Group 7

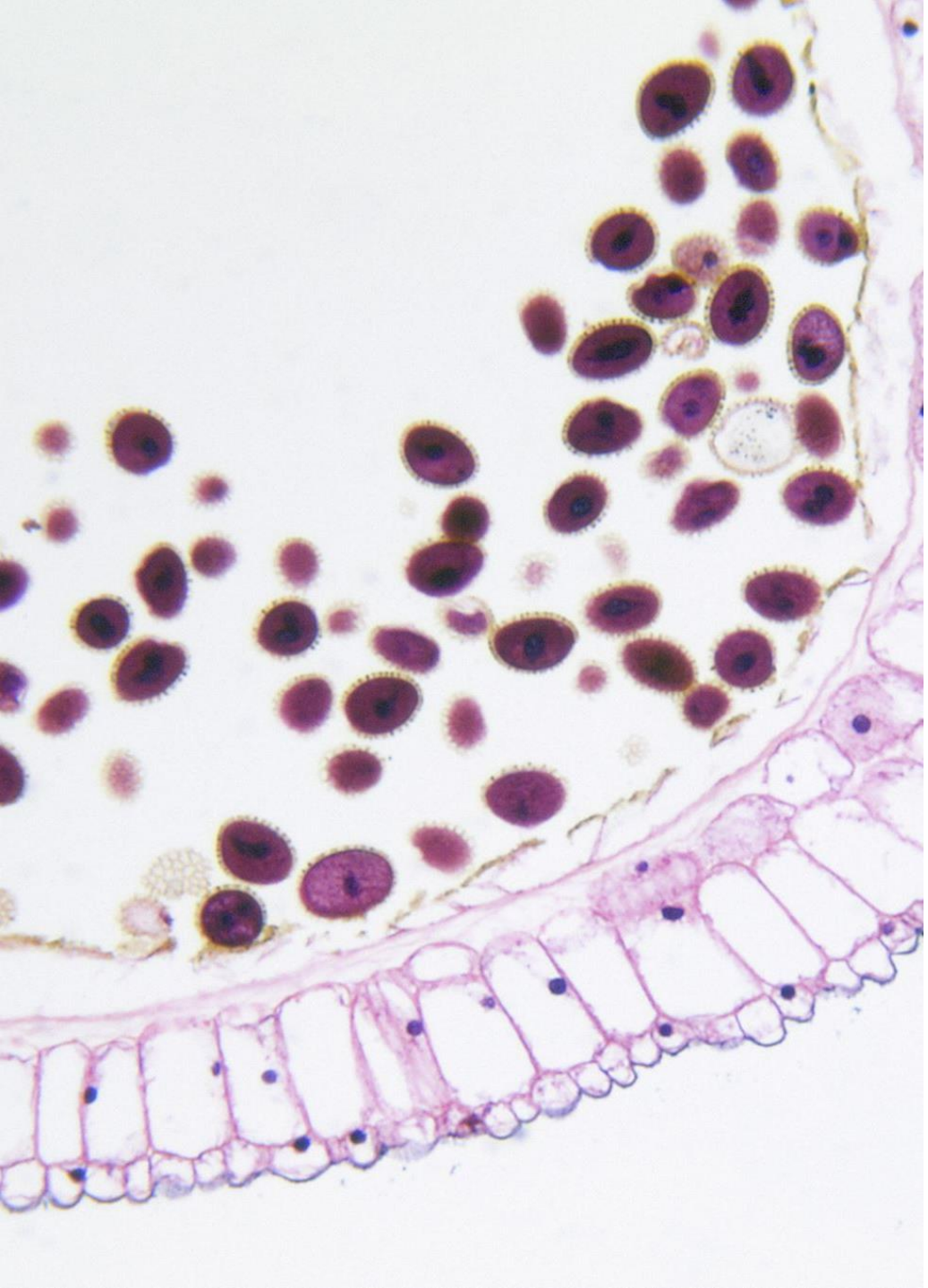
SQL Capstone Project

Analysis of DermAI Diagnostics Skin Cancer Dataset

AMADI JERRY

Background Information/Facts

- There are **1,088** distinct patients in the Patients_Info Table (**Table 1)**
- Similarly, there are also **1,088** lesion IDs in Lesion_Info Table **(Table 2)**
- There are six (6) types of Skin Lesions/Cancers affecting patients from the dataset
    - They are: **ACK, BCC, MEL, NEV, SCC & SEK**
- **All Patients** from the given dataset have one type of the six (6) listed skin lesions/cancers
- All Patients have one of the seven (7) types of **skin pigmentation** *(i.e. the **Fitzpatrick Skin Type**)*
    - The Fitzpatrick scale is a system used to classify skin types based on their pigmentation and response to ultraviolet (UV) radiation.
    - The scale ranges from **0 *(Very Very Fair Skin)*** to **Type 6 *(Deeply pigmented Brown or Black Skin)***



Key Facts on Skin Lesions

According to **The American Academy of Dermatology**, skin lesions are either **Benign** or **Malignant skin cancers**.

- **Benign** skin lesions are non-cancerous growths that typically do not spread to other parts of the body as they tend to remain confined to their original location. They are generally harmless and may only require observation or cosmetic treatment.

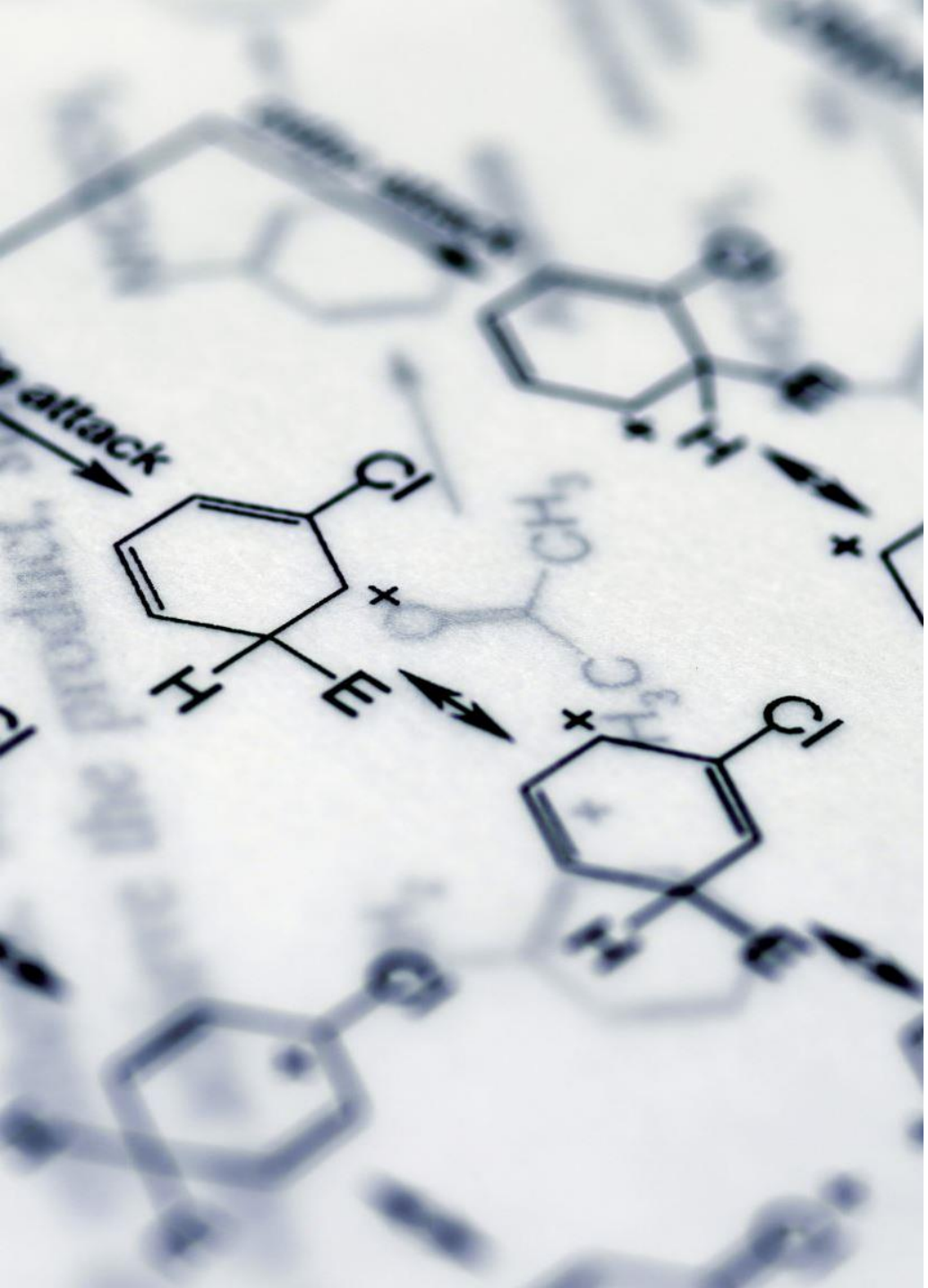
  Examples are:

  i. **NEV** – Nevus *(commonly known as moles)*
  ii. **SEK** – Seborrheic Keratosis

- **Malignant** skin lesions are cancerous and metastasizes to other organs and tissues through the bloodstream or lymphatic system. They can be life-threatening & terminal if left untreated. They grow rapidly and most times require aggressive treatment.

  Examples are:

  i. **BCC** – Basal Cell Carcinoma
  ii. **SCC** – Squamous Cell Carcinoma
  iii. **ACK** – Actinic Keratosis
  iv. **MEL** – Melanoma



Considerations In Analyzing The Dataset

- What proportion of the patients from dataset have malignant skin cancers?
- What factors/drivers predisposes patients to the malignant forms of cancers
  - Demographic Drivers ***(****e.g.* ***Ethnicity, Gender, Age and Skin Pigmentation)***
  - Lifestyle Choices/Behaviors *e.g.* ***Smoking, drinking***
  - Environmental Risk Factors *e.g.* ***Exposure to pesticides, access to piped water & sewage system***
  - Characteristics of the Skin Lesions ***(****e.g.* ***Lesion Itches, Pain, Bleeding, Color*** *&* ***Size)***
  - Medical History e.g. ***Previous Skin Cancer Diagnosis*** *&* ***History of Cancers in the Family***
- How would findings/data from the analysis enhance early-stage diagnosis?
- Create a machine learning- ready data set to support early detection of skin cancers

Demographics Analysis on Severity of Skin Lesions

*Analysis of **Ethnicity, Gender, Age, Previous Skin Cancer Diagnosis** & **Family History** in Relation to Skin Lesion Diagnostic Types*

Query | Query History

```
209
210  -- Disaggregate all Patients By Gender (Sex)
211
212  SELECT
213      Gender,
214      COUNT(DISTINCT Patient_id) AS patient_count,
215      ROUND(
216          100.0 * COUNT(DISTINCT Patient_id) / SUM(COUNT(DISTINCT Patient_id)) OVER (),
217          2
218      ) AS percentage_of_total
219  FROM Table1
220  GROUP BY Gender;
221
```

Data Output | Messages | Notifications

Showing rows: 1 to 2

	gender character varying (10)	patient_count bigint	percentage_of_total numeric
1	FEMALE	362	33.27
2	MALE	726	66.73

Query | Query History

```sql
--- Analysis (Count & Rate) of Skin Lesions Among Female Patients

SELECT
    diagnostic_type,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
FROM (
    SELECT DISTINCT
        t1.Patient_ID,
        CASE
            WHEN t2.Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
            ELSE 'Non-Cancerous'
        END AS diagnostic_type
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.Gender = 'FEMALE'
) AS categorized
GROUP BY diagnostic_type
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 2

	diagnostic_type text	patient_count bigint	percentage numeric
1	Cancerous	292	80.66
2	Non-Cancerous	70	19.34

Query | Query History

```sql
--- Disaggregation of the Types of Skin Lesions (Diagnostics) By Gender
SELECT
    t2.Diagnostic,
    COUNT(*) AS diagnostic_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS diagnostic_percentage
FROM Table1 t1
JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
WHERE t1.Gender = 'FEMALE'
  AND t2.Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC', 'NEV', 'SEK')
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	diagnostic_count bigint	diagnostic_percentage numeric
1	ACK	135	37.29
2	BCC	122	33.70
3	NEV	42	11.60
4	SEK	28	7.73
5	SCC	25	6.91
6	MEL	10	2.76

Query | Query History

```sql

--- Analysis (Count & Rate) of Skin Lesions Among Male Patients

SELECT
    diagnostic_type,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
FROM (
    SELECT DISTINCT
        t1.Patient_ID,
        CASE
            WHEN t2.Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
            ELSE 'Non-Cancerous'
        END AS diagnostic_type
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.Gender = 'MALE'
) AS categorized
GROUP BY diagnostic_type
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 2

	diagnostic_type text	patient_count bigint	percentage numeric
1	Cancerous	515	70.94
2	Non-Cancerous	211	29.06

Query | Query History

```sql
--- Disaggregation of the Types of Skin Lesions (Diagnostics) By Gender
SELECT
    t2.Diagnostic,
    COUNT(*) AS diagnostic_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS diagnostic_percentage
FROM Table1 t1
JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
WHERE t1.Gender = 'MALE'
  AND t2.Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC', 'NEV', 'SEK')
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	diagnostic_count bigint	diagnostic_percentage numeric
1	ACK	326	44.90
2	BCC	151	20.80
3	SEK	109	15.01
4	NEV	102	14.05
5	SCC	31	4.27
6	MEL	7	0.96

Query | Query History

```
634
635
636  --Age Groups
637
638  SELECT DISTINCT table2.diagnostic, SUM(CASE WHEN table1.age <= 20 then 1
639  else 0 END) AS "<= 20years", SUM(CASE WHEN table1.age BETWEEN 21 AND 40 then 1
640  else 0 END) AS "21 to 40 Yrs",
641      SUM(CASE WHEN table1.age > 40 then 1
642  else 0 END) AS "> 40years"
643  FROM table2
644  LEFT JOIN table1 on table1.patient_id = table2.patient_id
645  GROUP BY table2.diagnostic
646  ORDER BY table2.diagnostic;
647
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	<= 20years bigint	21 to 40 Yrs bigint	> 40years bigint
1	ACK	1	30	430
2	BCC	0	17	256
3	MEL	0	0	17
4	NEV	22	79	43
5	SCC	0	2	54
6	SEK	0	6	131

```
312
313
314  -- Count & % Percentage of Patients By Paternal Ethnicity
315  SELECT
316      Background_father,
317      COUNT(DISTINCT Patient_id) AS patient_count,
318      ROUND(
319          100.0 * COUNT(DISTINCT Patient_id) / SUM(COUNT(DISTINCT Patient_id)) OVER (),
320          2
321      ) AS percentage_of_total
322  FROM Table1
323  GROUP BY Background_father
324  ORDER BY percentage_of_total Desc
325  LIMIT 10;
326
```

	background_father character varying (255)	patient_count bigint	percentage_of_total numeric
1	UNKNOWN	584	53.68
2	POMERANIA	183	16.82
3	GERMANY	156	14.34
4	ITALY	75	6.89
5	BRAZIL	39	3.58
6	UNK	33	3.03
7	NETHERLANDS	6	0.55
8	PORTUGAL	6	0.55
9	BRASIL	2	0.18
10	POLAND	2	0.18

Total rows: 10 | Query complete 00:00:00.317

Query | Query History

```sql
-- Count & % Percentage of Patients By Maternal Ethnicity
SELECT
    Background_mother,
    COUNT(DISTINCT Patient_id) AS patient_count,
    ROUND(
        100.0 * COUNT(DISTINCT Patient_id) / SUM(COUNT(DISTINCT Patient_id)) OVER (),
        2
    ) AS percentage_of_total
FROM Table1
GROUP BY Background_mother
ORDER BY percentage_of_total Desc
LIMIT 10;
```

Data Output | Messages | Notifications

Showing rows: 1 to 10

	background_mother character varying (255)	patient_count bigint	percentage_of_total numeric
1	UNKNOWN	583	53.58
2	POMERANIA	186	17.10
3	GERMANY	161	14.80
4	ITALY	75	6.89
5	BRAZIL	34	3.13
6	UNK	30	2.76
7	NETHERLANDS	9	0.83
8	PORTUGAL	6	0.55
9	NORWAY	1	0.09
10	SPAIN	1	0.09

Total rows: 10 | Query complete 00:00:00.199

Query | Query History

```
194
195  --- Disaggregation of Diagnostics Counts/Percentage by Skin Pigmentation (Fitzpatrick Classifca
196  SELECT
197      Fitspatrick,
198      COUNT(*) AS diagnostic_count,
199      ROUND(
200          100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
201          2
202      ) AS diagnostic_percentage
203  FROM Table2
204  WHERE Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC', 'NEV', 'SEK')
205  GROUP BY Fitspatrick
206  Order By diagnostic_percentage Desc;
```

Data Output | Messages | Notifications

SQL

Showing rows: 1 to 7

	fitspatrick integer	diagnostic_count bigint	diagnostic_percentage numeric
1	0	579	53.22
2	2	285	26.19
3	3	142	13.05
4	1	46	4.23
5	4	28	2.57
6	5	7	0.64
7	6	1	0.09

Query | Query History

```sql
--- Rate of Cancerous Lesions by Skin Pigmentation

SELECT
    fitspatrick,
    COUNT(*) AS cancerous_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS cancerous_percentage
FROM Table2
WHERE Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC')
GROUP BY fitspatrick
Order By cancerous_percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	fitspatrick integer	cancerous_count bigint	cancerous_percentage numeric
1	0	328	40.64
2	2	273	33.83
3	3	135	16.73
4	1	43	5.33
5	4	22	2.73
6	5	6	0.74

Query | Query History

```
206
207
208  -- Count & % Percentage of Skins Lesions By Body Regions Affected
209  SELECT
210      Region,
211      COUNT(DISTINCT Patient_id) AS patient_count,
212      ROUND(
213          100.0 * COUNT(DISTINCT Patient_id) / SUM(COUNT(DISTINCT Patient_id)) OVER (),
214          2
215      ) AS percentage_of_total
216  FROM Table2
217  GROUP BY Region
218  ORDER BY percentage_of_total Desc
219  LIMIT 10;
220
```

Data Output | Messages | Notifications

SQL

Showing rows: 1 to 10

	region character varying (255)	patient_count bigint	percentage_of_total numeric
1	FACE	278	25.55
2	FOREARM	219	20.13
3	CHEST	124	11.40
4	BACK	105	9.65
5	ARM	92	8.46
6	NOSE	68	6.25
7	HAND	55	5.06
8	NECK	37	3.40
9	THIGH	35	3.22
10	EAR	32	2.94

Total rows: 10 | Query complete 00:00:00.290

Query | Query History

```sql
--- Disaggregation of Top 5 Regions (Body location of Skin Lesions) By Gender
SELECT
    t2.Region,
    COUNT(*) AS region_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS region_percentage
FROM Table1 t1
JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
WHERE t1.Gender = 'MALE'
  AND t2.Region IS NOT NULL
GROUP BY t2.Region
ORDER BY region_percentage DESC
LIMIT 5;
```

Data Output | Messages | Notifications

Showing rows: 1 to 5

	region character varying (255)	region_count bigint	region_percentage numeric
1	FACE	186	25.62
2	FOREARM	164	22.59
3	CHEST	87	11.98
4	BACK	70	9.64
5	ARM	57	7.85

Query | Query History

```sql
--- Disaggregation of Top 5 Regions (Body location of Skin Lesions) By Gender
SELECT
    t2.Region,
    COUNT(*) AS region_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS region_percentage
FROM Table1 t1
JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
WHERE t1.Gender = 'FEMALE'
  AND t2.Region IS NOT NULL
GROUP BY t2.Region
ORDER BY region_percentage DESC
LIMIT 5;
```

Data Output | Messages | Notifications

Showing rows: 1 to 5

	region character varying (255)	region_count bigint	region_percentage numeric
1	FACE	92	25.41
2	FOREARM	55	15.19
3	NOSE	41	11.33
4	CHEST	37	10.22
5	ARM	35	9.67

Effect of Lifestyle Choices on Severity of Skin Lesions/Cancers

*Analysis of **Smoking** & **Drinking** in Relation to Skin Lesion Diagnostic Types*

Query | Query History

```sql

--- Degree/Scale of Cancerous Skin Lesions on Patients Who SMOKE

SELECT
    diagnostic_type,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
FROM (
    SELECT DISTINCT
        t1.Patient_ID,
        CASE
            WHEN t2.Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
            ELSE 'Non-Cancerous'
        END AS diagnostic_type
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.Smoke = 'TRUE'
) AS categorized
GROUP BY diagnostic_type
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 2

	diagnostic_type text	patient_count bigint	percentage numeric
1	Cancerous	60	96.77
2	Non-Cancerous	2	3.23

Total rows: 2 | Query complete 00:00:00.117

Query | Query History

```sql
SELECT
    diagnostic,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.smoke = 'TRUE'
GROUP BY diagnostic
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	patient_count bigint	percentage numeric
1	BCC	30	48.39
2	SCC	15	24.19
3	ACK	13	20.97
4	MEL	2	3.23
5	SEK	1	1.61
6	NEV	1	1.61

Query | Query History

```sql
--- Analysis (Count & Rate) of Cancerous Skin Lesions Severity Among Non-Smoking Patients

SELECT
    diagnostic_type,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
FROM (
    SELECT DISTINCT
        t1.Patient_ID,
        CASE
            WHEN t2.Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
            ELSE 'Non-Cancerous'
        END AS diagnostic_type
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.Smoke = 'FALSE'
) AS categorized
GROUP BY diagnostic_type
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 2

	diagnostic_type text	patient_count bigint	percentage numeric
1	Cancerous	747	72.81
2	Non-Cancerous	279	27.19

Total rows: 2 | Query complete 00:00:00.224

Query | Query History

```sql
SELECT
    diagnostic,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.smoke = 'FALSE'
GROUP BY diagnostic
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	patient_count bigint	percentage numeric
1	ACK	448	43.66
2	BCC	243	23.68
3	NEV	143	13.94
4	SEK	136	13.26
5	SCC	41	4.00
6	MEL	15	1.46

Query | Query History

```sql

--- Degree/Scale of Cancerous Skin Lesions on Patients Who Drink

SELECT
    diagnostic_type,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
FROM (
    SELECT DISTINCT
        t1.Patient_ID,
        CASE
            WHEN t2.Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
            ELSE 'Non-Cancerous'
        END AS diagnostic_type
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.Drink = 'TRUE'
) AS categorized
GROUP BY diagnostic_type
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 2

	diagnostic_type text	patient_count bigint	percentage numeric
1	Cancerous	131	94.93
2	Non-Cancerous	7	5.07

Query | Query History

```sql

--- Analysis (Count & Rate) of Cancerous Skin Lesions Severity Among Non-Drinking Patients

SELECT
    diagnostic_type,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
FROM (
    SELECT DISTINCT
        t1.Patient_ID,
        CASE
            WHEN t2.Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
            ELSE 'Non-Cancerous'
        END AS diagnostic_type
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.Drink = 'FALSE'
) AS categorized
GROUP BY diagnostic_type
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 2

	diagnostic_type text	patient_count bigint	percentage numeric
1	Cancerous	676	71.16
2	Non-Cancerous	274	28.84

Environmental Exposure on Severity of Skin Lesions

*Analysis of **Exposure to Pesticides, Access to Piped Water & Sewage System** in Relation to Skin Lesion Diagnostic Types*

Query | Query History

```
41
42
43
44
45
46
47
48  --- Count the total number of patients exposed to pesticides in the Patients table
49
50
51  SELECT COUNT(Patient_ID) AS Patients_Exposed_To_Pesticides
52  FROM Table1 WHERE Pesticide = 'TRUE';
53
```

Data Output | Messages | Notifications

Showing rows: 1 to 1

	patients_exposed_to_pesticides bigint
1	223

Query | Query History

```sql
-- Calculate % Percentage of Patients Exposed to Pesticides

SELECT
(SELECT
(SELECT COUNT(Patient_ID) FROM Table1 WHERE Pesticide = 'TRUE') * (1.0) / COUNT(*)
FROM Table1) * 100 AS "Percentage Exposure";

SELECT
    ROUND(
        100.0 * COUNT(*) FILTER (WHERE Pesticide = 'TRUE') / COUNT(*),
        2
    ) AS "Percentage Exposure"
FROM Table1;
```

Data Output | Messages | Notifications

Showing rows: 1 to 1

	Percentage Exposure numeric
1	20.50

Query | Query History

```sql
--- Effects of Patients' Exposure to Pesticides In Relation to Cancerous Skin Lesions

SELECT
    diagnostic,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.Pesticide = 'TRUE'
GROUP BY diagnostic
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	patient_count bigint	percentage numeric
1	BCC	127	56.95
2	ACK	62	27.80
3	SCC	22	9.87
4	MEL	7	3.14
5	NEV	3	1.35
6	SEK	2	0.90

Query | Query History

```sql
--- Effects of Patients' Exposure to Pesticides In Relation to Cancerous Skin Lesions

SELECT
    diagnostic_type,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
FROM (
    SELECT DISTINCT
        t1.Patient_ID,
```

Data Output | Messages | Notifications

Showing rows: 1 to 2

	diagnostic_type text	patient_count bigint	percentage numeric
1	Cancerous	218	97.76
2	Non-Cancerous	5	2.24

Query | Query History

```
272
273
274 --- Analysis (Count & Rate) of Cancerous Skin Lesions on Patients NOT Exposed To Pestiicides
275
276 SELECT
277     diagnostic_type,
278     COUNT(*) AS patient_count,
279     ROUND(
280         100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
281         2
282     ) AS percentage
283 FROM (
284     SELECT DISTINCT
285         t1.Patient_ID,
```

Data Output | Messages | Notifications

Showing rows: 1 to 2

	diagnostic_type text	patient_count bigint	percentage numeric
1	Cancerous	589	68.09
2	Non-Cancerous	276	31.91

Query | Query History

```sql
--- Effects of Patients' Access to Piped Water In Relation to Cancerous Skin Lesions

SELECT
    diagnostic,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.has_piped_water = 'FALSE'
GROUP BY diagnostic
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	patient_count bigint	percentage numeric
1	ACK	371	47.44
2	NEV	132	16.88
3	SEK	128	16.37
4	BCC	123	15.73
5	SCC	20	2.56
6	MEL	8	1.02

Query | Query History

```sql
--- Effects of Patients' Access to Sewage System In Relation to Cancerous Skin Lesions

SELECT
    diagnostic,
    COUNT(*) AS patient_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS percentage
    FROM Table1 t1
    JOIN Table2 t2 ON t1.Patient_ID = t2.Patient_ID
    WHERE t1.has_sewage_system = 'FALSE'
GROUP BY diagnostic
ORDER BY percentage Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	patient_count bigint	percentage numeric
1	ACK	378	46.38
2	BCC	146	17.91
3	NEV	134	16.44
4	SEK	128	15.71
5	SCC	21	2.58
6	MEL	8	0.98

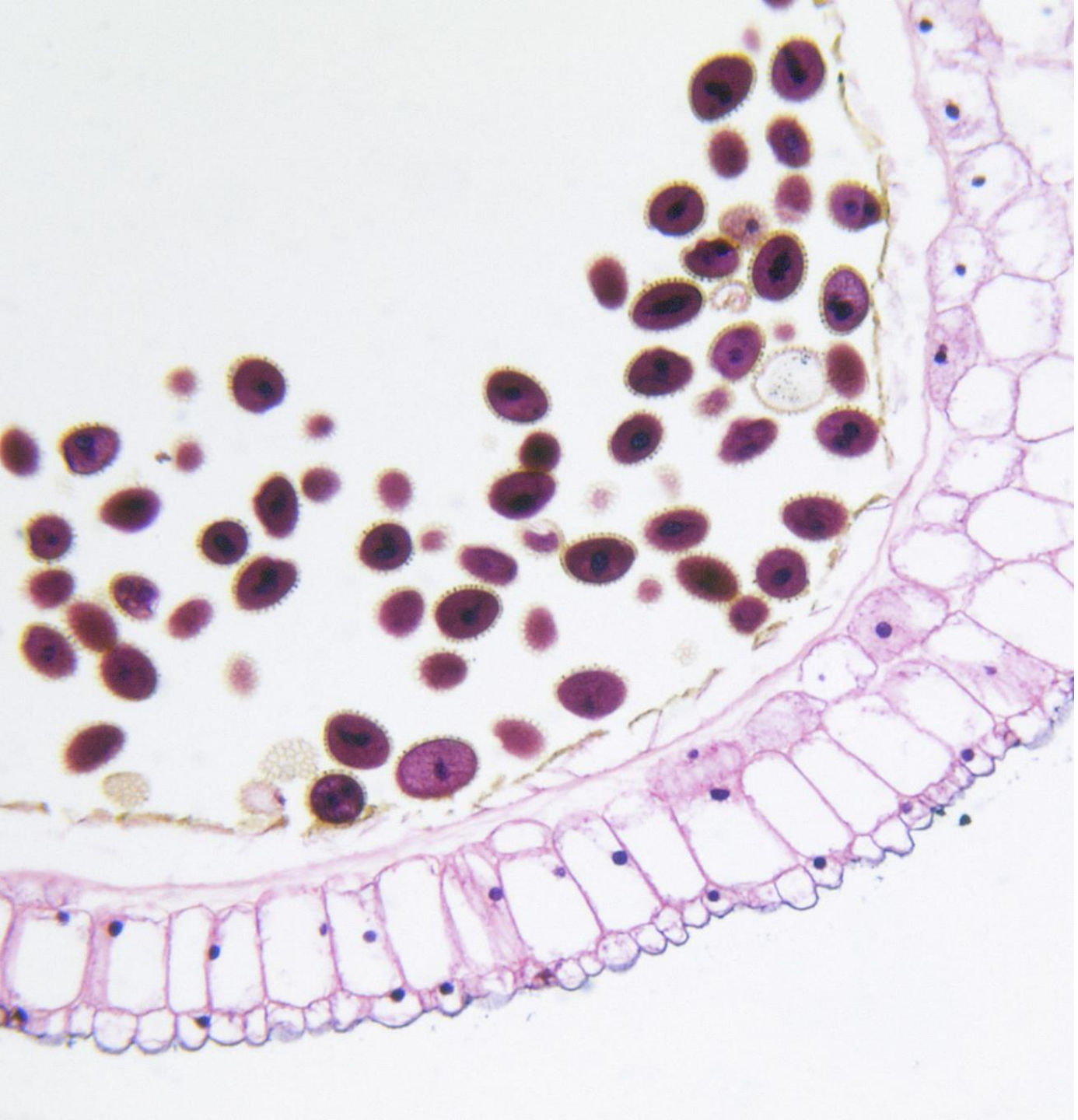
```sql
-- Environmental Factors  Vs Skin Lesions Types


SELECT table2.diagnostic, SUM(CASE WHEN table1.pesticide = 'TRUE' then 1
else 0 END) AS pesticides,
    SUM(CASE WHEN table1.has_piped_water = 'FALSE' then 1
else 0 END) AS no_piped_water,
    SUM(CASE WHEN table1.has_sewage_system = 'FALSE' then 1
else 0 END) AS no_sewage_system
FROM table2
LEFT JOIN table1 on table1.patient_id = table2.patient_id
GROUP BY table2.diagnostic
ORDER BY no_sewage_system Desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	pesticides bigint	no_piped_water bigint	no_sewage_system bigint
1	ACK	62	371	378
2	BCC	127	123	146
3	NEV	3	132	134
4	SEK	2	128	128
5	SCC	22	20	21
6	MEL	7	8	8



Lesion Characteristics To Different Types of Skin Lesions

*Analysis of **Lesion Itches, Pain, Bleeding, Color** & **Size** in Relation to Skin Lesion Diagnostic Types*

Query | Query History

```sql
--- Analysis of Itchy Lesion As An Indicator of Malignant Cancers

SELECT
    Diagnostic,
    CASE
        WHEN Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
        WHEN Diagnostic = 'NEV' THEN 'Non-Cancerous'
        ELSE 'Non-Cancerous'
    END AS diagnostic_type,
    COUNT(*) AS diagnostic_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS diagnostic_percentage
FROM Table2
WHERE Itch = 'TRUE'
GROUP BY Diagnostic, diagnostic_type
ORDER BY diagnostic_percentage DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	diagnostic_type text	diagnostic_count bigint	diagnostic_percentage numeric
1	ACK	Cancerous	350	52.24
2	BCC	Cancerous	212	31.64
3	SEK	Non-Cancerous	43	6.42
4	SCC	Cancerous	43	6.42
5	NEV	Non-Cancerous	17	2.54
6	MEL	Cancerous	5	0.75

Query | Query History

```sql
--- Analysis of Itchy Lesion As An Indicator of Malignant Cancers

SELECT
    Diagnostic,
    CASE
        WHEN Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
        WHEN Diagnostic = 'NEV' THEN 'Non-Cancerous'
        ELSE 'Non-Cancerous'
    END AS diagnostic_type,
    COUNT(*) AS diagnostic_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS diagnostic_percentage
FROM Table2
WHERE Itch = 'FALSE'
GROUP BY Diagnostic, diagnostic_type
ORDER BY diagnostic_percentage DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	diagnostic_type text	diagnostic_count bigint	diagnostic_percentage numeric
1	NEV	Non-Cancerous	127	30.38
2	ACK	Cancerous	111	26.56
3	SEK	Non-Cancerous	94	22.49
4	BCC	Cancerous	61	14.59
5	SCC	Cancerous	13	3.11
6	MEL	Cancerous	12	2.87

Query | Query History

```sql
--- Analysis of Painful Lesions As An Indicator of Malignant Cancers

SELECT
    Diagnostic,
    CASE
        WHEN Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
        WHEN Diagnostic = 'NEV' THEN 'Non-Cancerous'
        ELSE 'Non-Cancerous'
    END AS diagnostic_type,
    COUNT(*) AS diagnostic_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS diagnostic_percentage
FROM Table2
WHERE Hurt = 'TRUE'
GROUP BY Diagnostic, diagnostic_type
ORDER BY diagnostic_percentage DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 5

	diagnostic character varying (255)	diagnostic_type text	diagnostic_count bigint	diagnostic_percentage numeric
1	BCC	Cancerous	109	72.19
2	SCC	Cancerous	19	12.58
3	ACK	Cancerous	18	11.92
4	NEV	Non-Cancerous	3	1.99
5	SEK	Non-Cancerous	2	1.32

Query | Query History

```sql
--- Analysis of Painful Lesions As An Indicator of Malignant Cancers

SELECT
    Diagnostic,
    CASE
        WHEN Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
        WHEN Diagnostic = 'NEV' THEN 'Non-Cancerous'
        ELSE 'Non-Cancerous'
    END AS diagnostic_type,
    COUNT(*) AS diagnostic_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS diagnostic_percentage
FROM Table2
WHERE Hurt = 'FALSE'
GROUP BY Diagnostic, diagnostic_type
ORDER BY diagnostic_percentage DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	diagnostic_type text	diagnostic_count bigint	diagnostic_percentage numeric
1	ACK	Cancerous	443	47.28
2	BCC	Cancerous	164	17.50
3	NEV	Non-Cancerous	141	15.05
4	SEK	Non-Cancerous	135	14.41
5	SCC	Cancerous	37	3.95
6	MEL	Cancerous	17	1.81

Query | Query History

```sql
--- Analysis of Growth of Lesions As An Indicator of Malignant Cancers

SELECT
    Diagnostic,
    CASE
        WHEN Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
        WHEN Diagnostic = 'NEV' THEN 'Non-Cancerous'
        ELSE 'Non-Cancerous'
    END AS diagnostic_type,
    COUNT(*) AS diagnostic_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS diagnostic_percentage
FROM Table2
WHERE Grew = 'TRUE'
GROUP BY Diagnostic, diagnostic_type
ORDER BY diagnostic_percentage DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	diagnostic_type text	diagnostic_count bigint	diagnostic_percentage numeric
1	BCC	Cancerous	206	40.39
2	NEV	Non-Cancerous	100	19.61
3	ACK	Cancerous	77	15.10
4	SEK	Non-Cancerous	74	14.51
5	SCC	Cancerous	37	7.25
6	MEL	Cancerous	16	3.14

Query | Query History

```sql
--- Analysis of Growth of Lesions As An Indicator of Malignant Cancers

SELECT
    Diagnostic,
    CASE
        WHEN Diagnostic IN ('MEL', 'SCC', 'ACK', 'BCC') THEN 'Cancerous'
        WHEN Diagnostic = 'NEV' THEN 'Non-Cancerous'
        ELSE 'Non-Cancerous'
    END AS diagnostic_type,
    COUNT(*) AS diagnostic_count,
    ROUND(
        100.0 * COUNT(*) / SUM(COUNT(*)) OVER (),
        2
    ) AS diagnostic_percentage
FROM Table2
WHERE Grew = 'FALSE'
GROUP BY Diagnostic, diagnostic_type
ORDER BY diagnostic_percentage DESC;
```

Data Output | Messages | Notifications

Showing rows: 1 to 6

	diagnostic character varying (255)	diagnostic_type text	diagnostic_count bigint	diagnostic_percentage numeric
1	ACK	Cancerous	384	66.44
2	BCC	Cancerous	67	11.59
3	SEK	Non-Cancerous	63	10.90
4	NEV	Non-Cancerous	44	7.61
5	SCC	Cancerous	19	3.29
6	MEL	Cancerous	1	0.17


Early-Stage
Detection and
AI/Machine
Learning- Based
Decision Support

Query | Query History

```sql

--- Early Detection AI Modelling of Skin Cancer

SELECT t1.age, t2.fitspatrick, t2.region, t1.skin_cancer_history, t1.cancer_history,t2.diagnostic,t2.changed, t2.grew
  (CASE WHEN t2.changed = TRUE THEN 1 ELSE 0 END +
   CASE WHEN t2.grew = TRUE THEN 1 ELSE 0 END +
   CASE WHEN t2.itch = TRUE THEN 1 ELSE 0 END +
   CASE WHEN t2.bleed = TRUE THEN 1 ELSE 0 END +
   CASE WHEN t2.hurt = TRUE THEN 1 ELSE 0 END) AS symptom_count,
  CASE
    WHEN t2.diameter_1 > 6 AND
         (CASE WHEN t2.changed = TRUE THEN 1 ELSE 0 END +
          CASE WHEN t2.grew = TRUE THEN 1 ELSE 0 END +
          CASE WHEN t2.itch = TRUE THEN 1 ELSE 0 END +
          CASE WHEN t2.bleed = TRUE THEN 1 ELSE 0 END +
          CASE WHEN t2.hurt = TRUE THEN 1 ELSE 0 END) >= 3 AND
         (t1.age > 40 OR t1.cancer_history = TRUE OR t1.skin_cancer_history = TRUE)
      THEN 'High Risk'
    WHEN t2.diameter_1 > 6 AND
         (CASE WHEN t2.changed = TRUE THEN 1 ELSE 0 END +
          CASE WHEN t2.grew = TRUE THEN 1 ELSE 0 END +
          CASE WHEN t2.itch = TRUE THEN 1 ELSE 0 END +
          CASE WHEN t2.bleed = TRUE THEN 1 ELSE 0 END +
          CASE WHEN t2.hurt = TRUE THEN 1 ELSE 0 END) BETWEEN 1 AND 2 AND
         (t1.age > 40 OR t1.cancer_history = TRUE)
      THEN 'Medium Risk'
    ELSE 'Low Risk'
  END AS risk_status
FROM table1 t1
JOIN table2 t2 ON t1.patient_id = t2.patient_id;
```

Data Output | Messages | Notifications

Showing rows: 1 to 1000 | Page No: 1 of 2

	age integer	fitspatrick integer	region character varying (255)	skin_cancer_history boolean	cancer_history boolean	diagnostic character varying (255)	changed boolean	grew boolean	itch boolean	bleed boolean	hurt boolean	biopsed boolean	symptom_count integer	risk_status text
1	14	1	FACE	false	true	NEV	false	false	false	false	false	true	0	Low Risk
2	65	0	EAR	false	false	SEK	false	false	false	false	false	false	0	Low Risk
3	63	0	FOREARM	false	false	ACK	false	false	false	false	false	false	0	Low Risk
4	64	0	FACE	false	false	SEK	false	true	false	false	false	false	1	Low Risk
5	59	2	NECK	false	false	BCC	false	true	true	true	true	true	4	High Risk
6	43	0	BACK	false	false	ACK	false	false	true	false	false	false	1	Low Risk
7	41	0	FOREARM	false	false	ACK	false	false	true	false	false	false	1	Low Risk
8	67	0	FACE	false	false	ACK	false	false	true	false	false	false	1	Low Risk
9	37	0	ARM	false	false	ACK	false	false	false	false	false	false	0	Low Risk
10	48	0	ARM	false	false	ACK	false	false	true	false	false	false	1	Low Risk
11	41	0	CHEST	false	false	ACK	false	false	false	false	false	false	0	Low Risk
12	37	0	NOSE	false	false	ACK	false	false	true	false	false	false	1	Low Risk
13	39	0	FACE	false	false	ACK	false	false	true	false	false	false	1	Low Risk
14	52	0	FACE	false	false	ACK	false	false	true	false	false	false	1	Low Risk
15	38	0	FACE	false	false	ACK	false	false	false	false	false	false	0	Low Risk
16	47	0	NOSE	false	false	ACK	false	false	true	false	false	false	1	Low Risk
17	35	0	FOREARM	false	false	ACK	false	false	true	false	false	false	1	Low Risk
18	41	0	FACE	false	false	ACK	false	false	true	false	false	false	1	Low Risk
19	55	0	ARM	false	false	ACK	false	false	true	false	false	false	1	Low Risk
20	47	0	FACE	false	false	ACK	false	false	false	false	false	false	0	Low Risk
21	45	0	FOREARM	false	false	ACK	false	false	false	false	false	false	0	Low Risk
22	79	2	ARM	true	true	ACK	false	false	true	false	false	true	1	Medium Risk
23	40	0	HAND	false	false	ACK	false	false	false	false	false	false	0	Low Risk
24	36	0	FOREARM	false	false	SEK	false	false	false	false	false	false	0	Low Risk

Total rows: 1088 | Query complete 00:00:00.799 | CRLF | Ln 662, Col 1

Summary of Findings

Lesions with three or more symptoms are strongly linked to malignancy, making them critical for early detection and AI risk scoring.

Pesticide exposure and poor sanitation significantly increase the likelihood of biopsy-confirmed and malignant skin lesions.

Males, patients over 40, and those with Fitzpatrick skin types 3–4 are more likely to develop severe or malignant lesions.

A personal or family history of cancer is a strong risk factor for melanoma, especially when combined with age and visible symptoms.

Lesions on the head and neck, especially when painful, are more frequently confirmed as malignant through biopsy.

Lesions that change in colour or size account for a high proportion of melanoma cases and should be prioritised in screening.

Recommendations

Integrate Red-Flag Symptom Rules into Clinical Triage Tools
Use combinations like lesion size > 6mm with ≥3 symptoms (bleed, hurt, change, etc.) to flag high-risk cases for immediate review.

Prioritise Surveillance for High-Risk Groups
Target patients with skin cancer history, age > 40, and confirmed family history for regular dermatological checks.

Customise AI Models with Demographic & Environmental Data
Train ML models using variables like Fitzpatrick skin type, gender, pesticide exposure, and lack of sewage/piped water access, all shown to correlate with cancer severity.

Improve Diagnostic Awareness in Underserved Areas
Findings show higher percentages of malignant lesions in patients without sewage systems. Public health campaigns should prioritise these regions.

Use Symptom Clustering for Pre-Consult Screening
Lesions exhibiting multiple symptoms (pain, growth, colour change, elevation, bleeding) should automatically escalate during triage or telederm consultations.

Recommend Clinical Validation of AI Risk Labels
Labels such as “High Risk,” “Medium Risk,” and “Low Risk” based on SQL logic should be evaluated by dermatologists for clinical adoption.