**1. Introduction**

With the recent renewal of the MBTI's popularity, posts or analyses about it have become widespread on public platforms. It seems that MBTI has become a way people use at first meetings.

The Myers-Briggs Type Indicator (MBTI) is a psychological framework primarily used to describe and measure personality traits by categorizing individuals into four dichotomous dimensions. In essence, it aims to reflect how people perceive the world and make decisions based on their internal experiences (Chen & Chen, 2015, as cited in Yang, 2022).

And the test is conducted through asking a series of questions, and the type of question is kind of like asking about the tendency to do something. Based on the responses, individuals receive a result represented by four letters, each corresponding to one of the four bipolar dimensions. According to Pittenger (2005), these four dichotomous dimensions classify individuals as either extraverted (E) or introverted (I), sensing (S) or intuitive (N), thinking (T) or feeling (F), and judging (J) or perceiving (P). Combinations of the four preferences generate one of 16 personality types (e.g., ESFJ, ENFP, INTP, ISFJ), each associated with distinct behavioral tendencies, reflecting differences in attitudes, orientation, and decision-making styles. Also, on the official website, it will provide the future career options for you, like INFP, which may be more suitable for being an author. The percentage of each standard will also show in the final result.

The MBTI provides a widely recognized and accessible way to understand personality, making it a useful foundation for further behavioral and data-driven analysis.

**2. Objective**

(1) Evaluate the reliability of the 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.

**3. 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. Data Collection**

1. “MBTI Personality Type Twitter Dataset”

Tweets were originally harvested from the public Twitter API by a third-party collector and later released on Kaggle by Mazlumi (<https://www.kaggle.com/datasets/mazlumi/mbti-personality-type-twitter-dataset>).

• 8,600 Twitter users, ∼1 million raw tweets.

• Each user record contains the self-declared MBTI type (e.g., “ENFP”) taken from the user’s bio.

• The text has not been further cleaned or filtered—URLs, emojis, hashtags, and retweets remain. Researchers must therefore perform their preprocessing (tokenisation, stop-word removal, emoji handling, etc.) before analysis.

2. “KPMIRU Questionnaires Data”

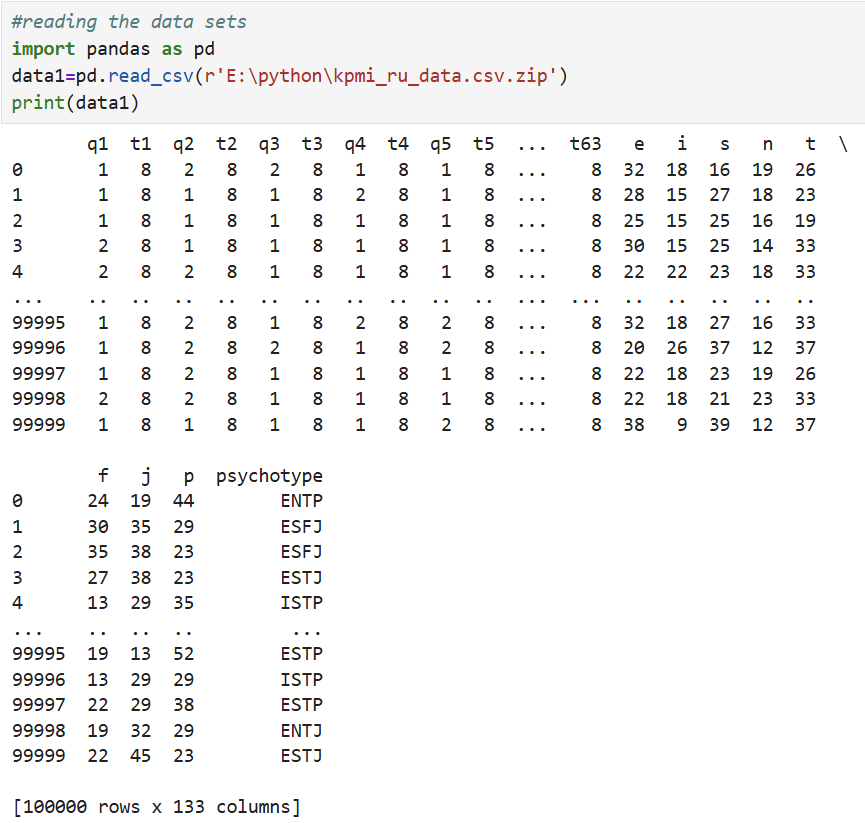
Questionnaire responses compiled and shared on Kaggle by Pmenshih (<https://www.kaggle.com/datasets/pmenshih/kpmiru-questionnaires-data>).

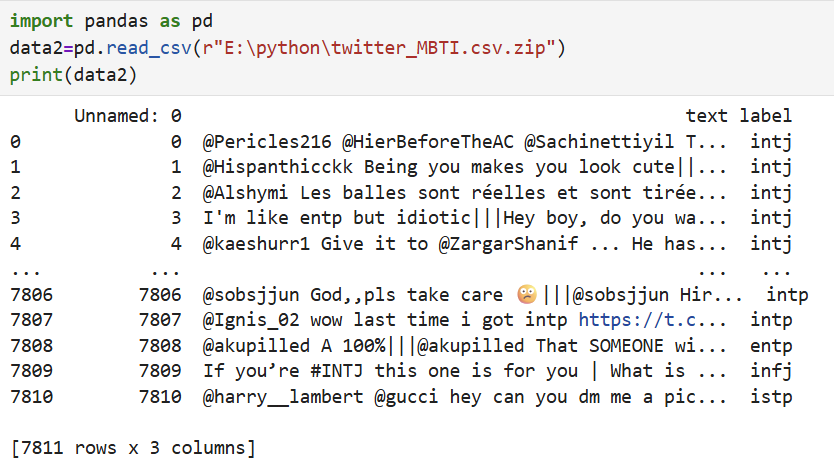
• Contains every participant’s item-level answers to the full KPMIRU personality inventory (several dozen Likert-scale questions).

• Provides the scored results for all four MBTI dimensions—reported as continuous scores (0–100 per axis) as well as the final type label (e.g., “ISTJ”).

• Demographic fields (age range, gender, education) are included, enabling richer statistical controls.

Together, the Twitter dataset supplies large-scale, real-world language samples with self-reported types, while the KPMIRU dataset offers clean, psychometrically scored questionnaire data. The two sources complement each other for training and validating our emotion-aware MBTI models.





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

**References**

Yang, Y. (2022, April). *Research on the Application of MBTI in Organizations*. In *2022 7th International Conference on Social Sciences and Economic Development (ICSSED 2022)* (pp. 1751–1754). Atlantis Press. <https://doi.org/10.2991/aebmr.k.220405.292>

Pittenger, D. J. (2005). Myers-Briggs Type Indicator (MBTI): Some psychometric limitations. *Australian Psychologist, 40*(1), 62–72. <https://doi.org/10.1080/00050060512331317265>