ETC1010 Group Project (T10_Tue_gganimate)

Group Members: Rounak Agarwal, Jason Ching Yuen Siu, Yingqi Luo, James Bathgate

Dataset:

Source of Detention dataset:

https://github.com/themarshallproject/cbp-migrantchildren-detention-data

License: *MIT License Copyright © 2020 The Marshall Project*

Source of dataset for Reddit comments (Scrapped using *RedditExtractorR* library):

https://drive.google.com/file/d/1opbouhgp-cjeNXzUpJF75nA14RDshaN1/view?usp=sharing

Original post - 60,000 child migrants detained by US in last 40 days

https://www.reddit.com/r/politics/comments/c4xi7j/60000_child_migrants_detained_by_us_in_last_40/

Questions:

Detention Data

1. We firstly analyse the children detained more than 72 hours by grouping them on the basis of nationality and age group.
2. Find the average number of hours the children were detained over time (The proportion of the children who were detained more than 72 hours and less than 72 hours)
3. By how many hours did the detention exceed the 72 hours limit on average?
4. By how much did the detention barrier increase by, over the past four years?
5. Could we find any trajectory for the total detentions over the limit against time? What would have been the number of wrong detentions in America in Jan 2020 if this case wouldn't have come into the limelight in May 2019?

Reddit data

6. Analyse the sentiment of the people from a reddit post posted after the news regarding the detention broke out to the public?
7. What are the most important keywords?

Description of the data:

1. We get some relevant data and the data contains two key sources from the U.S. Customs and Border Protection:
   a. CBP Office of Field Operations (OFO) child detentions: Detentions of children at ports of entry by the U.S. Customs and Border Protection Border Patrol during mid-January of 2017 and late January of 2020.
   b. CBP Border Patrol (BP) child detentions: Detentions of children between ports of entry by the U.S. Customs and Border Protection Office of Field Operations during mid-January of 2017 and mid-June of 2020.

2. For Analysing the sentiment, we scrape the comments from a post from Reddit using *RedditExtractorR* library in R.

Reference:

Flagg, A. (2020). 500,000 Kids, 30 Million Hours: Trump's Vast Expansion of Child Detention Retrieved from https://www.themarshallproject.org/2020/10/30/500-000-kids-30-million-hours-trump-s-vast-expansion-of-child-detention