DIVVY CHICAGO

Author : Frederic Petit ©2023

Executive Summary

- Bike Sharing is not a monolithic thing.
- It is made of “segments”, “profiles” or “types” of many things :
  - Types of stations,
  - Profiles of usages,
  - Segments of riders ….
- This is what we are exploring in this capstone project with at the end a specific focus on the casual vs. member business question of the project
- By identifying the segments of this market, we think to be giving the company directions where efforts to develop the business can be pushed forward

Outline

Part I : Overview

Part II : (Divvy) Infrastructure

Part III : Riders Experience

Part IV : Data Engineering considerations

Let's Talk Big !!!

Your System :

- Stations
- Bikes
- Riders
- Trips

Your Big Numbers :

- 600+ Stations
- 90K trips a month
- 50K Hours ride a month

Dashboard presenting your data

2 Faces : Infrastructure / Customer

Infrastructure

- Stations
- Traffic
- **Controls the cost**
- Where cost reduction may happen
- Optimization/Rationalisation

Challenge : Make the administration/management more efficient, more cost effective

Consumer Experience

- **Determines Revenues**
- Improved rider experience should boost revenues
- Trip Length
- Favorites Stations
- Attractivity (Member vs. Casual)
- What we don’t have is Loyalty information for the members

Challenge : Make the rider experience more enjoyable to attract more rides and riders

Infrastructure

Controlling Your Costs

All Stations Are NOT Equal

“**Departing**” vs. “**Arriving**” :

- Stations by Traffic **Volumes**
  - High Volume Traffic
  - Low or nil traffic volumes
  - Seasonality
- **Deficit/Surplus** (more “arriving” trips than “departing” ones or inverse): asymmetry
  - Structural/Random/Variability
  - Seasonality

All Stations Are NOT Equal

"**Departing**" vs. "**Arriving**" Stations

- High **Volumes** of Rides and Riders
- **Deficit/Surplus** of Bikes

Overall Use of the Stations

Top 20 “Departing” stations

Top 20 “Arriving” stations

Recommendations for Stations management

- **Low volumes Stations** : Make possible hard decision to close them or some of them
- **High Volumes** Stations: Provide higher level of service.
- **Management of the Asymmetry** : JIT
  - to reduce the overall size of inventory
  - to optimize the transferts from Excess to Deficit stations

Stations Asymmetry

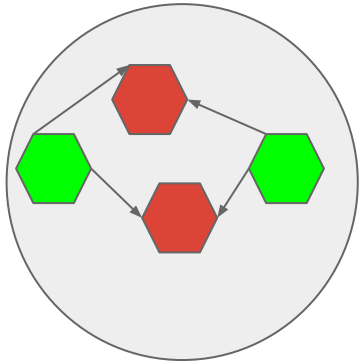
Top 20 “Surplus” stations

Top 20 “Deficit” stations

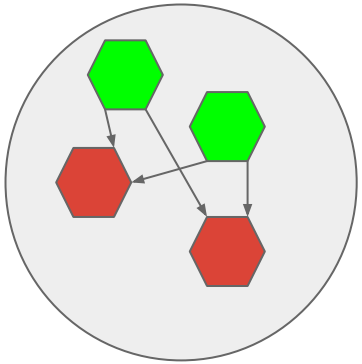
Optimizing the transfers:
- Grouping stations by sectors
- Limiting the transfer to within a sector

Map of Surplus and Deficit Stations

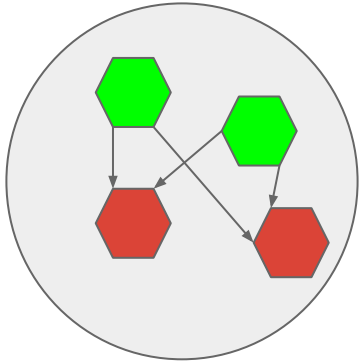
Sector 1



Sector 3



Sector 2



Surplus station

Surplus station

Traffic Per Neighborhoods

MostPopularRoutes5$Neighborhood.From <chr>	RidesFromNeighborhood <int>
Lincoln Park	1858
Northalsted	1769
Magnificient Mile	1686
Gold Coast	1355
River North	1335
West Loop	698
Lake View	554
Wicker Park	469
Metra Ogilvie	398
UnKnown	397

Rider Experience

Controlling Your Revenue

Approach : Types Of Riders and Rides - Use cases

USE-CASES (from the data):

- “Loop” ride (Return to the SAME starting station) vs. “Commute” ride (Return to a different starting station)
- Casual vs Member Riders

USE-CASES (requiring more data):

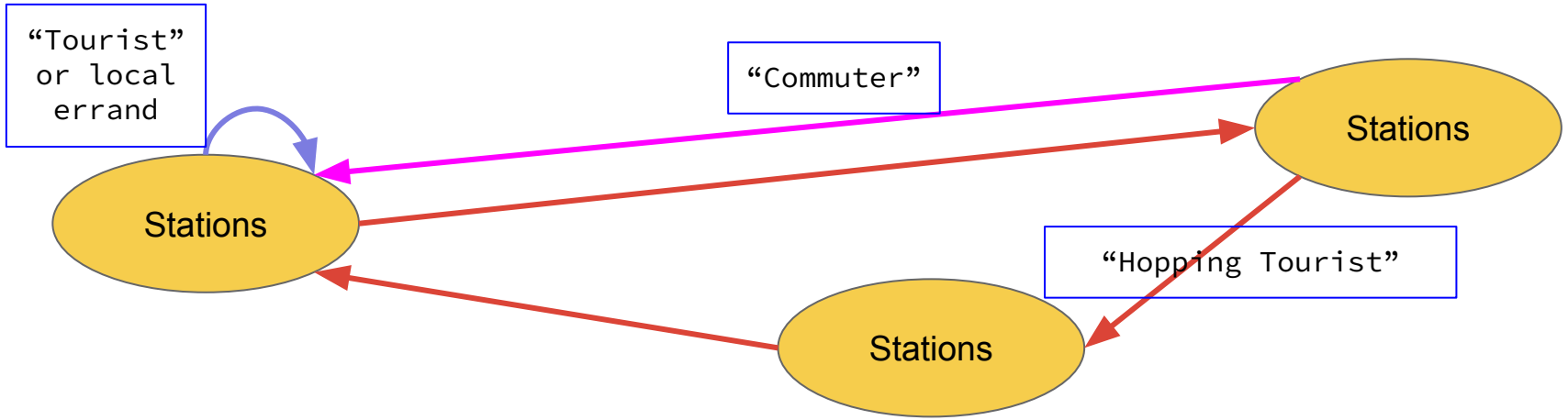
Identifying more use-cases requires additional investigation and data Enrichment :

- Purpose of the ride (pleasure, practical, professional etc…)
- Type of Riders (student, couriers, tourist etc…)
- Use per Riders (total time spent on bike etc…)

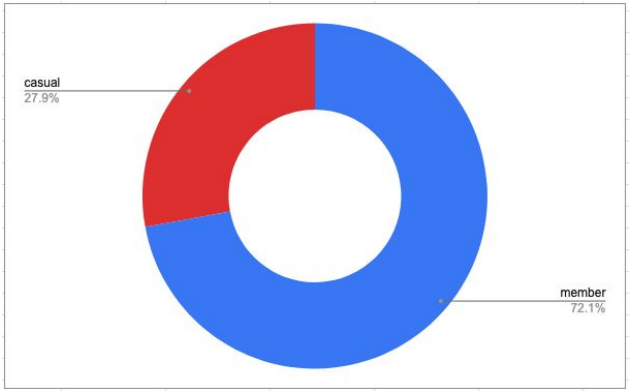
Are All Riders the Same ?

Distinguishing groups and types of Riders to “understand” them

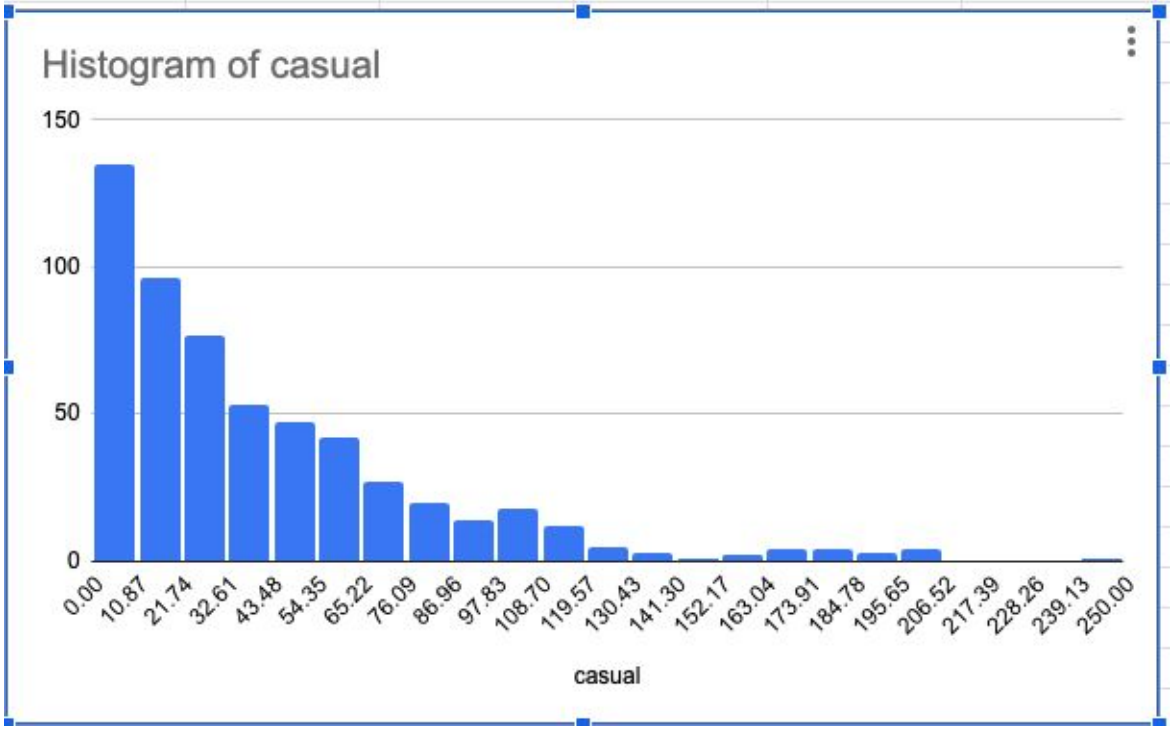
- One you know well already : “Member” vs. “Casual”
- Another is the type of trip w/ respect to start and finish station



Casual vs. Member

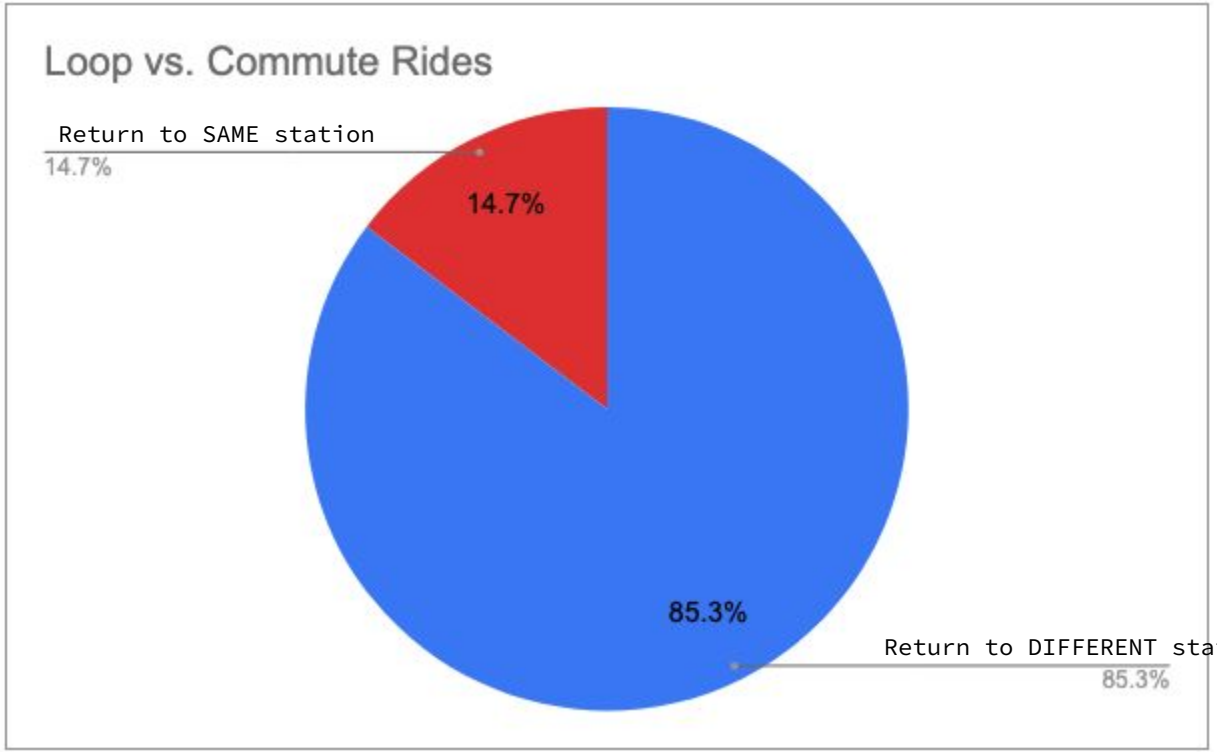


Nbr of stations for classes of Casual members

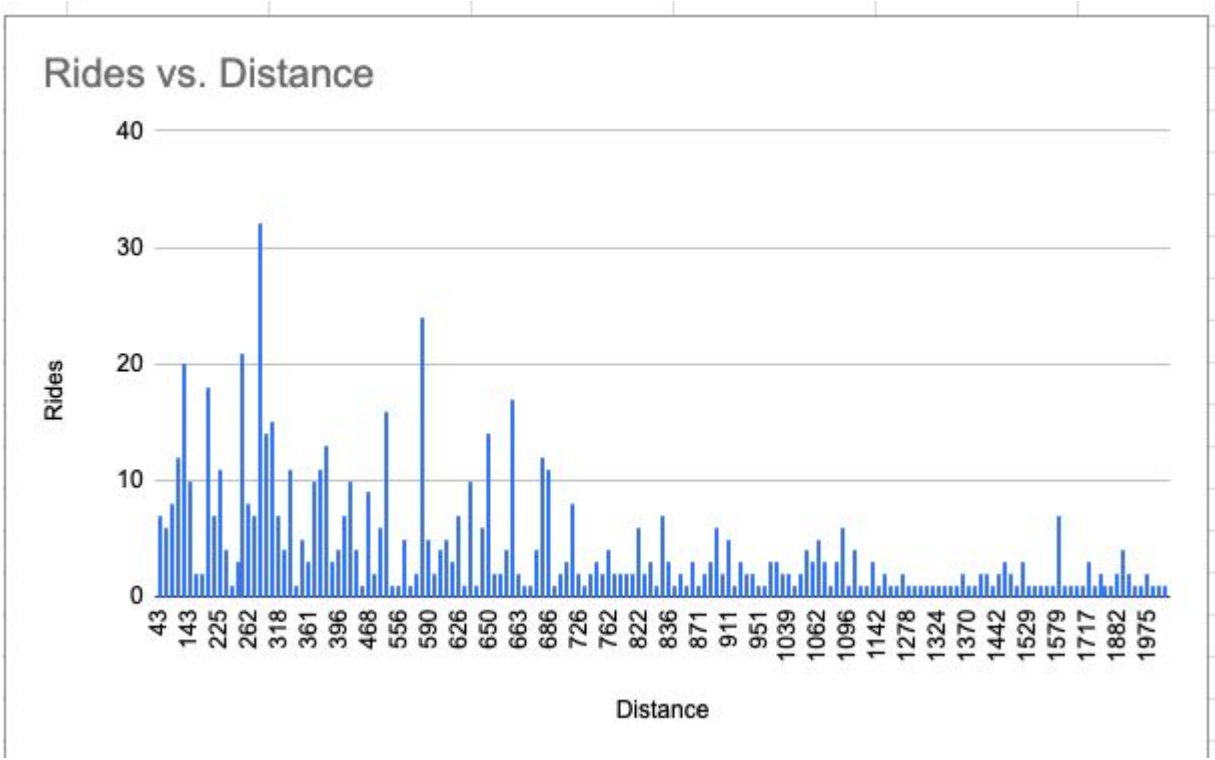


Stations with high level of “casual” likely to be “touristic” where it may be worth providing dedicated additional services

"Loop" vs. "Commute" Rides



Distances made by Riders



Roughly, the longer the distance (“From-To” station) the less Riders there is.

But it may be possible to distinguish **short-distance riders** to **long-distances riders**. For example students (dorm-classes) or couriers and local traffic.

It may be worth adding some specific services to those types of riders

Most Popular Rides By Neighborhoods

MostPopularRoutes5$Neighborhood.From <chr>	MostPopularRoutes5$Neighborhood.To <chr>	RidesPerNeighborhoodFromTo <int>
Northalsted	Northalsted	814
Lincoln Park	Lincoln Park	724
Magnificient Mile	Magnificient Mile	636
River North	River North	538
Gold Coast	Gold Coast	453
Northalsted	Lincoln Park	253
Lincoln Park	Northalsted	248
West Loop	West Loop	242
Gold Coast	Lincoln Park	235
Lincoln Park	Gold Coast	222

1–10 of 436 rows

Prev

Riders Behavior Analysis

Different Behaviors –> Specific Services and Assistances

- Identify Behaviors
- Define Additional services for each behavior

Riders Behavior Analysis and Better Rider Experience

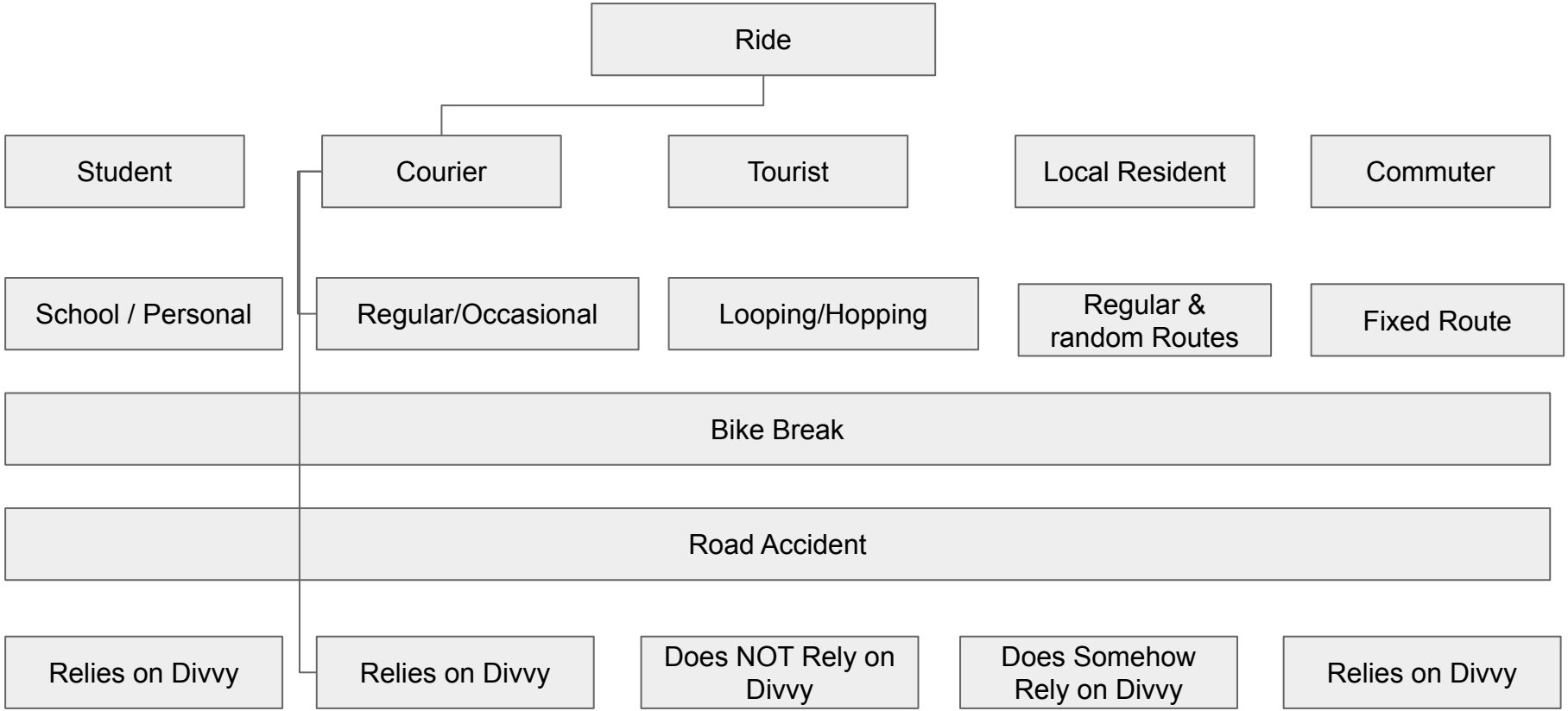
For example **a courier** may appreciate to have quick bike replacement in case the bike break

Bikes that never made to the destination station may be due **break of the bike** or **accident**. Increasing the safety of the riders could/should be improved

How to get more Insightful Data ?

- Leveraging the App (review, trip distance)
- In busiest stations put in place an interactive, friendly and fun human presence
- Characterizing stations
  - Tourist
  - Professional
  - Dwelling
  - Venues
  - Landmarks
  - Etc…
- Using Tree Method

How the Tree Method Can Help Figuring Behaviors ?



Insights and Recommendations

Insights :

- Riders have different motivations and needs
- Rides have different purposes
- Classes of :
  - Riders
  - Rides
  - Stations
- But, today only Limited insights

Recommendations :

- Increase visibility :
  - Use cases ?
  - Socio-economic ?
  - Feedback from riders ?
  - Well covered usages vs. badly covered use cases ?
- Design development plan :
  - Use cases To be promoted
  - Differentiate Stations to develop specific services and assistances

Casual vs. Member

Analysis

From data sample to full-blown data

Do Casual vs. Members have a preference for the type of Bike ?

member_casual	classic_bike	docked_bike	electric_bike	Grand Total
casual	520	95	334	949
member	1009		443	1452
Grand Total	**1529**	**95**	**777**	**2401**

	classic_bike	docked_bike	electric_bike	Grand Total
Casual	54.79%	10.01%	35.19%	100.00%
Member	69.49%	0.00%	30.51%	100.00%

Note that for the month of June 2022 a very large number of records had the Starting station empty, and even though it's not needed for that analysis, we eliminated them. We have only 2401 rides for the month. The data needs be fixed.

Favorite Stations,departing

start_station_name	casual	member	Grand Total
Clarendon Ave & Junior Ter	134	136	270
May St & Taylor St	42	197	239
Blue Island Ave & 18th St	58	83	141
Clark St & Randolph St	73	65	138
Stockton Dr & Wrightwood Ave	56	64	120
Sheffield Ave & Kingsbury St	32	44	76
Bissell St & Armitage Ave	8	63	71
Montrose Harbor	39	31	70
Southport Ave & Waveland Ave	23	45	68
Halsted St & Maxwell St	12	44	56
Wolcott Ave & Polk St	5	48	53
Racine Ave & 18th St	13	35	48
Milwaukee Ave & Fullerton Ave	30	16	46
Clinton St & Jackson Blvd	14	32	46
Streeter Dr & Grand Ave	31	10	41
Broadway & Wilson - Truman College Vaccination Site	12	25	37

Favorite Stations, Arriving

end_station_name	casual	member	Grand Total
Sheffield Ave & Kingsbury St	137	256	393
Clark St & Randolph St	142	241	383
Clarendon Ave & Junior Ter	203	172	375
May St & Taylor St	112	228	340
Blue Island Ave & 18th St	76	115	191
Loomis St & Lexington St	6	116	122
Lincoln Ave & Waveland Ave	52	60	112
Michigan Ave & Oak St	65	42	107
Milwaukee Ave & Fullerton Av	50	35	85
Broadway & Barry Ave	31	51	82
Racine Ave & 18th St	13	57	70
Sheridan Rd & Irving Park Rd	23	40	63
Mies van der Rohe Way & Ch	7	14	21
Orleans St & Elm St	4	14	18
Broadway & Granville Ave	4	5	9
Cicero Ave & Quincy St	9		9

What can be done to convert Casual into Member

Casuals are made of segments we cannot really identified. We know that at least there 2 kinds of casuals : tourists and locals.

Converting tourists seems difficult

Converting locals should be the target. We can identify that group as the casual NOT starting at “for tourist” starting station and/or NOT ending in known “tourist” location.

The Stations information needs be enriched to provide that level.

Once this is done, we need to figure out what is dissuading local casual from membership. Here again additional information is needed like, feeling of convenience, frequency of use (is it worth it to take a membership.. etc…) all things we lack information about.

Data Engineering

Limited amount of data were processed for this capstone project. What if real project ?

From data sample to full-blown data

Scaling Up

Amount of Data for 12 months being between 700MB and 1 GB, it is too large for a laptop and/or Google Sheet. It why we limited ourselves to a sample (april 2020 & June 2022). Even July 2022 required massive clean up to get rid of massive amount of records with no Start Station information.

Options for Storing the large amount of Divvy Data :

- BigQuery and Google Analytics tool. But this assumes that the file format will change. If it does, BigQuery and standard RDBMS are not a solution anymore
- On prem like PostGreSQL, but same issue it the file format changes
- ELK - looking the best option, as being NoSQL and great Analytics GUI

The End

Last updated : June 29 2023

Contacts :

- frederic.guy.petit@gmail.com
- fred@context22.com
- info@context22.com
- www.context22.com