**DS 1007 Project**

**Project Outline**

----

**Project goal:** Compare the sentiment of Twitter conversation across four competitive companies using python data visualization techniques and analysis.

**Data Source:** Twitter data pulled through [Talkwalker](https://www.talkwalker.com/)

**Process:**

The team will:

* Gather the data from Talkwalker
* Cleanse and prepare each competitor dataset
* Visualize the data by sentiment category (positive, neutral negative) including variations by competitor, over time, etc.
* Analyze the outputs \*\*\*

---

\*\*\* Do we need a business goal? I am not sure how much of that sort of thing is required?

---

**For our team**

**The submission:**

We have to submit the proposal as a PDF on Monday - the project brief also lists “slides quality” as part of the grading rubric. I didn’t realize we needed slides for the presentation so let me know if I am reading that properly. Happy to build out a deck if necessary if we align on the above.

It also says “duration” meaning we have to fill the 5 minutes. @Emma you said you would be cool with presenting?

**Note on complexity:**

As of right now I think this is a little too simple of a project. I could complete this end-to-end in about two hours. I would like to discuss some potential add ons:

* Simple: Adding categorical analysis to add more visualizations. It might not be thrilling but it would be a simple way to demonstrate more work and use different python methods besides just visualization methods.
* Simple: Displaying most shared/Retweeted posts by duplicate text in the content field
* Medium: Image frequency analysis (Talkwalker provides direct links to images so we could look at duplicates here somehow)
* Hard: Plotting post locations
  + Would require data transformation and learning new mapping visualization

**About the data:** I have highlighted some of the key columns but we can also discuss incorporating more features if we want. The sentiment outputs a number, I forget the scale but it ranges from most negative to most positive, I will look this up and update you. We can use those numbers to classify as positive/negative/neutral or we can do more breakdowns.

Here is some documentation on the automated sentiment provided by TW

Technical: <https://talkwalker.digitalexcellencecenter.com/home/sentiment>

Analysis: <https://www.talkwalker.com/blog/sentiment-analysis-guide>

**For team to discuss: Select a group of companies**

I (Jenna) personally am leaning toward some of the smaller brands in the meal kit delivery service and personal care categories, partially because they might have more sentiment attached and volume is lower.

With some of the larger companies the twitter convo will be dominated by shares of news articles that may be of less interest.

Open to more suggestions as well!

Pharma:

* Johnson & Johnson
* AstraZeneca
* Pfizer
* Gilead

Tech

Tech is pretty broad and big companies like Google, Apple, Amazon are challenging to analyze so I would recommend if we want to go with tech we discuss a more specific vertical.

Gaming:

* Playstation
* Atari
* Nintendo
* Xbox

Automotive:

* Tesla
* GM
* Ford
* Fiat Chrysler

Ice Cream:

* Haagen Dazs
* Ben & Jerrys
* Breyer’s
* Talenti

Meal Kit Services:

* Blue Apron
* Hello Fresh
* Green Chef
* Marley Spoon (Martha Stewart’s new service)

Disruptor personal care:

* Birchbox
* Stitchfix
* FabFitFun
* Ipsy