CE 539 Project Outline

**Title:**

A Data-Driven Direct Ridership Model Approach to Prioritizing New Bike Share Station Locations: A Lexington, Kentucky Case Study

**Abstract:**

Several studies have been conducted to understand which factors are influencing and estimate bike share ridership [1]–[3]. This research takes a similar approach to build a direct ridership model estimating daily station-level bike share ridership. This study does not focus on one region, but rather utilizes publicly available datasets covering station and ridership data for 5 major US cities. The model finds that bike share ridership is most influenced by x, y, and z. A and B are found to still correlate with high ridership totals, but are less significant. C and D are found to have a negative relationship with bike share ridership. The model is applied to forecast ridership totals for a proposed bike share system in the city of Lexington, Kentucky. A quarter mile grid of point locations are laid out as potential locations, and the locations with the highest forecasted ridership are prioritized as the best potential location. A separate GIS based analysis is performed to prioritize locations that provide the best connectivity with the current bike facilities of Lexington, Kentucky. The two sets of proposed locations are fused to provide a set of proposed new bike share stations that are prioritized based on demand drivers and street connectivity. The model predicts that XXX daily riders will use the new bike share system and it will generate a revenue of $XXX, assuming an average fare of $X.

**1 Introduction**

* Introductory paragraph giving some background on the 5 cities and the bike share companies that operate in them.
* Background on the modeling methodologies that other research has used to estimate ridership for bike share stations
  + List out the current modeling techniques that research is using to estimate bike share ridership.
  + Provide the regionally specific limitation of previous research
* Provide what is found to be statistically significant correlated with bike share ridership
* List out what this research contributes
  + Estimation of a bike share direct ridership model of 5 US cities
  + An application of multi-city bike share ridership model for an independent city.
  + A new methodology that fuses a GIS based connectivity approach with a direct ridership model approach.

**2 Data**

Bike share Station and Ridership Data

* Data for Chicago, Minneapolis, New York, San Francisco, and Washington D.C. can be found here: <https://github.com/BetaNYC/Bike-Share-Data-Best-Practices/wiki/Bike-Share-Data-Systems>

LEHD Employment

* Census block level data for employment variables in the bike share direct ridership model
* The ridership data is only available for all 5 cities for 1 year (2017), and the most recent LEHD data is only available for 2015.

American Community Survey

* Data for Socio-Demographic variables in the bike share direct ridership model
* Either 1 year estimates of census tract level or 5 year estimates of census block group level data can be used.

Lexington Street Infrastructure (for GIS analysis)

* Bike trails from: <https://data.lexingtonky.gov/dataset/bike-trails>
* Street speeds from: <https://data.lexingtonky.gov/dataset/street>
* Location of traffic signals from: <https://data.lexingtonky.gov/dataset/traffic-signal>
* Vacant Land from: <https://data.lexingtonky.gov/dataset/vacant_land_2010>
* Street locations from: <https://data.lexingtonky.gov/dataset/street_2016>
* Land use from: <https://data.lexingtonky.gov/dataset/land_use_2000>
* Grade/bike facilities/traffic from: <https://transportation.ky.gov/Planning/Pages/HIS-Extracts.aspx>

**3 Bike Share Direct Ridership Prioritization Methodology**

* Assumes the ridership at each station is a function of the demand around it.
  + Build buffers to assign demand data based on how much of the census geography’s area is within the buffer.
    - Half of the census block is within the buffer half of the employees are assigned.
* Estimate daily station level ridership as a function of the demand within walking distance of it.
* Apply the model to a grid of potential locations in Lexington Kentucky spaced out by ¼ mile.
* Produce a heatmap showing the areas that would capture the most demand.

**4 GIS Based Connectivity Prioritization Methodology**

* Assumes the ridership at each station is a function of how well the station can be accessed by bike.
  + This is more of user experience approach to estimating bike share ridership.
  + Brings in the aspect of how well will a new station connect with the current bike infrastructure.
  + Accounts for grade of the streets, where the street is located, the speed of the streets, where vacant land is available, etc.
  + A raster (GIS heatmap) is created for each factor and is ultimately overlaid on top of each other to produce a final raster.

**5 Fusion of Both Prioritization Methodologies**

* I am blanking on a good way to do this right now, but visually using both heatmaps and engineering judgement is the best method I can think of.
* Ultimately a set of new station locations are going to be made
  + If a rough estimate of the cost to build a new station and the average fare of bike shares are known, then we can assume that only stations with a benefit cost ratio greater than 1 will be built.
* The ridership for the proposed system can be forecasted using the bike share direct ridership model as well.

**6 Conclusions**

* The final product is a list of new bike share stations for the city of Lexington, KY that have been prioritized based on their connectivity/accessibility and their ability to capture demand.
* A rough estimate of the revenue that will be generated based on an assumed average fare and the forecasted ridership
* Any other lessons that we may learn along the way ☺