OCD Patient Dataset: Demographics & Clinical Data

Tools used in building the project:

1. Excel
2. SQL Server
3. Power BI

Objective of the Project:

The goal of this project is to perform an exploratory data analysis (EDA) on a dataset containing demographic and clinical data of OCD patients. The analysis will focus on understanding the relationships between various demographic factors and clinical outcomes.

Dataset Overview:

The dataset includes the following columns:

- Patient ID: Unique identifier for each patient.
- Age: Age of the patient.
- Gender: Gender of the patient.
- Ethnicity: Ethnicity of the patient.
- Marital Status: Marital status of the patient.
- Education Level: Level of education attained by the patient.
- OCD Diagnosis Date: Date when OCD was diagnosed.
- Duration of Symptoms (months): Duration for which the patient has been experiencing symptoms.
- Previous Diagnoses: Any previous diagnoses before OCD.
- Family History of OCD: Whether the patient has a family history of OCD.
- Obsession Type: Type of obsessions experienced by the patient.
- Compulsion Type: Type of compulsions experienced by the patient.
- Y-BOCS Score (Obsessions): Y-BOCS score related to obsessions.
- Y-BOCS Score (Compulsions): Y-BOCS score related to compulsions.
- Depression Diagnosis: Whether the patient has been diagnosed with depression.
- Anxiety Diagnosis: Whether the patient has been diagnosed with anxiety.
- Medications: Medications the patient is currently taking

Steps involved in building the project:

1. Study the dataset visually in Excel

2. Create a new date table in excel using the OCD diagnosis Date
   Year, Quarter, Month, Month number, Day, Day of Week and Day number are extracted in different columns.

3. Import the tables in SQL Server Management Studio

4. Verify the data imported in the OCD_Patient_Database
   SQL Query:
   SELECT * FROM OCD_Data;
   SELECT * FROM OCD_Date;

5. Clean the Data and prepare for export

   a. **CHECKING FOR NULL VALUES IN EACH OF THE COLUMNS**
      SQL Query:

```
-- Checking for NULL values in Age
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Age IS NULL;

Checking SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Gender IS NULL;

-- Checking for NULL values in Ethnicity
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Ethnicity IS NULL;

-- Checking for NULL values in Marital_Status
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
```

```
WHERE Marital_Status IS NULL;

-- Checking for NULL values in Education_Level
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Education_Level IS NULL;

-- Checking for NULL values in OCD_Diagnosis_Date
SELECT COUNT(*) AS Null_Count
FROM OCD_Data WHERE OCD_Diagnosis_Date IS NULL;

-- Checking for NULL values in Duration_of_Symptoms_months
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Duration_of_Symptoms_months IS NULL;

-- Checking for NULL values in Previous_Diagnoses
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Previous_Diagnoses IS NULL;

-- Checking for NULL values in Family_History_of_OCD
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Family_History_of_OCD IS NULL;

-- Checking for NULL values in Obsession_Type
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Obsession_Type IS NULL;
```

```
-- Checking for NULL values in Compulsion_Type
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Compulsion_Type IS NULL;

-- Checking for NULL values in Y_BOCS_Score_Obsessions
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Y_BOCS_Score_Obsessions IS NULL;

-- Checking for NULL values in Y_BOCS_Score_Compulsions
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Y_BOCS_Score_Compulsions IS NULL;

-- Checking for NULL values in Depression_Diagnosis
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Depression_Diagnosis IS NULL;

-- Checking for NULL values in Anxiety_Diagnosis
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Anxiety_Diagnosis IS NULL;

-- Checking for NULL values in Medications
SELECT COUNT(*) AS Null_Count
FROM OCD_Data
WHERE Medications IS NULL;
```

b. CHECKING FOR DUPLICATE ROWS

SQL QUERY:

```
WITH CTE_Duplicates AS (
  SELECT
    *,
    ROW_NUMBER() OVER (
      PARTITION BY
        Patient_ID,
        Age,
        Gender,
        Ethnicity,
        Marital_Status,
        Education_Level,
        OCD_Diagnosis_Date,
        Duration_of_Symptoms_months,
        Previous_Diagnoses,
        Family_History_of_OCD,
        Obsession_Type,
        Compulsion_Type,
        Y_BOCS_Score_Obsessions,
        Y_BOCS_Score_Compulsions,
        Depression_Diagnosis,
        Anxiety_Diagnosis,
        Medications
      ORDER BY Patient_ID
    ) AS RowNum
  FROM OCD_Data
)
SELECT *
FROM CTE_Duplicates
```

```
WHERE RowNum > 1;
```

c. CATEGORIZING THE YBOCS SCORES, AGE AND DURATION:

SQL QUERY:

```
-- Categorizing YBOCS Obsession Score
ALTER TABLE OCD_Data
ADD Y_BOCS_Obsessions_Severity VARCHAR(50);

UPDATE OCD_Data
SET Y_BOCS_Obsessions_Severity = CASE
  WHEN Y_BOCS_Score_Obsessions BETWEEN 0 AND 7 THEN 'Subclinical'
  WHEN Y_BOCS_Score_Obsessions BETWEEN 8 AND 15 THEN 'Mild'
  WHEN Y_BOCS_Score_Obsessions BETWEEN 16 AND 23 THEN 'Moderate'
  WHEN Y_BOCS_Score_Obsessions BETWEEN 24 AND 31 THEN 'Severe'
  WHEN Y_BOCS_Score_Obsessions BETWEEN 32 AND 40 THEN 'Extreme'
  ELSE 'Unknown'
END;

-- Categorizing YBOCS Compulsion Score
ALTER TABLE OCD_Data
ADD YBOCS_Compulsions_Severity VARCHAR(50);

UPDATE OCD_Data
SET YBOCS_Compulsions_Severity = CASE
  WHEN Y_BOCS_Score_Compulsions BETWEEN 0 AND 7 THEN 'Subclinical'
  WHEN Y_BOCS_Score_Compulsions BETWEEN 8 AND 15 THEN 'Mild'
  WHEN Y_BOCS_Score_Compulsions BETWEEN 16 AND 23 THEN 'Moderate'
  WHEN Y_BOCS_Score_Compulsions BETWEEN 24 AND 31 THEN 'Severe'
  WHEN Y_BOCS_Score_Compulsions BETWEEN 32 AND 40 THEN 'Extreme'
  ELSE 'Unknown'
END;
```

```
--Categorizing the Age

ALTER TABLE OCD_Data
ADD Age_Category VARCHAR(50);

UPDATE OCD_Data
SET Age_Category = CASE
   WHEN Age BETWEEN 18 AND 34 THEN 'Young Adult'
   WHEN Age BETWEEN 35 AND 49 THEN 'Middle-Aged Adult'
   WHEN Age BETWEEN 50 AND 64 THEN 'Older Adult'
   WHEN Age >= 65 THEN 'Senior'
   ELSE 'Unknown'
END;

--Categorizing the duration of symptoms

ALTER TABLE OCD_Data
ADD Duration_Category VARCHAR(50);

UPDATE OCD_Data
SET Duration_Category = CASE
   WHEN Duration_of_Symptoms_months BETWEEN 0 AND 6 THEN 'Short-Term'
   WHEN Duration_of_Symptoms_months BETWEEN 7 AND 12 THEN 'Mid-Term'
   WHEN Duration_of_Symptoms_months BETWEEN 13 AND 36 THEN 'Long-Term'
   WHEN Duration_of_Symptoms_months BETWEEN 37 AND 120 THEN 'Chronic'
   WHEN Duration_of_Symptoms_months > 120 THEN 'Very Chronic'
   ELSE 'Unknown'
END;
```

6. Import the database to Power BI

7. Build the measures required for the project.

DAX Queries of the measures:

```
1. Average Age = AVERAGE(OCD_Data[Age])
2. Average Duration of Symptoms =
   AVERAGE(OCD_Data[Duration_of_Symptoms_months])
3. Average YBOCS Compulsion =
   AVERAGE(OCD_Data[Y_BOCS_Score_Compulsions])
4. Average YBOCS Obsession =
   AVERAGE(OCD_Data[Y_BOCS_Score_Obsessions])
5. Average YBOCS Score = AVERAGE(OCD_Data[YBOCS_Score])
6. Extreme patients =
   DIVIDE(COUNTROWS(FILTER(OCD_Data,OCD_Data[YBOCS_Severity]="E
   xtreme")),COUNTROWS(OCD_Data),0)
7. Most_Common_Compulsion =
       SELECTCOLUMNS(
           TOPN(
               1,
               SUMMARIZE(
                   OCD_Data,
                   OCD_Data[Compulsion Type],
                   "CompulsionCount", COUNT(OCD_Data[Compulsion
   Type])
               ),
               [CompulsionCount],
               DESC
           ),
           "MostCommonCompulsion", OCD_Data[Compulsion Type]
       )

8. Most_Common_Obsession =
       SELECTCOLUMNS(
           TOPN(
               1,
               SUMMARIZE(
                   OCD_Data,
                   OCD_Data[Obsession Type],
                    "ObsessionCount", COUNT(OCD_Data[Obsession
Type])
               ),
               [ObsessionCount],
               DESC
           ),
           "MostCommonObsession", OCD_Data[Obsession Type]
       )

9. Most_Prescribed_Medication =
   VAR Top_Med =
```

```
        TOPN(
            1,
          SUMMARIZE(
                OCD_Data,
                OCD_Data[Medications],
                "PrescriptionCount",
    COUNT(OCD_Data[Medications])
            ),
            [PrescriptionCount],
            DESC
        )
    RETURN
       MAXX(Top_Med, OCD_Data[Medications])
10.   Most_Prescribed_Medication_Anxiety =
    VAR Top_Med_Anxiety =
        TOPN(
            1,
            SUMMARIZE(
                FILTER(OCD_Data, OCD_Data[Anxiety Diagnosis] =
    "Yes"),
                OCD_Data[Medications],
                "PrescriptionCount",
    COUNT(OCD_Data[Medications])
            ),
            [PrescriptionCount],
            DESC
        )
    RETURN
     MAXX(Top_Med_Anxiety, OCD_Data[Medications])
11.   Most_Prescribed_Medication_Depression =
    VAR Top_Med_Depression =
        TOPN(
            1,
            SUMMARIZE(
                FILTER(OCD_Data, OCD_Data[Depression Diagnosis]
    = "Yes"),
                OCD_Data[Medications],
                "PrescriptionCount",
    COUNT(OCD_Data[Medications])
            ),
            [PrescriptionCount],
            DESC
        )
    RETURN
        MAXX(Top_Med_Depression, OCD_Data[Medications])
12.   Percent_Anxiety&Depression = [Total patients with both
    anxiety and depression]/[Total patients]
```

```
13.  Percent_history_of_OCD =
DIVIDE(CALCULATE(COUNTROWS(OCD_Data),OCD_Data[Family History
of OCD] = "Yes"), [Total patients])
14.  Percent_Married =
DIVIDE(
    COUNTROWS(
        FILTER(OCD_Data, OCD_Data[Marital Status] =
"Married")
    ),
    COUNTROWS(OCD_Data),
    0
)
15.  Percent_with_anxiety = DIVIDE([Total Patients with
Anxiety],[Total patients])
16.  Percent_with_depression = DIVIDE([Total patients with
Depression], [Total patients])
17.  Total patients = COUNTROWS(OCD_Data)
18.  Total Patients with Anxiety =
CALCULATE(COUNTROWS(OCD_Data), OCD_Data[Anxiety
Diagnosis]="Yes")
19.  Total patients with both anxiety and depression =
CALCULATE(COUNTROWS(OCD_Data), OCD_Data[Anxiety Diagnosis] =
"Yes", OCD_Data[Depression Diagnosis] = "Yes")
20.  Total patients with Depression =
CALCULATE(COUNTROWS(OCD_Data), OCD_Data[Depression
Diagnosis]="Yes")
21.  Total patients with OCD family history = CALCULATE(
COUNTROWS(OCD_Data), OCD_Data[Family History of OCD] ="Yes")
```

8. Build the Visuals.

Pages of the Project:

1. **Home Page:** This page acts as the contents page of the project.

2. **Demographic Overview:** This page analyses the demographic distribution of OCD patients and their characteristics.
   Insights:
   - Total number of patients is 1,500 with an average age of 47 years.
   - Gender distribution is nearly equal (49.8% male, 50.2% female).
   - Most patients have "Some College" (394) or a "Graduate Degree" (376).
   - Age category distribution shows the largest group as Young Adults (437).
   - Marital status insights indicate that 34% of patients are married.
   - The ethnic breakdown shows a fairly even distribution, with Caucasian being the largest group (26.53%).

3. **Time Based Analysis:** This page examines OCD symptom trends and duration patterns over time.
   Insights:
   - Average symptom duration is 122 months, with the average YBOCS score being 20.
   - Diagnosis trends over time show peaks in 2018 and a decline by 2022.
   - Symptom duration is longest among Seniors and Young Adults (125 months) and shortest for Older Adults (117 months).
   - All ethnicities have a similar average symptom duration (~122 months).
   - Duration categories indicate that 50.2% of patients are in the "Chronic" group.

4. **Clinical Outcomes:** This page analyzes OCD severity, comorbid conditions, and their distribution across demographic and clinical factors.
   Insights:
   - Patients with extreme and subclinical compulsion severity have higher numbers compared to mild and moderate cases.
   - 50% of patients have anxiety, and 51% have depression.
   - 25% of patients suffer from both anxiety and depression.
   - 51% of patients have a family history of OCD, indicating a potential genetic predisposition.
   - Average YBOCS scores are consistent across ethnicities but vary slightly by age group.

- Younger adults have the highest YBOCS scores compared to other age groups.

5. **OCD Symptoms & Severity:** This page identifies common OCD symptoms and assess severity levels among patients.
   Insights:
   - Most common obsession is Harm-related (333 patients), and the most common compulsion is Washing (321 patients).
   - 51% of patients have a family history of OCD, and 10% fall into the "Extreme" category.
   - Co-occurring disorders are common, with the highest being MDD (345 patients), followed by Panic-Disorder (313).
   - Severe obsession severity is associated with moderate compulsion severity across most categories.
   - Duration and severity mapping shows that most moderate cases are Chronic (271 patients), while extreme cases are more prevalent in Very Chronic (74 patients).

6. **Treatment & Medications:** This page evaluates the effectiveness and distribution of medications prescribed based on OCD severity, comorbidities, and demographics.
   Insights:
   - Benzodiazepines are the most prescribed for OCD, anxiety, and depression.
   - Young adults and middle-aged adults are more likely to receive benzodiazepines and SNRIs.
   - Moderate and severe YBOCS cases are primarily treated with benzodiazepines and SNRIs.
   - Very chronic and chronic cases receive benzodiazepines slightly more frequently than other medication types.
   - Patients with both anxiety and depression are treated across all three medication types (Benzodiazepines, SNRIs, SSRIs), with benzodiazepines being the most common.