# Workflow

1. Wrangle, prepare, cleanse data
   - overview
     - shape
   - Deal with Null value
     - UNKNOWN/blank → NaN
   - Extreme value
     - age (how to deal with >100 & <5)

2. Define, Analyse & Explore
   - numerical
     - count, 50% quartile, avg., max, min
   - categorical
     - unique, top, freq.
   - Extract gender from facebook
     - extract gender from info
     - check match rate
     - fill facebook_gender to gender

   - add income level effect
     - get income table based on zip_code from IRS
     - data cleaning (group zip, avg_income)

- — append to df
- — analyze relationship bw income/zipcode and engagement (6).
  - \* user_count
  - \* num_video_viewed
    active_time
  - \* last view – register
- extract data from view_hour.
  - — sum_view_times (add col)
  - — % view times / ttl view times (6)
  - — freq_time (most freq view hour category)
  - — relationship bw freq_time & age, gender ...
  - — heatmap (push notification time).
- Define Y (stickness)  — biweek1,2,3,4 after 4
  - — find retention rate (overall)
  - — pick up sticky users
  - — define Y : 1, 0
- Location based on lat & long.

- Analyze age-group, gender, country, region, active time, stickness
  - — analyze one-by-one
  - — relationship with stickness
  - — heatmap.