Exercise 2

RQ: How does social media influence perceptions of safety and risk in skiing?

Exercise 2a

How will you create your dataset? Which general biases and issues does your dataset have? How can you mitigate these?

- **Data Collection:** Use posts from Instagram and TikTok where users share ski videos or images and posts which are related to ski accidents and safety.
- **General Biases:**
  - **Population Bias:** Over representation of younger or tech savvy users, therefore under representation of older or less active social media users
  - **Content Bias:** More extreme or sensational events are likely to be shared, leading to an increased emphasis on risky behavior
  - **Redundancy Bias:** Many posts on these platforms are reposts from the creators or other sites, which means especially the extreme posts, which are shared often could appear multiple times
- **Mitigation of Biases:**
  - Collect also from alternative sources to balance the different age groups, for example with surveys
  - Use multiple platforms to collect the data to not just let two different algorithms decide which content is included
  - Using a tool to compare if duplicate videos or photos are in the dataset

Exercise 2b

Which issues at the data source/origin level does the dataset have? How can you mitigate these?

- **Functional Bias:** Platform-specific features, like Instagram's mainly image focus or TikTok's video-driven content, may distort how people report skiing risks. There is no clear text option to inform people about for example risks like on X.
- **Normative Bias:** Unwritten norms on platforms for example showing mainly glamorous or exciting moments could skew the data toward highlighting fun and this would downplay the risks
- **Mitigation of Biases:**
  - Include diverse platforms like forums or blogs where more detailed discussions are taking place
  - Use a filter to not just find the most famous posts in this area

Exercise 2c

Does your data collection and processing introduce additional biases? If so, which ones and how can you counteract these?

- **Selection Bias:** Users who engage in more extreme or risky skiing behaviors may be more likely to post about their experiences, leading to an over representation of riskier content.
- **Mitigation of Biases:**
  - Encourage posts from a wider range of skiing experiences, including more moderate or casual skiing activities, to balance the dataset.

Exercise 2d

Which type of analyses would you choose for your study? Can a combination of different types mitigate the risk of a biased insight?

- **Descriptive Statistics:** The plan is to use descriptive statistics to analyze the dataset, focusing on identifying key patterns such as the frequency of safety-related posts and the different kinds of incidents. This approach will help to reveal trends and behaviors that are crucial for understanding how individuals interpret safety and risk in skiing.
- **Qualitative Analysis:** The next step involves qualitative analysis, where these patterns will be investigated in more detail. This will allow to explore deeper insights into how and why people interact with safety/risk and skiing.
- **Combining Methods:** By combining both methods, the aim is to enhance the deep dive into the dataset and find concrete reasons for the actions which are taking by the users.

Exercise 2e

How do you plan to evaluate and interpret the findings? Which issues may occur and how can you mitigate these?

- **Evaluation**: Combining descriptive statistics with qualitative analysis to identify trends and deeper insights about safety and risk in skiing.
- **Issues**: Potential bias in interpreting trends, especially to the overrepresentation of extreme content or specific locations of the participants (e.g. just analyzing the greater Innsbruck area might bias the results).
- **Mitigation**: Use multiple analysis methods (Descriptive Statistics and Qualitative Analysis) and ensure differentiation in content sources to balance interpretations and reduce bias.