1. **A link to the data set**

My data set is from X using the Developer portal X Developers and the following python library Tweepy . The Api allows me to retrieve tweets that match the specified criteria programmatically. The specific version used is API v2 found here : Introduction - X

2. **A detailed description of the data set**

When querying and receiving data from the X API you get the following
Tweet Content: The text of the tweet.
Tweet Metadata:

- **Creation Timestamp**: When the tweet was posted.
- **Tweet ID**: A unique identifier for the tweet.
- **Public Engagement Metrics**: Counts of retweets, replies, likes, and quotes.

User Information:

- **User ID**: A unique identifier for the user.
- **Username**: The handle of the user (e.g., @username).
- **Display Name**: The name the user has chosen to display.
- **Profile Information**: Such as bio, location (if provided), and profile image URL.

3. **A note on how the data de-identifies or otherwise protects the users' when making the data available for public consumption.**

X offers access to public user data via its API without anonymizing it. This indicates that when developers access data through the API, they obtain it in its original format, which includes information like usernames, user IDs, and tweet content. The obligation to anonymize or de-identify this data falls upon the developers and researchers who utilize it.

Developers should apply anonymization methods before sharing or analyzing the data to protect user privacy and adhere to data protection laws. This involves eliminating or disguising personally identifiable information (PII) and consolidating data to avoid recognizing individual users. It's crucial to mention that although X's API doesn't anonymize data automatically, developers must comply with X's Developer Agreement and Policy, which details the rules regarding data use and distribution. This entails honouring user privacy and guaranteeing that any shared information is suitably anonymized to avert misuse or re-identification.

In conclusion, X fails to anonymize user data in its API results; developers must apply appropriate anonymization methods to safeguard user privacy when managing and distributing this data.

4. A note on how the data could be exploited otherwise

If not handled responsibly, the data obtained from X's API can be misused in several ways:

Re-identification: Some parameters can be used to reverse engineer the tweet's source, as explained in part 2; there is the tweet ID and user ID; one way to prevent that is just returning the tweet without the other parameters specified in the metadata.

Harassment or Doxing: Revealing identifiable information in public can subject users to harassment or unsolicited attention.

Violation of Privacy:Disclosing confidential information without permission can violate users' privacy rights.