Bike Share Efficiency Amongst Modern City Transportation

Josh Spicer (s1883623@ed.ac.uk)
Teammate: Tom Bain

***Abstract** - One of the biggest responsibilities of any city is to provide means of transportation to its residents and visitors. Traditionally "providing transportation" meant maintaining a road network, and perhaps in addition offering some sort of metro and/or bus network for its residents to utilize. In recent years, numerous other modes of transportation have emerged, giving rise to more choice by consumers. This paper explores over 300,000 bikeshare journeys to draw conclusions on when a consumer should, and should not, utilize bike share. The paper concludes by discovering that journeys under 1.5 miles can be done fastest with a bike. Other conclusions are that while trips in the center of a city can generally be done much faster on a bike, trips several miles away from the city center can generally be done in comparable time regardless of mode of transportation.*

1. Introduction

1.1. Background

In recent years, the rise of bike sharing networks has provided another layer of connectivity for members of city communities. Bike sharing networks provide a quick, green alternative for short trips across the city.

In London, bike docking stations are typically spaced 400 meters apart, and located in key location beside major roads, near tourist attractions, or beside Tube stations.[1] According to the dataset utilized, there are 838 active use docking stations in London. Adoption of cycling public transportation has grown significantly in recent years with the number of cycling trips growing by 154% since 2000. There are over 730,000 journeys cycled each and every day within London, making it a growing competitor to the Tube.

1.2. Problem Statement

This paper analyzes the effectiveness of bike sharing networks in a major city using other modes of transportation, like Uber and rail, as baselines. The goal of such analysis is to inform consumers on the best method of transportation across their city by observing features of real journey data. This paper concludes with tangible information that a consumer could use when planning his or her travesl.

Transportation networks are complex and constantly evolving. People utilize transport networks in countless different ways, making it is quite challenging to provide general information about the network. Nevertheless, these findings are useful for anybody traveling throughout a city to keep in mind.

2. Related Works

One original strategy for this paper was to utilize TFL's journey planning API (although over 1 month ago we requested access and are still place 800+ in the queue). TFL provides the data that makes findproperly.co.uk's interactive trip map possible. This webapp illustrates where it is faster to bike opposed to utilizing public transit.

The work done in this paper relies on real bike travel data sampled over several months, opposed to TFL's simulated data. In addition, this paper aims to provide "rules of thumb" generalizations for consumers to use to decide mode of transport in a moment's notice. While this tool is interesting, it does not provide a "blanket" answer that this paper achieves.

3: The Datasets

3.1. Bike Journey Data

London's transit authority (Transport For London) publishes monthly anonymized raw journey data detailing the movement of bikes for every trip within a given period[2]. This dataset provides several interesting pieces of data including: start location, end location, trip duration, physical bike used, and more.

The dataset is composed of a total of 311,811 unique trips taken from February 28th-March 6th (2018) and April 19th-April 25th (2017). Given more time and computation power, years of trip data could be aggregated to provide a more representative picture of bike usage, irrespective of factors such as the seasons.

3.2. Tube Travel Data

We leveraged an existing dataset to simulate train travel times, and to leverage as a baseline for the bike journey data[3]. The dataset simulates the duration of a ride between stations, as well as the time necessary to switch platforms if need be. An attempt was made to acquire real Tube passenger data, but

---

1 https://tfl.gov.uk/corporate/about-tfl/what-we-do/cycling

2 https://cycling.data.tfl.gov.uk/

3 http://markdunne.github.io/2016/04/10/The-London-Tube-as-a-Graph/

the data was either behind a login page, or too difficult to parse.

3.3. Uber Travel Data

Uber travel data was more difficult to acquire. The data utilized in this paper was acquired by querying Uber's developer API, simulating trips on their `/estimate` endpoint. Given start and end coordinates, this endpoint provides, among other things, attributes such as: travel_time, travel_distance, estimated_cost, and pickup_estimate (estimated wait time). The value used for total trip duration is the sum of the travel_duration and pickup_estimate.

This data limited the total number of trials since the API endpoint only allows up to 100 requests per hour. This data was collected over numerous days, and its collection done via the script provided. API requests were done at various times throughout the day and throughout the week.

3.4 Calculation Algorithm

The datasets acquired all provided information on geographic location (latitude and longitude), or were indexed in such a way that geographic location could be extracted. Various techniques were then used to correlate the quickest travel time between a given origin coordinate, and a destination coordinate for each of the three modes of transportation.

The first technique was to generate purely random coordinates allowed to vary within an area bounded several miles around central London. This idea presented many challenges, one of which was calculating, from the dataset of bike journeys, the closest trip based on both origin and destination. This process was computationally time consuming, and therefore scrapped in favor of just randomly selecting one bike journey from the set of 300,000. Given more time, the simulation could be optimized and then run using purely random coordinates (unoptimized code is provided for just this).

From here, an API call was made to calculate the travel duration for an Uber ride. Other information, such as Uber base price, was also recorded for future analysis. For the London Underground, the baseline train journey time is acquired by running dijkstra's shortest path algorithm on the weighted Tube graph.

The three travel times are saved alongside the euclidean distance between the origin and destination. The distance from the origin and destination, to a predetermined center of London (51.510776 N, 0.115638 W), is also recorded. This process is iterated many times to provide the final dataset ready for visualization and analysis.

3.5 Visualization

This collected data is then fed into a graph where nodes represent the origins and destinations (placed onto a plane at normalized latitude and longitude positions). Edges are then drawn between nodes to represent journeys. These edges are colored based on which mode of transport was quickest.

4. Analysis

4.1 Constrained Minimal Travel Time

Computing the total travel time for a given journey from origin to destination is the most logical metric to equate the effectiveness of public transportation, and a key factor in a person's decision of whether or not to utilize an available mode of transportation.

In this section, the travel time for each computed journey is evaluated against two constraints: one being the total travel distance from origin to destination, and the other the journey existing within some variable proximity to London's city center. Of the three potential modes of transport, the minimum travel time for each journey is then reported.

4.1.1 Total Travel Distance

For bike networks to be utilized, one important factor is for them to be on par, or even quicker than other modes of transport available for the relative distance being traveled.

In the following table, the *fastest* mode of transportation is calculated for ranges of trip distances. The network graphs presented in the Appendix provide visual representations of these findings.

Total Trip Distance	Bike	Uber	Train
0.0mi - 0.5mi	67%	4%	28%
0.5mi - 1.0mi	61%	9%	28%
1.0mi - 1.5mi	52%	8%	39%
2.0mi - 2.5mi	34%	22%	43%
2.5mi - 3.0mi	28%	17%	54%
3.0mi+	29%	9%	62%

Fastest mode of travel by total trip distance

4.1.2 Proximity to City Center

In the following table, the *fastest* mode of transportation is calculated when only traveling within the given range of proximity to the city center. The network graphs presented in the Appendix provide visual representations of these findings.

Distance from City Center	Bike	Uber	Train
0 - 0.5mi	75%	8%	2%
0.5mi - 1mi	63%	11%	26%
1mi - 1.5mi	50%	4%	46%
2.0mi - 2.5mi	35%	25%	40%
2.5mi - 3.0mi	46%	25%	29%
3.0mi - 3.5mi	45%	33%	21%
4mi+	31%	30%	39%

Fastest mode of travel by proximity to city center

4.2 Potential Future Approaches

There are numerous other experiments that can be conducted with this dataset, provided more time and additional space in this report.

Calculations comparing discussed metrics vs. price is one such experiment. Santander bikes, for instance, are £2 per up to 30 minutes of riding. Compared to the range that Uber returns via its API endpoint, and the cost of a train ticket per Tube zone, an interesting price map could be created. From this data, it may also be possible to predict at what distance the price of an Uber is no longer worth it for a consumer.

Another set of experiments could include mixing the constraints in 4.1.1 and 4.1.2, providing more detailed trip analysis.

5. Conclusions

The first experiment in section 4.1.1 clearly showed that as trip distance increases, the usage of bikes as the quickest means of transportation quickly decreases. What surprised me, however, was that Ubering consistently ranked the *least* quick mode of transport regardless of travel distance. From the data collected, the tipping point for when bikes become less effective as the quickest mode of transport occurs after about 1.0mi-1.5mi of total trip distance. As distance increases, so does the ineffectiveness of relatively quick bike rides. It is important to note that in a congested city like London, that finding is not obvious.

The next experiment in section 4.2.2 provided data on the relative speed of various modes of transportation given initial distance from the city center. The data suggests that trips in the center of a city can generally be done much faster on a bike (as one might assume due to traffic and congestion). What is interesting, however, is that trips several miles away from the city center can generally be done in comparable time regardless of mode of transportation.

Observing the graph data available in the appendix, it is interesting to note that after about 2 miles of distance from the center of London, most trips in the bike dataset occurred without crossing through the city center. Bikes, therefore, are most likely utilized when the rider imagines the ride occuring intra-neighborhood, despite actually traveling a long enough distance to visit another neighborhood. This is another feature that would be interesting to explore further.

An hypothesis for why Uber's values are so low is that perhaps their API estimation returns modest (big) values (to ensure customers are "happily surprised" instead of angry when the trip does not take the estimated time). Ideally, getting access to both raw train journey data and uber journey data would provide a stronger baseline for the experiment. Nonetheless, the findings definitely provide riders insight on when and when not to utilize rideshare services.

See attached file for Appendix