2.0 Objective

(1) Evaluate the reliability of MBTI from a statistical perspective.

(2) Explore the potential application of MBTI in social media behavior analysis.

(3) Help people better understand personality traits and behavioral patterns.

(4) Helping people eliminate stereotypes caused by MBTI personality types.

(4) Quantifying the distribution of MBTI types among different populations, and exploring the changes in test results over time and situations.

(5) By analyzing the posts on social media, identify the key words and expression characteristics of different personality types.（这两点其实没有）

3.0 Problem Statements

1. Are the results of the MBTI personality test statistically robust and reliable?

2. Do the four dimensions of MBTI work independently, or are they connected in some way?

3. Do people with different MBTI types have different levels of activity on the internet?

4. Do significant differences exist in the interest preferences and behavioral patterns of different MBTI personality types on social networking sites?

4.0 Data Collection

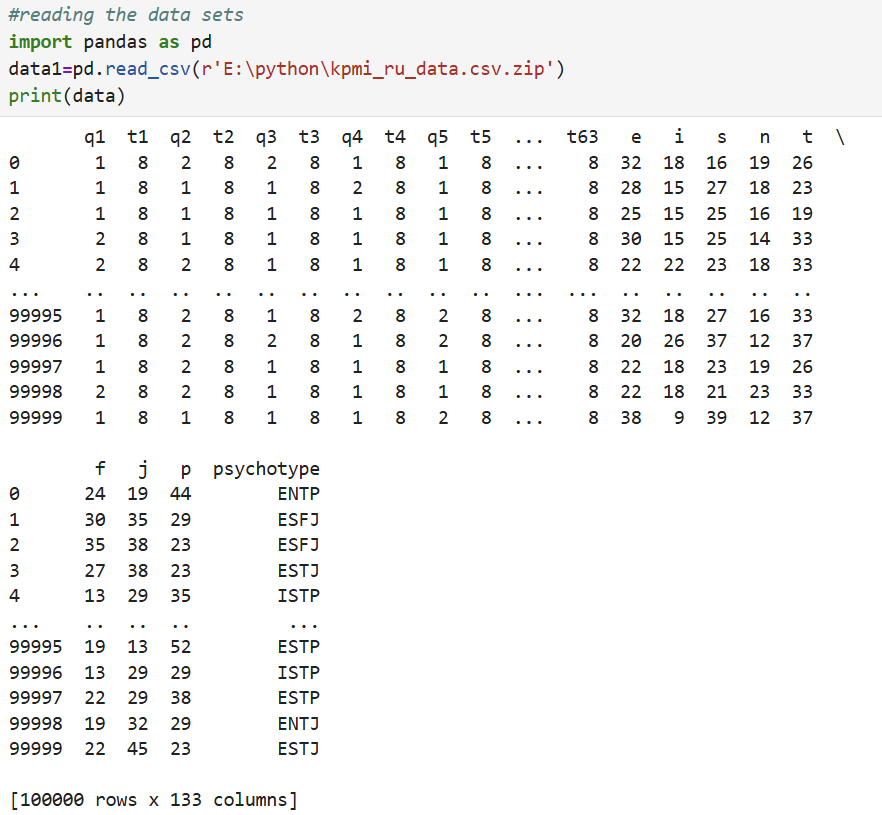
The datasets used in this study are downloaded from two Kaggle repositories:

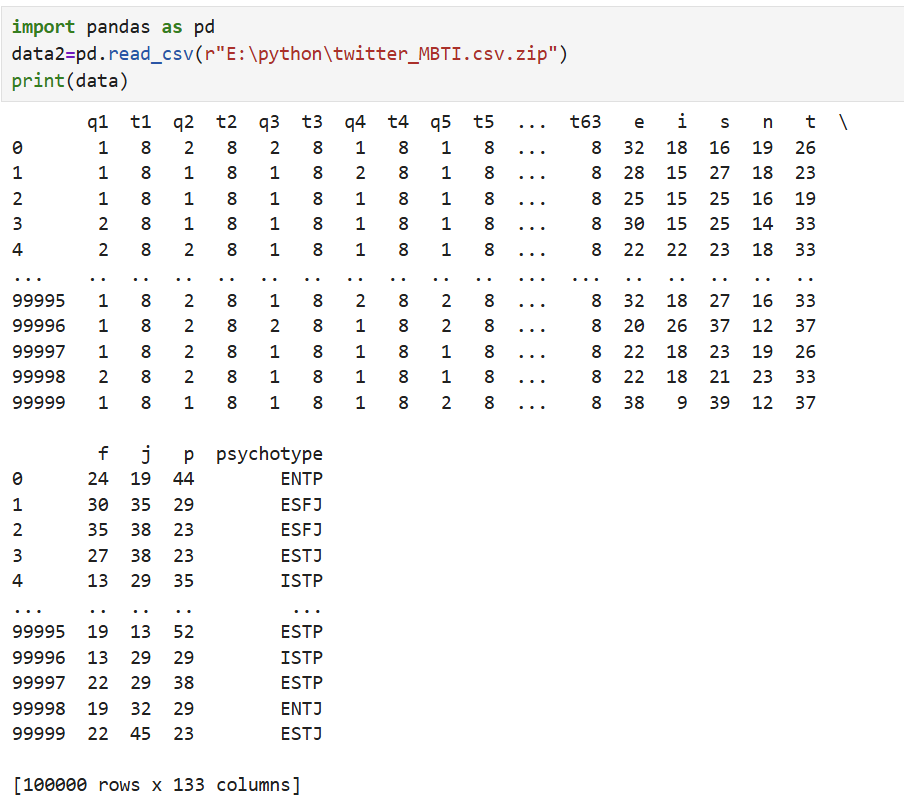
1. “MBTI Personality Type Twitter Dataset” – tweets originally gathered through the Twitter API and subsequently curated on Kaggle by Mazlumi <https://www.kaggle.com/datasets/mazlumi/mbti-personality-type-twitter-dataset>
2. “KPMIRU Questionnaires Data” – questionnaire responses compiled and shared on Kaggle by Pmenshih

<https://www.kaggle.com/datasets/pmenshih/kpmiru-questionnaires-data>

（这里可以详细说一下数据集，比如第一个是他人调用Twitter API从Twitter上爬取的，同时包含了用户自己声称的mbti人格类型，但是没有经过清洗；第二个可以说包含每一题的答案和测试结果，测试结果精确到分数而不仅仅是类型）

Together, these sources constitute the whole set of data analyzed in our work.





这里代码错了导致打印的数据集也有问题

As illustrated in the two Figures above, the combined Kaggle sources provide information on:

1. Self-reported MBTI types for each respondent

2. Raw Twitter posts and basic tweet metadata linked to those MBTI labels

3. Demographic and psychometric questionnaire answers (KPMIRU survey)

4. Behavioral metrics such as posting frequency and topic keywords extracted from the tweets