
University of Texas

Data Analytics and Visualization Bootcamp

Group Six - final_project

https://github.com/BooneyBeCoding/final_project.git

Description

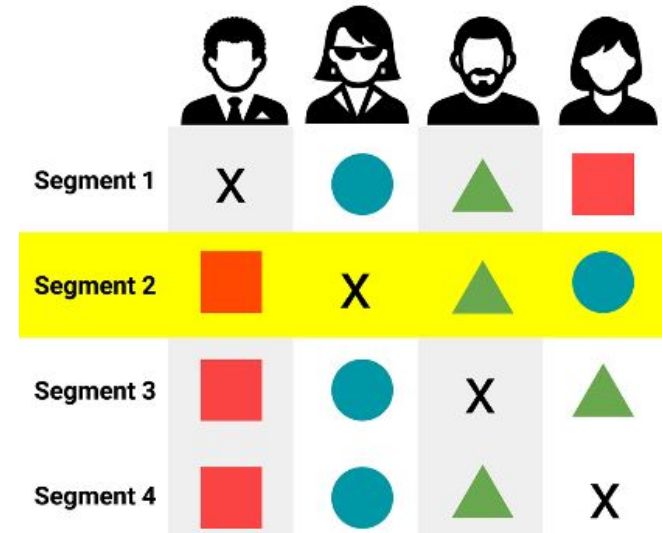
Investigate whether or not sentiment analysis data, generated by VADER, can be used in a basic machine learning model to predict post origins across multiple Reddit discussion groups (subreddits) that share similar topics but different cultures.



Group Six Team

The Group Six team consisted of four members:

- Adam Boone, Square Role
- Brendan de Vidal, Triangle Role
- Paul Smith, Circle Role
- Richard Wyatt, X Role



1. Proposal

- The 3 most prominent subreddits (investing, stocks, & wallstreetbets), whose communities are built around the topic of stock market investing each have different constructions of culture.
- By running a sentiment-gauging cross-analysis over the headline/title of posts of each, can we teach a Machine Learning model to discern which of the subreddits was most likely the subreddit of origin?

2. Technologies

1. **Kaggle** - datasets
2. **Jupyter** - virtual environment
3. **Github** - version management
4. **Python** - scripting language
5. **Quickbase App** - ERD
6. **PostgreSQL** - database housing and management
7. **Vader** - Python package for Sentiment Analysis
8. **Tableau** - visualization
9. **HTML & CSS** - visualization
10. **VS Code Basic** - file creation, idea structure, and organization

	title	score	subreddit	url	num_comments	body	date
0	UPVOTE so everyone sees we got SUPPORT	265029	wallstreetbets	https://i.redd.it/sgoqy8nyt2e61.png	11825	NaN	2021-01-29 00:40:34
1	GME YOLO	23532	wallstreetbets	https://i.redd.it/opzucppb15e61.png	23532	NaN	2021-01-29 08:06:23
2	CLASS ROBINHOOD	18318	wallstreetbets	https://www.reddit.com/r/wallstreetbets/comment...	18318	LEAVE ROBINHOOD. They dont deserve to make mon...	2021-01-29 00:49:11
3	GME YOLO update — Jan 27 2021 ---	185949	wallstreetbets	https://i.redd.it/a309gkm5yxd61.png	15495	NaN	2021-01-28 08:15:35
4	Can we all take a moment and appreciate the Mo...	184517	wallstreetbets	https://www.reddit.com/r/wallstreetbets/comment...	7105	NaN	2021-01-28 11:57:32

1st Segment

During the first week we:

- ❑ Built the team,
- ❑ Decided our topic,
- ❑ Decided the technologies,
- ❑ Built mockups of the Dataset and ML

2nd Segment

During the second and third week we:

- ❑ Continue to build on our Dataset and ML

