**CYBER/DATASCI 233 Project Proposal** Karl Eirich, Kasha Muzila

TOTAL POINTS

**85 / 100**

QUESTION 1

**1** Proposal Body **85 / 100**

Group members (10 points)

**✓ + 10 pts Members of the group listed.**

**+ 0 pts** Members of the group are not listed.

Work description (20 points)

**+ 20 pts** Adequate description of proposed work. **✓ + 15 pts Description of proposed work can be improved.**

Related work (20 points)

**+ 20 pts** Adequate survey of related work. **✓ + 15 pts Related work section listed but can be improved.**

**+ 10 pts** Related work section listed but is incomplete.

**+ 5 pts** Related work section listed but is significantly incomplete.

Preliminary results/Feasibility/Proof of concept (30 points)

**+ 30 pts** Adequate description of preliminary results/justification for feasibility.

**✓ + 25 pts Preliminary results/justification for feasibility section listed but can be improved.**    **+ 20 pts** Incomplete preliminary results/feasibility description.

**+ 0 pts** Section/description is missing.

Milestones (10 points)

**✓ + 10 pts Proposed milestones listed.**

Description of potential contributions of team members (10 points)

**✓ + 10 pts Adequate description of potential contributions of each group member.**

**+ 10 pts** Solo project

**+ 0 pts** No description of potential contributions of team members.

Page limit

**✓ + 0 pts Within page limit.**

**- 2 pts** Minor overflow (up to 0.5 extra pages)    **- 5 pts** Moderate overflow (up to 1 extra pages)    **- 10 pts** Major overflow (up to 2 extra pages)    **- 15 pts** Beyond major overflow (more than 2 extra pages)

**- 5 pts** Late submission

      I am excited about this topic. Very creative. **1**

      Good identification and clarification of the risks **2** you're interested in, but the scope is a little large. Maybe focus on a few?

      I would have liked to seen more on the analysis **3** you completed with the Venmo data. Which variables did you consider or are available? What was the dataset size (all 7 million transactions)? Which quasi identifiers did will you consider? How are you measuring your sensitive attribute (you describe a few, generally, as identity disclosure)? What time range will you consider?

      I like this presentation, it has a lot of information **4** you can use for your presentation. But, it was hard to tell how your proposal scope was different from this presentation's final product. For example, are you going to clean the data the same way? The presentation authors cited 125 hours to search user names of 45,356 users + social media mapping. Are you doing the same thing? Can you clarify which sensitive attribute you'll be looking at (i.e. both membership disclosure, identity disclosure and/or inference threats)?

**5**

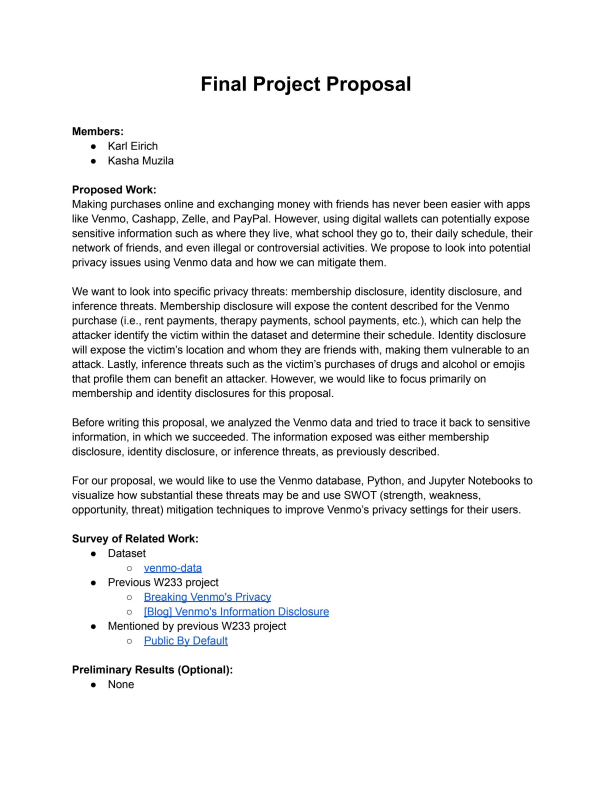
      Are there any published papers on privacy using this data? A quick Google Scholar search of "venmo privacy data" has some cool papers. Check out Tandon et al. (2022) "I know what you did on Venmo." I didn't review this piece, but Yao et al. (2018) "Beware of What you Share: Inferring User Locations in Venmo" sounds cool. Again, I'd like to have seen more desciprtions on how your proposal is different or simliar to related work? That would have helped me understand the scope of this proposal. It still feels a bit large and ambitous.

      These milestones are good and illustrates an **6** understanding of the phases of research that will need to be accomplished. However, I would have liked to seen more description in your scope of work for what exactly you intend to do. For example, what kind of insights do you suspect to find or

hypothesize? I am sort of assuming this is some devaiation or development from your related work citations.

      Draw out more of this thought: "mitigation **7** techniques to improve Venmo's privacy settings" how will your results provide this recommendation to Venmo?

Page 2

**1** 

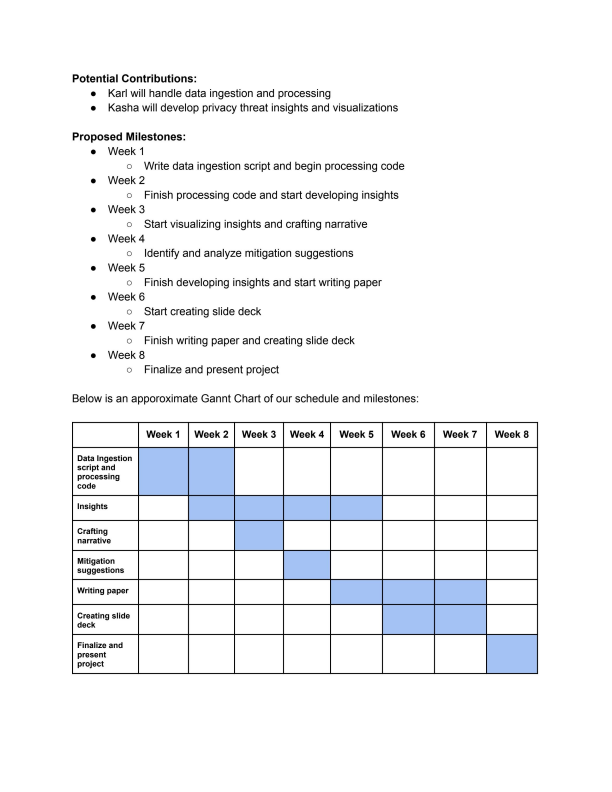
**2**

**3**

**7**

**4**

**5**

**6**

**1** Proposal Body **85 / 100**

Group members (10 points)

**✓ + 10 pts Members of the group listed.**

**+ 0 pts** Members of the group are not listed.

Work description (20 points)

**+ 20 pts** Adequate description of proposed work.

**✓ + 15 pts Description of proposed work can be improved.**

Related work (20 points)

**+ 20 pts** Adequate survey of related work.

**✓ + 15 pts Related work section listed but can be improved.**

**+ 10 pts** Related work section listed but is incomplete.

**+ 5 pts** Related work section listed but is significantly incomplete.

Preliminary results/Feasibility/Proof of concept (30 points)

**+ 30 pts** Adequate description of preliminary results/justification for feasibility.

**✓ + 25 pts Preliminary results/justification for feasibility section listed but can be improved.**    **+ 20 pts** Incomplete preliminary results/feasibility description.

**+ 0 pts** Section/description is missing.

Milestones (10 points)

**✓ + 10 pts Proposed milestones listed.**

Description of potential contributions of team members (10 points)

**✓ + 10 pts Adequate description of potential contributions of each group member.**

**+ 10 pts** Solo project

**+ 0 pts** No description of potential contributions of team members.

Page limit

**✓ + 0 pts Within page limit.**

**- 2 pts** Minor overflow (up to 0.5 extra pages)

**- 5 pts** Moderate overflow (up to 1 extra pages)

**- 10 pts** Major overflow (up to 2 extra pages)

**- 15 pts** Beyond major overflow (more than 2 extra pages)

**- 5 pts** Late submission

      I am excited about this topic. Very creative. **1**

      Good identification and clarification of the risks you're interested in, but the scope is a little large. Maybe **2** focus on a few?

      I would have liked to seen more on the analysis you completed with the Venmo data. Which variables did you **3** consider or are available? What was the dataset size (all 7 million transactions)? Which quasi-identifiers did will you consider? How are you measuring your sensitive attribute (you describe a few, generally, as identity disclosure)? What time range will you consider?

**4**

      I like this presentation, it has a lot of information you can use for your presentation. But, it was hard to tell how your proposal scope was different from this presentation's final product. For example, are you going to clean the data the same way? The presentation authors cited 125 hours to search user names of 45,356 users + social media mapping. Are you doing the same thing? Can you clarify which sensitive attribute you'll be looking at (i.e. both membership disclosure, identity disclosure and/or inference threats)?

      Are there any published papers on privacy using this data? A quick Google Scholar search of "venmo privacy **5** data" has some cool papers. Check out Tandon et al. (2022) "I know what you did on Venmo." I didn't review this piece, but Yao et al. (2018) "Beware of What you Share: Inferring User Locations in Venmo" sounds cool. Again, I'd like to have seen more desciprtions on how your proposal is different or simliar to related work? That would have helped me understand the scope of this proposal. It still feels a bit large and ambitous.

      These milestones are good and illustrates an understanding of the phases of research that will need to be **6** accomplished. However, I would have liked to seen more description in your scope of work for what exactly you intend to do. For example, what kind of insights do you suspect to find or hypothesize? I am sort of assuming this is some devaiation or development from your related work citations.

      Draw out more of this thought: "mitigation techniques to improve Venmo's privacy settings" how will your **7** results provide this recommendation to Venmo?

Page 6