

Predicting When CitiBike Stations Will Go Out of Stock



Source: New York Times

by Solomane Sirleaf, Jeff Guo, Xuhui Fan, Zihao Feng, Matthew Coughlin, Zuda Yan and Kaiwen Zhang

Hypothesis

Null Hypothesis:

It is not possible to predict when CitiBike stations are expected to become empty.

Alternative Hypothesis:

It is possible to predict when CitiBike stations are expected to become empty.

Motivation

According to an article, *Citi Bike Struggles to Keep Up With New Yorkers' Love of Cycling*, published by Ana Ley in the New York Times, states that since the onset of the pandemic, bike ridership has increased significantly throughout New York City. However, while many New Yorkers have their own bicycles, for many, the annual CitiBike membership or occasional ride via the CitiBike app is a much more convenient and cost-effective option for alternative transportation. However, with so many citizens using CitiBikes these days, it can be

difficult for the general public and policymakers alike to predict when CitiBike stations will become empty and subsequently when they are likely to be sufficiently restocked. We believe building a model to predict when CitiBike stations will go empty will solve two problems: 1) it will provide the public with more transparent information to guide their decision-making processes, and 2) it will provide policymakers with valuable information as to which CitiBike stations have higher frequencies of stock shortages, thus providing optimal neighborhood locations for additional bike stations.

Research Question

Is it possible to predict when CitiBike docking stations are expected to go out of stock so that stakeholders can plan accordingly?

Methods

There are several methods available to answer the following research question: **Is it possible to predict when CitiBike docking stations are expected to go out of stock?** So, stakeholders can plan accordingly. These are the methods we plan to use to answer our research question: deplore a time-series analysis model to find peak time in Citibike stations usage and frequently use stations; find the optimal variables that are relevant in yielding the best results to our research question.

Data Sets:

synthetic population <https://dw.tandoncsmart.com/dataset/synthetic-population>

Citibike tripdata: <https://s3.amazonaws.com/tripdata/index.html>

Cibike rebalanced movement dataset: <https://github.com/ckran/bikeshare/blob/main/202009-citibike-reblance.parquet>

Objectives

Primary Quantitative:

1. Build a predictive model which estimates to the smallest feasible time interval when a CitiBike station will become empty AND when they are expected to restock
 - a. Based on historical CitiBike data for 2019 - 2021, thus providing pre-pandemic and mid-pandemic insights (possibly more years).
 - b. We could use 2022 as test data based on findings, and maybe 2-3 months from each year as validation data (the remaining 9-10 months being training data)

Secondary Quantitative:

1. Identify potential neighborhoods (census tract or zip code) for the installation of additional bike stations based on model findings.
2. Discover how Rebalancing Movements between stations works (when, where), (see reference).

Roles:

Team Members	Team Roles
Jeff Guo, Solomane Sirleaf	Data acquisition and cleaning (Preprocessing)
Jeff Guo, Zihao Feng	Descriptive/Exploratory Analysis
Solomane, Zihao Feng	Time Series Analysis
Zuda, Xuhui	Additional Modeling (PCA, RF, etc.)
All team members	Final Conclusion and Analysis
Matthew, Kaiwen	Final Data Visualization
Kaiwen Zhang	Writing and Presenting Report
Matthew, Zuda	Drafting, Editing, and Finalizing Report

References

1. <https://nyc.streetsblog.org/2021/11/02/asking-for-a-friend-is-citi-bike-being-held-hostage-by-dot-fear-of-inconveniencing-drivers/>
2. <https://www.nytimes.com/2021/12/02/nyregion/citi-bike-parking-docking-station.html#:~:text=Since%20it%20started%20in%20Manhattan,sharing%20system%20in%20the%20country.>