# EDA for Your Everyday

Rashmi Panse, Nick Miller, Kabir Nawani

Data Exploration

- Understanding our data
    - Age (Numeric)
    - Do you usually eat breakfast everyday (Categorical Binary)
    - What time do you usually eat breakfast (Numeric)
    - What do you usually have for breakfast (Text feature)
    - How often do you repeat the same breakfast (Numeric, # of days in a week)
    - Do your breakfasts differ on weekdays vs weekends
- Summary statistics/information
    - Response size
    - Average age: 25
    - Categories of "do you usually eat breakfast everyday": Yes, No
    - Range of time people eat breakfast: 6am - 12pm
    - How often do you repeat the same breakfast: max = 7 days/week

Data Cleaning

- Need reliable data to extract information

- Missing values:
    - Kabir never eats breakfast
    - Might need to impute with a common breakfast item so we don't lose data about that student

- Duplicate data
    - Multiple entries from the same student

- Data type conversion
    - Date could be recorded as text

How many John Smiths really?
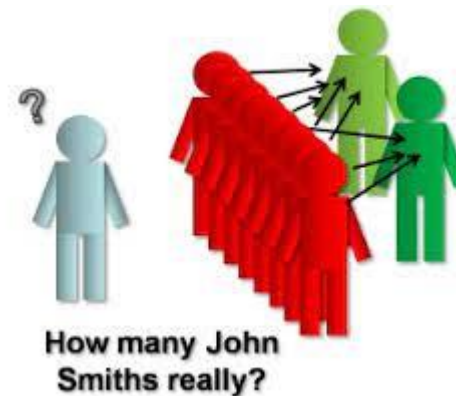
Data Cleaning

- Outliers
  - Rashmi eats avocado toast **EVERY** day
  - How extraordinary are the outliers?

- Cleaning free text fields (What do you eat for breakfast?)
  - Standardize words (capitalization, abbreviation)
  - Remove stop words
  - Tokenization
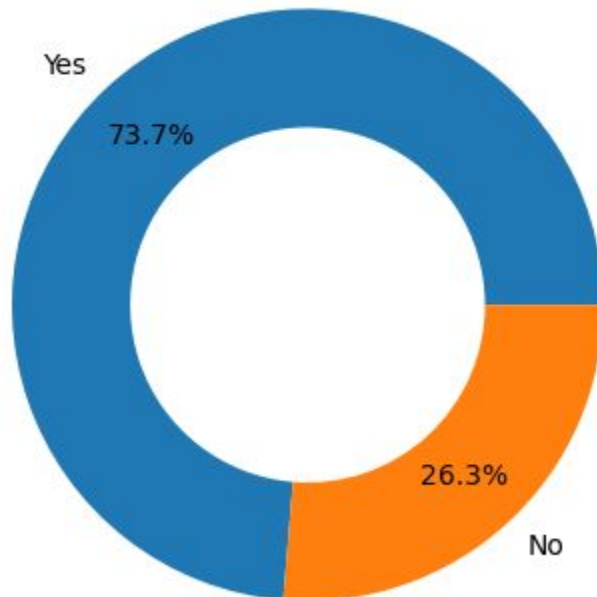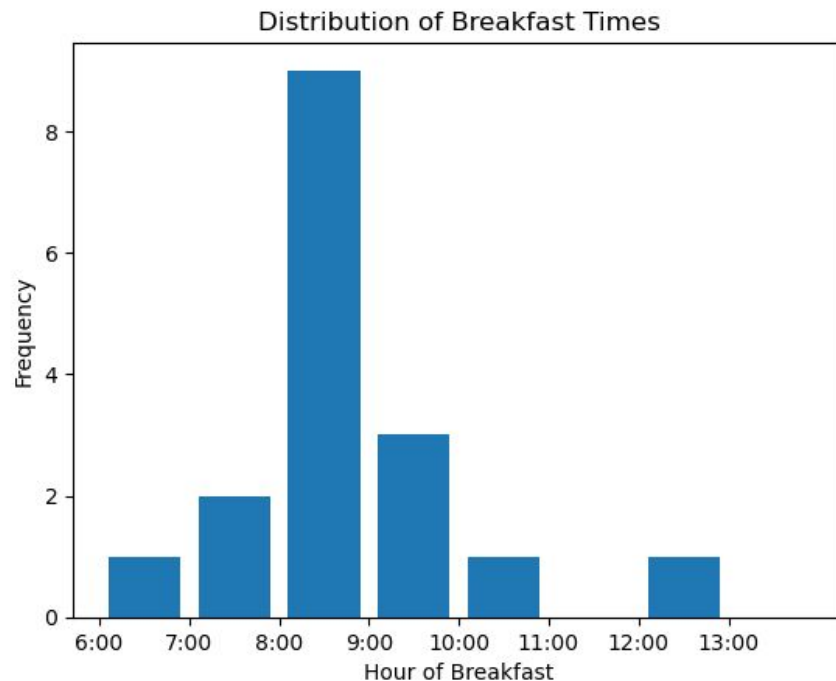   - "Eggs, Toast, Hot Sauce, Coffee"

"Avocado Toast"

"avocado toast"   "av0cdo Tost"

Do you usually eat breakfast everyday?
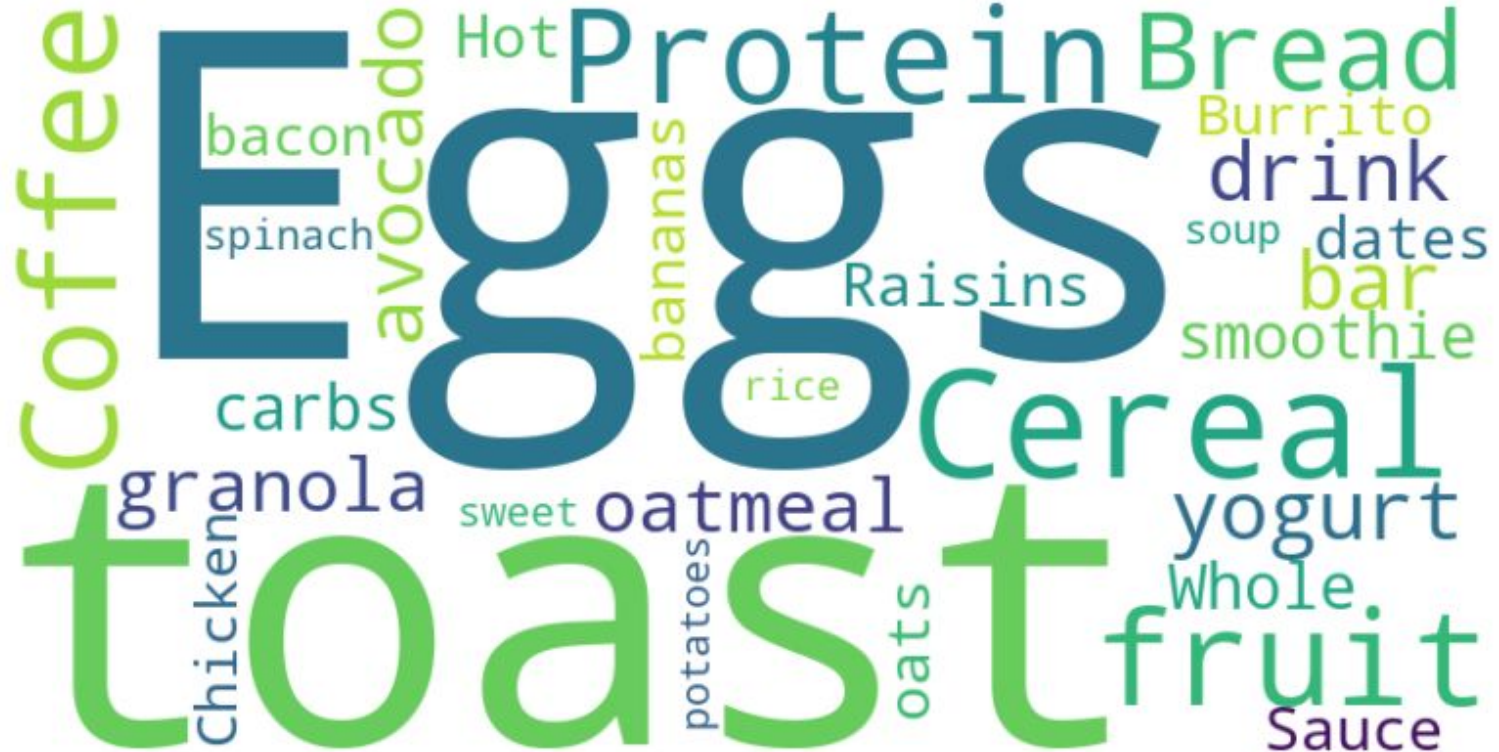
Breakfast Eating Habits

What time do you eat your breakfast?

Distribution of Breakfast Times

What do you usually have for breakfast?

Word Cloud of Breakfast Choices

Drawbacks to EDA

- Needs the right amount of data to be useful

- Can be an endless process if goal is not defined

- Subjectivity

Conclusions

- This information has so much potential!

- Why do people eat these things? Are they the quickest? Are they the cheapest?

- Is there a relationship between what someone has for breakfast and GPA?