

Sep 18, 2023

Attendees: Abdul Wasay Siddiqui, Tianliang Zhao, Manuel Amaya

#### Notes

- What we need?
  - fetch\_data\_from\_tech.py
  - Data\_preprocessing.py
  - to\_db.py
  - Meeting notes
  - A pdf version of the Readme

#### Action items

- ☒ ~~Initial Discussion~~
- ☐ Discuss the distribution of work
- ☐ Tiangling: fetch\_data\_from\_tech.py
- ☐ Abdul : Data\_preprocessing.py
- ☐ Manuel : meeting notes
- ☐ Manuel : code review
- ☐ All: final checks

Sep 19, 2023

Attendees: Abdul Wasay Siddiqui, Tianliang Zhao, Manuel Amaya

#### Notes

- The distribution of work is agreed upon based on previously done work as it is very similar to a previous homework assignment
- Tianling starts working with praw

#### Action items

- ☒ ~~Initial Discussion~~
- ☒ ~~Discuss the distribution of work~~
- ☐ Tiangling: fetch\_data\_from\_tech.py
- ☐ Abdul : Data\_preprocessing.py
- ☐ Manuel : to\_db.py
- ☐ Manuel : meeting notes
- ☐ Manuel : code review
- ☐ All: final checks

## Sep 20, 2023

Attendees: Abdul Wasay Siddiqui, Tianliang Zhao, Manuel Amaya

### Notes

- Initial concerns about pulling large data using PRAW
- Initial csv is produced from PRAW

### Action items

- ☒ ~~Initial Discussion~~
- ☒ ~~Discuss the distribution of work~~
- ☐ Tianling: fetch\_data\_from\_tech.py
- ☐ Abdul : Data\_preprocessing.py
- ☐ Manuel : to\_db.py
- ☐ Manuel : meeting notes
- ☐ Manuel : code review
- ☐ All: final checks

## Sep 21, 2023

Attendees: Abdul Wasay Siddiqui, Tianliang Zhao, Manuel Amaya

### Notes

- Tianliang finishes fetch\_data\_from\_tech.py and produces final version of csv
- Possible duplication issue

### Action items

- ☒ ~~Initial Discussion~~
- ☒ ~~Discuss the distribution of work~~
- ☒ ~~Tianling: fetch\_data\_from\_tech.py~~
- ☐ Abdul : Data\_preprocessing.py
- ☐ Manuel : to\_db.py
- ☐ Manuel : meeting notes
- ☐ Manuel : code review
- ☐ All: final checks

## Sep 22, 2023

Attendees: Abdul Wasay Siddiqui, Tianliang Zhao, Manuel Amaya

### Notes

- Abdul starts work on data\_preprocessing.py

- Discussion about Post text columns as only 800 values approx out of 5000 have values. Majority empty. It is concluded that it should be kept with NaN
- Discussion about PRAW limitation leading to data quality issues

#### Action items

- ☒ ~~Initial Discussion~~
- ☒ ~~Discuss the distribution of work~~
- ☒ ~~Tiangling: fetch\_data\_from\_tech.py~~
- ☒ ~~Abdul : Data\_preprocessing.py~~
- ☐ Manuel : to\_db.py
- ☐ Manuel : meeting notes
- ☐ Manuel : code review
- ☐ All: final checks

Sep 23, 2023

Attendees: Abdul Wasay Siddiqui, Tianliang Zhao, Manuel Amaya

#### Notes

- Abdul finishes data\_preprocessing.py
- Manuel finishes to\_db.py
- Manuel performs code review, notices abdul forgot to check latest code. Also notices readme isnt created
- Everyone works on readme
- All: final checks
- Abdul does submission

#### Action items

- ☒ ~~Initial Discussion~~
- ☒ ~~Discuss the distribution of work~~
- ☒ ~~Tiangling: fetch\_data\_from\_tech.py~~
- ☒ ~~Abdul : Data\_preprocessing.py~~
- ☒ ~~Manuel : to\_db.py~~
- ☒ ~~Manuel : meeting notes~~
- ☒ ~~Manuel : code review~~
- ☒ ~~All: final checks~~