FYP Presentation, 25th March

Topic: Analysis of Behavioural Data from a mHealth App for Digital Phenotyping

Presented by: Lee Kiat Kai

Contents

1. Introduction

Project objective, Project description

DIAMANTE trial

Diabetes and Mental Health Adaptive Notification Tracking and Evaluation

Depression and diabetes often co-occur together, are major causes of global disability

Increasing evidence suggesting that physical activities can help target both diseases together

Mobile applications have been found effective in helping patients engage in physical activities

Main Aim of DIAMANTE:

- To test a smartphone intervention that generates adaptive messaging, uniform random messaging intervention and control condition
- Primary aim: Improvements in physical activities at 6-month follow-up defined by daily steps counts

Project Objective

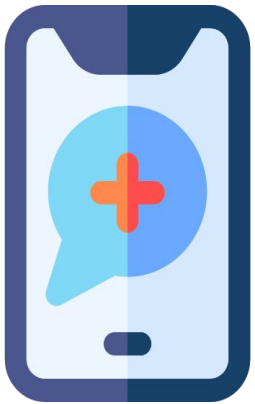
🎯 **Aim:**

To compare smartphone intervention that generates adaptive messaging and uniform messaging from the mobile health application DIAMANTE and identify which intervention is more effective in increasing daily walking

- Find clusters of similar characteristics between the different patients
- Find behavioural patterns within each of these clusters
- Find potential digital phenotypes that could be used for the personalisation of intervention.

Digital Phenotyping

Digital phenotyping is the moment-by-moment in situ quantification of the individual-level human phenotype using data from personal digital devices



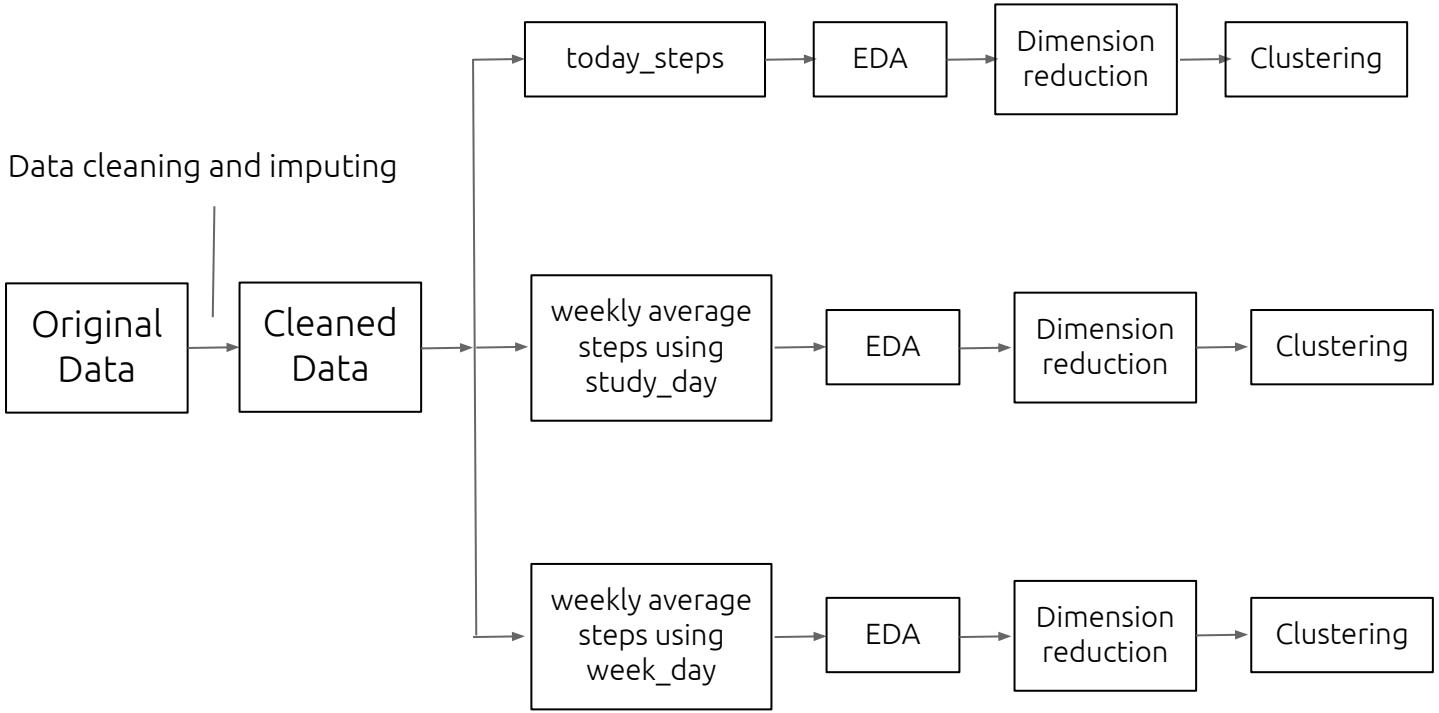
Advantages:

- Allows for a cost-efficient intervention method
- Allows for continuous measurement of human behaviours for prolonged period of time
- Allows for personalisation of individual interventions

Digital Phenotyping

- Fairly new term introduced in 2016
- Many papers only explain how digital phenotyping could be used in future clinical settings and its huge potential for patient health
- However, only few papers have developed an actual framework for the use of digital phenotyping in the area of walking steps

Framework


today_steps
EDA
Dimension reduction
Clustering
Data cleaning and imputing
Original Data
Cleaned Data
weekly average steps using study_day
EDA
Dimension reduction
Clustering
weekly average steps using week_day
EDA
Dimension reduction
Clustering

2. Exploratory Descriptive Analysis

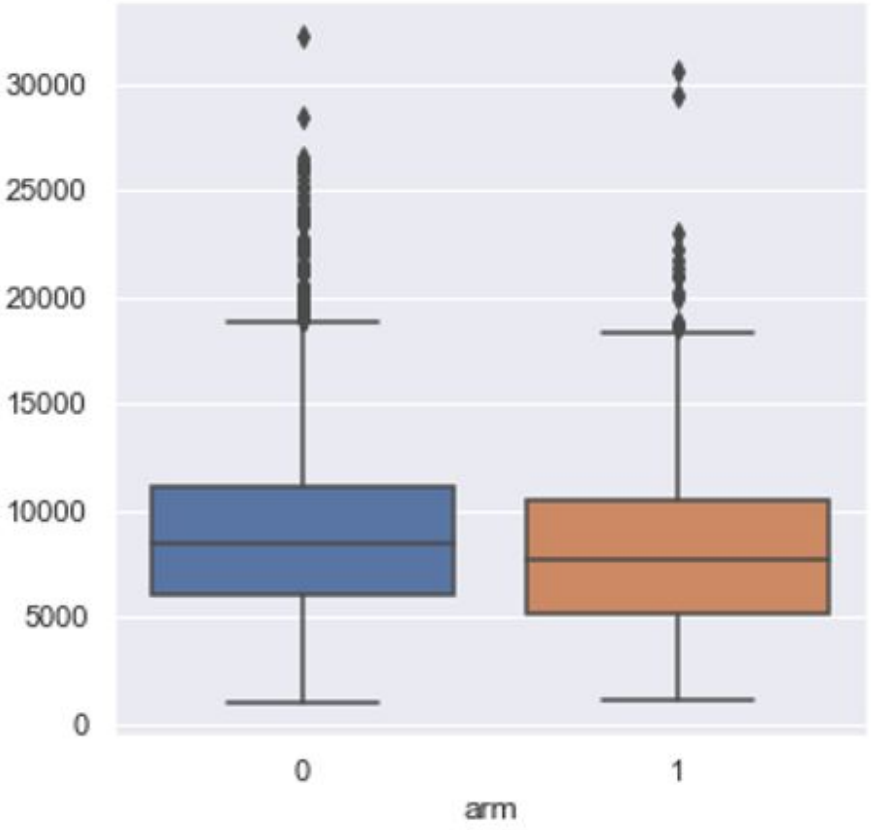
Background, features and response metrics

DIAMANTE Study Dataset

- Original dataset:
  - 3770 rows, 121 columns
  - 83 numerical variables, 38 categorical variables
  - 29 columns with missing features
  - Total of 84 participants aged between 18 to 22 years old

- Data Imputing:
  - Imputed the dataset on variable *today_steps* using a running mean of 5 days to reduce the number of missing steps
  - 560 missing today_steps rows reduced to 275 missing rows

Intervention groups



Adaptive intervention arm (arm 1)

- feedback/motivational messages at selected timing based on reinforcement learning
- Mobile phone application DIAMANTE learns from the daily patient data to personalise intervention

Uniform Random Intervention arm (arm 0)

- Patients receives random feedback/motivational message at randomly selected time intervals

Categories for Feedback/Motivational/Timing

Feedback Messages

- F0: No message
- F1: Reaching goal
- F2: Steps walked yesterday
- F3: Walked more/less today than yesterday
- F4: Steps walked yesterday, plus a positive/negative motivational message

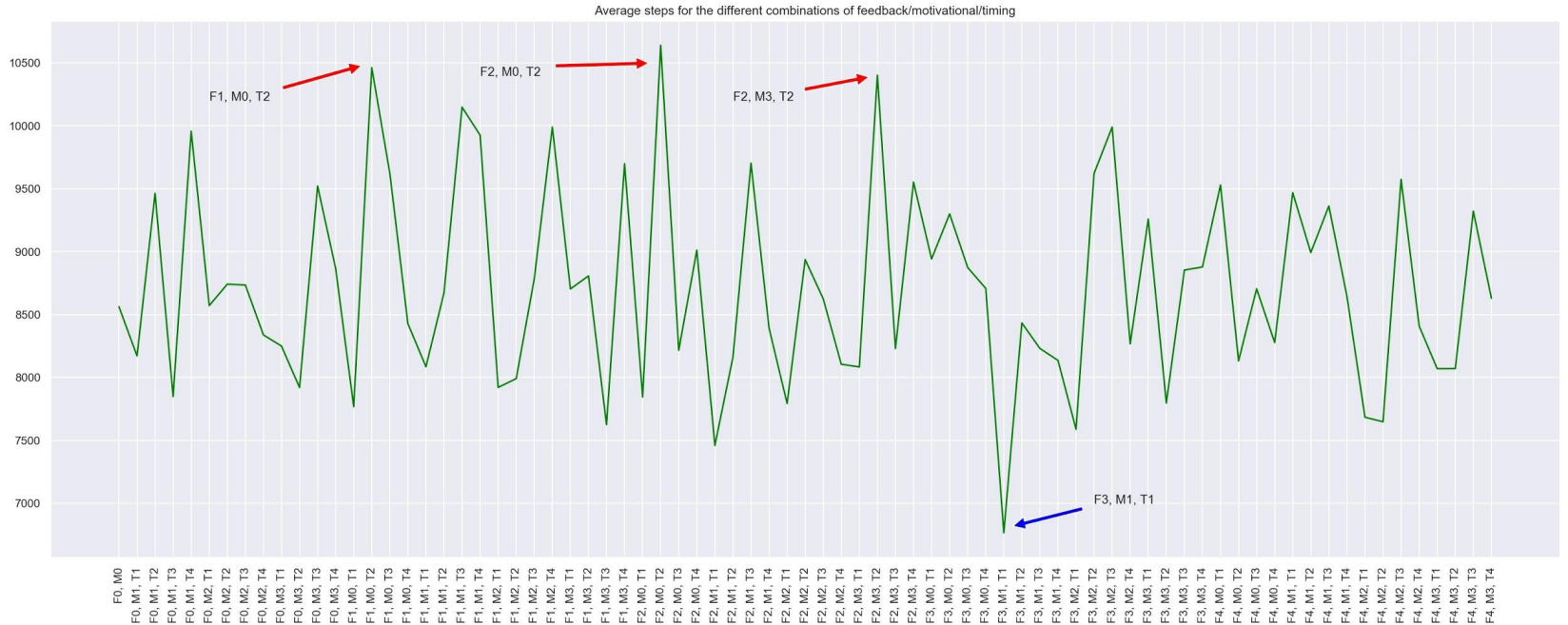
Motivational Messages

- M0: No message
- M1: Capability
- M2: Motivation
- M3: Opportunity

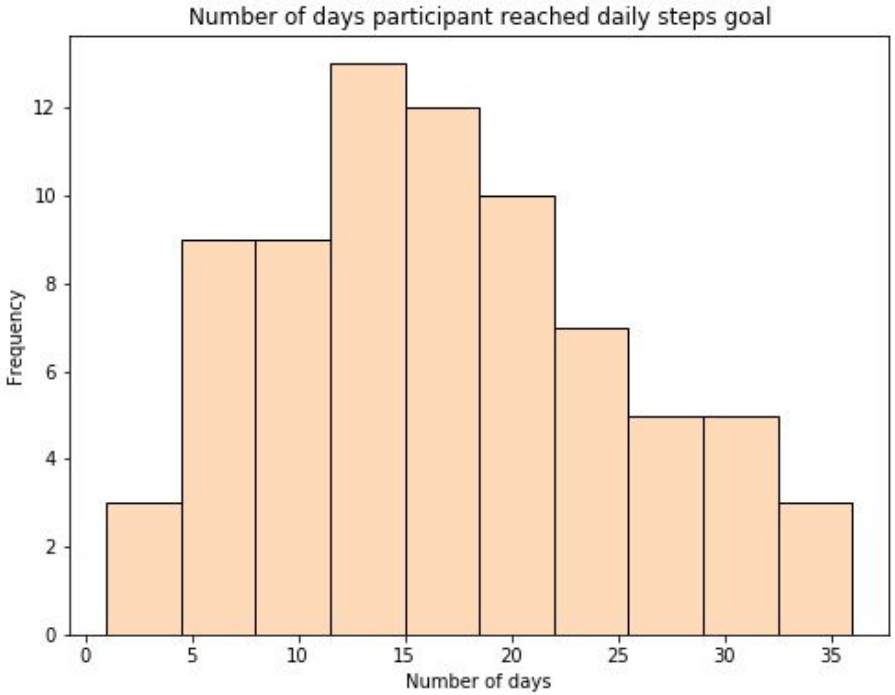
Time Period of the day

- T0: No message
- T1: 09:00 - 11:30
- T2: 11:30 - 14:00
- T3: 14:00 - 16:30
- T4: 16:30 - 19:00

Average steps for each combination of feedback/motivational/timing



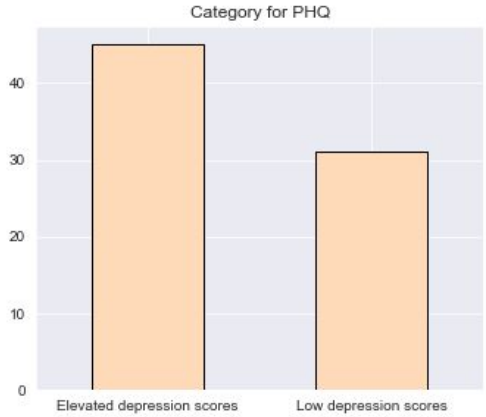
Number of days a participant reached their daily step goals

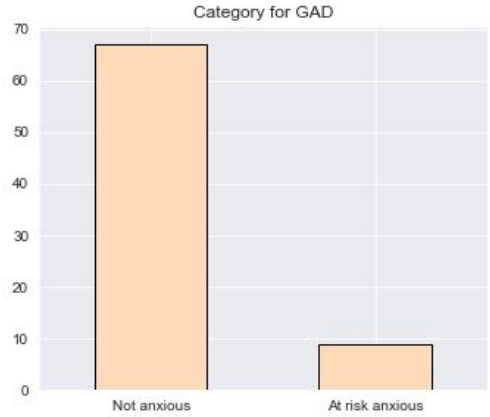


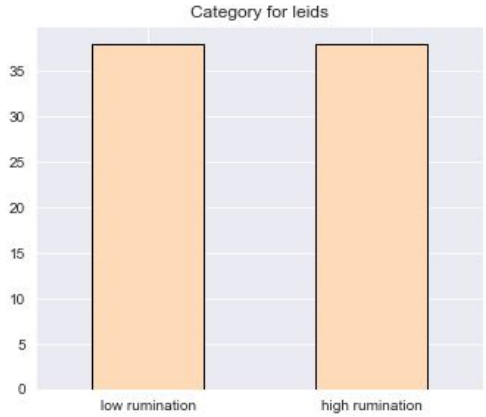
Observations:

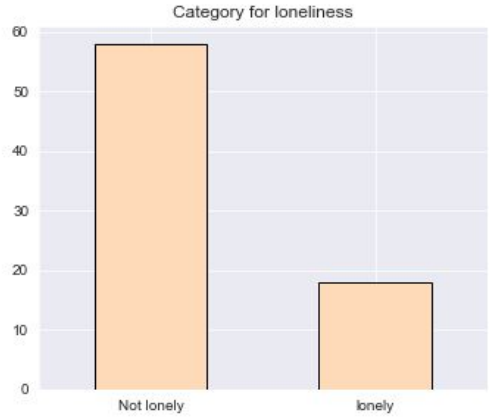
- Majority of the participants reached their daily step goals of about 11 to 15 days
- Only a minority few managed to reached their target daily steps throughout the entire duration of their study

Health Variables









Observations:

- Majority of the participants recruited have elevated depression scores, are not anxious, and are not lonely
- There is equal number of participants how have low and high rumination

Response metrics

today_steps

- Original variable in the dataset
- Dataset:
  - 3164 rows
  - 41 columns

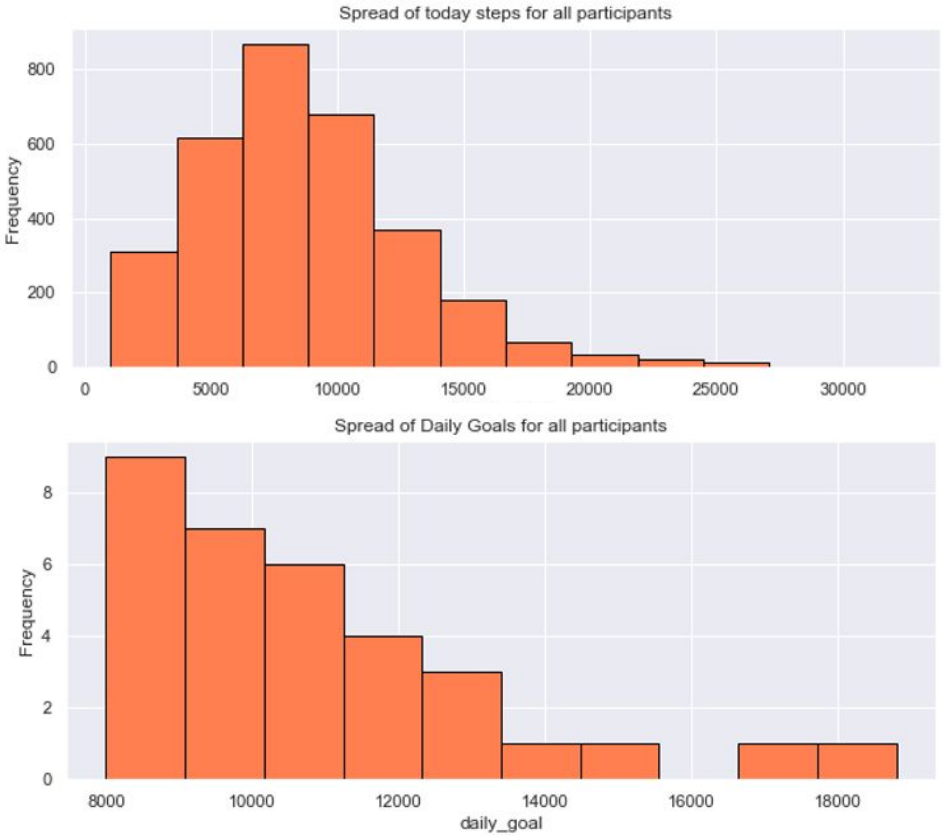
weekly average steps using study_day

- Aggregated using study_day
- E.g. day 1 to day 7 is week 1, day 8 to day 14 is week 2
- Dataset:
  - 496 rows
  - 43 columns

weekly average steps using week_day

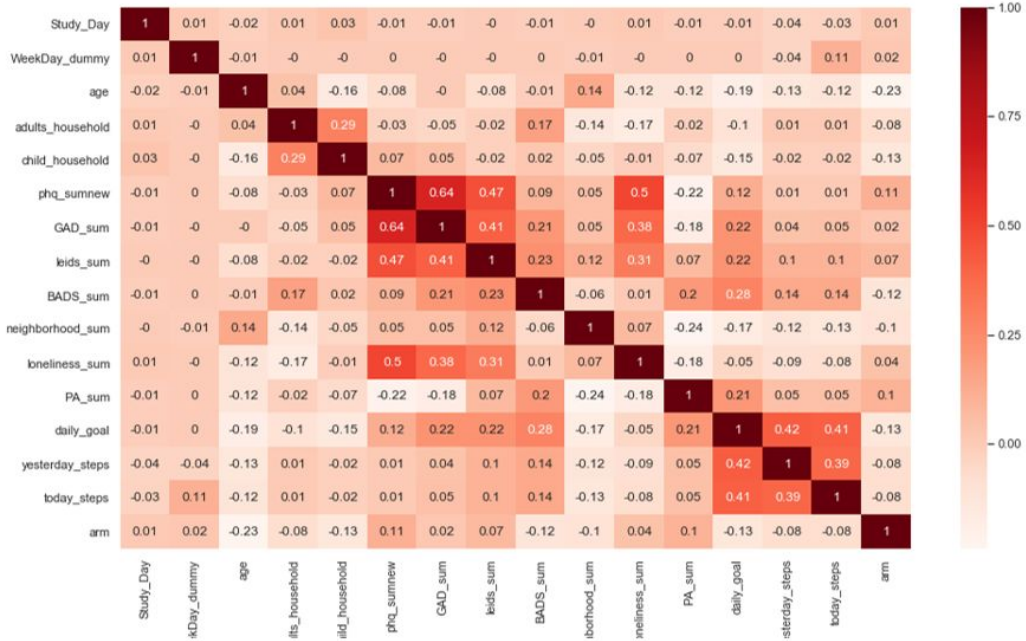
- Aggregated using week_day
- E.g. Monday to Sunday is week 1, next Monday to Sunday is week 2
- Dataset:
  - 518 rows
  - 43 columns

today_steps



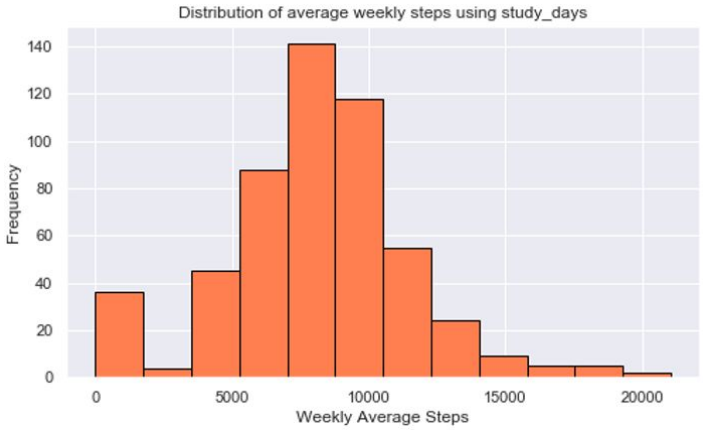
- Most participants on average walked about 7000 steps
- Only a minority of participants walked more than 10,000 steps daily
- Most participants had an average daily goal of about 8,000 steps

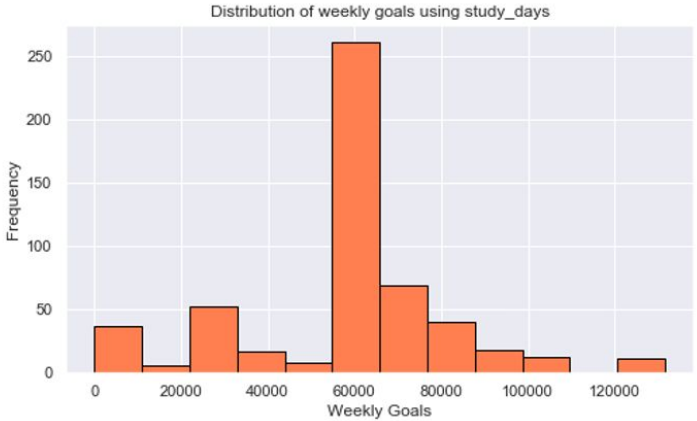
Correlation plot for today_steps



- Health variables such as loneliness score and depression score have higher correlation with each other
- Steps variables such as daily_goal, today_steps and yesterday_steps are highly correlated with each other

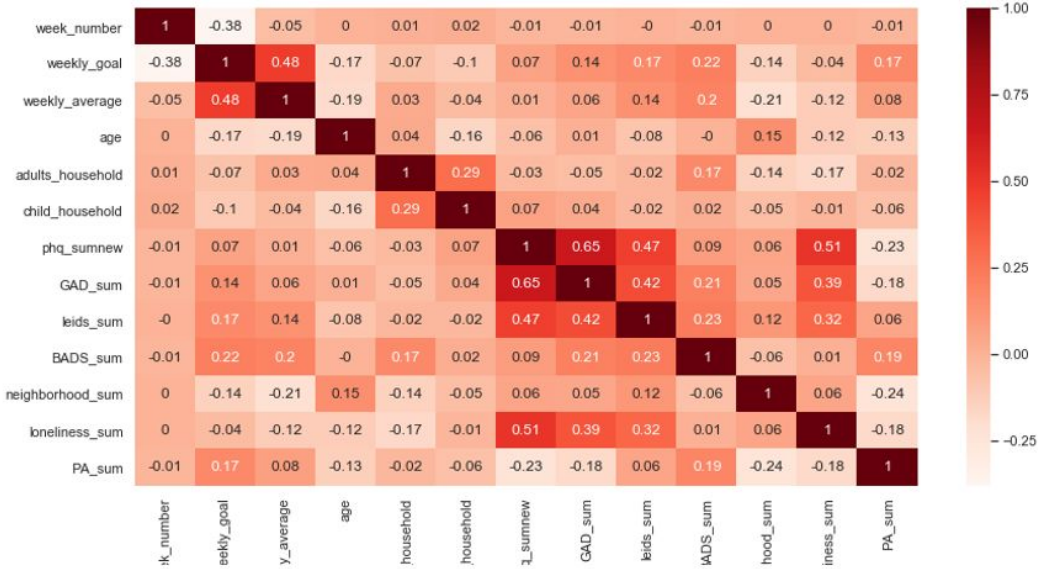
weekly average steps using study_day





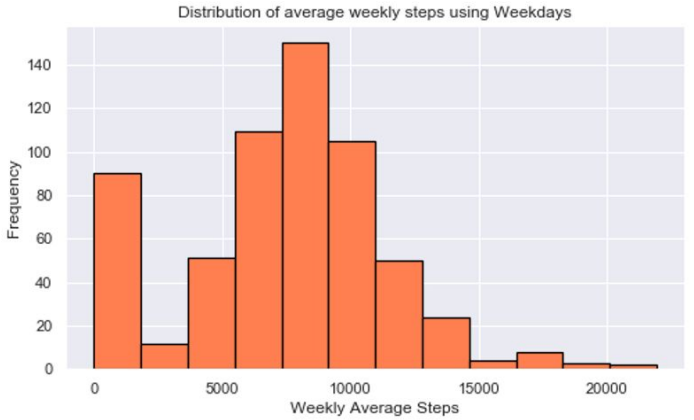
- Most participants on average walked about 8000 steps
- Only a minority of participants walked more than 12,500 steps daily
- Most participants had an average weekly goal of about 60,000 steps

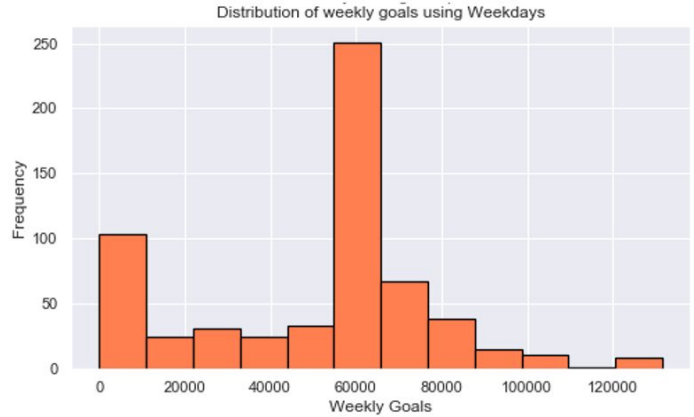
Correlation plot for weekly average steps using study_day



- Health variables such as loneliness score and depression score have higher correlation with each other
- Steps variables such as weekly_goal weekly_average are highly correlated with each other

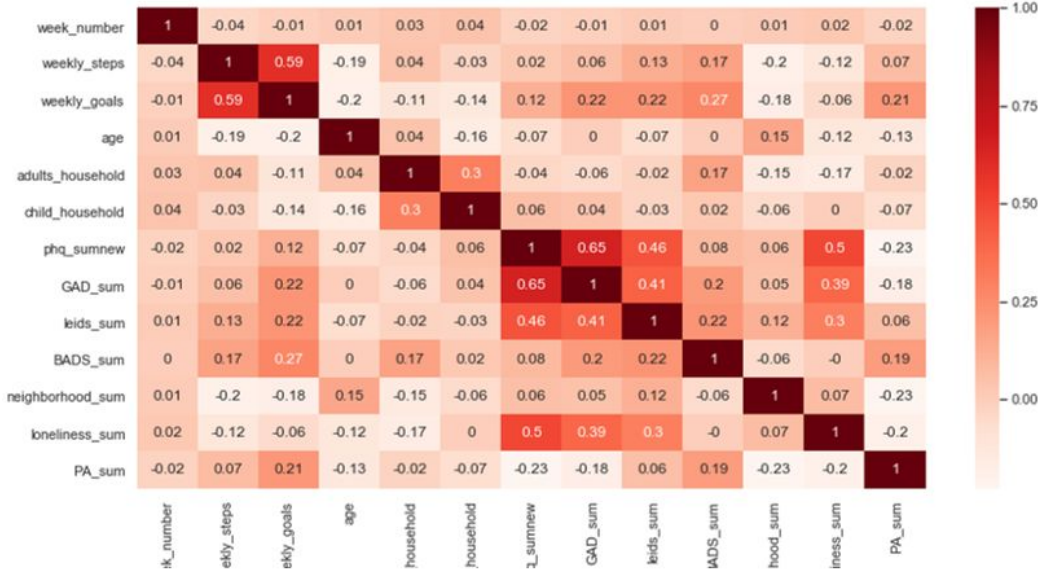
weekly average steps using week_day





- Most participants on average walked about 8000 steps
- Only a minority of participants walked more than 12,500 steps daily
- Most participants had an average weekly goal of about 60,000 steps

Correlation plot for weekly average steps using week_day



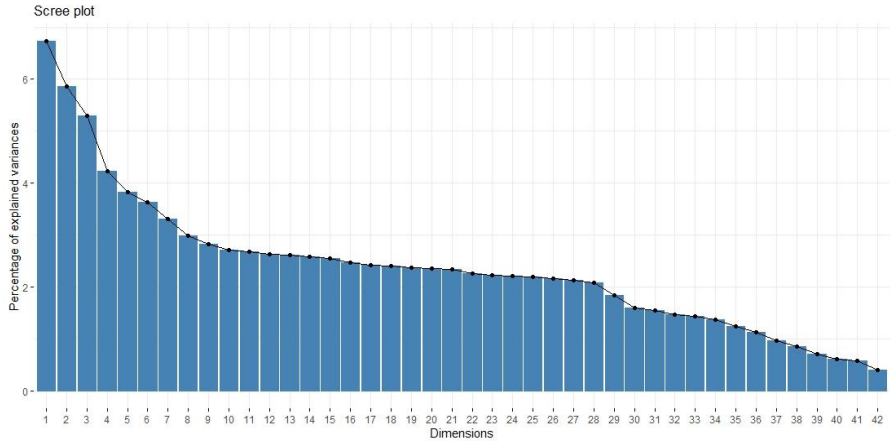
- Health variables such as loneliness score and depression score have higher correlation with each other
- Steps variables such as weekly_goal weekly_average are highly correlated with each other

3. Clustering

Feature selection using Random Forest, Clustering using hierarchical clustering and K-Prototypes

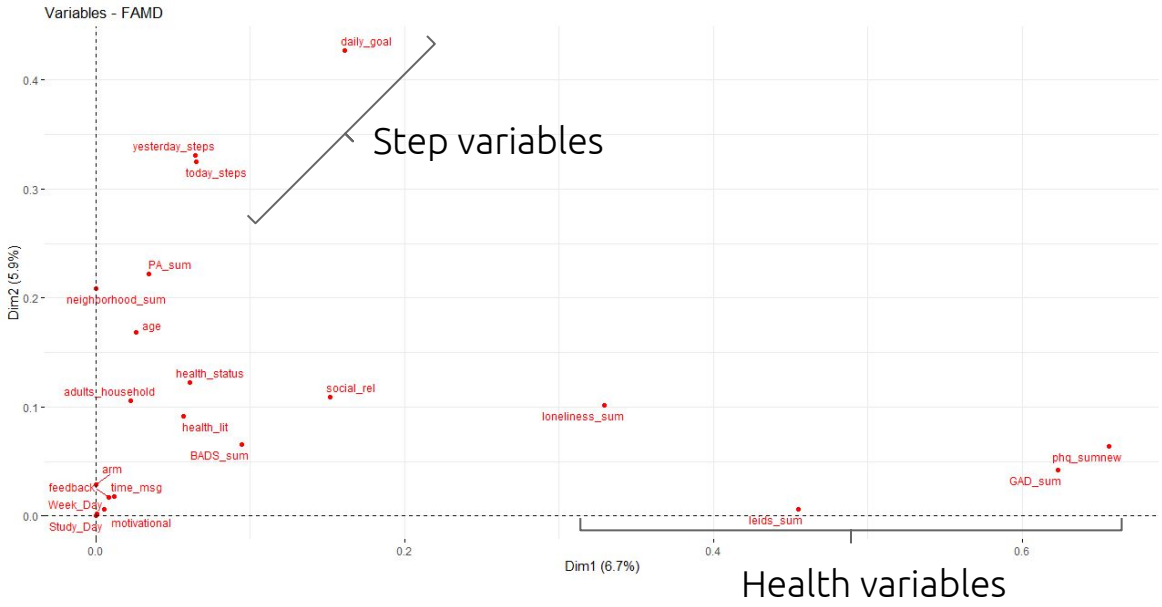
Factor Analysis of Mixed Data & Hierarchical Clustering

Factor Analysis of Mixed Data (FAMD) for today_steps



- 100% of cumulative variance explained is at 42 components
- At 28 components, the cumulative variance explains approximately 84.19% of the total variance

Hierarchical clustering for today_steps (at least 85% cumulative variance)



- Variables relating to steps contribute more towards principal dimension 2
- Variables relating to health variables such as depression score and anxiety score contributes more towards principal dimension 1.

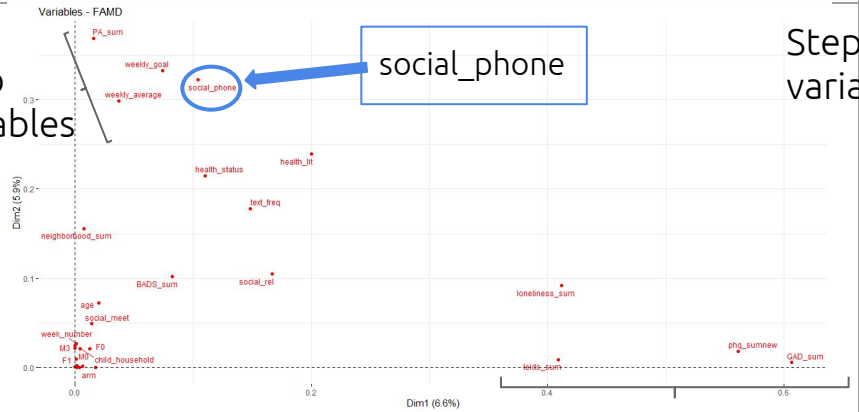
Factor Analysis of Mixed Data (FAMD)

Response metrics:	today_steps	weekly average steps using study_day	weekly average steps using week_day
Number of components that explains about 85% of total variance:	28 components 84.19%	28 components 85.29%	27 components 84.57%

Factor Analysis of Mixed Data (FAMD)

Weekly average steps using study_day

Step variables



social_phone

Health variables

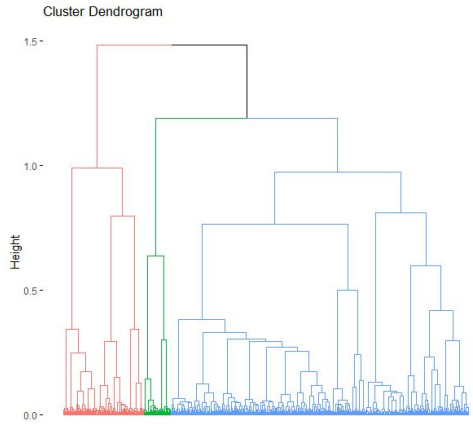
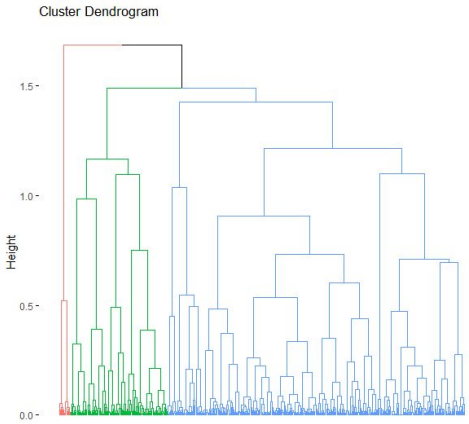
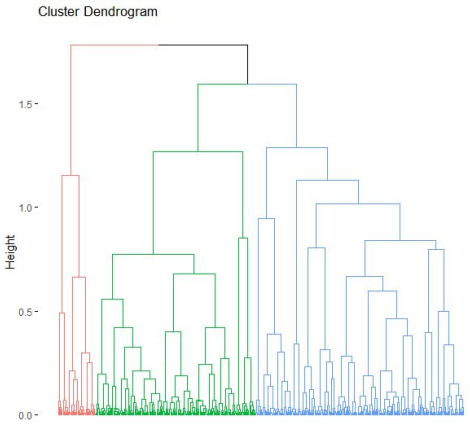
Weekly average steps using week_day

Step variables

social_meet

Health variables

Hierarchical clustering

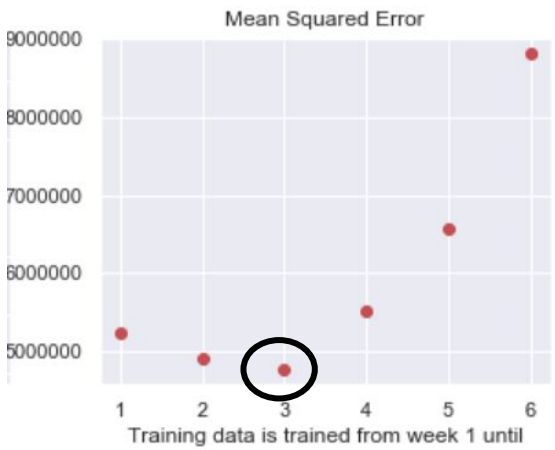
today_steps	weekly average steps using study_day	weekly average steps using week_day
		

Hierarchical clustering

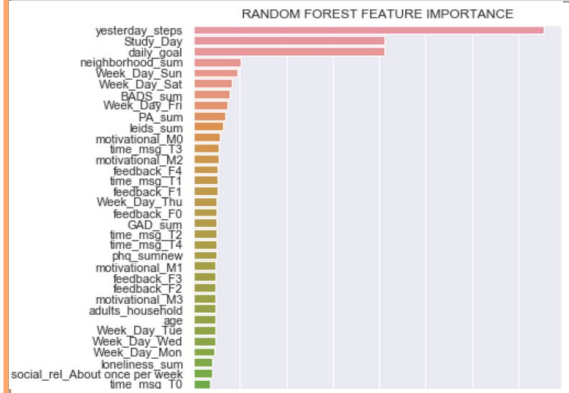
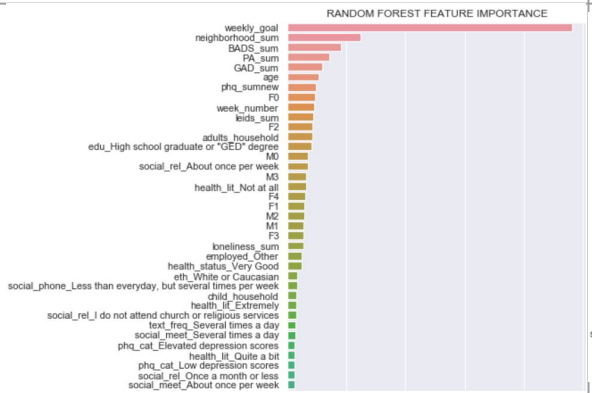
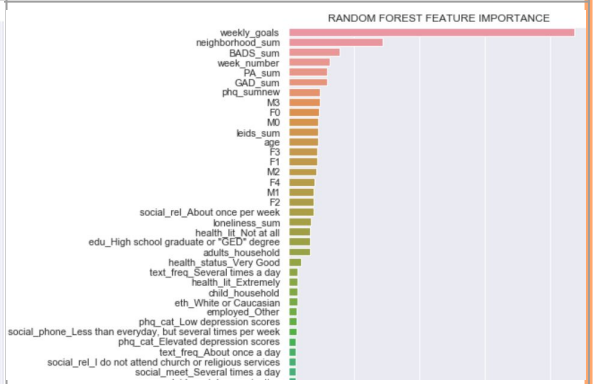
today_steps	weekly average steps using study_day	weekly average steps using week_day
Patients in cluster 3 have elevated depression scores, anxiety and are lonely - Have an average physical activity score of 169 minutes per week only, lower than the average of 183 minutes - Attend religious services about once per week	Patients in cluster 2 have elevated depression scores, anxiety and are lonely - Have average physical activities of only 132 minutes each week, lower than average of 175 minutes each week - Attend religious services about once a month or less	Patients in cluster 3 have elevated depression scores, anxiety and are lonely - Have an average physical activities of 188 minutes each week, higher than average of 180 minutes each week - Attend religious services several times per week

Random Forest & K-Prototypes

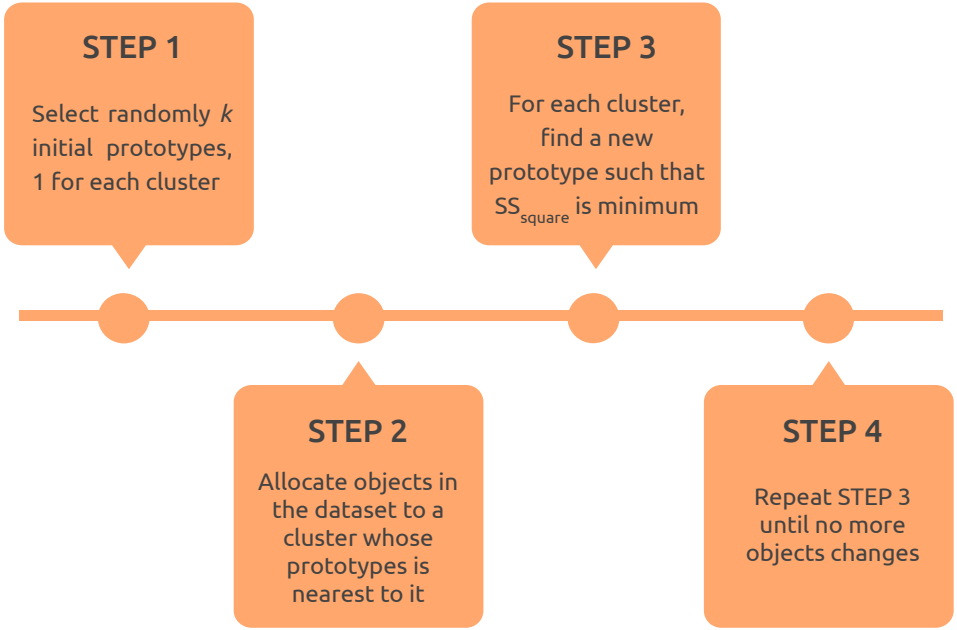
Random Forest

today_steps	weekly average steps using study_day	weekly average steps using week_day
		

Random Forest

today_steps	weekly average steps using study_day	weekly average steps using week_day
		
• 21 variables selected	• 29 variables selected	• 29 variables selected

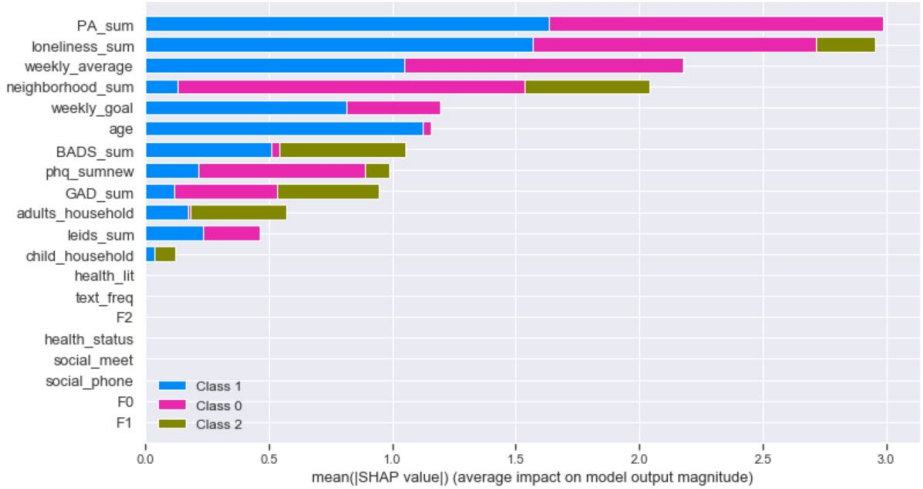
K-Prototypes



K-Prototypes for weekly average steps using study_day (For arm 0)

Number of clusters	F1 score
2	0.790272
3	0.879707
4	0.773807
5	0.830003
6	0.821258
7	0.772569
8	0.769600
9	0.690353
10	0.775129
11	0.799146
12	0.787004
13	0.795706
14	0.788145
15	0.870861

- Optimal number of clusters: 3 clusters
- F1 score: 0.880



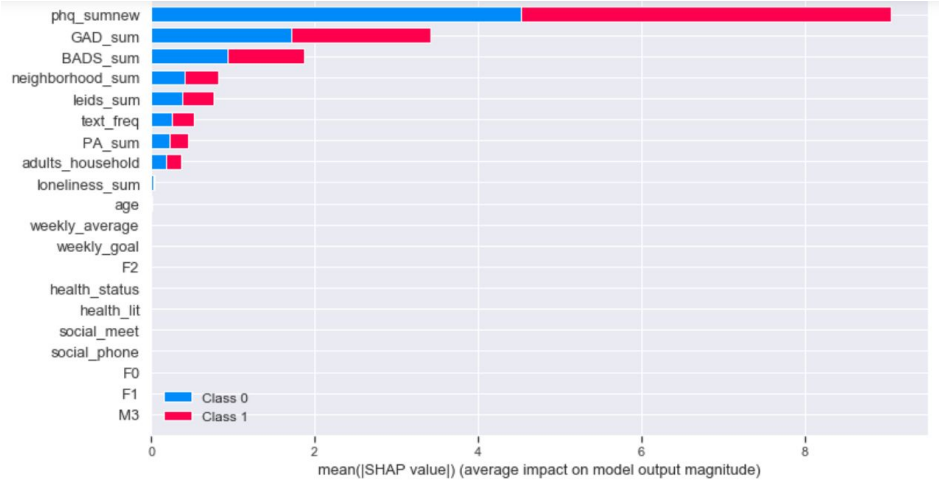
Observations:

- Duration of physical activities each week and loneliness score are the top two most important factors for clustering
- Only cluster 0 comprises of participants with elevated depression scores, anxiety, and are lonely. They have low physical activities of about 81 minutes per week as compared to at least 180 minutes each week for participants with no conditions

K-Prototypes for weekly average steps using study_day (For arm 1)

Number of clusters	F1 score
2	0.969597
3	0.804735
4	0.808309
5	0.718842
6	0.897326
7	0.963768
8	0.910769
9	0.756361
10	0.947546
11	0.834799
12	0.931484
13	0.906886

- Optimal number of clusters: 2 clusters
- F1 score: 0.970



Observations:

- Depression score and anxiety score are the top two most important factors for clustering
- Only cluster 1 comprises of participants with elevated depression scores, anxiety, and are lonely. They have lower physical activities of about 175 minutes per week as compared to 202 minutes for participants with no conditions

K-Prototypes Results (For arm 0)

Response metrics:	weekly average steps using *study_day*	weekly average steps using *week_day*
Number of clusters:	**3**	**5**
Observations:	Patients with elevated depression scores, anxiety and are lonely - have the lowest weekly average steps - lowest physical activities sum	Patients with elevated depression scores, anxiety and are lonely - have the highest weekly average steps - higher physical activities sum
	- Call less than everyday but several times a week - Have meeting once a week	- Call about once a day - Have meeting once a week

K-Prototypes Results (For arm 1)

Response metrics:	weekly average steps using *study_day*	weekly average steps using *week_day*
Number of clusters:	**2**	**2**
Observations:	Patients with elevated depression scores, anxiety and are lonely - have a higher weekly average steps - lower physical activities sum compared to patients without those conditions	
	- Call several times a day - Have social meeting once a day	- Call several times a day - Have social meeting once a day

Comparison of results

Response metrics:	Weekly average steps using study_day	Weekly average steps using week_day
K-Prototypes	F1 for arm 0: 0.880 F1 for arm 1: 0.970	F1 for arm 0: 0.886 F1 for arm 1: 0.551

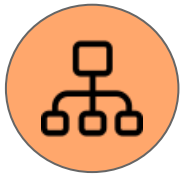
Performance of walking steps:	7,000 ~ 8,000 steps	5,000 ~ 10,000 steps
Average duration of physical activities:	175 ~ 215 minutes	75 ~ 215 minutes

Conclusion

Overall Digital Phenotypes

For patients with:

- **Elevated depression scores** 
- **with anxiety** 
- **are lonely** 



Hierarchical Clustering

1. Have fewer social meetings of about once a week
2. Text and call less than everyday, but several times per week
3. Attend social religious activities regularly at least several times each week

Overall Digital Phenotypes for K-Prototypes

ARM 0: UNIFORM MESSAGING

1. Have fewer social meetings of about once per week
2. text several times a day and call about once a day
3. Do not attend any social religious services at all

ARM 1: ADAPTIVE MESSAGING

1. Have frequent social meetings of about once a day
2. Text and call several times a day
3. Do not attend any social religious services at all

Digital Phenotyping

- Cost-efficient smartphone intervention
- Approximately half of the world population owns a smartphone device
- Developing a framework using digital phenotyping to cater to individuals' needs, allows for significant progression towards reducing risk of depression & diabetes by increasing physical activities

Thank you!