Stress Detection

Group 15:

Ben Butler, Ian Franda, Jack Grosskreuz, Annika Kennerhed, Sara Bay

Introduction to dataset

3000 rows of daily data (100 participants, 30 days)

Perceived stress scores: range 0-40, (the dataset only ranges from 10-40)

- low: 0-13
- moderate: 13-26
- high: 27-40

Key variables include personality traits, phone usage, sleep metrics, skin conductance and physical activity.

Proposal questions

Our analysis will focus on the following questions:

- Which factors are most correlated with stress levels?
    - Which personality traits are most correlated with high stress?
    - How can stress levels vary throughout the week?
    - Is phone usage correlated with stress level?
    - How is sleep/physical activity correlated with stress levels?

We will use the results from our data to address these questions:

- Are these determining factors easy to control or manage?
- What would be potential ways to mitigate high stress levels?

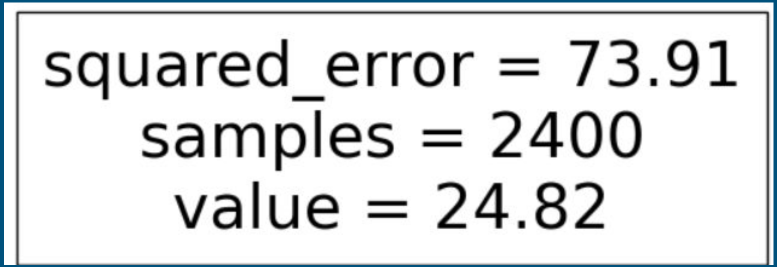
Methods

- No missing values
- Split data 80/10/10 by participant IDs into train/val/test
- Perceived Stress Score (PSS) column used as y values

Attempt #1: Decision Tree Regressor

- Reasons for using DecisionTreeRegressor:
  - All variables are numeric
  - Decision tree will choose most relevant variables and sort out the least useful variables
  - Many variables are on a scale, which means linearity might not make sense
- Attempted DecisionTreeRegressor
  - Used grid search to find best parameters
    - max_depth = 1
    - min_impurity_decrease = 0.45
    - Finds that the best tree has no splits
  - Guesses 24.82 (mean) for every point

Our Decision “Tree”



squared_error = 73.91
samples = 2400
value = 24.82

Attempt #2: Decision Tree Classifier

- Bin stress levels into high stress level (27-40) and low/moderate stress (10-26)
- Attempted DecisionTreeClassifier
  - Score of 0.53 on validation data
  - 57% of validation data is low/moderate stress
  - Guessing low/moderate every time makes a better model than the Decision Tree Classifier

Attempt #3: Linear models

- Linear model using “Openness”, “Neuroticism” and “mobility_distance”
- Lasso chooses “Openness”, “call_duration” and “num_sms”

Performance of regression models on validation data

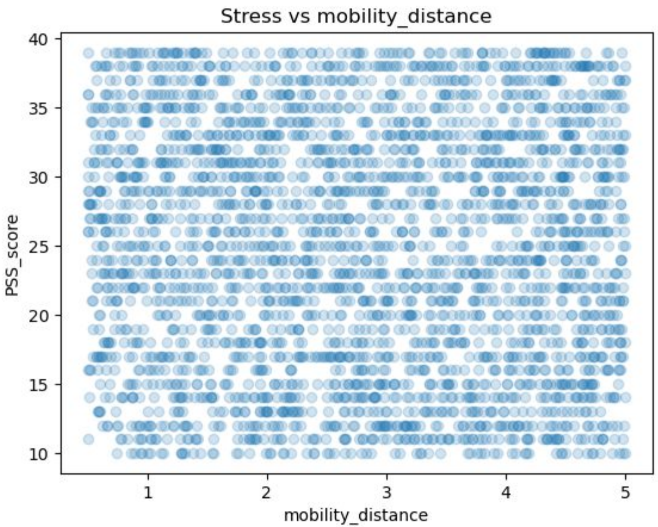
Linear regression does best on validation data, but difference is very small

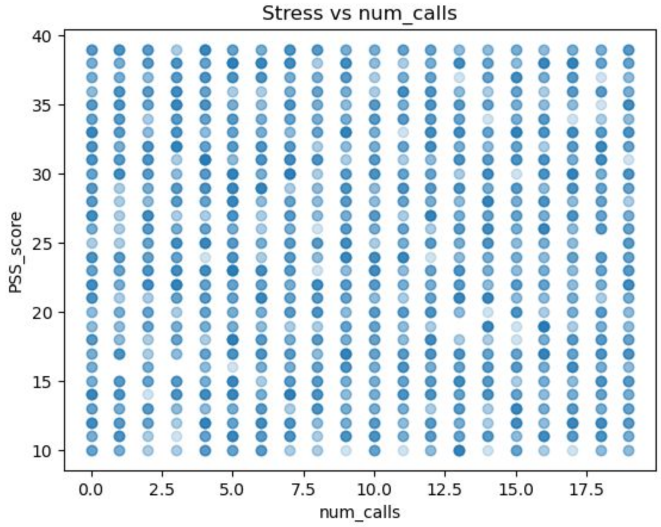
Since the performance measures are so close, we looked more into correlation between features

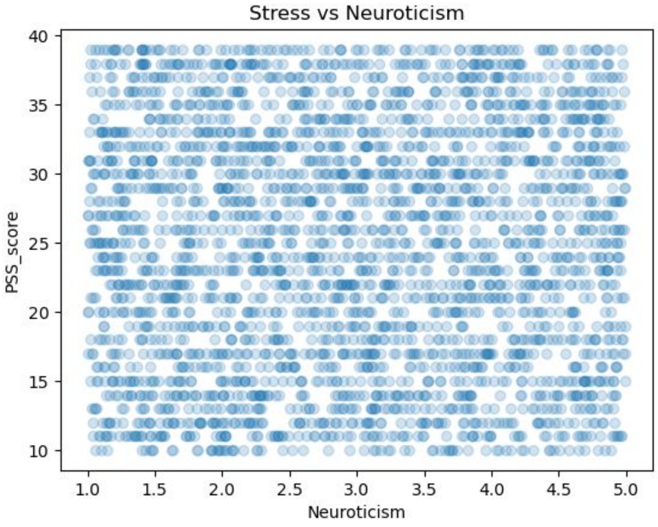
Method	MSE on validation data
Decision tree regressor (guess mean every time)	75.535
LASSO	75.532
Linear Regression	75.498

Correlation check

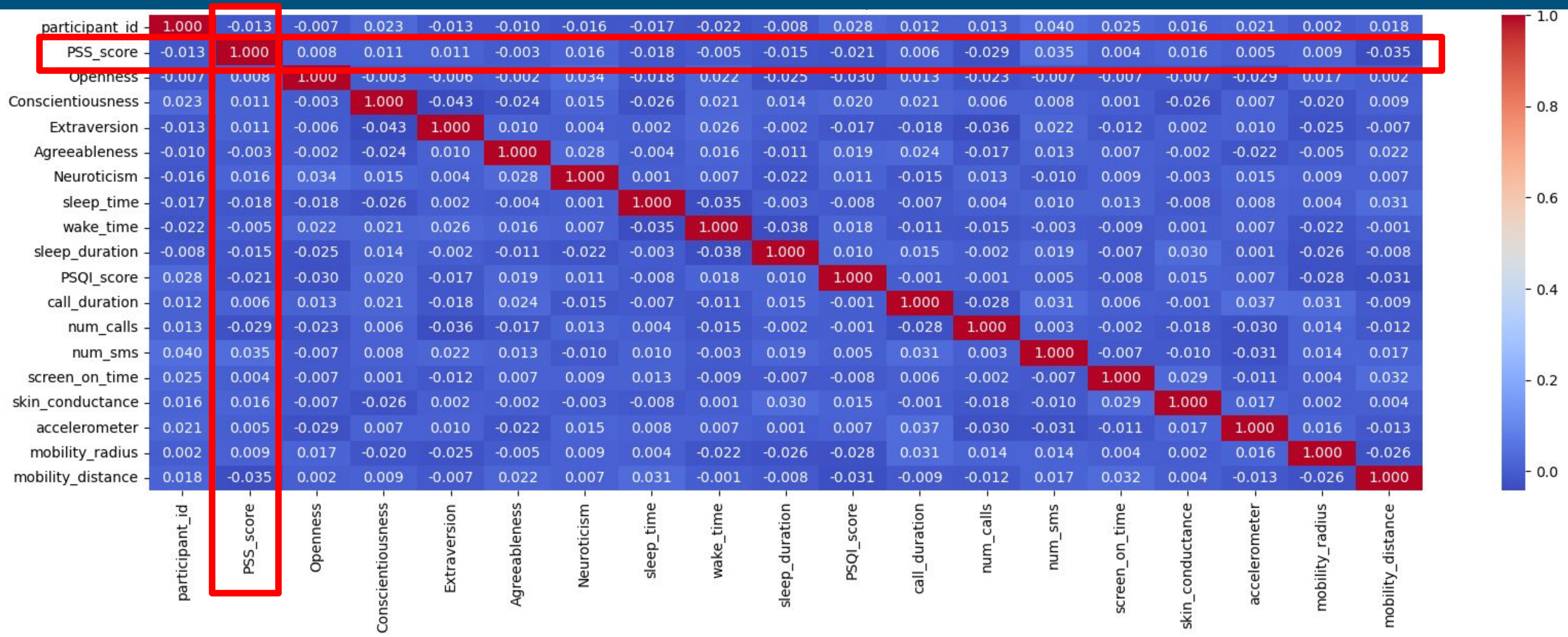
Plots show no correlation between PSS_score and mobility_distance, num_calls and neuroticism.







Correlation Heatmap

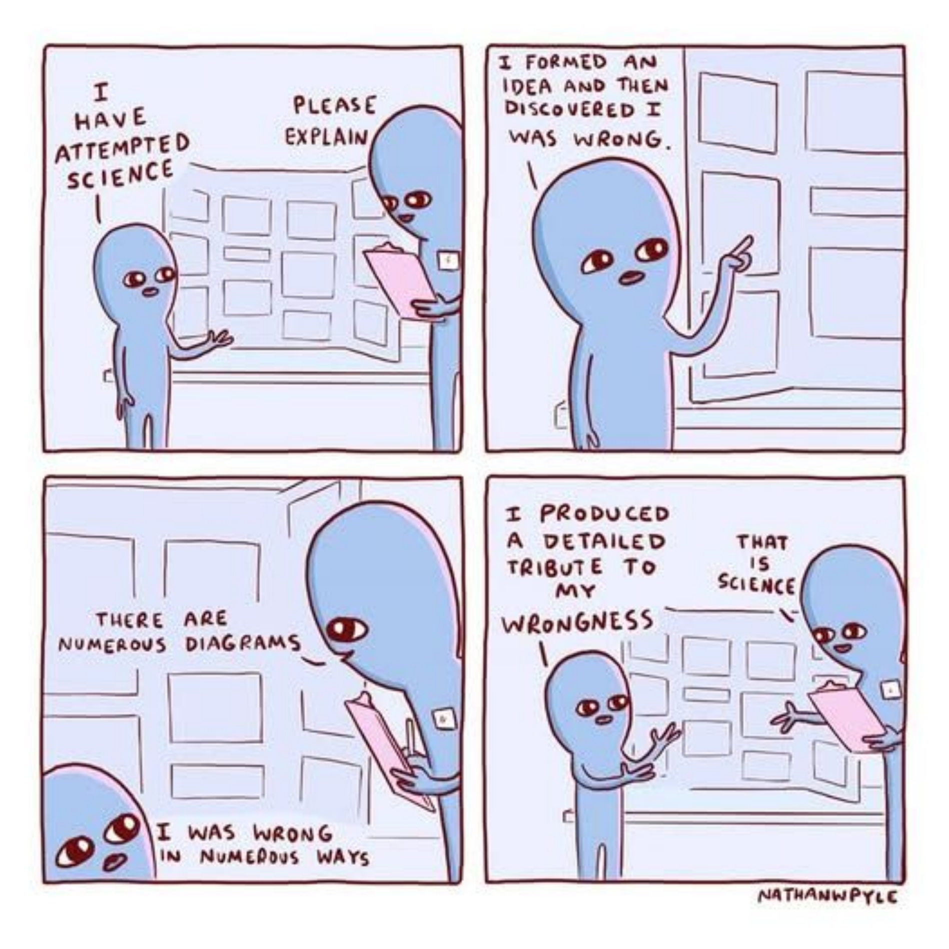


	participant_id	PSS_score	Openness	Conscientiousness	Extraversion	Agreeableness	Neuroticism	sleep_time	wake_time	sleep_duration	PSQI_score	call_duration	num_calls	num_sms	screen_on_time	skin_conductance	accelerometer	mobility_radius	mobility_distance
participant_id	1.000	-0.013	-0.007	0.023	-0.013	-0.010	-0.016	-0.017	-0.022	-0.008	0.028	0.012	0.013	0.040	0.025	0.016	0.021	0.002	0.018
PSS_score	-0.013	1.000	0.008	0.011	0.011	-0.003	0.016	-0.018	-0.005	-0.015	-0.021	0.006	-0.029	0.035	0.004	0.016	0.005	0.009	-0.035
Openness	-0.007	0.008	1.000	-0.003	-0.006	-0.002	0.034	-0.018	0.022	-0.025	-0.030	0.013	-0.023	-0.007	-0.007	-0.007	-0.029	0.017	0.002
Conscientiousness	0.023	0.011	-0.003	1.000	-0.043	-0.024	0.015	-0.026	0.021	0.014	0.020	0.021	0.006	0.008	0.001	-0.026	0.007	-0.020	0.009
Extraversion	-0.013	0.011	-0.006	-0.043	1.000	0.010	0.004	0.002	0.026	-0.002	-0.017	-0.018	-0.036	0.022	-0.012	0.002	0.010	-0.025	-0.007
Agreeableness	-0.010	-0.003	-0.002	-0.024	0.010	1.000	0.028	-0.004	0.016	-0.011	0.019	0.024	-0.017	0.013	0.007	-0.002	-0.022	-0.005	0.022
Neuroticism	-0.016	0.016	0.034	0.015	0.004	0.028	1.000	0.001	0.007	-0.022	0.011	-0.015	0.013	-0.010	0.009	-0.003	0.015	0.009	0.007
sleep_time	-0.017	-0.018	-0.018	-0.026	0.002	-0.004	0.001	1.000	-0.035	-0.003	-0.008	-0.007	0.004	0.010	0.013	-0.008	0.008	0.004	0.031
wake_time	-0.022	-0.005	0.022	0.021	0.026	0.016	0.007	-0.035	1.000	-0.038	0.018	-0.011	-0.015	-0.003	-0.009	0.001	0.007	-0.022	-0.001
sleep_duration	-0.008	-0.015	-0.025	0.014	-0.002	-0.011	-0.022	-0.003	-0.038	1.000	0.010	0.015	-0.002	0.019	-0.007	0.030	0.001	-0.026	-0.008
PSQI_score	0.028	-0.021	-0.030	0.020	-0.017	0.019	0.011	-0.008	0.018	0.010	1.000	-0.001	-0.001	0.005	-0.008	0.015	0.007	-0.028	-0.031
call_duration	0.012	0.006	0.013	0.021	-0.018	0.024	-0.015	-0.007	-0.011	0.015	-0.001	1.000	-0.028	0.031	0.006	-0.001	0.037	0.031	-0.009
num_calls	0.013	-0.029	-0.023	0.006	-0.036	-0.017	0.013	0.004	-0.015	-0.002	-0.001	-0.028	1.000	0.003	-0.002	-0.018	-0.030	0.014	-0.012
num_sms	0.040	0.035	-0.007	0.008	0.022	0.013	-0.010	0.010	-0.003	0.019	0.005	0.031	0.003	1.000	-0.007	-0.010	-0.031	0.014	0.017
screen_on_time	0.025	0.004	-0.007	0.001	-0.012	0.007	0.009	0.013	-0.009	-0.007	-0.008	0.006	-0.002	-0.007	1.000	0.029	-0.011	0.004	0.032
skin_conductance	0.016	0.016	-0.007	-0.026	0.002	-0.002	-0.003	-0.008	0.001	0.030	0.015	-0.001	-0.018	-0.010	0.029	1.000	0.017	0.002	0.004
accelerometer	0.021	0.005	-0.029	0.007	0.010	-0.022	0.015	0.008	0.007	0.001	0.007	0.037	-0.030	-0.031	-0.011	0.017	1.000	0.016	-0.013
mobility_radius	0.002	0.009	0.017	-0.020	-0.025	-0.005	0.009	0.004	-0.022	-0.026	-0.028	0.031	0.014	0.014	0.004	0.002	0.016	1.000	-0.026
mobility_distance	0.018	-0.035	0.002	0.009	-0.007	0.022	0.007	0.031	-0.001	-0.008	-0.031	-0.009	-0.012	0.017	0.032	0.004	-0.013	-0.026	1.000

Outputs/Analysis

- The dataset shows that stress levels cannot necessarily be predicted from:
    - Personality traits
    - Sleep metrics
    - Skin conductance
    - Phone usage
    - Physical activity
- Possible explanations:
    - PSS_scores are conducted from a questionnaire
        - Not precise
        - Different interpretations of questions
    - Different people have different causes/symptoms for stress

Conclusion


I HAVE ATTEMPTED SCIENCE
PLEASE EXPLAIN
I FORMED AN IDEA AND THEN DISCOVERED I WAS WRONG.
THERE ARE NUMEROUS DIAGRAMS
I WAS WRONG IN NUMEROUS WAYS
I PRODUCED A DETAILED TRIBUTE TO MY WRONGNESS
THAT IS SCIENCE
NATHANWPYLE