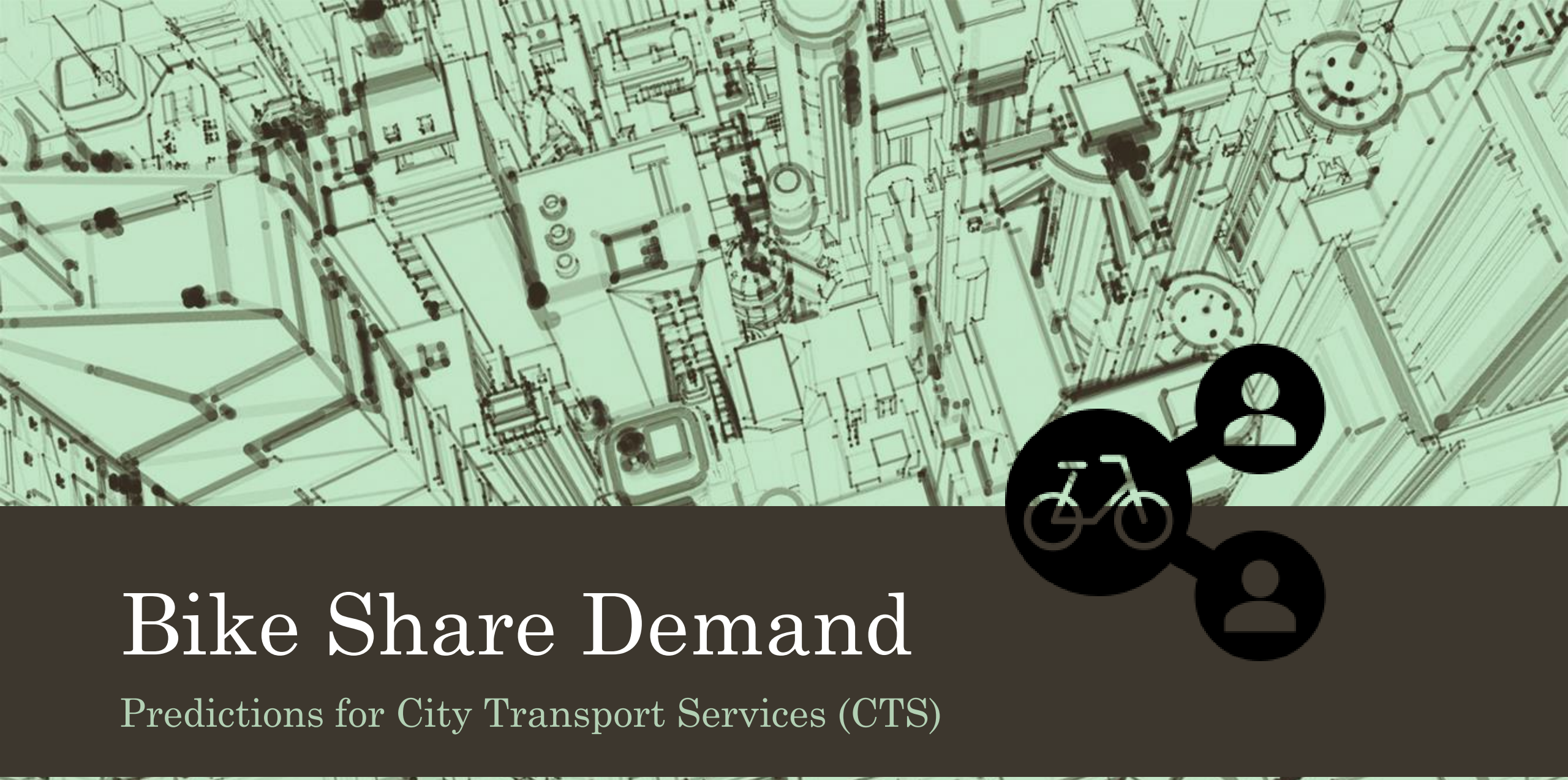


Bike Share Demand

Predictions for City Transport Services (CTS)

Summary

City Transport Services (CTS) wish to better understand mobility in the city, in particular the bike sharing systems and the characteristics of the data generated by these systems. Hence, it is expected that most of the important events in the city could be detected via monitoring these data.

By better understanding this data, CTS wish to be able to forecast mobility and plan for future events. This project analysed various attributes associated to daily bike rentals. Specifically, we assessed seasonality, weather conditions and types of users.

From this, we found that:

- The data is mainly skewed to registered riders
- Bike rentals are highest during warmer season (Spring and Summer) and lowest during Winter
- Rentals seem to be highest during 'good' weather conditions (Clear with few clouds) and drops roughly 50% during 'poor' weather (light snow, light ran + thunderstorm + scattered clouds, light rain + scattered clouds)
- Temp and atemp have a relatively strong relationship with bike rental counts

Outline

- **Business Problem**
  Objective and approach
- **Data**
  Data adopted and analysis
- **Methods**
  Preparation, analysis and modelling data
- **Results and Conclusion**
  Interpretations and recommendations

Business Problem

Due to CTS having limited knowledge of bike rental demand, we ask the following questions to help us solve the business problem

We assessed various attributes in particular season, weathersit and temp/atemp to aid us in solving which variables significantly impact bike rental demand.

By answering these questions, CTS can make better assumptions on how to control transport means and resources, when to supply more or less and how to efficiently run operations within the city transport network.

Data

Bike sharing systems are a new generation of traditional bike rentals where whole process from membership, rental and return back has become automatic. Through these systems, people can easily rent a bike from a particular position and return back at another position. Currently, there are about over 500 bike-sharing programs around the world which is composed of over 500 thousands bicycles. Today, there is huge demand for these systems due to their importance in traffic, environmental and health issues.

Apart from interesting real-world applications of bike sharing systems, the characteristics of data being generated by these systems make them attractive for the research. Opposed to other transport services such as bus or subway, the **duration of travel, departure and arrival position is explicitly recorded in these systems**. This feature turns bike sharing system into a virtual sensor network that can be used for sensing mobility in the city. Hence, it is expected that most of important events in the city could be detected via monitoring these data.

- One daily dataset 'day.csv' from the Capital bikeshare system with corresponding weather and seasonal information from 2011-2012 was used for this analysis
- We targeted the following independent variables; season (1: winter, 2: spring, 3: summer, 4: fall), weathersit (1: Clear, Few clouds, Partly cloudy, Partly cloudy, 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist, 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds, 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog), temp (normalised temp in celscius), atemp (normalised feeling temperature in celscius) and our dependent variable 'cnt' (count of total rental bikes including both casual and registered)

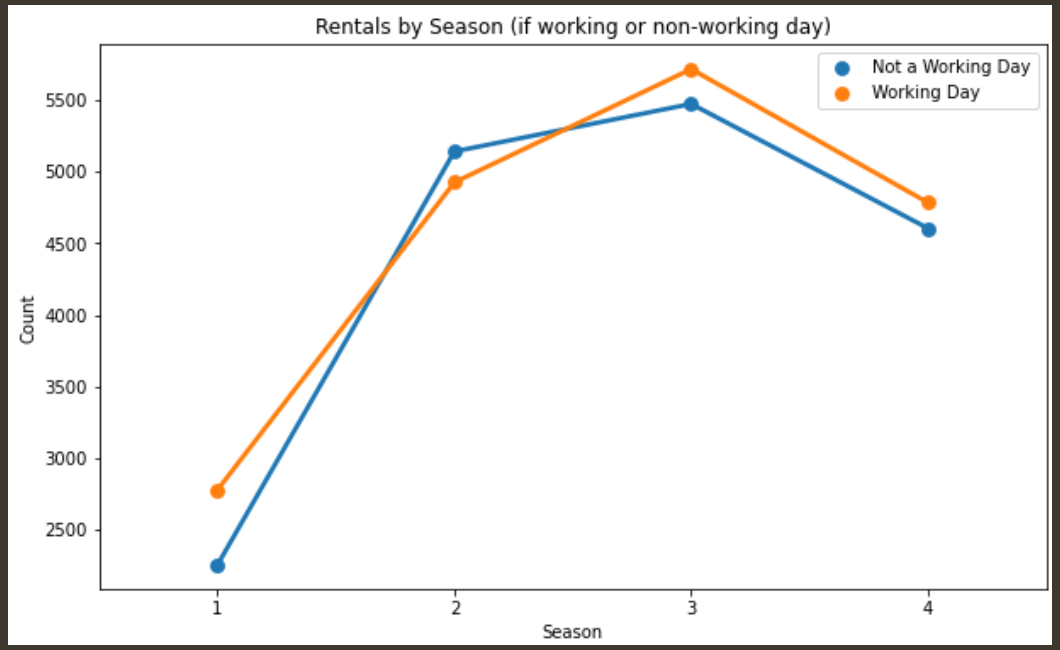
Method

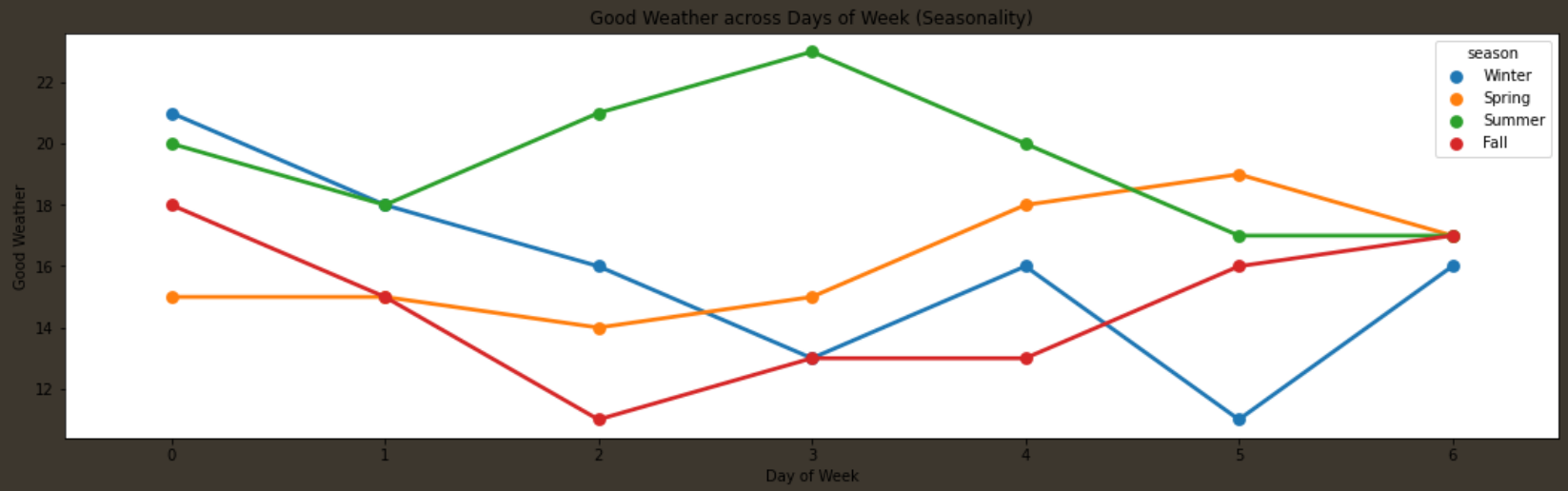
Preparation, EDA and modelling the data. Cross correlation, visualization, feature engineering, train/test data, assumptions of linearity, 27 OLS Regression models, prediction using model 27.

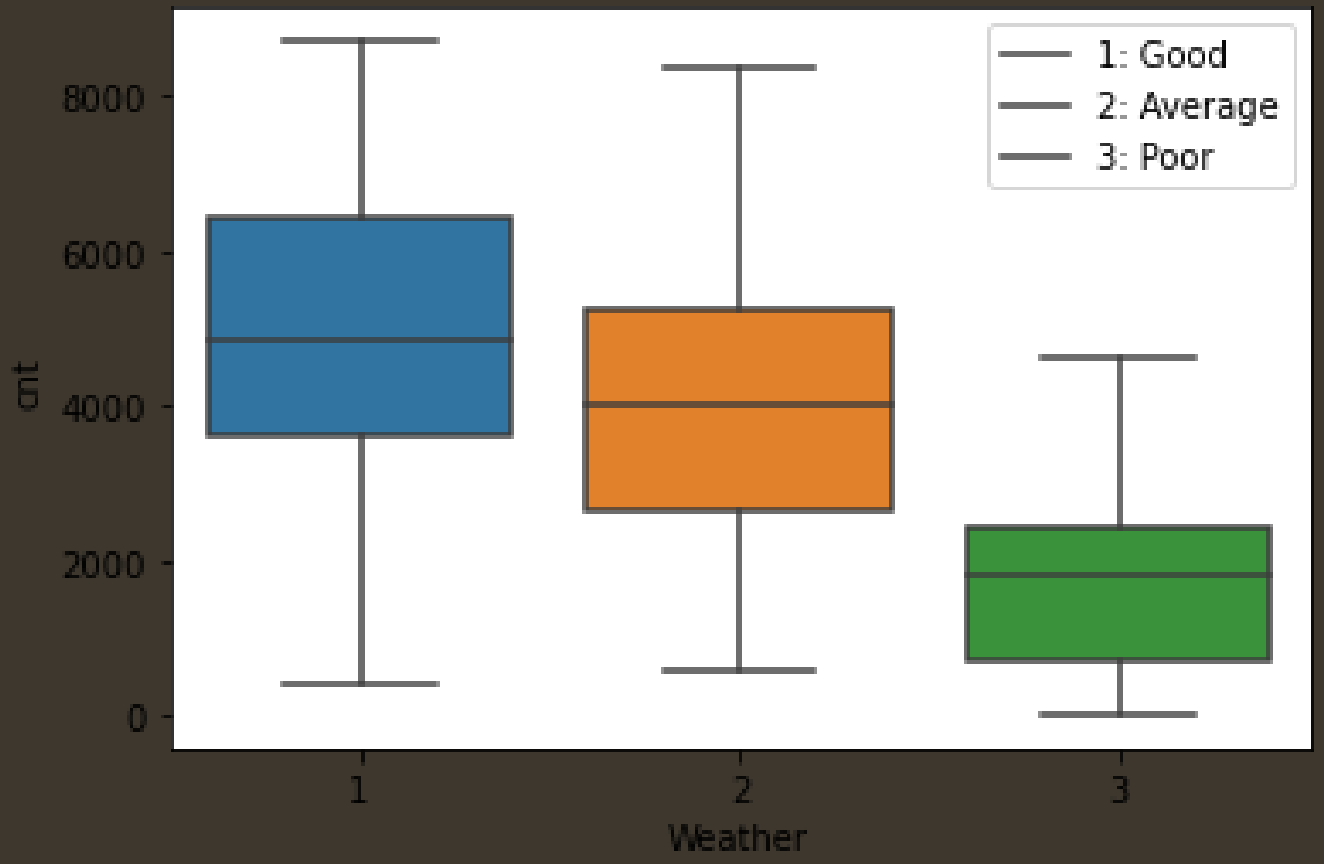
- Dataset was imported, 'dteday' coverted to Date time, followed by 'mnth' with names of month. Data had no duplicates or null values. We were able to immediately identify from the mean that there are 3656 registered riders vs. 848 casual, therefore most of this data is skewed to registered riders
- We ran a quick cross correlation to get a sense of variable correlations, our visualisation indicated strong correlations between 'cnt' and 'registered', 'yr', 'instant', 'atemp', 'temp' and 'casual'
- We plot average count across 'season' between working vs. non working days to see if anything interesting, we did the same for casual vs registered riders across days of week
- Assessed for outliers using boxplot. 'windspeed' and 'hum' both had outliers, we used violin plot to zoom in of the spread. We checked for distribution of the data to see if any other variables showed signs that could impact rental 'cnt'
- We know that 'registered' and 'casual' are included in 'cnt', so we dropped those
- Dummy variables were created for categorical data for ease of splitting our train and test data. We applied feature scaling to reduce bias towards the feature which has values higher in magnitude
- A baseline model was created and based on the Q-Q plot suggesting model qualities for assumptions of linearity. Baseline R-Square was low, we continued to create more models using statsmodels by removing variables until we found highest R-Squared and P-value is <0.05 for as many features as we can
- We validated our model for assumptions in regression and finally we made a prediction Dataframe using our final model.

Results

Interpreting the results, we found that bike rentals are highest during warmer season (Spring and Summer), lowest during Winter



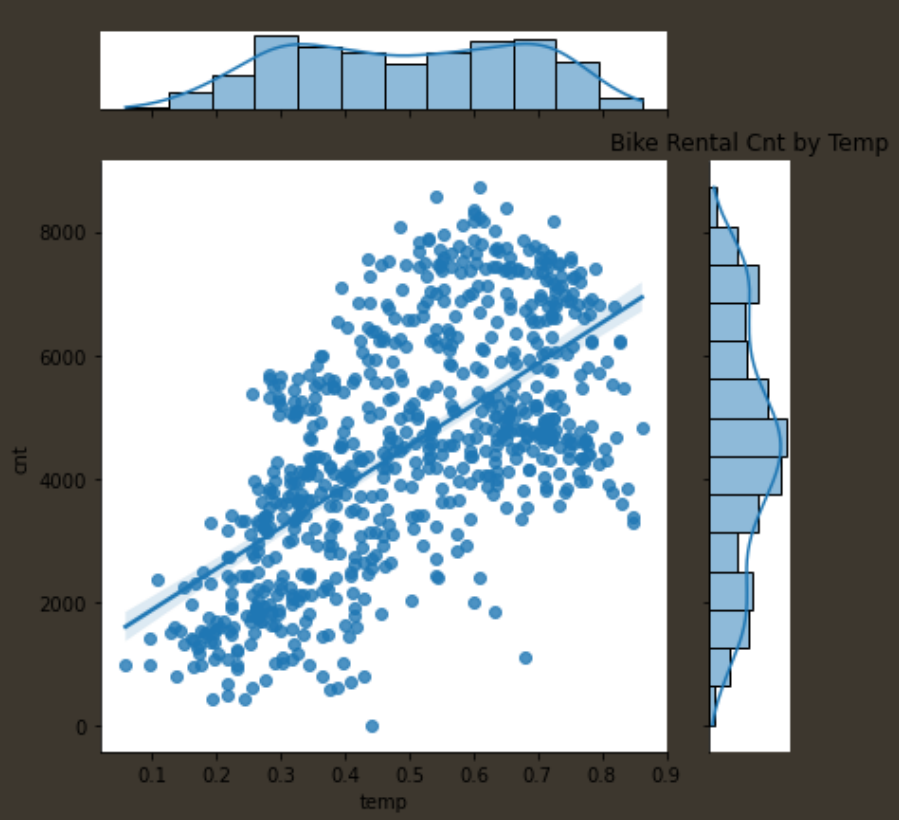




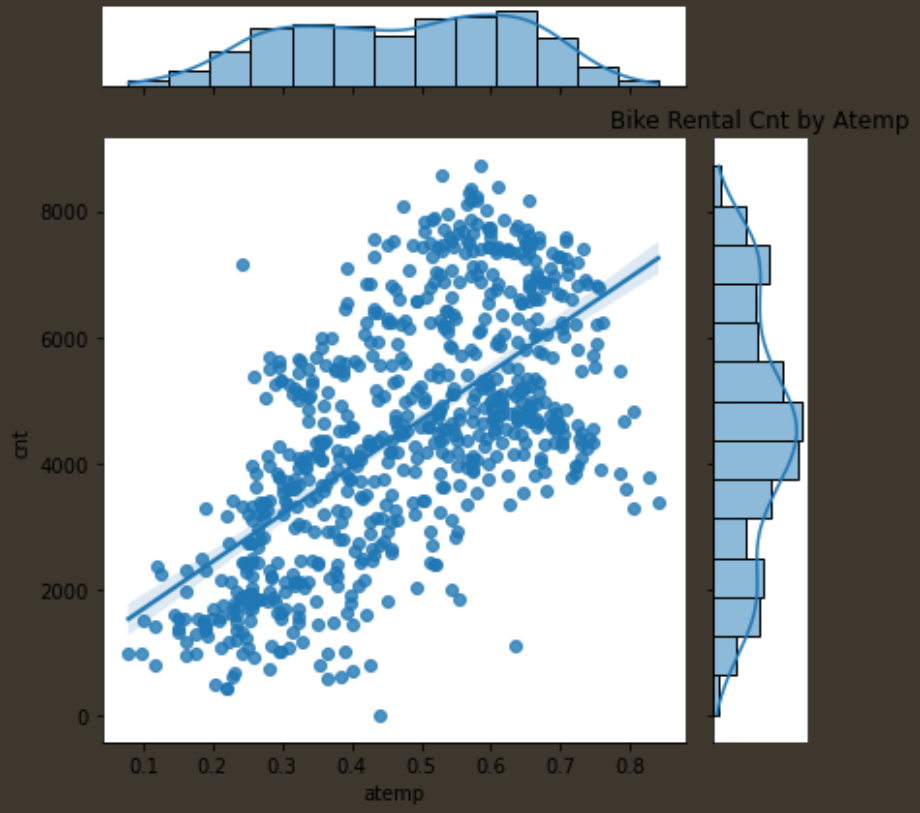
Results - Weather

Rentals seem to be highest during 'good' weather conditions (Clear with few clouds) and drops roughly 50% during 'poor' weather (light snow, light ran + thunderstorm + scattered clouds, light rain + scattered clouds)

Results - Temperature



Temp and atemp whilst not a perfect correlation with bike rental counts, there certainly is a relatively relatively strong relationship



Conclusions

From the results in our analysis, we recommend the following:

- Recommendation 1

* Reduce supply/frequency of other transport means and resources during Spring and Summer and increase during Winter

- Recommendation 2

* Assess weather forecast frequently for a more accurate read. Increase supply/frequncy of other transport means during poor weather conditions

- Recommendation 3
- * Assess climate data from previous years to predict temp and bike rental demand for the future. Similarly, to the above proposed, more transport options available during cold days (and vice versa)

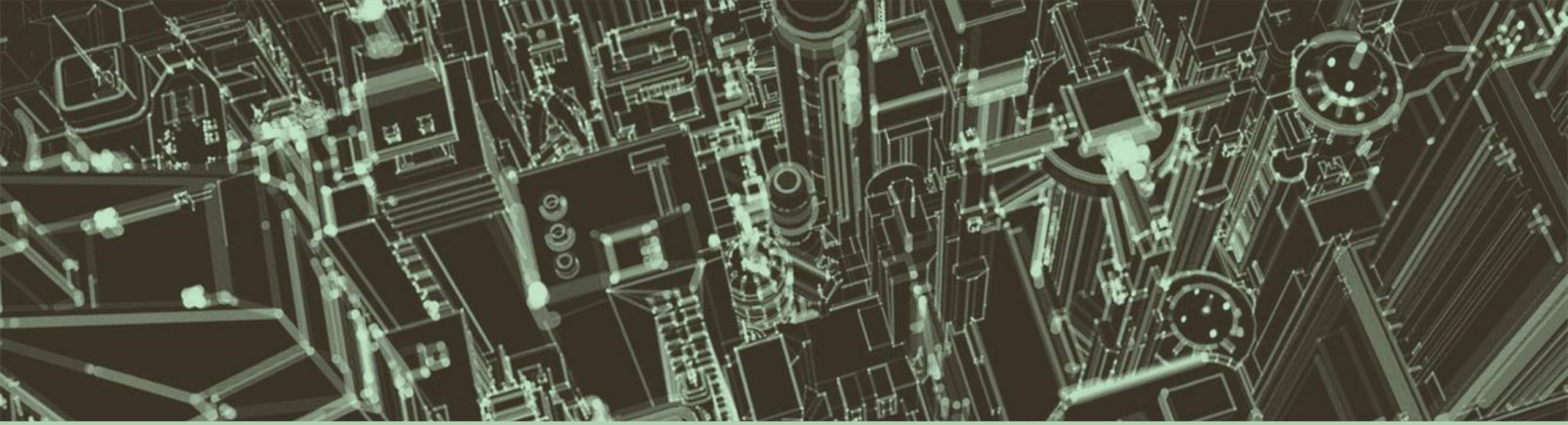
Next Steps

•There could be other variables that CTS would be interested in understanding to better understand bike rental mobility in the city.

•More recent data (2013-2022) could be concatenated to provide a better view of trends, as well as geographical data, pick up and drop off stations, bike routes, times etc.

•By understanding more intrinsic data such as time of pick up vs. time of drop off we can better understand the trends and routes people take. This will give CTS a clearer view of which routes are busiest and where other public transport means are required (or reduced).

•Having a bigger picture of the landscape, we can drill down further. We hypothesis that there may be a correlation between certain bike rental locations, for example, suburbs with bike routes, less parking, more narrow routes, high density and close to points of interest (shops, schools, parks etc.)are likely to have more rentals.



Thank You

annieliu1989@yahoo.com

https://github.com/annieliug/Bike-Share-Demand.git