



Boston Bluebikes Usage Analysis and Seasonal Demand Forecasting

Group 4 -

Rishab Jaiprakash Khuba

Prerna Chander

Pranamy Dinesh

Debankita Basu

Saurav Krishna



Agenda

- Introduction
- Methods
 - Data Tidying and Transformation
 - Exploratory Data Analysis
 - Data Modeling
- Results
- Discussion
 - Conclusion
 - Future Work
- References



Introduction

- Bike rentals are cost-effective, environmentally friendly alternative transport options that are gaining popularity
- Utilized publicly available Boston Bluebikes dataset across the years 2019 - 2022 for analysis
- Goals
 - Focusing on Boston Bluebikes program to understand the seasonal and time period effect on the utilization of the bicycles
 - Narrowing on areas around Northeastern University in order to visualize the demand during school terms
 - Performing seasonal demand forecasting using ARIMA and LSTM models

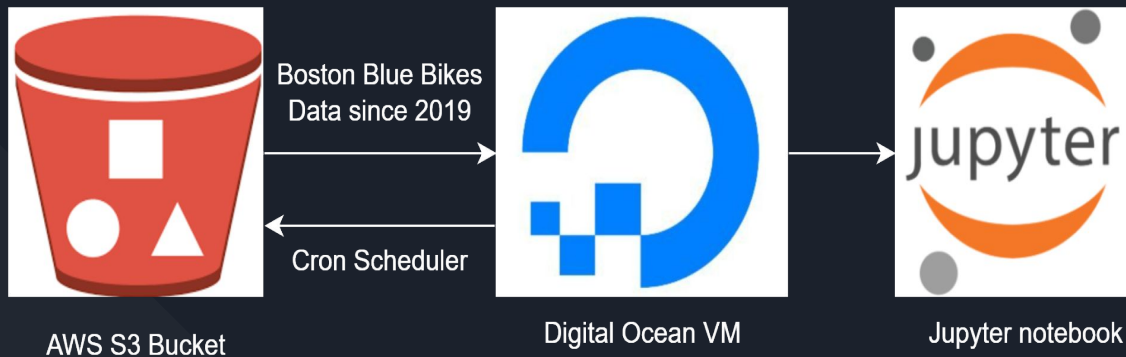


Methods


- Data Tidying and Transformation
- Exploratory Data Analysis (EDA)
- Data Modeling

Data Tidying and Transformation

- Data Compilation -



Data Tidying and Transformation

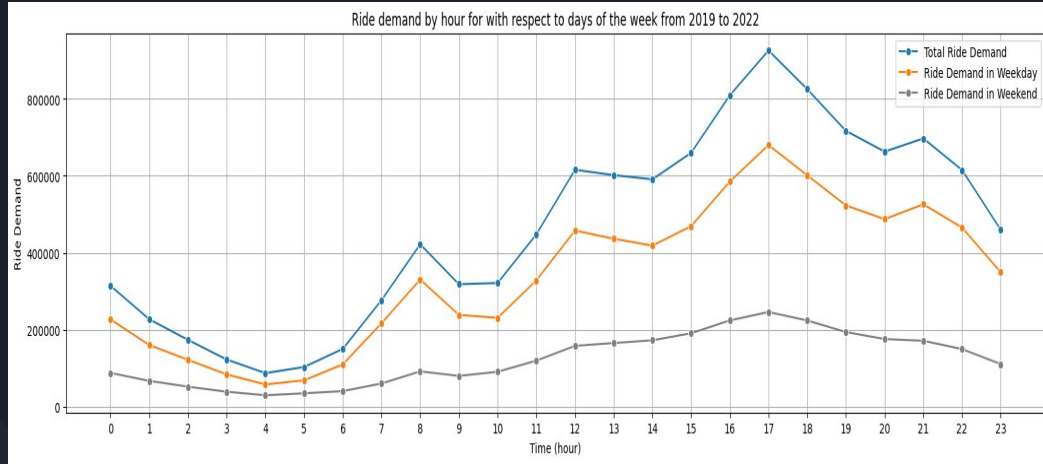


Features	Description	Type
Start Time	Timestamp of when the bike ride began (eg: 1/1/2019 10:09:47 AM)	Interval
Trip Duration	Duration of ride in seconds (eg: 371.5 s)	Continuous
User Type	Is the user a regular customer or a subscriber	Categorical
Start/End Station Latitude	Coordinates of the dock stations from where the ride began and ended (eg: 42.358100 and -71.093198)	Continuous
Start/End Station Longitude	Coordinates of the dock stations from where the ride began and ended (eg: 42.358100 and -71.093198)	Continuous

Data Tidying and Transformation

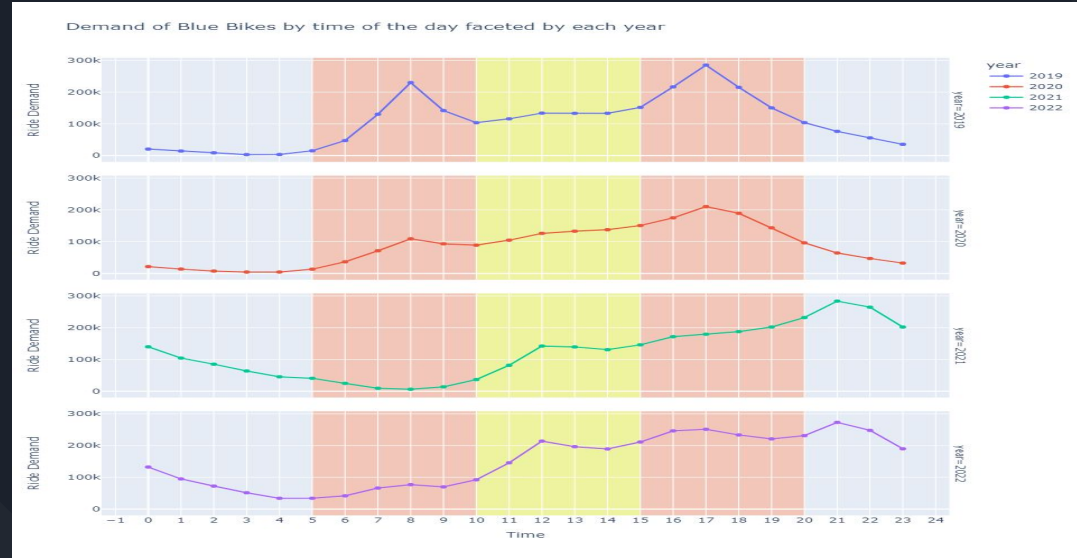
- Data Cleaning and Manipulation
 - Removal of NULL values and insignificant features
 - The dataset was subsetting into multiple dataframes for specific analysis
 - Dataframes were based on rentals per hour on weekdays and weekends, demand of rentals per season and dock stations around Northeastern

Exploratory Data Analysis



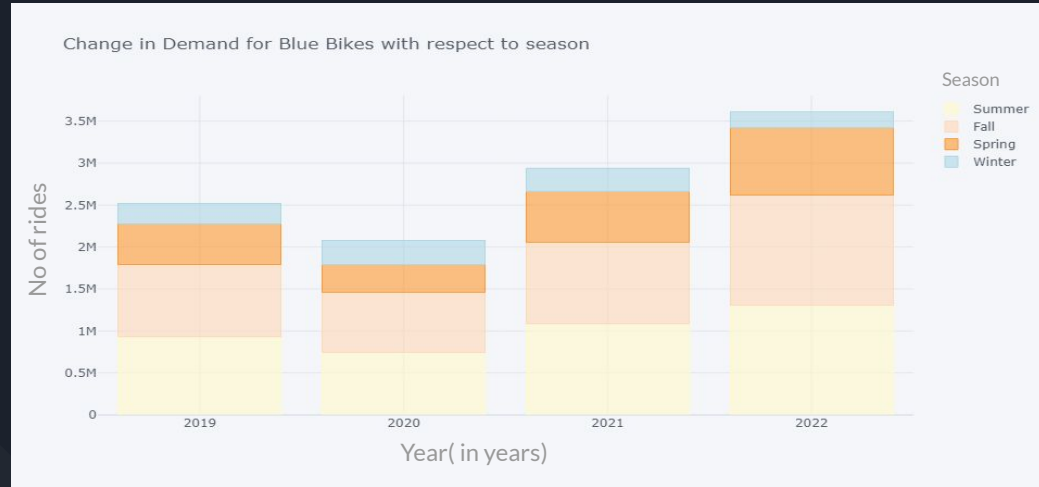
- Ride demand on weekdays versus weekends split by hours of day
- The ride demand reaches its peak at 5 PM
- More demand is seen during regular working hours compared to weekends

Exploratory Data Analysis



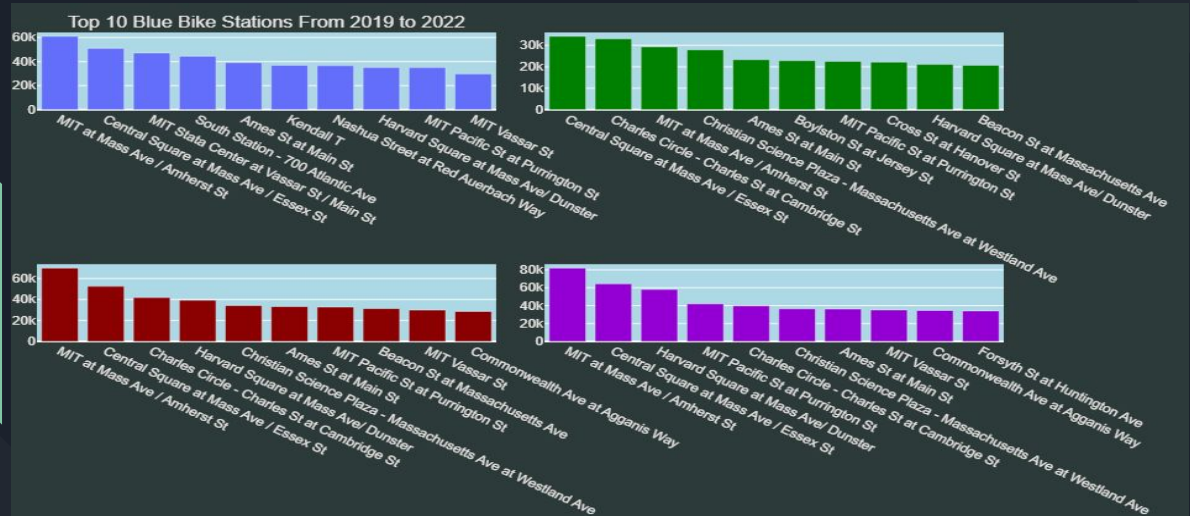
- The COVID-19 pandemic results in a fall in the demand for Bluebikes in 2020 .
- In 2021 and 2022 we see the most demand in non- regular working hours as opposed to 2019.
- This may be attributed to the increase in remote jobs since COVID.

Exploratory Data Analysis



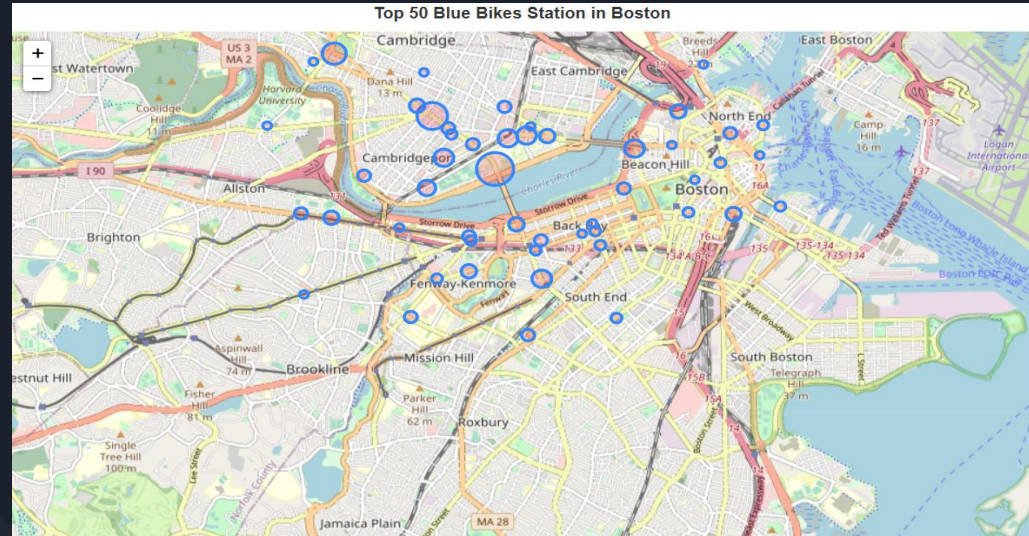
- Summer and Fall seasons seem to be the time when most people utilize Bluebikes
- During harsh winter conditions, the demand falls drastically
- Barring 2020 there is a steady increase in demand over the years

Exploratory Data Analysis



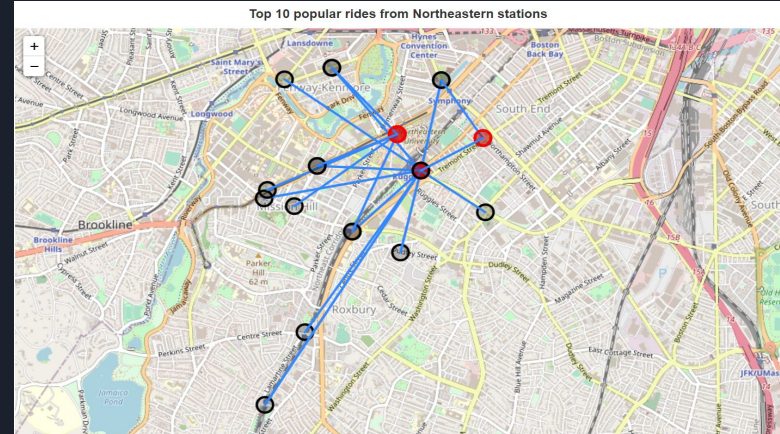
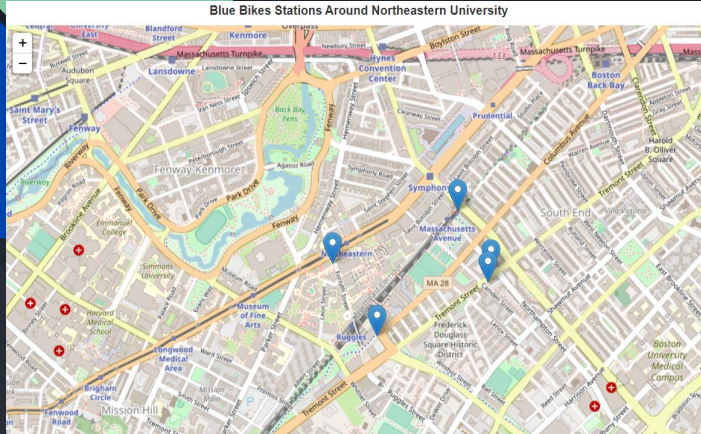
- MIT at Mass Ave/Amherst St is the most popular dock station over the years.
- Possibly used a lot by students of MIT.
- For the better part of 2020, MIT remained closed and this resulted in MIT at Mass Ave/Amherst St being less popular that year.

Exploratory Data Analysis



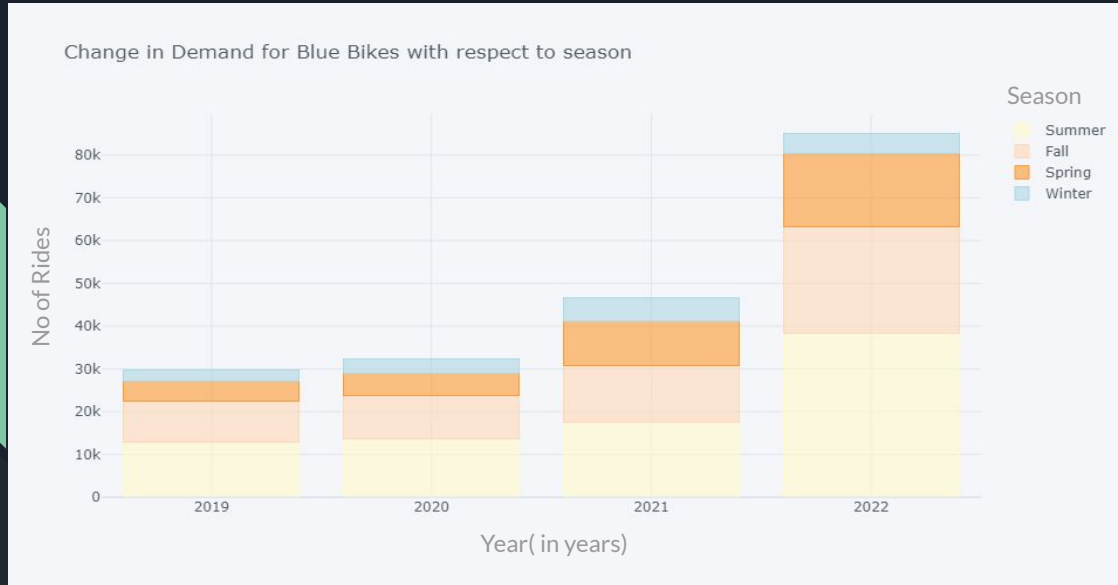
- Highest number of total rides taken from the top 50 popular stations.
- Ruggles, one of the stations near Northeastern University being in the top 24 among those popular Bluebike stations.

Exploratory Data Analysis



- Analyzing blue bike stations in and around Northeastern University
- Bluebike stations near Northeastern University - Ruggles T Stop, Forsyth St at Huntington Ave, Mass Ave T Station and 2 stations in Tremont St are used the most
- Ruggles station is the most popular source dock station

Exploratory Data Analysis



- Summer and Fall seasons have most Demand
- Demand for Bluebikes near NEU doubled by 2022.

Data Modeling

- Dataset containing information about dock stations around Northeastern University was used for data modeling to perform demand forecasting.
- The data was converted into a time series format for modeling.
- Two models - Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) were used

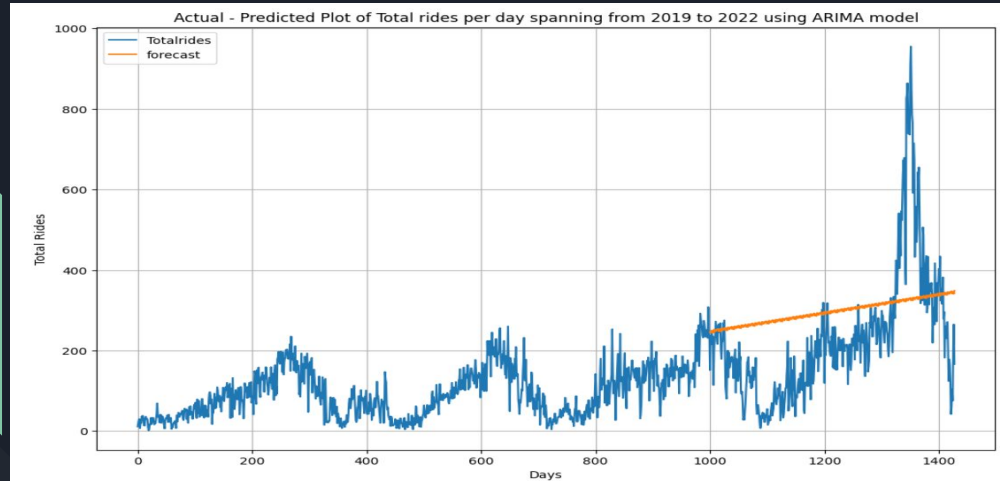
Data Modeling - ARIMA

- ARIMA is a statistical forecasting algorithm that uses the past values of the time series to predict future values
- Augmented Dickey Fuller Test (ADF Test) was performed before parsing the data through the ARIMA model
- It determines whether the data is stationary
- Stationary data implies that mean, variance and autocorrelation are constant over time

Data Modeling - LSTM

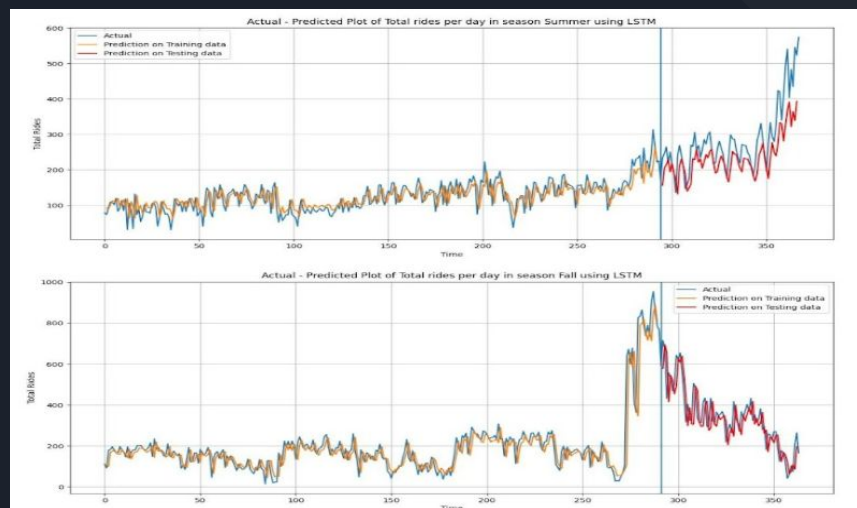
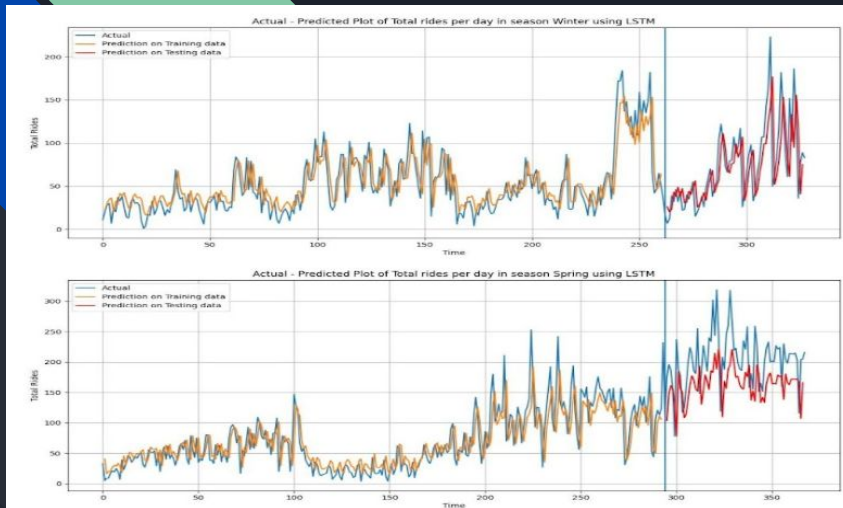
- LSTM is a long short term memory neural network that remembers the long-term dependencies of the dataset
- Has the ability of adding or removing information from the cell-state using gates
- The Gates are made up of numerous pointwise addition and multiplication operations as well as activation layers
- The LSTM unit consists of an input, output and hidden gate

Results -ARIMA



- Demand forecasting using ARIMA model for Bluebikes stations around NEU
- Difference in the actual demand for Bluebikes in the dock stations around NEU and the forecasted demand using our ARIMA model
- The prediction turns out to be approximately linear and it fails to account for the drop in demand seen during off season and also the huge spike up seen around the 1300 day mark

Results - LSTM



- To get better results, a LSTM neural network model was used and the data was split into four seasons
- The LSTM presented a high accuracy rate
- It had a lower RMSE as compared to the ARIMA model

Discussion- Conclusion

- We forecasted a high demand for Bluebikes around Northeastern during the fall season
- Beneficial to open up more dock stations or add more bikes during fall
- Steep decline of bike usage in winters is seen, would be wise to spend less effort on maintenance of dock stations
- Bluebikes when it was offered for free saw a sharp rise during that month
- Opportunity to use cost optimization techniques to increase usage of blue bikes



Discussion- Future Work

- Using hourly data from Blue Bikes API
- Analytical tool which would compare bike rentals across different cities
- Build a complete data pipeline to streamline the whole process

References

- <https://www.thecrimson.com/article/2022/12/1/Bluebikes-boston-logistics/>
- <https://www.Bluebikes.com/blog/2022-winter-challenge>
- <https://www.Bluebikes.com/system-data>
- <https://www.boston.com/news/local-news/2022/09/19/Bluebikes-ridership-records-orange-line-shutdown/>
- <https://analyticsindiamag.com/a-complete-guide-to-lstm-architecture-and-its-use-in-text-classification/>