

Stage 1. Find and critique a dataset

When it comes to a healthy lifestyle, what is the first thing that comes to your mind? If you thought of physical health and exercise, you are part of the majority of people. Which brings me to my real question, why are people not thinking more about mental health?

Having a mental illness has a bad reputation to it where people get discriminated against having mental health issues, some forms of discrimination include denying someone housing or turning down someone for a job based on their mental health.

An article from McLean hospital states that “people are afraid of being judged, or worse, that they could lose their job.” and “Almost half of the population will be diagnosed with a mental illness at some point in their life. Even if half of a family experiences a mental illness, the whole family feels its impact.”

Hence, I found this dataset which is a survey on mental health in the tech industry. This dataset provides numerous factors which when combined, provide a calculated assumption as to why they are suffering from mental illness. A simple example would be if gender plays a role in an individual suffering from mental illness and if so why?

This leads me to 3 main questions I would like to answer:

1. How much people are affected by mental health according to gender?
2. Which factors have a bigger role in having a mental illness?
3. What is the number of individuals that are affected according to each factor?

As the dataset did not specifically mention, which individual has been suffering from mental illness. Hence, I will be using the “treatment” factor to assume. Those that have indicated ‘Yes’ will be treated as suffering from a mental illness.

Assessing data set

Quality

The data set is a survey done by Open Sourcing Mental Illness (OSMI) taken in 2014. OSMI is a corporation dedicated to raising awareness, educating, and providing resources to support mental wellness in the tech and open-source communities.

They spread awareness and educate the community on mental health and are very dedicated to what they do. From their website, they have been doing a survey based on mental health in the tech industry every year, This shows that they are serious in what they do and the dataset is reliable.

Level of detail

The dataset provides many different factors regarding mental health. Hence, it is useful in providing crucial information to a high degree. This will allow me to identify the more impactful factors that lead to mental health.

Documentation

The documentation is easy to understand as all the words they used are fully explained in detail on Kaggle. An example would be "tech_company" which means "Is your employer primarily a tech company/organization?"

Interrelation

The dataset could be useful as a subset of Mental Health, as this is more specific towards the tech industry. However, most of the questions asked are quite general and can also be applied to others by removing some factors.

Use

The dataset will be used to test for assumptions that certain factors play a role in mental illness whether it is biological like age, gender. Data visualisation can be used on this to have a better understanding of how certain factors attribute to mental health, but as this will only be a simple web application, it will not be implemented.

The survey is done by individuals so there may be different answers or missing data. If the percentage of missing data is low, it will be filled. However, if there is too much missing data, that particular factor will not be used.

Discoverability

This dataset in particular was easy to find. Kaggle provided a huge variety of datasets, this particular dataset was chosen in particular due to having more factors compared to others. This provided the opportunity to select the factors which I wanted to investigate has a bigger impact on mental illness.

License

From Kaggle, the dataset is stated as CC BY-SA 4.0 which allowed it to be shared and adapted as long as credit was given. Hence, the dataset is allowed to be used for this assignment with a reference link to the dataset provided at the bottom.

Stage 2. Model your data

Figure 1 shows the entity-relationship model where the relationship of table Person is shown with the other tables consisting of one to many or zero, one to man. For example, Person can have zero to many conditions.

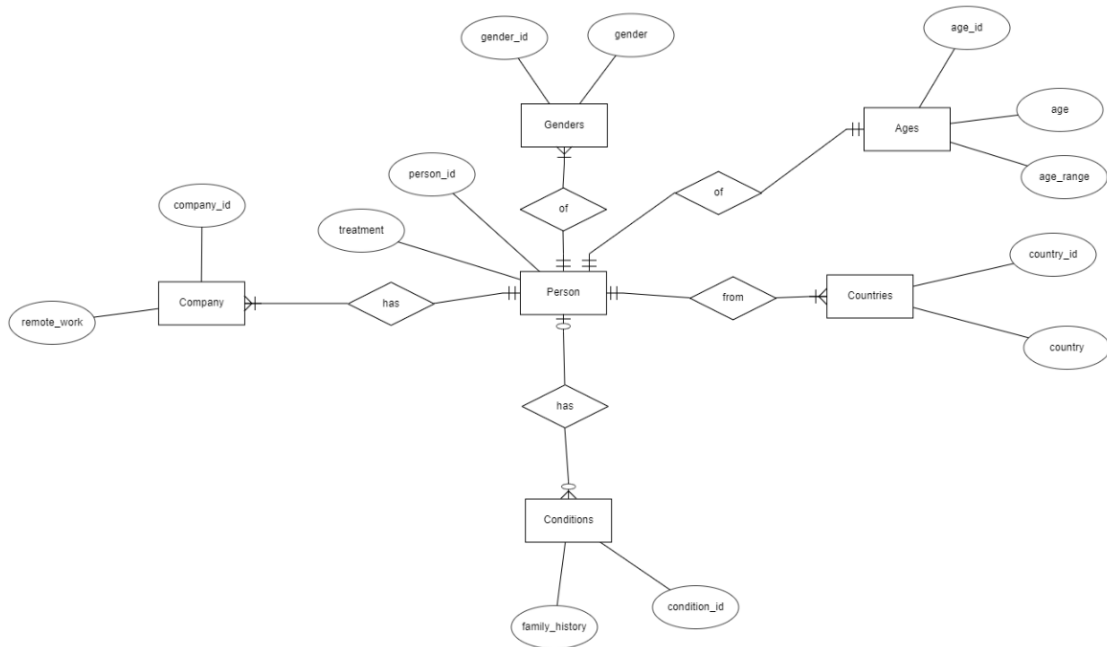


Figure 1: ER model

Figure 2 shows how each table is connected using foreign keys.

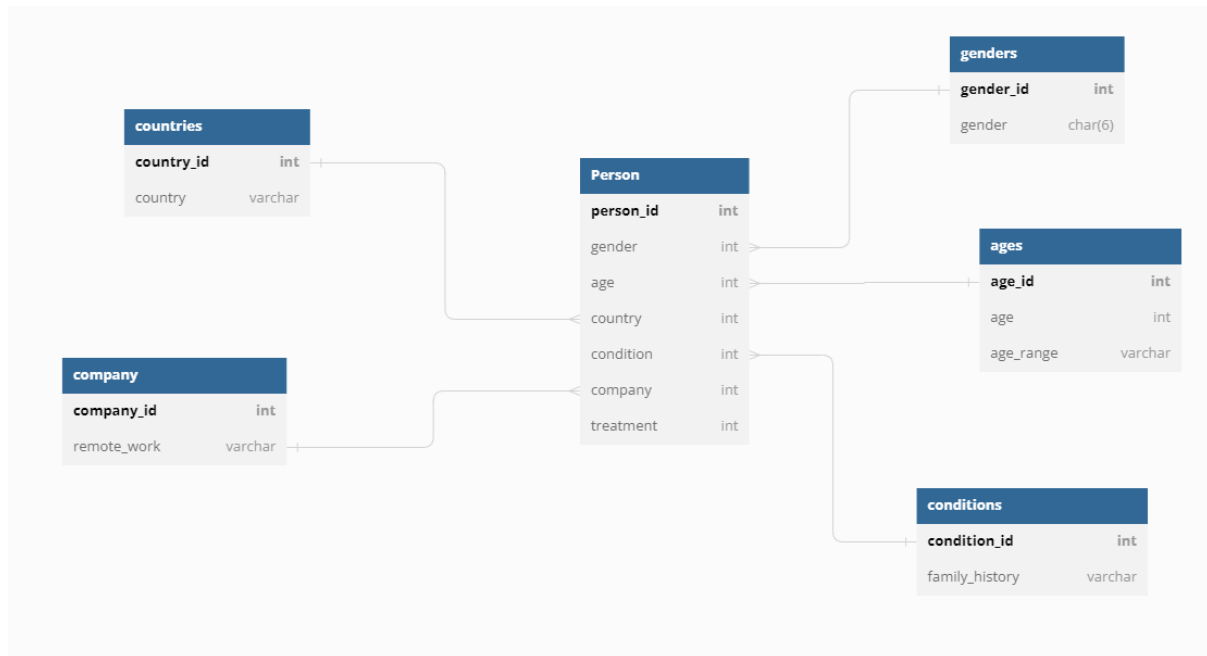


Figure 2: ER diagram

Figure 3 shows the columns that were available in the original dataset. However, only a subset of columns was used due to choosing the relevant attributes. Those that were relevant may also have missing data which was handled using python and pandas, however, some had too much missing data and were not chosen eventually.

Attributes	Description
Timestamp	-
Age	-
Gender	-
Country	-
State	-
Self employed	Are you self-employed?
Family history	Do you have a family history of mental illness?

Treatment	Have you sought treatment for a mental health condition?
Work interfere	If you have a mental health condition, do you feel that it interferes with your work?
No. employees	How many employees does your company or organization have?
Remote work	Do you work remotely (outside of an office) at least 50% of the time?
Tech company	Is your employer primarily a tech company/organization?
benefits	Does your employer provide mental health benefits?
Care options	Do you know the options for mental health care your employer provides?
Wellness program	Has your employer ever discussed mental health as part of an employee wellness program?
Seek help	Does your employer provide resources to learn more about mental health issues and how to seek help?
anonymity	Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?
leave	How easy is it for you to take medical leave for a mental health condition?
Mental health consequence	Do you think that discussing a mental health issue with your employer would have negative consequences?
Phys health consequence	Do you think that discussing a physical health issue with your employer would have negative consequences?
coworkers	Would you be willing to discuss a mental health issue with your coworkers?
supervisor	Would you be willing to discuss a mental health issue with your direct supervisor(s)?
Mental health interview	Would you bring up a mental health issue with a potential employer in an interview?
Phys health interview	Would you bring up a physical health issue with a potential employer in an interview?
mentalvsphysical	Do you feel that your employer takes mental health as seriously as physical health?
Obs consequence	Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?
comments	Any additional notes or comments

Figure 3: columns of original data

The database has been normalised to 3NF as shown in Figure 1 and 2. As every non-primary-key attribute is fully functionally dependent on the primary key and no non-primary-key attribute is transitively dependent on the primary key

Stage 3. Create the database

Below is a list of CREATE commands that were used to create the tables in MySQL.

CREATE commands	Description
CREATE DATABASE mental_health;	Using MySQL, a database called mental_health is first created using
CREATE USER admin IDENTIFIED BY '123';	A new user is created known as “admin” to get access to our database mental_health
GRANT SELECT on Person TO admin; GRANT SELECT on genders TO admin; GRANT SELECT on ages TO admin; GRANT SELECT on conditions TO admin; GRANT SELECT on countries TO admin; GRANT SELECT on company TO admin;	Gives the admin the ability to use SELECT command to the tables in mental_health.
CREATE TABLE raw_data(`id` int PRIMARY KEY AUTO_INCREMENT, `age` int, `gender` char(6), `country` varchar(50), `self_employment` varchar(10), `family_history` varchar(10), `treatment` int, `remote_work` varchar(10), `tech_company` varchar(10), `age_range` varchar(10));	Create a table to contain all raw data from the CSV file.
CREATE TABLE genders (`gender_id` int PRIMARY KEY AUTO_INCREMENT, `gender` varchar(6));	Create table called genders with attributes gender_id (PK), gender
CREATE TABLE ages (`age_id` int PRIMARY KEY AUTO_INCREMENT, `age` int NOT NULL, `age_range` varchar(10) NOT NULL);	Create table called ages with attributes age_id (PK), age, age_range
CREATE TABLE countries (`country_id` int PRIMARY KEY AUTO_INCREMENT, `country` varchar(50) NOT NULL);	Create table called countries with attributes country_id (PK), country
CREATE TABLE conditions (`condition_id` int PRIMARY KEY AUTO_INCREMENT, `family_history` varchar(10) NOT NULL);	Create table called conditions with attributes condition_id (PK), family_history

<pre>CREATE TABLE company (`company_id` int PRIMARY KEY AUTO_INCREMENT, `remote_work` varchar(10) NOT NULL);</pre>	<p>Create table called conditions with attributes remote_id (PK), remote_work</p>
<pre>CREATE TABLE Person (`person_id` int PRIMARY KEY AUTO_INCREMENT, `gender` int NOT NULL, `age` int NOT NULL, `country` int NOT NULL, `condition` int NOT NULL, `company` int NOT NULL, `treatment` int NOT NULL, FOREIGN KEY (`gender`) REFERENCES `genders` (`gender_id`), FOREIGN KEY (`age`) REFERENCES `ages` (`age_id`), FOREIGN KEY (`country`) REFERENCES `countries` (`country_id`), FOREIGN KEY (`condition`) REFERENCES `conditions` (`condition_id`), FOREIGN KEY (`company`) REFERENCES `company` (`company_id`));</pre>	<p>Create table called Person with attributes person_id (PK), gender (FK), age (FK), country (FK), condition (FK), treatment</p>

Figure 4: CREATE commands

Enter instance data

1. The first 1000 lines of raw data from the CSV file were loaded into the table to not overload the lab environment which might cause it to slow down. 1000 lines will be sufficient to answer our questions. After the raw_data table has been created, data is loaded using the following command:

LOAD DATA

```
INFILE '/home/coder/project/mid-term/data/mental_health_edited.csv'
REPLACE
INTO TABLE raw_data
FIELDS
  TERMINATED BY ','
  OPTIONALLY ENCLOSED BY '"'
  ESCAPED BY '"'
IGNORE 1 ROWS;
```


2. Data is then inserted from raw_data into the genders, ages, countries, and conditions. All the tables have the data inserted in a similar way except Person. 2 examples will be shown:

```
INSERT INTO ages (age,age_range)
SELECT 'age','age_range'
FROM raw_data
LIMIT 1000;
```

```
INSERT INTO countries (country)
SELECT 'country'
FROM raw_data
LIMIT 1000;
```

The Person table's data will be inserted by using the id from the other tables. The following command was used:

```
INSERT INTO Person (gender,age,country,condition)
SELECT genders.gender_id,ages.age_id, countries.country_id ,conditions.condition_id
FROM genders
JOIN ages ON genders.gender_id = ages.age_id
JOIN countries ON countries.country_id =ages.age_id
JOIN conditions ON conditions.condition_id = countries.country_id;
```

I later realize i needed to add the treatment information directly into Person table, so i had to manually add in using the following command:

```
INSERT INTO Person (treatment)
SELECT 'treatment'
FROM raw_data
LIMIT 1000;
```

Reflect on how well the database reflects the data

The data has been normalised and separated so that any changes made on the Person table will not directly impact its other attributes such as if age is removed. One thing I did not consider was the values under 'remote_work' in the company table, it made it such that when combining the queries from the other tables, the result will not look polished. Which made me decide to not include it in the end

List SQL commands that answer questions identified in Stage 1/Step 3

Q1: How much people are affected by mental health according to gender?

This shows the number according to gender that is affected and not affected by mental health. Where 0 and 1 under the treatment column refers to not affected by mental illness and affected by mental illness respectively (Figure 4). The following is the command:

```
SELECT COUNT(genders.gender) AS count, genders.gender, Person.treatment
FROM genders JOIN Person ON genders.gender_id = Person.gender
GROUP BY treatment, genders.gender
ORDER BY treatment;
```

count	gender	treatment
57	female	0
447	male	0
4	trans	0
131	female	1
348	male	1
13	trans	1

Figure 4

Q2: Which factors have a bigger role in having a mental illness?

Q3: How many people are affected according to each factor?

Q2 and Q3 are both answered using the following command where the top 20 factors are shown in descending order with the count showing how many people are affected according to each factor. Example: 31 people from Canada are affected by mental illness and 30 people of age 32 are affected by mental illness. (Figure 5)

```
Select *
From
(SELECT COUNT(g.gender_id) AS count, g.gender as factors
FROM genders g JOIN Person p on g.gender_id = p.gender
WHERE p.treatment = 1
GROUP BY g.gender
UNION ALL
SELECT COUNT(a.age_id) AS count, a.age
FROM ages a JOIN Person p on a.age_id = p.gender
WHERE p.treatment = 1
GROUP BY a.age
UNION ALL
SELECT COUNT(c.country_id) AS count, c.country
FROM countries c JOIN Person p on c.country_id = p.gender
WHERE p.treatment = 1
GROUP BY c.country) as sum
ORDER BY count DESC
LIMIT 20;
```

count	factors
348	male
326	United States
131	female
71	United Kingdom
33	33
31	Canada
30	32
28	28
28	29
27	31
26	26
26	30
26	35
25	27
24	25
23	34
19	37
17	23
16	Germany
16	38

Figure 5

4. Create a simple web application

The web application allows users to choose different factors to see how many people are affected by mental illness, there is also an overall ranking to show which factors are the one that affects people the most. Using this web application, I want to identify what factors are affecting mental illness the most using tables.

To use the web application, the user has to enter 'node index.js' in the terminal and run on port 8088. index.js will be used to make queries from the database and the web application will be using the Express framework and Mustache templating engine.

Note: Users should ensure they are in the '/home/coder/project/mental_health' before using 'node index.js'.

```
1  // Import libraries
2  const express = require('express');
3  const bodyParser = require('body-parser');
4  const mysql = require('mysql');
5  const mustacheExpress = require('mustache-express');
6
7  // Initialise objects and declare constants
8  const app = express();
9  const webPort = 8088;
10
11  const con = mysql.createConnection({
12    host: "localhost",
13    user: "root",
14    password: "",
15    database: "mental_health"
16  });
17
18  con.connect((err) => {
19    if (err) {
20      throw err;
21    }
22    console.log("Connected to database");
23  });
24
25  app.engine('html', mustacheExpress());
26  app.set('view engine', 'html');
27  app.set('views', './templates');
28  app.use(bodyParser.urlencoded({ extended: true }));
29
30  app.listen(
31    webPort,
32    () => console.log('web application listening on port ' + webPort) // success callback
33  );
```

Figure 6

Note: I had some issues (Figure 7) connecting to the server with the user that was created in Figure 4. I tried different methods to solve the issue like upgrading MySQL but it did not work, hence I resorted to using the root user to access the database.

```
at Protocol._parsePacket (/home/coder/project/mental_health/node_modules/mysql/lib/protocol/Protocol.js:291:23)
at Parser._parsePacket (/home/coder/project/mental_health/node_modules/mysql/lib/protocol/Parser.js:433:10)
at Parser.write (/home/coder/project/mental_health/node_modules/mysql/lib/protocol/Parser.js:43:10)
at Protocol.write (/home/coder/project/mental_health/node_modules/mysql/lib/protocol/Protocol.js:38:16)
at Socket.<anonymous> (/home/coder/project/mental_health/node_modules/mysql/lib/Connection.js:88:28)
at Socket.<anonymous> (/home/coder/project/mental_health/node_modules/mysql/lib/Connection.js:526:10)
at Socket.emit (events.js:315:20)
at addChunk (_stream_readable.js:309:12)
-----
at Protocol._enqueue (/home/coder/project/mental_health/node_modules/mysql/lib/protocol/Protocol.js:144:48)
at Protocol.handshake (/home/coder/project/mental_health/node_modules/mysql/lib/protocol/Protocol.js:51:23)
at Connection.connect (/home/coder/project/mental_health/node_modules/mysql/lib/Connection.js:116:18)
at Object.<anonymous> (/home/coder/project/mental_health/index.js:18:5)
at Module._compile (internal/modules/cjs/loader.js:1063:30)
at Object.Module._extensions..js (internal/modules/cjs/loader.js:1092:10)
at Module.load (internal/modules/cjs/loader.js:928:32)
at Function.Module._load (internal/modules/cjs/loader.js:769:14)
at Function.executeUserEntryPoint [as runMain] (internal/modules/run_main.js:72:12)
at internal/main/run_main_module.js:17:47 {
  code: 'ER_NOT_SUPPORTED_AUTH_MODE',
  errno: 1251,
  sqlMessage: 'Client does not support authentication protocol requested by server; consider upgrading MySQL client',
  sqlState: '08004',
  fatal: true
}
```

Figure 7

The queries are made using the following command in Figure 8. The query is then used to render Mustache templates in index.html (Figure 9). The names used in the table are the column names in MySQL after the table is displayed. For example, 'count' in Figure 9 line 14 is the 'count' column in Figure 5.

```
app.get('/', function (req, res) {
  const data = {};
  let query = "Select * From (SELECT COUNT(g.gender_id) AS count, g.gender as factors FROM genders g JOIN Person p on g.gender_id = p.gender WHERE";
  con.query(query, templateRenderer('index', res));
});
```

Figure 8

```

1 <!DOCTYPE html>
2 <html lang="en">
3 <style>
4     table,th,td {
5         border: 1px solid;
6     }
7     table {
8         width: 100%;
9         border-collapse: collapse;
10    }
11    th,td {
12        padding: 8px 16px;
13        text-align: center
14    }
15 </style>
16 <head>
17     <meta charset="utf-8" />
18     <title>Ranking</title>
19 </head>
20 <body>
21     <h1>Ranking by factors</h1>
22
23     <p>Top 20 factors that affect mental illness</p>
24     <p>Count: number of people </p>
25     <p>Factors: attributes that may cause mental illness </p>
26
27     <a href="/">ranking</a>
28     <a href="/gender">gender</a>
29     <a href="/age">age</a>
30     <a href="/country">country</a>
31
32     <table>
33         <tr>
34             <th>count</th>
35             <th>factors</th>
36         </tr>
37         {{#data}}
38         <tr>
39             <td>{{count}}</td>
40             <td>{{factors}}</td>
41         </tr>
42         {{/data}}
43     </table>
44 </body>
45 </html>

```

Figure 9

Using mustache and some HTML, each row in MySQL, will then be displayed (Figure 10).

localhost:8088/

Ranking by factors

Top 20 factors that affect mental illness

Count: number of people

Factors: attributes that may cause mental illness

[ranking](#) [gender](#) [age](#) [country](#)

count	factors
348	male
326	United States
131	female
71	United Kingdom
33	33
31	Canada
30	32

Figure 10

Besides the different queries, the other pages are done in a similar way. The examples of the different webpages are show below:

localhost:8088/gender

Ranking by genders

How does gender affect the number of people suffering from mental illness

Count: number of people

Treatment: 0 = No, 1 = Yes

[ranking gender age country](#)

Count	Gender	Treatment
57	female	0
447	male	0
4	trans	0
131	female	1
348	male	1

Figure 11

localhost:8088/age

Ranking by ages

Top 20 factors that affect mental illness

Count: number of people

[ranking gender age country](#)

Count	Ages
33	33
30	32
28	28
28	29
27	31
26	35

Figure 12

localhost:8088/country

Ranking by countries

Count: number of people

[ranking gender age country](#)

Count	Countries
326	United States
71	United Kingdom
31	Canada
16	Germany
9	Australia
6	Ireland
6	Netherlands

Figure 13

Conclusion

The web application provides a way to view the factors that affect mental illness in a concise manner. However, it does not mean that it should be used as any evidence for any kind of research etc, it is purely based on assumptions and for educational purposes.

References

Dataset: <https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey>

Mental health:

<https://www.mcleanhospital.org/essential/lets-face-it-no-one-wants-talk-about-mental-health>

Your shareable lab link:

<https://hub.labs.coursera.org:443/connect/sharedvpafchgd?forceRefresh=false&path=%2F%3Ffolder%3D%2Fhome%2Fcoder%2Fproject>