

1. Create the Table in MySQL

sql

```
CREATE TABLE IF NOT EXISTS mental_health_survey (  
    id INT AUTO_INCREMENT PRIMARY KEY,  
    Age INT,  
    Gender VARCHAR(20),  
    Country VARCHAR(50),  
    self_employed INT,  
    family_history INT,  
    treatment INT,  
    work_interfere VARCHAR(30),  
    no_employees VARCHAR(20),  
    remote_work INT,  
    tech_company INT,  
    benefits INT,  
    care_options INT,  
    wellness_program INT,  
    seek_help INT,  
    anonymity INT,  
    leave VARCHAR(30),  
    mental_health_consequence INT,  
    phys_health_consequence INT,  
    coworkers INT,  
    supervisor INT,  
    mental_health_interview INT,  
    phys_health_interview INT,  
    mental_vs_physical INT,  
    obs_consequence INT  
);
```

2. Import Your Cleaned Data

sql

```
LOAD DATA INFILE 'C:/path/to/survey_final_typed.csv'  
INTO TABLE mental_health_survey  
FIELDS TERMINATED BY ','  
ENCLOSED BY '"'  
LINES TERMINATED BY '\n'  
IGNORE 1 ROWS  
(Age, Gender, Country, self_employed, family_history, treatment, work_interfere,  
no_employees, remote_work, tech_company, benefits, care_options, wellness_program,  
seek_help, anonymity, leave, mental_health_consequence, phys_health_consequence,  
coworkers, supervisor, mental_health_interview, phys_health_interview,  
mental_vs_physical, obs_consequence);
```

3. Run Exploratory Data Analysis (EDA) Queries

a. Prevalence of Mental Health Issues by Company Size

sql

```
SELECT
    remote_work,
    AVG(treatment) AS avg_mental_health_issue
FROM
    mental_health_survey
GROUP BY
    remote_work;
```

Sample Output:

no_employees	total_respondents	affected_employees	pct_affected
1-5	120	48	40.00
6-25	80	28	35.00
26-100	60	18	30.00
100-500	100	25	25.00
500-1000	50	10	20.00
More than 1000	90	15	16.67

Interpretation:

Smaller companies tend to have a higher percentage of employees reporting mental health treatment.

b. Correlation Between Remote Work and Mental Health Issues

sql

```
SELECT
    remote_work,
    AVG(treatment) AS avg_mental_health_issue
FROM
    mental_health_survey
GROUP BY
    remote_work;
```

Sample Output:

remote_work	total_respondents	affected_employees	pct_affected
1 (Yes)	200	60	30.00
0 (No)	300	120	40.00

Interpretation:

Employees without remote work options report more mental health issues.

c. Segmentation by Age, Gender, and Work Interference

Sample Output:

Age	Gender	work_interfere	total_responses	affected_responses	pct_affected
28	Male	Often	5	5	100.00
31	Female	Often	3	3	100.00
24	Male	Sometimes	7	6	85.71
29	Female	Rarely	6	5	83.33
35	Other	Often	2	1	50.00

Interpretation:

Younger employees and those who report work often interfering with mental health are most likely to seek treatment.

4. Benefits and Mental Health

sql

```
SELECT
    benefits,
    COUNT(*) AS total_respondents,
    SUM(treatment) AS affected_employees,
    ROUND(100.0 * SUM(treatment) / COUNT(*), 2) AS pct_affected
FROM
    mental_health_survey
GROUP BY
    benefits;
```

Sample Output:

benefits	total_respondents	affected_employees	pct_affected
1 (Yes)	250	60	24.00
0 (No)	150	60	40.00
NULL	100	30	30.00

Interpretation:

Companies offering mental health benefits see a lower percentage of employees needing treatment.