**MEDICINE/DRUGS DATASET ANALYSIS**

**-- Clean data --**

**-- Remove 'mg' from Strength column**

UPDATE medicine\_dataset

SET Strength = cast(REMOVE(Strength, ' mg','')AS UNSIGNED);

**-- set Strength column to Integer**

ALTER TABLE medicine\_dataset

MODIFY COLUMN Strength INT;

**-- change 'name' column to medicine\_name**

ALTER TABLE medicine\_dataset

CHANGE COLUMN name medicine\_name CHAR (100);

**-- change 'dosage form' column to 'dosage\_form'**

ALTER TABLE medicine\_dataset

CHANGE COLUMN `dosage form` dosage\_form CHAR(100);

**-- 1. Most Common Medicines in Each Category:**

**-- Identify which specific medicines are most frequently listed within each therapeutic category.**

WITH common\_medicine as

(SELECT category, medicine\_name, count(\*) as medi\_count

from medicine\_dataset

GROUP BY category, medicine\_name

),

rank\_med AS

(SELECT category,medicine\_name, medi\_count,

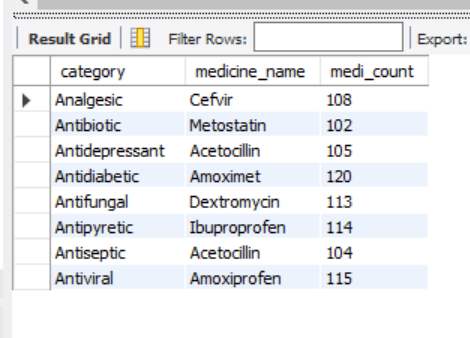
ROW\_NUMBER () OVER (PARTITION BY category ORDER BY medi\_count DESC) AS `rank`

FROM common\_medicine)

SELECT category, medicine\_name, medi\_count

FROM rank\_med

where `rank` = 1;



**-- 2. Average Strength by Dosage Form: Determine the average strength of medicines for each dosage form.**

WITH average\_med AS

( SELECT dosage\_form, avg(strength) AS avg\_strength

FROM medicine\_dataset

GROUP BY dosage\_form

),

rank\_avg AS

(SELECT dosage\_form, avg\_str,

ROW\_NUMBER() OVER (PARTITION BY dosage\_form ORDER BY avg\_str desc) as `rank`

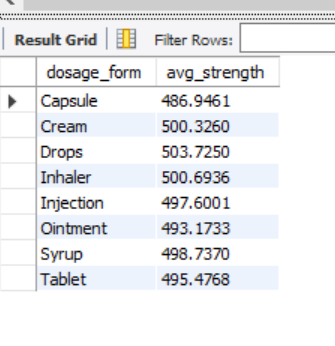
FROM average\_med

)

SELECT dosage\_form, avg\_str

FROM rank\_avg

where `rank` = 1;



**-- 3. Distribution of Medicines by Manufacturer: Count how many different medicines each manufacturer produces.**

WITH manufacturer AS

( SELECT manufacturer, count(\*) as medicine\_count

FROM medicine\_dataset

GROUP BY manufacturer

),

rank\_manufacturer AS

( select manufacturer, medicine\_count,

ROW\_NUMBER () OVER (PARTITION BY manufacturer ORDER BY medicine\_count desc) AS `rank`

FROM manufacturer

)

SELECT manufacturer, medicine\_count

FROM rank\_manufacturer

where `rank` = 1;



**-- 4. Count of Medicines by Indication: Find out how many medicines are available for each medical indication.**

WITH med\_indication AS

(SELECT indication, count(\*) as medname\_count

FROM medicine\_dataset

GROUP BY indication

),

rank\_indication AS

( SELECT indication, medname\_count,

ROW\_NUMBER () OVER (PARTITION BY indication ORDER BY medname\_count desc) AS `rank`

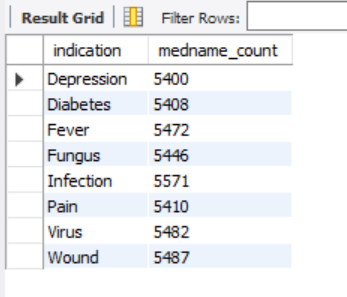
FROM med\_indication

)

SELECT indication, medname\_count

FROM rank\_indication

WHERE `rank` = 1;



**-- 5. Prescription vs. Over-the-Counter Medicines by Category:**

**-- Compare the number of prescription and over-the-counter medicines within each therapeutic category.**

WITH med\_classification AS

(select classification, count(\*) as medname\_count

FROM medicine\_dataset

GROUP BY classification

),

rank\_classification AS

(SELECT classification, medname\_count,

ROW\_NUMBER() OVER (PARTITION BY classification ORDER BY medname\_count desc) AS `rank`

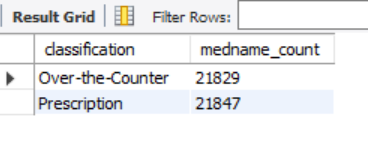
FROM med\_classification

)

SELECT classification, medname\_count

FROM rank\_classification

WHERE `rank` = 1;



**-- 6. Top Manufacturers for Specific Indications:**

**-- Identify which manufacturers produce the most medicines for particular medical conditions.**

WITH manu\_indication AS

(SELECT indication, manufacturer, count(\*) AS medname\_count

FROM medicine\_dataset

GROUP BY indication, manufacturer

),

rank\_manu\_indication AS

(SELECT indication, manufacturer, medname\_count,

ROW\_NUMBER() OVER (PARTITION BY indication ORDER BY medname\_count desc) as `rank`

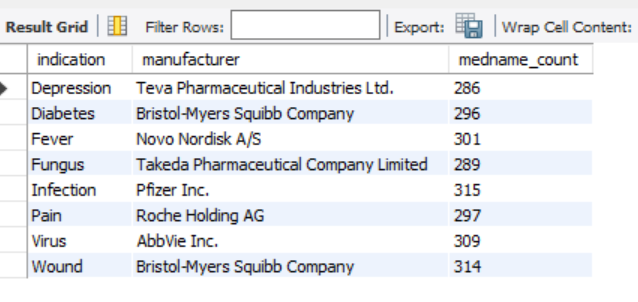
FROM manu\_indication

)

SELECT indication, manufacturer, medname\_count

FROM rank\_manu\_indication

WHERE `rank` = 1;

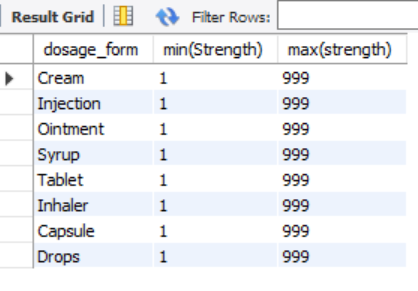


**-- 7. Strength Range for Each Dosage Form: Determine the range of strengths available for each dosage form.**

SELECT dosage\_form, min(Strength), max(strength)

FROM medicine\_dataset

GROUP BY dosage\_form;



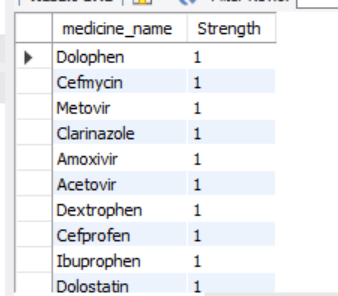
**-- 8. Medicines with the Highest and Lowest Strengths: Identify which medicines have the highest and lowest strengths.**

**-- Lowest**

SELECT medicine\_name, Strength

FROM medicine\_dataset

WHERE Strength = (SELECT min(Strength) FROM medicine\_dataset);

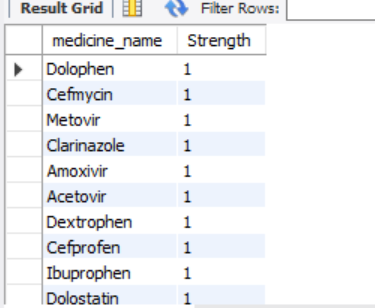


**-- Highest**

SELECT medicime\_name, Strength

FROM medicine\_dataset

WHERE Strength = (SELECT max(Strength) FROM medicine\_dataset);



**-- 9. Most Common Dosage Forms for Specific Categories:**

**-- Find out which dosage forms are most commonly associated with each therapeutic category.**

WITH dosage\_category AS

(SELECT dosage\_form, category, count(\*) AS dosage\_count

FROM medicine\_dataset

GROUP BY dosage\_form, category

),

rank\_dosage\_category AS

(SELECT dosage\_form, category, dosage\_count,

ROW\_NUMBER () OVER (PARTITION BY category ORDER BY dosage\_count desc) AS `rank`

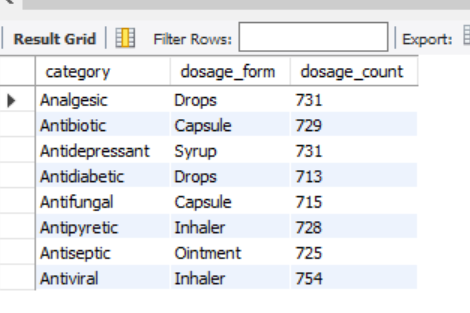
FROM dosage\_category

)

SELECT category, dosage\_form, dosage\_count

FROM rank\_dosage\_category

WHERE `rank` = 1;



Data Source: https://www.kaggle.com/datasets/ujjwalaggarwal402/medicine-dataset

GitHub Repository: <https://github.com/Florence93/Medicine-Drugs-Data-Analysis>

Author: ©2024 Florence Idowu