

## OCD Patients Analysis

I found a dataset on Kaggle, that provided demographic and some clinical information on patients diagnosed with OCD.

I want to find out the following insights:

- The Number/Percentage of Males and Females diagnosed with OCD
- The Number of patients diagnosed with OCD filtered by Ethnicities.
- The Number of Patients diagnosed with OCD on a monthly basis.
- The Most Common Obsession Type amongst patients.
- The Most Common Compulsion Type amongst patients.

To answer these insights, I setup a MySQL database, where the patient dataset is stored and performed queries to narrow down and attain focused views of the data.

### The Number/Percentage of Males and Females diagnosed with OCD

To identify the total of males and females diagnosed with OCD, I ran the following query:

```
-- Patients By Gender
SELECT
gender,
COUNT(GENDER) AS count,
ROUND(((COUNT(GENDER)/1500) * 100),2) AS gender_percentage,
AVG(`Y-BOCS Score (Obsessions)`)
AS avg_obs_score FROM ocd_patient_dataset
GROUP BY 1
ORDER BY 2;
```

I ran a query that counts the number of patients and groups the count by gender. So, in this case it will display a row with a count for males and another row with a count for females. I also performed a calculated column, which determines the percentage of each gender. To determine the total number of patients there were in the dataset (1500 patients), I simply ran a count without a group.

```
SELECT COUNT(GENDER) as total FROM ocd_patient_dataset;
```

### **The Number of Patients diagnosed with OCD filtered by Ethnicity**

To identify the patients by ethnicity, we have to select the ethnicity column and perform a count of the patients and group the results by the ethnicity field. This is done with the following query:

```
-- Patients By Ethnicity
SELECT ethnicity,
count(*) AS patients_count,
AVG(`Y-BOCS Score (Obsessions)`) AS avg_obs_score
FROM ocd_patient_dataset
GROUP BY 1
ORDER BY 2;
```

We then order the results by the count, so that we can see the ethnicities with the highest prevalence of patients diagnosed first.

### **The Number of Patients diagnosed with OCD filtered by Ethnicity and link to Family History of OCD**

To determine if there is an ethnicity that is more predisposed to OCD, we can see if there is a genetic link by looking at the patients of each ethnicity's Family history, to judge whether they have relatives/ancestors with OCD.

To do this, I created a general query that selects the Patient Id, the patient's family history of OCD, and ethnicity.

```
-- Patients' Family History of OCD and Ethnicity
SELECT
`Patient ID`,`Family History of OCD`,Ethnicity
FROM ocd_patient_dataset
;
```

### **The Number of Patients diagnosed with OCD on a Monthly Basis**

To discover how many patients have been diagnosed monthly, I wrote the following query:

```
-- Number of People diagnosed with OCD Monthly
SELECT
`OCD Diagnosis Date`,
COUNT(`Patient ID`) as patients_diagnosed
FROM ocd_patient_dataset
GROUP BY 1;
```

This again counts the number of patients, but this time groups the count by the dates found by the Diagnosis Dates found on the dataset.

### **The Most Common Obsession Type amongst patients**

To view the most common obsession type amongst patients, I create a query where it performs a count which is then grouped by the obsession type. It is then ordered by the number of patients in an ascending order.

```
-- Most Common Obsession Type

SELECT
`Obsession Type`,
COUNT(*) AS patients_diagnosed,
AVG(`Y-BOCS Score (Obsessions)`) AS avg_obs_score
FROM ocd_patient_dataset
GROUP BY 1
ORDER BY 2;
```

### **The Most Common Compulsion Type amongst patients**

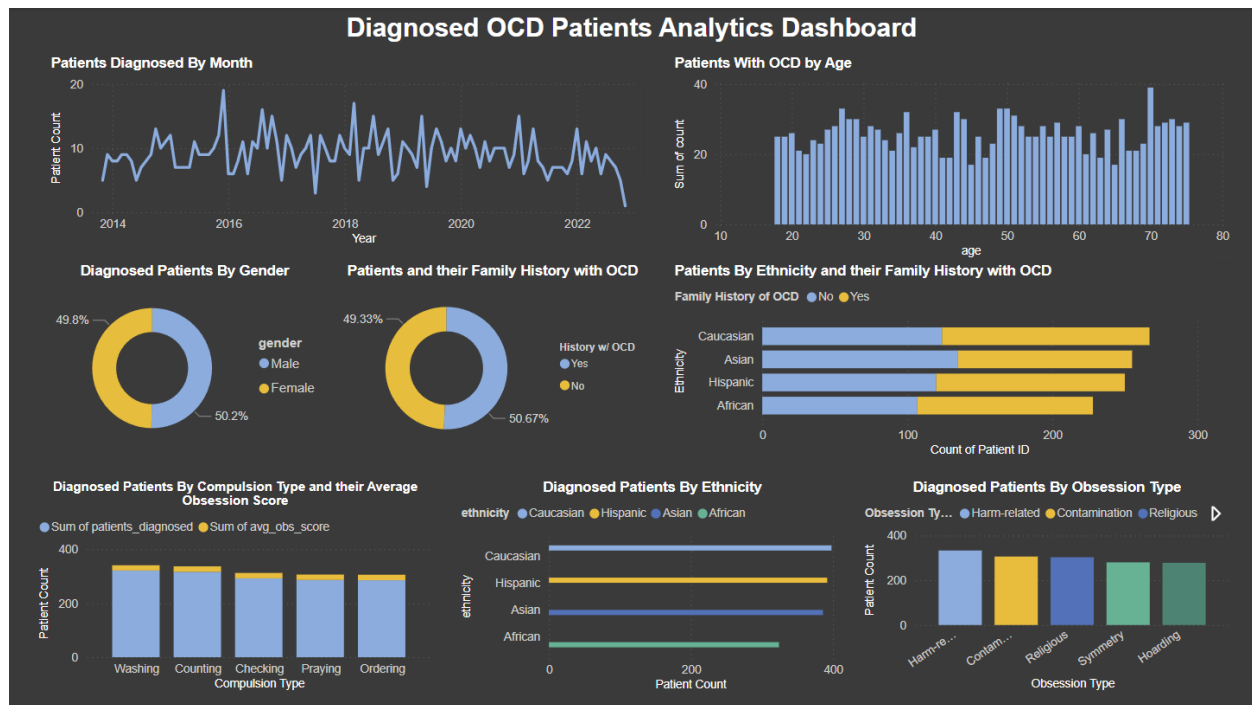
To view the most common compulsion type amongst patients, I create a query where it performs a count which is then grouped by the compulsion type. It is then ordered by the number of patients in an ascending order. I also perform an average on the Y-BOCS Score, which focuses on Obsessions for additional insights.

```
-- The most common Compulsion Type and its respective Average Obsession Score

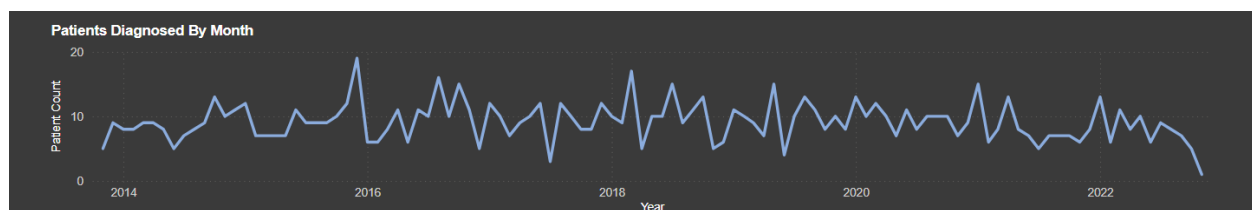
SELECT
`Compulsion Type`,
COUNT(*) AS patients_diagnosed,
AVG(`Y-BOCS Score (Obsessions)`) AS avg_obs_score
FROM
ocd_patient_dataset
GROUP BY 1
ORDER BY 2;
```

## Visualisations

To visualise the views of data that has been collected, I ran the queries against the database and saved each of the tables that were generated. I then loaded these tables onto PowerBI, so that I can create a static dashboard with visuals to provide useful information.



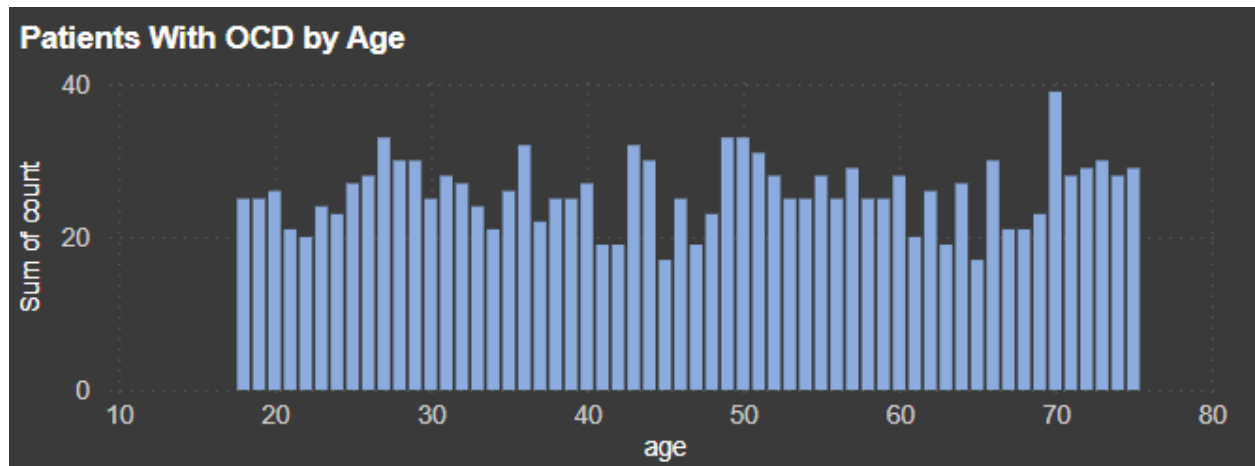
### Visual 1 – Patients Diagnosed by Month



This visual displays the entire duration of the data that has been recorded. For each month, the sum of patients diagnosed is displayed. With this, the user can identify which periods of time had spikes in diagnosis, and which periods had dips in diagnosis. This could be used in further insights, to identify birthdates that correlate with patients that have been diagnosed, to see which age group are being diagnosed with OCD.

With this graph, we can see there is a slow long-term decrease in patients being diagnosed, as the peak from the month before 2016 is the largest peak recorded. From 2016 onwards, the largest peak decreases over time, indicating that the number of patients with OCD is decreasing. However, with a small data set of 1500, we cannot say that for certain.

**Visual 2 – OCD Patients by Age**



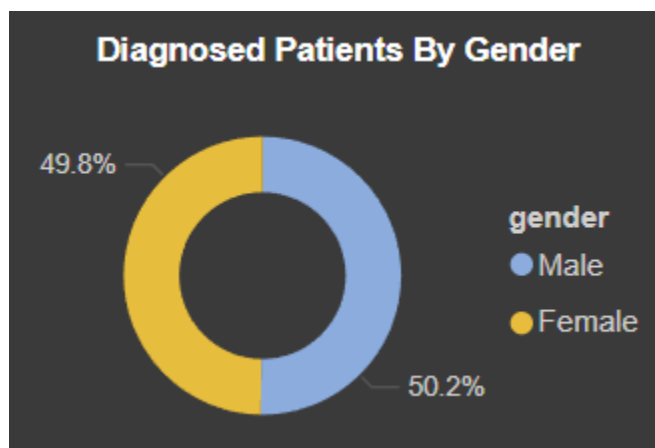
This visual distributes the patients by age. From this, we can see that those aged 70 are most likely to have developed OCD. With this, researchers can do a study solely on people aged 70, to determine whether the surprising volume of individuals diagnosed with OCD at this age could be due to a lack of diagnosis earlier in their life, or other issues such as retirement, reduced daily structure, or more assessments being performed, as there is greater healthcare access in older age.

Ages 45 and 65 have the lowest count of patients sitting at 17, and several ages around ages 40 – 47 are below the mean.

The spike in the late 20s could be linked to increased stressors, such as career pressures, leaving university, and major life transitions.

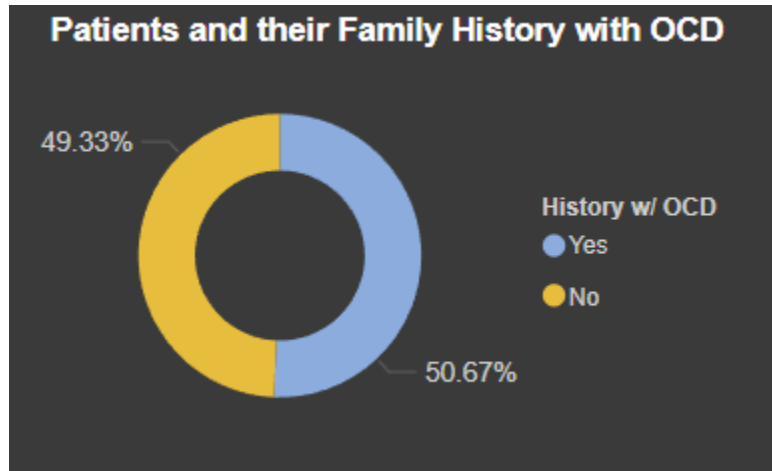
From ages 40 to 45, we see a dip in patients, as this age group possibly carry out coping mechanisms, are preoccupied with family care, and work to get diagnosed. This can also indicate that researchers must explore this age range, so that they can examine possible barriers to diagnosis or reporting.

**Visual 3 – Diagnosed Patients by Gender**



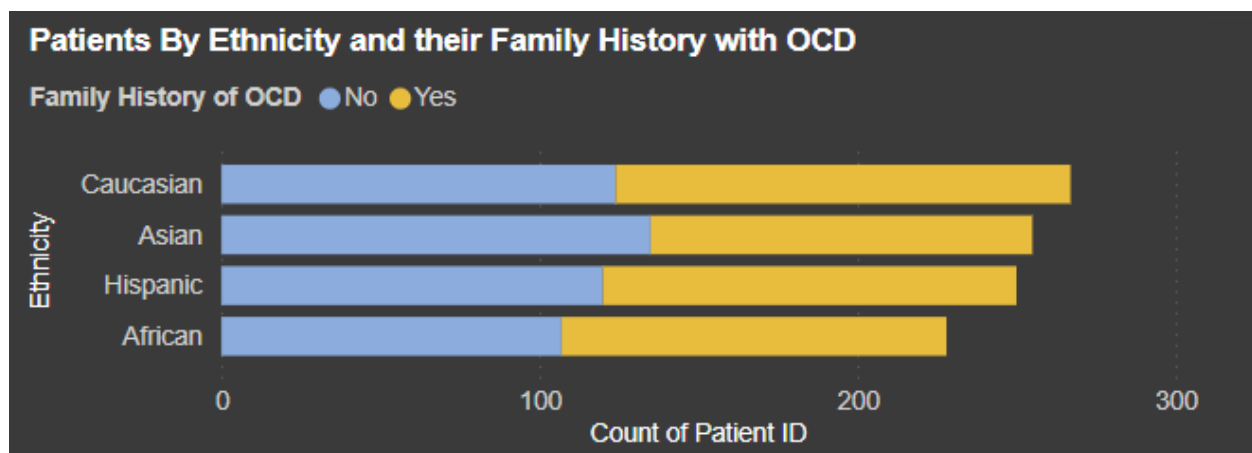
In this visual, it displays the number of patients by gender. Here we can see there isn't a significant difference between the number of male patients and number of female patients. As such, we can state from the dataset that OCD doesn't significantly affect one gender more than the other.

Visual 4 – Patients and their Family History With OCD



This visual shows the percentage of patients diagnosed with OCD and whether they have a family history with OCD or not. From the data, we can see there is no significant difference between patients having a family history or not. Hence, we cannot say there is a genetic familial link for OCD.

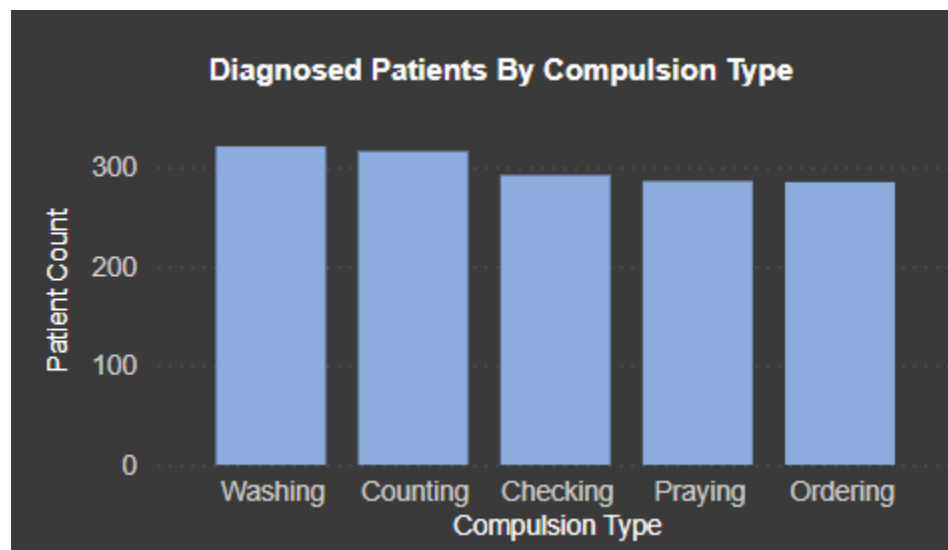
**Visual 5 – Patients By Ethnicity and their Family History with OCD**



For this visual, I used a stacked horizontal bar chart. It displays the number of patients diagnosed with OCD for each ethnicity. With this, we can see that Caucasians are the most diagnosed patients with OCD, whilst the ethnicity with the lowest number of patients with OCD are Africans. With this information, an investigation on diets consumed by Caucasians can be performed, in comparison to the diets consumed by Africans, to understand nutritional differences that could be linked to the development of OCD.

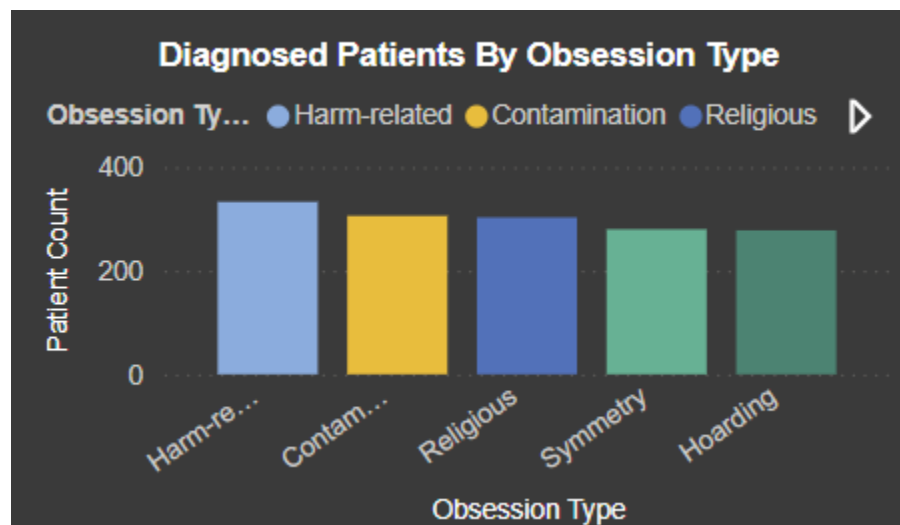
Furthermore, we can see that Caucasians seem to have more of a genetic predisposition to OCD than other ethnicities, as they have 143 patients with a family history with OCD, followed closely by Hispanics, who have 130 patients with a family history with OCD. This is reinforced by the fact that over 50% of the patients that are Caucasian or Hispanic have a family history with OCD. This can support research into epigenetics of Caucasian or Hispanic people, that could cause the increased predisposition of OCD.

**Visual 6 – Number of Patients by Compulsion Type and their Average Obsession Score**



From the bar chart above, we can see that Washing is the most common compulsion type, closely followed by Counting. This is understandable, as people with OCD often experience fears of contamination, which is an innate aversion to dirt and filth, to prevent infections and diseases. Those with OCD have an overactive instinct, leading to compulsive washing behavior. Counting follows closely, as it is a way in which people with OCD can express order and control. It aids in compelling a person to feel temporary relief or reduced anxiety in chaotic situations.

**Visual 7 – Patients By Obsession Type**



From this visual, harm-related obsessions are the most common type among individuals with OCD. This prevalence is understandable upon further research, as those with OCD often experience an intolerance of uncertainty and an intense need to control situations and thoughts. In OCD, the brain assigns unnecessary importance to intrusive thoughts, amplifying distress and preoccupation. This is further exacerbated by a heightened sense of responsibility, leading individuals to constantly monitor their actions and thoughts, to ensure they do not cause harm, whether intentionally or unintentionally.

From the chart we can see that contamination is another obsession type, which was previously discussed. The natural rejection of filth and contamination, that is instinctual to humans, is heightened for those with OCD. It pairs with their need for complete control and the elevated compulsion to neutralise completely.