DATA 698 Project v1

Predicting Citi Bike Availability in NYC

Philip Tanofsky

2022-11-16

Problem and Objective

Introduce yourselves and describe your problem. Explain your objectives, challenges of your work, proposed methodologies, and the assumptions you made while conducting modeling and/or analysis. Provide an overview of your approach and/or conceptual model (please do not present your code directly). Describe the results you obtain and summarize the current achievements and possibility of future works.

Challenges and Assumptions

- Too much data
  - Over 3.5 million trips in month of Sept. 2022
- Rebalancing identification
- User friendly approach
  - Inputs to output

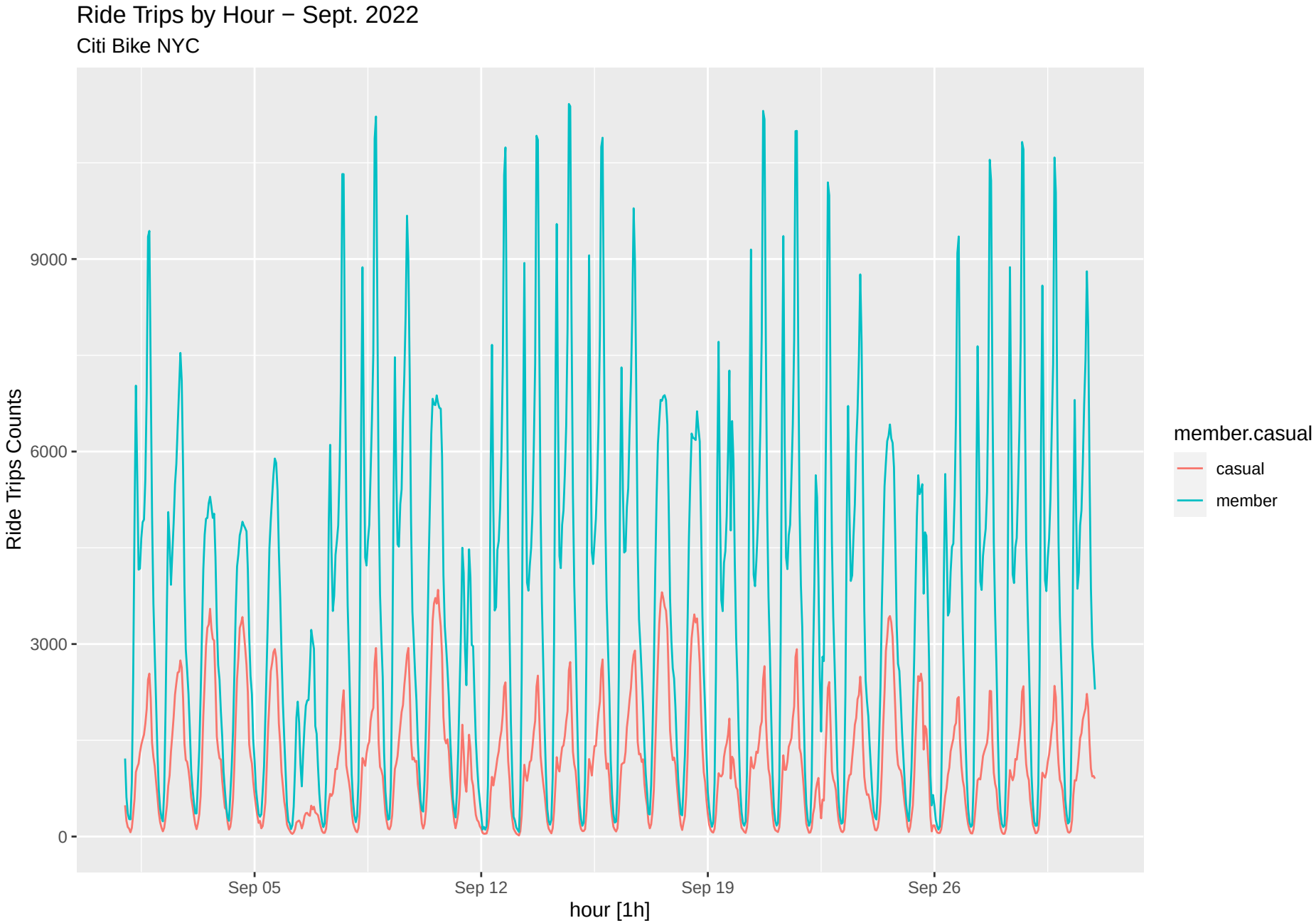
Proposed Methodologies

- Inputs
  - Latitude and longitude
  - Day of the Week
  - Time of Day
  - Number of Bikes
- Citi Bike offers live map of availability
- Lyft provides real-time availability
- Time series model
- Poisson distribution and Negative Binomial given the over-dispersion

Overview of Approach

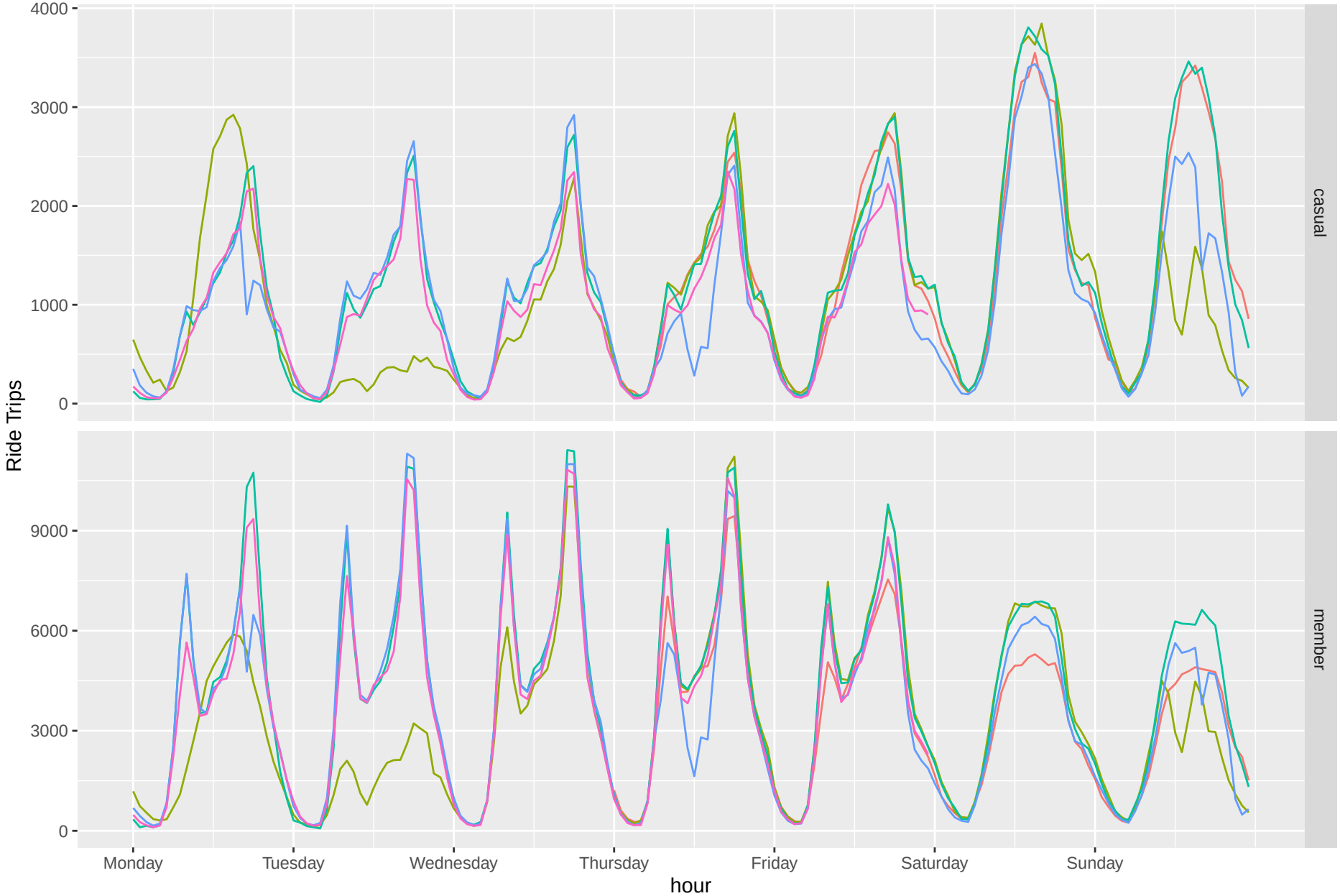
- Data from API call every 15 minutes for two weeks
  - Citi bike availability at each station
  - Two weeks is small interval to predict
    - ★ Valid limitation of model

EDA 1


Ride Trips by Hour − Sept. 2022
Citi Bike NYC
Ride Trips Counts
9000
6000
3000
0
Sep 05
Sep 12
Sep 19
Sep 26
hour [1h]
member.casual
casual
member

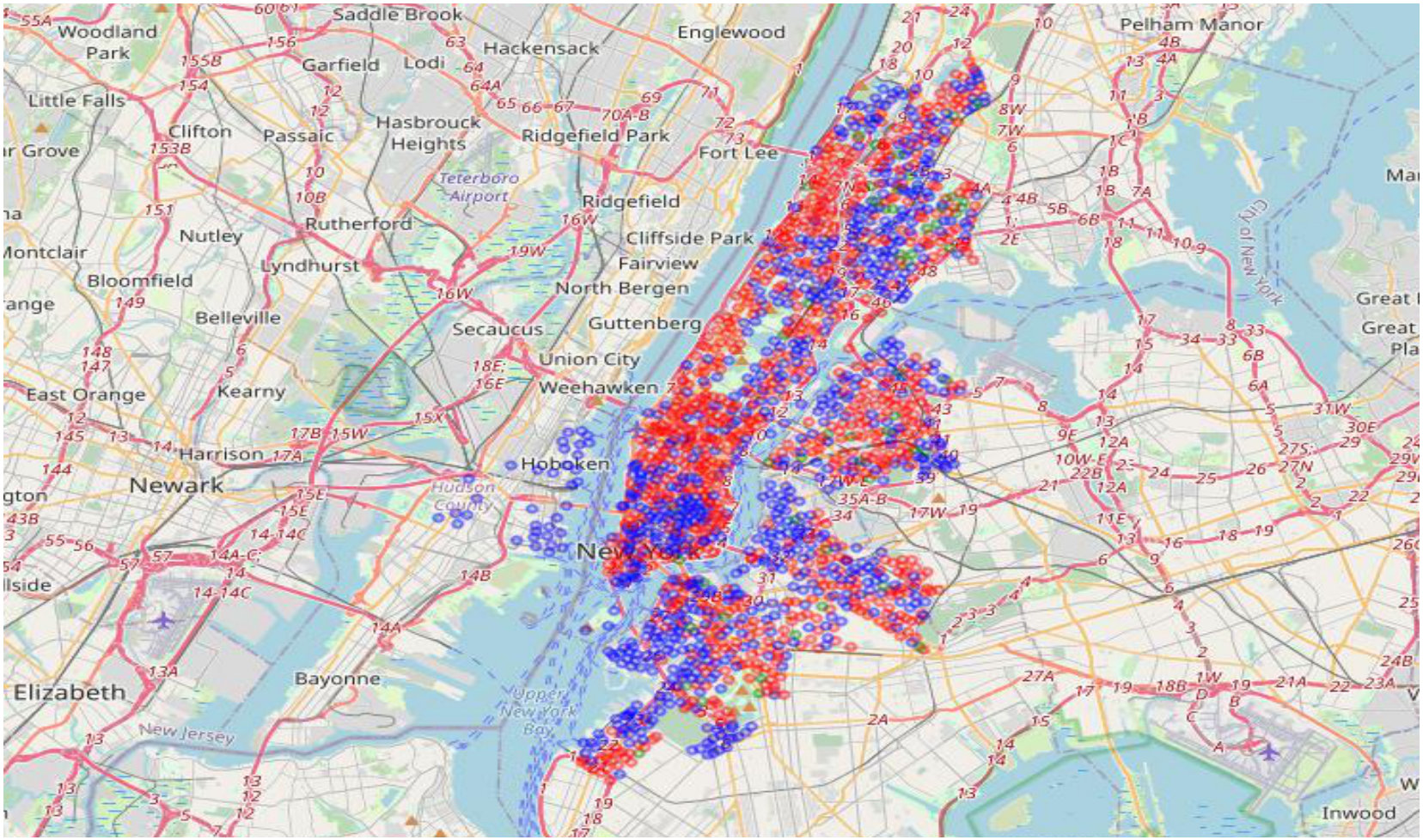
EDA 2

Seasonal plot: Weekly Trip Counts for Sept. 2022



EDA 3

Overall Monthly Surplus by Docking Station



focus on Crown Heights

Conceptual Model: Step 1

Clustering

Conceptual Model: Step 2

Modeling

Model Results

- Results table

Slide 10

Current Achievements

- Clustering
- Model prediction

Slide 12

Future Works

- Weather . . . actually, can I predict weather? would that really work?
- Subway stations: Citi Bike offers valet
- Model of all NYC
- Real-time clustering would be better

Slide 14

Slide 15