385	37753	45221	48061	59349	55505	40956	314230
Grand Total	54749	61471	66923	72686	92885	96532	89899	535145

ACT

Key findings

1. More collective minutes spent by Casual riders during rides in every season as compared to Members.
2. The highest number of trips is in Fall followed by Summer.
    The lowest is in Winter followed by Spring.
3. More casual riders are using the bikes during any given day of the week in all seasons as compared to members.

Recommendations

✓ Encourage members to join by marketing discounts on trips.
✓ Marketing should also focus on selling season-inspired wearables to casual riders or give discounts on warm exercise clothes in Winter and Spring.

✓ Create an app that gives reminders on daily exercise with a built-in subscription for casual riders and free for members.

Conclusions

These 3 recommendations will definitely encourage a large number of casual riders to join the annual subscription as members.