

NeuroFlow Data project

Yushan Cai|yc844@drexel.edu

Project A

Assumption

1. The earliest date is the first time to be recorded before treatment.
2. Users always will accept the treatment after the first record.
3. There is no other treatment during the record.
4. There is no the force manure event.

Business question

- 1.visualize the rates
- 2.networking
- 3.business insights

Information we need in the future

1. What kind of treatment users did.
2. when the users accept the treatment.
3. More tress record
4. The patients' demographic information(age, gender...)

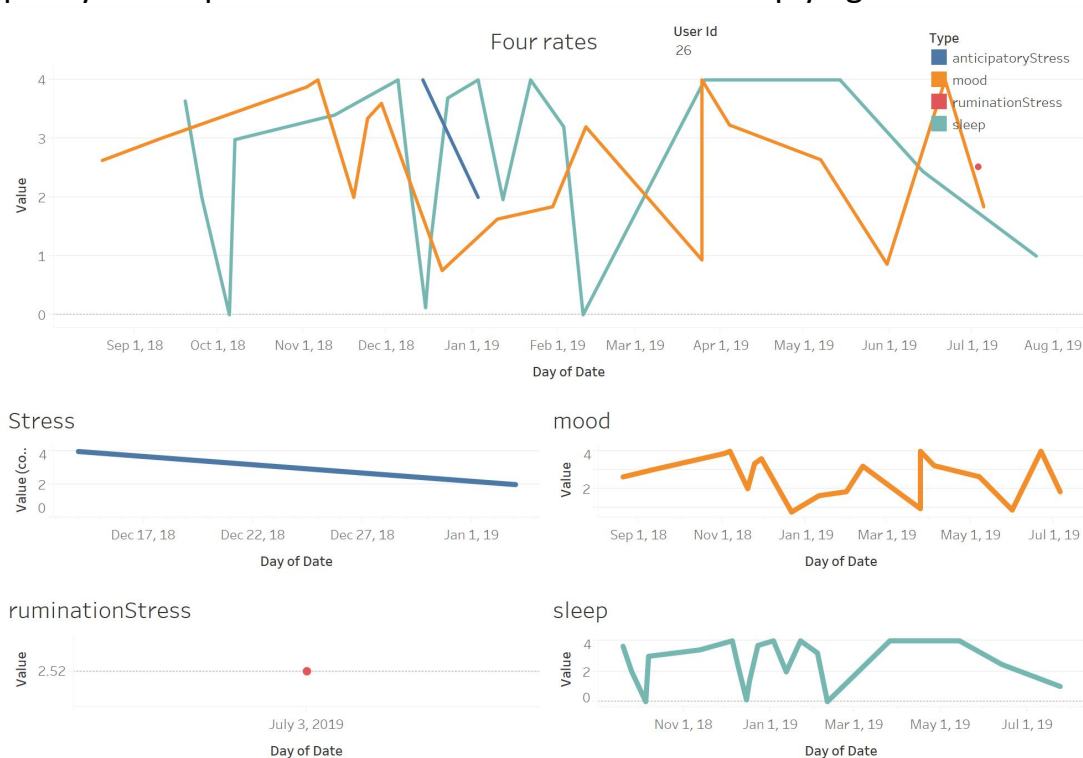
Solution

1. visualize the rates

1.1 Tableau(work book)

A. personal rates records (dashboard 1)

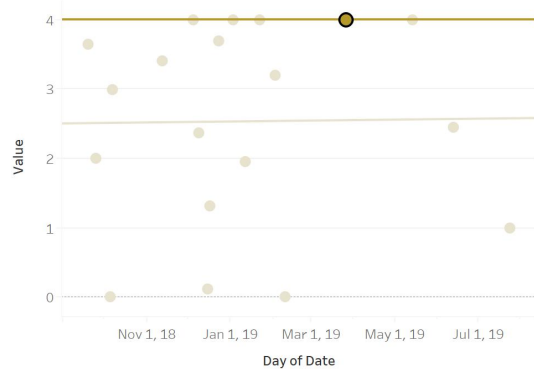
for example, the 26th user almost record twice per month in moon and sleep. This user has insomnia, but the average sleep rate after April,2019 goes well. The anticipatory stress is serious in Dec,2018 to Jan,2019 , the quality of sleep and the rate of mood decrease sharply again.



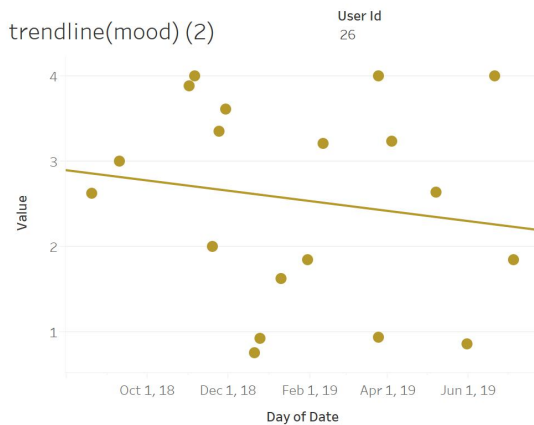
B. personal trend(dash2)

The line is the linear regression.

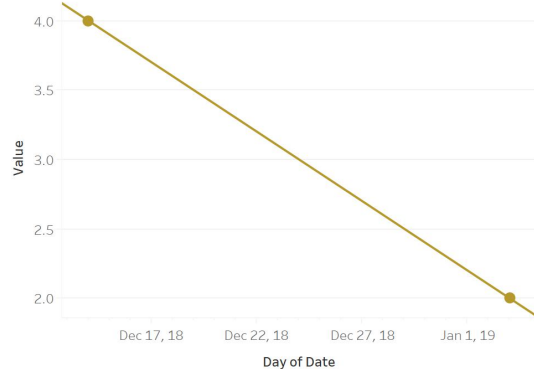
trendline(sleep)



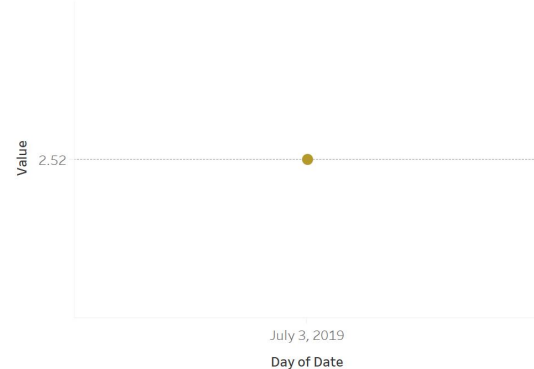
trendline(mood) (2)



trendline(anti) (3)



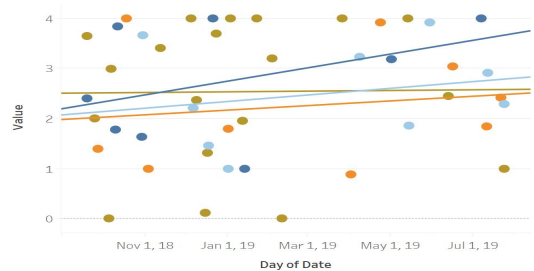
trendline(rumi) (4)



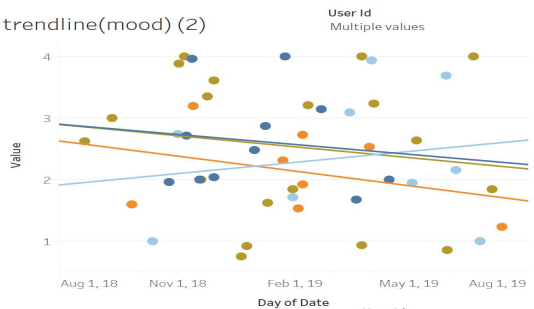
C. Multiple person trend id=0,1,2,26

According to the trend and record we can see the 1 users have good trend on sleep, mood and anticipatory stress. The 26 users have declining trend on mood and anticipatory Stress. The users 1,2,26 can networking with each other to communicate the situations they have.

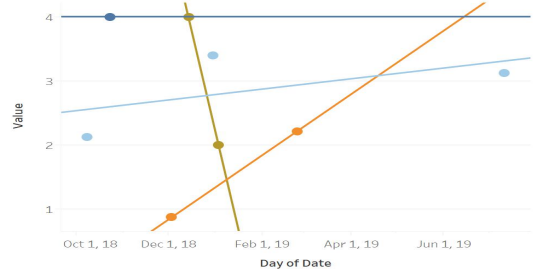
trendline(sleep)



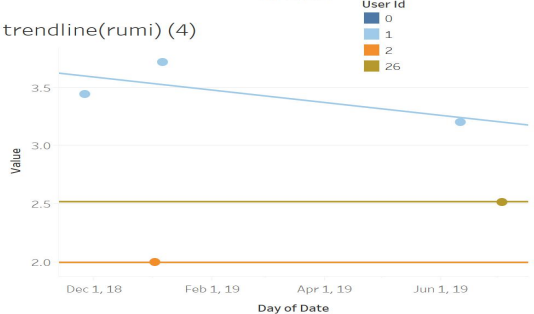
trendline(mood) (2)



trendline(anti) (3)



trendline(rumi) (4)



2. Python (notebook or github)

I also have the version in python, but due to the discrete record, we can not have the very clear visualization. If the data is integrated, we can also use python to generate the plot at once.

Project2

1. How many users completed an exercise in their first month per monthly cohort?

Assume you have two tables in our company's database:

- 'users' table, with columns 'user_id', 'created_at'
- 'exercises' table, with columns 'exercise_id', 'user_id', 'exercise_completion_date'

Assumption: in the first month means they finish the exercise in the same month.

```
DECLARE mm INT;  
SET mm = 0;  
WHILE mm <= 12  
    BEGIN  
        (SELECT COUNT(*)  
         FROM users left join exercises on users(user_id) = exercises(user_id)  
         WHERE MONTH(created_at) = MONTH(exercise_completion_date)  
         and MONTH(created_at)=mm;)  
        / (SELECT COUNT(*)FROM users WHERE MONTH(created_at)=mm;)  
        SET mm = mm + 1 AS "one_month"  
  
        PRINT(mm,one_month)  
    END;  
  
GO
```

2. Assume you have two tables in our company's database:

- 'Providers' table that contains 'provider_id', 'organization_id', and 'organization_name'
 - 'Phq9' table that contains 'patient_id', 'provider_id', 'score', 'datetime_created'
- For context, A phq score ranges from 0-27 and anything 20 or above is considered severe. Write a single query that finds the top five organizations that have the highest average phq9 score per patient.

SELECT TOP 5 organization_name

```
FROM Providers left join Phq9 on Providers(provider_id) =  
Phq9(provider_id)  
Order by AVG(score);
```