Project Title: Make BART Great Again

Analize BART's efficiency compared to the top three metro lines from around the world.

Team:
Eric
Mana
Matt
Rohan

Project Description/Outline:
How to make BART more efficient by analyzing the top 3 metro systems around the world. According to this CNN article, the top 3 metro systems around the world are: New York, London, and Sâo Paulo according to user satisfaction. What makes these systems so user friendly compared to BART? How does BART incorporate some of the key findings we will uncover.

We will take a look at these 3 cities and compare them to BART in order to make suggestions how to improve the public transportation offering in the Bay Area.

$\mathbf{H_0}$ = BART is the worst subway system in our sample.

Research Questions to answer:

a. How many destinations does BART serve compared to other metro systems?
    i. eric
b. What is the total track length that BART offers compared to the other metro systems?
    i. eric
c. BART (Rider Customer Complaints Pie Chart)
    i. Mana
d. Operating cost per passenger per mile comparison
    i. Matthew
e. Use metro population compared to total city population ~~and total track length (coverage/service area) as a metric to normalize the metro usage of each city.~~
    i. Matthew will do a stacked bar chart for this visual

Datasets to be Used:

1. https://www.kaggle.com/citylines/city-lines
2. ~~https://www.kaggle.com/cityofLA/los-angeles-traffic-collision-data~~
3. https://www.kaggle.com/PromptCloudHQ/world-happiness-report-2019
4. ~~http://media.metro.net/about_us/vision-2028/Report_2017_Customer_Survey_Final_2018-0103.pdf~~
5. BART Rider Reports
6. BART Open Data
7. https://www.kaggle.com/thiagodsd/sao-paulo-metro

8. https://www.kaggle.com/new-york-state/nys-metropolitan-transport-authority-mta-data
9. https://developer.transportapi.com/ (UK transport API)
10. https://data.cityofnewyork.us/browse?Dataset-Information_Agency=Metropolitan+Transportation+Authority+%28MTA%29 (NYC subway API)
11. http://www.metro.sp.gov.br/en/pdf/sustainability-report-2015.pdf (Sao Paulo Survey)
12. https://www.metrotransit.org/data/sites/1/media/blog/metro-transit-rider-survey-2014---final.pdf (Sao Paulo Survey)
13. ~~http://media.metro.net/about_us/vision-2028/Report_2017_Customer_Survey_Final_2018-0103.pdf (Los Angeles Customer Sats Survey - good comp info.)~~
14. https://developer.transportapi.com/documentation/tube-information (Wait times for London)
15. https://www.londontoolkit.com/briefing/underground.htm (use to calculate London underground fare)
16. http://content.tfl.gov.uk/tfl-annual-report-and-statement-of-accounts-2018-19.pdf (annual report tfl)
17. https://en.wikipedia.org/wiki/New_York_City_Subway - NY City system track length in miles and km.
18. http://web.mta.info/nyct/facts/ridership/ - Subway Traffic breakout
19. http://web.mta.info/mta/news/hearings/pdf/MTA_Reinvention_Report_141125.pdf - MTA "Reinvention Report"
20. https://new.mta.info/about-us/the-mta-network - MTA at a glance - 2018
21. http://dashboard.mta.info/ - MTA dashboard of metrics info
22. https://ny.curbed.com/2017/10/13/16455880/new-york-city-subway-mta-operating-cost-analysis - operating cost per line NYC
23. http://content.tfl.govNYC subway operating costs: an analysis.uk/csopp-20170713-part-1-item12-tfl-international-benchmarking-report.pdf - TfL - London "TfL International Benchmarking Report"
24. https://www.findproperly.co.uk/tube-map.php#.XY1DdkZKhPY - pricing for housing per line around London Tube - would like to find the raw data
25. https://en.m.wikipedia.org/wiki/London_Underground_rolling_stock - track length, seating capacity
26. https://tfl.gov.uk/info-for/open-data-users/our-open-data?intcmp=3671#on-this-page-2 - tube info
27. https://www.bart.gov/sites/default/files/docs/BART%20SCOA%20Final%20Report%20June%202013.pdf

~~**Visuals-**~~

~~**Research Answers Visualized/Statistics:**~~

~~Average wait time per station - Normalize data, standard mean? Standard error of mean?~~

~~Cleanliness (Rider satisfaction scores)- Pie chart of grouping~~

~~City happiness rank with metro cities in our study - GDP Country Bar Chart~~

~~Population Density compared to rider data to understand a baseline of % of population of riders - scatter plot? Stacked bars?~~

~~Use metro population compared to total city population and total track length (coverage/service area) as a metric to normalize the metro usage of each city - comparison statistics?~~

~~[T-Test and/or ANOVA to double check result.]~~

Breakdown of Tasks:

Technical Requirements:

Use Pandas to clean and format your data set(s)
Matt

Create a Jupyter Notebook describing the **data exploration and cleanup** process
Matt

Create a Jupyter Notebook illustrating the **final data analysis**
Rohan

Use Matplotlib to create a total of 6-8 visualizations of your data (ideally, at least 2 per "question" you ask of your data)
Eric
Mana
Rohan - will do data/statistical analysis

- *Time series anomaly detection** - (unable to provide without historical data)
- Forecasting (if historical dataframe is provided)
- ttest/ztest/anova

* [ ] Save PNG images of your visualizations to distribute to the class and instructional team, and for inclusion in your presentation
Eric
Mana

* [ ] Optionally, use at least one API, if you can find an API with data pertinent to your primary research questions
Rohan

* [ ] Create a write-up summarizing your major findings. This should include a heading for each "question" you asked of your data, and under each heading, a short description of what you found and any relevant plots.
Eric can Lead / Team

Presentation Outline:

- Background/Executive Summary
- Null hypothesis
- Data sources/cleaning
- Key question clusters
    - Visuals and statistics
- *ML Analysis*
- Reject/DNR Null hypothesis
- Holes in data, what is incomplete, challenges
    - Small sample size (just the top systems)
    - Limited datasets

BART = Bay Area Rapid Transport

- Does not cover the entire bay area
- Not rapidly transport