Digital Phenotyping in DIAMANTE

FYP PRESENTATION 22/12/20

Objectives

1) To test if there is an increased in step counts in the adaptive messaging arm group compared to the uniform messaging arm group

2) To find clusters of similar characteristics and find behavioral pattern within these clusters

Contents

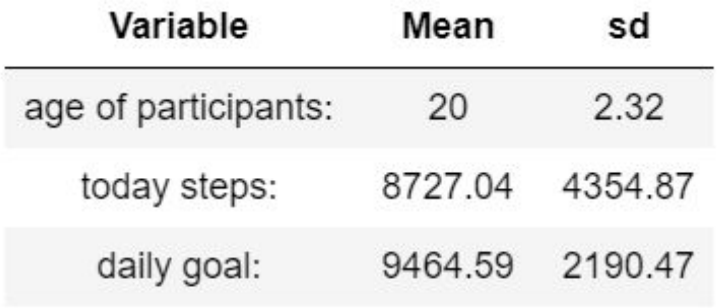
1 Exploratory Descriptive Analysis

2 Multilevel Statistical Model

1. Exploratory Descriptive Analysis

Original Data

- 3770 rows, 121 columns
- Total of 84 participants
- Each participant have a maximum study duration of 45 days and minimum 41 days

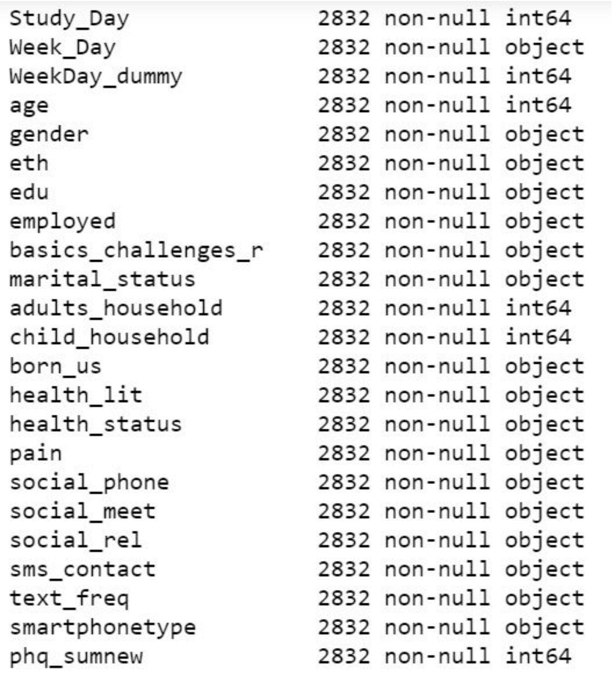


Variable	Mean	sd
age of participants:	20	2.32
today steps:	8727.04	4354.87
daily goal:	9464.59	2190.47

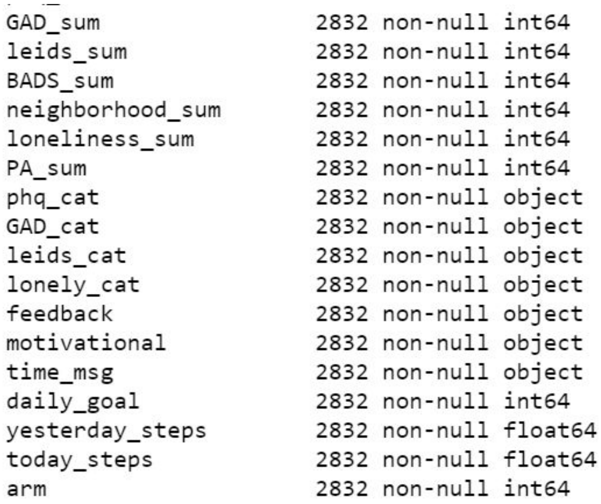
Data Cleaning

- Removed all columns with missing values
- Removed rows with missing value for today_steps
- Replaced NA’s in variable time_msg with T0 to represent that they did not receive any message at all
- Removed outliers for variables time_msg, today_steps, eth, gender
- Removed 78 variables, left with 43 columns
- <u>Cleaned data</u>: 2832 rows, 42 columns

Remaining variables in cleaned data

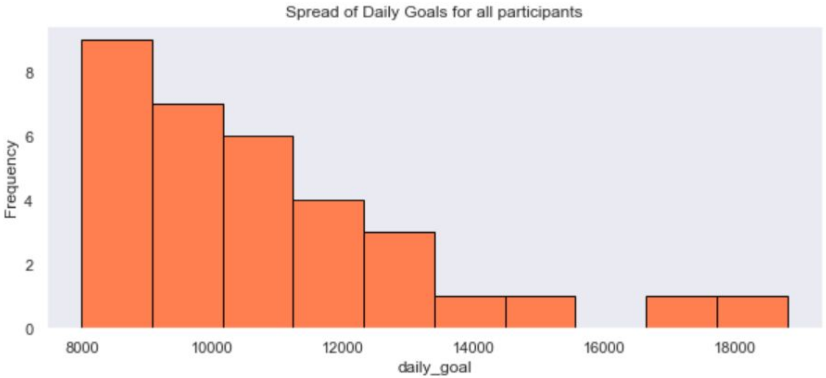


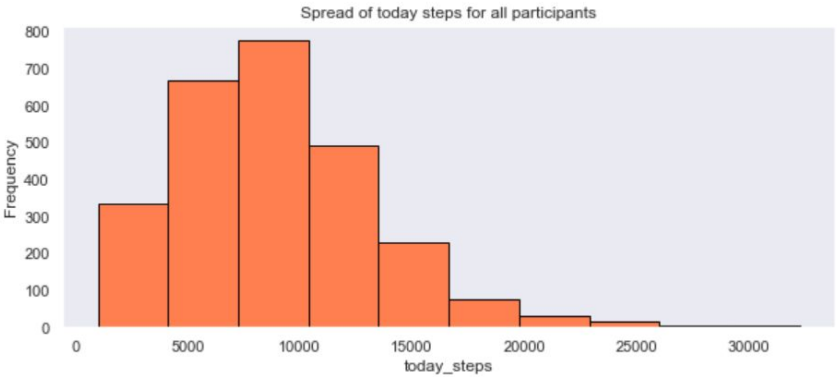
```
Study_Day              2832 non-null int64
Week_Day               2832 non-null object
WeekDay_dummy          2832 non-null int64
age                    2832 non-null int64
gender                 2832 non-null object
eth                    2832 non-null object
edu                    2832 non-null object
employed               2832 non-null object
basics_challenges_r    2832 non-null object
marital_status         2832 non-null object
adults_household       2832 non-null int64
child_household        2832 non-null int64
born_us                2832 non-null object
health_lit             2832 non-null object
health_status          2832 non-null object
pain                   2832 non-null object
social_phone           2832 non-null object
social_meet            2832 non-null object
social_rel             2832 non-null object
sms_contact            2832 non-null object
text_freq              2832 non-null object
smartphonetype         2832 non-null object
phq_sumnew             2832 non-null int64
```



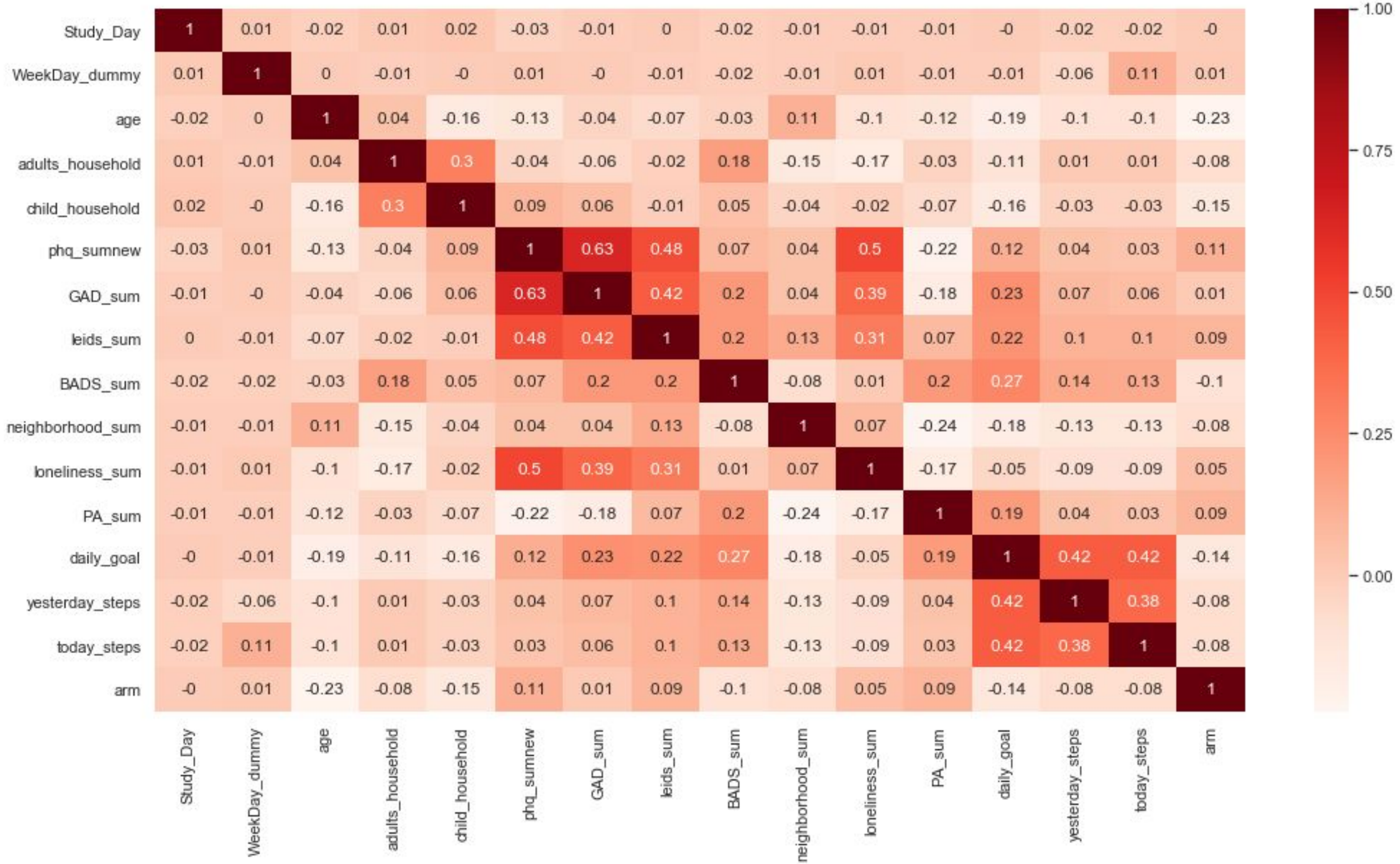
```
GAD_sum                2832 non-null int64
leids_sum              2832 non-null int64
BADS_sum               2832 non-null int64
neighborhood_sum       2832 non-null int64
loneliness_sum         2832 non-null int64
PA_sum                 2832 non-null int64
phq_cat                2832 non-null object
GAD_cat                2832 non-null object
leids_cat              2832 non-null object
lonely_cat             2832 non-null object
feedback               2832 non-null object
motivational           2832 non-null object
time_msg               2832 non-null object
daily_goal             2832 non-null int64
yesterday_steps        2832 non-null float64
today_steps            2832 non-null float64
arm                    2832 non-null int64
```

Distribution of daily_goal and today_steps for all participants



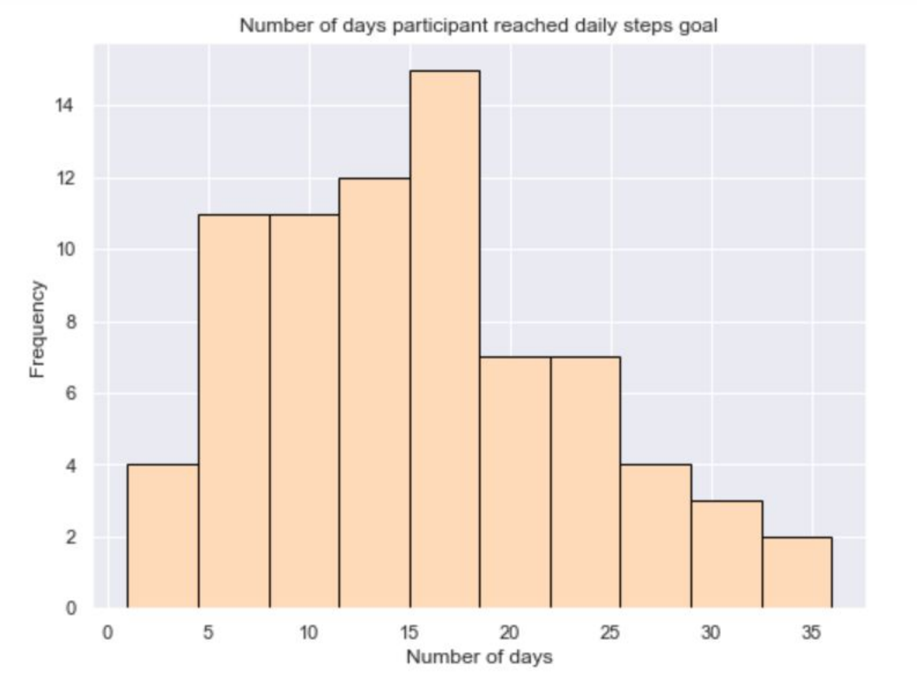


Correlation Matrix for numerical variables



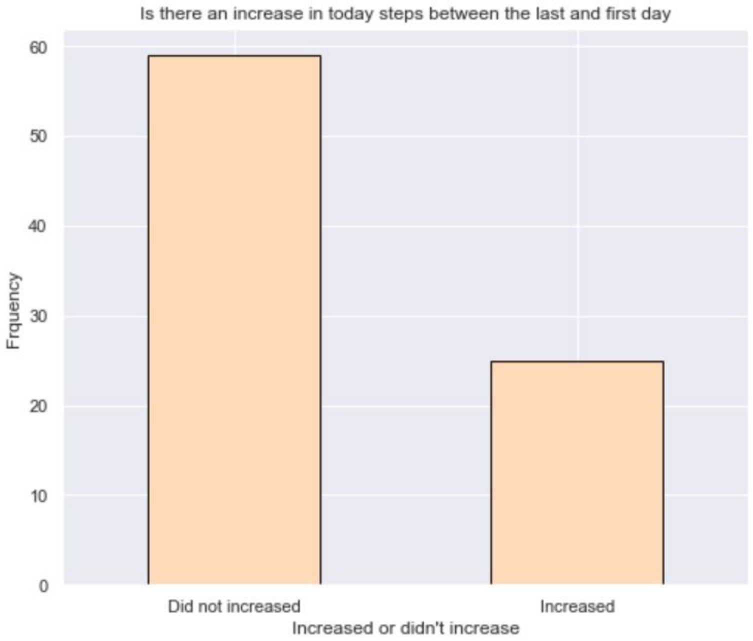
	Study_Day	WeekDay_dummy	age	adults_household	child_household	phq_sumnew	GAD_sum	leids_sum	BADS_sum	neighborhood_sum	loneliness_sum	PA_sum	daily_goal	yesterday_steps	today_steps	arm
Study_Day	1	0.01	-0.02	0.01	0.02	-0.03	-0.01	0	-0.02	-0.01	-0.01	-0.01	-0	-0.02	-0.02	-0
WeekDay_dummy	0.01	1	0	-0.01	-0	0.01	-0	-0.01	-0.02	-0.01	0.01	-0.01	-0.01	-0.06	0.11	0.01
age	-0.02	0	1	0.04	-0.16	-0.13	-0.04	-0.07	-0.03	0.11	-0.1	-0.12	-0.19	-0.1	-0.1	-0.23
adults_household	0.01	-0.01	0.04	1	0.3	-0.04	-0.06	-0.02	0.18	-0.15	-0.17	-0.03	-0.11	0.01	0.01	-0.08
child_household	0.02	-0	-0.16	0.3	1	0.09	0.06	-0.01	0.05	-0.04	-0.02	-0.07	-0.16	-0.03	-0.03	-0.15
phq_sumnew	-0.03	0.01	-0.13	-0.04	0.09	1	0.63	0.48	0.07	0.04	0.5	-0.22	0.12	0.04	0.03	0.11
GAD_sum	-0.01	-0	-0.04	-0.06	0.06	0.63	1	0.42	0.2	0.04	0.39	-0.18	0.23	0.07	0.06	0.01
leids_sum	0	-0.01	-0.07	-0.02	-0.01	0.48	0.42	1	0.2	0.13	0.31	0.07	0.22	0.1	0.1	0.09
BADS_sum	-0.02	-0.02	-0.03	0.18	0.05	0.07	0.2	0.2	1	-0.08	0.01	0.2	0.27	0.14	0.13	-0.1
neighborhood_sum	-0.01	-0.01	0.11	-0.15	-0.04	0.04	0.04	0.13	-0.08	1	0.07	-0.24	-0.18	-0.13	-0.13	-0.08
loneliness_sum	-0.01	0.01	-0.1	-0.17	-0.02	0.5	0.39	0.31	0.01	0.07	1	-0.17	-0.05	-0.09	-0.09	0.05
PA_sum	-0.01	-0.01	-0.12	-0.03	-0.07	-0.22	-0.18	0.07	0.2	-0.24	-0.17	1	0.19	0.04	0.03	0.09
daily_goal	-0	-0.01	-0.19	-0.11	-0.16	0.12	0.23	0.22	0.27	-0.18	-0.05	0.19	1	0.42	0.42	-0.14
yesterday_steps	-0.02	-0.06	-0.1	0.01	-0.03	0.04	0.07	0.1	0.14	-0.13	-0.09	0.04	0.42	1	0.38	-0.08
today_steps	-0.02	0.11	-0.1	0.01	-0.03	0.03	0.06	0.1	0.13	-0.13	-0.09	0.03	0.42	0.38	1	-0.08
arm	-0	0.01	-0.23	-0.08	-0.15	0.11	0.01	0.09	-0.1	-0.08	0.05	0.09	-0.14	-0.08	-0.08	1

Distribution of the number of days participants today_steps reached their daily goal



- Majority of the participant reached on average a total of 15 daily_goals throughout the study
- Very few participants managed to complete most days of the study duration

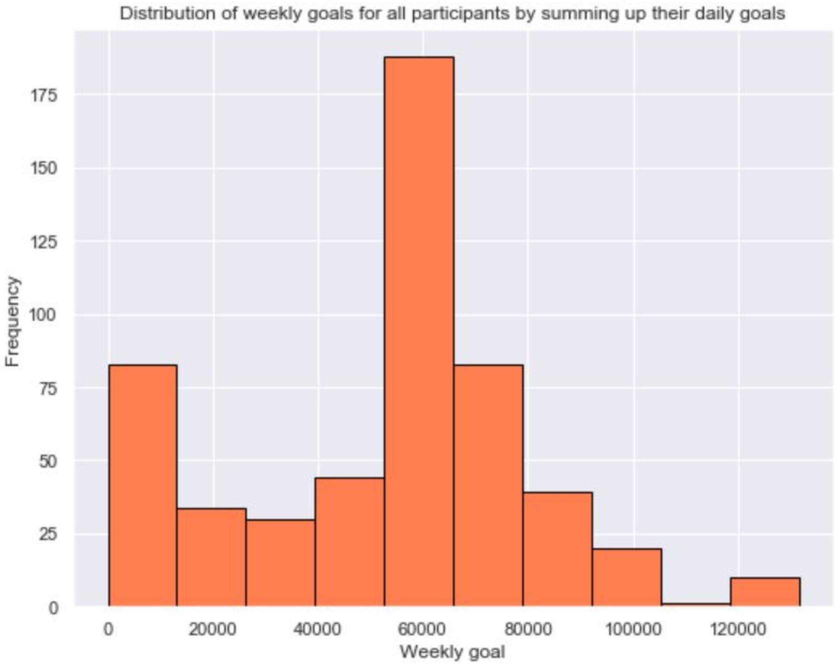
Distribution of increase/decrease in today steps on the last day compared to first day for each participants



- Majority of the participant did not have a positive increase in the today_steps on their last day as compared to the first day

- Only about less than ⅓ of the participants did increase their today steps on their last day compared to their first day

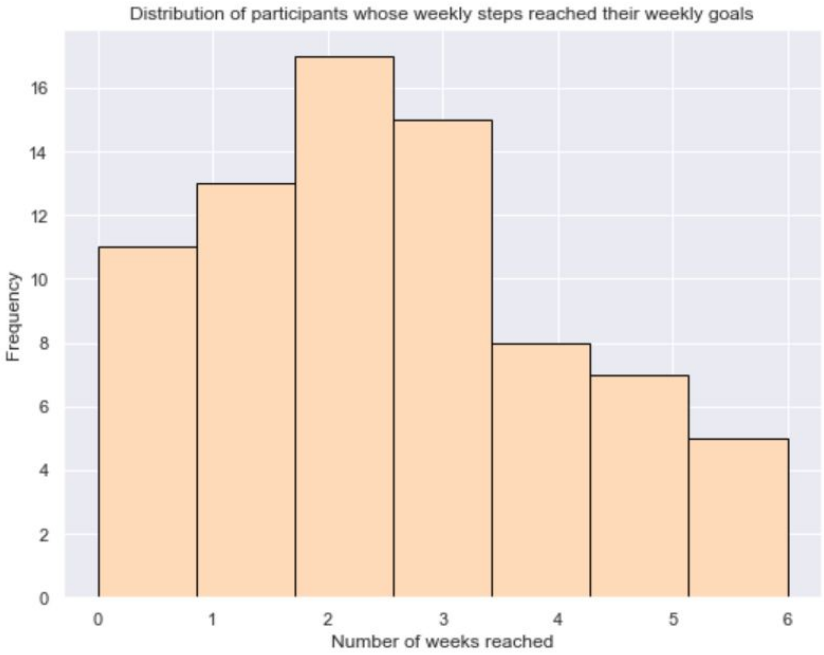
Distribution of weekly_goal for each participant



- weekly_goal is calculated by summing up their daily_goal for each week (e.g. day 1 to day 7 is week 1, day 8 to day 14 is week 2, etc)

- From the plot, majority of the participants have a weekly target goal of around 60,000 steps

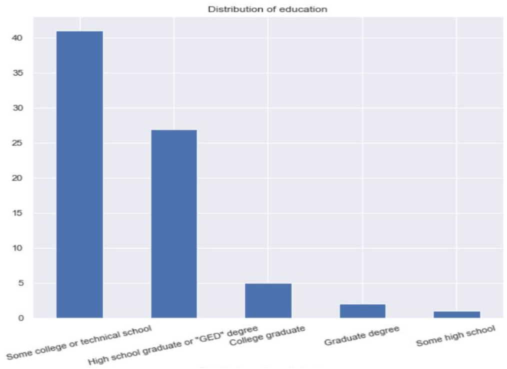
Distribution of the number of participants whose weekly steps reached their weekly goal

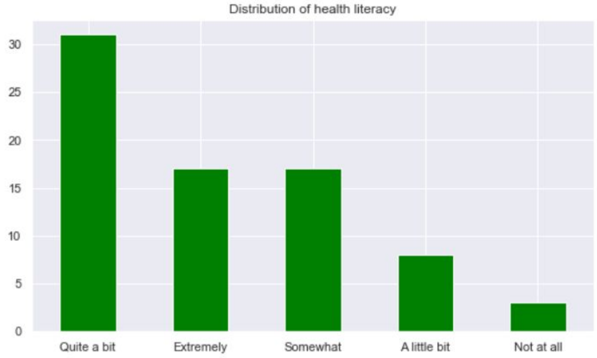


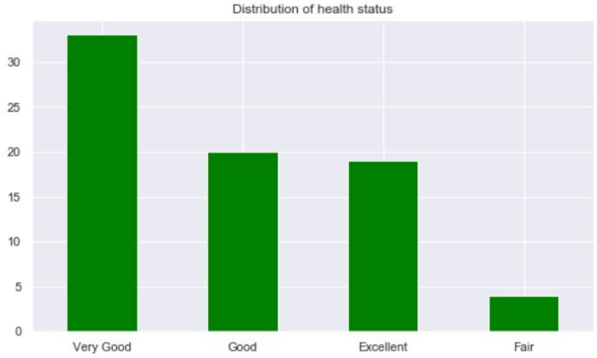
- We obtain weekly_steps in the same way as we did for weekly_goal by summing up all the today_steps for each week

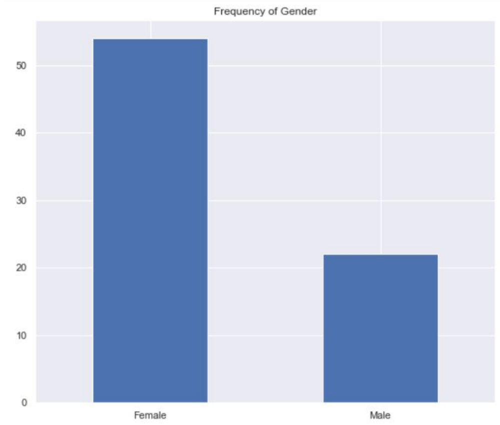
- Majority of the participants reached about 2 weeks of their targets

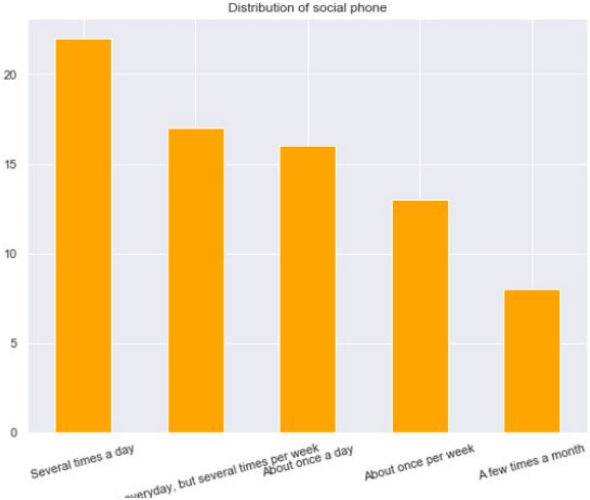
Some other descriptive analysis of other variables

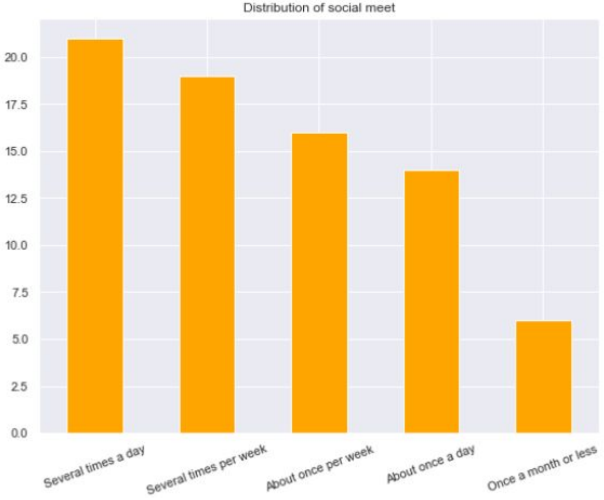

Distribution of education


Distribution of health literacy

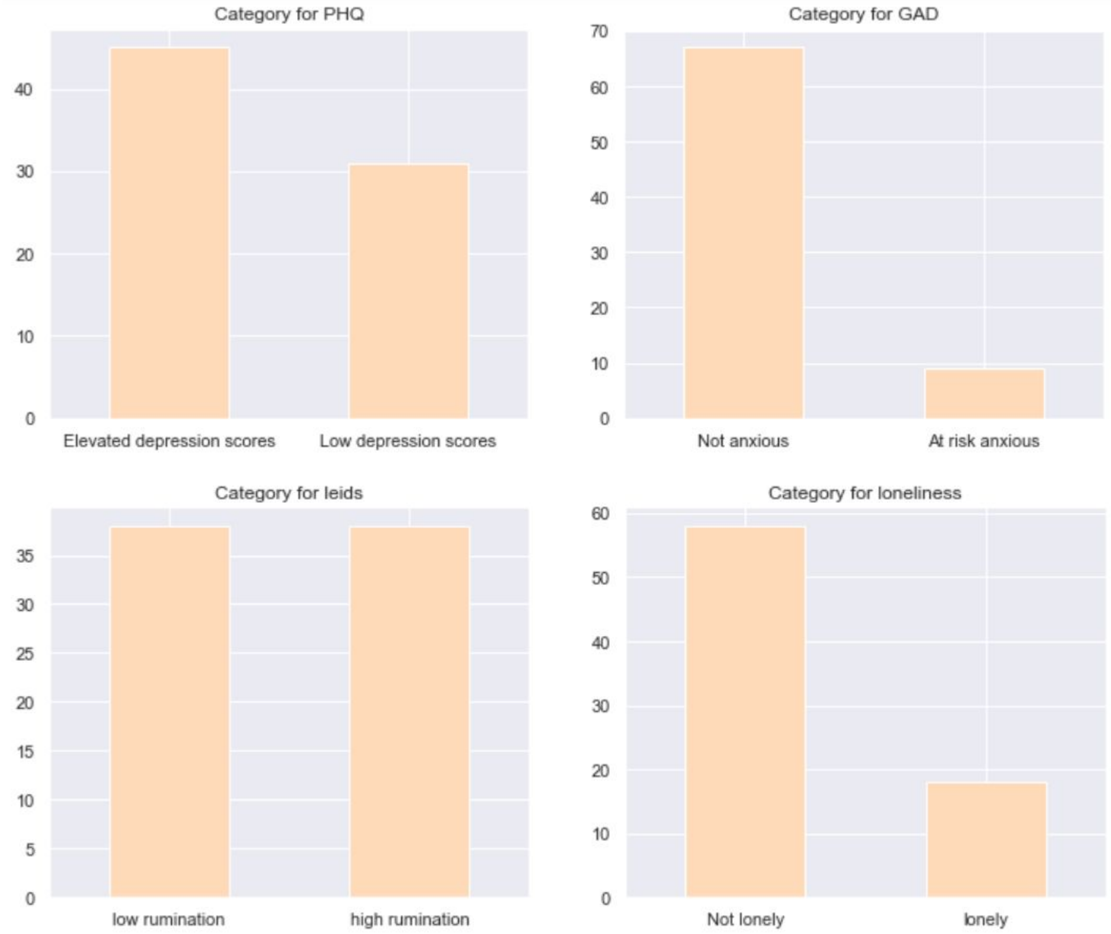

Distribution of health status


Frequency of Gender

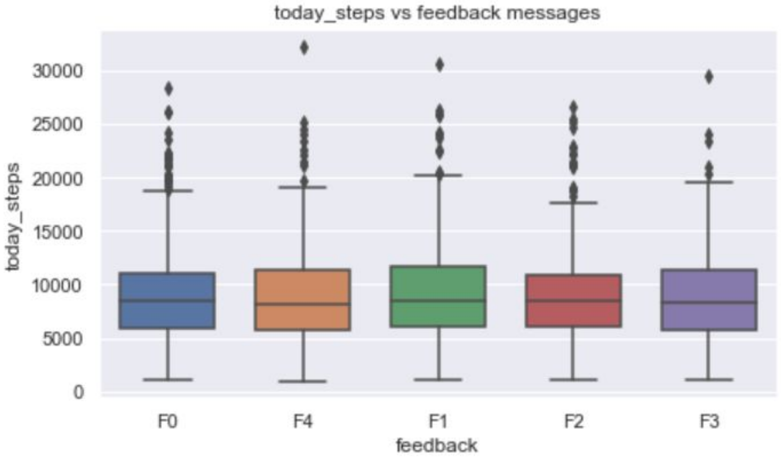

Distribution of social phone


Distribution of social meet

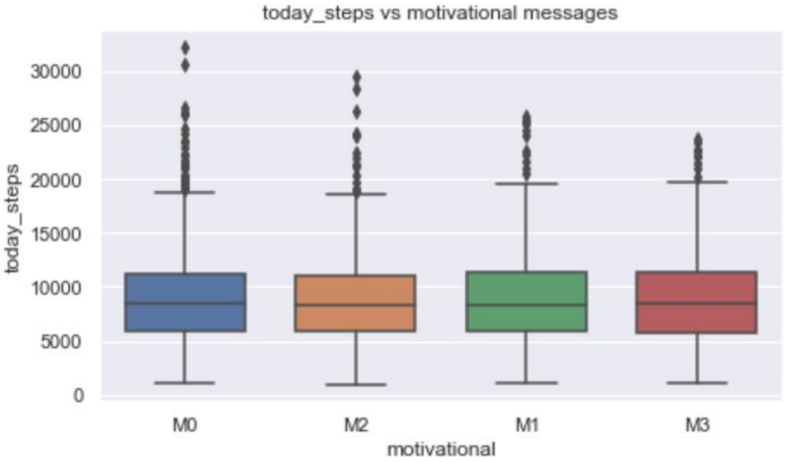
Distribution of health questionnaires variables



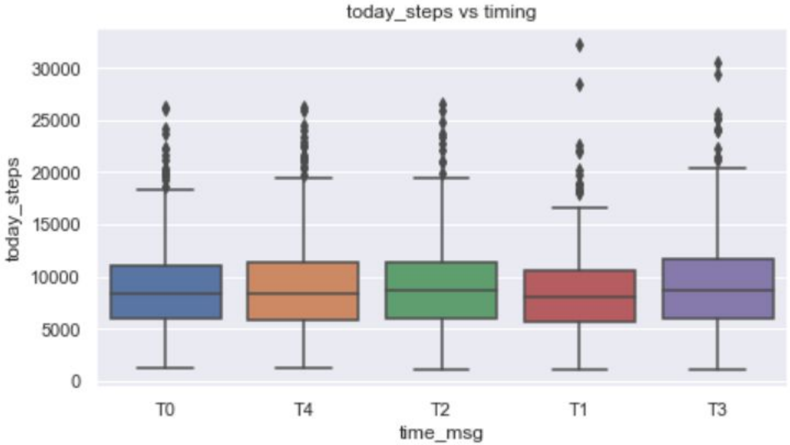
Average today steps against feedback, motivational & time_msg



today_steps vs feedback messages

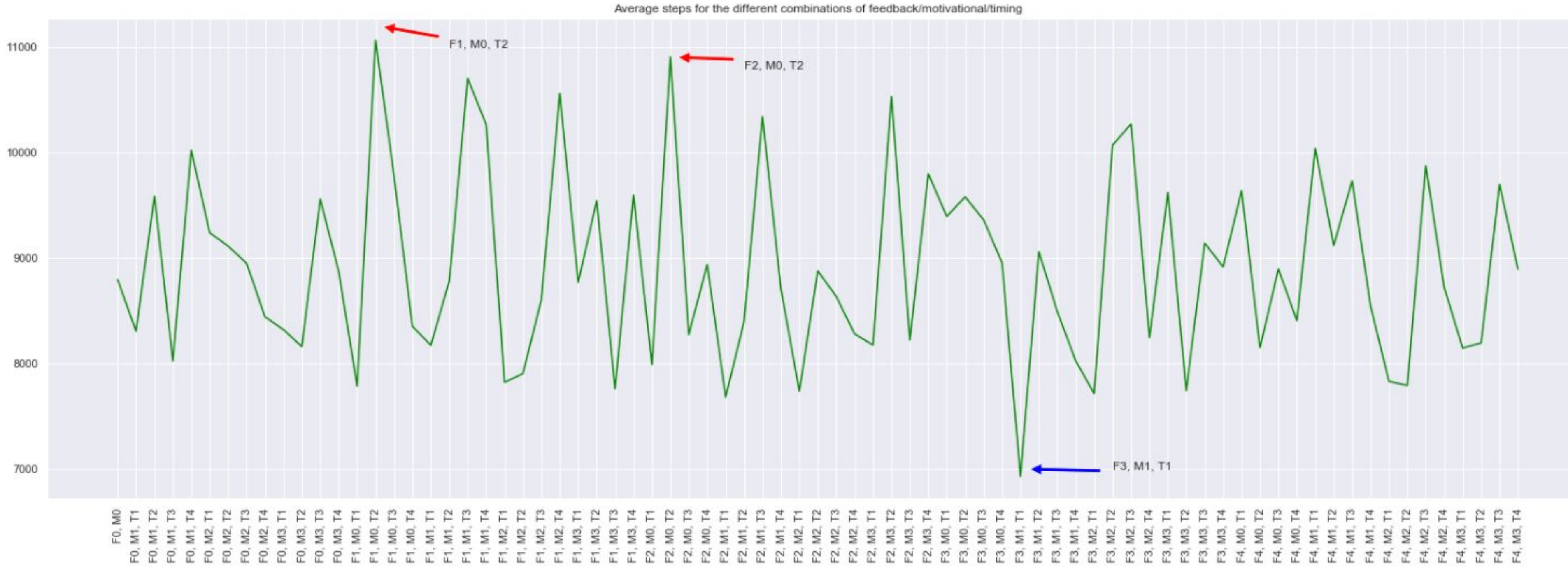


today_steps vs motivational messages

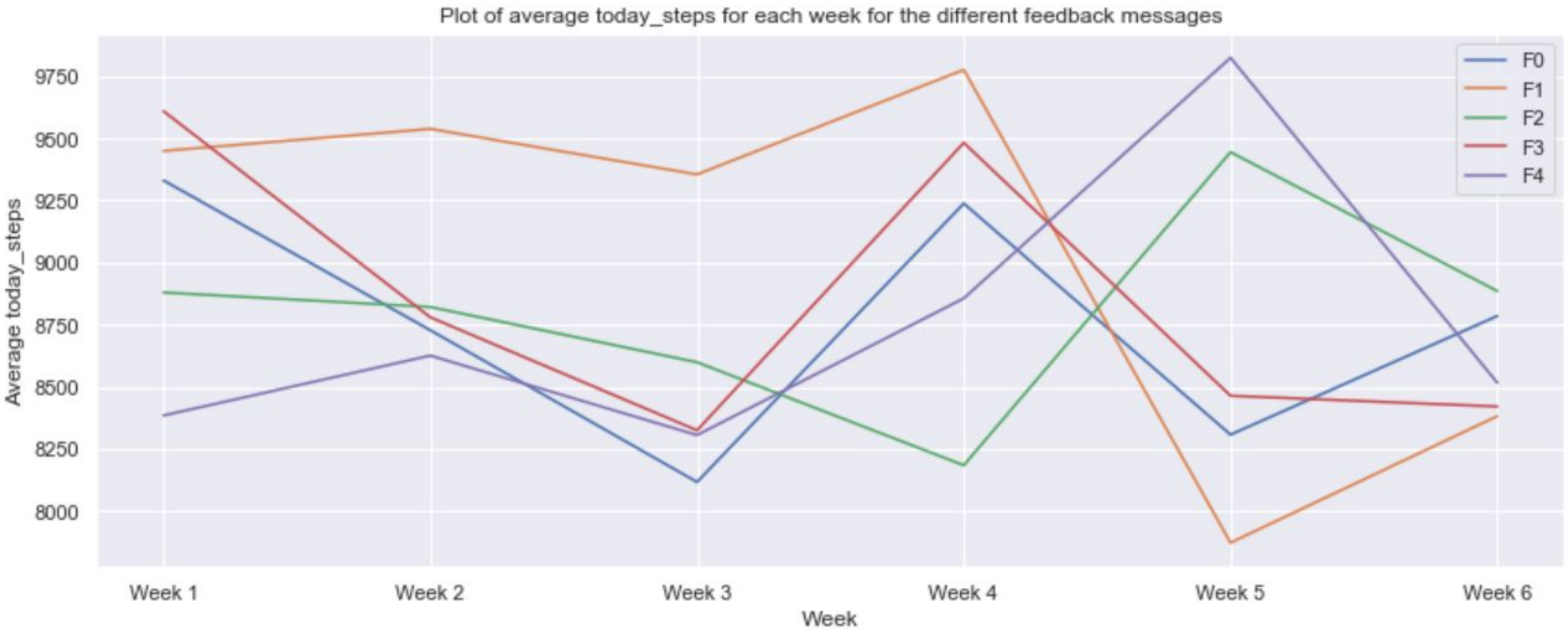


today_steps vs timing

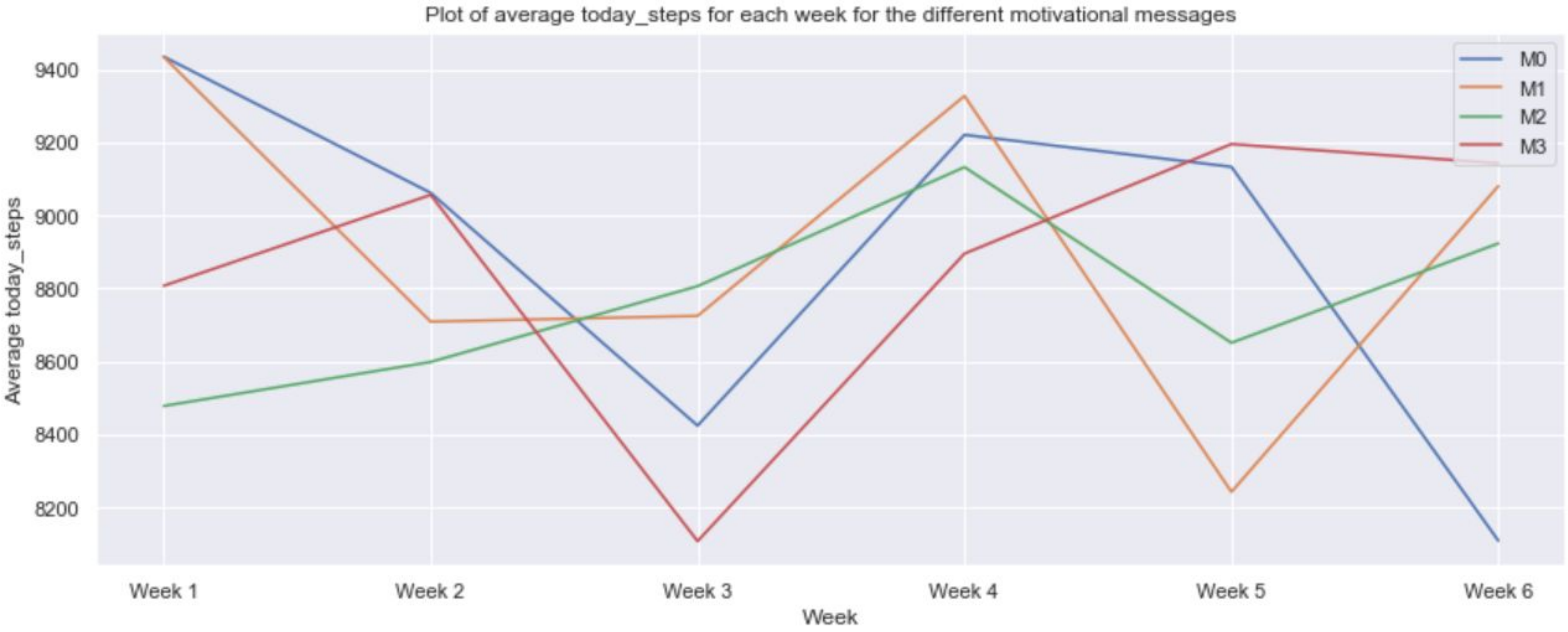
Plot of average steps for the different combinations of feedback/motivational/time_msg



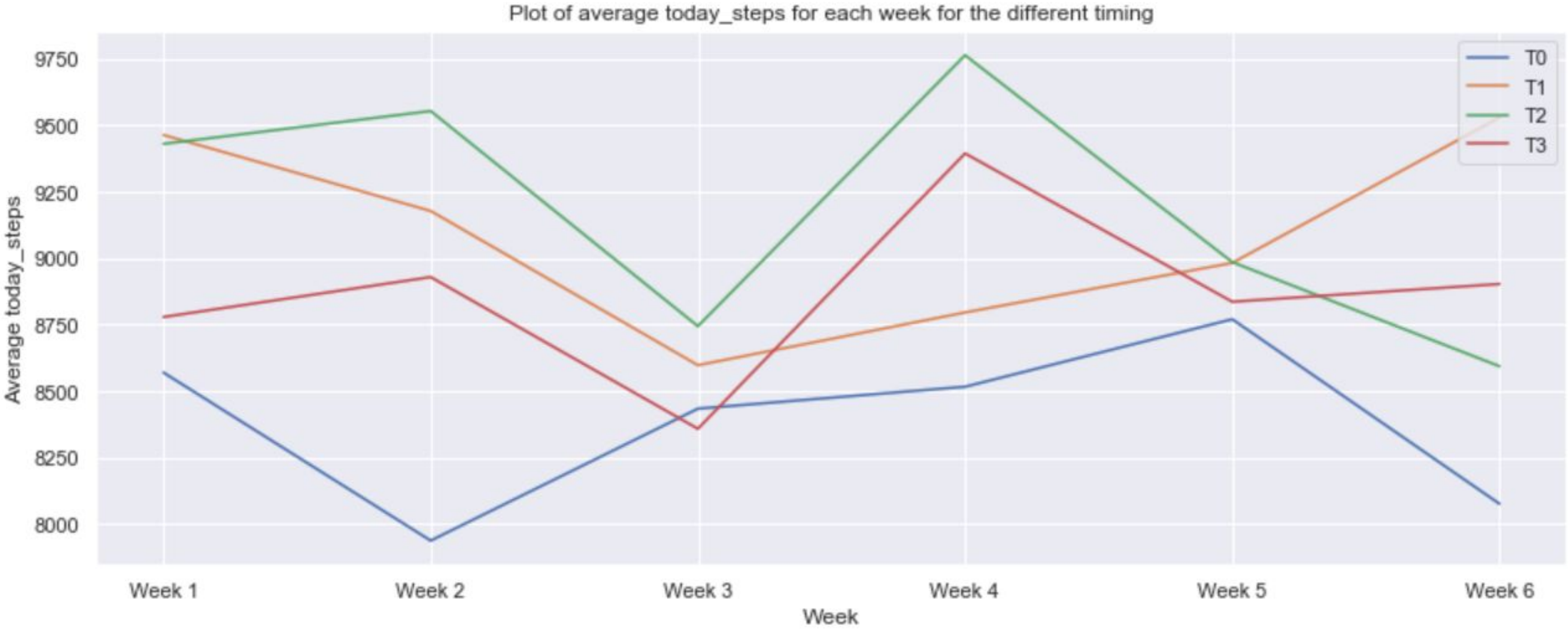
Plot of average today_steps for the different feedback messgaes against each week



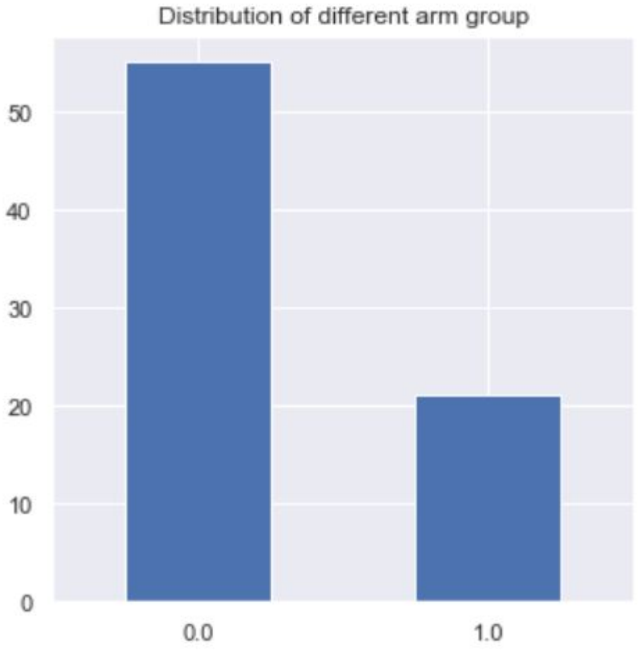
Plot of average today_steps for the different motivational messages against each week



Plot of average today_steps for the different time_msg against each week

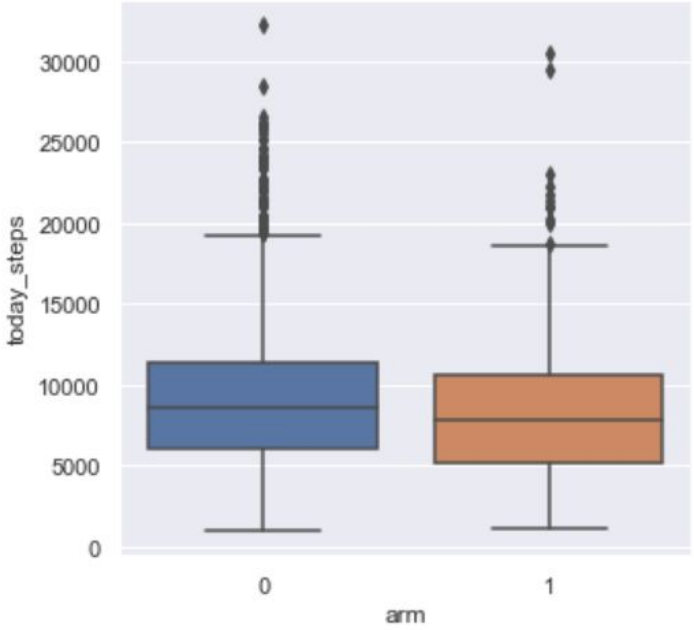


Distribution of arm group



- 55 participants belong to the group receiving uniform random messaging
- 21 participants belong to the group receiving adaptive messaging

Distribution of today_steps for different arm group



- Average today_steps in the arm 0 group (uniform messaging) is 9095.62
- Average today_steps in the arm 1 group (adaptive messaging) is 8272.42

2. Multilevel Statistical Model

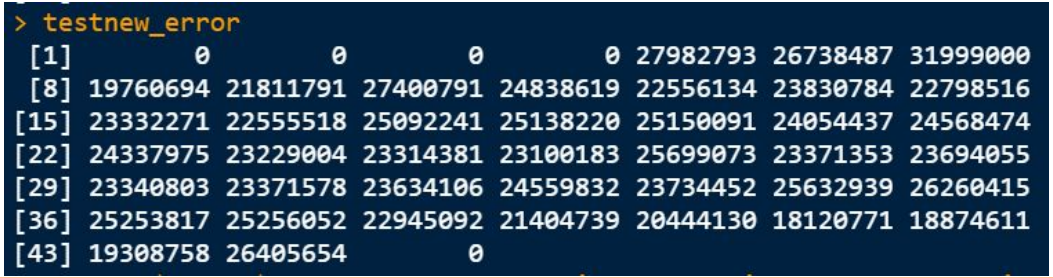
Decision Trees

Decision tree

- First, I used decision tree as they are simple to explain as it mirrors human decision-making and can be easily interpreted even by someone of no domain knowledge.

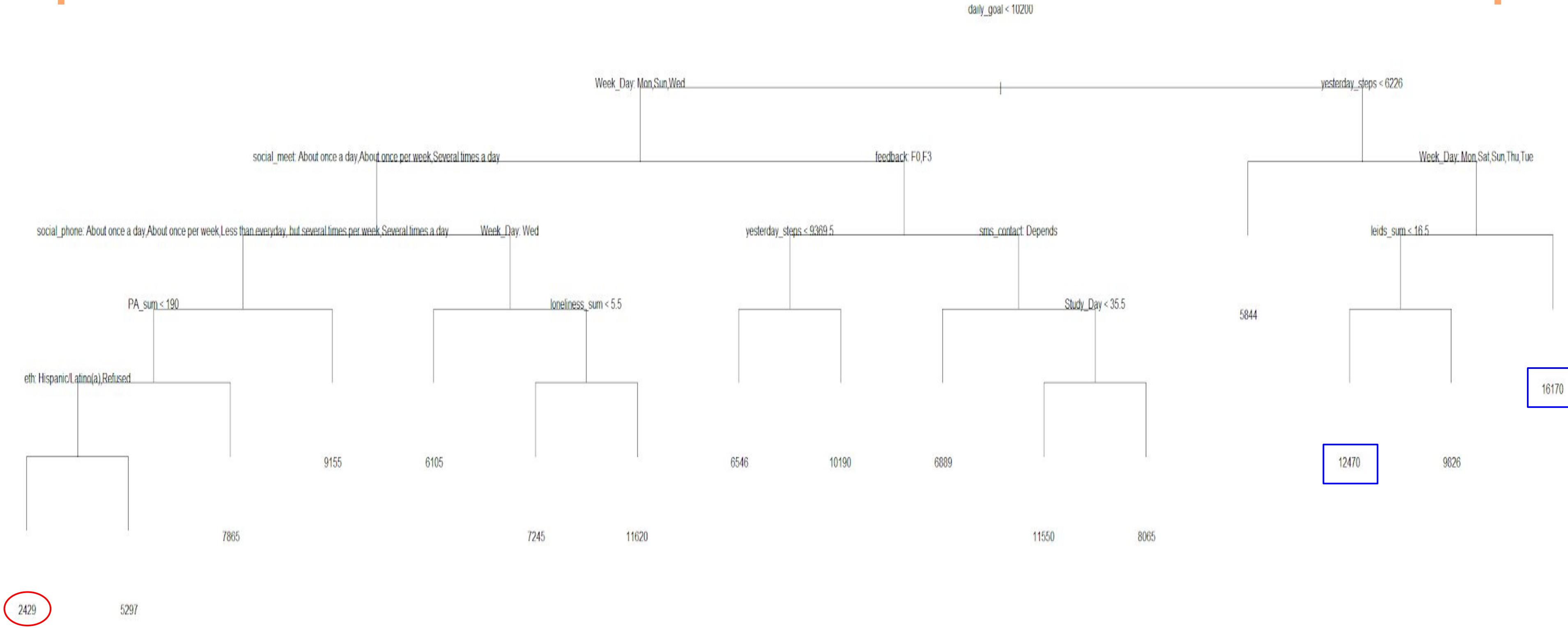
- Trained the model using different Study_Day from day 1 to day t and tested the model from day t+1 to day 45 where t is 5 to 43

- minimum test error of 18120771 is obtained when training on day 1 to 40 and test on day 41 to 45

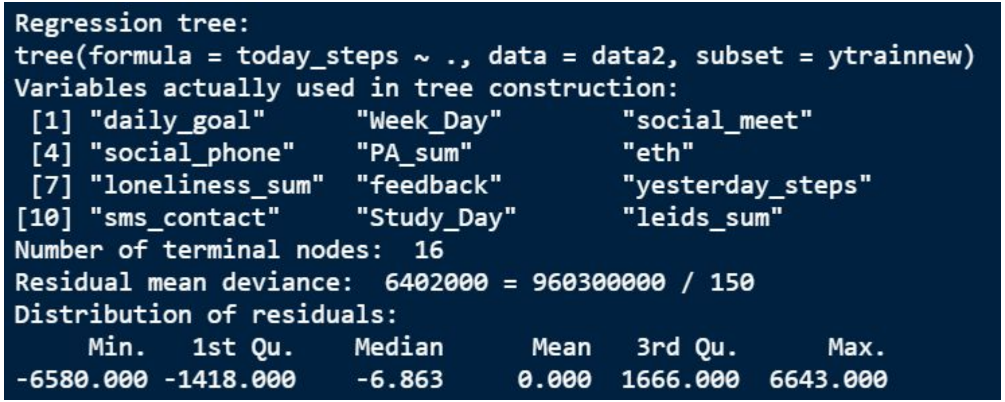


```
> testnew_error
 [1]        0        0        0        0 27982793 26738487 31999000
 [8] 19760694 21811791 27400791 24838619 22556134 23830784 22798516
[15] 23332271 22555518 25092241 25138220 25150091 24054437 24568474
[22] 24337975 23229004 23314381 23100183 25699073 23371353 23694055
[29] 23340803 23371578 23634106 24559832 23734452 25632939 26260415
[36] 25253817 25256052 22945092 21404739 20444130 18120771 18874611
[43] 19308758 26405654        0
```

Regression tree with today_steps as response



Summary

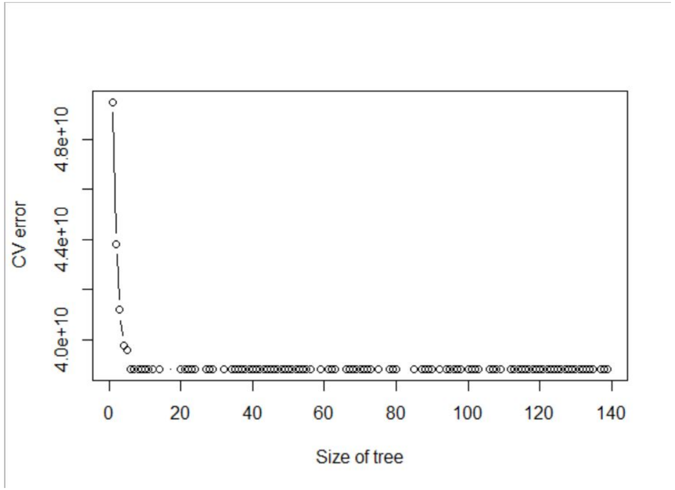


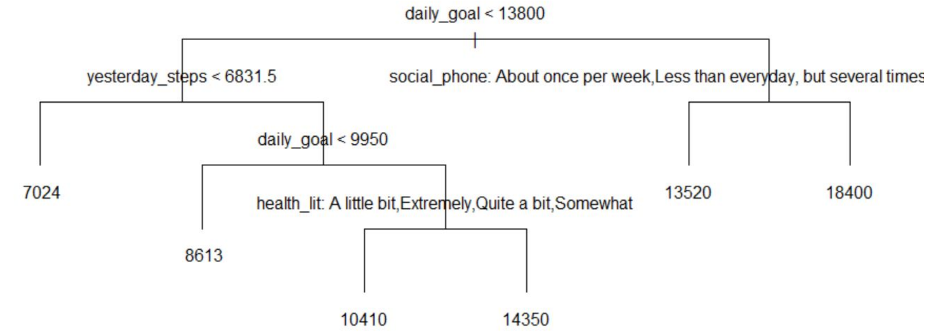
```
Regression tree:
tree(formula = today_steps ~ ., data = data2, subset = ytrainnew)
Variables actually used in tree construction:
 [1] "daily_goal"      "Week_Day"         "social_meet"
 [4] "social_phone"    "PA_sum"           "eth"
 [7] "loneliness_sum"  "feedback"         "yesterday_steps"
[10] "sms_contact"     "Study_Day"        "leids_sum"
Number of terminal nodes:  16
Residual mean deviance:  6402000 = 960300000 / 150
Distribution of residuals:
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-6580.000 -1418.000    -6.863     0.000  1666.000  6643.000
```

- From the summary, 12 variables were used to fit the regression tree
- There are a total of 16 terminal nodes

Fit an larger tree and prune it to obtain a subtree

- Performed a 10-fold cross validation to choose the size of the subtree and the tree with 6-nodes onwards has the lowest CV error
- Choose the parsimonious 6-node tree and obtained the following subtree





daily_goal < 13800
yesterday_steps < 6831.5
social_phone: About once per week,Less than everyday, but several times
daily_goal < 9950
7024
health_lit: A little bit,Extremely,Quite a bit,Somewhat
13520
18400
8613
10410
14350

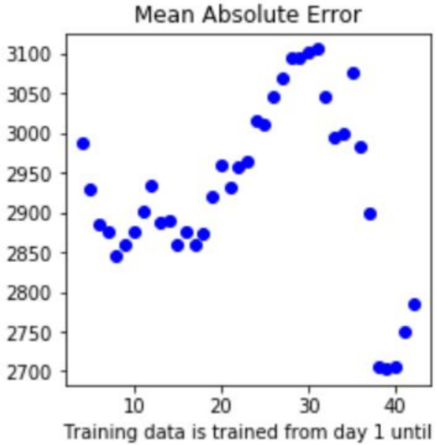
Limitations of decision trees

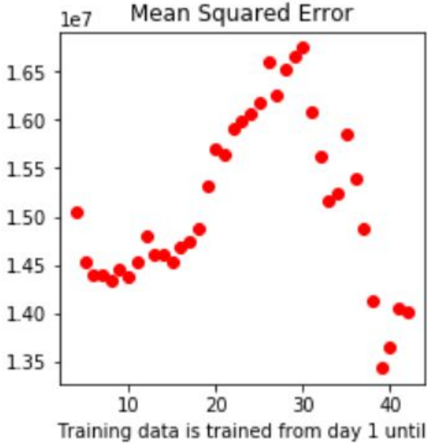
- Decision trees may be simple and useful for interpretation
- However, trees can be very non-robust where a small change in the data can cause a huge change in the final estimated tree
- Not competitive in prediction accuracy

Random Forests

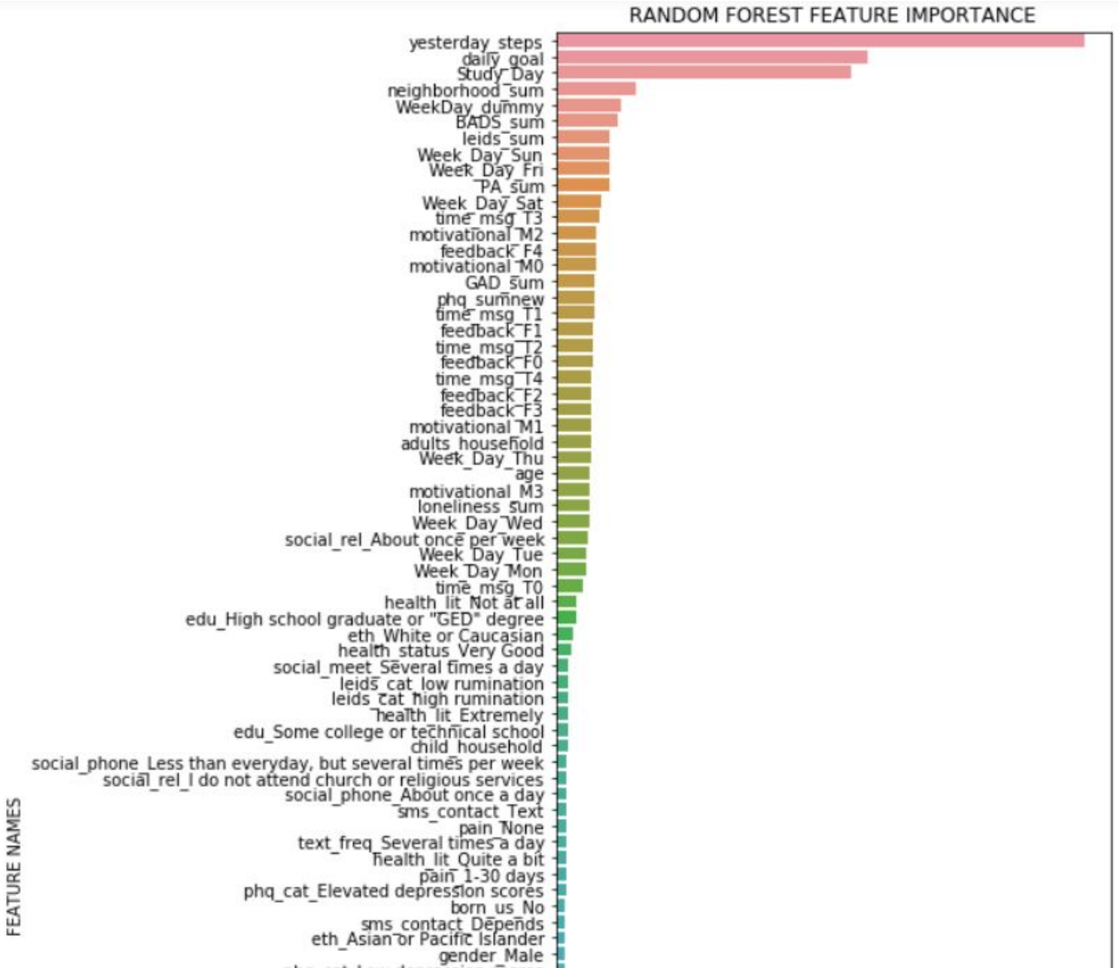
Regression using random forest

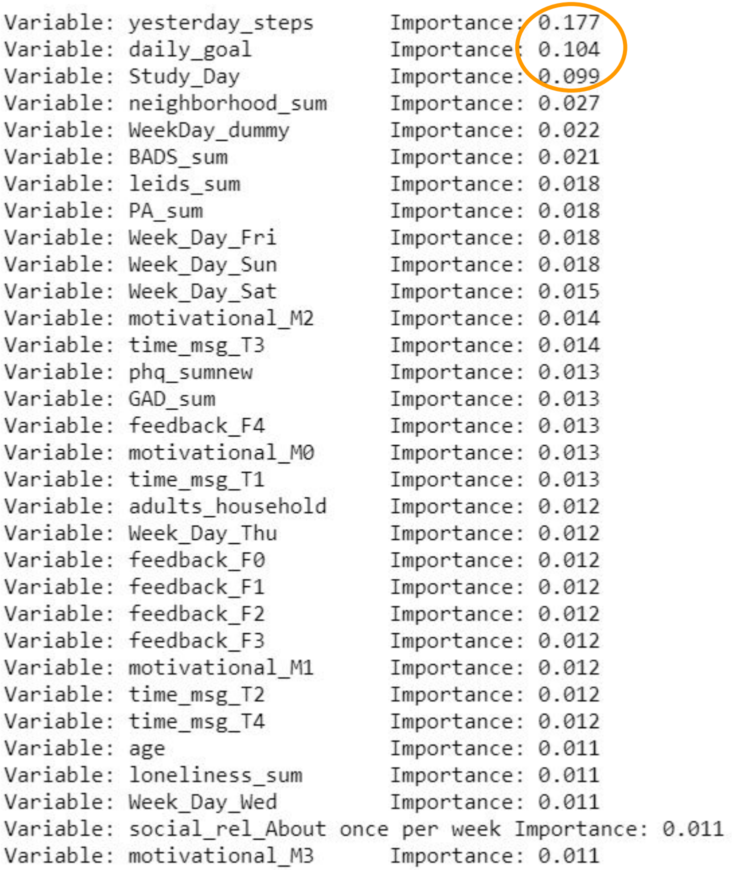
- Trained the model using different Study_Day from day 1 to day t and tested the model from day t+1 to day 45 where t is 4 to 43.
- Lowest MSE of 13432450.64 is obtained when trained from day 1 to day 39 and tested the model from day 40 to day 45





Feature Importance from random forest

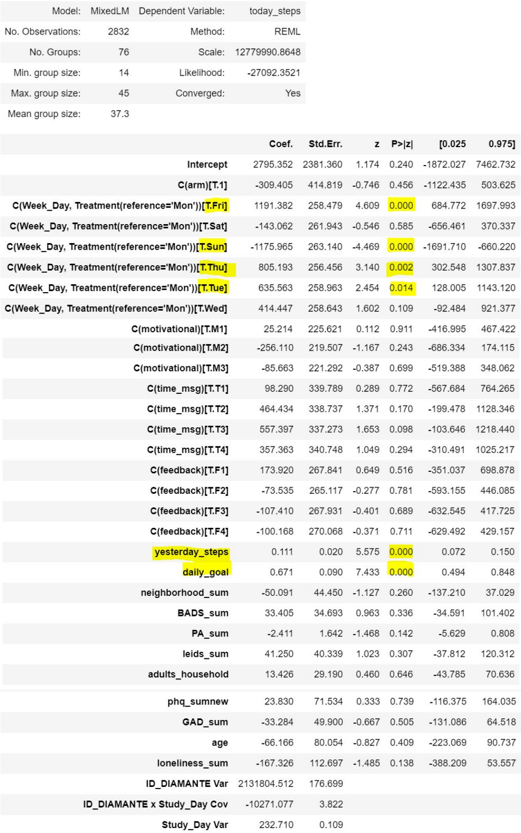




```
Variable: yesterday_steps      Importance: 0.177
Variable: daily_goal           Importance: 0.104
Variable: Study_Day            Importance: 0.099
Variable: neighborhood_sum     Importance: 0.027
Variable: WeekDay_dummy        Importance: 0.022
Variable: BADS_sum             Importance: 0.021
Variable: leids_sum            Importance: 0.018
Variable: PA_sum               Importance: 0.018
Variable: Week_Day_Fri         Importance: 0.018
Variable: Week_Day_Sun         Importance: 0.018
Variable: Week_Day_Sat         Importance: 0.015
Variable: motivational_M2      Importance: 0.014
Variable: time_msg_T3          Importance: 0.014
Variable: phq_sumnew           Importance: 0.013
Variable: GAD_sum              Importance: 0.013
Variable: feedback_F4          Importance: 0.013
Variable: motivational_M0      Importance: 0.013
Variable: time_msg_T1          Importance: 0.013
Variable: adults_household     Importance: 0.012
Variable: Week_Day_Thu         Importance: 0.012
Variable: feedback_F0          Importance: 0.012
Variable: feedback_F1          Importance: 0.012
Variable: feedback_F2          Importance: 0.012
Variable: feedback_F3          Importance: 0.012
Variable: motivational_M1      Importance: 0.012
Variable: time_msg_T2          Importance: 0.012
Variable: time_msg_T4          Importance: 0.012
Variable: age                  Importance: 0.011
Variable: loneliness_sum       Importance: 0.011
Variable: Week_Day_Wed         Importance: 0.011
Variable: social_rel_About once per week Importance: 0.011
Variable: motivational_M3      Importance: 0.011
```

Fitted a multilevel model with variables up till feature importance of

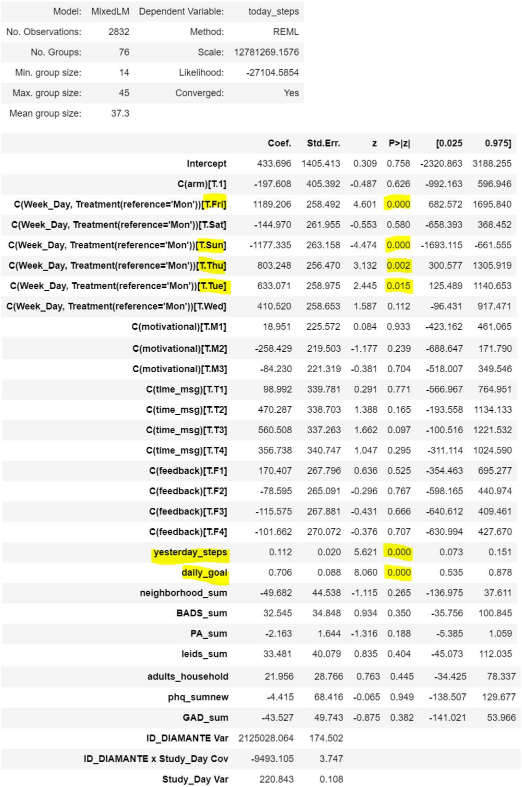
0.011



Model:	MixedLM	Dependent Variable:	today_steps
No. Observations:	2832	Method:	REML
No. Groups:	76	Scale:	12779990.8648
Min. group size:	14	Likelihood:	-27092.3521
Max. group size:	45	Converged:	Yes
Mean group size:	37.3		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	2795.352	2381.360	1.174	0.240	-1872.027	7462.732
C(arm)[T.1]	-309.405	414.819	-0.746	0.456	-1122.435	503.625
C(Week_Day, Treatment(reference='Mon'))[T.Fri]	1191.382	258.479	4.609	0.000	684.772	1697.993
C(Week_Day, Treatment(reference='Mon'))[T.Sat]	-143.062	261.943	-0.546	0.585	-656.461	370.337
C(Week_Day, Treatment(reference='Mon'))[T.Sun]	-1175.965	263.140	-4.469	0.000	-1691.710	-660.220
C(Week_Day, Treatment(reference='Mon'))[T.Thu]	805.193	256.456	3.140	0.002	302.548	1307.837
C(Week_Day, Treatment(reference='Mon'))[T.Tue]	635.563	258.963	2.454	0.014	128.005	1143.120
C(Week_Day, Treatment(reference='Mon'))[T.Wed]	414.447	258.643	1.602	0.109	-92.484	921.377
C(motivational)[T.M1]	25.214	225.621	0.112	0.911	-416.995	467.422
C(motivational)[T.M2]	-256.110	219.507	-1.167	0.243	-686.334	174.115
C(motivational)[T.M3]	-85.663	221.292	-0.387	0.699	-519.388	348.062
C(time_msg)[T.T1]	98.290	339.789	0.289	0.772	-567.684	764.265
C(time_msg)[T.T2]	464.434	338.737	1.371	0.170	-199.478	1128.346
C(time_msg)[T.T3]	557.397	337.273	1.653	0.098	-103.646	1218.440
C(time_msg)[T.T4]	357.363	340.748	1.049	0.294	-310.491	1025.217
C(feedback)[T.F1]	173.920	267.841	0.649	0.516	-351.037	698.878
C(feedback)[T.F2]	-73.535	265.117	-0.277	0.781	-593.155	446.085
C(feedback)[T.F3]	-107.410	267.931	-0.401	0.689	-632.545	417.725
C(feedback)[T.F4]	-100.168	270.068	-0.371	0.711	-629.492	429.157
yesterday_steps	0.111	0.020	5.575	0.000	0.072	0.150
daily_goal	0.671	0.090	7.433	0.000	0.494	0.848
neighborhood_sum	-50.091	44.450	-1.127	0.260	-137.210	37.029
BADS_sum	33.405	34.693	0.963	0.336	-34.591	101.402
PA_sum	-2.411	1.642	-1.468	0.142	-5.629	0.808
leids_sum	41.250	40.339	1.023	0.307	-37.812	120.312
adults_household	13.426	29.190	0.460	0.646	-43.785	70.636
phq_sumnew	23.830	71.534	0.333	0.739	-116.375	164.035
GAD_sum	-33.284	49.900	-0.667	0.505	-131.086	64.518
age	-66.166	80.054	-0.827	0.409	-223.069	90.737
loneliness_sum	-167.326	112.697	-1.485	0.138	-388.209	53.557
ID_DIAMANTE Var	2131804.512	176.699				
ID_DIAMANTE x Study_Day Cov	-10271.077	3.822				
Study_Day Var	232.710	0.109				

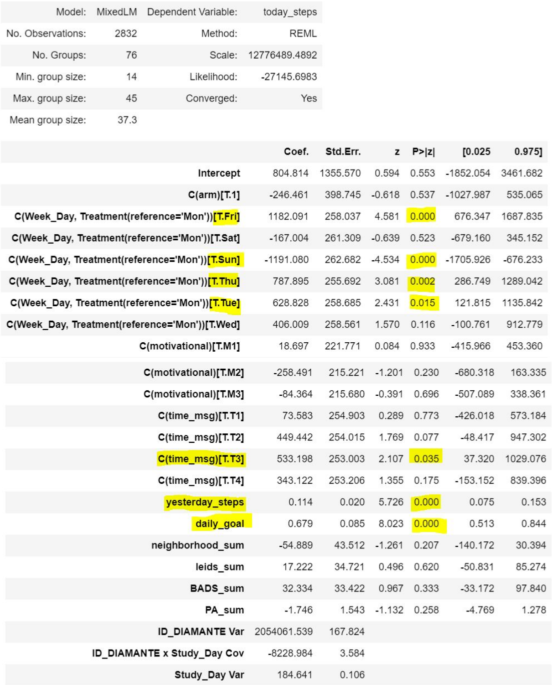
0.012



Model:	MixedLM	Dependent Variable:	today_steps
No. Observations:	2832	Method:	REML
No. Groups:	76	Scale:	12781269.1576
Min. group size:	14	Likelihood:	-27104.5854
Max. group size:	45	Converged:	Yes
Mean group size:	37.3		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	433.696	1405.413	0.309	0.758	-2320.863	3188.255
C(arm)[T.1]	-197.608	405.392	-0.487	0.626	-992.163	596.946
C(Week_Day, Treatment(reference='Mon'))[T.Fri]	1189.206	258.492	4.601	0.000	682.572	1695.840
C(Week_Day, Treatment(reference='Mon'))[T.Sat]	-144.970	261.955	-0.553	0.580	-658.393	368.452
C(Week_Day, Treatment(reference='Mon'))[T.Sun]	-1177.335	263.158	-4.474	0.000	-1693.115	-661.555
C(Week_Day, Treatment(reference='Mon'))[T.Thu]	803.248	256.470	3.132	0.002	300.577	1305.919
C(Week_Day, Treatment(reference='Mon'))[T.Tue]	633.071	258.975	2.445	0.015	125.489	1140.653
C(Week_Day, Treatment(reference='Mon'))[T.Wed]	410.520	258.653	1.587	0.112	-96.431	917.471
C(motivational)[T.M1]	18.951	225.572	0.084	0.933	-423.162	461.065
C(motivational)[T.M2]	-258.429	219.503	-1.177	0.239	-688.647	171.790
C(motivational)[T.M3]	-84.230	221.319	-0.381	0.704	-518.007	349.546
C(time_msg)[T.T1]	98.992	339.781	0.291	0.771	-566.967	764.951
C(time_msg)[T.T2]	470.287	338.703	1.388	0.165	-193.558	1134.133
C(time_msg)[T.T3]	560.508	337.263	1.662	0.097	-100.516	1221.532
C(time_msg)[T.T4]	356.738	340.747	1.047	0.295	-311.114	1024.590
C(feedback)[T.F1]	170.407	267.796	0.636	0.525	-354.463	695.277
C(feedback)[T.F2]	-78.595	265.091	-0.296	0.767	-598.165	440.974
C(feedback)[T.F3]	-115.575	267.881	-0.431	0.666	-640.612	409.461
C(feedback)[T.F4]	-101.662	270.072	-0.376	0.707	-630.994	427.670
yesterday_steps	0.112	0.020	5.621	0.000	0.073	0.151
daily_goal	0.706	0.088	8.060	0.000	0.535	0.878
neighborhood_sum	-49.682	44.538	-1.115	0.265	-136.975	37.611
BADS_sum	32.545	34.848	0.934	0.350	-35.756	100.845
PA_sum	-2.163	1.644	-1.316	0.188	-5.385	1.059
leids_sum	33.481	40.079	0.835	0.404	-45.073	112.035
adults_household	21.956	28.766	0.763	0.445	-34.425	78.337
phq_sumnew	-4.415	68.416	-0.065	0.949	-138.507	129.677
GAD_sum	-43.527	49.743	-0.875	0.382	-141.021	53.966
ID_DIAMANTE Var	2125028.064	174.502				
ID_DIAMANTE x Study_Day Cov	-9493.105	3.747				
Study_Day Var	220.843	0.108				

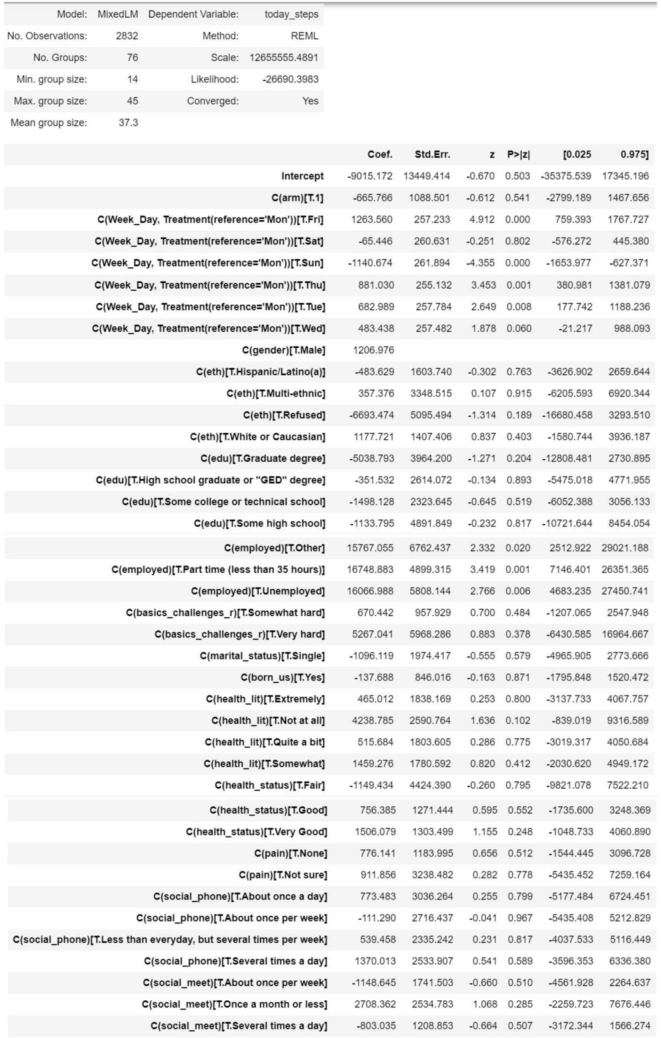
0.014



Model:	MixedLM	Dependent Variable:	today_steps
No. Observations:	2832	Method:	REML
No. Groups:	76	Scale:	12776489.4892
Min. group size:	14	Likelihood:	-27145.6983
Max. group size:	45	Converged:	Yes
Mean group size:	37.3		

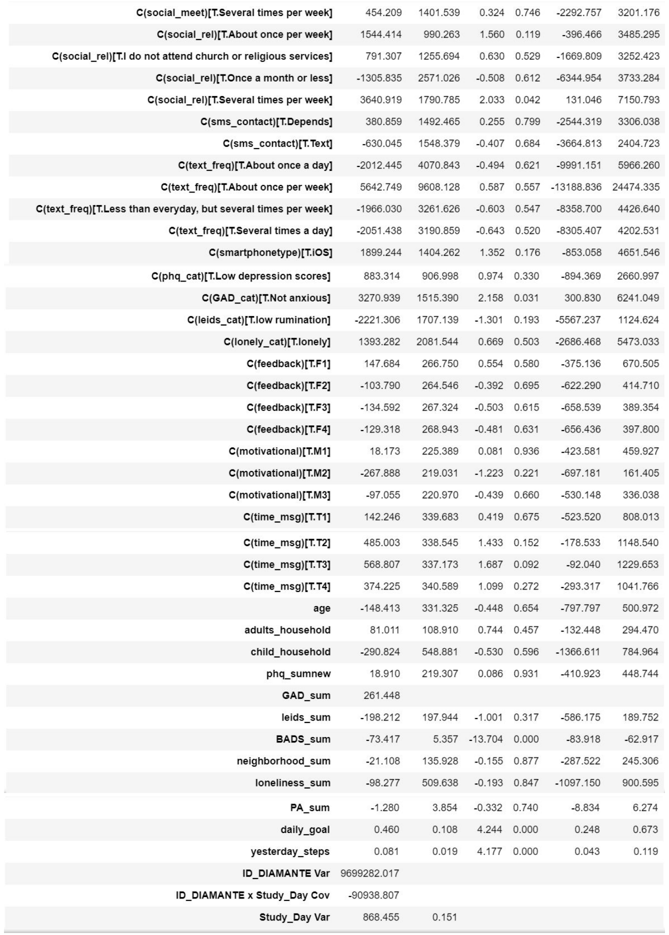
	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	804.814	1355.570	0.594	0.553	-1852.054	3461.682
C(arm)[T.1]	-246.461	398.745	-0.618	0.537	-1027.987	535.065
C(Week_Day, Treatment(reference='Mon'))[T.Fri]	1182.091	258.037	4.581	0.000	676.347	1687.835
C(Week_Day, Treatment(reference='Mon'))[T.Sat]	-167.004	261.309	-0.639	0.523	-679.160	345.152
C(Week_Day, Treatment(reference='Mon'))[T.Sun]	-1191.080	262.682	-4.534	0.000	-1705.926	-676.233
C(Week_Day, Treatment(reference='Mon'))[T.Thu]	787.895	255.692	3.081	0.002	286.749	1289.042
C(Week_Day, Treatment(reference='Mon'))[T.Tue]	628.828	258.685	2.431	0.015	121.815	1135.842
C(Week_Day, Treatment(reference='Mon'))[T.Wed]	406.009	258.561	1.570	0.116	-100.761	912.779
C(motivational)[T.M1]	18.697	221.771	0.084	0.933	-415.966	453.360
C(motivational)[T.M2]	-258.491	215.221	-1.201	0.230	-680.318	163.335
C(motivational)[T.M3]	-84.364	215.680	-0.391	0.696	-507.089	338.361
C(time_msg)[T.T1]	73.583	254.903	0.289	0.773	-426.018	573.184
C(time_msg)[T.T2]	449.442	254.015	1.769	0.077	-48.417	947.302
C(time_msg)[T.T3]	533.198	253.003	2.107	0.035	37.320	1029.076
C(time_msg)[T.T4]	343.122	253.206	1.355	0.175	-153.152	839.396
yesterday_steps	0.114	0.020	5.726	0.000	0.075	0.153
daily_goal	0.679	0.085	8.023	0.000	0.513	0.844
neighborhood_sum	-54.889	43.512	-1.261	0.207	-140.172	30.394
leids_sum	17.222	34.721	0.496	0.620	-50.831	85.274
BADS_sum	32.334	33.422	0.967	0.333	-33.172	97.840
PA_sum	-1.746	1.543	-1.132	0.258	-4.769	1.278
ID_DIAMANTE Var	2054061.539	167.824				
ID_DIAMANTE x Study_Day Cov	-8228.984	3.584				
Study_Day Var	184.641	0.106				

Full Model using all variables from the data_dropnan2 data set



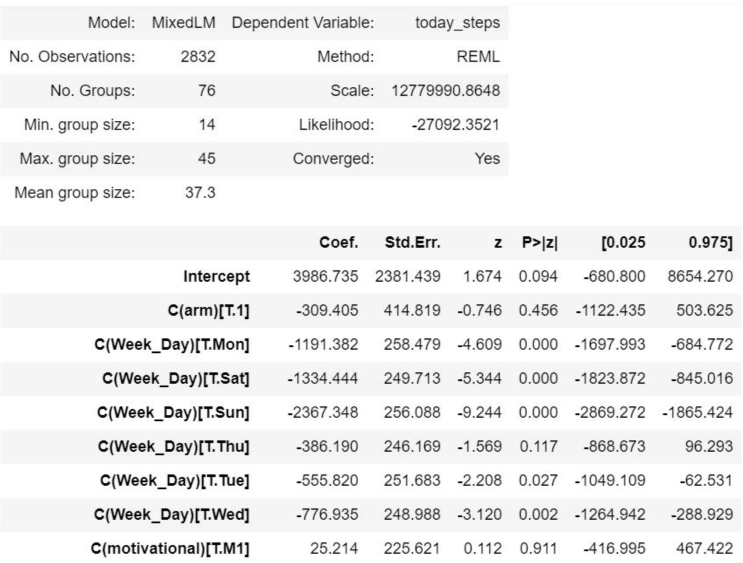
Model:	MixedLM	Dependent Variable:	today_steps
No. Observations:	2832	Method:	REML
No. Groups:	76	Scale:	12655555.4891
Min. group size:	14	Likelihood:	-26690.3983
Max. group size:	45	Converged:	Yes
Mean group size:	37.3		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	-9015.172	13449.414	-0.670	0.503	-35375.539	17345.196
C(arm)[T.1]	-665.766	1088.501	-0.612	0.541	-2799.189	1467.656
C(Week_Day, Treatment(reference='Mon'))[T.Fri]	1263.560	257.233	4.912	0.000	759.393	1767.727
C(Week_Day, Treatment(reference='Mon'))[T.Sat]	-65.446	260.631	-0.251	0.802	-576.272	445.380
C(Week_Day, Treatment(reference='Mon'))[T.Sun]	-1140.674	261.894	-4.355	0.000	-1653.977	-627.371
C(Week_Day, Treatment(reference='Mon'))[T.Thu]	881.030	255.132	3.453	0.001	380.981	1381.079
C(Week_Day, Treatment(reference='Mon'))[T.Tue]	682.989	257.784	2.649	0.008	177.742	1188.236
C(Week_Day, Treatment(reference='Mon'))[T.Wed]	483.438	257.482	1.878	0.060	-21.217	988.093
C(gender)[T.Male]	1206.976					
C(eth)[T.Hispanic/Latino(a)]	-483.629	1603.740	-0.302	0.763	-3626.902	2659.644
C(eth)[T.Multi-ethnic]	357.376	3348.515	0.107	0.915	-6205.593	6920.344
C(eth)[T.Refused]	-6693.474	5095.494	-1.314	0.189	-16680.458	3293.510
C(eth)[T.White or Caucasian]	1177.721	1407.406	0.837	0.403	-1580.744	3936.187
C(edu)[T.Graduate degree]	-5038.793	3964.200	-1.271	0.204	-12808.481	2730.895
C(edu)[T.High school graduate or "GED" degree]	-351.532	2614.072	-0.134	0.893	-5475.018	4771.955
C(edu)[T.Some college or technical school]	-1498.128	2323.645	-0.645	0.519	-6052.388	3056.133
C(edu)[T.Some high school]	-1133.795	4891.849	-0.232	0.817	-10721.644	8454.054
C(employed)[T.Other]	15767.055	6762.437	2.332	0.020	2512.922	29021.188
C(employed)[T.Part time (less than 35 hours)]	16748.883	4899.315	3.419	0.001	7146.401	26351.365
C(employed)[T.Unemployed]	16066.988	5808.144	2.766	0.006	4683.235	27450.741
C(basics_challenges_r)[T.Somewhat hard]	670.442	957.929	0.700	0.484	-1207.065	2547.948
C(basics_challenges_r)[T.Very hard]	5267.041	5968.286	0.883	0.378	-6430.585	16964.667
C(marital_status)[T.Single]	-1096.119	1974.417	-0.555	0.579	-4965.905	2773.666
C(born_us)[T.Yes]	-137.688	846.016	-0.163	0.871	-1795.848	1520.472
C(health_lit)[T.Extremely]	465.012	1838.169	0.253	0.800	-3137.733	4067.757
C(health_lit)[T.Not at all]	4238.785	2590.764	1.636	0.102	-839.019	9316.589
C(health_lit)[T.Quite a bit]	515.684	1803.605	0.286	0.775	-3019.317	4050.684
C(health_lit)[T.Somewhat]	1459.276	1780.592	0.820	0.412	-2030.620	4949.172
C(health_status)[T.Fair]	-1149.434	4424.390	-0.260	0.795	-9821.078	7522.210
C(health_status)[T.Good]	756.385	1271.444	0.595	0.552	-1735.600	3248.369
C(health_status)[T.Very Good]	1506.079	1303.499	1.155	0.248	-1048.733	4060.890
C(pain)[T.None]	776.141	1183.995	0.656	0.512	-1544.445	3096.728
C(pain)[T.Not sure]	911.856	3238.482	0.282	0.778	-5435.452	7259.164
C(social_phone)[T.About once a day]	773.483	3036.264	0.255	0.799	-5177.484	6724.451
C(social_phone)[T.About once per week]	-111.290	2716.437	-0.041	0.967	-5435.408	5212.829
C(social_phone)[T.Less than everyday, but several times per week]	539.458	2335.242	0.231	0.817	-4037.533	5116.449
C(social_phone)[T.Several times a day]	1370.013	2533.907	0.541	0.589	-3596.353	6336.380
C(social_meet)[T.About once per week]	-1148.645	1741.503	-0.660	0.510	-4561.928	2264.637
C(social_meet)[T.Once a month or less]	2708.362	2534.783	1.068	0.285	-2259.723	7676.446
C(social_meet)[T.Several times a day]	-803.035	1208.853	-0.664	0.507	-3172.344	1566.274



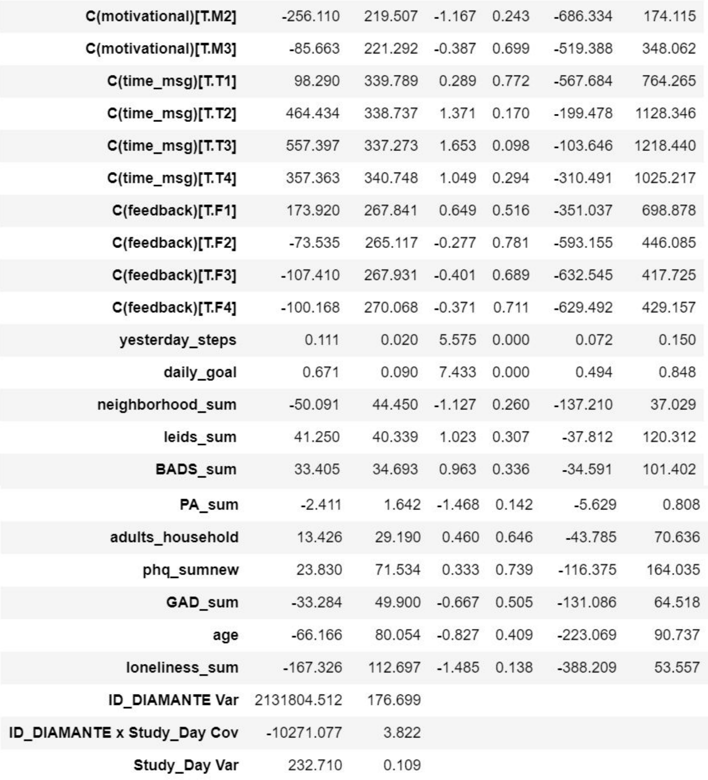
C(social_meet)[T.Several times per week]	454.209	1401.539	0.324	0.746	-2292.757	3201.176
C(social_rel)[T.About once per week]	1544.414	990.263	1.560	0.119	-396.466	3485.295
C(social_rel)[T.I do not attend church or religious services]	791.307	1255.694	0.630	0.529	-1669.809	3252.423
C(social_rel)[T.Once a month or less]	-1305.835	2571.026	-0.508	0.612	-6344.954	3733.284
C(social_rel)[T.Several times per week]	3640.919	1790.785	2.033	0.042	131.046	7150.793
C(sms_contact)[T.Depends]	380.859	1492.465	0.255	0.799	-2544.319	3306.038
C(sms_contact)[T.Text]	-630.045	1548.379	-0.407	0.684	-3664.813	2404.723
C(text_freq)[T.About once a day]	-2012.445	4070.843	-0.494	0.621	-9991.151	5966.260
C(text_freq)[T.About once per week]	5642.749	9608.128	0.587	0.557	-13188.836	24474.335
C(text_freq)[T.Less than everyday, but several times per week]	-1966.030	3261.626	-0.603	0.547	-8358.700	4426.640
C(text_freq)[T.Several times a day]	-2051.438	3190.859	-0.643	0.520	-8305.407	4202.531
C(smartphonetype)[T.iOS]	1899.244	1404.262	1.352	0.176	-853.058	4651.546
C(phq_cat)[T.Low depression scores]	883.314	906.998	0.974	0.330	-894.369	2660.997
C(GAD_cat)[T.Not anxious]	3270.939	1515.390	2.158	0.031	300.830	6241.049
C(leids_cat)[T.low rumination]	-2221.306	1707.139	-1.301	0.193	-5567.237	1124.624
C(lonely_cat)[T.lonely]	1393.282	2081.544	0.669	0.503	-2686.468	5473.033
C(feedback)[T.F1]	147.684	266.750	0.554	0.580	-375.136	670.505
C(feedback)[T.F2]	-103.790	264.546	-0.392	0.695	-622.290	414.710
C(feedback)[T.F3]	-134.592	267.324	-0.503	0.615	-658.539	389.354
C(feedback)[T.F4]	-129.318	268.943	-0.481	0.631	-656.436	397.800
C(motivational)[T.M1]	18.173	225.389	0.081	0.936	-423.581	459.927
C(motivational)[T.M2]	-267.888	219.031	-1.223	0.221	-697.181	161.405
C(motivational)[T.M3]	-97.055	220.970	-0.439	0.660	-530.148	336.038
C(time_msg)[T.T1]	142.246	339.683	0.419	0.675	-523.520	808.013
C(time_msg)[T.T2]	485.003	338.545	1.433	0.152	-178.533	1148.540
C(time_msg)[T.T3]	568.807	337.173	1.687	0.092	-92.040	1229.653
C(time_msg)[T.T4]	374.225	340.589	1.099	0.272	-293.317	1041.766
age	-148.413	331.325	-0.448	0.654	-797.797	500.972
adults_household	81.011	108.910	0.744	0.457	-132.448	294.470
child_household	-290.824	548.881	-0.530	0.596	-1366.611	784.964
phq_sumnew	18.910	219.307	0.086	0.931	-410.923	448.744
GAD_sum	261.448					
leids_sum	-198.212	197.944	-1.001	0.317	-586.175	189.752
BADS_sum	-73.417	5.357	-13.704	0.000	-83.918	-62.917
neighborhood_sum	-21.108	135.928	-0.155	0.877	-287.522	245.306
loneliness_sum	-98.277	509.638	-0.193	0.847	-1097.150	900.595
PA_sum	-1.280	3.854	-0.332	0.740	-8.834	6.274
daily_goal	0.460	0.108	4.244	0.000	0.248	0.673
yesterday_steps	0.081	0.019	4.177	0.000	0.043	0.119
ID_DIAMANTE Var	9699282.017					
ID_DIAMANTE x Study_Day Cov	-90938.807					
Study_Day Var	868.455	0.151				

Fitted a multilevel model with variables up till importance of 0.011



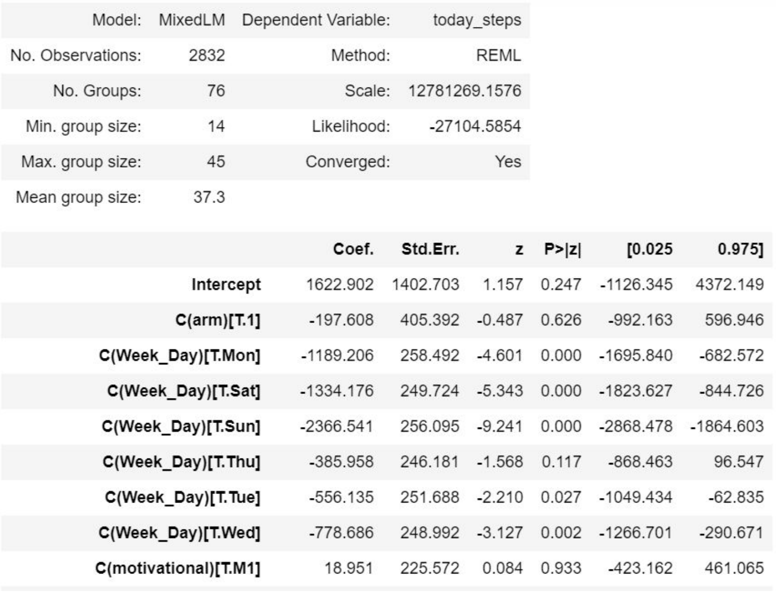
Model:	MixedLM	Dependent Variable:	today_steps
No. Observations:	2832	Method:	REML
No. Groups:	76	Scale:	12779990.8648
Min. group size:	14	Likelihood:	-27092.3521
Max. group size:	45	Converged:	Yes
Mean group size:	37.3		

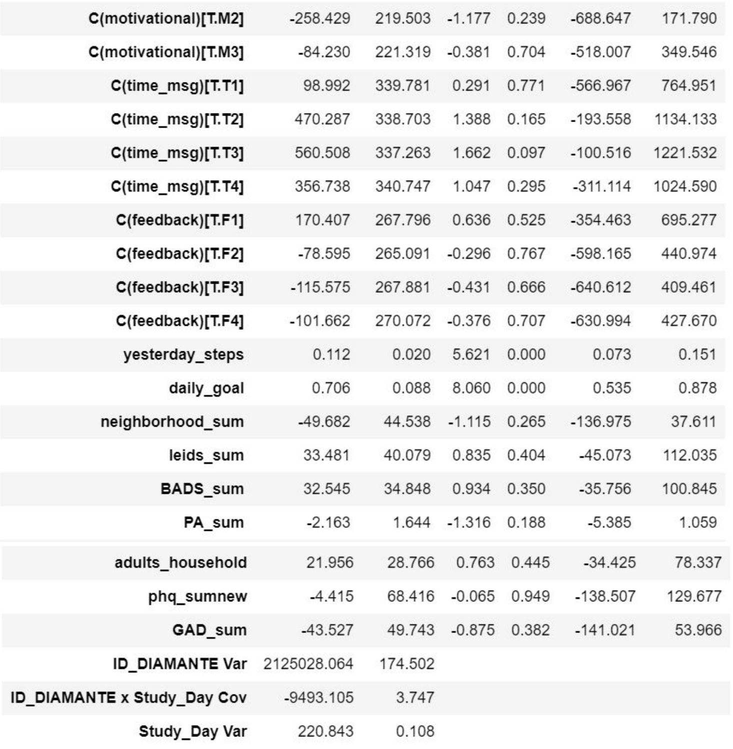
	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	3986.735	2381.439	1.674	0.094	-680.800	8654.270
C(arm)[T.1]	-309.405	414.819	-0.746	0.456	-1122.435	503.625
C(Week_Day)[T.Mon]	-1191.382	258.479	-4.609	0.000	-1697.993	-684.772
C(Week_Day)[T.Sat]	-1334.444	249.713	-5.344	0.000	-1823.872	-845.016
C(Week_Day)[T.Sun]	-2367.348	256.088	-9.244	0.000	-2869.272	-1865.424
C(Week_Day)[T.Thu]	-386.190	246.169	-1.569	0.117	-868.673	96.293
C(Week_Day)[T.Tue]	-555.820	251.683	-2.208	0.027	-1049.109	-62.531
C(Week_Day)[T.Wed]	-776.935	248.988	-3.120	0.002	-1264.942	-288.929
C(motivational)[T.M1]	25.214	225.621	0.112	0.911	-416.995	467.422



C(motivational)[T.M2]	-256.110	219.507	-1.167	0.243	-686.334	174.115
C(motivational)[T.M3]	-85.663	221.292	-0.387	0.699	-519.388	348.062
C(time_msg)[T.T1]	98.290	339.789	0.289	0.772	-567.684	764.265
C(time_msg)[T.T2]	464.434	338.737	1.371	0.170	-199.478	1128.346
C(time_msg)[T.T3]	557.397	337.273	1.653	0.098	-103.646	1218.440
C(time_msg)[T.T4]	357.363	340.748	1.049	0.294	-310.491	1025.217
C(feedback)[T.F1]	173.920	267.841	0.649	0.516	-351.037	698.878
C(feedback)[T.F2]	-73.535	265.117	-0.277	0.781	-593.155	446.085
C(feedback)[T.F3]	-107.410	267.931	-0.401	0.689	-632.545	417.725
C(feedback)[T.F4]	-100.168	270.068	-0.371	0.711	-629.492	429.157
yesterday_steps	0.111	0.020	5.575	0.000	0.072	0.150
daily_goal	0.671	0.090	7.433	0.000	0.494	0.848
neighborhood_sum	-50.091	44.450	-1.127	0.260	-137.210	37.029
leids_sum	41.250	40.339	1.023	0.307	-37.812	120.312
BADS_sum	33.405	34.693	0.963	0.336	-34.591	101.402
PA_sum	-2.411	1.642	-1.468	0.142	-5.629	0.808
adults_household	13.426	29.190	0.460	0.646	-43.785	70.636
phq_sumnew	23.830	71.534	0.333	0.739	-116.375	164.035
GAD_sum	-33.284	49.900	-0.667	0.505	-131.086	64.518
age	-66.166	80.054	-0.827	0.409	-223.069	90.737
loneliness_sum	-167.326	112.697	-1.485	0.138	-388.209	53.557
ID_DIAMANTE Var	2131804.512	176.699				
ID_DIAMANTE x Study_Day Cov	-10271.077	3.822				
Study_Day Var	232.710	0.109				

Fitted a multilevel model with variables up till importance of 0.012

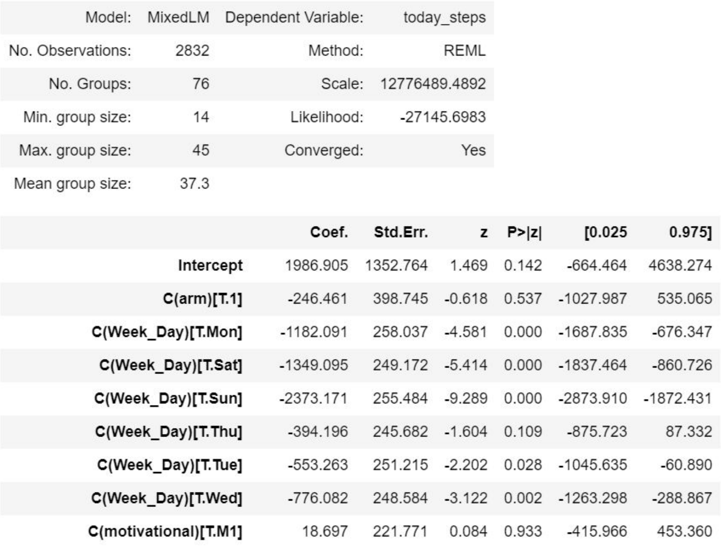




Model:	MixedLM	Dependent Variable:	today_steps
No. Observations:	2832	Method:	REML
No. Groups:	76	Scale:	12781269.1576
Min. group size:	14	Likelihood:	-27104.5854
Max. group size:	45	Converged:	Yes
Mean group size:	37.3		

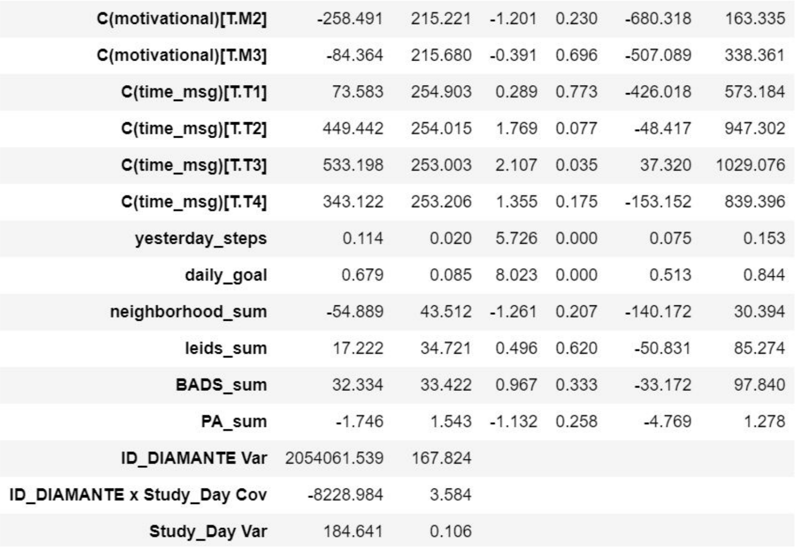
	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	1622.902	1402.703	1.157	0.247	-1126.345	4372.149
C(arm)[T.1]	-197.608	405.392	-0.487	0.626	-992.163	596.946
C(Week_Day)[T.Mon]	-1189.206	258.492	-4.601	0.000	-1695.840	-682.572
C(Week_Day)[T.Sat]	-1334.176	249.724	-5.343	0.000	-1823.627	-844.726
C(Week_Day)[T.Sun]	-2366.541	256.095	-9.241	0.000	-2868.478	-1864.603
C(Week_Day)[T.Thu]	-385.958	246.181	-1.568	0.117	-868.463	96.547
C(Week_Day)[T.Tue]	-556.135	251.688	-2.210	0.027	-1049.434	-62.835
C(Week_Day)[T.Wed]	-778.686	248.992	-3.127	0.002	-1266.701	-290.671
C(motivational)[T.M1]	18.951	225.572	0.084	0.933	-423.162	461.065
C(motivational)[T.M2]	-258.429	219.503	-1.177	0.239	-688.647	171.790
C(motivational)[T.M3]	-84.230	221.319	-0.381	0.704	-518.007	349.546
C(time_msg)[T.T1]	98.992	339.781	0.291	0.771	-566.967	764.951
C(time_msg)[T.T2]	470.287	338.703	1.388	0.165	-193.558	1134.133
C(time_msg)[T.T3]	560.508	337.263	1.662	0.097	-100.516	1221.532
C(time_msg)[T.T4]	356.738	340.747	1.047	0.295	-311.114	1024.590
C(feedback)[T.F1]	170.407	267.796	0.636	0.525	-354.463	695.277
C(feedback)[T.F2]	-78.595	265.091	-0.296	0.767	-598.165	440.974
C(feedback)[T.F3]	-115.575	267.881	-0.431	0.666	-640.612	409.461
C(feedback)[T.F4]	-101.662	270.072	-0.376	0.707	-630.994	427.670
yesterday_steps	0.112	0.020	5.621	0.000	0.073	0.151
daily_goal	0.706	0.088	8.060	0.000	0.535	0.878
neighborhood_sum	-49.682	44.538	-1.115	0.265	-136.975	37.611
leids_sum	33.481	40.079	0.835	0.404	-45.073	112.035
BADS_sum	32.545	34.848	0.934	0.350	-35.756	100.845
PA_sum	-2.163	1.644	-1.316	0.188	-5.385	1.059
adults_household	21.956	28.766	0.763	0.445	-34.425	78.337
phq_sumnew	-4.415	68.416	-0.065	0.949	-138.507	129.677
GAD_sum	-43.527	49.743	-0.875	0.382	-141.021	53.966
ID_DIAMANTE Var	2125028.064	174.502				
ID_DIAMANTE x Study_Day Cov	-9493.105	3.747				
Study_Day Var	220.843	0.108				

Fitted a multilevel model with variables up till importance of 0.014



Model:	MixedLM	Dependent Variable:	today_steps
No. Observations:	2832	Method:	REML
No. Groups:	76	Scale:	12776489.4892
Min. group size:	14	Likelihood:	-27145.6983
Max. group size:	45	Converged:	Yes
Mean group size:	37.3		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	1986.905	1352.764	1.469	0.142	-664.464	4638.274
C(arm)[T.1]	-246.461	398.745	-0.618	0.537	-1027.987	535.065
C(Week_Day)[T.Mon]	-1182.091	258.037	-4.581	0.000	-1687.835	-676.347
C(Week_Day)[T.Sat]	-1349.095	249.172	-5.414	0.000	-1837.464	-860.726
C(Week_Day)[T.Sun]	-2373.171	255.484	-9.289	0.000	-2873.910	-1872.431
C(Week_Day)[T.Thu]	-394.196	245.682	-1.604	0.109	-875.723	87.332
C(Week_Day)[T.Tue]	-553.263	251.215	-2.202	0.028	-1045.635	-60.890
C(Week_Day)[T.Wed]	-776.082	248.584	-3.122	0.002	-1263.298	-288.867
C(motivational)[T.M1]	18.697	221.771	0.084	0.933	-415.966	453.360



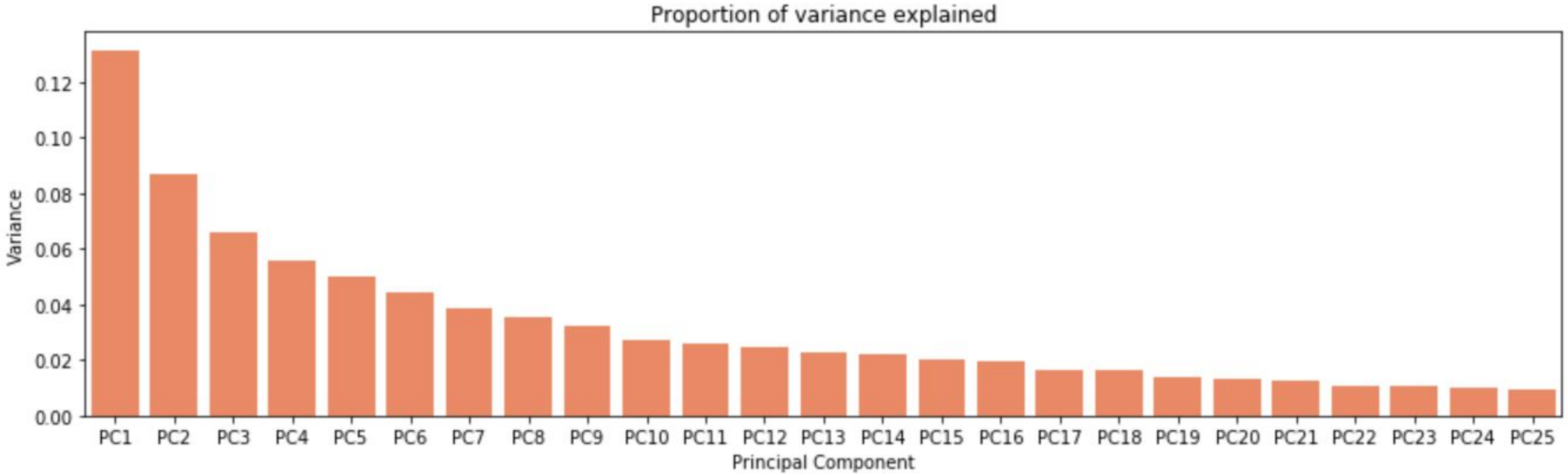
C(motivational)[T.M2]	-258.491	215.221	-1.201	0.230	-680.318	163.335
C(motivational)[T.M3]	-84.364	215.680	-0.391	0.696	-507.089	338.361
C(time_msg)[T.T1]	73.583	254.903	0.289	0.773	-426.018	573.184
C(time_msg)[T.T2]	449.442	254.015	1.769	0.077	-48.417	947.302
C(time_msg)[T.T3]	533.198	253.003	2.107	0.035	37.320	1029.076
C(time_msg)[T.T4]	343.122	253.206	1.355	0.175	-153.152	839.396
yesterday_steps	0.114	0.020	5.726	0.000	0.075	0.153
daily_goal	0.679	0.085	8.023	0.000	0.513	0.844
neighborhood_sum	-54.889	43.512	-1.261	0.207	-140.172	30.394
leids_sum	17.222	34.721	0.496	0.620	-50.831	85.274
BADS_sum	32.334	33.422	0.967	0.333	-33.172	97.840
PA_sum	-1.746	1.543	-1.132	0.258	-4.769	1.278
ID_DIAMANTE Var	2054061.539	167.824				
ID_DIAMANTE x Study_Day Cov	-8228.984	3.584				
Study_Day Var	184.641	0.106				

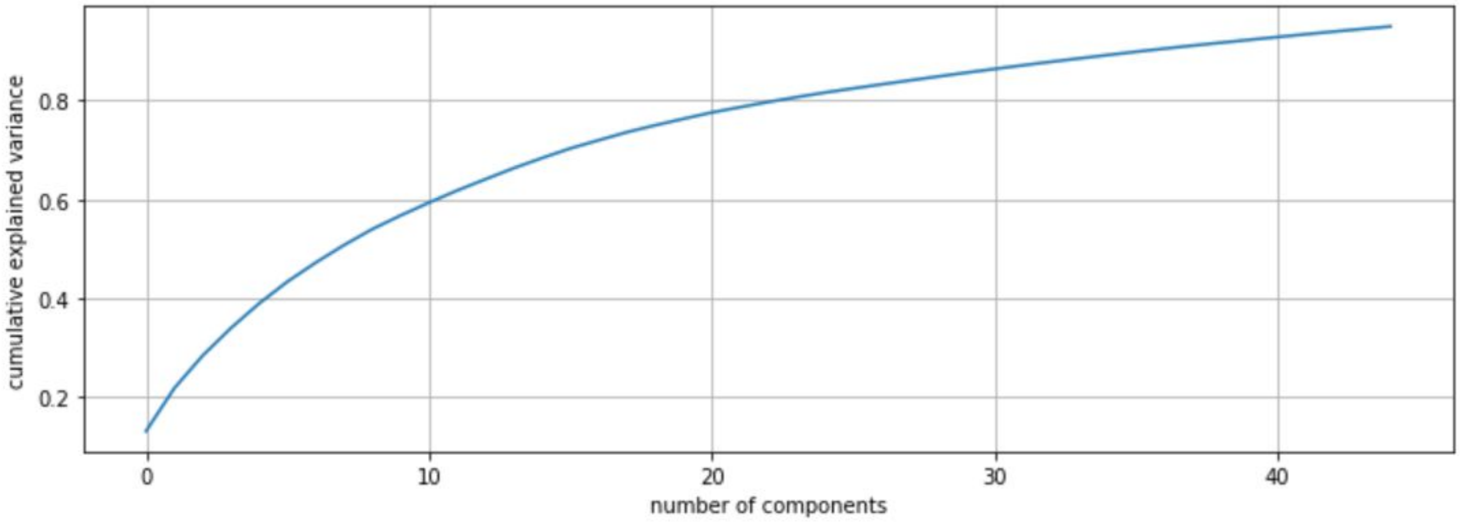
Principal Component Analysis

PCA

- Principal component summarizes a large set of correlated variables using a smaller set of variables that explain most of the variability in the original set.

Proportion of variance explained by the Principal Components



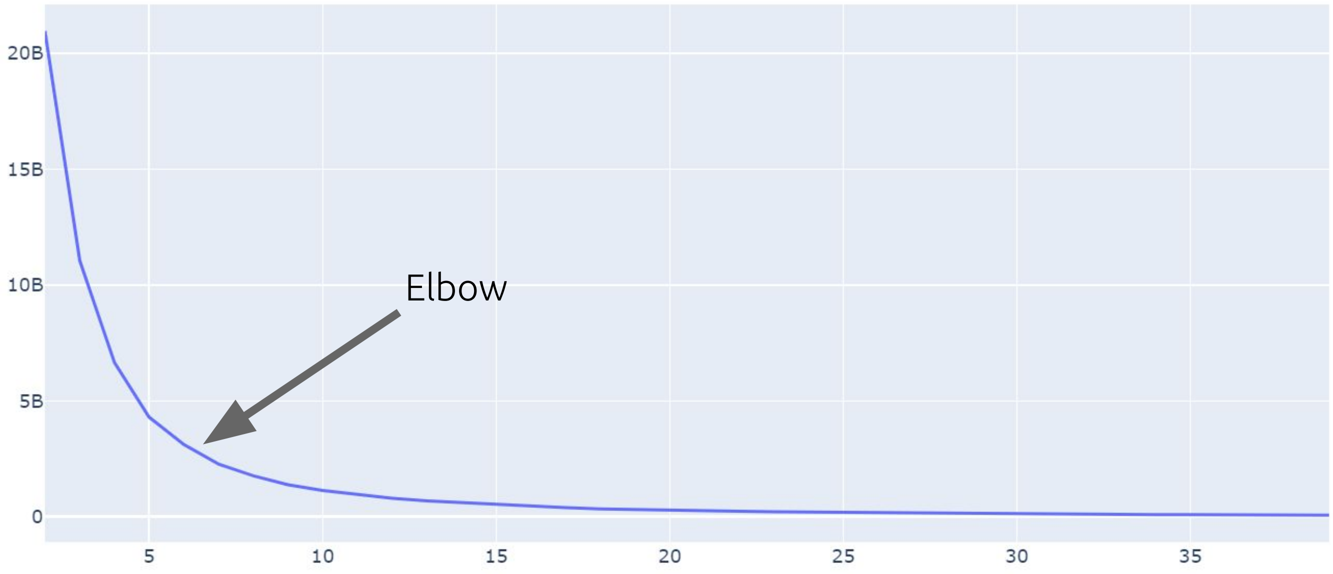

cumulative explained variance
0.8
0.6
0.4
0.2
0
10
20
30
40
number of components

K-means

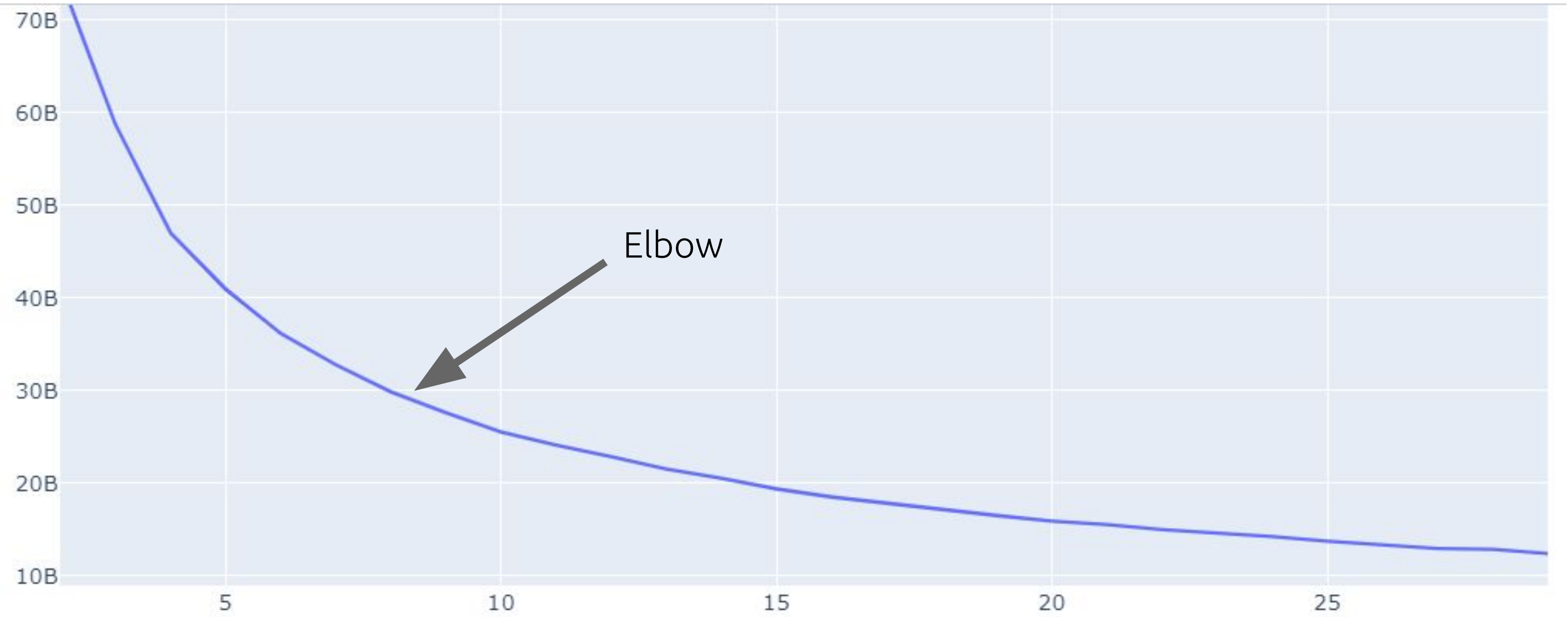
K-means

- Use K-means clustering for partitioning the data set into K distinct non-overlapping clusters

Kmeans



K-modes



Plan

- Currently reading up on clustering methods and apply them to the dataset

Thank you!