Project Check Point

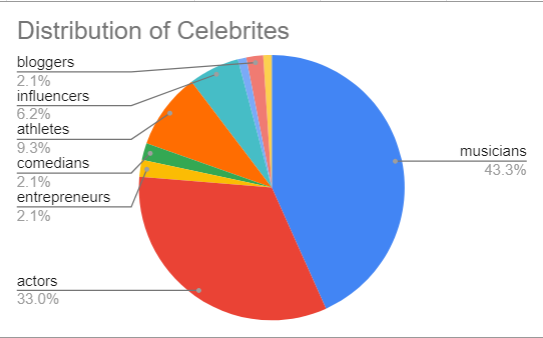
Team Abnormal Distribution: Charles Phelan, Krish Shah-Nathwani, Jacob van Steyn, Matthew Reynolds

Our Data:

We gathered the username and 270 tweets each from almost 400 US politicians and the top 100 most followed Twitter users in America. It is important to note that we excluded brands, for example, The New York Times, and decided to only focus on actual people. We decided to only take data from US celebrities, as we are comparing their political leanings to US politicians. To clean our politician data, we removed extraneous information, and deleted politicians who were missing any information. This left us with 200 Democrats and 197 Republicans.

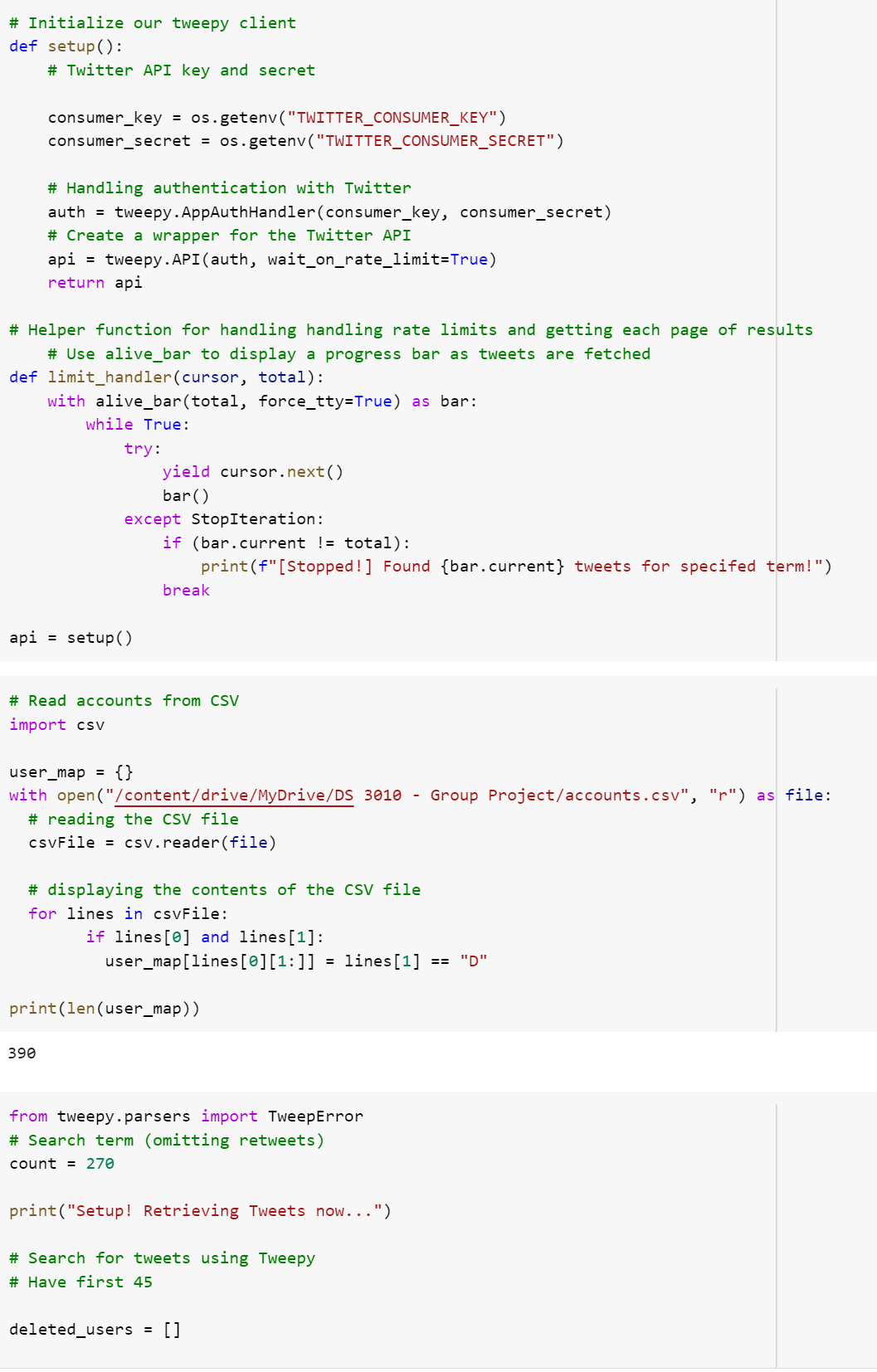
Insights:

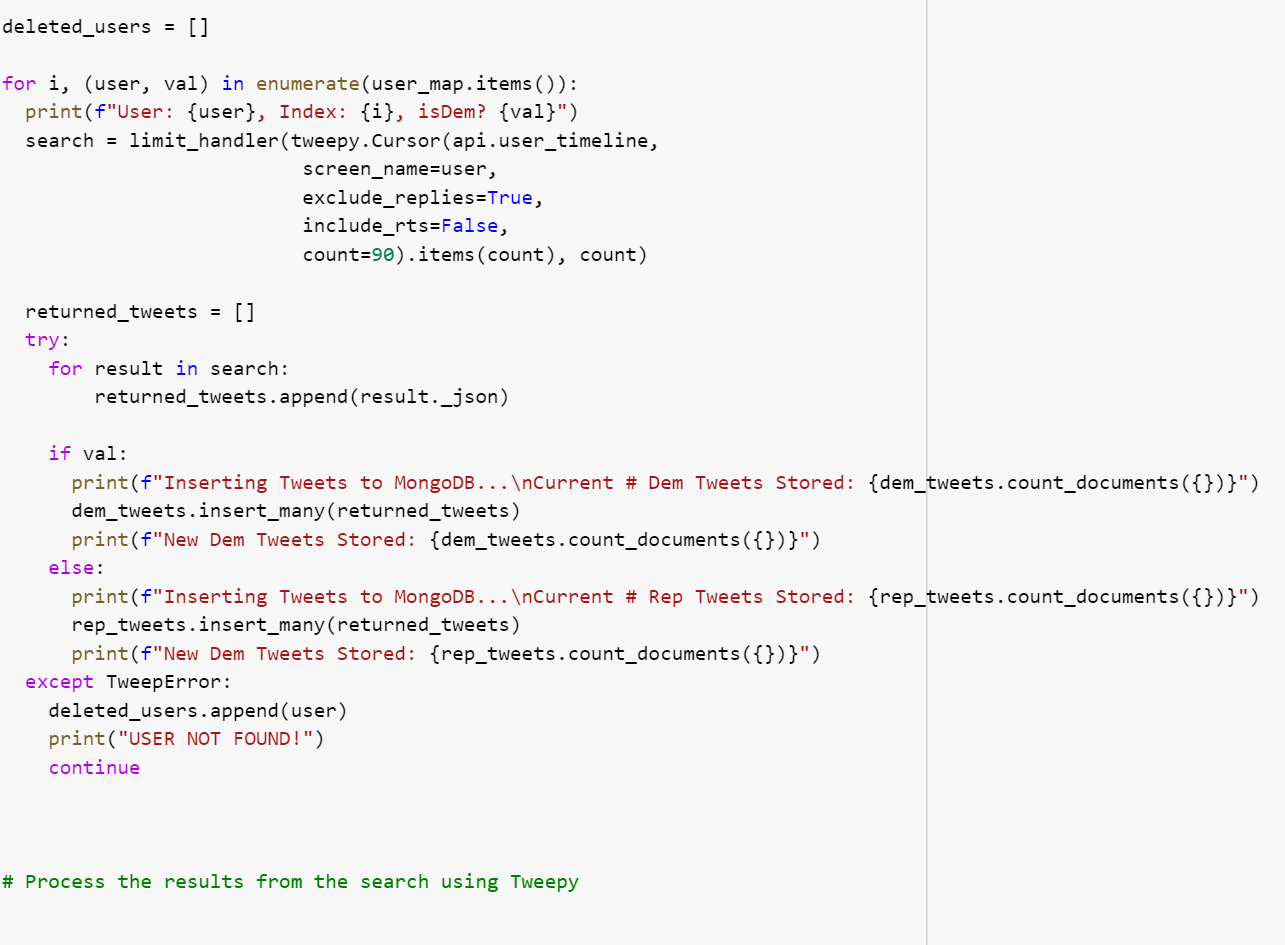
There are 45057 Republican tweets stored, and 45327 Democrat tweets stored. This is a fairly even distribution given the sheer amount. We have also categorized each of the celebrities we will be investigating. It is important to note that the small blue unlabeled sliver is the amount of authors.



The aspect of this data that sticks out the most is that the majority of our celebrities are musicians and actors. We would like to further investigate if profession has any sort of effect on political leaning.

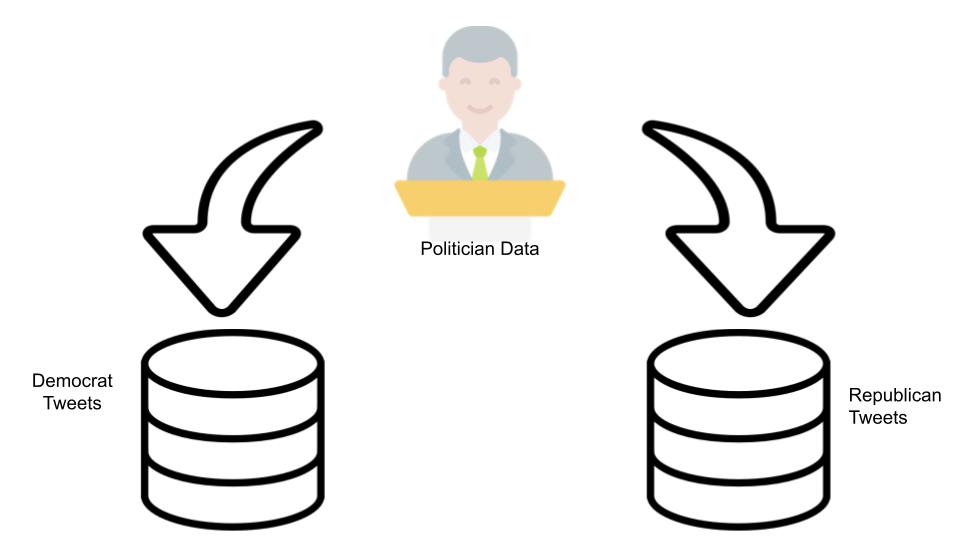
ScreenShots:



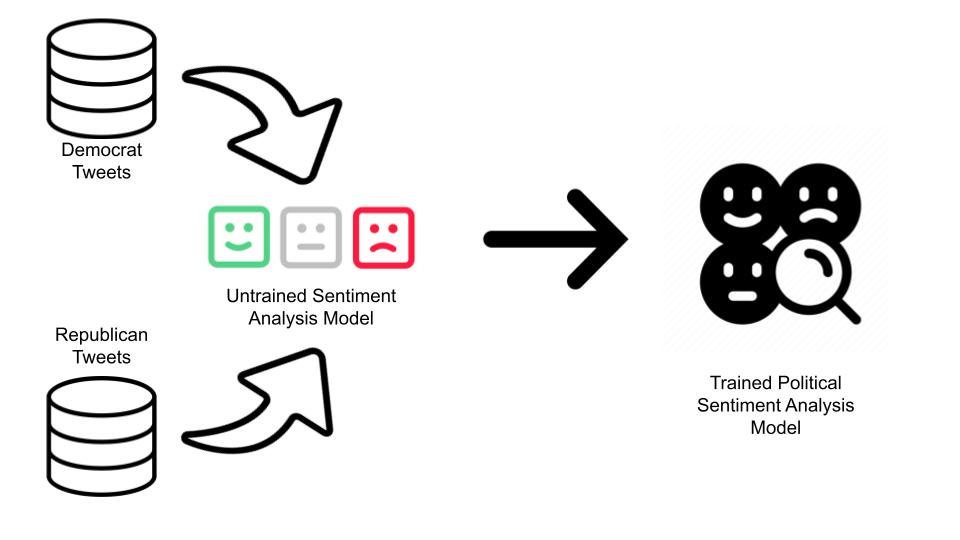


Sketches of Interactions:

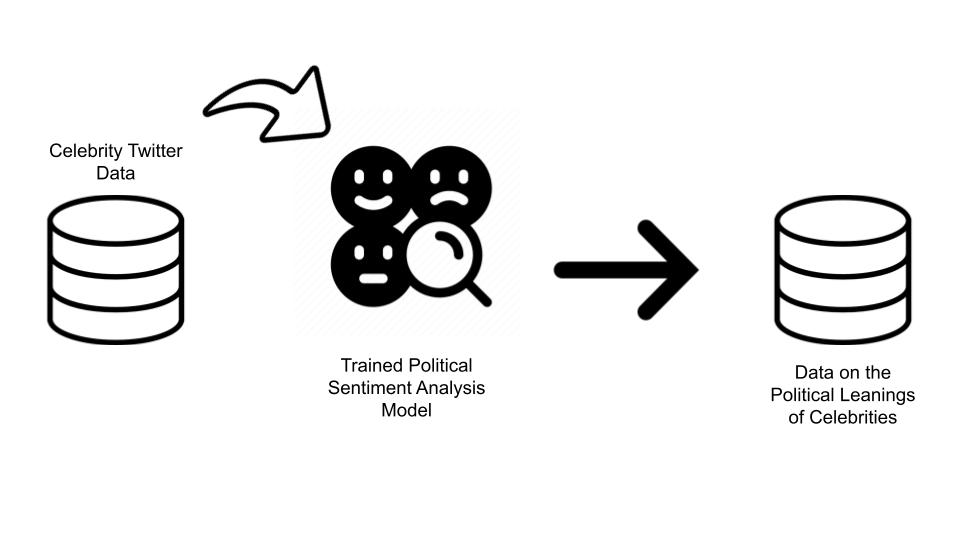
Split Politician Data for the Analysis Model



Train Analysis Model with Politician Tweets



Use Trained Model to Gather Data on the Political Leanings of Generic Celebrities



Next Steps:

For our next steps in the project we will take the data we took from the politicians using the twitter api and run it through our model in order to train it. Then we’ll run the data we gathered from the top American celebrities on twitter through it and classify them as either a Republican or a Democrat. Then after that we’ll dive into the analytics and see if things like gender or age have any correlation to their political alignment.