

Boulder Bicycle Traffic Forecasting

JACOB MUNOZ

University of Colorado Boulder

jamu0075@colorado.edu

Abstract

Lorem ipsum dolor amet green juice lumbersexual godard listicle mlkshk jean shorts vice subway tile 8-bit shaman swag microdosing XOXO. Neutra biodiesel bespoke irony kinfolk la croix cold-pressed kale chips. Biodiesel 8-bit tbh flannel chia. Direct trade cornhole art party, selfies beard hashtag keytar venmo. Tattooed tbh intelligentsia blue bottle marfa brooklyn. 90's cornhole everyday carry try-hard.

I. INTRODUCTION

Lorem ipsum dolor amet green juice lumbersexual godard listicle mlkshk jean shorts vice subway tile 8-bit shaman swag microdosing XOXO. Neutra biodiesel bespoke irony kinfolk la croix cold-pressed kale chips. Biodiesel 8-bit tbh flannel chia. Direct trade cornhole art party, selfies beard hashtag keytar venmo. Tattooed tbh intelligentsia blue bottle marfa brooklyn. 90's cornhole everyday carry try-hard.

Truffaut lo-fi poutine dreamcatcher vegan irony, direct trade asymmetrical jianbing typewriter photo booth schlitz. Lo-fi hot chicken af taiyaki waistcoat. Pug PBR tbh stumptown williamsburg offal tattooed try-hard. DIY flexitarian edison bulb, chia wolf salvia pinterest chicharrones activated charcoal aesthetic 90's subway tile locavore cliché cronut. Truffaut ramps man bun microdosing selfies la croix mixtape. Mustache occupy celiac single-origin coffee. Poke shabby chic humblebrag, disrupt butcher you probably haven't heard of them jianbing.

Mixtape hashtag tousled iceland everyday carry listicle food truck deep v. Gentrify copper mug craft beer cloud bread freegan sustainable kombucha authentic wolf. Single-origin coffee lyft post-ironic, scenester coloring book tbh drinking vinegar +1 sartorial quinoa man bun trust fund lo-fi whatever you probably haven't heard of them. Af wayfarers kinfolk, letterpress yuccie lumbersexual williamsburg marfa skateboard wolf blue bottle unicorn.

II. DATA

Two data sets were used throughout this project. The first being bicycle counts at various intersections throughout Boulder and the second being daily weather data. The bicycle data is obtained from the City of Boulder website that has publicly available data that is updated regularly. The weather data is obtained from the National Oceanic and Atmospheric Administration (NOAA) website and is also updated regularly.

The bicycle data includes the count of bikes observed at each intersection every 15 minutes. The intersections being observed are highly trafficked and capture many common routes in Boulder. Every intersection has data up until the current day and begin between 2011 and 2015. It is uncertain how the data is collected but it is believed to be a simple count of objects passing through each intersection's bicycle lane. This would capture other modes of transportation such as skateboards or scooters but the vast majority of traffic in these lanes are bicycles.

The weather data includes the daily temperature minimum and maximum, snow cover, and precipitation beginning in 1897. This information is updated on a monthly basis and comes directly from NOAA's observations here in Boulder.

III. METHODS

To begin, plots were created to view bike counts grouped by day and month. The monthly counts showed a clear seasonal correlation while the daily count followed this trend in a more sporadic manner. The monthly counts remained mostly constant year after year though the total counts are rising slightly over the years. Due to the clear seasonal relation, the weather data was plotted next.

Daily temperature and monthly means were plotted to show a very consistent weather pattern. For the last 10 years the monthly means for low and high temperature has remained very consistent with only a few degrees of variation. When monthly means were plotted alongside monthly bicycle counts there was a clear visual correlation between the two. For the purposes of building a model, the Folsom & BoulderCreek Path intersection was focused on for the longevity of the data(2011-present) and its importance as a major entry/exit point for the Boulder Creek Multi-Purpose Path.

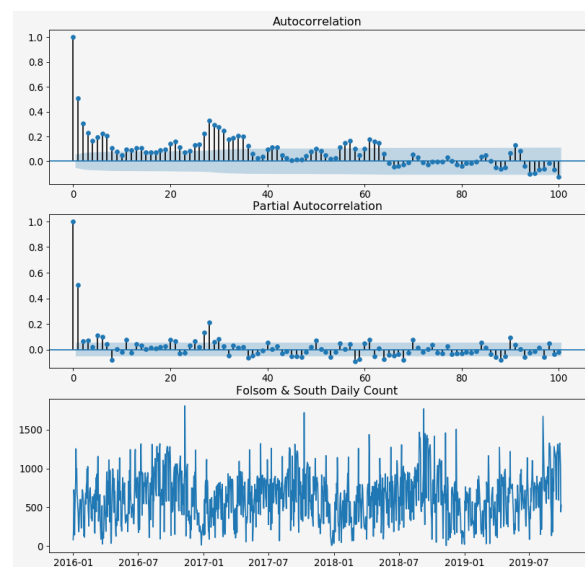
Before a model could be built the bicycle count data had to be cleaned. The number of 0 counts was around 9.8% over 8 years. When looking at the data the 0 counts could be observed to be grouped by days, often 3 or more in a row, and repeated monthly. This is mostly impossible in reality and most likely meant the counting mechanism was offline for some scheduled maintenance. For this reason all counts of 0(within the daily grouping) were removed. Furthermore the data had only 3 outliers and each were 4 or more standard deviations out. These dates did not appear to be any known holiday or event(i.e. Boulder bike to work day) and thus were removed. After these changes the skew for the Folsom & BoulderCreek Path count was 0.47 and was not adjusted. The skew of the weather data(daily max) was -0.39 and was similarly not adjusted.

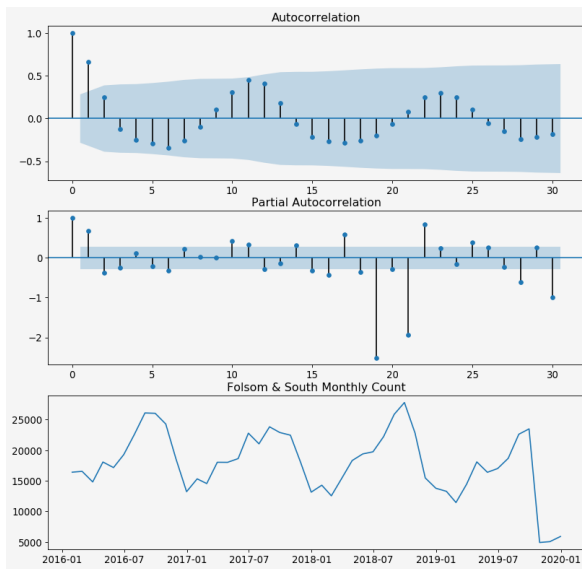
Next, a simple linear regression model was built using the daily temperature high. Daily high was used after observing the correlation table for the total and features: daily low, daily high, precipitation, snow, and snow cover. The

null hypothesis, there is no correlation between weather and bicycle count, was safely rejected after receiving a p-value of $6.89e-67$. The model has a room mean squared error of 347.57 and an R^2 value of 0.16. Thus, the model does not perform as well as I would like and I will move forward with multiple linear regression using precipitation and snow cover. However, I don't suspect this model will perform significantly better given the correlation value of these features are 0.09 and 0.14 respectively.

(I may go into some more depth of MLR here but it did not as prove successful as I hoped)

Given the behaviour of the data time series analysis was the next step. When observing the auto correlation and partial auto correlation we can see clear seasonal trends, especially amongst the monthly counts.





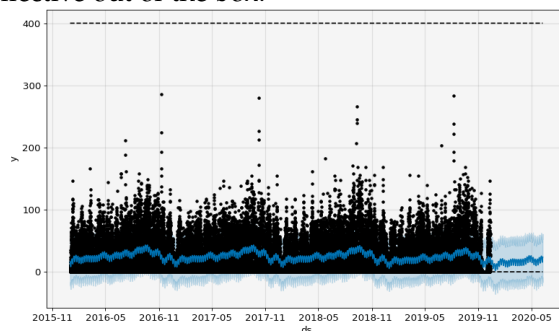
(all images need to be updated for readability)

Within the daily auto correlation there seems to be something affecting bike counts every 30 days, each month. When looking at the monthly correlation there is a clear trend every 6 months, perhaps due to the changing of seasons. This suggests that time series forecasting will likely yield better results. FBprophet was used for time series forecasting due to its versatility and customization options.

IV. RESULTS

(I still need to move some stuff from methods to here regarding regression.)

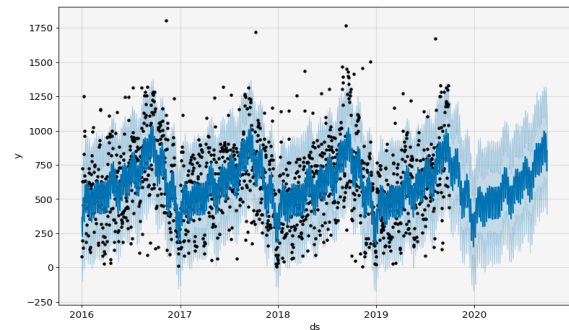
The goal is to predict bicycle activity in the near future, sub-daily, given previous observations. Initial hourly forecasting proved ineffective out of the box:



Hourly bicycle traffic is heavily seasonal within a day with heavy traffic before and af-

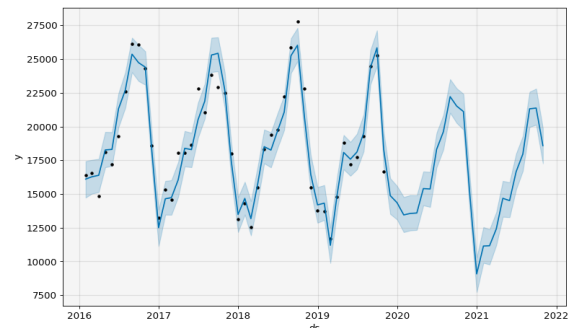
ter work hours (include graph here, not sure where it went but I'm not just making this up...) This suggests that with more time and tweaking this model could prove to be successful by predicting base off of the previous hour or two. It will also be worth attempting this same predictions with a smaller time frame of 15 minute intervals if hourly is proved to be successful.

Next, daily bicycle traffic was modeled, yielding a more accurate model, but still much room for improvement.



As you can see, daily counts fluctuate quite dramatically making it hard to predict. The model was able to capture the daily trend throughout the year however the value range is less than 100 up to over 1200, thus creating an unhelpfully large range of predictions. Again, given more time and model adjustments this could prove to be more useful, however, sub-daily data may be even more helpful given how short the time range can be and how activity is clustered throughout a day.

Moving on to the monthly model, it proved far more successful than smaller scale predictions out of the box due to its simplicity. Here you can the confidence intervals are much smaller and are able to capture most of the data.



While this model may prove helpful in predicting total bicycle traffic months are years into the future(given more data) that is not the goal of this research. This model may be worth tweaking and improving in the future but for now the focus is short term predictions for daily use.

(I have accuracy metrics such as MAPE and RMSE for each of these models but no visuals. I need more time with each to fully understand/talk about them. I also plan on going more in depth with each model and the decisions I made but currently do not have the performance to warrant that discussion at this time.)

V. DISCUSSION AND CONCLUSIONS

Lorem ipsum dolor amet offal bitters venmo 90's brunch thundercats synth messenger bag. Kombucha bespoke tumblr pabst. Seitan four loko hell of pabst, mustache meh shabby chic hella YOLO cardigan pok pok gluten-free vinyl iceland. Tumblr chicharrones vaporware, farm-to-table jianbing gochujang normcore woke

kickstarter before they sold out activated charcoal dreamcatcher.

Vape retro VHS PBR chartreuse readymade chambray gastropub artisan etsy organic la croix. VHS shabby chic squid kale chips lo-fi fashion axe green juice tofu affogato fam hella 8-bit gentrify. 90's ethical skateboard, listicle narwhal everyday carry taxidermy. Tumblr you probably haven't heard of them letterpress chia live-edge seitan literally iceland.

REFERENCES

- [City of Boulder, 2019] City of Boulder Bicycle Traffic Counts (2019) <https://bouldercolorado.gov/open-data/bicycle-traffic-counts/>
- [NOAA, 2019] NOAA Boulder Daily Data (2019). <https://www.esrl.noaa.gov/psd/boulder/getdata.html>
- [Hyndman, R.J., Athanasopoulos, G, 2018] Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. (2019) [OTexts.com/fpp2](https://www.otexts.com/fpp2)