

CAPSTONE WEEKLY REPORT

Greenspace Team 3 -Therapeutic Alliance

Zheng Zhang

Bingshen Yang

Zerui Zhang

Kohsin Chen

2024.MAY.21

Team Achievements

- All team members are on track with respective tasks.
- **Data preparation & loading:** Assessment scores and therapeutic alliance scores SQL Query to CSV file.
- **Model discussion:** RF and linear regression

Week 3 Tasks

Assessment score is not enough.
We need more info

Questions/Response

Tag

Count of assessments

Last Week Contribution -Bingshen

Extract key patient characteristics from the Question and Response table, focusing on age, gender, and occupation based on their responses. More info coming...

```
dob_mask = data['QUESTION_TITLE'].str.contains('date of birth', case=False, na=False)
data_dob = data[dob_mask]

data_dob['Date of Birth'] = pd.to_datetime(data_dob['RESPONSES'], errors='coerce')
current_year = datetime.now().year
data_dob['Age'] = current_year - data_dob['Date of Birth'].dt.year
| 

gender_mask = data['QUESTION_TITLE'].str.contains('gender', case=False, na=False)
data_gender = data[gender_mask]
data_gender['Gender'] = data['RESPONSES']

occupation_mask = data['QUESTION_TITLE'].str.contains('occupation', case=False, na=False)
data_occupation = data[occupation_mask]
data_occupation['Occupation'] = data['RESPONSES']
```

Next Step

Working on Mother tone
and Ethnicity classification

White North American
White
White European
Do Not Know
Prefer Not To Answer

Asian – South, Asian – East, Black–African, White European
Black–Caribbean, Middle Eastern
Indigenous/Aboriginal, Latin American, White North American
White European, White North American, First Nations
Indian–Caribbean, White North American
Name: Ethnicity, Length: 283, dtype: int64

English	20263
French	1855
Other	973
Arabic	60
Spanish	33
...	
russian	1
Mohawk	1
Somalian	1
pidgin	1
Indonesian	1
Name: Mother Tongue, Length: 122, dtype: int64	

Last Week Contribution -Zheng

- 1 Built regression model and random forest model to predict therapeutic alliance scores using assessment scores and assessment code.
- 2 Use SQL queries to merge all relevant data encompassing questions, responses, and patient information from the database.
- 3 Classify patients response and then labels to patients.

Model Result

X = [assessment score,
assessment code, therapist
avg therapeutic alliance
score]
y = therapeutic alliance score

----linear regression model----

Mean Squared Error: 250.20060576340484
 r^2 : 0.16003207240927675

----random forest regressor----

Mean Squared Error: 262.0827954537188
 r^2 : 0.12014144856781894

Patient Classification Result

		PATIENT_ID	Age	Gender	Occupation
0	dddc3762dfee68c2c45cade1cbab4351	17.0	Male	Unemployed	
3	59c1260c7ddc0e823ba16b5ed9ab9190	21.0	Male	Student	
4	7f20d2c0d25e39383a41329b6bbe335f	45.0	Male	Unemployed	
6	9310b2ea0cf757be714ff1575daa8d1f	25.0	Female	Health Care	
7	e168b9d918df9df8f60df4eec6ac09a4	29.0	Female	Health Care	
...
203745	4dc3dd39eae0e8c0145dc20d192c4e36	63.0	Female	Education	
203746	ec615e02c1553dfbecb0bc47cdb58852	72.0	Female	Secretarial	
203747	64cdbd92952fb339588a6cdca7dc6e47	27.0	Female	Health Care	
203749	5859571e61882f7615439f0a27352a0f	30.0	Female	Marketing	
203759	1443e53c2d8b71b342008f2c8c063bbe	40.0	Woman	Psychotherapist	
85017 rows × 4 columns					

Challenges

- Which are useful questions
- Answers are vary

	0
0	Can a confidential message be left at this num...
1	Address*
2	I consent to receive email communications and ...
3	Apartment or P.O. Box
4	Client provided consent for participation in a...
...	...
2124	What is your gender identity (or the client's)...
2125	First name:
2126	Last name:
2127	Do either of your parents have any medical con...
2128	Alternate Phone
2129	rows × 1 columns

False
No
Substance use or addiction
True
Trauma/Abuse: Past

Low mood or intense sadness, Stress or anxiety
Risky/Impulsive/Self-harm behavior, Difficulty
Thinking or memory difficulties, Seeing/hearing
Low mood or intense sadness, Stress or anxiety
Substance use or addiction, Thinking or memory
Name: symptom, Length: 26841, dtype: int64

Next Step

1. Focus on extracting and organizing **questions from big clinics** by frequency of occurrence
2. **Identify and select** the most useful questions
3. **Classify** patients based on their responses data
4. Try to build a model predicting TA scores with these refined features

Last Week Contribution

-Zerui

1

Categorized and analyzed tags attached to patient records.

2

The tags were normalized and grouped into categories for common mental health conditions.

3

Additionally, tags were categorized based on treatment methods, such as CBT, DBT, mindfulness, and so on.

Examples:

- Recognized tags that will not be included in further analysis

```
inoffice  
archived  
followup  
pseudo  
kelowna  
...  
community-living-huntsville  
ref-hospitalmedical-centre  
block-5  
waiting
```

1

Identify various tags and keywords associated with them.

2

Assign tags to these treatment groups based on keyword matching.

- Tags categorized based on treatment methods

cbt	[cbt-program, group-cbt-anxiety-depression, tr...
dbt	[group-dbt, modality-dbt, dbt, approach-dbt, d...
group_therapy	[group-coping-w-chronic-illness, erst-group, s...
medication	[on-medication, eap-medsleep, medication-manag...
mindfulness	[modality-mindfulness, approach-mindfulness, m...

Challenges

- **5.4K Tags in total**
- **Accurately group tags despite variations in wording and terminology**
- **We are unable to confirm the use of many tags (eg. inoffice)**
- **Complex tags contain more than one keyword**

group-cbt-anxiety-depression
anxiety-and-depression-group
pp-child-anxiety-depression
pp-anxiety-no-depression
pp-depression-no-anxiety
pp-anxiety-and-depression
ax-anxiety-and-depression
ax-anxiety-no-depression
ax-depression-no-anxiety

Next Step

- Carefully split complex tags into smaller tags, ensuring each tag contains only one keyword.
- Focus on data from the three largest clinics, apply the developed grouping strategy to evaluate its effectiveness.
- Finally, the tags will be used as predictors in building models to investigate the relationship with therapeutic alliance.

Last Week Contribution

-Kohsin

- 1 Utilized SQL queries to download assessment data as csv
- 2 EDA to explore the relationship between initial assessment score and therapeutic alliance (TA)
- 3 Built random forest and regression models
- 4 Extracted the counts of completed and missing assessments to see if these could predict TA

SQL Query EDA Model Building

	PATIENT_THERAPEUTIC_ALLIANCE_SCORE	SCORE_NORMALIZED
PATIENT_THERAPEUTIC_ALLIANCE_SCORE	1.000000	0.136195
SCORE_NORMALIZED	0.136195	1.000000

```
SELECT
    T.THERAPIST_ID,
    T.THERAPEUTIC_ALLIANCE_SCORE,
    P.PATIENT_ID,
    PAR.PARTICIPANT_ID,
    P.THERAPEUTIC_ALLIANCE_SCORE AS PATIENT_THERAPEUTIC_ALLIANCE_SCORE,
    CAPS_ASSESSMENTS_ASSIGNED.ASSESSMENT_ID,
    CAPS_ASSESSMENTS.CODE,
    CAPS_ASSESSMENTS.ASSESSMENT_CATEGORY_ID,
    RES.CREATED,
    RES.SCORE,
    RES.SCORE_NORMALIZED,
    ROW_NUMBER() OVER (PARTITION BY P.PATIENT_ID, CAPS_ASSESSMENTS_ASSIGNED.ASSESSMENT_ID ORDER BY
RES.CREATED ASC) AS Rank
FROM CAPS_PARTICIPANTS AS PAR
JOIN CAPS_PATIENTS AS P ON P.OWNER_PARTICIPANT_ID = PAR.PARTICIPANT_ID
JOIN CAPS_ASSESSMENTS_ASSIGNED ON PAR.PARTICIPANT_ID = CAPS_ASSESSMENTS_ASSIGNED.PARTICIPANT_ID
JOIN CAPS_ASSESSMENTS ON CAPS_ASSESSMENTS_ASSIGNED.ASSESSMENT_ID = CAPS_ASSESSMENTS.ASSESSMENT_ID
JOIN CAPS_RESPONSES AS RES ON RES.ASSESSMENT_ASSIGNED_ID =
CAPS_ASSESSMENTS_ASSIGNED.ASSESSMENT_ASSIGNED_ID
JOIN CAPS_PATIENTS_ASSIGNED AS PA ON PA.PATIENT_ID = P.PATIENT_ID
JOIN CAPS_THERAPISTS AS T ON T.THERAPIST_ID = PA.PROVIDER_ID
WHERE P.THERAPEUTIC_ALLIANCE_SCORE IS NOT NULL
```

```
# Random forest
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

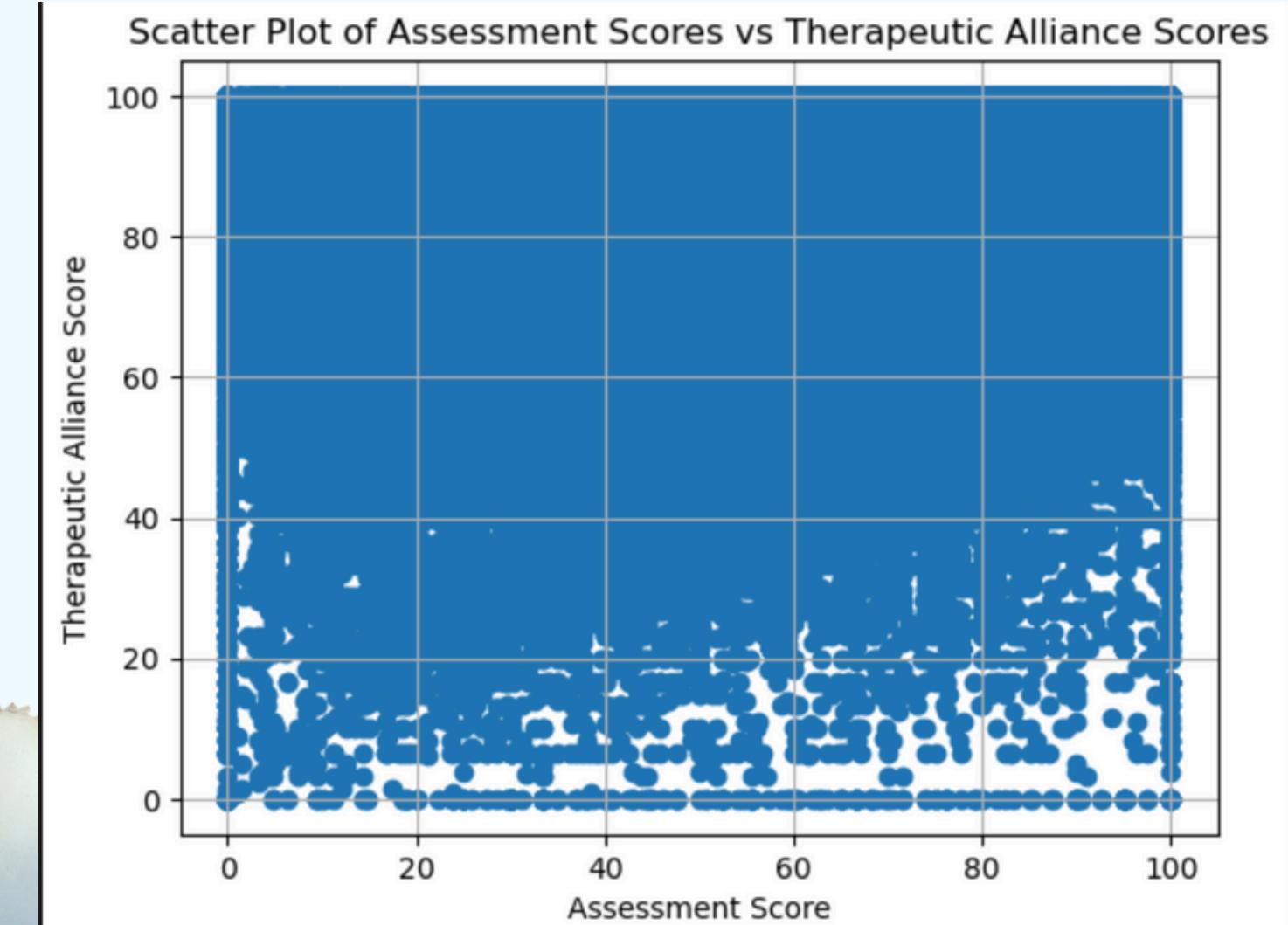
y_pred = model.predict(X_test)

# evaluation
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f'Mean Squared Error: {mse}')
print(f'R^2 Score: {r2}')

3m 17.8s
```

Mean Squared Error: 250.15098136523346
R² Score: 0.13900246778517888



Challenges

Currently, the calculated counts for completed assessments and missing assessments do not seem to match the number of assessments assigned to each person.

Next Step

- Try to fix the SQL query to ensure that the sum of the completed assessment count and the missing assessment count equals the total number of assessments.
- Build model to examine the relationship between counts of missing/completed assessments and TA.
- Attempt to identify the main diagnosis based on the types of assessments assigned to each patient.
- Try to determine diagnosis based on the most frequently assigned assessment.



Collaboration

Regular team meetings with occasional additional meetings, to discuss progress, challenges, and next steps.

Share regular updates on everyone's progress during these meetings.

Discuss insights from our meetings with Greenspace and split our tasks accordingly.

Match Score Logic

- high average TA (therapeutic alliance score)
- high average TA of clients with similar characteristics (demographic info, first assessment response or intake assessment)
- low caseload
- improvement within 6 sessions with patients of similar characteristics
- success with 5 similar patients based on their clinical outcomes



A thank you card featuring a central 'THANK YOU' message in a large, bold, dark blue serif font. The background is a light cream color with abstract blue washes at the top and bottom. Gold leaf patterns, including a large hexagon on the left and a large triangle on the right, frame the text. Small gold dots and circles are scattered throughout the design.

THANK YOU