### Group Members

* Karl Eirich
* Kasha Muzila

### Proposed Work

* Identify privacy threats resulting from a public venmo profile
  + Membership Disclosure
    - Payment note content
    - Frequent or repeated transactions
      * Ex: rent, therapy payments, can build a schedule with this
    - Sender or Receiver in context of payment note
  + Identity Disclosure
    - The ability to find the specific person
      * Use location hints from payment note
      * The social network of the person
        + Like who’s associated with them
        + Ex: person is in frat but don’t know which school. But know that he has a friend named kevin so we can narrow down who that might be
  + Inference Threat
    - Identify specific attributes
    - Any of the above
    - Ex: person is buying drugs or alcohol (was done already in a previous project), child predator symbols
* Contextualize the impact of the privacy threats on individuals
  + Encourage use of privacy settings
  + Outline the data, showed what they can do with the data, and how to improve privacy based on that information.
  + Even the cleanest person can be trailed onto other platforms that they do use. You might be clean but friends might not be.
* Provide mitigation suggestions to protect users
  + SWOT (streangth, weakness, opportunity, threat) analyze the mitigations
    - What will it do to the usability, if too protected can’t use the app as well
    - Venmo doesn’t allow specific words (i.e., blow job, sex work, drugs, etc.), can bring more privacy here bc people will use slang or emojis to describe it so it doesn’t actually protect you. Can use blow job as a joke. Opportunity instead of banning words, can make these specific purchases private and bucket them into specific categories so they can have better generalization. Threat is being able to find these people and what they are purchasing.

### Related Work

* Dataset
  + [venmo-data](https://github.com/sa7mon/venmo-data)
* Previous W233 project
  + [Breaking Venmo's Privacy](https://cdn-uploads.piazza.com/paste/j3l2ilc13xs4mw/bdec9514d002207ade3034850cf98dce8059e3a27cee024ae6c007e20d6d91a7/PI_JL_DB_Project_Presentation.pdf)
  + [[Blog] Venmo's Information Disclosure](https://medium.com/berkeleyischool/venmos-information-disclosure-5d05174982ce)
* Mentioned by previous W233 project
  + [Public By Default](https://22-8miles.com/public-by-default/)

### Preliminary Results (Optional)

* None

### Member Contributions

* Karl will handle data ingestion and processing
* Kasha will develop privacy threat insights and visualizations

### Proposed Milestones

* Week 7
  + Write data ingestion script and begin processing code
* Week 8
  + Finish processing code and start developing insights
* Week 9
  + Start visualizing insights and crafting narrative
* Week 10
  + Identify and analyze mitigation suggestions
* Week 11
  + Finish developing insights and start writing paper
* Week 12
  + Start creating slide deck
* Week 13
  + Finish writing paper and creating slide deck
* Week 14
  + Finalize and present project

### [EXTRA] Other Info

* Proposal Doc
  + <https://github.com/zachChilders/deanonymization>
* Presentations
  + [ANALYSIS OF PRIVACY RISKS FOR TIKTOK USERS](https://cdn-uploads.piazza.com/paste/j3l2ilc13xs4mw/3dbbe3772997be4febcfae9c68300b9ace160c819cfec6a8fabb8b6bef3ac231/TB_NR_PN_GN_Project_Presentation.pdf)
  + [A PRIVACY-FORWARD APPROACH TO PII IN THE PUBLIC DOMAIN](https://cdn-uploads.piazza.com/paste/j3l2ilc13xs4mw/c4df85e9a0c9f486ddccbef2b0ee44f2c6145a42ad25824b86093c013ec551cb/AF_EK_CH_Project_Presentation.pdf)
  + [Resume Obfuscation](https://cdn-uploads.piazza.com/paste/j3l2ilc13xs4mw/a8f90d73a56a9f1314ef9bd596b196e23a094b5ae85c580c41e3f2f1151e573f/JS_MR_SD_VDC_KK_Project_Presentation.pdf)
  + [Privacy in Digital Identification](https://cdn-uploads.piazza.com/paste/j3l2ilc13xs4mw/bd61c37c19f84dadebd81561e0014259a3069ffce799a042385c6cabac0214cc/AA_GL_SB_Project_Presentation.pdf)
* Datasets
  + TikTok
    - [TikTok Video Comments - David Dobrik's top videos | Kaggle](https://www.kaggle.com/datasets/jhosn13/tiktok-video-comments-david-dobriks-top-videos)
    - <https://github.com/krkhalil/Tiktok>
    - [OSF | TikTok Metadata from 2020](https://osf.io/hfuaw/)
  + Instagram
    - [Instagram Images with Captions | Kaggle](https://www.kaggle.com/datasets/prithvijaunjale/instagram-images-with-captions)
  + Youtube
    - [Just Dance @ YouTube: Multi-label Text + Analytics | Kaggle](https://www.kaggle.com/datasets/renatojmsantos/just-dance-on-youtube)
    - <https://github.com/MayurDeshmukh10/youtube_analysis>
  + Twitter
    - <https://github.com/jorancornelisse/Twitter-Attribute-Inference>
  + OnlyFans
    - <https://github.com/fpjnijweide/passion_economy>
  + Amber Heard/Johnny Depp
    - [Instagram\_Data Amber Heard Social Network Analysis | Kaggle](https://www.kaggle.com/datasets/amberhearddata/data-instagram-amber-heard-social-network-analysis)
    - [TikTok\_Data Amber Heard - Social Media 2022](https://www.kaggle.com/datasets/amberhearddata/tiktok-data-amber-heard-social-media-2022)
    - [Johnny Depp Trial 2022 - Live Chat](https://www.kaggle.com/datasets/maxmnemo1010/johnny-depp-trial-livechat)