See Jane | Research Project Data Normalization

Final Project Presentation

1. Team & Project Introduction

- **Team Structure**
  - Sponsors: Meredith Conroy, Romeo Perez
  - Project Manager: Grace Yoon
  - Team Members: Dayong Wu (Team Lead), Emmanouil Kritharakis, Yan Tong, Junfei Huang, Yuanli Wang
- **Project Motivation**
  - Create gender balance, foster inclusion and reduce negative stereotyping in family entertainment media.
- **Project Goal**
  - Normalize the datasets & refine the codebooks
  - Analyze industry differences in representation in advertisements
  - Summarize the change in representation in advertisements
  - Detect the trends of representations in advertisements
- **Backgrounds Needed to Understand the Project**
  - Prior knowledge to the entertainment industry
  - Sense of social responsibility of gender balance, race equality, diversity and inclusion, etc
  - Comprehensive understanding of the codebooks
  - Python Pandas, Numpy, Matplotlib (etc), and Excel skills are prerequisites

2. Merging Datasets

- **What data have you used? If you collected data, how did you do that?**
  - Mars 2020-2021 & Cannes datasets.
  - Include answers over questionnaires on gender representation in advertisements
  - Data were provided by the Geena Davis Institute.

- **What challenges did you encounter working with data?**
  - People's responses do not match with the multiple choice answers provided by questionnaires.
  - Questionnaires slightly alter over the years.
  - Take into consideration multiple column answers to create a new one.
  - Abiding communication with clients to conclude over which final columns need to be merged correctly.
  - Fixing final dataset codebook along with the clients.

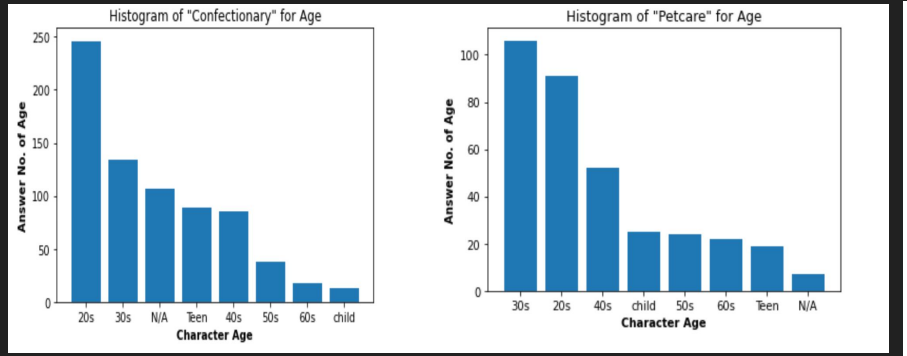
3. Analysis & Methods

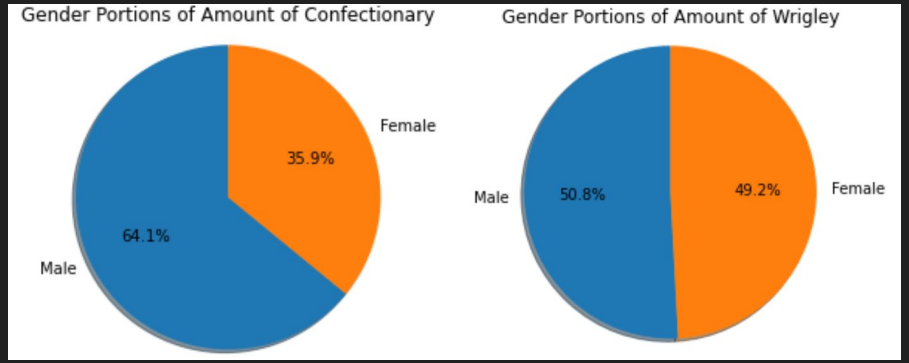
- Main analysis
  - Are there industry differences in representation in advertisements?
  - Is there a change in representation in advertisements over time?
  - What are the trends?
- Analysis Method
  - Select specific features
  - Visualize selected features
  - Do analysis based on visualization figures

4 Are there industry differences?

- **What results or observations do you have?**
  - The largest number of people work in “Confectionary” industry regardless of gender.
  - We find that the average age of the “Food” industry is the youngest.
  - And for “Food” industry, compared with other races, the number of "White" is the largest.
  - We observe that among the Race column, for value "White", the number of "Confectionary" industry and "Petcare " industry which are far more than "Wrigley" and "Food" industry. Therefore, the majority of "White" work in these two industries.
  - By the statistics, we find that the largest amount of data whatever their race is in the "Confectionary" industry.

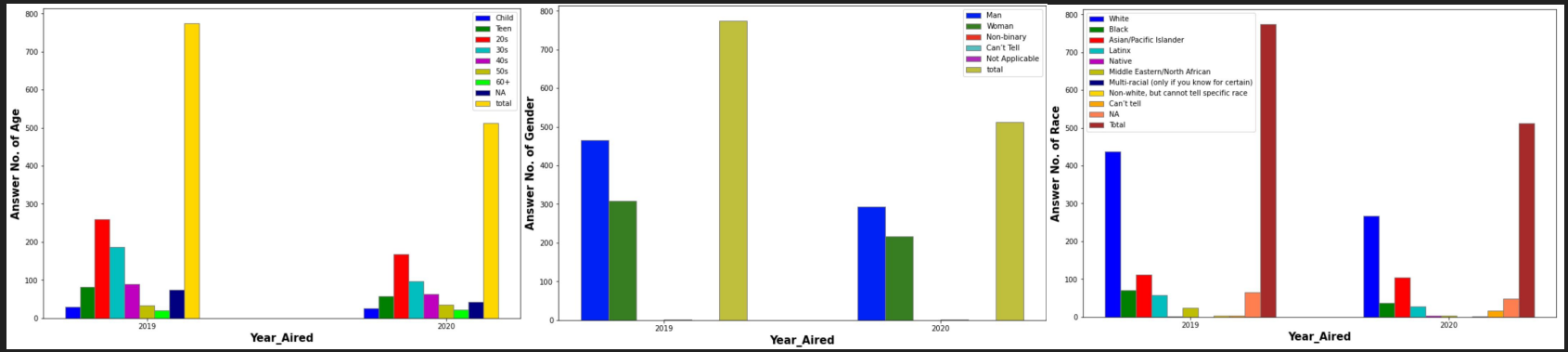
- **What visualizations can you show?**





5. What are the trends?

- Interpretation of this key question:
  - The data moving tendency in terms of the columns (e.g. age, gender, race) over time (i.e. year_aired).
- Trends of Age:
  - The total number decreases from 2019 to 2020.
  - “20s” is the largest value between two years.
  - Some didn't tell or are unwilling to disclose the information of age.
- Trends of Gender:
  - We find that the total number decreases from 2019 to 2020.
  - Male numbers are larger than female numbers in both two years.
- Trends of Race:
  - We find that the total number of Race decreases from 2019 to 2020.
  - Among all different answers, option “White people”, is the largest in both two years.
  - The total number of rest options is almost the same as the number of option “White people”.



6. Challenges & Conclusions

- **Any additional challenges you have encountered?**
  - Not enough data
  - Scope of key questions
  - Not strong consistent between code book and dataset
- **What limitations have you encountered? What assumptions have you had to make?**
  - Limitation: Too many features to choose from
  - Assumption: Only use the most interpretable features to do analysis (e.g. sex, age, race)
- **Conclusions**
  - Analyzed industry differences and summarized changes in representation in advertisements
  - Detected the trends of representations in advertisements