What is behind depression?

Flore Perillat
Sophia Ouyang
Michelle Brier
Vincent Chen
Romain Kakko-Chiloff

Problems and Questions Addressed

- What is the profile of a depressed patient compared to a non-depressed patient?
- What is the profile of a patient with sleep disorder compared to a patient without sleep disorder?
- What is the profile of a patient with anxiety compared to a patient without anxiety disorder?
- How has the patient profile evolved from the first to the last clinical encounter?
- What are correlates of depression in terms of social habits?
- Is there any overlap in correlates of depression and those of sleep disorder and/or anxiety?
  - Do people with insomnia display similar social histories as people with depression?
  - Do people with anxiety display similar social histories as people with depression?
- Can we develop predictive models for depression?

Our Approach

- Reviewed data dictionary to get an overview of existing data features and features we would need to engineer
- Data preprocessing:
  - Merged features from demographics, social history, and problem list datasets based on patient ID into one dataset
  - Created new data features to determine whether patients' disorders were resolved, the duration of their treatment, and whether they were diagnosed with both mental and sleep disorders
- Performed data analysis and visualization on the general patient population as well as specific patients afflicted with mental disorders and sleep disorders
- Searched for correlation in medical and/or social histories between patients with different mental disorders

Architecture of Solution

1. **General demographic analysis to gain a general understanding of patient pool**
   - Broke down all patients' responses to social history questions based on age group and patient pool size
   - Compared demographics and social histories of patients who are currently deceased to patients who are still alive (age group, ethnicity, gender, smoker status)

Architecture of Solution

2. **Determined patient responses to clinical attention in terms of social habits in order to identify any differences in response between patients with different mental disorders**
   - Compared social habits between first and last medical appointments for all patients with mental disorders by analyzing patients' answers to social history questions during their first appointment and their last appointment

Architecture of Solution

3. **Focused on patients with depression, sleep disorder, and anxiety**
   - Distributions of time required to resolve their respective mental disorders
   - Resolve rate for patients with depression, sleep disorder, and anxiety
   - Overlap diagnoses for patients with depression, sleep disorder, and anxiety
   - Percentage of patients with depression, sleep disorder, and anxiety out of total patients with mental disorders
   - Correlation between these 3 mental disorders and alcohol consumption, drug use, sexual activity, and smoking

Architecture of Solution

4. **Developed predictive models for depression**
    ○   Models with only demographic features
    ○   Models with demographic features and anxiety diagnosis features
    ○   Models with demographic features and sleep disorder diagnoses features
    ○   Models with all of the above in addition to social habit features

The Data Files: Run & Choose

Demographics	Problem List	Encounters	Encounters diagnosis	Social History
Medication records	Medication Order	Medication Admin	Labs	Surgery
Surgery History	Billing Procedures	Procedures Orders		
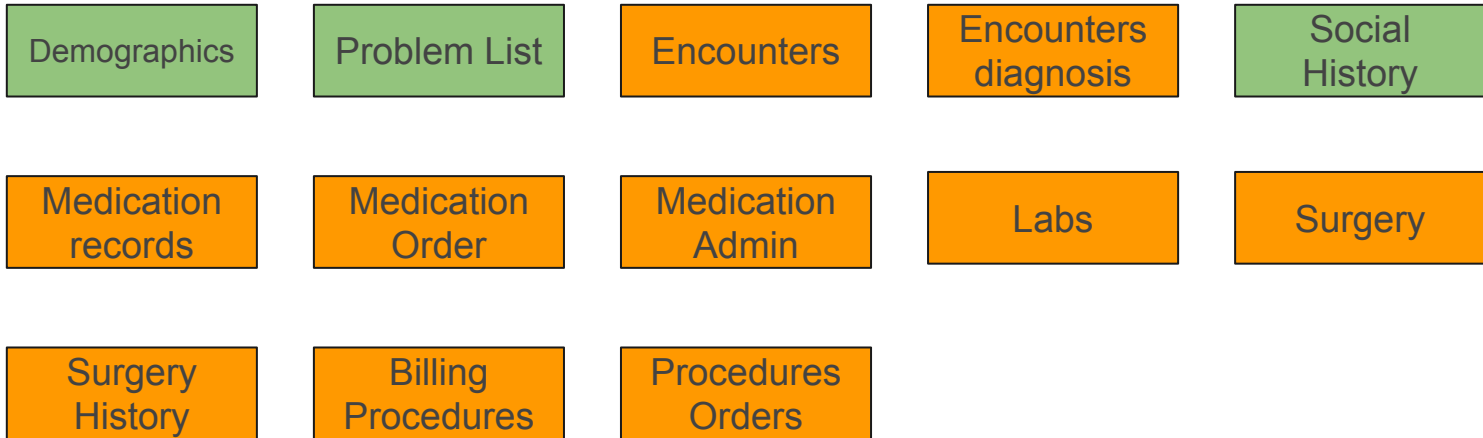
Data Preprocessing

- **Problem List** dataset
  - For each patient, all the problems he / she faced are listed
  - Feature to 1 or 0 based on whether or not the problems have been resolved
- Each problem is associated with an icd-10 code (a medical classification)
  - F32 → Depression
  - F41 → Anxiety
  - G47 → Sleep Disorders
- Now able to see the patients that are interesting for our study

studyid	prob_noted_dt_m	prob_resolved_dt_m	icd10	prob_dx_nm	
PT149256	2002-10-06	NaN	F32.9	DEPRESSIVE DISORDER, NOT ELSEWHERE CLASSIFIED	
PT149263	2008-12-25	NaN	F17.200	TOBACCO USE DISORDER	
PT149267	2015-11-03	2016-11-16	F43.23	ADJUSTMENT REACTION WITH ANXIETY AND DEPRESSION	
PT14927	2011-01-28	NaN	F32.9	DEPRESSIVE DISORDER, NOT ELSEWHERE CLASSIFIED	
PT149275	2007-03-25	NaN	F41.1	ANXIETY STATE	

Data Preprocessing

- **Demographics** dataset
  - Patients with Mood Disorders informations
  - Labeling categorical variables : sex, age, race, id

Age	Label_Age
0 - 25	0
25 - 40	1
40 - 60	2
60 +	3

Sex	Label_Sex
Female	0
Male	1
Unknown	2

Ethnicity	Label_Ethnicity
White	0
Latino	1
Black	2
Asian	3
Natives	4
Unknown	5

Data Preprocessing

- **Social History** dataset:
  - List all the people appointments and habits at the time of the appointments
  - Labeling categorical variables: drug use, alcohol consumption, sexual activity
  - Find age of people at the time of appointments
- Merge the datasets with first appointments, last appointments and most current habits

Smoker Status	Label_Smoker
No Smoker	0
Light Smoker	1
Heavy Smoker	2
Unknown	3

Final Results

1. *Demographics*

Number of Patients Over Time

Patients over time
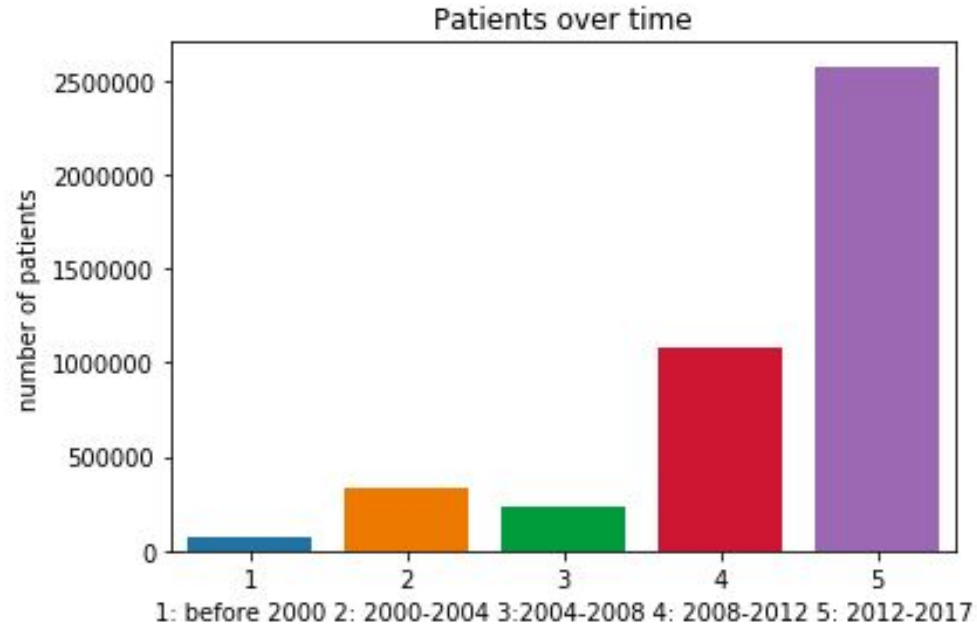
Age Repartition of Patients

Smoking Habits Depending on Age

Smoking habits

0: Age 0-25
1: Age 25-40
2: Age 40-60
3: Age 60+
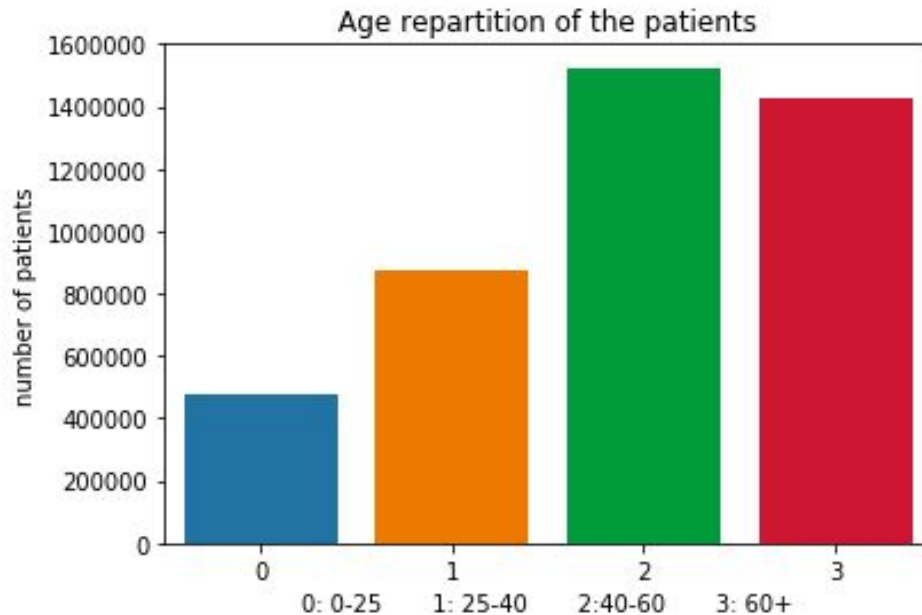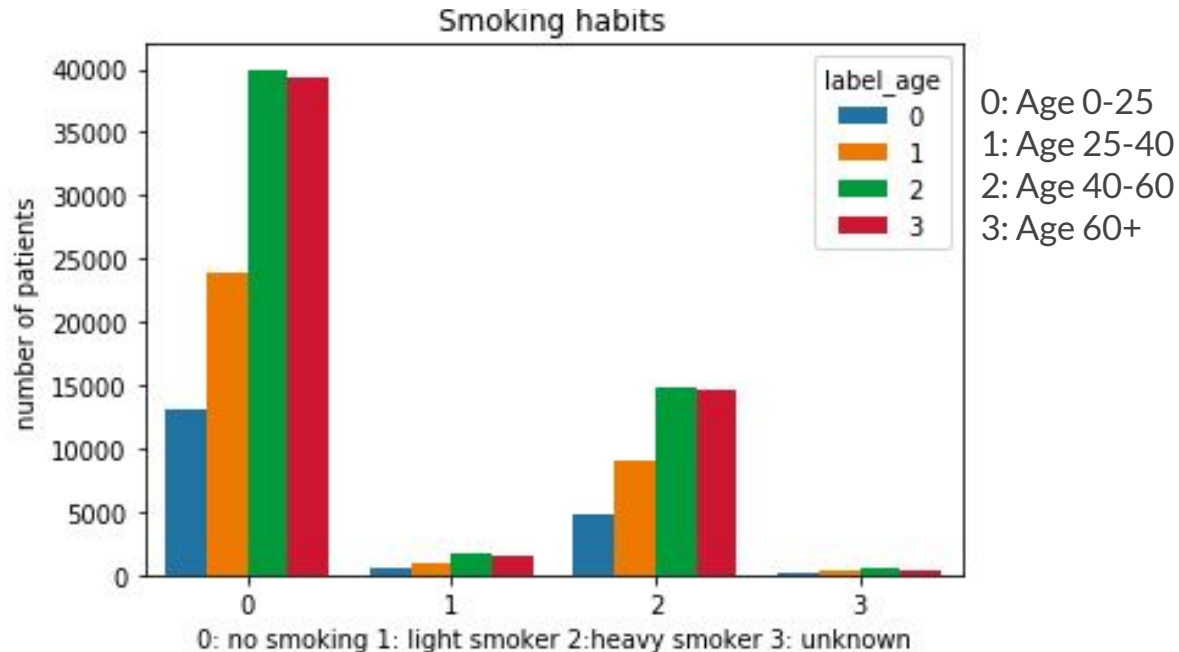
Drinking Habits Depending on Age
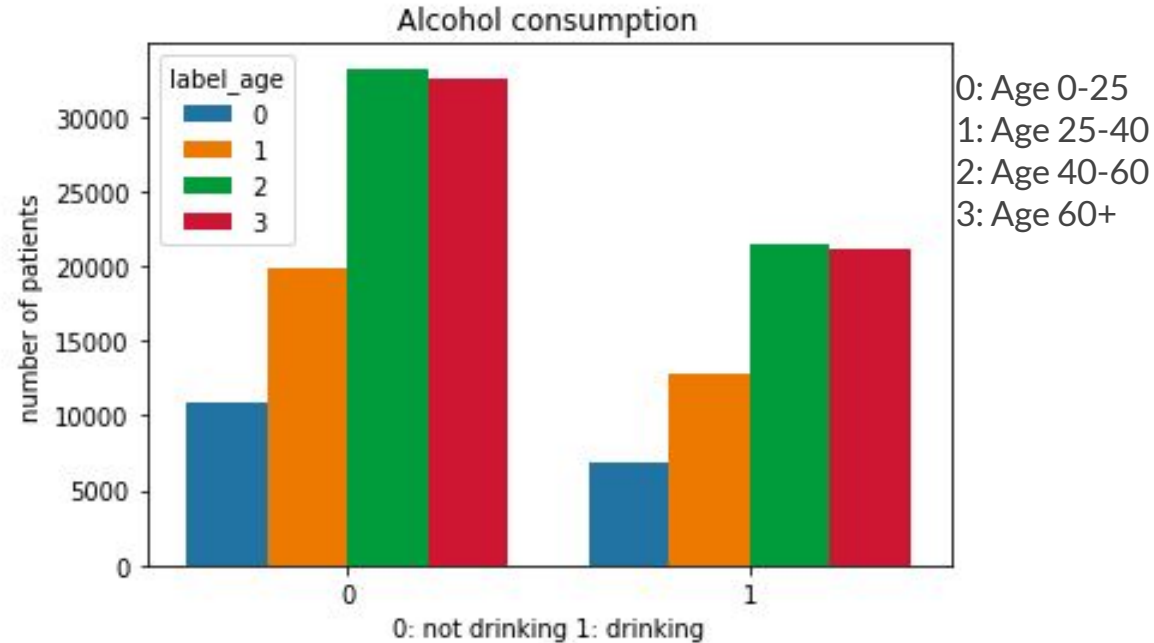
Drug Consumption Depending on Age

Drug consumption

0: Age 0-25
1: Age 25-40
2: Age 40-60
3: Age 60+
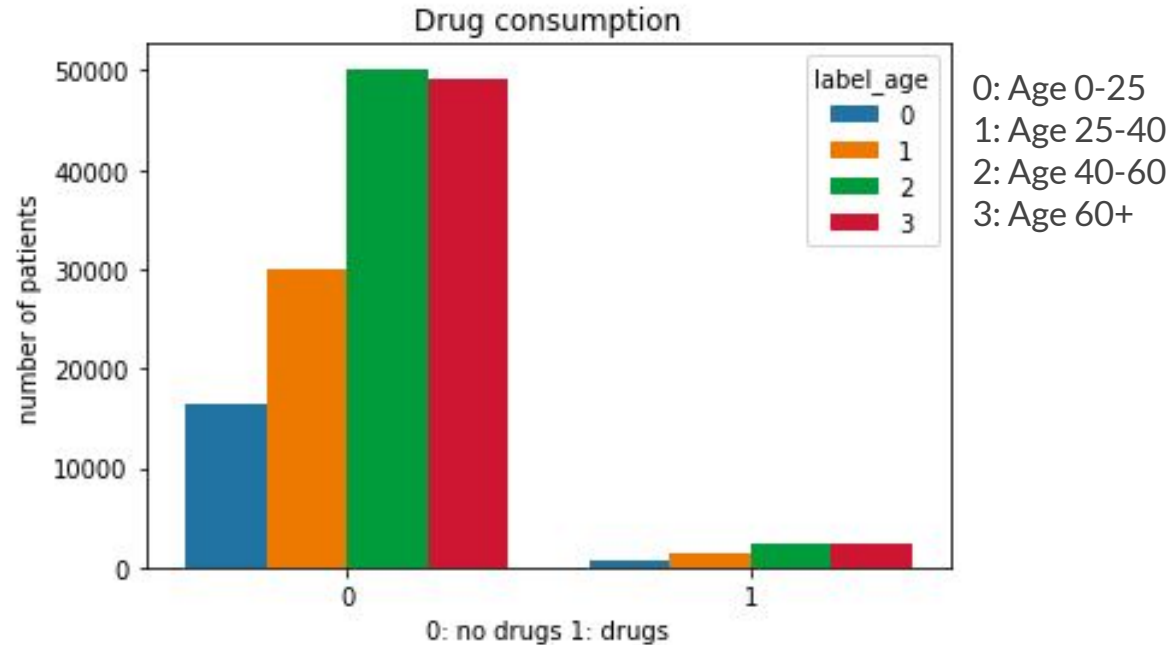
Sexual Activity Depending on Age

Sexual activity

0: Age 0-25
1: Age 25-40
2: Age 40-60
3: Age 60+
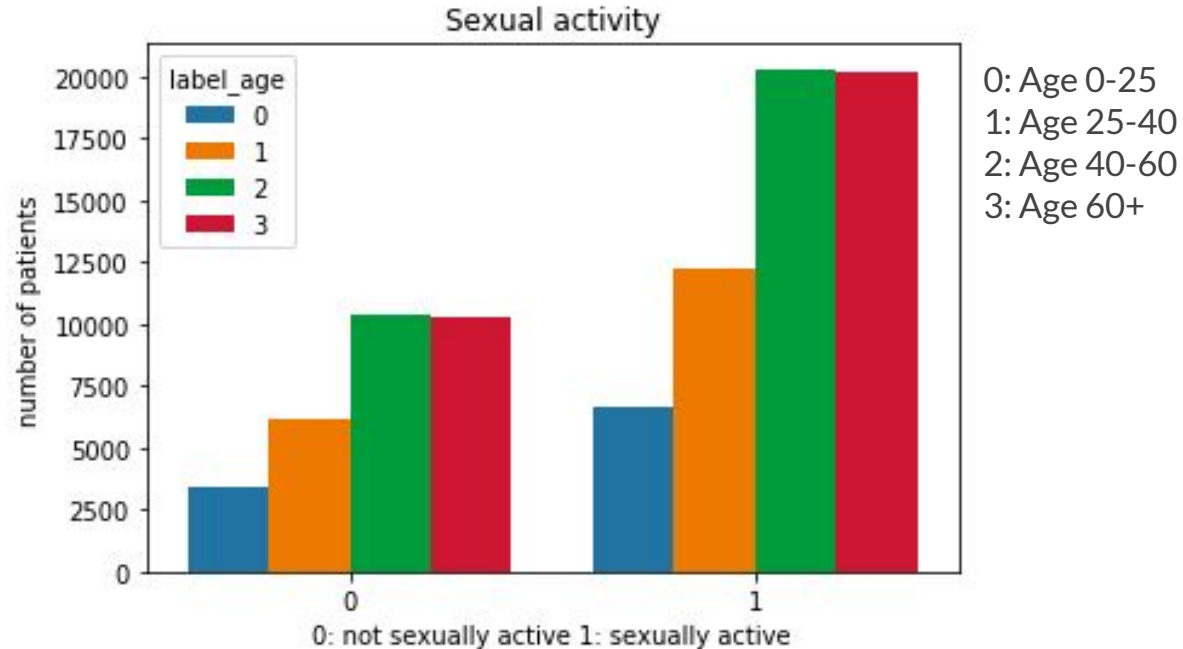
Deceased vs Alive Patients: Age
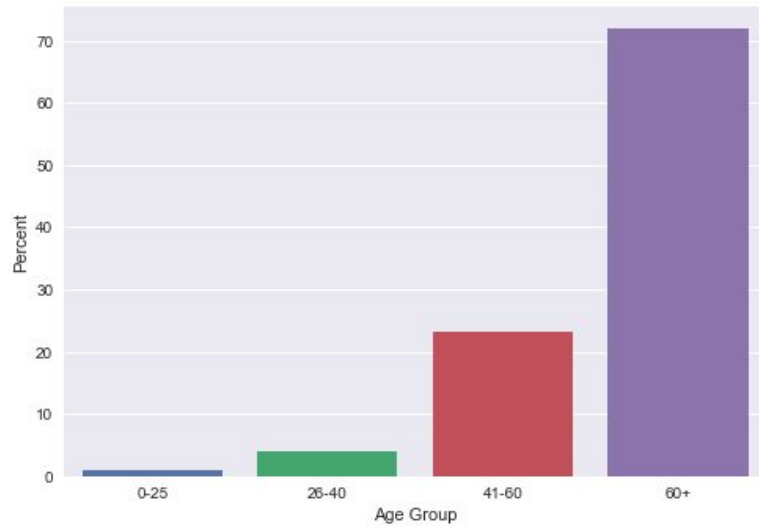
Deceased
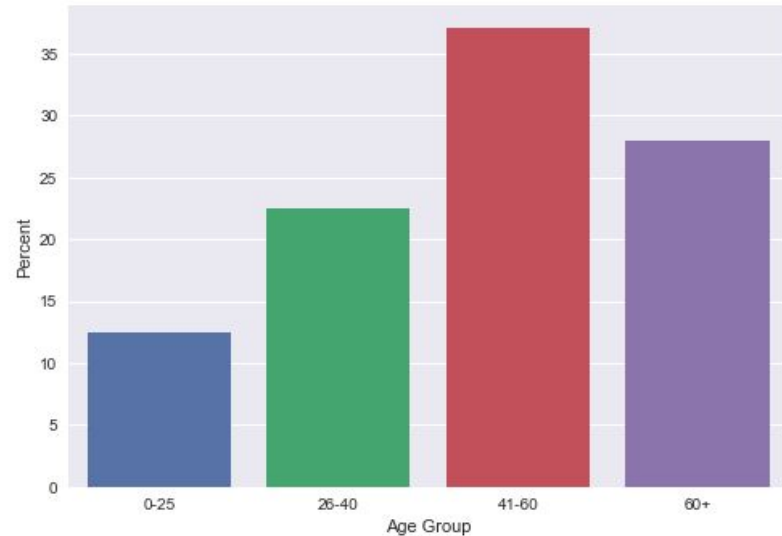
Alive

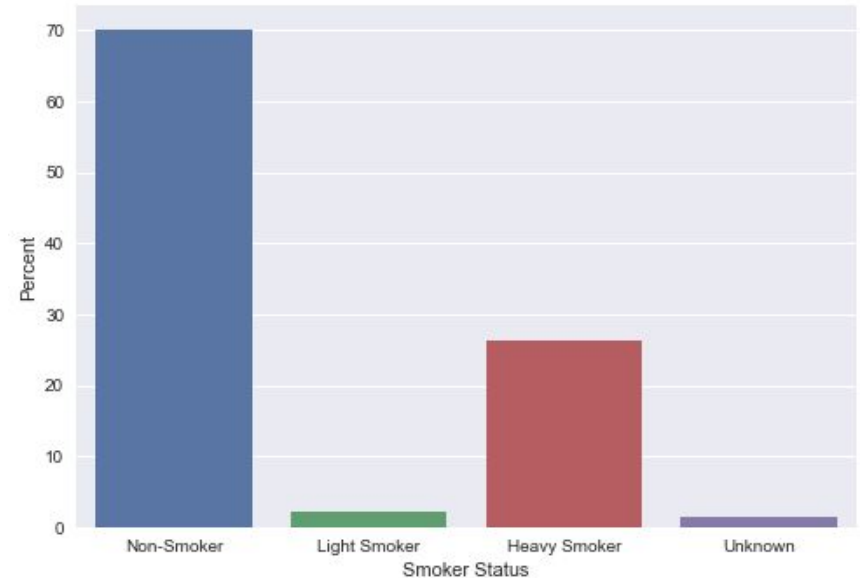# Deceased vs Alive Patients: Smoker Status
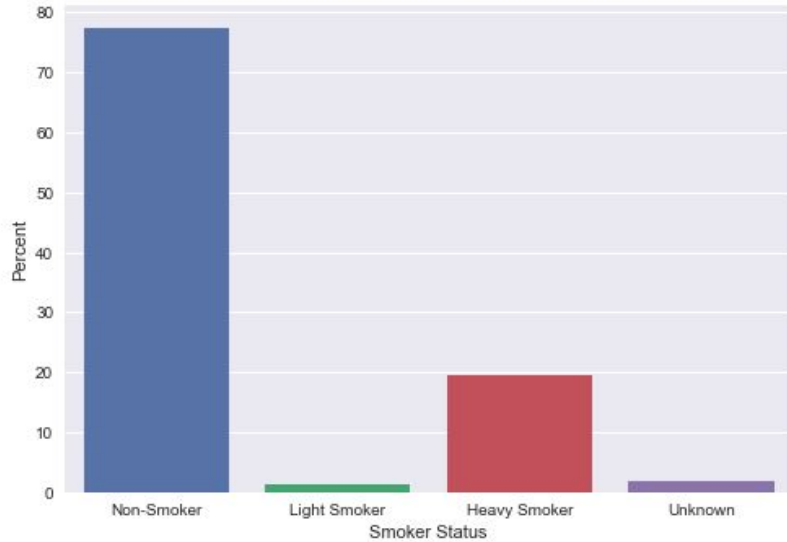
Deceased
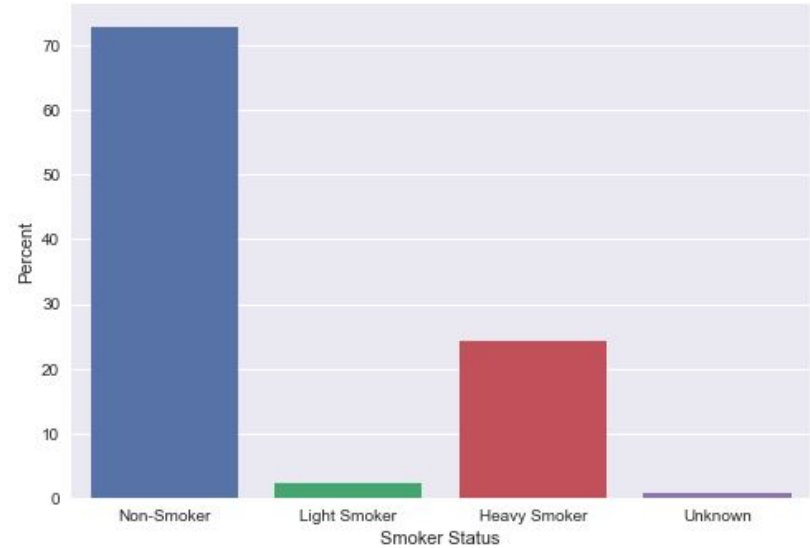
Alive

Final Results

2. *Change in Social Habits Between First and Last Clinical Appointments*

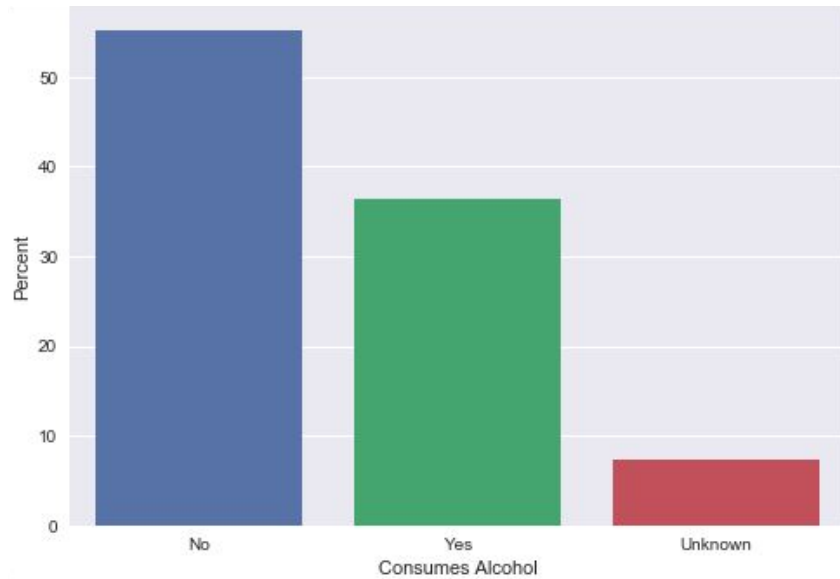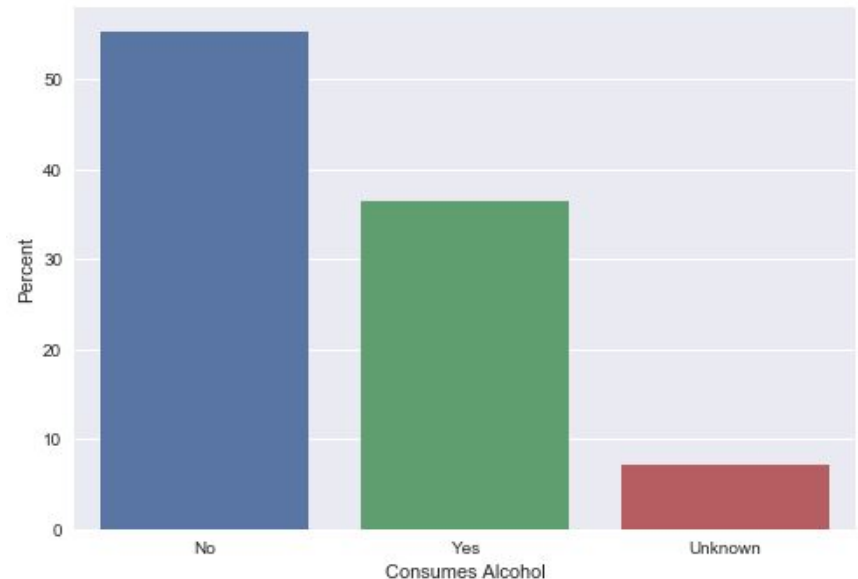# First vs Last Appt: Smoker Status
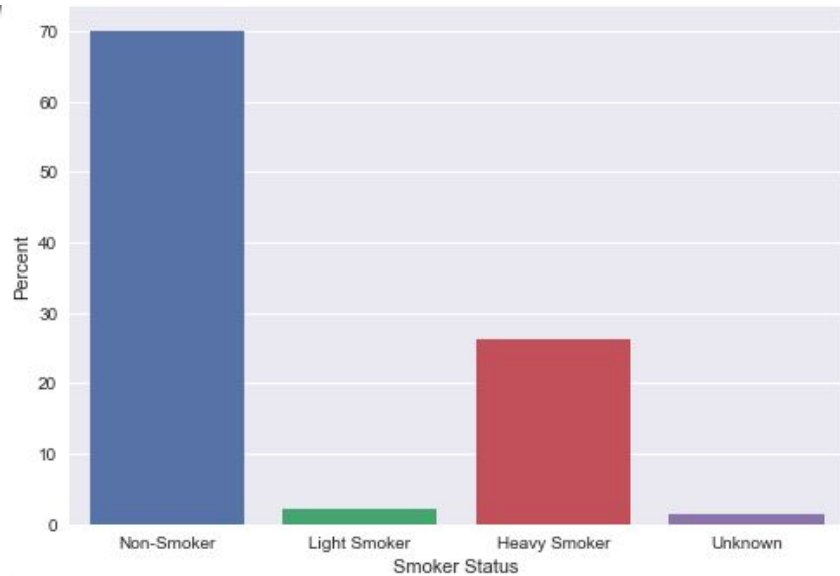
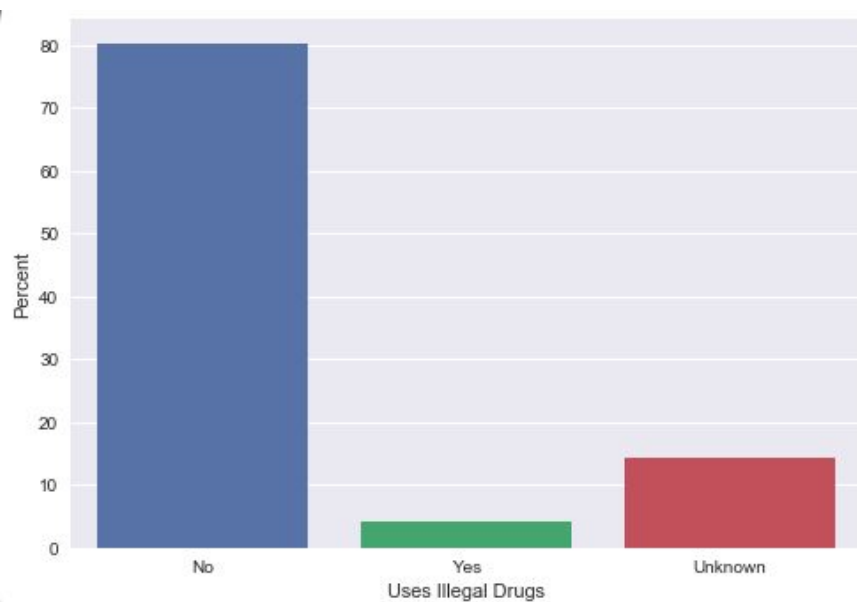First

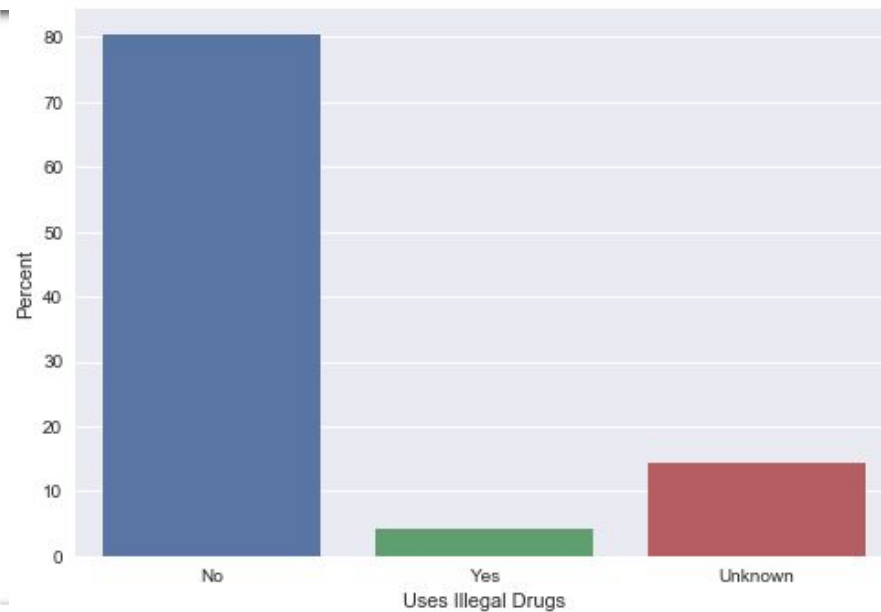Last

First vs Last Appt: Consumes Alcohol

First

Last

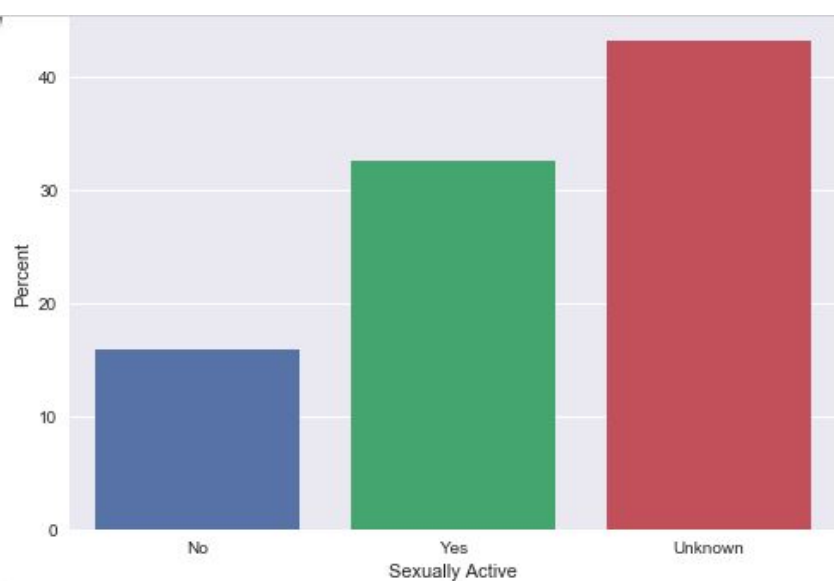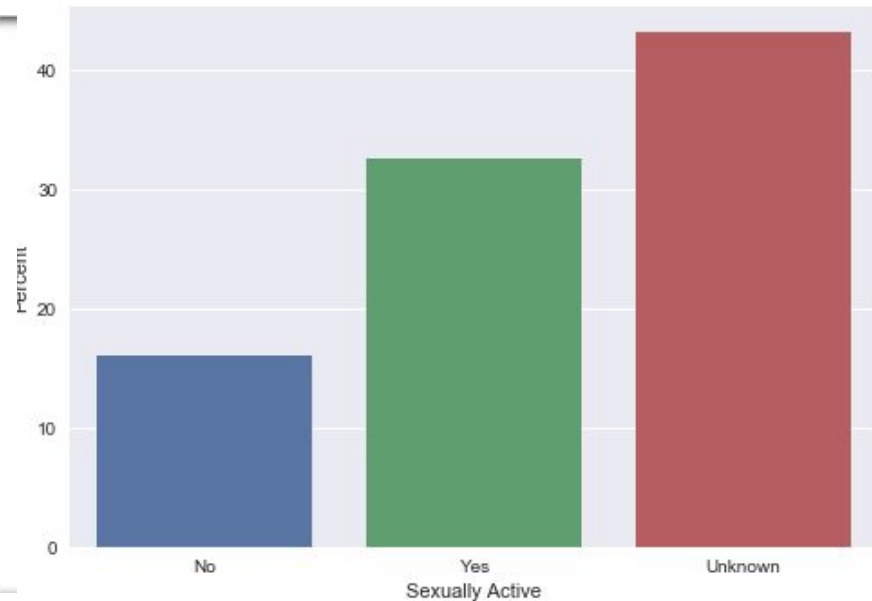# First vs Last Appt: Uses Illegal Drugs

First

Last

First vs Last Appt: Sexually Active

First

Last

First vs Last Appointment: Conclusion

Conclusion: No change at all in responses to social habit questions (smoker status, alcohol consumption, illegal drugs, and sexual activity) between the first and last clinical appointments when analyzing all patient responses, so decided not to break down analysis into subgroups based on mental disorder.

Social habits for all mental disorders were unaffected by clinical treatment.
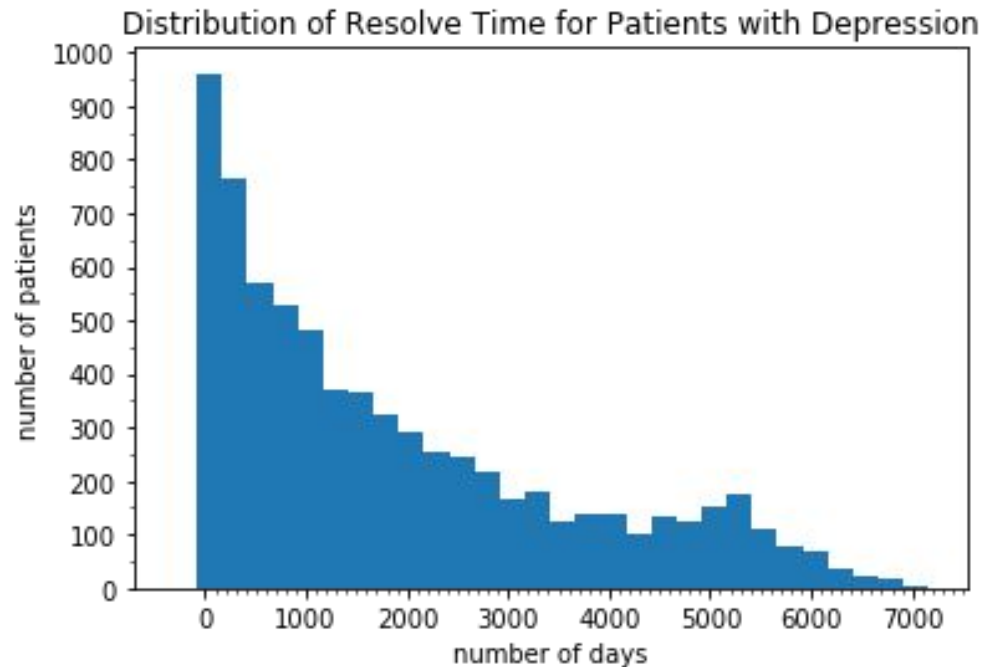
Final Results

3. Patients with Depression, Sleep Disorder, and Anxiety

Cleaning and Analyzing Problem List Data

- Group patients by disorder, diagnostic code and problem status
- Look at time it took for patients diagnosed with depression/sleep disorder/anxiety to have problem resolved
- Calculate percentage of patients who are diagnosed with depression/sleep disorder/anxiety and have problem marked as resolved
- Find percentage of people who were also diagnosed with sleep disorder/anxiety among those diagnosed with depression

Time to Resolve Depression

Distribution of Resolve Time for Patients with Depression

Time to Resolve Sleep Disorder

Distribution of Resolve Time for Patients with Sleep Disorder

Time to Resolve Anxiety
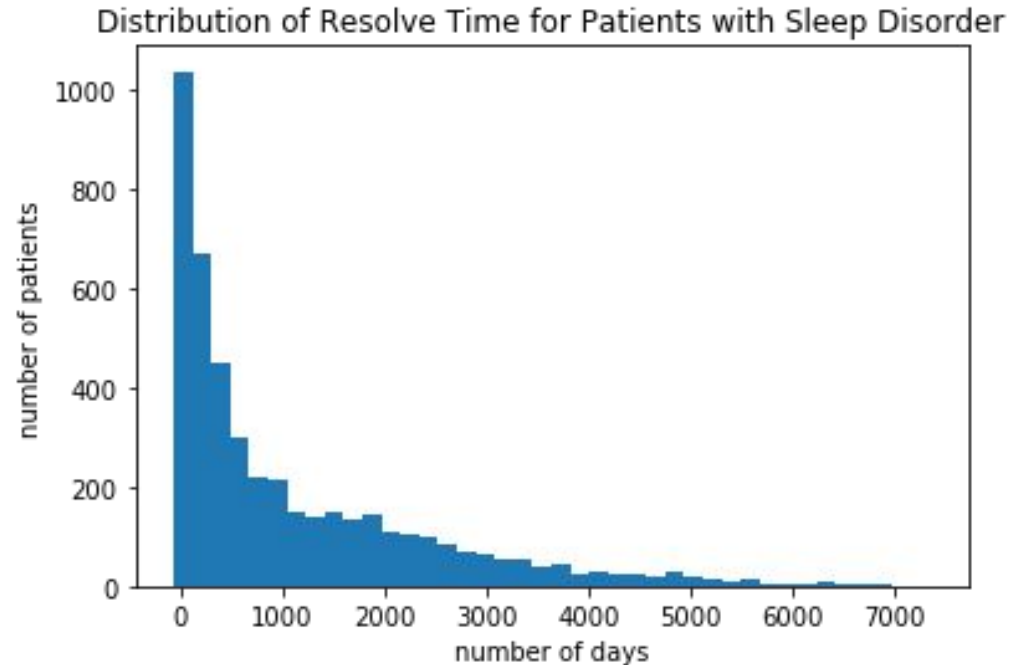
Distribution of Resolve Time for Patients with Anxiety
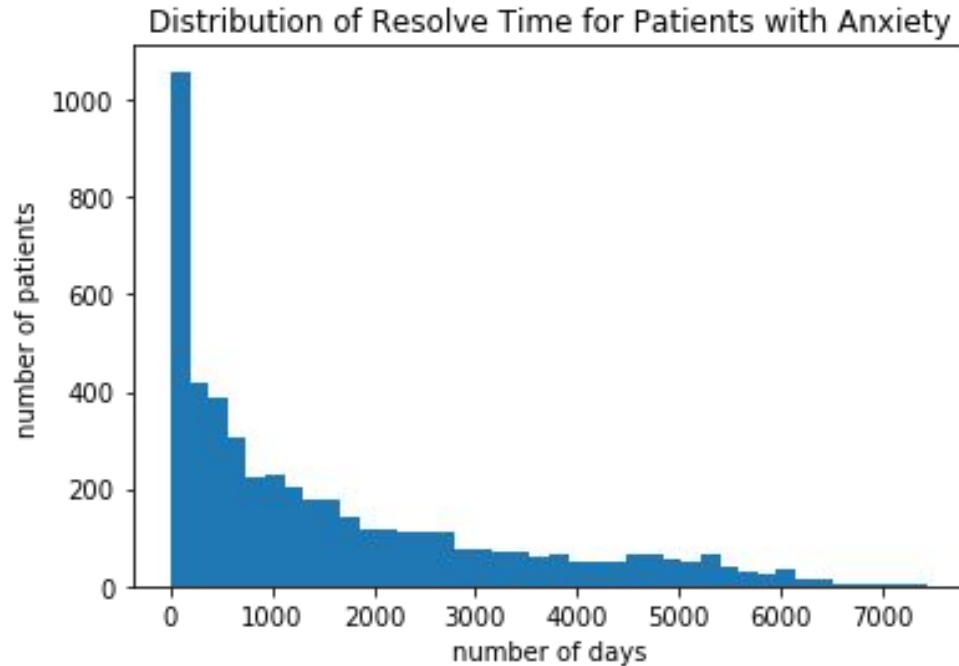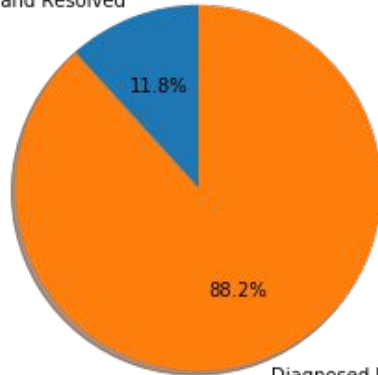
Resolve Rate for Depression, Sleep Disorder, and Anxiety

Patients Diagnosed with Depression
Diagnosed and Resolved
11.8%
88.2%
Diagnosed but not resolved

Patients Diagnosed with Sleep Disorder
Diagnosed and Resolved
15.2%
84.8%
Diagnosed but not resolved
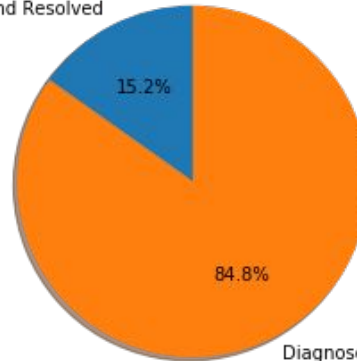
Patients Diagnosed with Anxiety
Diagnosed and Resolved
5.5%
94.5%
Diagnosed but not resolved
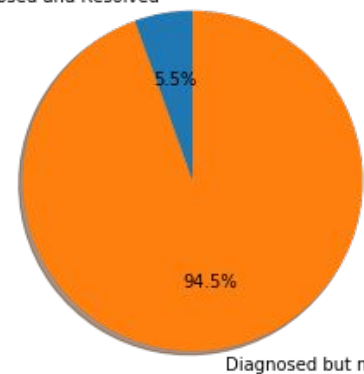
Patients Diagnosed with Depression

Patients Diagnosed with Depression

Also Diagnosed with Anxiety

35.7%

64.3%

No Anxiety

Patients Diagnosed with Depression

Also Diagnosed with Sleep Disorder

24.2%

75.8%

No Sleep Disorder
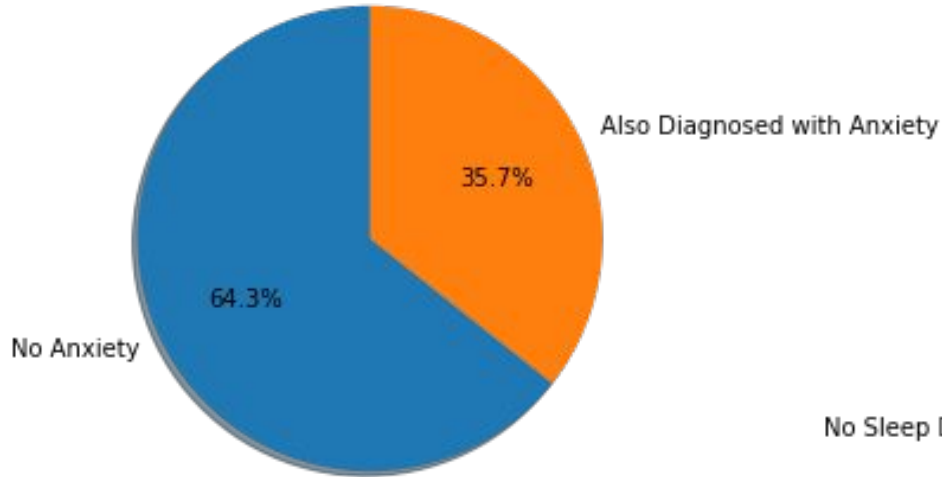
Patients Diagnosed with Depression
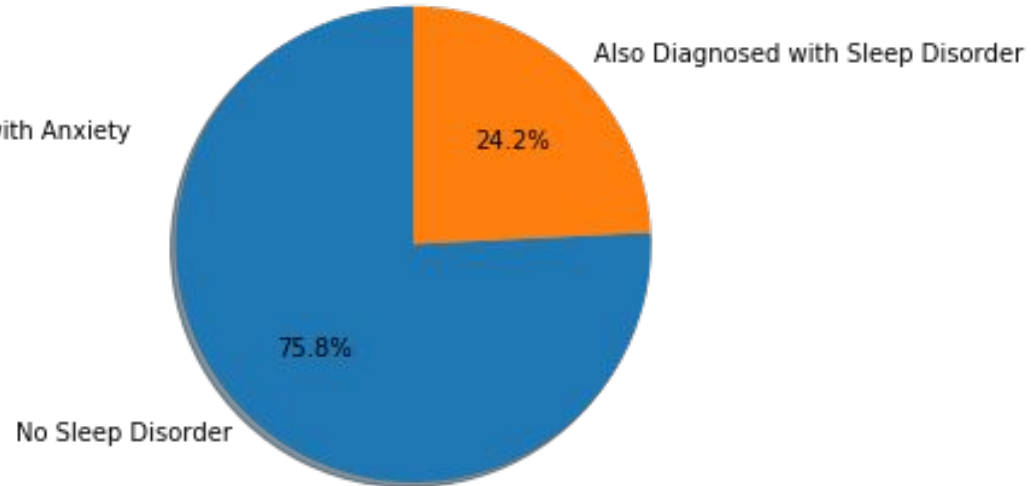
Patients Diagnosed with Depression
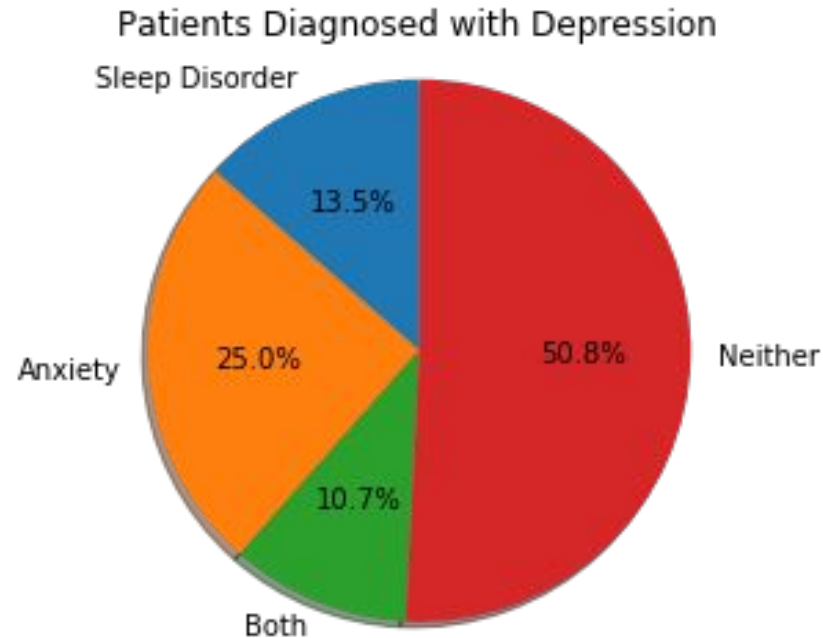
Depression and Correlates

Searching for correlation between depression and alcohol consumption, drug use, sexual activity, and smoker status

Behavior of Depressed Patients Compared to Non Depressed Patients

Depression and Correlates: Overall Findings

Depressed Patients (including resolved and unresolved)

➤ A slightly higher percentage of depressed patients use illegal drugs and heavily smoke compared to non-depressed patients

Sleep Disorder and Correlates

Searching for correlation between sleep disorder and alcohol consumption, drug use, sexual activity, and smoker status

Behavior of Patients with Sleep Disorder Compared to Patients with No Sleep Disorder

Behaviours depending on the sleep disorder
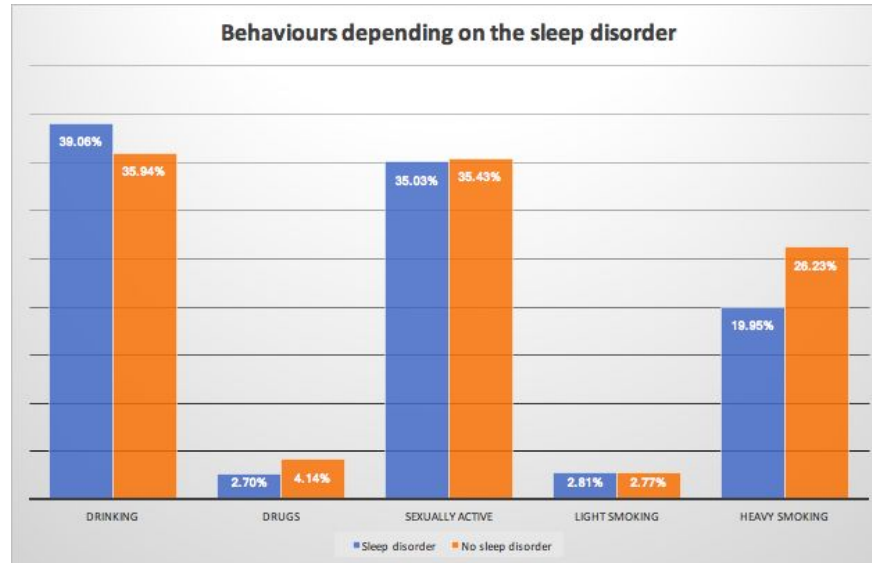
Sleep Disorder and Correlates: Overall Findings

Patients with Sleep Disorder (including resolved and unresolved)

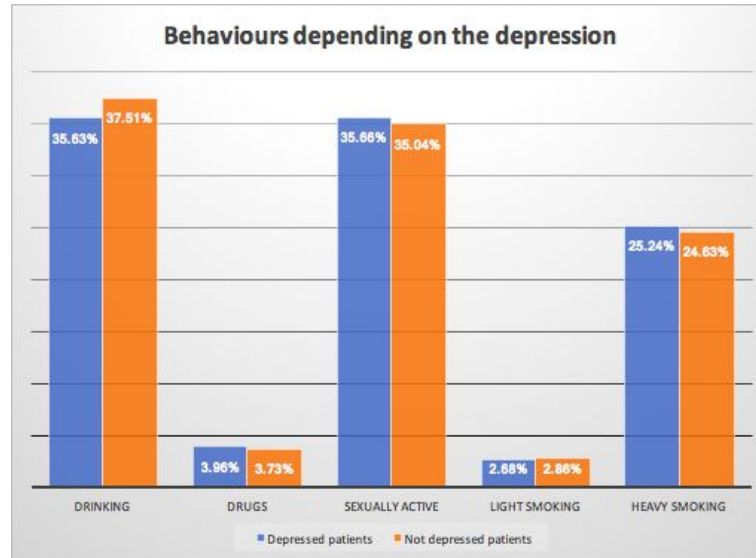→ A higher percentage of patients with sleep disorder consume alcohol compared to patients without sleep disorder

Anxiety and Correlates

Searching for correlation between anxiety and alcohol consumption, drug use, sexual activity, and smoker status

Behavior of Anxious Patients Compared to Non-Anxious Patients

Behaviours depending on anxiety

DRINKING: Anxious 38.33%, Non anxious 34.09%
DRUGS: Anxious 3.28%, Non anxious 4.84%
SEXUALLY ACTIVE: Anxious 38.60%, Non anxious 33.55%
LIGHT SMOKING: Anxious 2.78%, Non anxious 2.77%
HEAVY SMOKING: Anxious 24.03%, Non anxious 26.22%

■ Anxious ■ Non anxious
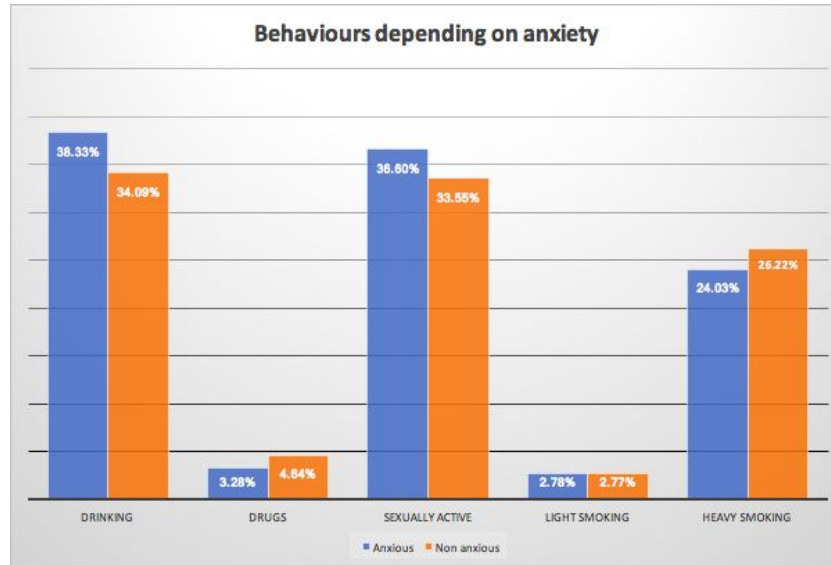
Anxiety and Correlates: Overall Findings

Patients with Anxiety (including resolved and unresolved)

A higher percentage of patients with anxiety consume alcohol and are sexually active compared to patients without anxiety disorder

Depression, Sleep Disorder, & Anxiety and Correlates: Conclusions

- **Depression Correlates:** Drug use, heavy smoking (very weak correlation)
- **Sleep Disorder Correlates:** Consume alcohol
- **Anxiety Correlates:** Consume alcohol, more sexually active

Conclusion: No strong, conclusive correlations found between social habits and depression, sleep disorder, or anxiety and no overlap in social indicators between depression and sleep disorder or depression and anxiety

Final Results

4. Predictive Models for Depression

Predicting Depression
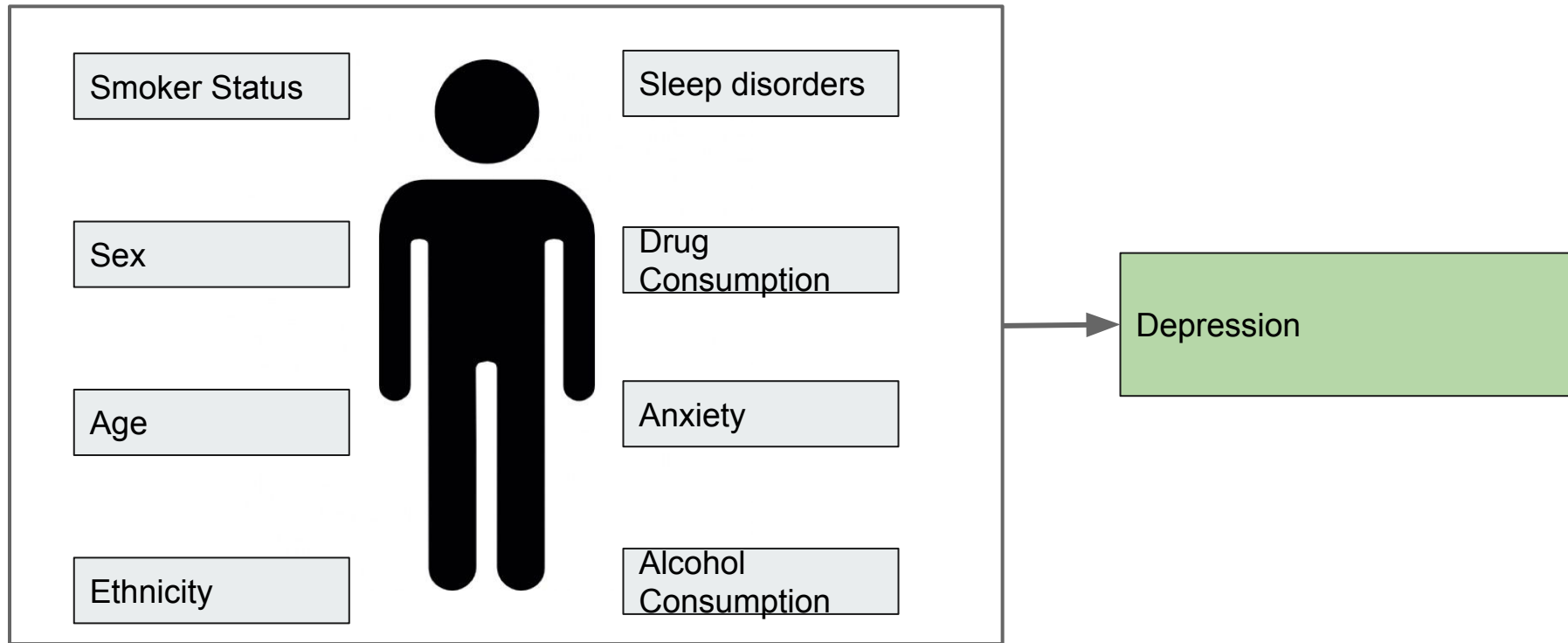
Predicting Depression

Table with results for models

Models	Test Accuracy
Logistic Regression	72.37 %
k-NN	68.05 %
Perceptron	69.63 %
XG Boost	72.49 %
Random Forest	72.36 %

Test Accuracy vs the different methods
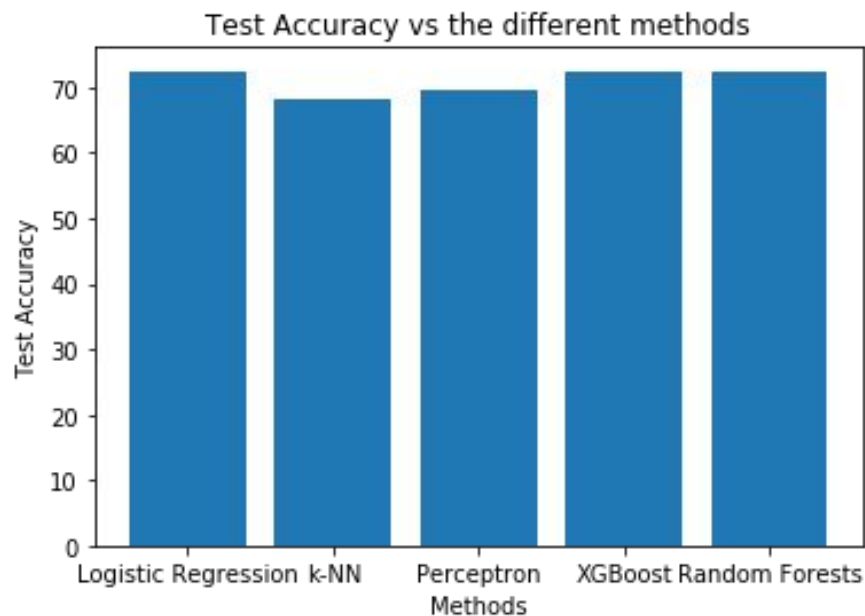
Model with only Demographic features

Low accuracies indicate that demographics do not strongly predict depression

```python
X = final[['label_ethnicity',
           'label_sex', 'label_age']]
Y = final['isDepression']
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.20)
for n in range(0, len(test)):
    model = test[n]
    model.fit(X_train, Y_train)
    print(name[n])
    print("Training accuracy: ", model.score(X_train, Y_train))
    print("Test accuracy: ", model.score(X_test, Y_test))
```

```
Logistic Regression
Training accuracy:  0.54386638148
Test accuracy:  0.54267740312
SVC
Training accuracy:  0.547754151988
Test accuracy:  0.5504781077
Perceptron
Training accuracy:  0.490953699044
Test accuracy:  0.49411172622
KNN
Training accuracy:  0.497005535984
Test accuracy:  0.501233014595
XGB
Training accuracy:  0.547867388022
Test accuracy:  0.550427780574
Random Forest
Training accuracy:  0.547924006039
Test accuracy:  0.550176144942
```

Model with only Sleep Disorder and Demographic features

Low accuracies indicate that sleep disorder also does not strongly predict depression

```python
X = final[['isSleepDisorder', 'isActiveSleepDisorder','label_ethnicity',
           'label_sex', 'label_age']]
Y = final['isDepression']
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.20)
for n in range(0, len(test)):
    model = test[n]
    model.fit(X_train, Y_train)
    print(name[n])
    print("Training accuracy: ", model.score(X_train, Y_train))
    print("Test accuracy: ", model.score(X_test, Y_test))
```

```
Logistic Regression
Training accuracy:  0.554365878208
Test accuracy:  0.554051333669
SVC
Training accuracy:  0.554994967287
Test accuracy:  0.553925515853
Perceptron
Training accuracy:  0.491696024157
Test accuracy:  0.491142425767
KNN
Training accuracy:  0.524458983392
Test accuracy:  0.522898842476
XGB
Training accuracy:  0.555171112229
Test accuracy:  0.555435329643
Random Forest
Training accuracy:  0.555712128837
Test accuracy:  0.555737292401
```

Model with only Anxiety and Demographic features

Higher accuracies indicate that anxiety may better predict depression than sleep disorders

```python
X = final[['isAnxiety', 'isActiveAnxiety','label_ethnicity',
           'label_sex', 'label_age']]
Y = final['isDepression']
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.20)
for n in range(0, len(test)):
    model = test[n]
    model.fit(X_train, Y_train)
    print(name[n])
    print("Training accuracy: ", model.score(X_train, Y_train))
    print("Test accuracy: ", model.score(X_test, Y_test))
```

```
Logistic Regression
Training accuracy:  0.73224710619
Test accuracy:  0.731303472572
SVC
Training accuracy:  0.733543029693
Test accuracy:  0.732360342224
Perceptron
Training accuracy:  0.719929542023
Test accuracy:  0.718746854555
KNN
Training accuracy:  0.636273276296
Test accuracy:  0.639003522899
XGB
Training accuracy:  0.733498993457
Test accuracy:  0.732511323603
Random Forest
Training accuracy:  0.733794665325
Test accuracy:  0.732536487167
```

Learning Path

Early/Mid-Semester

- Appointment with mentor, Professor Roberto Zicari, to learn about project specifications and datasets we would be working with
- Reviewed existing medical literature to educate ourselves on factors leading to depression and insomnia -- all reported strong link between Major Depressive Disorder and symptoms of sleep disorder

Learning Path

Mid-Semester/End of Semester

- Obtained dataset from Professor Zicari (in partnership with Geisinger)
- Skype appointment with Professor Maurice Ohayon from Stanford to get ideas on what features of data to explore and periodic check ins with Professor Zicari
- Reviewed data analysis and data visualization tools introduced in class to determine which tools could most benefit our research
- Preprocessed datasets
- Analyzed demographics, searched for correlations between mental disorders and associated social/medical histories, and created data visualizations
- Expanded breadth of search to anxiety disorder in addition to sleep disorder after preliminary analysis

Learning Path

Present

- Unable to find strong social correlates of depression, anxiety, and sleep disorder and no overlap between the three mental disorders in the correlates found
- Found a considerable correlation between patients diagnosed with depression and anxiety (46%) and depression and sleep disorder (24.2%)
- Developed predictive models for depression and identified which features were more important than others when predicting depression

To be continued

- Link our study with the medicines that people take

- Deepen our work while adding insights on Time-Series Analysis

- Go through the other datasets : information in other datasets is useful for studying other correlations

- Develop better predictive and classifier models using feature engineering from deep learning

GitHub Repository

https://github.com/michellebrier/predict-mood-disorders-datax