SWAN Data Analysis

SUMMER 2022

IN PARTNERSHIP WITH:

hey freya

Overview

1. Introduction to Research Questions
2. Data Design & Processing
3. Findings by Research Question
4. Conclusions
5. Additional Areas of Interest
6. Dashboard Demonstration
7. Individual Project Proposals
8. Questions

hey freya

Research questions at outset of project

Hey Freya is interested in patterns and/or correlations between sleep quality & quantity, and/or self-reported lack of energy with nutritional supplementation patterns, sleep medication patterns, dietary estimates of nutrient consumption, and friendship support patterns in all participants who have not yet been assigned as menopausal, related to the day of the cycle. We'd like to see these in the context of age, race, socio-economic status, and disease status.

Data Design

Raw Data Set Challenges

- Column name variance between years
- New variables introduced and removed
- High null rate
- Inconsistent responses such as “yes,” “Yes," and "1: Yes"

Final Dataset Build

- Data spans all years of survey
- Removed columns with more than 50% null values
- Resolved response inconsistency
- Dataset shape: 236 Columns by 28,789 Rows

hey freya

Assessing the indicators of lifestyle stress, hormone levels and sleep quality/quantity

INDEX CATEGORIES

Data Columns Utilized

- Sleep Quality (SLEEPQL)
- Trouble Sleeping (TRBLSLE)
- Wakeup (WAKEUP)
- Wakeup Early (WAKEEARL)

Design

- Binary Variable to Describe those having three or more Reported Sleep Problems

RAW INDEX DISTRIBUTION (66% | 34%)

Proportion	Count	% of Total
0.00%	5,228	19.33%
25.00%	3,925	14.51%
33.33%	2,210	8.17%
50.00%	4,110	15.19%
67.67%	2,393	8.85%
75.00%	3,878	14.34%
100%	5,308	19.62%

XG boost Modeling

Model Advantages

- Handles missing values
- More powerful than a basic decision tree model

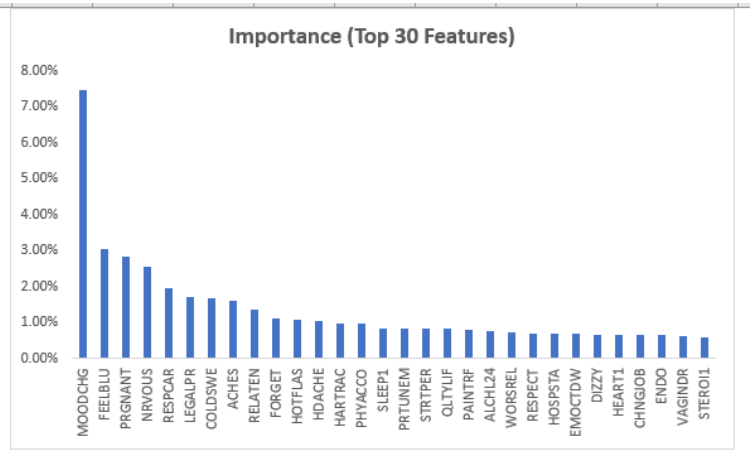
Sleep Index Model Outcome

- Weighted accuracy of 72.52%
- Better at Predicting those without sleep issues than those with sleep issues

hey freya

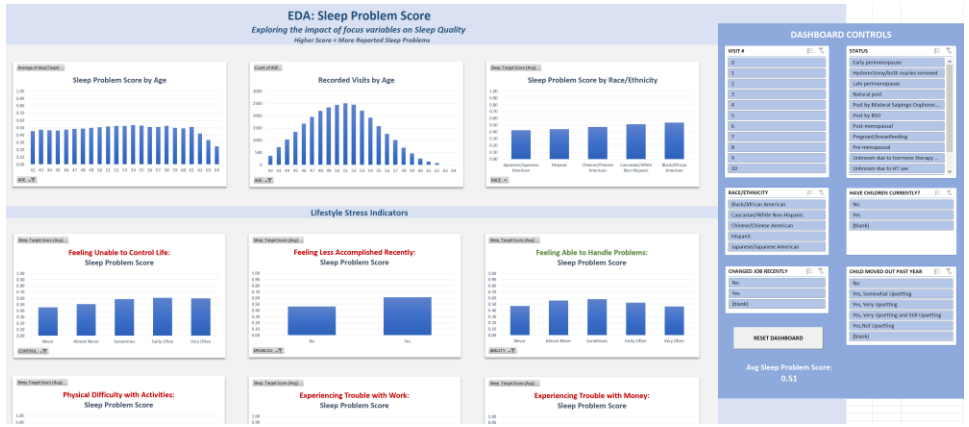
Model Feature Importance

Metric	Importance (Scale of 100%)	Text
MOODCHG	7.46%	Freq mood changes past 2 weeks
FEELBLU	3.03%	Feeling blue past 2 weeks
PRGNANT	2.85%	Pregnant since last visit
NRVOUS	2.54%	Tense/nervous past 2 weeks
RESPCAR	1.96%	Responsibility for care - how upsetting past year
LEGALPR	1.70%	Legal problems - how upsetting past year
COLDSWE	1.67%	Cold sweats past 2 weeks
ACHES	1.61%	Back aches/pains past 2 weeks
RELATEN	1.36%	Ended relationship - how upsetting past year
FORGET	1.12%	Forgetfulness past 2 weeks
HOTFLAS	1.08%	Hot flashes past 2 weeks
HDACHE	1.04%	Headaches past 2 weeks
HARTRAC	0.98%	Heart pounding/racing past 2 weeks
PHYACCO	0.98%	Accomplished less past month due to health
SLEEP1	0.84%	OTC Sleep med #1 taken 2x/wk last mo
PRTUNEM	0.84%	Partner unemployed - how upsetting past year
STRTPER	0.83%	Start Period in Last Week
QLTYLIF	0.82%	Quality of life
PAINTRF	0.79%	Pain interfere w/ work past month
ALCHL24	0.76%	Alcohol in Last 24 hours
WORSREL	0.71%	Worsening relationship - how upsetting past year
RESPECT	0.70%	Treated w/ less respect than others
HOSPSTA	0.69%	Hospital stays since last visit
EMOCTDW	0.68%	Cut down on activities/work past month due to emotional problems
DIZZY	0.67%	Dizzy spells past 2 weeks
HEART1	0.67%	Heart med #1 taken 2x/wk for last mo
CHNGJOB	0.66%	Change in job since last visit
ENDO	0.65%	Endometriosis difficult
VAGINDR	0.62%	Vaginal dryness past 2 weeks
STEROI1	0.60%	Steroid #1 taken 2 times/week for the last month



Sleep Index Dashboard

Enables user control for data exploration



- Slicers (aka filters) provide ability to drill into specific subsets based on features such as ethnicity, hormonal status, or presence of children in the household
- Select any combination of features and observe the effect had on observed traits within the subset
- Quickly compare between variables of interest, such as lifestyle or support indicators, and the target variable (Sleep Index Score)

hey freya

Sleep Index Dashboard

Example Findings

Those who scored with the most sleep issues were also the most likely to answer:

- Feeling the most out of control in life
- Feeling the least accomplished recently
- Experiencing and feeling very upset about money or work-related issues

Those who scored with the least sleep issues were also the most likely to answer:

- Always having a listener or confidant available to them
- Not having serious family issues
- Being either single or currently married (rather than divorced, separated, or widowed)

hey freya

Longitudinal patterns of lifestyle stress indicators with forms of medications, nutritional supplementation, or alternative medicines

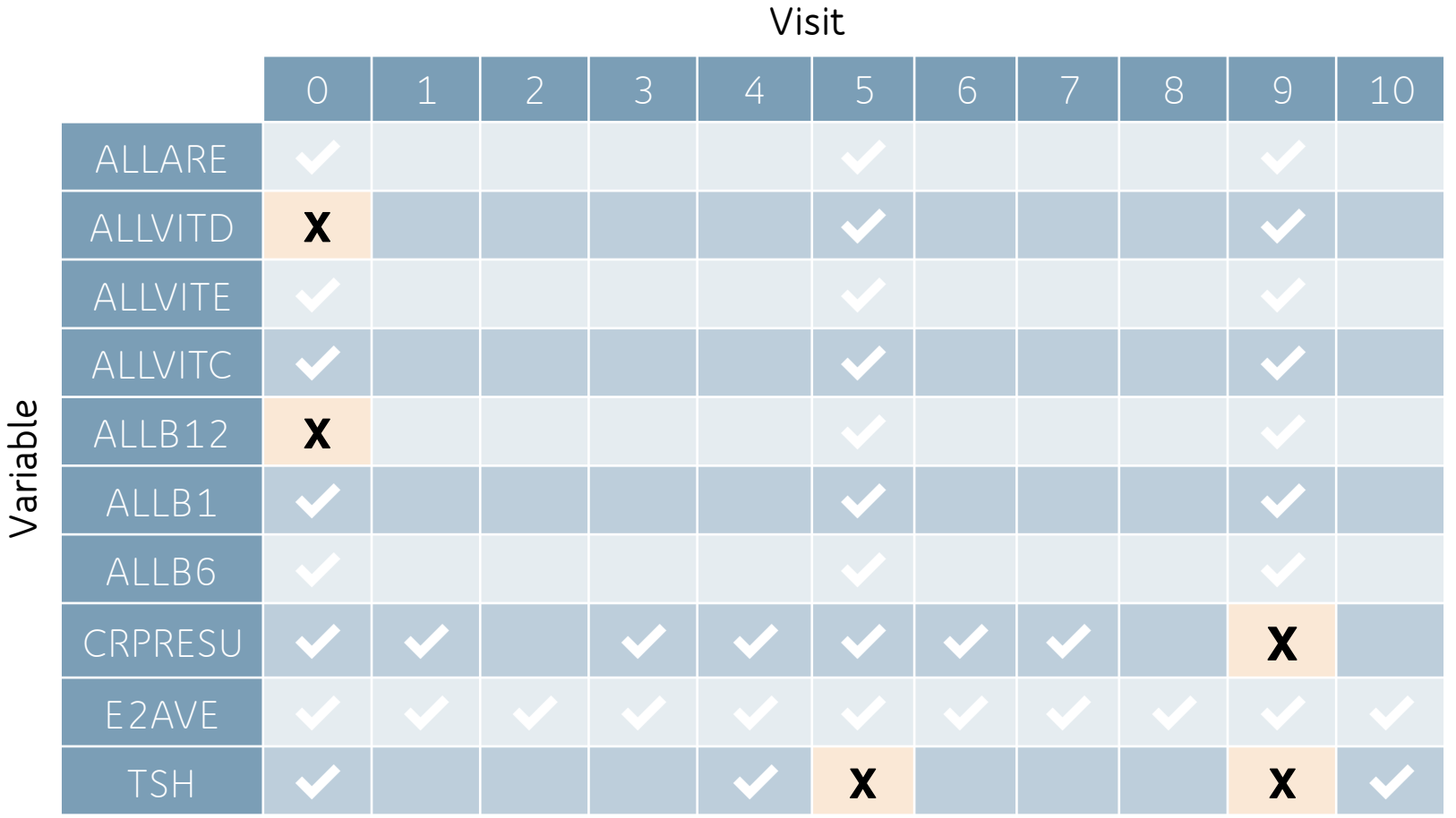
MIXED EFFECTS BINOMIAL REGRESSION

- Predictor variables used:
  - SWANID, VISIT, STATUS, ALLARE, ALLVITC, ALLVITD, ALLVITE, ALLB1, ALLB6, ALLB12, ALLCALC, ALLFOL, ALLIRON, ALLZINC, E2AVE
- Response variable(s) used:
  - TRBLSLE, WAKEUP, WAKEEARL
  - Binarized to represent the presence of sleep issues (where 0 indicates the participant did not report sleep disturbances and 1 indicates the participant reported, at minimum, one of the above sleep disturbances three or more times a week)

- Overall, the model yielded **60.57% accuracy**
  - Sensitivity: 50.58%, Specificity: 69.10%
- Results showed that **ALLVITC, ALLIRON, and early perimenopausal STATUS had a statistically significant effect** on sleep quality
  - Each unit increase of Vitamin C (ALLVITC) is associated with ***a decrease of 12%*** in the odds of experiencing sleep disturbance.
  - The early perimenopausal stage (STATUSEarly Perimenopausal) is associated with ***a 54% reduction*** in the relative risk of sleep disturbance.
  - Each unit increase in iron (ALLIRON) is associated with ***an increase of 22%*** in the odds of experiencing sleep disturbance.

Baseline, midline, endline patterns of nutritional markers and if deficiencies were successfully addressed by supplementation

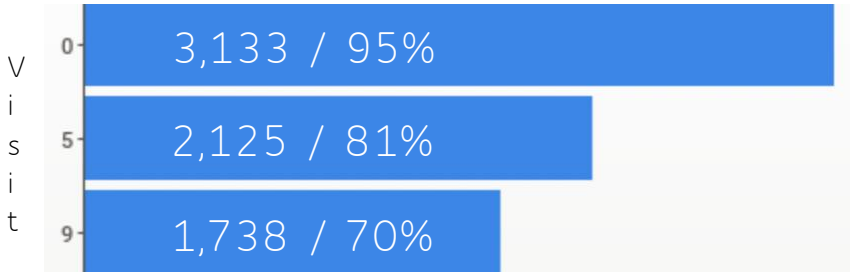
- A complete exploration of each supplement is available in the appendix of the deliverables.
- Model-building would benefit from another round of refining the question b/c inconsistent data.
- Understanding more about what constitutes a deficiency is key to fully addressing the question.

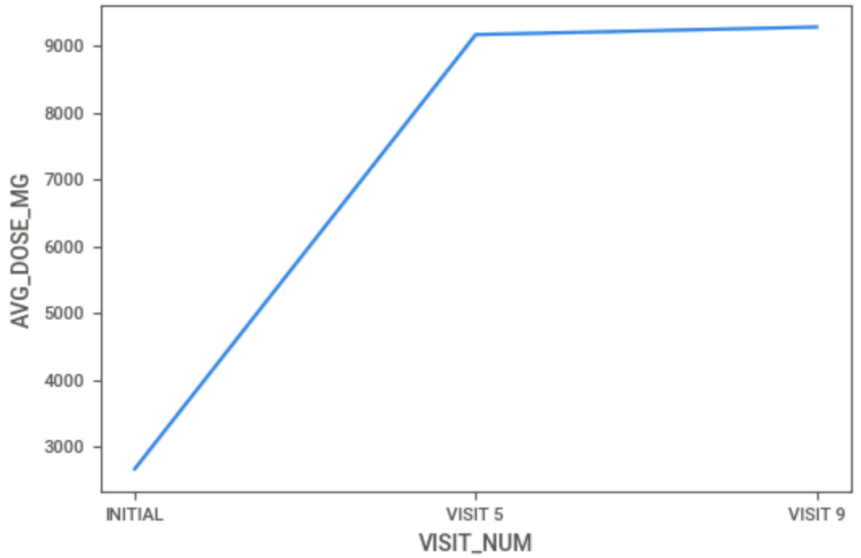


hey freya

ALLARE – Diet + Supp Vit A, retinol equiv. | Visits 0,5,9

Number of patients taking ALLARE by visit / percent of patients who report taking ALLARE when ALLARE is tracked

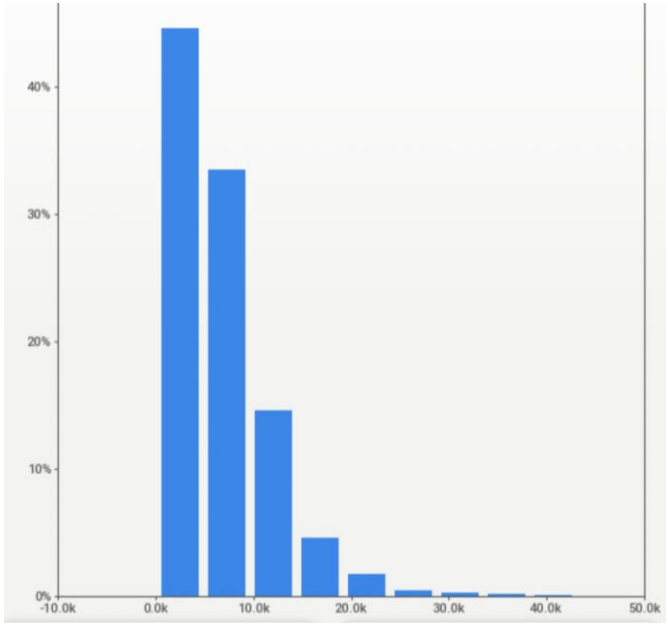




MOST FREQUENT VALUES		
8473.67	2	<0.1%
1804.62	2	<0.1%
1022.27	2	<0.1%
6209.82	2	<0.1%
1562.04	2	<0.1%
5846.88	2	<0.1%
423.69	2	<0.1%
453.66	2	<0.1%
1197.75	2	<0.1%
733.59	2	<0.1%
5467.84	2	<0.1%
6116.27	2	<0.1%
642.13	2	<0.1%
957.58	2	<0.1%
803.31	2	<0.1%

SMALLEST VALUES		
125.62	1	<0.1%
173.35	1	<0.1%
181.86	1	<0.1%
194.23	1	<0.1%
196.15	1	<0.1%
204.63	1	<0.1%
207.39	1	<0.1%
213.79	1	<0.1%
216.28	1	<0.1%
223.12	1	<0.1%
224.26	1	<0.1%
236.46	1	<0.1%
246.09	1	<0.1%
248.99	1	<0.1%
251.32	1	<0.1%

LARGEST VALUES		
47919.17	1	<0.1%
43477.67	1	<0.1%
43272.7...	1	<0.1%
42843.85	1	<0.1%
42796.6...	1	<0.1%
39566.53	1	<0.1%
38630.71	1	<0.1%
38476.04	1	<0.1%
36662.32	1	<0.1%
35688.48	1	<0.1%
35664.12	1	<0.1%
35430.29	1	<0.1%
35294.82	1	<0.1%
35143.83	1	<0.1%
35039.51	1	<0.1%



Reported dosage unit for Vit. A is unclear

Reported dosage unit unclear

The impact of motherhood shifts or career stage changes on stress, sleep and nutritional markers

Data Columns Utilized

- Responsible for the care of another? Is it upsetting? (RESPCAR)
- Have children/stepchildren? (CHILDRE)
- Role as a mother rewarding? (REWRDMO)
- Role as a mother stressful? (STRSSMO)
- Current job stressful? (STRSSJO)
- Current relationship stressful? (STRSREL)
- Perceived stress score (P_STRESS)
- Caregiver role stressful for older adult or disabled family member? (STRSCAR)

LIMITATIONS: MISSING DATA, NO INDICATORS OF MOTHERHOOD SHIFTS OR CAREER CHANGES

Percent Missing Data:

RESPCAR: 6.07% (Visits 0-10)

CHILDRE: 41.87% (Visits 1-6, 8)

REWRDMO: 52.16% (Visits 1-6, 8)

STRSSMO: 52.15% (Visits 1-6, 8)

STRSSJO: 54.52% (Visits 1-6, 8)

STRSREL: 52.04% (Visits 1-6, 8, 10)

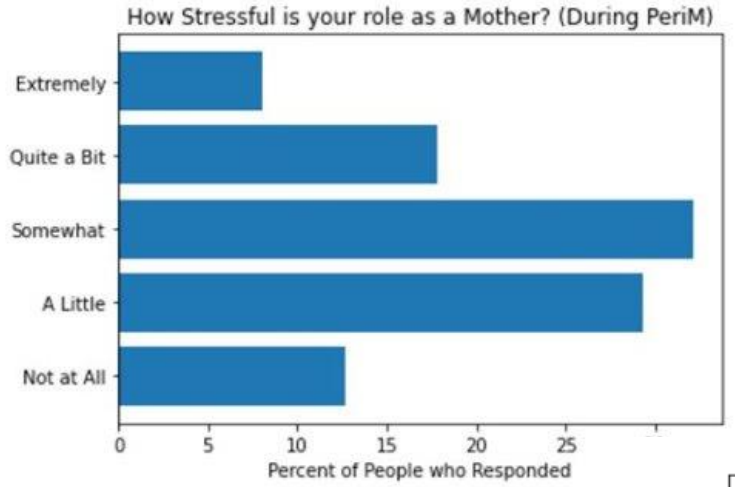
P_STRESS: 34.9% (4 variables that are in Visits 1-6, 8, 10)

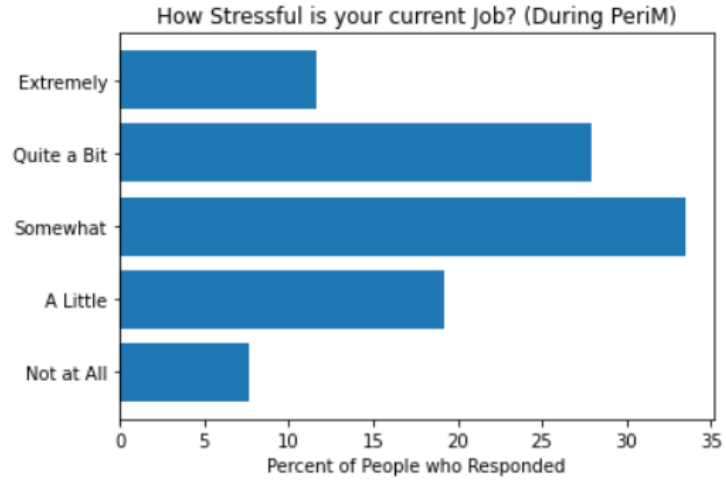
STRSCAR: 92.67% (Visits 1-6, 8)

hey freya

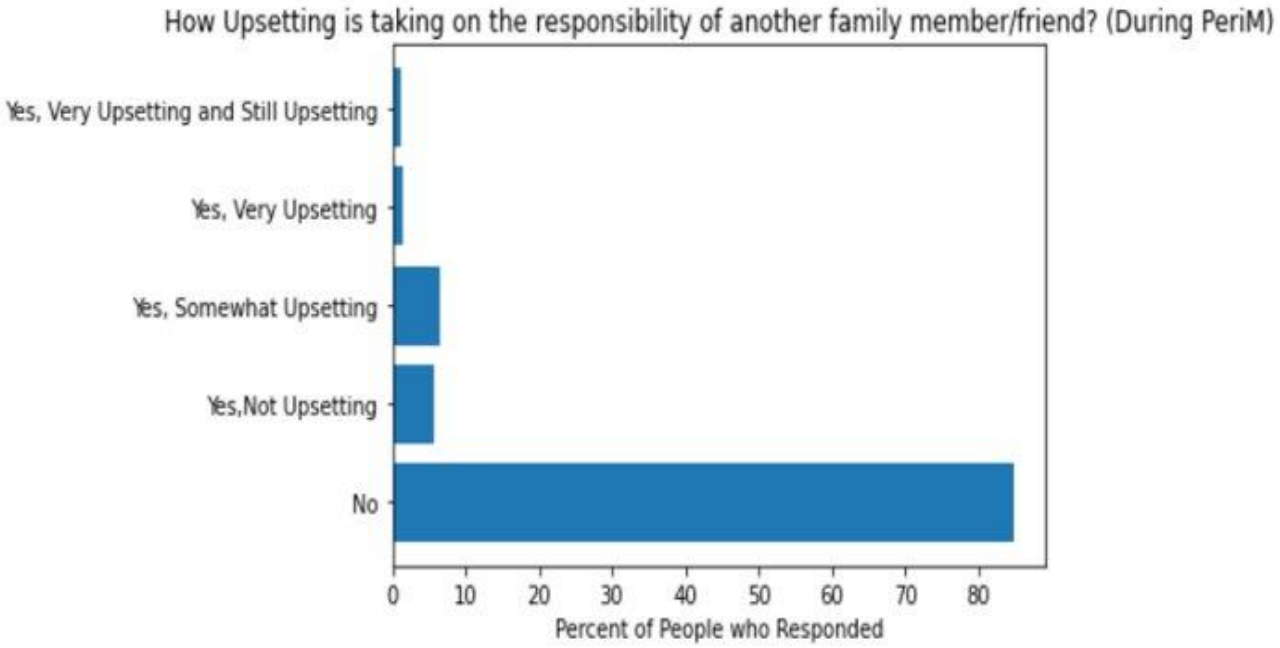
Motherhood and Career Impacts on Stress

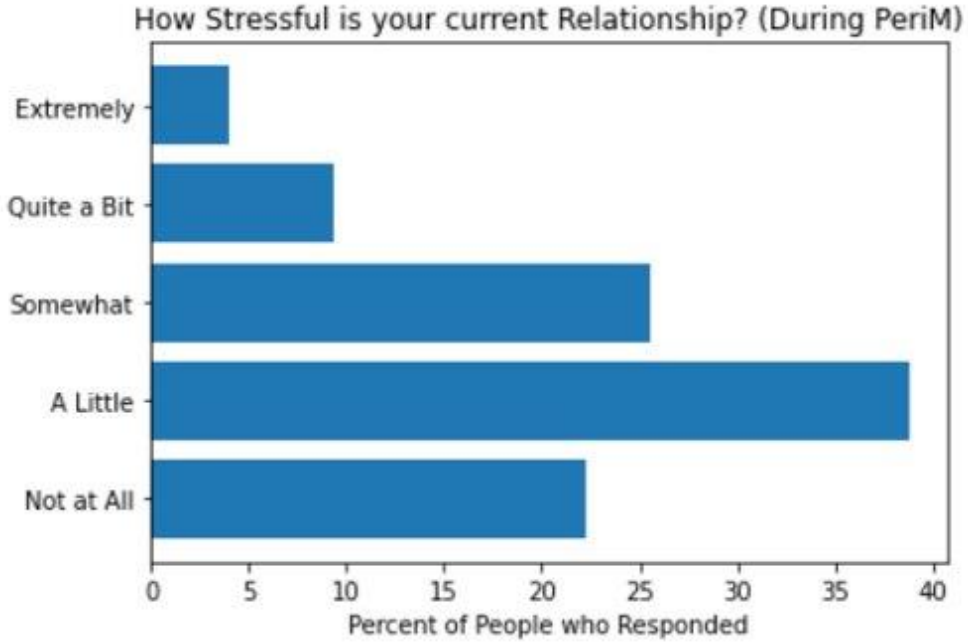
- Directly asked in some of the questions
- Marginal differences in distributions for these responses between women currently going through perimenopause and those who weren't (mostly post-menopause).
- Main differences were:
  - Women going through perimenopause are slightly more stressed from their role as a mother
  - More women not going through perimenopause were not at all stressed by their current relationship





Motherhood and Career Impacts on Stress (cont'd)





Motherhood and Career Impacts on Sleep

METHODS

- Used XG Boost modeling with previous created sleep index.
- Predictors used were the variables of interest (except for STRSCAR) as well as the RACE variable to see if the impact on sleep that these variables have varies for different races.

RESULTS:

- 78.7% accurate
- Most important features: RESPCAR, RACE, P_STRESS, and CHILDRE (STRSSMO is valued highly if P_STRESS isn't used in the model)..
- Different races have different trends in sleep quality in general.
- Caregivers who are stressed by their role trended towards having a worse sleep score.
- Women going through perimenopause who indicated that they have children averaged a slightly worse sleep score than those who don't have children.

** Nutritional markers were hard to track with this data

Additional Areas of Interest: Exercise

EXERCISE

Model: XGBoost Classifier

Predicted whether participant wakes up often

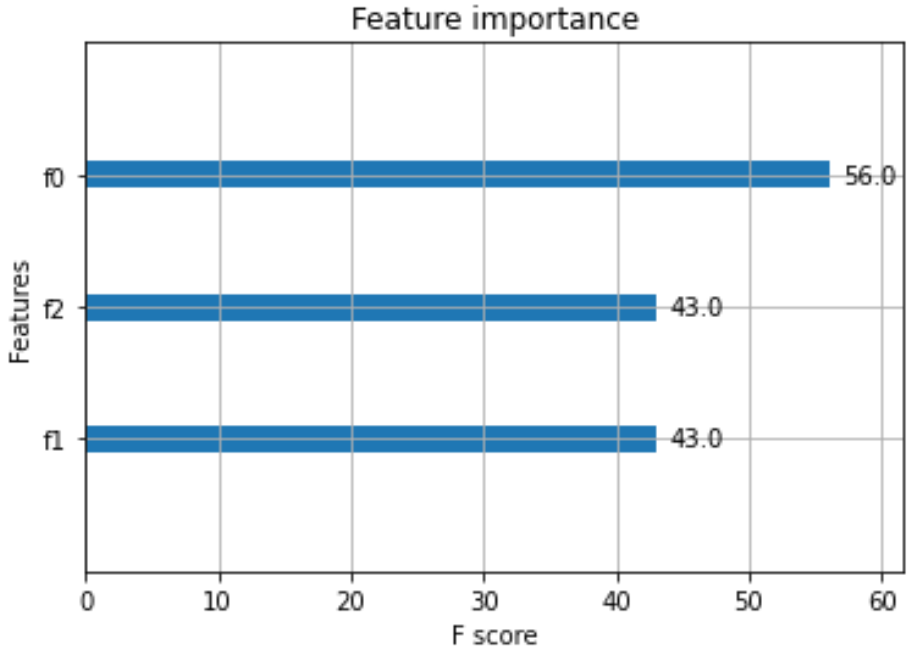
- Accuracy: 71% using just exercise attributes
- Model is better at predicting those with wake-up problems than without.

Variables of Consideration

- Mild exercise

LIMITATIONS:

- Exercise variables were frequently left out of visits.



hey freya

Conclusions

MODEL SUMMARIES:

- Motherhood and careers data showed what you would expect. Need better tracking over time.
- Mild exercise predicts whether a participant wakes up, but exercise overall was not a strong indicator of the "Sleep Target" variable.
- Regression results showed that iron and estradiol had a statistically significant effect on sleep quality

SUGGESTIONS FOR DATA COLLECTION:

1. Obtain NIA BioBank data which contains bio markers from blood and urine samples. (Contact SWAN for permission)
2. Design proprietary studies. The HF application can enable uniform predictor and target variable collection that adheres to consistent intervals.

hey freya

Interactive Dashboard Demo

hey freya

Individual Proposals

Individual project proposal – Food Deserts

Hey Freya indicated that it would be meaningful to view their project questions within the context of race and socio-economic status in addition to age and the presence of disease.

In recent years, "food deserts" have become a much more widely discussed class and race issue in America. Knowing also, that the number one disease prevention methodology (worldwide) is a multi-vitamin, we propose designing an extension of the SWAN study that geo-locates respondents in order to better understand the inherent geographic challenges that study participants may face in the context of nutritional data, its interpretation, and its inferences/conclusions.

A variety of classification models could be used to predict regional health outcomes if the right data were able to be obtained.

Individual Project Proposal – Social inconsistencies of self-reporting

Bias and unethical data collection is a serious concern related to health research. It is well supported that there are differences in the accuracy and understanding of data between men's and women's health, but it has become increasingly important to better understand the differences in subpopulations based on their unique features and dynamics. For example, it can be unclear whether observed trends in self-reported symptoms are based on sex, ethnicity, socioeconomic status, cultural acceptance of medicine, or other factors.

To better identify and support those who may need treatment, the intent of this research proposal is to investigate differences between population subsets regarding the accuracy of, and likelihood of, self-reported symptoms which may result in the diagnosis of a medical or psychological condition.

Individual Project Proposal – World and cultural event influence on health

External factors impact people in a longitudinal manner, but these are not addressed in the SWAN data. For instance, how did the same people change before during and after the events of 9-11? How did people respond to the housing crash of 2007? Are there statistically significant variances in any of the subsequent years by participant? If so, what are those changes and how can they be mitigated? Are there individuals who navigated these issues better than others? What did they have in common? A greater study into these events and can help doctors provide the appropriate type of treatment dependent on the existing environment. Look no further than current times to demonstrate how this type of research would have been incredibly valuable for the existing pandemic and inflationary issues we are currently experiencing.

Individual Project Proposal – Trial to Study the Effects of Preventative Perimenopausal Treatments on Post-Menopausal Sleep Conditions

Drawing on the work of our investigation into the Studies of Women's Health Across the Nation (SWAN) data, the following proposal seeks to evaluate the effect of preventative treatments for post-menopausal sleep disorder. This work is founded in research that asserts the perimenopausal sleep condition as the greatest predictor of post-menopausal sleep disturbance, one of the most frequently cited health complaints among aging women (Ciano et. al., 2017). Using evidence-based sleep therapies, this experimental design compares the long-term effects of early intervention during the perimenopausal stage. In doing so, the study hopes to identify promising care practices and improve women's quality of life overall.

Individual Project Proposal – Using Resilience as an Indicator for Perimenopause Symptoms

There are existing studies that demonstrate a women's resilience to stressors to be a key indicator of the severity and frequency of perimenopause symptoms such as VMS. There are also studies that identify certain hormones and psychosocial factors as being heavily linked with resilience. Hey Freya can recreate these longitudinal studies to verify their results and use them to guide the design of their own future longitudinal studies as they test treatments for menopausal symptoms in perimenopausal women.

Some of these psychosocial variables are optimism, self-esteem, emotional stability, etc. Several medications are already used to treat these areas and can be juxtaposed with studies that analyze natural remedies.

Hey Freya could consider reconstructing some of these studies to validate their conclusions because a) these were early studies done into women's health that had some hiccups, and b) some of the data presented in these studies have obvious flaws.

Individual Project Proposal –
Research on the Impact of Career Stage Change in Women between 30-50

A subset of the research that Hey Freya would like to focus on is the impact of career stage changes on stress, sleep, and nutritional markers. I'm proposing that this impact also be looked at in the context of the industry women work in, whether they are in a particularly high-stress industry, one that is dominated by women or men, or one that is disproportionately affected by the gender pay gap.

There are several studies that illustrate the gender pay gap, showing that women, on average, earn $.80 for every dollar a man is paid. Additionally, a study about startups by Carta, a silicon Valley firm that helps start-ups manage equity shares, discovered that for every $1 in company equity held by men, women hold just 47 cents (The Washington Post, 2018). Some industries are disproportionately affected by the gender pay gap. For example, Health Care and Social Assistance, Professional and Technical Services, and Financial and Insurance have the largest gaps. At the same time, Construction, Arts, Entertainment and Recreation, and Agriculture and Related Industries are relatively equal.

Thank You.
Questions?