**CS 5128/6028 Large Scale Software Engineering**

**Graduate Student Project**

Austin Conn ([connat@mail.uc.edu](mailto:connat@mail.uc.edu))

John Brown ([brow2jp@mail.uc.edu](mailto:brow2jp@mail.uc.edu))

**Overview:**

The graduate student project is composed of two parts: exploratory analysis of a chosen feature dataset, and comparative analysis of the dataset to the two datasets used throughout the semester, specifically the feature datasets from Zoom and WebEx for the year of 2022. Our group, consisting of Austin Conn and John Brown, chose the feature dataset from Discord. This dataset was taken directly from Discord’s change log, publicly available on Discord’s developer portal, <https://discord.com/developers/docs/change-log>.

**Exploratory Analysis:**

The feature dataset for Discord was created manually, by scraping the webpage. As of this report’s creation, the feature dataset is current and comprehensive, with entries dating back to the year 2017. There are 259 total feature entries. Many of these features were released together, creating 85 unique release days over the five plus years spanning from 2017 through 2023. In general, Discord did not report bug fixes or resolved issues within their change logs. Because of this fact, no feature entries were removed prior to exploratory analysis. This decision was made to hopefully increase the ability to compare our Discord dataset to the Zoom and WebEx datasets.

…. Talk about first graph …

(total feature entries)

**Figure 1: Discord Feature Count By Month (2017-2023)**

… talk about second graph …

**Figure 2: Discord Feature Count By Month (2017-2023)**

… talk about third graph …

(single day)

**Figure 3: Discord Release Day Count By Year (2017 -2023)**

… talk about fourth graph …

**Figure 3: Discord Release Day Count By Month (2017 -2023)**

… talk about fifth graph …

**Figure 5: Discord Feature Count Depreciating Old Features By Year (2017 -2023)**

… talk about sixth graph …

(Just 2022)

**Figure 6: Discord Feature Count By Month (2022)**

… Talk about figure 7 …

**Figure 7: Zoom Feature Count By Month (2022)**

… Talk about eighth picture …

**Figure 8: WebEx Feature Count By Month (2022)**

**Conclusions:**

… discuss conclusions …

\*talk about workflow of each company

\*\*consistent for Webex, zoom and discord have months with nothing

\*talk about release windows for discord/zoom

\*\* talk about quarters (of the year with like a business angle)

**Figure 9: Combined Feature Count By Month (2022)**

TODO:

* Formalize the spreadsheet
  + submit
* Write the report
  + Link to definition: https://en.wikipedia.org/wiki/Exploratory\_data\_analysis
  + Throw as much garbage as possible!
  + Submit
* Effort/peer evaluation
  + Create/submit

“Curate the features” -> organize

Exploratory analysis:

* Total # of features
* Total/average/mean/max min # of features per month
* # of tokens (voc size)
* # of sentences per feature
* Sentence length
* Feature length …

Draw meaningful comparisons with Zoom and WebEx:

* When are released happening (month)
  + Compare by year (2022 for zoom data)
  + “clean boundaries”
* Number of releases per year
* Sentence length