RedEye Service Final Report

1. Topic and questions:

For our final project, we decided to look into the RedEye service at Northeastern University. We unanimously choose this topic since we wanted to find something that pertains to everyone in the group as we are all students of Northeastern and also potential users of RedEye, we thought it would be interesting to delve into our school's services and learn a bit more about it through our data wrangling process.

Over the course of the project, we will see how effective the service is and if there is any room for improvement to further better the service. In addition, we will also look into how accessible the RedEye is, the safety of students on campus, and learn more about any other alternative commuting options available. Ultimately our analysis is something that will be important to the students (as we are the users) and pertinent to the stakeholder of Northeastern especially if the school is looking into making changes to transportation on campus.

2. Data:

The data sets that we will be using to help further guide our analysis is a mix of data that we collect on our own and from outside public sources.
The primary data set is what we collected ourselves through a survey that was created to learn more about the RedEye service and the interaction of students with the service. We find out more about how popular the RedEye is truly amongst our school community, the times that students leave the campus and all the alternative methods of transportation that they take to go back home. We looked into the zip codes that students provided to see how many students live outside the 2-mile radius.

Our secondary data sets are public data sets, with one being from Analyze Boston, which is the crime incident report and the satisfaction of passengers of MBTA, taken straight from the MBTA website. For the crime incident report, we look into the number of crimes that occurred from the start of 2022 to the present and merge that with our survey information. We look into the times that crimes occur and how that correlates with the time that most students leave the Northeastern campus to help us gather information on the safety of students within the city of Boston.
In terms of the data set taken from MBTA, we wanted to look more into how satisfied passengers are really with the MBTA service as our data tells us that majority of students rely on taking the T to get home from school, obtaining an understanding of people's sentiments to public transportation and seeing if we can leverage this to get students to be more interested in using the RedEye and find ways to improve the service given any shortcomings.

Data sources:

1. https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system
2. https://mbta-massdot.opendata.arcgis.com/datasets/MassDOT::mbta-service-customer-satisfaction/explore

3. Information quality:

Regarding the crime incident report data set, after running a Profiling analysis on Python, we obtained that it has 17 variables, 10 categorical, 5 numeric and 2 unsupported. Moreover, the data set contains 63,367 observations and 126,886 missing cell values which represent 11.8% of our data. However, 99% of the missing values are located in intern variables, which we are not going to use in our analysis. In the case of the RedEye survey, we established 12 variables and obtained 114 responses. Moreover, we have 19 missing value cells, which represent 1.4% of the total cells. The variable that contains more missing values is “Neighborhood”, however, although we do not have the neighborhood of each student, our variable of zip codes is complete, so we are using this variable for our wrangling process.

1. **Privacy and confidentiality:** One of our data sets is a survey from students about their commuting from school to their house and their knowledge about the RedEye service. In this survey, important information is the address of each student, so we can identify exactly if it is covered by the RedEye service. However, we were aware that asking for their specific address can lead to a breach of information and that many students would not be willing to share it. To overcome this problem, we will only ask for the zip codes, which will then be linked to our second data set of Boston neighborhoods. However, not having each student's exact address is a challenge to identify if the radius of the RedEye is covered or not.

2. **Barriers to data accessibility:** RedEye service does not have a website set up but is only available as an app. Other than information about the service provided within a 2 miles radius (with Snell Library as a central location) covered, we do not have much data. We have decided to use a map provided on the app to manually trace the areas covered and link them to our second data set of Boston neighborhoods.

3. **Interpretability and understandability:** Our third data set is crime rates across Boston neighborhoods, which we want to link with our survey of students to narrow down areas that are either safe or unsafe while also adding in the time factor. However, this data set does not include the zip code, just an “Area Code”. In order to assign the zip codes, we are going to use Python to generate that new data.

4. Methods and tools:

To understand students' familiarity with the RedEye service and which mode of transportation is being utilized most often daily, we created a survey to collect the responses from Northeastern students. We used the response file which is generated from *Google Forms* to collect the current zip code from students. This zip code information helped us with the following critical progress to inspect the safety in each neighborhood in Boston near the Northeastern campus.

First, the zip code information from the survey helped us match the 2 miles radius from the RedEye service. We used data formatting in this part to obtain a clear message in this survey.

Second, in the crime data set, there is information about which neighborhood has the most incidents, and we applied the information to see if there are matching residence zip codes with where most students live.

There was a challenging part for us to map out the residence zip code within 2 miles radius of RedEye Service since the zip code not only covers a residence address but a wider area as well. In this case, there was difficulty to match the boundaries of the 2 miles drop-off radius. We want to make sure the boundaries of the RedEye service are beneficial to Northeastern students, to tackle this problem, we adopt the SearchEngine, geocode, zip code, warnings, geodesic, DistanceMetric, radians to get the center of each zip code, and convert to the longitude and latitude, so that we can measure if specific zip code is covered by the 2 miles dropoff radius.

With regards to the MBTA data set we mainly used excel to clean data such as removing unnecessary columns, calculating the number of responses from percentages, reformatting the questions, and creating multiple subsets. This was then used to create charts for comparison and visualization.

The other main purpose is to check the regions where students are safe enough to commute by themselves via other public transportation. Thus, we applied the same measure method to obtain the distance with the location where most crimes happened.

5. Data Wrangling Process:

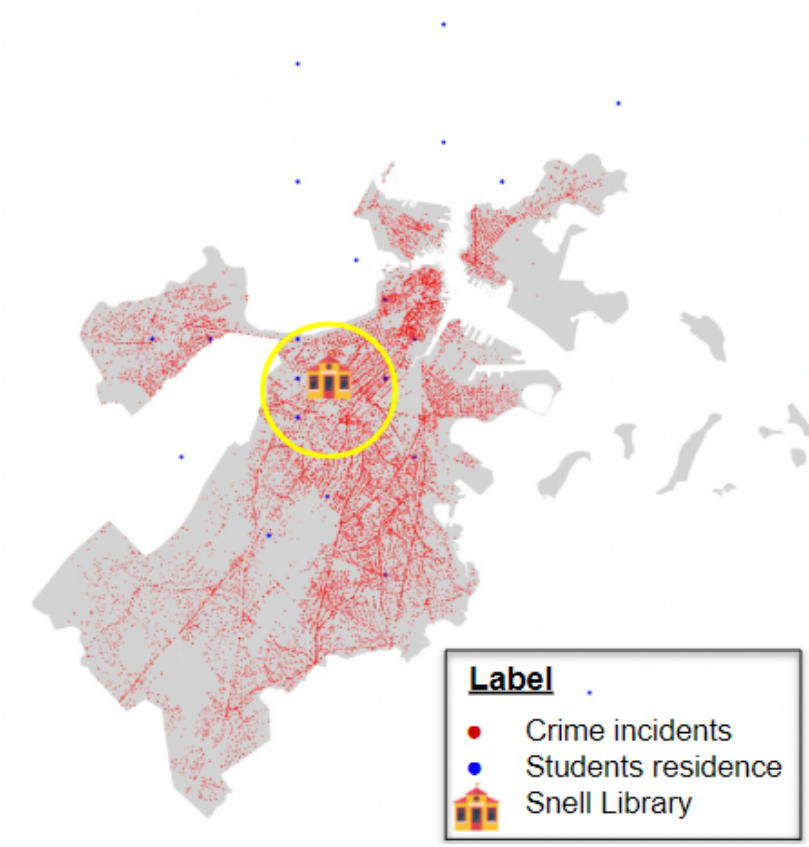
The data wrangling process involved a lot of cleaning from our end to be usable in python. At first, we had to rename columns from the redeye survey as it would be difficult and impractical to use the original column names (example: 3. What type of student are you?) every time we run a code in python.

Next, we dropped irrelevant columns from the crime dataset like districts or neighborhoods among others that were redundant or had too many missing values. A subset was created (Crime_22_subset) that stored all the relevant information like the latitudes and longitudes of the crimes committed that were grouped by the offense description.

Additionally, we encountered many invalid or outlier zip codes which was a hindrance when we plotted the heat map as the code accounted for all the outliers as well, making the visualization difficult to infer. We removed the outlier zip codes as they were not useful and standardized the invalid zip codes with correct zips from the red eye survey.

Lastly, to mark the crime incidents on the heat map, there were many (around 3000) invalid latitudes and longitudes from the crime dataset (having coordinates like (0,0)) that we had to remove as well. After all this wrangling we had two clean datasets we could work with to plot the heat map and gain more insights visually.

To finally plot the map, we needed the cleaned datasets and created subsets from each of them containing the latitudes and longitudes (by converting their zips using search engine()) and computing the minimum distances between the offense and students' residence.



The above map was plotted using GeoPandas after using the cleaned datasets (redeyeupdated.csv, crimeupdated.csv)

- The red markers represent the crimes committed
- The blue markers represent the students' residence
- The green marker represents Snell Library along with the red eye radius

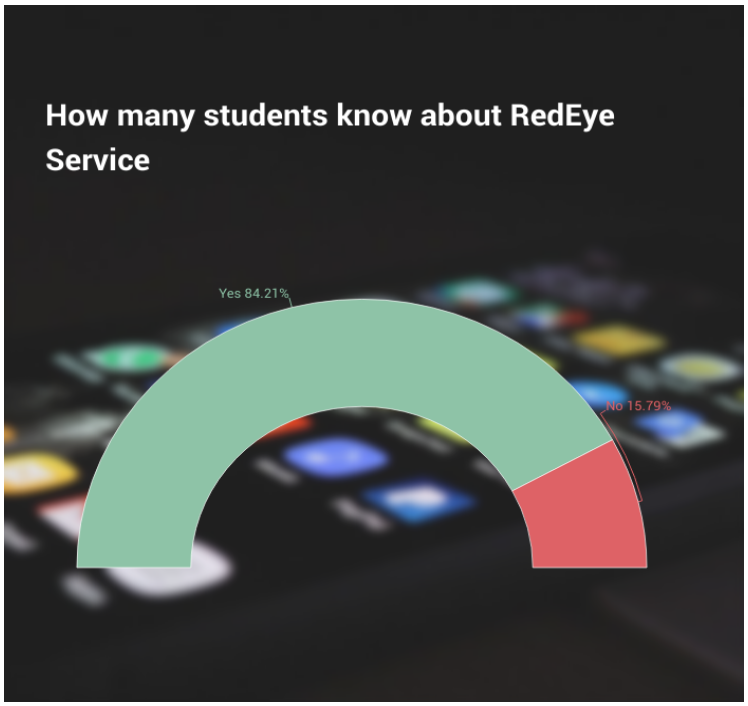
6. Analysis and results:

We have analyzed the crime data set, student survey, and MBTA data set.

Analysis of student survey:

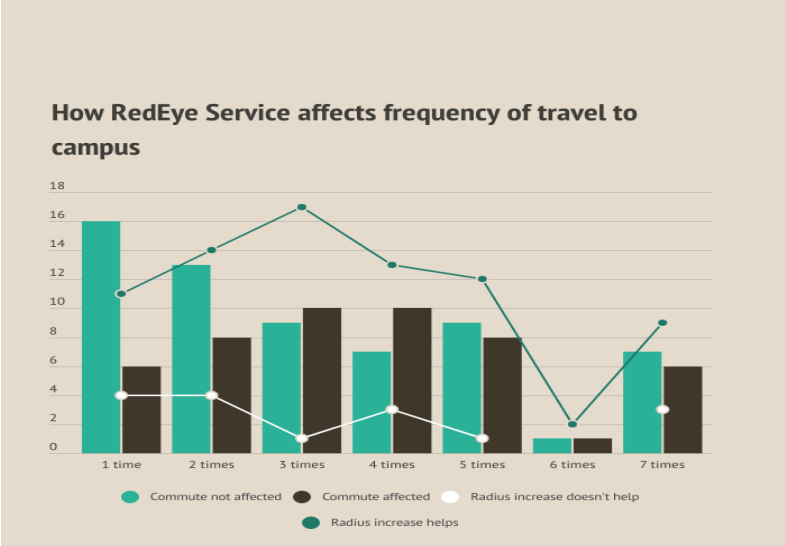
We received a total of 114 responses from the survey shared with all Northeastern students. Our questions mainly focused on their awareness of RedEye service, safety during their commute, accessibility, and alternate modes of commute.

Awareness:



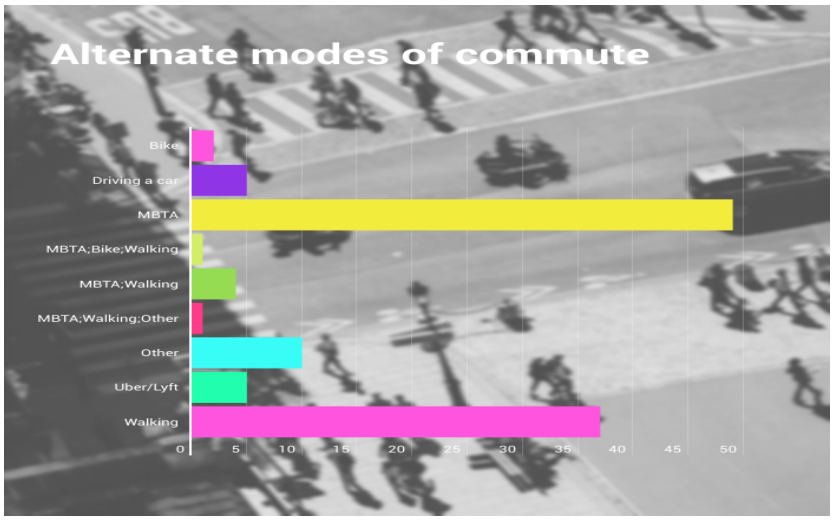
There are still about 15% of students who do not know about RedEye services. Mainly graduate students tend to stay back at the university until late at night and have a higher travel rate to campus.

Frequency of travel:



We wanted to check if there is a correlation between students' travel frequency and their options of commute at night. From the chart, it's clear to see that as the frequency of travel to university increases their commute is also affected.
Post this question on our survey we asked if RedEye were to increase their 2-mile coverage, would it help them or not and we can see that many students responded positively to this suggestion.

Alternate mode of transport:

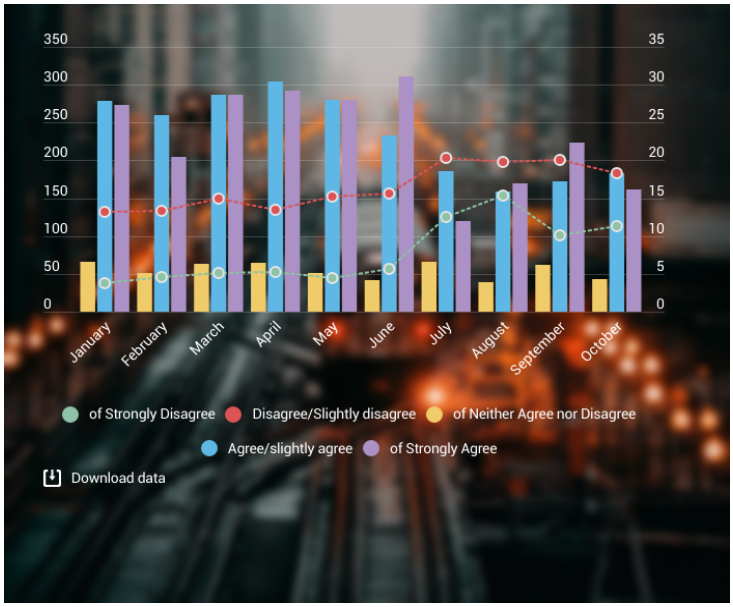


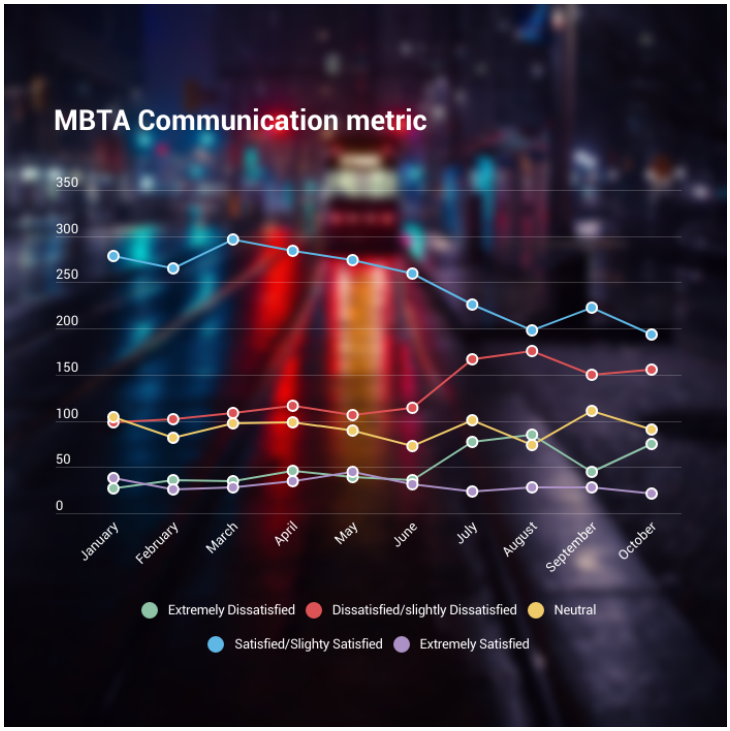
We wanted to check how students in a 2 mile radius are commuting to university, from the above bar chart we can see that the majority use the MBTA right after walking. Walking is still preferred by those who live nearby campus. Also students have chosen different commute options depending on the time of departure such as if they leave at 5pm they walk but the same person would use the MBTA if it's late at night.

Analysis of crime data set: *(Please refer to the heat map included under the data wrangling process)*

From our analysis of the student survey we found that 49% of them still feel unsafe while commuting. For this purpose we wanted to use our crime data set to find how far or the rate of crime is around their accommodation.
This would help us emphasize the need to extend RedEye service operations. From the plotted map we could analyze and gain insights on the offenses committed across the Boston area with the red markers. The students' residence is visualized by the blue markers, however we cannot visualize the population density as the zip codes stay the same for all the set of students in an area. To overcome this we created a pivot table on Python to group the areas by the student count. After this we could infer that around 48% of the people who took our survey reside in Roxbury, Mission Hill, Southend and Jamaica Plain. These areas stand out as having the most serious crimes committed as well. As there are many students who reside here it's safe to assume why the Redeye service is needed more in these areas.

Analysis of External material MBTA:





As the MBTA was chosen as one of the most used alternate commute options,we wanted to compare and analyze their passengers. The two metrics we used were “How good is the MBTA's communication ?” and “How reliable is MBTA service ?” For this, we used the public data set and cleaned/formatted it as per our needs to produce the charts shown.

The adjacent chart shows how reliable passengers think the MBTA is. We can see that there is a good portion of people who trust the service but still a good amount who disagree.

We can see that there was a steady increase in dissatisfaction of passengers with MBTA’s communication from June to August of thai year, this usually includes any announcements made regarding incidents/delays/future shutdown of lines.
Through this, we can conclude that though MBTA is a good alternative option for commuting students still would be affected by a slight change in weather or incidents.

7. External material:

The purpose of this analysis includes the safety surrounding the campus. We already knew there are numerous articles focusing on some neighborhoods nearby Northeastern that aren’t the safest, which is our main focus to know if RedEye service can benefit our students.

There is an article about the safety analysis surrounding Northeastern[1], it is clear to see in the student-related crime analysis, reported the overall crime rate of Northeastern, which reported 81.1% on campus crime data, and 17.8% of Boston campus neighborhood.
Northeastern University is situated in Roxbury and Fenway, there is the significant article mentioned Roxbury is reported as the most dangerous neighborhood throughout the city[2][3], Estimates suggest around 5000 crimes are committed per 100 thousand people according to the data in 2021. In this context, we can assume that most students who reside around campus may feel unsafe, especially while commuting to and from campus.

There is more discussion we can take regarding the safety around the Northeastern campus[4][5]. Most of the students/alumni who currently live in Boston gave comments that they felt fairly safe at night, but at Ruggle station which is situated within the campus, students felt unsafe when it gets dark, and they mentioned RedEye service is a good commute option to use at night.

And the other supportive resource is the Boston crime analysis report found online, which covered the incident records in each district[7], it also had a clear chart about the amounts of incidents that happened each month in a year, and deeply analyzes the amount of the incident throughout the week. And it is interesting to mention that data reported some other high crime rates on the streets, and the highest crime rate is on Washington street.

Since the majority of responses from the survey showed their commute option is MBTA, we found out various articles recently pointed out the safety concern[8], which has Boston

threatening the safety of riders and killing one passenger. The problems led to a rare month-long shutdown of a subway line for emergency repairs and upgrades. And it influenced students in Northeastern whether the MBTA is still the safest choice to commute.

The main MBTA line passing via the Northeastern campus is the orange line and green line. After the orange line shut down for 30 days for its construction project, they established some new slow zones to identify additional problems during the slowdown. And the agency reported that speed restriction delays will remain through the end of the year, which causes inconvenience for passengers[9]. I think it may have a negative influence on students since it is at least 11 miles away(From the farthest station - *Oak Grove* to the nearest station-*Ruggles*.) from Northeastern University.

Another MBTA-related data, which is about the prediction accuracy data of rapid transit and bus[10], regarding its accuracy, it mostly has achieved roughly 80% accuracy, but the waiting time varies, it is critical, if the waiting time fluctuates as students need to commute to their place.

8. Additional data and analysis:

Regarding the additional data, we would like to obtain permission from Northeastern so they can release the raw data collected from the RedEye app. In this way, we can have more information regarding the number of trips, the peak times, the duration of trips and the number of cars used daily. However, because it is a private resource from the university, we know that the permission is going to take time to get approved, and there is also a possibility that we would not get it. Moreover, we would like to replicate the MBTA Satisfaction Survey with students from Northeastern, especially because the MBTA is one of the most common commuting options that students use at night. In this way, it is important for us to know from each student their perspective of the MBTA rides, especially the orange and green lines.

In addition, another future development is related to the mobile app called Transit. We believe it would be beneficial to add the RedEye service as one of the commuting options. In this way, the students that log in using the NEU account can add RedEye as an option. In addition, we would suggest enhancing the function to track the RedEye car in the app (the app has a function to show the live location of the car or train depending on what you choose). In this way, students can have enough information to choose the best commuting option, as well as ensure their safety.

External references:

[1] https://www.collegefactual.com/colleges/northeastern-university/student-life/crime/
[2] https://www.reddit.com/r/NEU/comments/8a4h68/safety_on_and_around_campus/
[3] https://www.reddit.com/r/BostonU/comments/oitkc6/is_it_safe_around_campus_in_boston/
[4]https://estatousa.com/bad-neighborhoods-in-boston/#:~:text=Roxbury,the%20most%20hazardous%20in%20Boston
[5]https://huntnewsnu.com/59576/campus/nupd-develops-new-version-of-redeye-app-for-off-campus-students/
[6] https://www.areavibes.com/boston-ma/roxbury/crime/
[7]https://rstudio-pubs-static.s3.amazonaws.com/453629_6be32e64d25b4e7189bc3bbe6968bcaf.html#offense-category
[8]https://www.washingtonpost.com/transportation/2022/08/31/boston-transit-mbta/
[9]https://www.wbur.org/news/2022/11/14/mbta-t-orange-line-green-slow-zone-delays
[10]https://mbta-massdot.opendata.arcgis.com/datasets/MassDOT::rapid-transit-and-bus-prediction-accuracy-data/explore
[11]https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system
[12]https://mbta-massdot.opendata.arcgis.com/datasets/MassDOT::mbta-service-customer-satisfaction/explore