

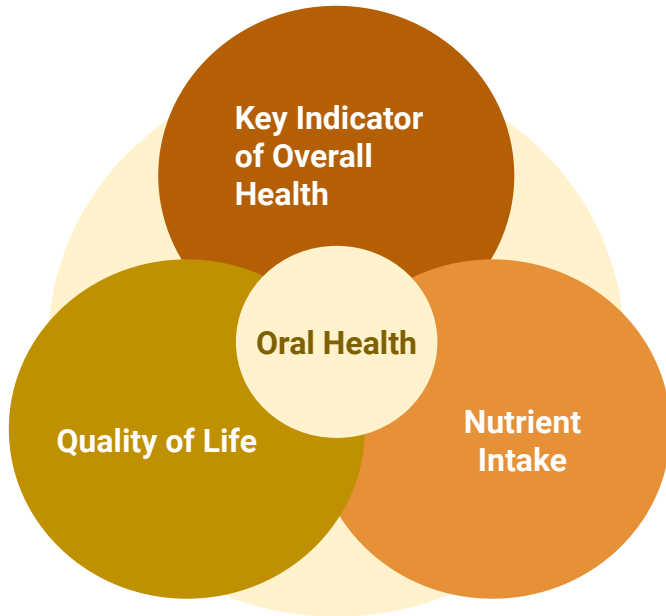


Analysis of Dietary Intake and Dental Health

Group 18: Xihe Gu, Chang Liu, Sree Bhargavi Balija, Parth Hiren Shah, Ameya Panse

Nov.23.2022

Introduction



Questions:

- What **kinds** of dietary supplement do the participants usually take?
- Any **habits** regarding nutrient intake for the participants?
- Any **correlation** between nutrient intake, dietary supplement, and dental health?

Dataset Overview and Preprocessing

National Health and Nutrition Examination Survey (NHANES)



15,000+ participants

Dietary supplement intake data: **34** attributes

Individual foods/Total nutrient intake data: **64** attributes

Oral health data: **6** attributes for **overall oral health**, **6** for **each** specific tooth

Pre-processing

Drop empty
attributes/entries



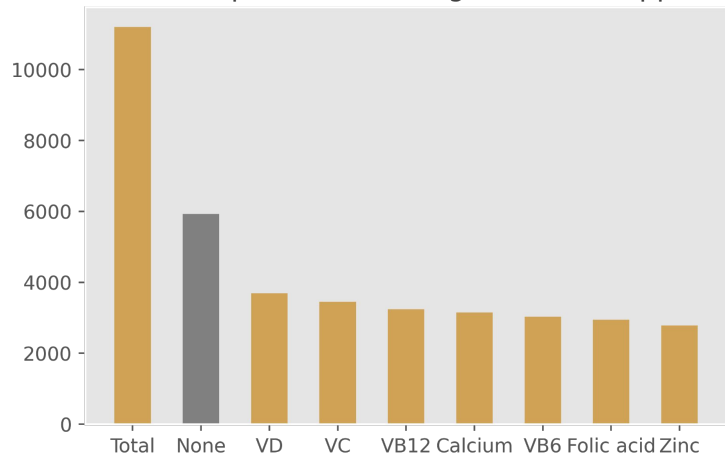
Fill in empty/"refused"
cells



Product labels by
combining dental data

Dietary Supplement Dataset

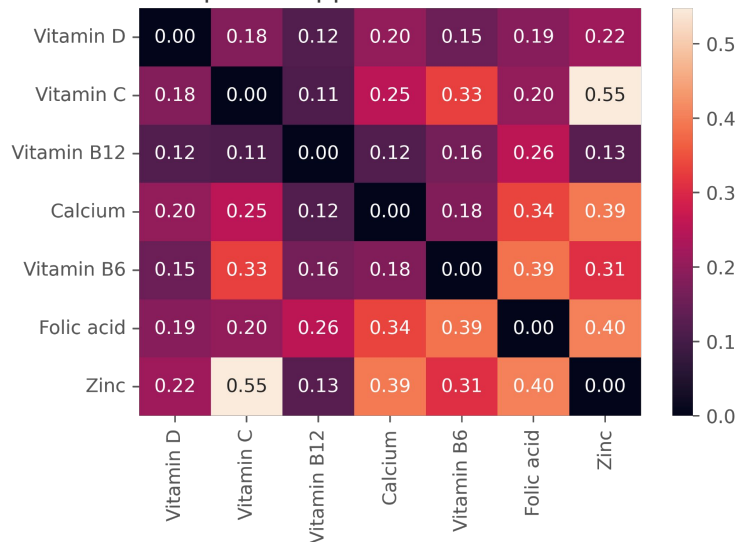
Number of Respondents Having Different Supplements



Mean Pearson Correlation:

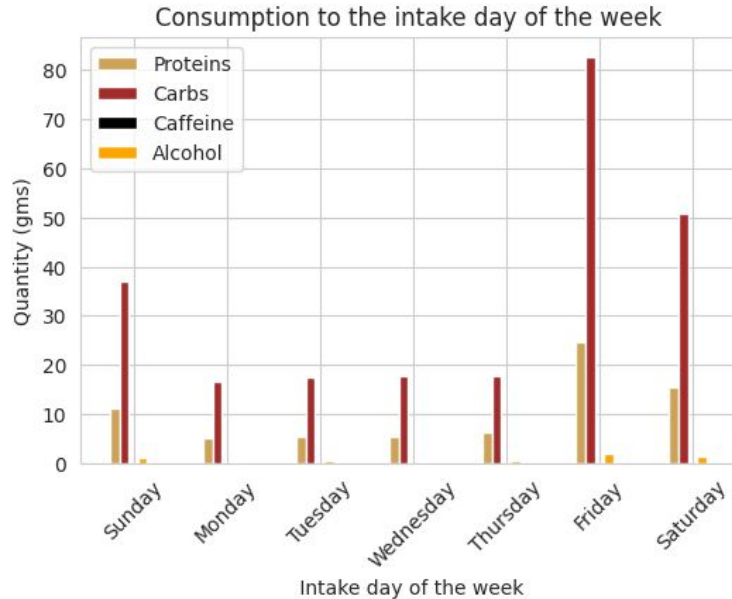
- All 34 supplements: 0.13
- Top 7 supplements: 0.24

7 Most Popular Supplements Intake Correlation

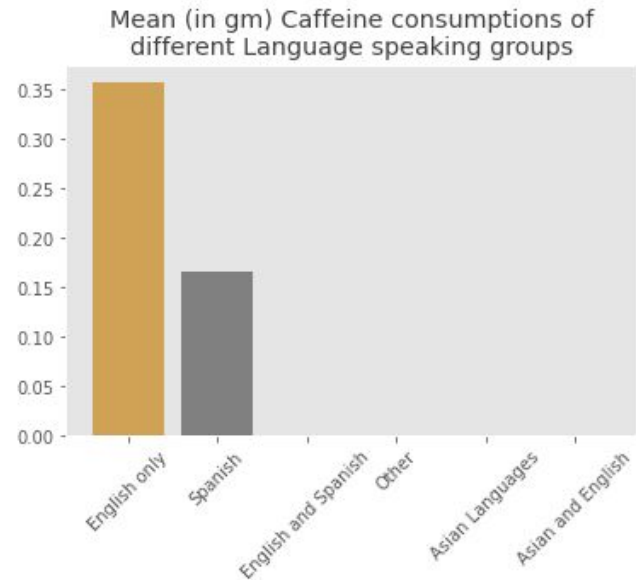


People who are taking one common supplement is more likely to take other common supplements at the same time.

Nutrient Intake - Data Insights

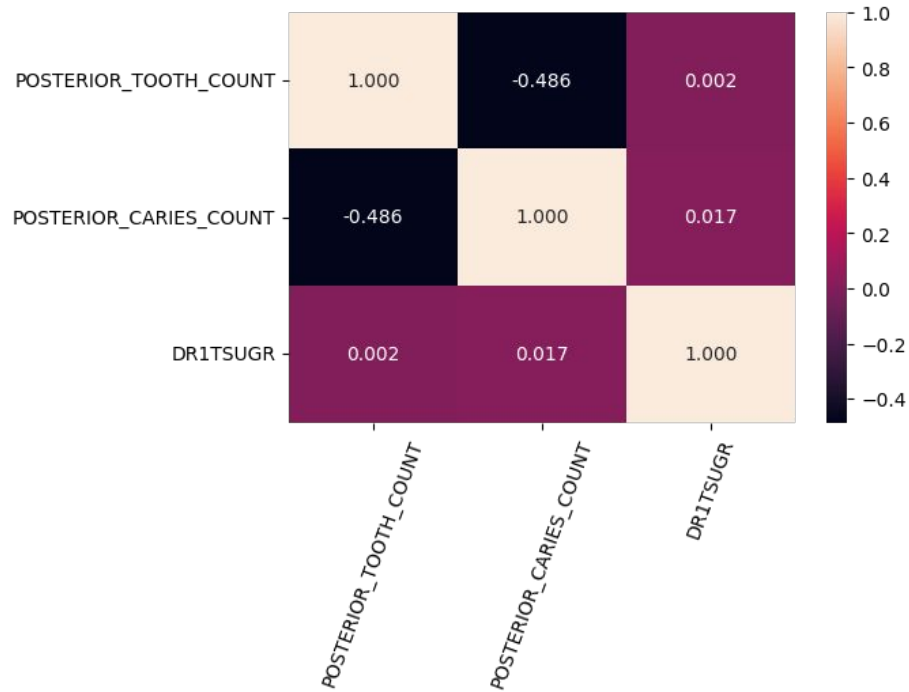
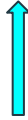


Respondents intake 140% more nutrients on Friday than the average of a week.



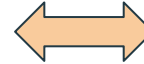
Caffeine consumptions in English and Spanish speaking countries.

Total Sugar \Leftrightarrow Dental Caries



Cross Dataset Insights

Number of dietary supplements and Number of foods/beverages reported ($\rho = 0.28$)

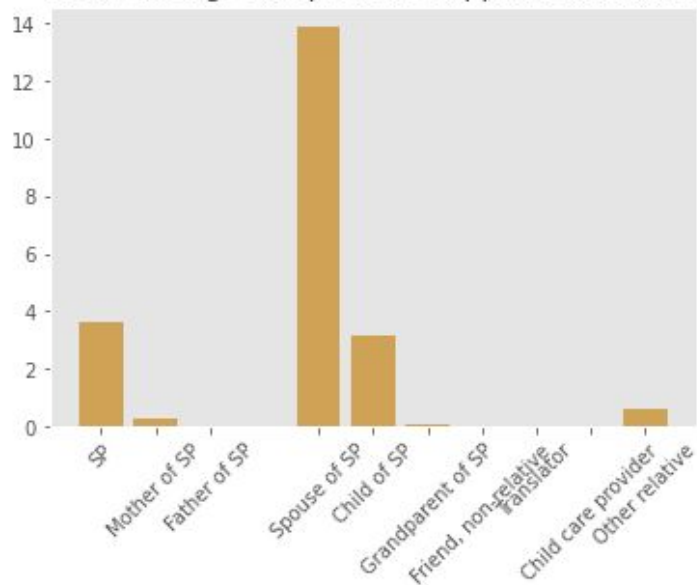


Caffeine and Root caries ($\rho = 0.28$)

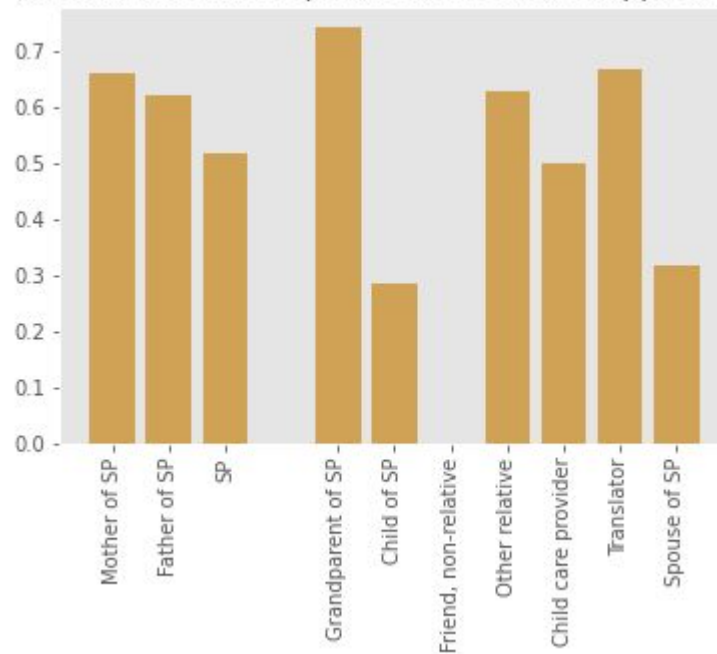


Cross Dataset

Mean (in mg) Phosphorous Supplements taken



% distribution of respondents who take supplements

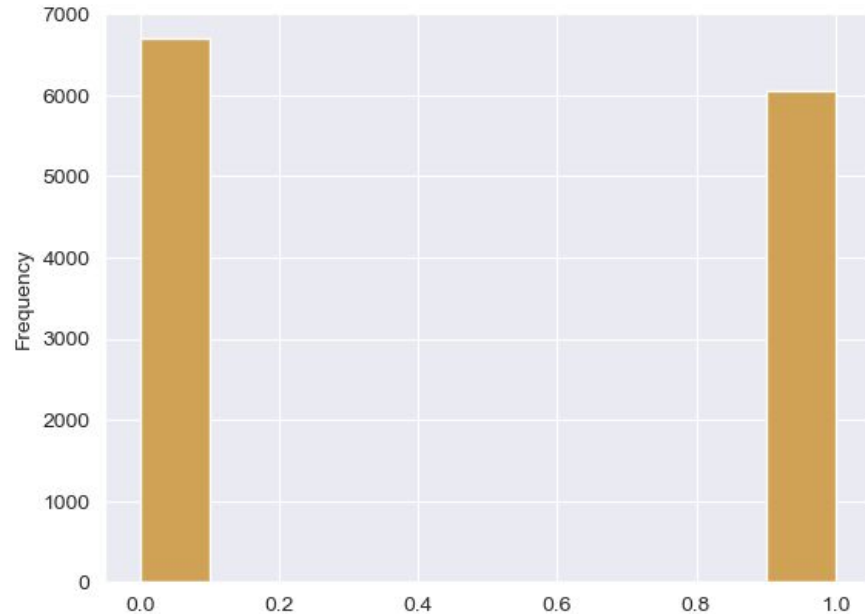


Dental Dataset - Labelling

Convert ~150 sparse columns to 0-1 labels

Serious Dental Health Issue :

- More than 6 Caries
- At least one Root Carie



Model

Pre-Processing

- Numerical Data
 - 0-1 Scaling
 - PCA
- Categorical Data
 - One Hot Encoding
- Train-Test Split : 80%-20%

Models

- Linear Models
- Random Forest
- Support Vector Machines
- ADA Boost

Score

- PCA shows Linear Separation Not Possible
- Random Forest : 66%
- SVM : 68%
- ADA Boost : 70%

Conclusion

Observed

- High nutrient intake during weekends
- Diet difference matters
- High caffeine consumption in western-speaking countries

Predicted

Serious dental issue: ADA Boost Accuracy 70%