**Project Plan Deadline 18th August**

*Aim: scrap twitter for Brexit related sentiments vs. UK bond prices & EURO bonds.*

Objectives

* Create automatic twitter scrape for #Brexit and other measures
* Stream prices
* Try to find relationships
* Process the language [sentiments]
* Create visualisations
* Segment secondary/primary/joe boggs
  + **What sources are the biggest influences**
* Which bond is most fluid

Create a report, using plotly: cleaned and plotted.

Scope

* Jupyter to test
* PyCharm to run

UK & Euro bonds sentiment on Brexit.

Include other indicators in scrape, but defer decision on using them for time being.

1st iteration: Brexit influences on prices.

Assumptions: that will be significant Brexit articles coming out in the next few weeks.

Approach

Min 2 weeks deadline

1. Planning:
   1. Research: finding sources, e.g. google finance:
      1. Scrolling through other peoples research; twitter sentiment projects
   2. Defining keywords
   3. Investigating python modules
   4. Assign roles
   5. Write contingency plan
2. Gathering data:
   1. Writing the scraper for twitter
      1. Storing in SQL serving
      2. Running on a backup
      3. Set up team view, check running
   2. Create & maintain SQL server/database
      1. How is this shared?
   3. Writing scraper for bond prices:
      1. Find website with relevant bonds
      2. Use beautifulSoup to grab tables
      3. Rearrange as required
      4. Append to SQL db with new row for each time point
3. Language Processing
   * 1. Understand Vader/Text blob
     2. Research around NLTK library for relevant sections to aid with analysis.
     3. Write a mini-report: what is relevant, how to us eit.
     4. Test on news article
     5. Write productionised code
4. Sklearn:
   * 1. **RETURN TO LATER**
5. Write analysis:
   * 1. Replan analysis
     2. Write productionised code
6. Create visualisations
   * 1. Divided between the three

Assumptions

* Bond prices can be streamed
* There is some meaningful relationship between bond prices and Brexit sentiment.
* Everything is possible (re. coding)

Dependency

* Before the scraper for productionised bond prices:
  + Define the source
  + Define the method of scraping
  + Running SQL database
* Before the productionised twitter:
  + Define keywords
  + Have running SQL database
* Language Processing:
  + Not dependent on other factors
* Sklearn:
  + Classes on Sklearn
* Analysis:
  + Dependent on everything before
* Visualisation:
  + Last thing we do

Estimating & Scheduling

See Gantt Chart