Homework 2

Aaryan Agarwal Fall 2023

# Part 1: Data Modelling

# About Data

The dataset that will be used for this homework assignment is an OCD Patient Dataset: Demographics & Clinical Data. It is a comprehensive collection of information pertaining of 1500 individuals diagnosed with OCD. Included in this dataset are key demographic details such as age, gender, ethnicity, marital status, and education level, offering a comprehensive overview of the sample population. Additionally, clinical information like the date of OCD diagnosis, duration of symptoms, and any previous psychiatric diagnoses are recorded, providing context to the patients' journeys. The dataset also delves into the specific nature of OCD symptoms, categorizing them into obsession and compulsion types. Severity of these symptoms is assessed using the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) scores for both obsessions and compulsions. Furthermore, it documents any co-occurring mental health conditions, including depression and anxiety diagnoses. Notably, the dataset outlines the medications prescribed to patients, offering valuable insights into the treatment approaches employed. It also records whether there is a family history of OCD, shedding light on potential genetic or environmental factors. Overall, this dataset serves as a valuable resource for researchers, clinicians, and mental health professionals seeking to gain a deeper understanding of OCD and its manifestations within a diverse patient population. The dataset was collected form Kaggle: [https://www.kaggle.com/datasets/ohinhaque/ocd-patient-dataset-demographics-and-clinical-data/data](https://www.kaggle.com/datasets/ohinhaque/ocd-patient-dataset-demographics-and-clinical-data/data%20)

# Schema:

A diagram of a system

Description automatically generated

# Part 2

# Cypher code:

// Creating nodes for patients and their attributes

LOAD CSV WITH HEADERS FROM

"file:///D:/Programs/Personal/MSDS/ADT/Homework2/archive/ocd\_patient\_dataset.csv" AS row

RETURN row

LIMIT 5

No constraints can be created as a single person has multiple different records.

//creating nodes and relationships.

LOAD CSV WITH HEADERS FROM

"file:///D:/Programs/Personal/MSDS/ADT/Homework2/archive/ocd\_patient\_dataset.csv"

AS row

WITH row

MERGE (p:Patient {id: row.ID, age: toInteger(row.Age), gender:row.Gender, education\_level:COALESCE(row.EducationLevel, 'Unknown')})

MERGE (f:Family\_History {value: row.FamilyHistoryofOCD})

MERGE (s:Symptoms {depression: row.DepressionDiagnosis,anxiety: row.AnxietyDiagnosis})

MERGE (m:Medications {medications: row.Medications})

MERGE (c:Compulsions {type: row.CompulsionType})

MERGE (o:Obsession {type: row.ObsessionType})

MERGE (pd:Previous\_Diagnosis {name: row.PreviousDiagnoses})

MERGE (d:Diagnosis\_Details {date: row.OCDDiagnosisDate,duration: row.DurationofSymptoms})

MERGE (p)-[:HAS\_FAMILY\_HISTORY]->(f)

MERGE (p)-[:HAS\_SYMPTOMS]->(s)

MERGE (p)-[:HAS\_MEDICATION]->(m)

MERGE (p)-[:HAS\_COMPULSION{score:tointeger(row.YBOCSScoreCompulsions)}]->(c)

MERGE (p)-[:HAS\_OBSESSION{score:tointeger(row.YBOCSScoreObsessions)}]->(o)

MERGE (p)-[:IS\_DIAGNOSED]->(d)

MERGE (p)-[:DIAGNOSED\_PREVIOUSLY]->(pd)

# Part 3

Call db.schema.visualization()

A screenshot of a computer

Description automatically generated

# Nodes:

Patient: id, gender, age, education\_level

A screenshot of a computer

Description automatically generatedCompulsions: type

A screenshot of a computer

Description automatically generated

Diagnosis\_Details: date, duration

A screenshot of a computer

Description automatically generated

Family\_History: value

A screenshot of a computer

Description automatically generated

Medications: medications

A screenshot of a computer

Description automatically generated

Obsession: type

A screenshot of a computer

Description automatically generated

Previous\_Diagnosis: name

A screenshot of a computer

Description automatically generated

Symptoms: anxiety, depression

A screenshot of a computer

Description automatically generated

# Relationships:

DIAGNOSED\_PREVIOUSLY: Relationship between Patient and Diagnosed\_Previously

A computer screen shot of a computer

Description automatically generated

HAS\_COMPULSION: Relationship between Patient and Compulsion with and OCD compulsions score

A screenshot of a computer

Description automatically generated

HAS\_FAMILY\_HISTORY: Relationship between Patient and Family\_History

A screenshot of a computer

Description automatically generated

HAS\_MEDICATION: Relationship between Patient and Medication

A screenshot of a computer

Description automatically generated

HAS\_OBSESSION: Relationship between Patient and Obsession with an OCS Obsession score

A screenshot of a computer

Description automatically generated

HAS\_SYMPTOMS: Relationship between Patient and Symptoms

A screenshot of a computer

Description automatically generated

IS\_DIAGNOSED: Relationship between Patient and Diagniosis\_Details

A screenshot of a computer

Description automatically generated

# Query 1

Patients with obsession score of more than 31 and need medical treatment.

A screenshot of a computer

Description automatically generated

# Query 2

ID of patients having multiple treatment records.

A screenshot of a computer

Description automatically generated

# Query 3

Number of patients with respect to each previous disorder

A screenshot of a computer

Description automatically generated

# Query 4

Number of patients with or without family history on Benzodiazepine

A screenshot of a computer program

Description automatically generated

# Query 5

Number of patients with respective compulsion disorders who need medical treatment

A screenshot of a computer

Description automatically generated