**Report on Mental Health Indicators Analysis Using SQL**

**Prepared by:**

Jeremy Madonna Sebastian Joe

Contents

[Introduction 3](#_Toc191239239)

[Goal of the project 3](#_Toc191239240)

[Data cleansing 3](#_Toc191239241)

[Relational Schema/ERD 3](#_Toc191239242)

[SQL queries 5](#_Toc191239243)

[Query 1: **Average values by age group and gender** (Basic Join and Aggregation) 5](#_Toc191239244)

[Query 2: **Finding indicators with high average value (Having Clause and Joins)** 6](#_Toc191239245)

[Query 3: **Display the highest mental health indicator in each year** (Max function) 8](#_Toc191239246)

[Query 4: **Grouping by gender to show average value (Group By)** 9](#_Toc191239247)

[Query 5: **Gender comparison by indicator (Multiple joins)** 10](#_Toc191239248)

[Query 6: **Categorize values for each indicator as high or low (Case statement)** 12](#_Toc191239249)

[Query 7: **Show top mental health indicator for each province (CTE)** 13](#_Toc191239250)

[Query 8: **Indicator Summary (View)** 15](#_Toc191239251)

[Query 9: **Demographic Risk Assessment** 16](#_Toc191239252)

[Query 10: **Regional Health Dashboard (View)** 18](#_Toc191239253)

[Conclusion 19](#_Toc191239254)

[References 20](#_Toc191239255)

# Introduction

Mental health data analysis plays a crucial role in understanding population well-being. This project leverages PostgreSQL database management to normalize and analyze Canadian mental health statistics from 2002-2022, creating an efficient system for tracking indicators, demographics, and regional variations in mental health conditions.

# Goal of the project

This project aims to examine key indicators influencing mental health across various demographics, including age, gender, and provinces. By identifying these factors, the goal is to ensure timely support and intervention for those in need of mental health assistance.

# Data cleansing

The dataset I have used for my analysis is found in Stats CAN page related to Mental health indicators (Government of Canada, Statistics Canada, 2024)

**URL:** https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310046501

To prepare the mental health dataset for analysis, I implemented several crucial cleaning steps. First, I removed Canada-wide statistics to focus on regional patterns. I cleaned up data quality issues by removing empty status fields, unknown values (".."), and redundant 'Total' categories in gender and age classifications. The indicators column required restructuring - I split it into two distinct components: the indicator name and its duration measurement. Similarly, I refined the characteristics column to retain only the primary descriptors. For data efficiency, I removed non-essential columns like Scalar\_id, Vector, and Coordinate. Throughout this process, I maintained data integrity by ensuring proper validation and retaining a minimum of 5000 records for meaningful analysis. There is enough data to conduct the analysis effectively even if not every table contain 50 rows.

# Relational Schema/ERD

The mental health dataset was normalized into five distinct tables: demographics, geography, indicators, characteristics, and mental\_health\_stats. This normalization eliminates data redundancy, maintains data integrity, and establishes clear relationships through primary and foreign key constraints.

A screenshot of a computer

AI-generated content may be incorrect.

*Fig:1 Mental Health Data ER-Diagram*

The primary keys, foreign keys are tabulated for reference below

|  |  |  |
| --- | --- | --- |
| **Table Name** | **Primary Key** | **Foreign Keys** |
| demographics | demographic\_id | None |
| geography | geo\_id | None |
| indicators | indicator\_id | None |
| characteristics | characteristic\_id | None |
| mental\_health\_stats | stat\_id | - demographic\_id  - geo\_id  - indicator\_id  - characteristic\_id |

# SQL queries

## Query 1: **Average values by age group and gender** (Basic Join and Aggregation)

SELECT

d.age\_group,

d.gender,

ROUND (AVG(mhs.value)::NUMERIC, 2) AS avg\_value

FROM

mental\_health\_stats mhs

INNER JOIN

demographics d

ON mhs.demographic\_id = d.demographic\_id

GROUP BY d.age\_group, d.gender

ORDER BY d.age\_group, d.gender

|  |  |  |
| --- | --- | --- |
| **age\_group** | **gender** | **avg\_value** |
| 15-24 | Men | 4008.15 |
| 15-24 | Women | 4018.99 |
| 25-44 | Men | 4009.99 |
| 25-44 | Women | 4010.93 |
| 45-64 | Men | 4010.52 |
| 45-64 | Women | 4015.34 |

**Insight:**

Women show higher average values across all age groups,with the biggest gap at 15-24 (4018.99 vs. 4008.15), indicating greater challenges for young women. Both genders see high values at 45-64, highlighting middle-age concerns. Men's values stay steady, while women's vary more, guiding targeted mental health support.

## Query 2: **Finding indicators with high average value (Having Clause and Joins)**

SELECT

i.indicator\_name,

ROUND(AVG(mhs.value)::NUMERIC, 2) AS avg\_value

FROM

mental\_health\_stats mhs

INNER JOIN

indicators i

ON mhs.indicator\_id = i.indicator\_id

GROUP BY

i.indicator\_name

HAVING

AVG (mhs.value) > 4010

ORDER BY

avg\_value DESC

|  |  |  |
| --- | --- | --- |
| **indicator\_name** | **avg\_value** | **total\_records** |
| Alcohol abuse or dependence | 4058.9 | 985 |
| Generalized anxiety disorder | 4035.93 | 985 |
| Major depressive episode | 4023.68 | 990 |
| Self-rated work stress | 4019.73 | 990 |
| Suicidal thoughts | 4014.93 | 990 |
| Any mood disorder | 4012.35 | 985 |
| Cannabis use | 4012.14 | 986 |

**Insight:**

* Alcohol abuse and dependence shows the highest average value (4058.9), significantly higher than other indicators, suggesting it might be the most prevalent mental health concern in the dataset.
* Anxiety and depression-related indicators (Generalized anxiety disorder at 4035.93 and Major depressive episode at 4023.68) rank second and third, highlighting the prominence of these conditions.
* All indicators in this filtered list show relatively high values (above 4010), with close measurement frequencies (985-990 records each), indicating these are consistently tracked major mental health concerns.

## Query 3: **Display the highest mental health indicator in each year** (Max function)

SELECT

DISTINCT ON (ref\_date) ref\_date as year,

i.indicator\_name,

MAX(value) as highest\_value

FROM

mental\_health\_stats mhs

INNER JOIN indicators i ON mhs.indicator\_id = i.indicator\_id

GROUP BY

ref\_date,

i.indicator\_name

ORDER BY

ref\_date,

highest\_value DESC

|  |  |  |
| --- | --- | --- |
| **year** | **indicator\_name** | **highest\_value** |
| 2002 | Eating disorder | 16116.7 |
| 2012 | Major depressive episode | 16093.8 |
| 2022 | Perceived mental health | 15820.6 |

**Insight:**

* Each year shows different leading indicators: eating disorders (2002), Major depressive episodes (2012), and Perceived mental health (2022), suggesting shifting patterns in mental health challenges over time.
* Notably, in 2022, 'Perceived mental health' emerged as the highest-value indicator, potentially reflecting increased awareness and self-reporting of mental health issues in recent years.
* This analysis reveals how the dominant mental health concerns have evolved over the two decades, with different conditions showing peak values in different years.

## Query 4: **Grouping by gender to show average value (Group By)**

SELECT

d.gender,

ROUND(

AVG(value):: numeric,

2

) as avg\_value,

COUNT(\*) as total\_records

FROM

mental\_health\_stats mhs

INNER JOIN demographics d ON mhs.demographic\_id = d.demographic\_id

GROUP BY

d.gender

ORDER BY

avg\_value DESC

|  |  |  |
| --- | --- | --- |
| **gender** | **avg\_value** | **total\_records** |
| Women | 4015.03 | 4476 |
| Men | 4009.56 | 5400 |

**Insight:**

* There is notably higher data collection for men with 5,400 records compared to women with 4,476 records, suggesting a more comprehensive sampling of male mental health statistics.
* Despite fewer records, women show a slightly higher average value (4015.03) compared to men (4009.56), indicating marginally elevated mental health indicator readings among women.

## Query 5: **Gender comparison by indicator (Multiple joins)**

SELECT

i.indicator\_name,

ROUND(

AVG(

CASE WHEN d.gender = 'Men' THEN mhs.value END

) :: Numeric,

2

) as men\_avg,

ROUND(

AVG(

CASE WHEN d.gender = 'Women' THEN mhs.value END

) :: Numeric,

2

) as women\_avg

FROM

mental\_health\_stats mhs

INNER JOIN demographics d ON mhs.demographic\_id = d.demographic\_id

INNER JOIN indicators i ON mhs.indicator\_id = i.indicator\_id

GROUP BY i.indicator\_name

|  |  |  |
| --- | --- | --- |
| **indicator\_name** | **men\_avg** | **women\_avg** |
| Any mood disorder | 4015.26 | 4008.82 |
| Alcohol abuse or dependence | 4041.85 | 4079.6 |
| Major depressive episode | 4027.95 | 4018.55 |
| Suicidal thoughts | 4007.43 | 4023.92 |
| Generalized anxiety disorder | 4011.95 | 4065.03 |
| Perceived mental health | 3999.99 | 4007.08 |
| Eating disorder | 3997.93 | 3951.94 |
| Cannabis use | 4016.28 | 4007.14 |
| Post-traumatic stress disorder | 3937.27 | 3992.87 |
| Self-rated work stress | 4039.64 | 3995.83 |

**Insight:**

* Dependence and abuse of alcohol had the highest averages for males and females, according to the data, with women reporting somewhat higher rates (4079.6) than men (4041.85).
* There is a significant gender gap in generalized anxiety disorder, with women averaging higher (4065.03) than men (4011.95), indicating that anxiety is more common in women.
* It's interesting to note that eating disorders have lower overall averages than other indicators; in contrast to typical clinical observations, men have a slightly higher average (3997.93) than women (3951.94).

## Query 6: **Categorize values for each indicator as high or low (Case statement)**

SELECT

i.indicator\_name,

CASE WHEN AVG(value) > 4000 THEN 'High' ELSE 'Low' END as category,

ROUND(

AVG(value):: numeric,

2

) as avg\_value

FROM

mental\_health\_stats mhs

INNER JOIN indicators i ON mhs.indicator\_id = i.indicator\_id

GROUP BY i.indicator\_name

|  |  |  |
| --- | --- | --- |
| **indicator\_name** | **category** | **avg\_value** |
| Any mood disorder | High | 4012.35 |
| Alcohol abuse or dependence | High | 4058.9 |
| Major depressive episode | High | 4023.68 |
| Suicidal thoughts | High | 4014.93 |
| Generalized anxiety disorder | High | 4035.93 |
| Perceived mental health | High | 4003.21 |
| Eating disorder | Low | 3977.15 |
| Cannabis use | High | 4012.14 |
| Post-traumatic stress disorder | Low | 3962.55 |
| Self-rated work stress | High | 4019.73 |

**Insight:**

* The analysis reveals most mental health indicators show high prevalence (above 4000), with alcohol abuse and dependence being the most significant at 4058.9.
* Only two indicators - eating disorders (3977.15) and PTSD (3962.55) - fall into the 'Low' category.
* Notably, anxiety, depression, and work stress all show consistently high values, suggesting these are prominent mental health challenges in the population.

## Query 7: **Show top mental health indicator for each province (CTE)**

WITH RegionStats AS (

SELECT

g.geo\_name,

i.indicator\_name,

ROUND(

AVG(mhs.value):: numeric,

2

) as avg\_value

FROM

mental\_health\_stats mhs

INNER JOIN geography g ON mhs.geo\_id = g.geo\_id

INNER JOIN indicators i ON mhs.indicator\_id = i.indicator\_id

GROUP BY

g.geo\_name,

i.indicator\_name

)

SELECT

DISTINCT ON (geo\_name) geo\_name,

indicator\_name,

avg\_value

FROM

RegionStats

ORDER BY

geo\_name,

avg\_value DESC

|  |  |  |
| --- | --- | --- |
| **geo\_name** | **indicator\_name** | **avg\_value** |
| Alberta | Major depressive episode | 4128.24 |
| Atlantic Provinces | Alcohol abuse or dependence | 4125.64 |
| British Columbia | Self-rated work stress | 4120.41 |
| Manitoba | Generalized anxiety disorder | 4143.67 |
| New Brunswick | Generalized anxiety disorder | 4117.52 |
| Newfoundland and Labrador | Alcohol abuse or dependence | 4106.33 |
| Nova Scotia | Generalized anxiety disorder | 4085.4 |
| Ontario | Alcohol abuse or dependence | 4178.96 |
| Prairie Provinces | Suicidal thoughts | 4179.63 |
| Prince Edward Island | Cannabis use | 4112.77 |
| Quebec | Alcohol abuse or dependence | 4213.39 |
| Saskatchewan | Generalized anxiety disorder | 4165.79 |

**Insight:**

* The analysis reveals significant regional variations in mental health indicators across Canada.

* Atlantic Provinces consistently show higher prevalence rates (particularly in depression and anxiety indicators), suggesting a potential need for increased mental health resources in these regions.

* This geographic disparity could be influenced by factors such as access to healthcare services, socioeconomic conditions, or demographic distributions unique to each region.

## Query 8: **Indicator Summary (View)**

CREATE

OR REPLACE VIEW indicator\_summary AS

SELECT

i.indicator\_name,

COUNT(\*) as total\_records,

ROUND(

AVG(value) :: NUMERIC, 2

) as average\_value

FROM

mental\_health\_stats mhs

JOIN indicators i ON mhs.indicator\_id = i.indicator\_id

GROUP BY i.indicator\_name

|  |  |  |
| --- | --- | --- |
| **indicator\_name** | **total\_records** | **average\_value** |
| Any mood disorder | 985 | 4012.35 |
| Alcohol abuse or dependence | 985 | 4058.9 |
| Major depressive episode | 990 | 4023.68 |
| Suicidal thoughts | 990 | 4014.93 |
| Generalized anxiety disorder | 985 | 4035.93 |
| Perceived mental health | 990 | 4003.21 |
| Eating disorder | 985 | 3977.15 |
| Cannabis use | 986 | 4012.14 |
| Post-traumatic stress disorder | 990 | 3962.55 |
| Self-rated work stress | 990 | 4019.73 |

**Insight/Rationale:**

The indicator view provides a high-level overview of each mental health indicator with its basic statistics. It helps stakeholders quickly understand:

1. How frequently each indicator is measured (total records)
2. The average prevalence of each condition (average\_value) This is particularly useful for healthcare administrators who need quick insights into the most common or severe mental health condition.

## Query 9: **Demographic Risk Assessment**

CREATE OR REPLACE VIEW demographic\_risk\_assessment AS

WITH IndicatorRanks AS (

SELECT

d.age\_group,

d.gender,

i.indicator\_name,

ROUND(AVG(mhs.value)::numeric, 2) as avg\_value,

RANK() OVER (PARTITION BY i.indicator\_name ORDER BY AVG(mhs.value) DESC) as severity\_rank

FROM mental\_health\_stats mhs

JOIN demographics d ON mhs.demographic\_id = d.demographic\_id

JOIN indicators i ON mhs.indicator\_id = i.indicator\_id

GROUP BY d.age\_group, d.gender, i.indicator\_name

)

SELECT

age\_group,

gender,

indicator\_name,

avg\_value,

CASE

WHEN severity\_rank <= 3 THEN 'High Risk'

WHEN severity\_rank <= 6 THEN 'Medium Risk'

ELSE 'Low Risk'

END as risk\_category

FROM IndicatorRanks

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **age\_group** | **gender** | **indicator\_name** | **avg\_value** | **risk\_category** |
| 45-64 | Women | Alcohol abuse or dependence | 4123.81 | High Risk |
| 15-24 | Men | Alcohol abuse or dependence | 4092.06 | High Risk |
| 45-64 | Men | Alcohol abuse or dependence | 4080.01 | High Risk |
| 15-24 | Women | Alcohol abuse or dependence | 4072.92 | Medium Risk |
| 25-44 | Women | Alcohol abuse or dependence | 4065.4 | Medium Risk |
| 25-44 | Men | Alcohol abuse or dependence | 3953.48 | Medium Risk |
| 15-24 | Women | Any mood disorder | 4061.11 | High Risk |
| 45-64 | Men | Any mood disorder | 4049.84 | High Risk |
| 15-24 | Men | Any mood disorder | 4011.7 | High Risk |
| 45-64 | Women | Any mood disorder | 3988.01 | Medium Risk |

**Insight/Rationale:**

Demographic Risk View helps identify vulnerable populations by:

* Breaking down mental health indicators by age and gender
* Ranking the severity of conditions across demographic groups
* Categorizing risk levels (High, Medium, Low) This information is crucial for healthcare providers and policymakers to target interventions and allocate resources to the most at-risk populations.

## Query 10: **Regional Health Dashboard (View)**

CREATE

OR REPLACE VIEW regional\_mental\_health\_dashboard AS

SELECT

g.geo\_name,

i.indicator\_name,

i.uom,

ROUND (

AVG(mhs.value):: numeric,

2

) as average\_value,

MIN (mhs.value) as minimum\_value,

MAX (mhs.value) as maximum\_value,

COUNT(\*) as number\_of\_measurements

FROM

mental\_health\_stats mhs

JOIN geography g ON mhs.geo\_id = g.geo\_id

JOIN indicators i ON mhs.indicator\_id = i.indicator\_id

GROUP BY

g.geo\_name,

i.indicator\_name,

i.uom

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **geo\_name** | **indicator\_name** | **uom** | **average\_value** | **minimum\_value** | **maximum\_value** | **number\_of\_measurements** |
| Alberta | Alcohol abuse or dependence | Number | 4075.7 | 0 | 15137.1 | 75 |
| Atlantic Provinces | Alcohol abuse or dependence | Number | 4125.64 | 0 | 14395.9 | 90 |
| British Columbia | Alcohol abuse or dependence | Number | 3948.77 | 0 | 14730.9 | 75 |
| Manitoba | Alcohol abuse or dependence | Number | 3930.22 | 0 | 14504.1 | 75 |
| New Brunswick | Alcohol abuse or dependence | Number | 4020.05 | 0 | 14681.4 | 90 |
| Newfoundland and Labrador | Alcohol abuse or dependence | Number | 4106.33 | 0 | 14333.8 | 90 |
| Nova Scotia | Alcohol abuse or dependence | Number | 3963.74 | 0 | 14402.6 | 90 |
| Ontario | Alcohol abuse or dependence | Number | 4178.96 | 0 | 14687.2 | 75 |
| Prairie Provinces | Alcohol abuse or dependence | Number | 4073.3 | 0 | 14209.5 | 75 |
| Prince Edward Island | Alcohol abuse or dependence | Number | 3977.04 | 0 | 15023.1 | 90 |

**Insight/Rationale:**

The Regional Dashboard view serves as a comprehensive regional analysis tool by showing:

* Geographic distribution of mental health conditions
* Range of values (minimum to maximum) for each indicator
* Measurement frequency by region
* Unit of measurement context: This helps regional health authorities understand local mental health patterns and make data-driven decisions about resource allocation.

# Conclusion

The analysis of Canadian mental health data (2002-2022) revealed significant trends across demographics and regions. Women, specifically in the age group of 15-24, consistently showed higher prevalence of mental health problems. Alcohol abuse was the most prevalent issue (4058.9), followed by anxiety (4035.93) and depression (4023.68). When compared to other regions, the Atlantic Provinces showed noticeably higher values for mental health indicators. These results can be utilized to inform targeted interventions and resource distribution in public health planning, particularly for at-risk populations.

# References

Government of Canada, Statistics Canada. (2024, December 12). *Mental health indicators*. <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310046501>

TheKnowledgeAcademy. (n.d.). *Basic and Advanced SQL Functions with Examples*. <https://www.theknowledgeacademy.com/blog/sql-functions/>

*Normalization in SQL DBMS: 1NF, 2NF, 3NF, and BCNF Examples | PopSQL*. (n.d.). <https://popsql.com/blog/normalization-in-sql>

Neon. (2024, February 1). *PostgreSQL CAST: Convert a value of One Type to Another*. Neon. https://neon.tech/postgresql/postgresql-tutorial/postgresql-cast#postgresql-cast-function-and-cast-operator-examples

Decomplexify. (2021, November 21). *Learn Database Normalization - 1NF, 2NF, 3NF, 4NF, 5NF* [Video]. YouTube. https://www.youtube.com/watch?v=GFQaEYEc8\_8

EDB. (2021, February 23). *PGAdmin: Entity-Relationship Diagram (ERD) Tool Demo - [PostgreSQL Tutorial]* [Video]. YouTube. <https://www.youtube.com/watch?v=2pxVCzRFGeg>

Page, D. (n.d.). *Import/Export Data Dialog — pgAdmin 4 9.0 documentation*. https://www.pgadmin.org/docs/pgadmin4/development/import\_export\_data.html