Lesley Liu, Charles Ma, Sara Glattstein, Tommy Suen
Professor Andrei Lapets
CS 591 L1 - Data Mechanics
01 May 2018

CS 591 Project Report

Introduction

As college students, every once in a while, we need to destress and get outside of campus. For most colleges, that is not a problem, as their surroundings are well mapped out, with convenient entertainment venues and restaurants to cater to students' needs. However, this is not the case for Boston, which is one of the oldest cities in the United States. As such, it was not very well planned out when being built, and retains its lackluster structure, as exemplified by its notorious reputation for being difficult to drive around.

Luckily, Boston offers a couple of other transportation methods that make it easy for poor, car-less college students like ourselves to get around: the MBTA and Hubway. These cheap options let ourselves get around, saving money and being environmentally-friendly while doing so. All that is left is to provide a convenient way for students to find a way to get to venues. That is why our team decided to develop _____, which visualizes the closest Hubway or MBTA station to each venue. As an added feature, because Boston is an urban environment, we calculated the average time it takes to get from a station to a venue, to give people the option of choosing places closer to their respective closest stations, minimizing the time they'd have to spend exposed to an urban environment.

Base Datasets

To start off, we pulled data on all entertainment venues in Boston, as well as all the dining joints in Boston, from Analyze Boston. This gave us all the data we needed to create a list of destinations students could be interested in going to. Next, we pulled data on all the MBTA stations and Hubway stations across data, from their respective sites, giving us the info we needed to map out environmentally friendly routes students could take. Lastly, although we did not pursue it further, we pulled crime data on various areas across Boston, which we had

originally intended to use to further gauge safety, but had decided against it because the areas through which people were travelling were probably heavily policed, especially on weekend nights.

Algorithmic and Statistical Challenges

The datasets we pulled from various websites had a lot of extraneous data that we did not need, and so we did a lot of projections to just pull the data points from each entry that we needed for our purpose. After doing those projections, we performed optimization-based k-means on our two unified datasets, destinations and stations, to see which station was closest to each destination, helping students get around. Lastly, to increase convenience and to take students' safety in an urban environment into account, we calculated the z-score of the time it takes to walk between each destination-station pairing, to give users a better perspective of how far they would have to walk in an urban environment at night. Of course, this does not give a full view of how safe they are, as to do so, we would have to also incorporate crime rates in various areas of Boston. However, since most entertainment venues are in well-populated areas, we assumed there would be a decent level of police at those areas, meaning a relatively uniform crime rate among most popular destinations.

Conclusion

The original purpose of this project was to provide an additional tool that may prove useful for students to have a fun night out. As such, there are definitely limitations to the information it can provide, but that information can probably be easily substituted by other map applications, such as Google Maps. Our project eliminates the need for the student to go and find the closest station for a variety of venues, and allows them to immediately consider a more environmentally friendly alternative to cars when deciding on a transportation method.

There are definitely improvements that could be made, such as actually incorporating the crime dataset that we imported, to more accurately account for safety, as well as optimizing the k-means algorithm to better pair stations and destinations. Besides improvements/optimizations to the current project, additional ideas to implement could include an integration with the

Hubway API, to let users see the availability of bikes and docks at each station, so they can make a more informed decision, as well as integrations with real-time MBTA buses, like on Google Maps, again to help users make more informed decisions.

mbta: "Berkeley St @ Columbus Ave"