

Healthcare data analysis

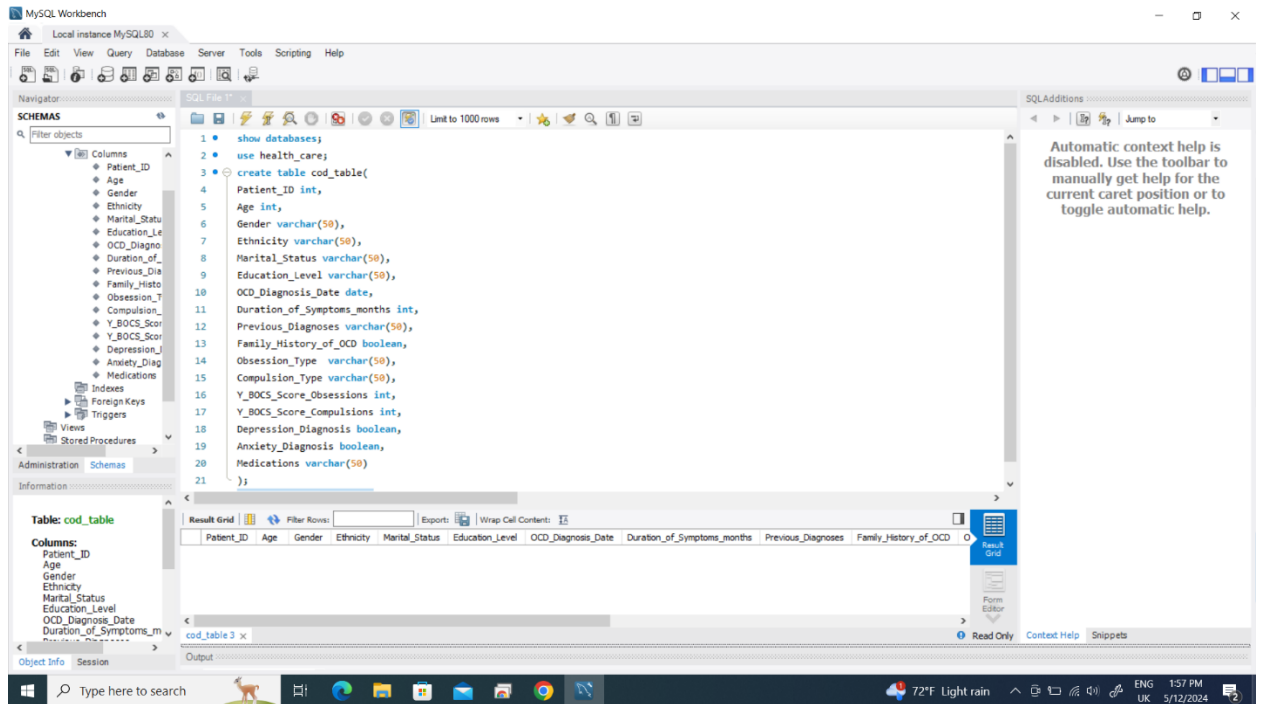
Data source: <https://www.kaggle.com/datasets/ohinhaque/ocd-patient-dataset-demographics-and-clinical-data/>

1. CSV dataset

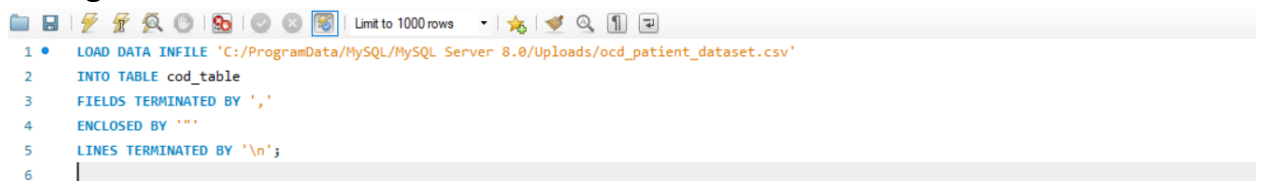
Patient ID	Age	Gender	Ethnicity	Marital Status	Education Level	OCD Diagnosis Date	Duration of Sympt	Previous Diagnosis	Family History of OCD	Obsession Type	Compulsion Type	Y-BOCS Score	Medications
1018	32	Female	African	Single	Some College	7/15/2016	203 MDD	No	Harm-related	Checking	17	10 Yes	Yes
2406	69	Male	African	Divorced	Some College	4/28/2017	180 None	Yes	Harm-related	Washing	21	25 Yes	Yes
1188	57	Male	Hispanic	Divorced	College Degree	2/2/2018	173 MDD	No	Contamination	Checking	3	4 No	No
6200	27	Female	Hispanic	Married	College Degree	8/25/2014	126 PTSD	Yes	Symmetry	Washing	14	28 Yes	Yes
5824	56	Female	Hispanic	Married	High School	2/20/2022	168 PTSD	Yes	Hoarding	Ordering	39	18 No	No
6946	32	Female	Asian	Married	College Degree	6/25/2016	46 GAD	No	Hoarding	Ordering	26	11 Yes	Yes
9861	38	Female	Hispanic	Single	College Degree	3/13/2017	110 MDD	No	Contamination	Praying	12	16 Yes	No
8396	57	Male	Hispanic	Divorced	College Degree	8/25/2015	157 PTSD	No	Religious	Ordering	31	4 Yes	No
9039	36	Male	Hispanic	Divorced	College Degree	9/19/2016	84 None	No	Harm-related	Praying	37	24 No	Yes
1580	72	Female	Hispanic	Divorced	Graduate Degree	7/13/2014	47 GAD	Yes	Contamination	Ordering	28	36 Yes	Yes
9672	21	Female	African	Married	College Degree	12/18/2014	95 GAD	Yes	Harm-related	Checking	40	26 Yes	No
7965	73	Female	Hispanic	Divorced	High School	1/13/2017	233 GAD	No	Religious	Counting	4	16 Yes	Yes
6237	23	Female	Hispanic	Married	High School	6/9/2015	48 Panic Disorder	No	Contamination	Praying	14	27 Yes	Yes
4456	68	Female	Caucasian	Single	Some College	4/13/2019	126 MDD	No	Contamination	Counting	10	39 Yes	No
9678	48	Female	Caucasian	Divorced	High School	8/24/2020	141 MDD	Yes	Harm-related	Washing	23	39 No	No
8070	29	Male	Hispanic	Married	College Degree	8/24/2022	108 PTSD	Yes	Symmetry	Praying	17	25 No	Yes
5003	25	Male	African	Married	College Degree	1/20/2021	206 Panic Disorder	Yes	Religious	Counting	7	21 Yes	No
7251	55	Female	African	Divorced	High School	1/23/2021	147 None	Yes	Religious	Checking	38	20 No	Yes
8271	50	Female	African	Single	Some College	11/4/2017	62 None	Yes	Hoarding	Washing	22	22 No	No
2637	66	Female	Asian	Divorced	College Degree	8/14/2018	73 Panic Disorder	No	Harm-related	Washing	0	12 No	Yes
6778	27	Male	Asian	Single	Graduate Degree	11/6/2015	220 PTSD	Yes	Religious	Ordering	30	4 Yes	Yes
9354	62	Male	Caucasian	Single	Graduate Degree	9/27/2022	89 PTSD	Yes	Religious	Counting	1	40 Yes	No
1022	28	Male	Caucasian	Single	College Degree	8/23/2016	208 PTSD	No	Contamination	Checking	32	16 No	No
8818	43	Male	Caucasian	Married	Graduate Degree	3/4/2022	39 GAD	No	Religious	Washing	0	18 Yes	Yes
3552	21	Female	Caucasian	Single	Some College	4/22/2016	14 Panic Disorder	No	Harm-related	Praying	29	5 Yes	No
4560	55	Female	Asian	Divorced	Graduate Degree	9/11/2017	152 MDD	No	Religious	Washing	19	25 Yes	Yes
5366	47	Male	Caucasian	Single	Some College	12/14/2016	134 GAD	No	Symmetry	Washing	9	36 Yes	No
8106	74	Male	Asian	Divorced	College Degree	9/9/2018	42 GAD	Yes	Hoarding	Checking	4	8 Yes	No
1213	60	Male	Asian	Divorced	College Degree	6/26/2014	226 MDD	No	Harm-related	Praying	18	0 Yes	No
7992	28	Male	Hispanic	Divorced	Graduate Degree	10/15/2019	151 GAD	Yes	Contamination	Counting	31	13 Yes	No
9536	22	Female	Asian	Divorced	College Degree	1/11/2021	90 Panic Disorder	No	Hoarding	Praying	20	20 No	Yes
6831	28	Male	Caucasian	Married	Some College	10/19/2021	58 Panic Disorder	Yes	Religious	Checking	27	4 No	No
4055	24	Male	African	Divorced	Some College	1/15/2017	58 Panic Disorder	Yes	Harm-related	Praying	34	0 No	Yes
7299	48	Male	Hispanic	Married	Some College	2/27/2021	193 None	Yes	Religious	Counting	23	12 Yes	Yes
2560	56	Male	Caucasian	Married	Some College	3/18/2020	49 None	Yes	Harm-related	Counting	22	22 Yes	No
2995	73	Female	Caucasian	Married	Graduate Degree	12/27/2017	112 Panic Disorder	No	Religious	Washing	17	38 Yes	No

SQL

2. Creating table into the database



3. Loading data into the table



4. Modifying table and ready to work

```
ALTER TABLE cod_table
MODIFY COLUMN Patient_ID int;

ALTER TABLE cod_table
MODIFY COLUMN Age int,
MODIFY COLUMN Duration_of_Symptoms_months int,
MODIFY COLUMN Y_BOCS_Score_Obsessions int,
MODIFY COLUMN Y_BOCS_Score_Compulsions int;

-- Convert the existing VARCHAR values to dates and update the column
UPDATE cod_table
SET OCD_Diagnosis_Date = DATE_ADD('1900-01-01', INTERVAL (CAST(OCD_Diagnosis_Date AS UNSIGNED) - 1) DAY);

ALTER TABLE cod_table
MODIFY COLUMN OCD_Diagnosis_Date date;
```

5. Calculating count of female and male that have ocd and average obsession score by gender

```
43
44 • SELECT Gender,
45         COUNT(Patient_ID) AS patient_count,
46         ROUND(AVG(Y_BOCS_Score_Obsessions), 2) AS avg_obs_scr
47 FROM cod_table
48 GROUP BY Gender
49 ORDER BY patient_count
50 ;
51
```

Result Grid			
Filter Rows:		Export:	Wrap Cell Content:
Gender	patient_count	avg_obs_scr	
Female	747	20.20	
Male	753	19.90	

6. Calculating percentage for male and female

```
24 • with data as (  
25     SELECT Gender,  
26           COUNT(Patient_ID) AS patient_count,  
27           ROUND(AVG(Y_BOCS_Score_Obsessions), 2) AS avg_obs_scr  
28     FROM cod_table  
29     GROUP BY Gender  
30     ORDER BY patient_count  
31 )  
32 select  
33     sum(case when Gender = 'Female' then patient_count else 0 end) as count_female,  
34     sum(case when Gender = 'Male' then patient_count else 0 end) as count_male,  
35     round((sum(case when Gender = 'Female' then patient_count else 0 end)/  
36           (sum(case when Gender = 'Female' then patient_count else 0 end)+  
37             sum(case when Gender = 'Male' then patient_count else 0 end))*100),2) as female_pct,  
38     round((sum(case when Gender = 'Male' then patient_count else 0 end)/  
39           (sum(case when Gender = 'Female' then patient_count else 0 end)+  
40             sum(case when Gender = 'Male' then patient_count else 0 end))*100),2) as male_pct  
41   from data  
42   ;  
43
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	count_female	count_male	female_pct	male_pct
▶	747	753	49.80	50.20

7. Calculating count and average of obsession by ethnicity

```
51  
52     -- count and average obsession by ethnicity that have cod  
53  
54 • select Ethnicity,  
55         count(Patient_ID) as Patient_count,  
56         avg(Y_BOCS_Score_Obsessions) as obs_score  
57     from cod_table  
58     group by Ethnicity  
59     order by Patient_count  
60     ;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Ethnicity	Patient_count	obs_score
▶	African	324	19.7593
	Asian	386	20.3161
	Hispanic	392	20.2832
	Caucasian	398	19.7915

8. Calculating number of people diagnosis with ocd month over month

```
61
62  -- number of people diag with ocd month over month
63 • select
64  date_format(OCD_Diagnosis_Date, '%Y-%m') as month,
65  count('Patient_ID') as patient_count
66  from cod_table
67  group by month
68  order by month
69  ;
```

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	month	patient_count			
▶	2013-11	9			
	2013-12	9			
	2014-01	13			
	2014-02	13			
	2014-03	15			
	2014-04	12			
	2014-05	11			
	2014-06	9			
	2014-07	11			
	2014-08	10			
	2014-09	12			
	2014-10	19			
	2014-11	18			
	2014-12	14			
	2015-01	16			
	2015-02	10			
	2015-03	12			
	2015-04	9			
	2015-05	10			

Result 25 x

9. most common obsession type (count) and its average obsession score

```
72 • select
73   Obsession_Type,
74   count(Patient_ID)as patient_count,
75   round(avg(Y_BOCS_Score_Obsessions),2) as obs_scr
76   from cod_table
77   group by Obsession_Type
78   order by patient_count;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Obsession_Type	patient_count	obs_scr
▶	Hoarding	278	21.01
	Symmetry	280	19.67
	Religious	303	19.23
	Contamination	306	19.67
	Harm-related	333	20.65

Result 29 x

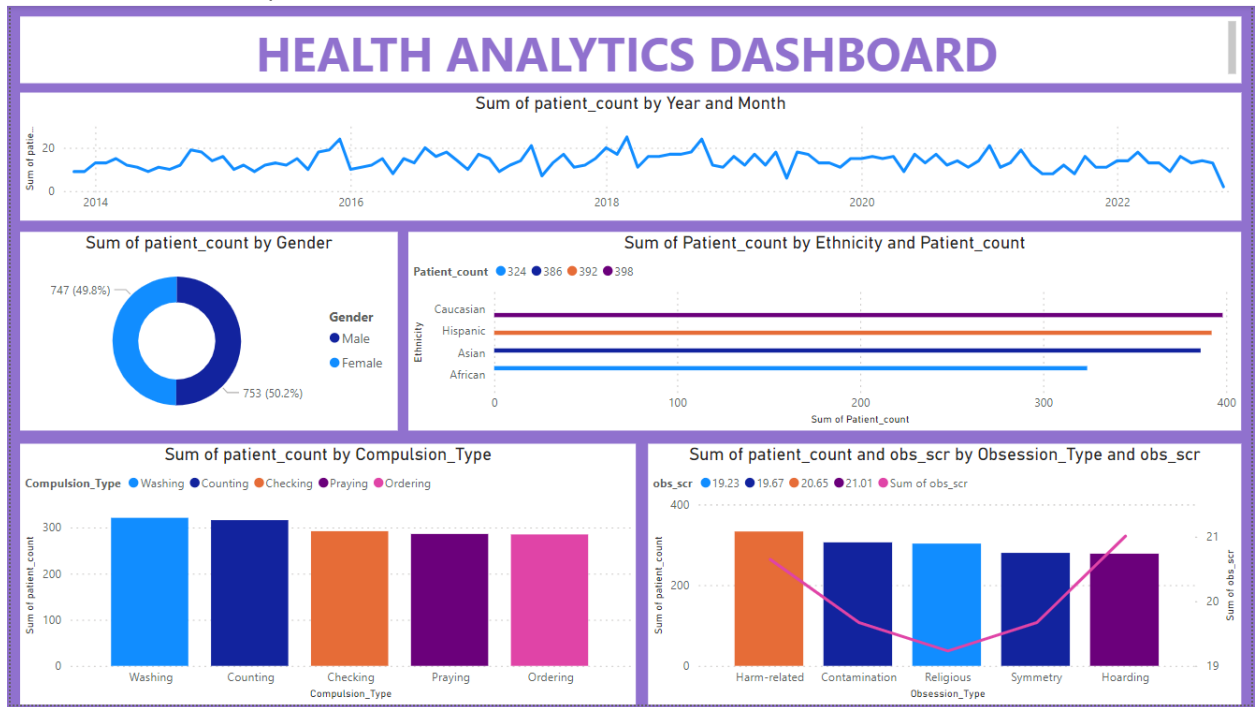
10. calculating most common compulsion type (count) and its average obsession score

```
80   -- calculating most common compulsion type (count) and its average obsession score
81 • select
82   Compulsion_Type,
83   count(Patient_ID)as patient_count,
84   round(avg(Y_BOCS_Score_Obsessions),2) as obs_scr
85   from cod_table
86   group by Compulsion_Type
87   order by patient_count;
```

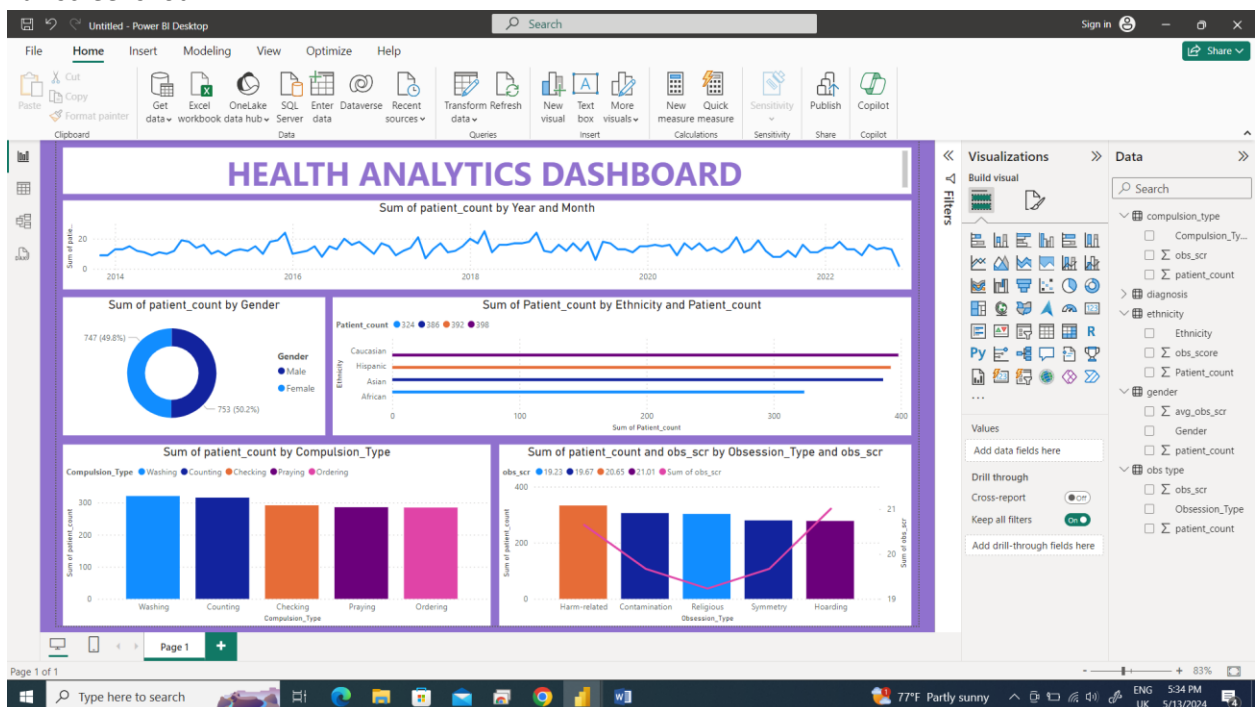
Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	Compulsion_Type	patient_count	obs_scr
▶	Ordering	285	20.40
	Praying	286	20.08
	Checking	292	19.99
	Counting	316	20.41
	Washing	321	19.40

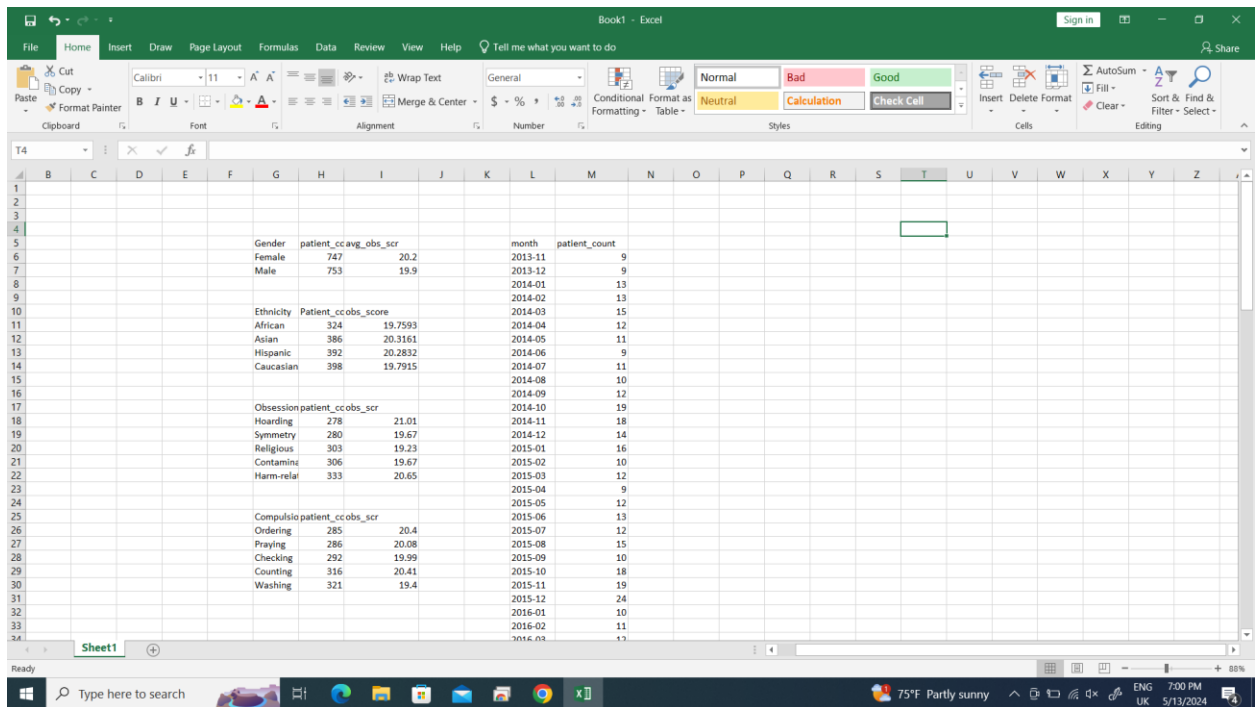
11. dashboard created in power bi



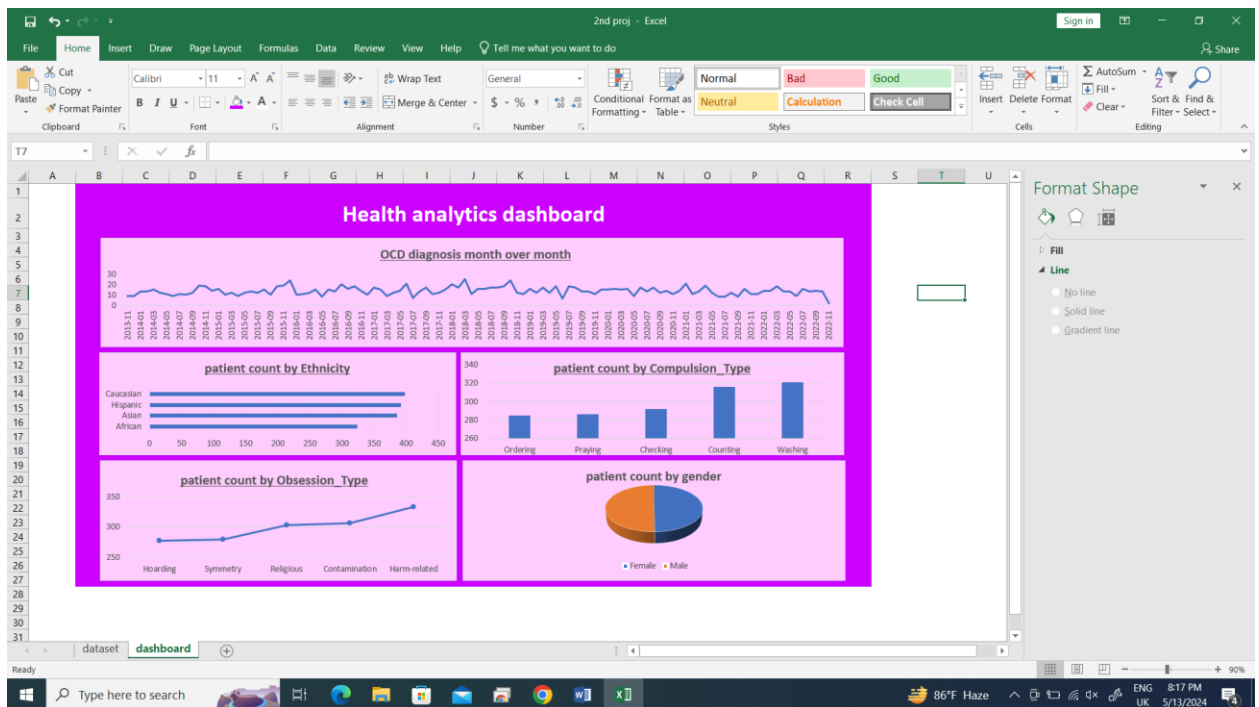
Full screenshot



12. Mini dataset loaded in the excel



13. Dashboard In excel



Thank you