

# DIVVY CHICAGO

**Author : Frederic Petit ©2023**

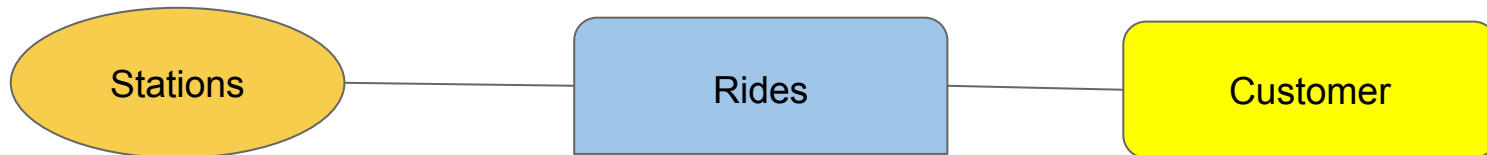
# LET'S TALK BIG !!!

## Your System :

- Stations
- Bikes
- Riders
- Trips

## Your Big Numbers :

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



# 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

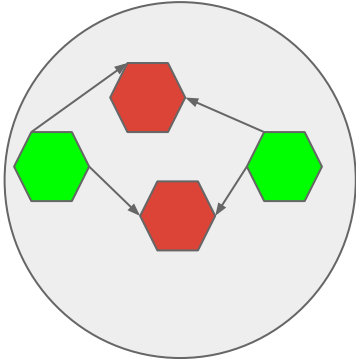
Top 20 “Surplus” stations

Top 20 “Deficit” stations

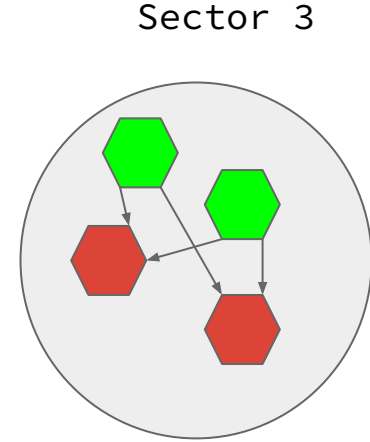
Optimizing the transfers:

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

# MAP OF SURPLUS AND DEFICIT STATIONS

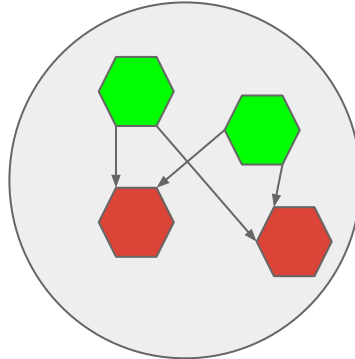


Sector 1



Sector 3

Sector 2



# TRAFFIC PER NEIGHBORHOODS

# 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):

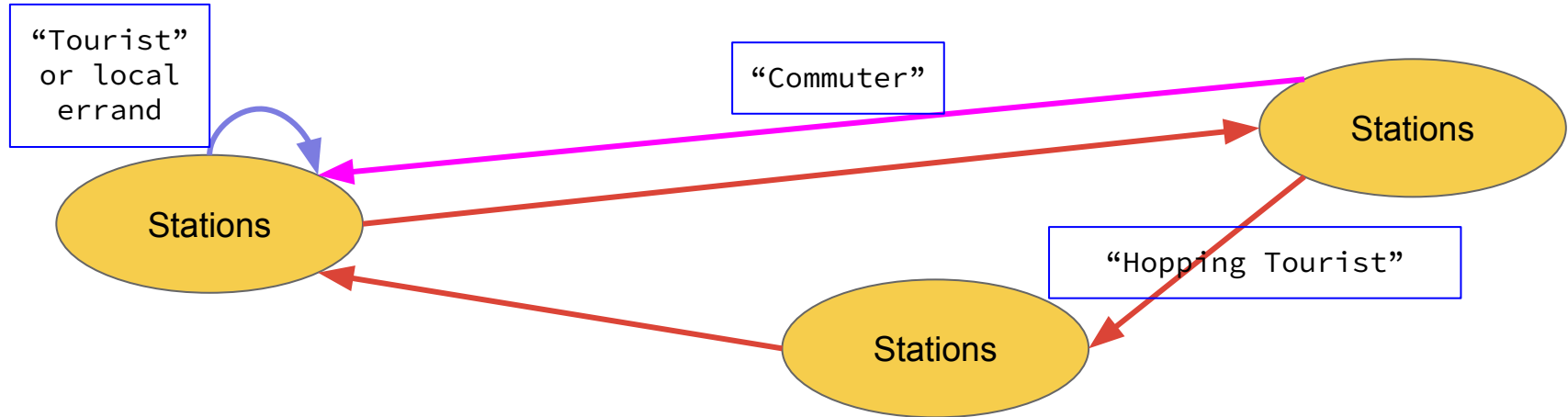
Identifying more use-cases requires additional investigation and data Enrichment :

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

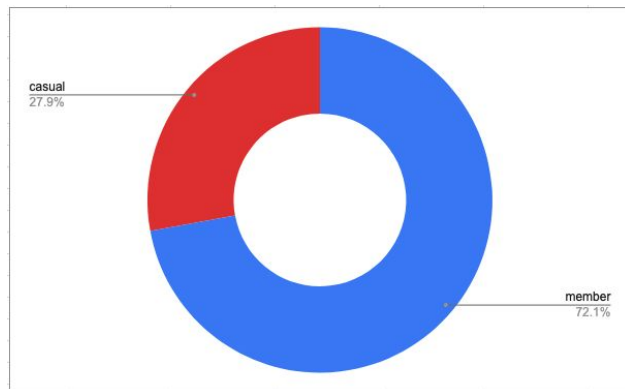
# ARE ALL RIDERS THE SAME ?

**Distinguishing groups and types** of Riders to “understand” them

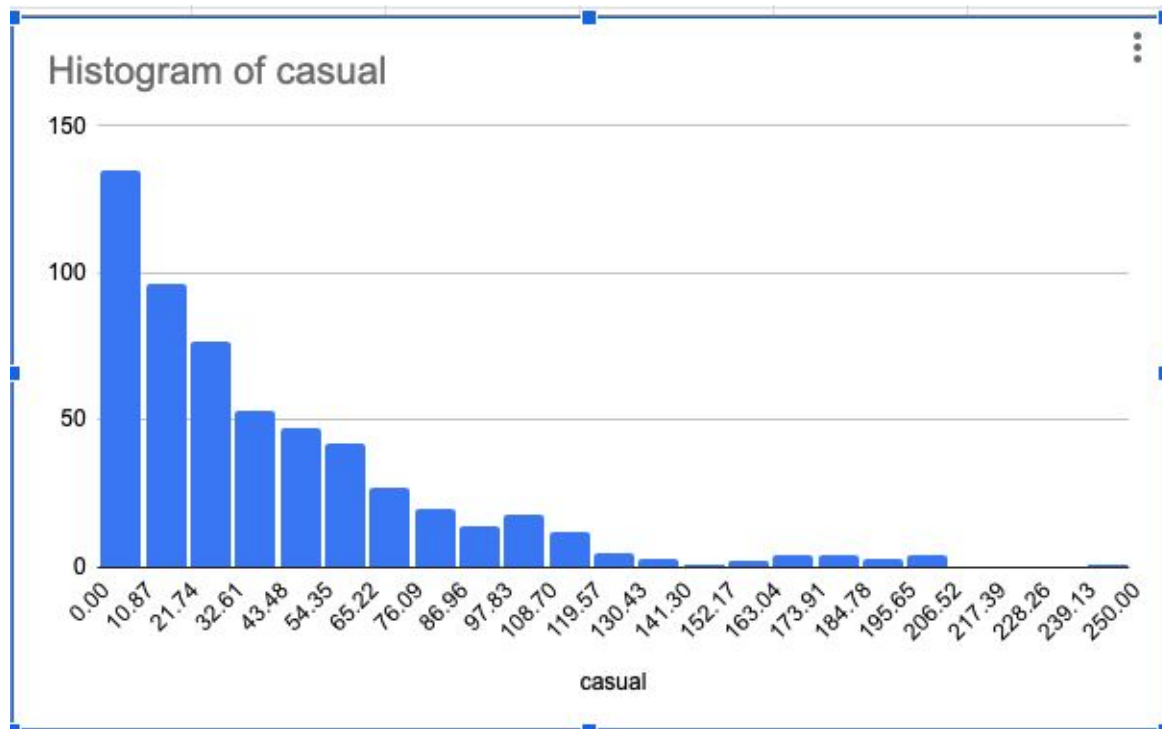
- One you know well already : “Member” vs. “Casual”
- Another is the type of trip w/ respect to start and finish station



# CASUAL VS. MEMBER

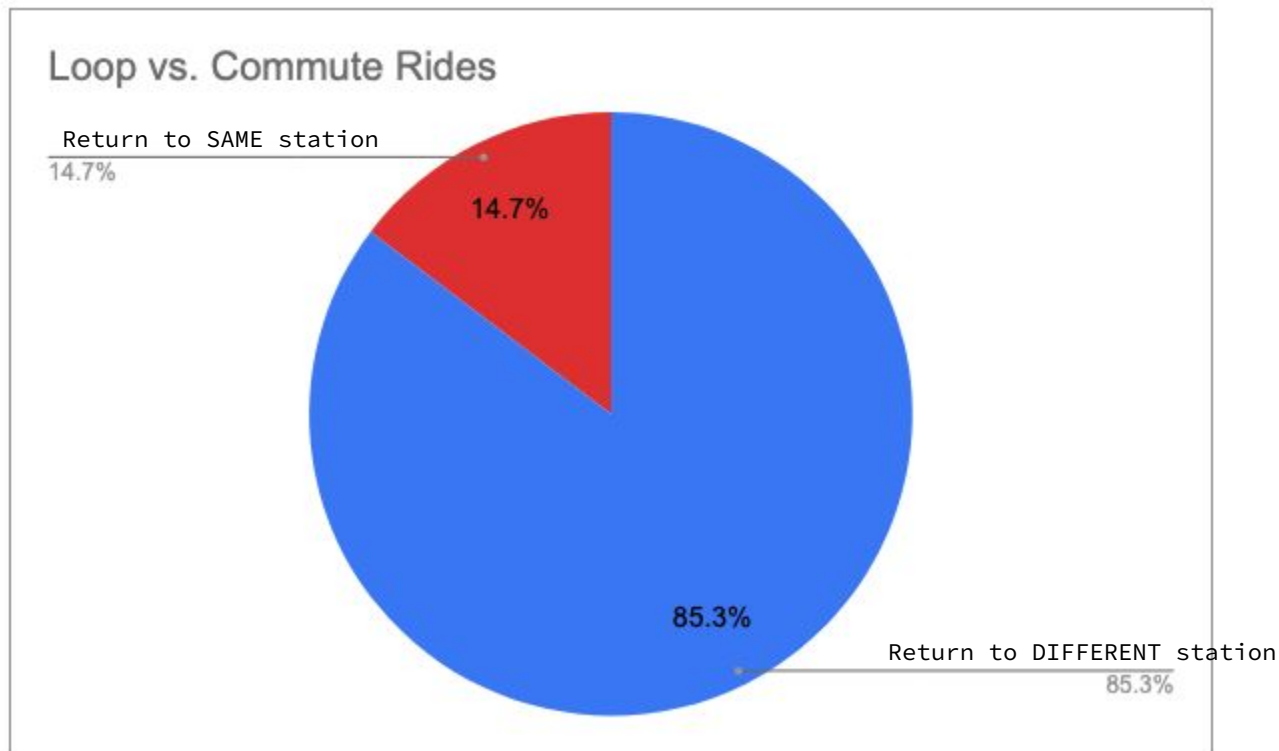


Nbr of stations for  
classes of Casual members



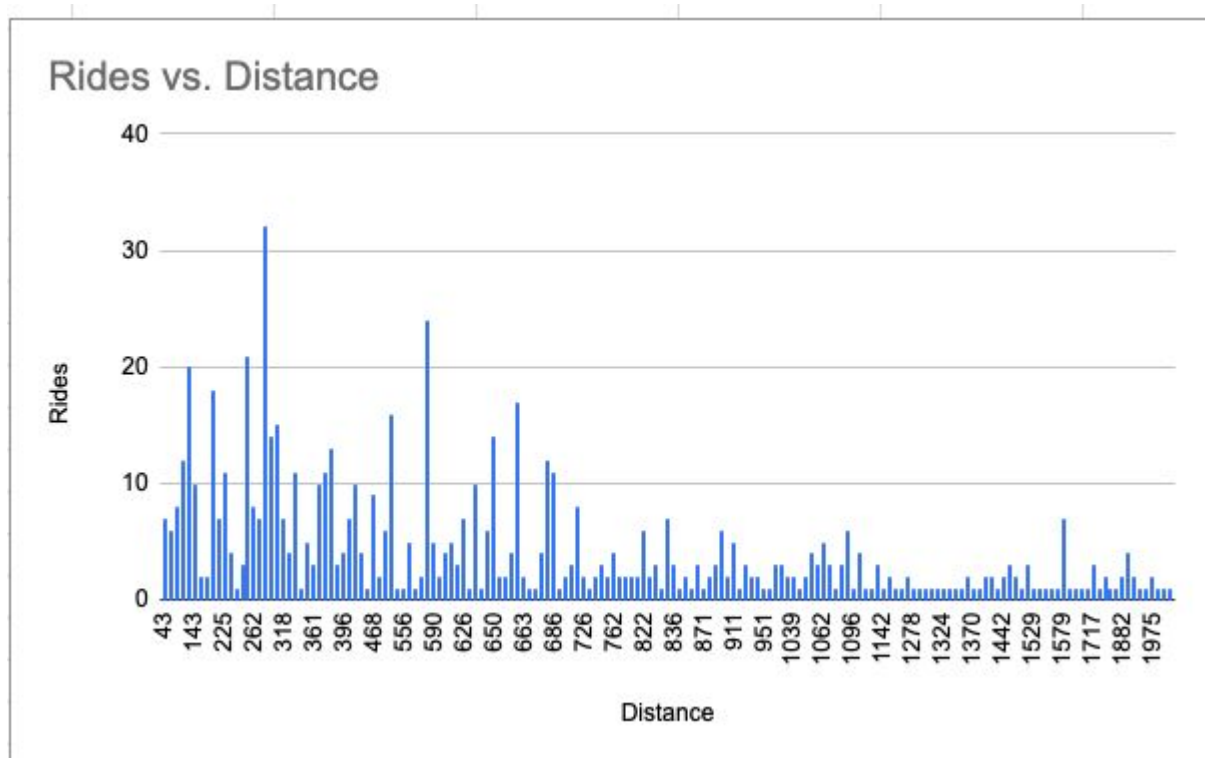
Stations with high level of “casual” likely to be “touristic” where it may be worth providing dedicated additional services

# "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

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

MostPopularRoutes5\$Neighborhood.From <chr>	MostPopularRoutes5\$Neighborhood.To <chr>	RidesPerNeighborhoodFromTo <int>
Northalsted	Northalsted	814
Lincoln Park	Lincoln Park	724
Magnificent Mile	Magnificent 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

# RIDERS BEHAVIOR ANALYSIS

**Different Behaviors → Specific Services and Assistances**

- Identify Behaviors
- Define Additional services for each behavior

# RIDERS BEHAVIOR ANALYSIS AND BETTER RIDER EXPERIENCE

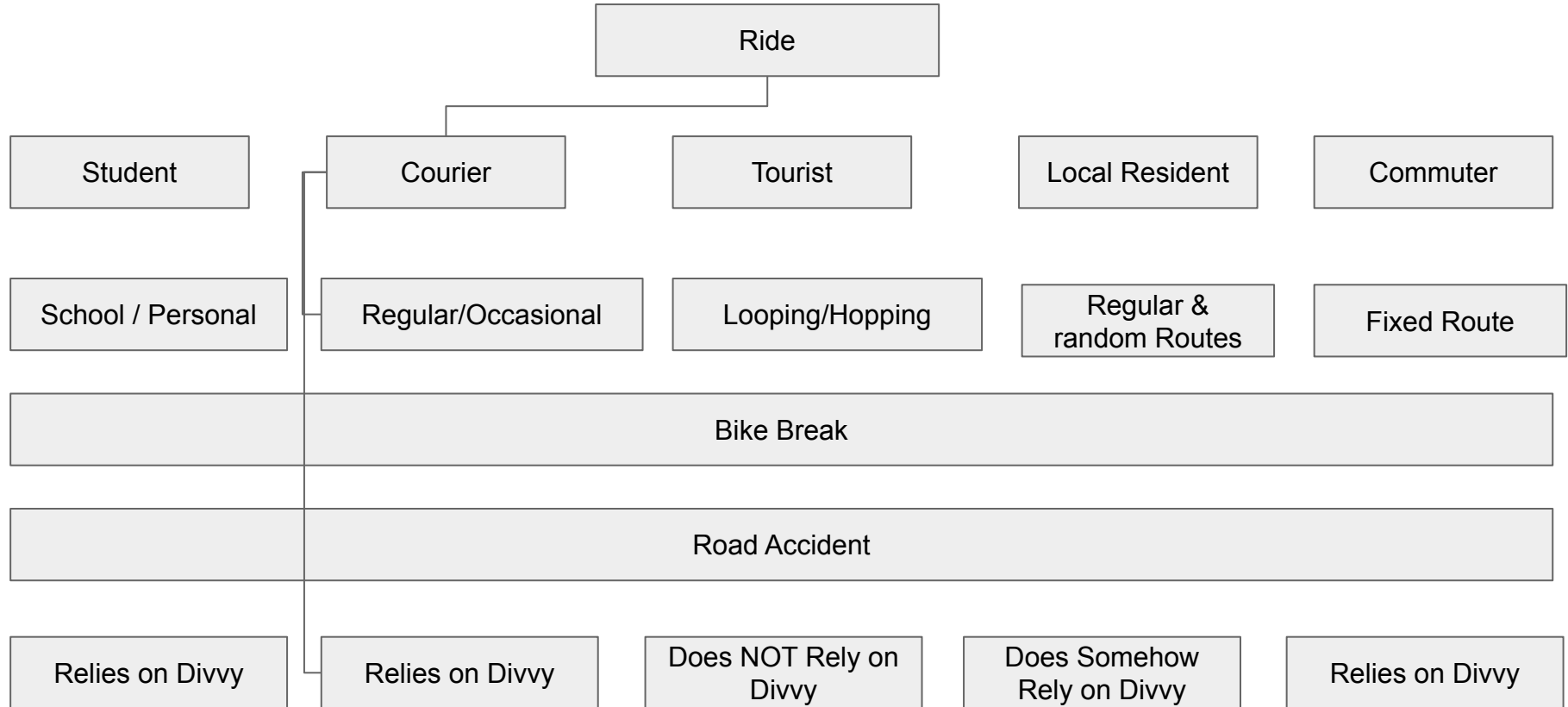
For example **a courier** may appreciate to have quick bike replacement in case the bike break

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

# THE END

Last updated : June 29 2023

## Contacts :

- [frederic.guy.petit@gmail.com](mailto:frederic.guy.petit@gmail.com)
- [fred@context22.com](mailto:fred@context22.com)
- [info@context22.com](mailto:info@context22.com)
- [www.context22.com](http://www.context22.com)



# 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

Options for Storing the large amount of Divvy Data :

-

# BACKGROUND SLIDES

---

# RIDE ASYMMETRY IN TIME

Why look at this features ?

- Moving bikes from excedent to deficit stations is needed
- Move should be efficient
- 600 stations is too big to be managed properly
- Need to group (cluster) them
- Recommended way to cluster : by geography

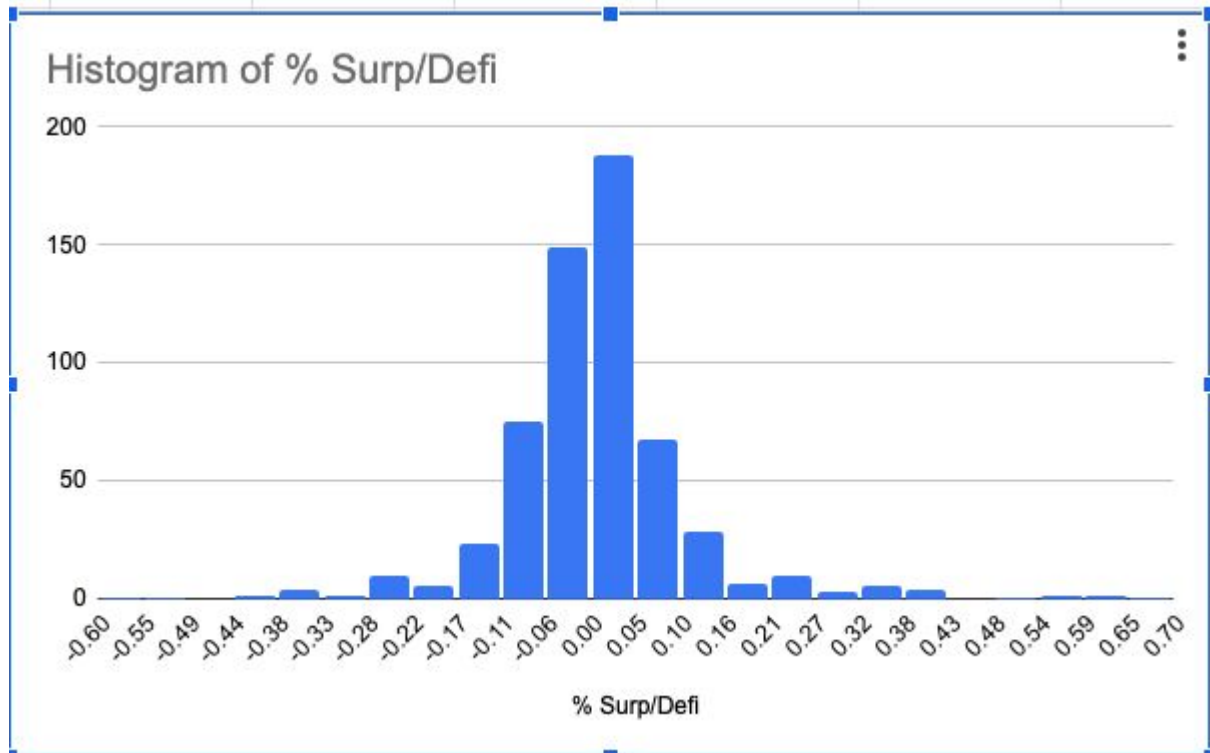
# APPROACH

Distinguishing among the riders and the rides different groups.

Figure out dimensions in the rides and riders:

- Reasons for a ride
- “Profile” of riders, for example, student vs. young professional, bachelors vs. families etc...
- Main issue : little amount of information

# BALANCE OF THE STATIONS - RIDES ASYMMETRY



# ARE ALL STATIONS AND TRIPS EQUAL ??

Should we consider all trips and station the same way ? Can we distinguish **“groups” of Stations ?**

Actually yes there are **differences** or **DIMENSIONS** :

A ride can be “departing” or “arriving”, therefore Stations can be seen as “Departing” or “Arriving”