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

1 I am excited about this topic. Very creative.

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

3 I would have liked to seen more on the analysis 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?

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)?

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.

6 These milestones are good and illustrates an 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.

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

Final Project Proposal

Members:

- Karl Eirich
- Kasha Muzila

Proposed Work:

Making purchases online and exchanging money with friends has never been easier with apps like Venmo, Cashapp, Zelle, and PayPal. However, using digital wallets can potentially expose sensitive information such as where they live, what school they go to, their daily schedule, their network of friends, and even illegal or controversial activities. We propose to look into potential privacy issues using Venmo data and how we can mitigate them.

We want to look into specific privacy threats: membership disclosure, identity disclosure, and inference threats. Membership disclosure will expose the content described for the Venmo purchase (i.e., rent payments, therapy payments, school payments, etc.), which can help the attacker identify the victim within the dataset and determine their schedule. Identity disclosure will expose the victim's location and whom they are friends with, making them vulnerable to an attack. Lastly, inference threats such as the victim's purchases of drugs and alcohol or emojis that profile them can benefit an attacker. However, we would like to focus primarily on membership and identity disclosures for this proposal.

Before writing this proposal, we analyzed the Venmo data and tried to trace it back to sensitive information, in which we succeeded. The information exposed was either membership disclosure, identity disclosure, or inference threats, as previously described.

For our proposal, we would like to use the Venmo database, Python, and Jupyter Notebooks to visualize how substantial these threats may be and use SWOT (strength, weakness, opportunity, threat) mitigation techniques to improve Venmo's privacy settings for their users.

Survey of Related Work:

- Dataset
  - venmo-data
- Previous W233 project
  - Breaking Venmo's Privacy
  - [Blog] Venmo's Information Disclosure
- Mentioned by previous W233 project
  - Public By Default

Preliminary Results (Optional):

- None

Potential Contributions:

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

Proposed Milestones:

- Week 1
  - Write data ingestion script and begin processing code
- Week 2
  - Finish processing code and start developing insights
- Week 3
  - Start visualizing insights and crafting narrative
- Week 4
  - Identify and analyze mitigation suggestions
- Week 5
  - Finish developing insights and start writing paper
- Week 6
  - Start creating slide deck
- Week 7
  - Finish writing paper and creating slide deck
- Week 8
  - Finalize and present project

Below is an apporoximate Gannt Chart of our schedule and milestones:

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Data Ingestion script and processing code	🟦	🟦						
Insights		🟦	🟦	🟦	🟦			
Crafting narrative			🟦					
Mitigation suggestions				🟦				
Writing paper					🟦	🟦	🟦	
Creating slide deck						🟦	🟦	
Finalize and present project								🟦

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

1 I am excited about this topic. Very creative.

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

3 I would have liked to seen more on the analysis 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?

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)?

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.

6 These milestones are good and illustrates an 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.

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