DSC96 Final Project Proposal

Prepared by: Scramble Eggs( Jun Linwu & Man-Fang-Liang)

# 

# Data:

This project would focus on the people’s behaviors, preferences, and others' perceptions according to their MBTI types. We would utilize the MBTI-Types & Enneagram Texts dataset from Kaggle (<https://www.kaggle.com/yamaerenay/mbtitypes>).

# Research Question:

* What are the most typical behaviors associated with each MBTI type?
* Which MBTI type receives the most positive perceptions?

# Brief Plan(tool, test, visualization):

* Data Cleaning: Before the analysis, we plan to remove duplicate text in the *texts* column using Python/ Excel/ Tableau.
* An open-coding analysis would be performed using Excel. Each author would work on 20% of the whole dataset individually to assign a sentiment score (-1, 0, 1; positive, neutral, negative accordingly) and a most relevant behavior or preference for each text. Those codes would be adjusted until Cohen's kappa value reaches over 0.80. Once the combined codebook is implemented, we would make a second dataset containing the overall percentage of the relevant behaviors and perceptions(positive, neutral, and negative) for each MBTI type.
* An additional sentiment analysis would be performed using NLP with Python This would serve as a sanity check that whether the result of the open-coding analysis is consistent with the result of NLP.
* We plan to visualize the data through Tableau.