capital bikeshare

Optimizing Capital Bikeshare Rebalancing

Thursday, September 7th, 2017

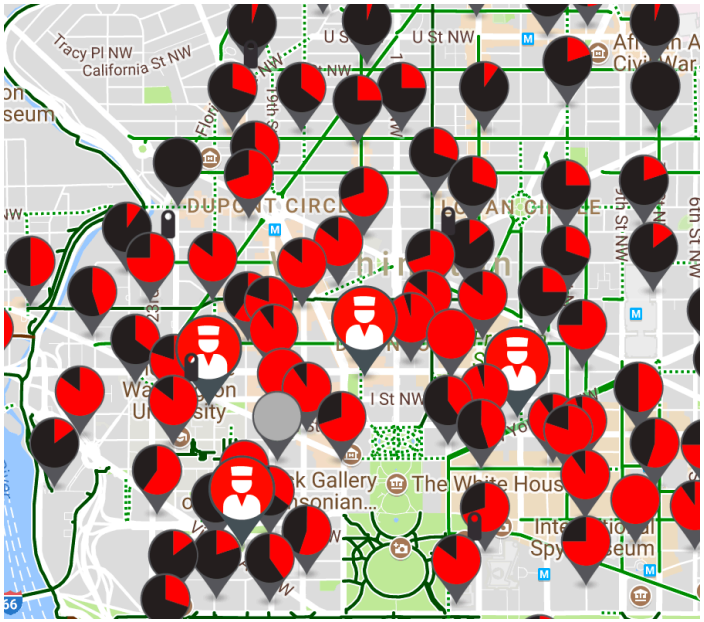
AGENDA

capital bikeshare

WHAT IS CAPITAL BIKESHARE?

Overview

- Capital Bikeshare is a public bikesharing system offering bike rental service within the DC Metro area
- Riders can pick up bikes at pre-determined stations, using either membership or single-ride payment options
- Capital Bikeshare was created in 2010 and has served 14.5 million trips



THE PROBLEM

CaBi has more bikes than spots, so it needs to optimally redistribute

With more than twice as many docks as bikes, CaBi needs to carefully monitor bike dispersion to ensure availability.

STATISTICS

Number of stations: 440

Number of docks: 7,555

Number of bikes: 3,700

REBALANCING

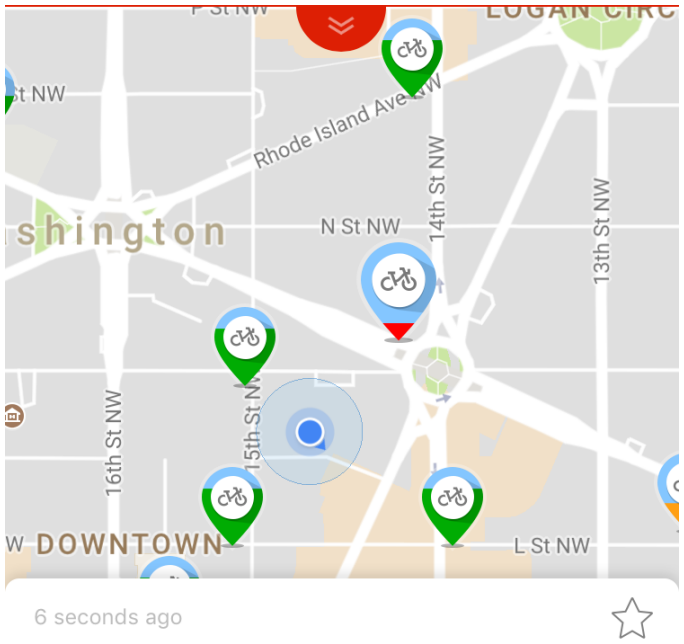
To ensure bike availability, CaBi "rebalances" bikeshare stations by **moving bikes to and from empty and full stations** respectively.

- Rebalancing typically takes place within 2 hours of empty/full
- **It's on one of the largest operating expenses for CaBi**

THE PROBLEM

Capital Bikeshare stations frequently run out of bikes

How does Capital Bikeshare (CaBi) know where to rebalance bikes, when bikeshare stations run out of bikes?



APPROACH

I built a model with CaBi trip data using Bayesian estimation to identify key stations

Data Collection	All trip data was made available on CaBi's website from 9/2010 - 3/2017
Data Munging	Trip data was cleaned and transformed into hourly statuses by station
Data Preparation	The transformed data was grouped into model data sets to prepare for analysis
Data Modeling	Each station was modeled to analyze net transactions at neighboring stations
Data Interpretation	Using the results, stations were ranked by priority for rebalancing efforts

THE DATA

CaBi publishes data on individual trips from September 2010

DATA FORMAT

- Data was organized by quarter and contained ~14 million trips
- Available Fields
  - Trip Start Date and Time
  - Trip End Date and Time
  - Trip Starting Location
  - Trip Ending Location
- Station characteristic data was available through an API and contained the longitude/latitudes and capacity of each station

ASSUMPTIONS

- Only used the **347 stations in DC/Arlington**
- Converted trip data to hourly count data by station
- Trips **beginning at a location differing from the preceding end location** were considered ”rebalanced” bikes

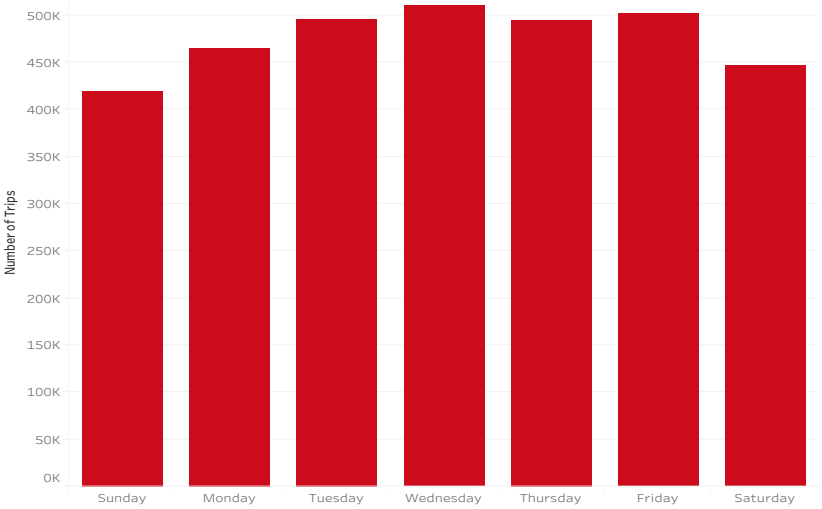
DATA EXPLORATION

I explored data on 2016 trips, to get a sense for recent trends

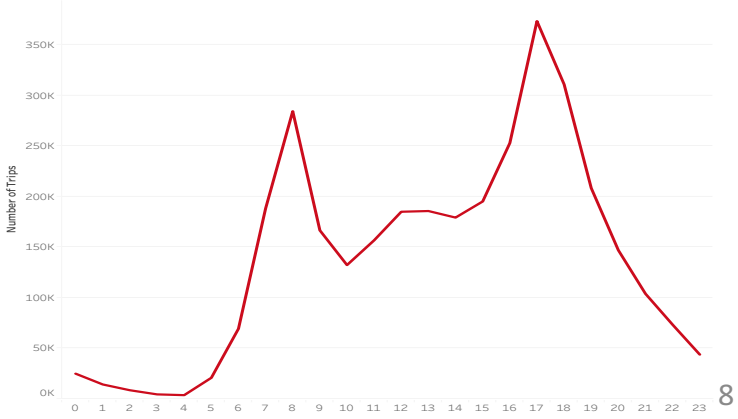
KEY FIGURES

- Wednesday had the most trips in 2016, and **weekends were clearly less popular**
- Trips were taken most frequently during the **morning and evening rush hour periods**
- The average single-ride trip time was **42 minutes vs. only 12 minutes** for members
- There were an **average of 750 trips** per unique bike

Trips per Weekday (2016)



Trips per Hour (2016)



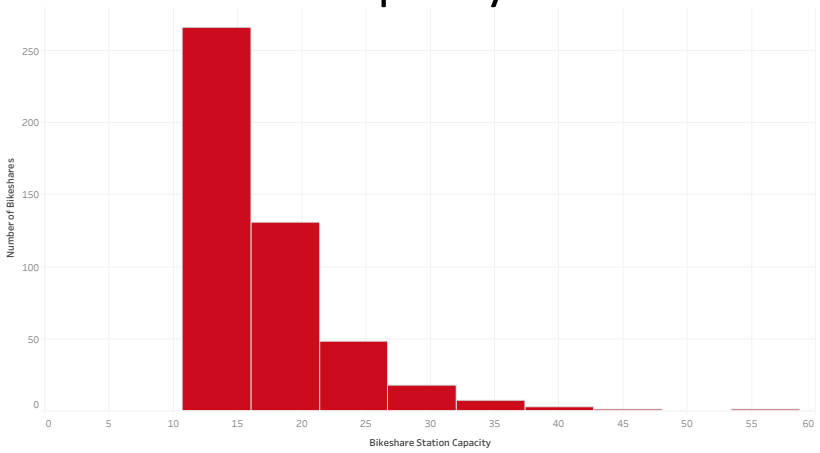
DATA EXPLORATION

I also explored some trends in my larger data set

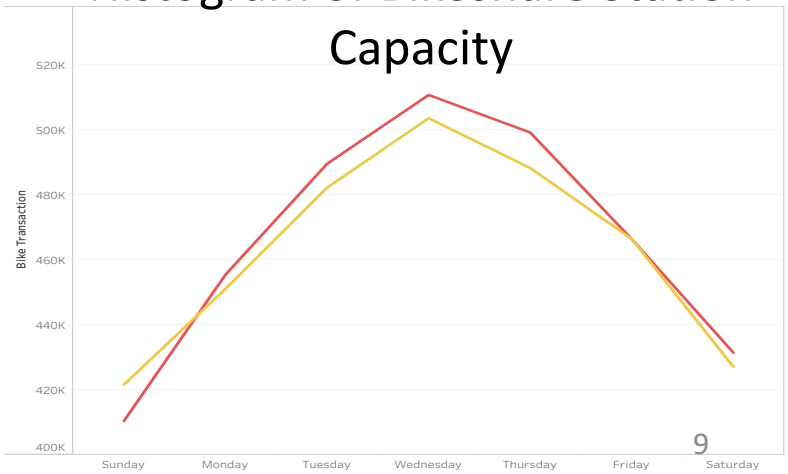
KEY FIGURES

- The **average station capacity is 17,** but the **median is 15** docks
- The average number of rebalanced bikes per period was **1 bike in per station** and **1 bike out per station**
- The station with **most bikes rebalanced in was at Union Station**
- The station with **the most bikes rebalanced out was at Dupont Circle** (also 2nd most in though)

Histogram of Bikeshare Station Capacity



Histogram of Bikeshare Station Capacity



TECHNICAL ANALYSIS

To model key stations, the 4 nearest stations were used

PROBLEM SETUP

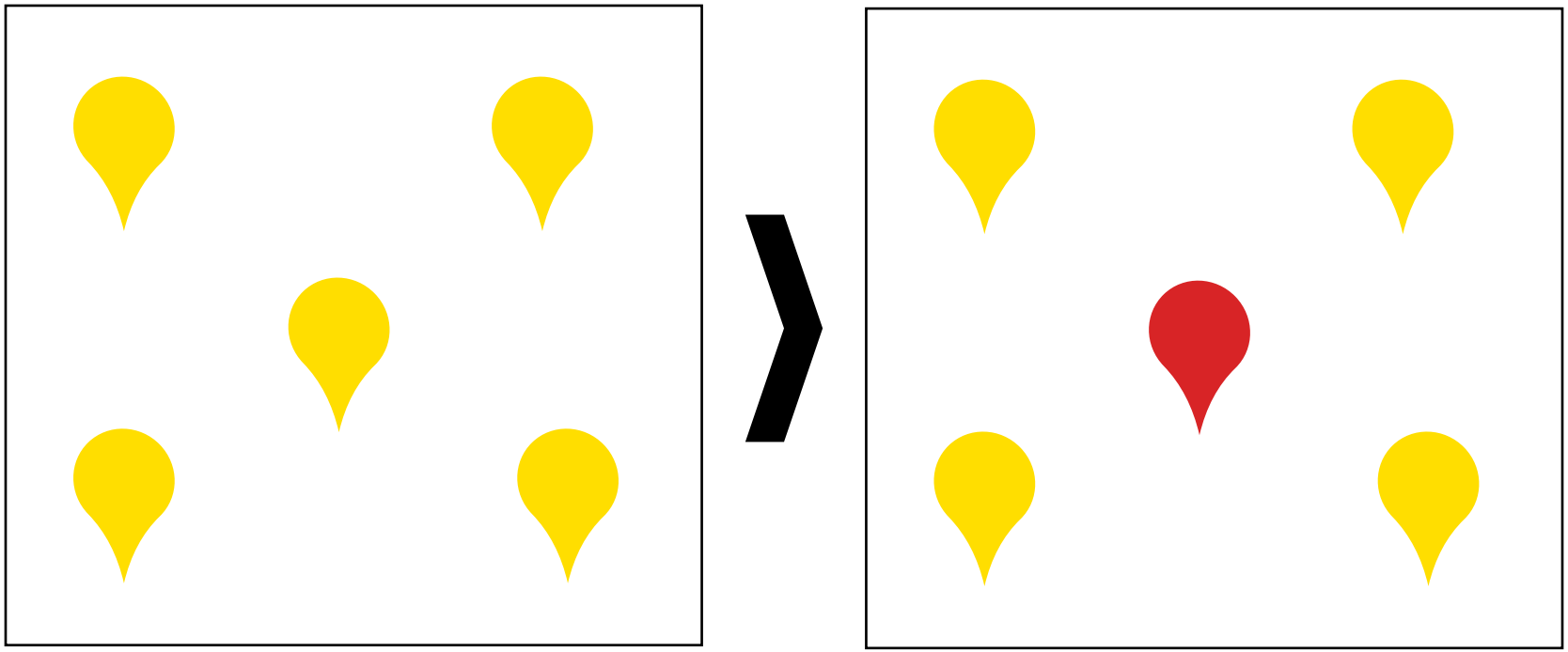
- CaBi only adds stations to areas that are **within .1 - .23 miles (2-5 blocks) of another station**
- On average there are **4 stations located within .35 miles** (7 minute walk) of a given station

OBSERVATIONS

- Observations were grouped into **6 hour time blocks** to determine:
  - Bikes brought in
  - Bikes taken out
  - Bikes ”rebalanced” in
  - Bikes “rebalanced” out
- Using the net transactions for a time block, each time block was calculated as either:
  - Empty - *net transactions is negative and exceeds the maximum available capacity*
  - Full – *net transactions is positive and exceeds the maximum available docks*
  - Regular – *neither empty nor full*

TECHNICAL ANALYSIS

I limited my analysis to the 229 stations with at least one empty observation



A regular observation, where the station of interest has bikes

An empty observation, where the station of interest has no bikes

Each observation is a 6 hour time block of the station status and the average hourly bike movement for the 4 nearest stations

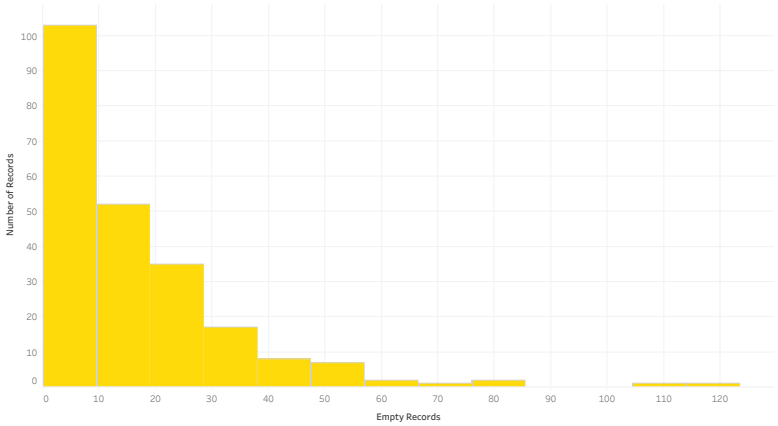
TECHNICAL ANALYSIS BREAKDOWN

I limited my analysis to the 229 stations with at least one empty observation

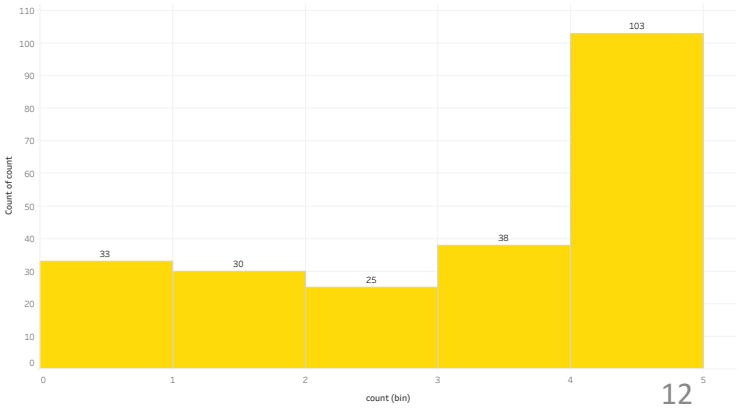
DATA DISTRIBUTION

- The average number of observations per station was 8,687
- The maximum number of empty periods was 123 at Adams Mill and Columbia Road NW
- The average number of empty periods was 16, with a median of 12
- The average rate for neighboring stations was 0, indicating that most stations have the the same number of bikes enter and leave

Distribution of Empty Periods



Distribution of Nearby Stations



TECHNICAL ANALYSIS METHOD

I used Bayesian Estimation to compare empty and regular periods

HOW DOES BAYESIAN ESTIMATION WORK?

- Bayesian estimation incorporates information on the prior belief of population, based on a distribution
- Using Bayesian estimation, we can **compare two populations**, based on our prior belief that **given the observed data, they are still the same**
- The Markov Chain Monte Carlo method helps us make this estimation

COMPARING EMPTY AND REGULAR PERIODS

Prior

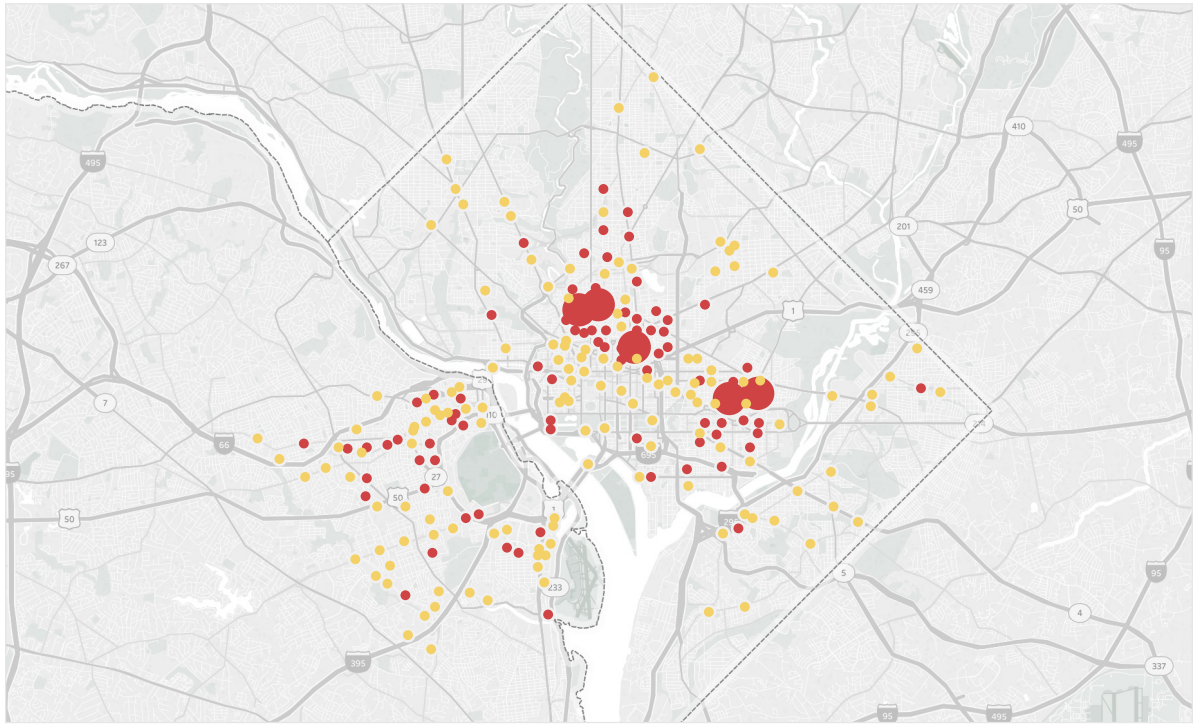
For the prior distribution, I used the mean and standard deviation of average rate

Likelihood

The Likelihoods were built off observed data for empty and regular periods

RESULTS

87 stations out of 229 had different results between empty and regular



Biggest Differences

1. 15th and F St NE
2. Maryland Ave and E St NE
3. 15th and W St NW
4. California St and Florida Ave NW
5. 8th and 0 St NW

CONCLUSIONS AND NEXT STEPS

The highest priority stations had an above average empty number of empty periods

RECOMMENDATIONS

- Rebalancing is obviously a challenge, and while some stations may be obvious, other less obvious stations see an **increase in demand at neighboring stations when they run out**, signaling that these stations should be higher priority for rebalancing
- The highest priority stations seem to have **similar rates of bikes coming in and out** as opposed to stations that have a greater imbalance
- They also seem to have a **high number of neighboring stations**, indicating that there is unmet demand when they run out

NEXT STEPS

- Studying how well CaBi rebalances by gathering point in time data on bikeshare station status
- Looking at how ’full’ stations differ from regular

capital bikeshare

QUESTIONS?