



ANALYSIS OF DERM AI DIAGNOSTICS

SKIN CANCER DATASET

ABEF Data Hub

BUSINESS OVERVIEW – DERMAI DIAGNOSTICS

DermAI Diagnostics is a pioneering health-tech company committed to transforming early skin cancer detection. By integrating **machine learning** with **clinical dermatology research**, DermAI aims to provide data-driven insights that empower medical professionals in diagnosing and treating skin lesions more accurately and efficiently.

PROJECT RATIONALE



**Bridging Data
and Medicine**



**Early
Detection &
Prevention**



**AI-Driven
Medical
Research**



**Real-World
Application**



**SQL Learning
Opportunity**

THE PROBLEM WE ARE SOLVING



Late Detection of Skin Cancer

Skin cancer is often diagnosed at a late stage, reducing treatment effectiveness.

Early detection drastically increases survival rates but remains challenging.



Misdiagnosis & Environmental Factors Overlooked

Limited access to dermatologists leads to high rates of misdiagnosis.

Key risk factors like sun exposure, pesticide contact, and family history are underutilized in diagnosis.



Lack of Structured, Queryable Datasets

Medical data is often unstructured, making analysis difficult.

There's a critical need for datasets that support SQL-based queries to uncover hidden patterns and support clinical decision-making.

PROJECT OBJECTIVES & METHODOLOGY



Clear Objectives

- Join clinical + lesion data via SQL
- Identify risk factors using queries
- Analyze patterns in lesion characteristics
- Structure data for machine learning



Methodology

- Import Dataset
- Join Tables
- Write Queries
- Analyze Results
- Provide Recommendations

THE DATA DESCRIPTION MARKET

Dataset Overview – Risks Categorization and Lesion Categorization

DATA DICTIONARY - DATA DESCRIPTION

Column Name	Risk Categorization	Column Name	Lesion Categorization
Patient_id	Unique identifier for each patient	NEV	Benign
Smoke	Patient smokes (TRUE/FALSE)	MEL	Malignant
Drink	Patient drinks alcohol (TRUE/FALSE)	BCC	Malignant
Background_father	Patient's paternal ethnicity	SEK	Benign
Background_mother	Patient's maternal ethnicity	ACK	Malignant
Age	Age of patient	SCC	Malignant
Pesticide	Exposure to pesticides (TRUE/FALSE)		
Gender	Gender (MALE/FEMALE)		
Skin_cancer_history	Previous skin cancer diagnosis (TRUE/FALSE)		
Cancer_history	Family history of cancer (TRUE/FALSE)		
Has_piped_water	Access to piped water (TRUE/FALSE)		
Has_sewage_system	Access to sewage system (TRUE/FALSE)		











SQL QUERIES & INSIGHTS





- Find correlation between variables and lesion type
- Analyze lesion characteristics (e.g., itching, growth) with diagnoses
- Environmental risk analysis (pesticide, smoke)







HOW DOES THE DEMOGRAPHIC FACTORS AFFECT THE LESION TYPE ?

	drink boolean 	diagnostic_type text 	total_count bigint 	percentage numeric 
1	false	Benign	274	29
2	false	Malignant	676	71
3	true	Malignant	131	95
4	true	Benign	7	5

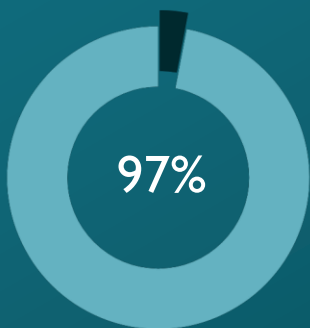
	smoke boolean 	diagnostic_type text 	total_count bigint 	percentage numeric 
1	false	Benign	279	27
2	false	Malignant	747	73
3	true	Malignant	60	97
4	true	Benign	2	3

	pesticide boolean 	diagnostic_type text 	total_count bigint 	percentage numeric 
1	false	Benign	276	32
2	false	Malignant	589	68
3	true	Malignant	218	98
4	true	Benign	5	2

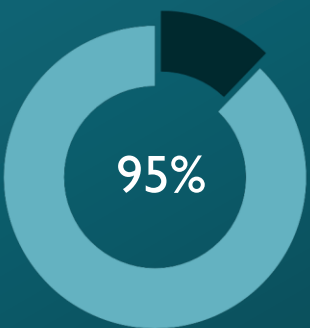
	gender character varying (10) 	diagnostic_type text 	total_count bigint 	percentage numeric 
1	FEMALE	Benign	70	19
2	FEMALE	Malignant	292	81
3	MALE	Malignant	515	71
4	MALE	Benign	211	29

EXAMINING LESION TRAITS TO IDENTIFY UNDERLYING PATTERNS

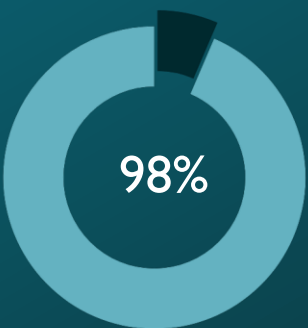
MULTIPLE VARIABLE SEGMENTATION



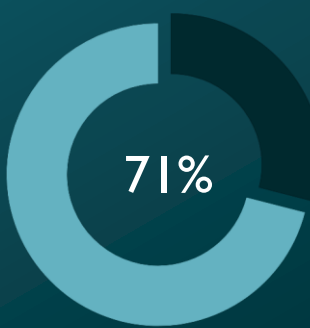
SMOKE



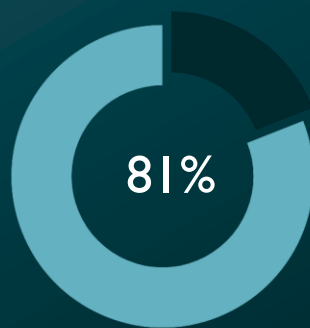
DRINK



PESTICIDE



MALE



FEMALE

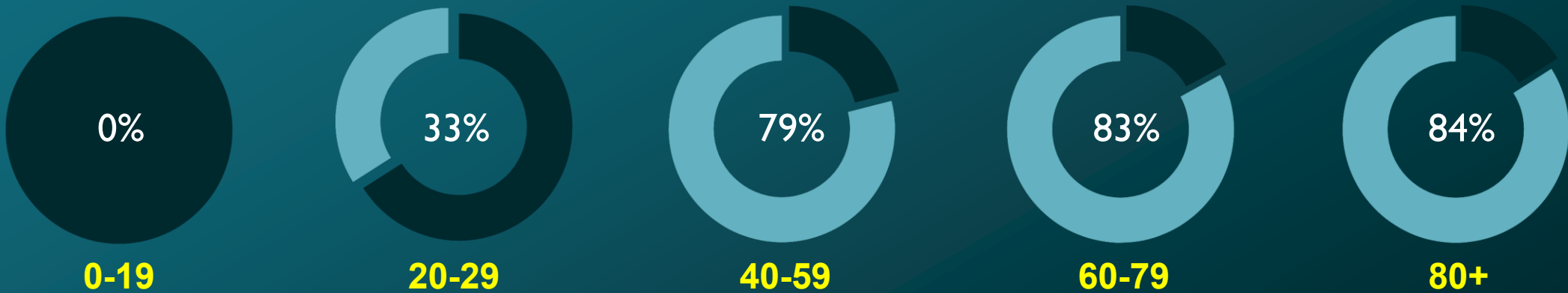
VARIABLES	MALIGNANT	BENIGN	% MALIGNANT	REMARKS
Smoke	60	2	97	High Risk Factor
Drink	131	7	95	High Risk Factor
Pesticide	218	5	98	High Risk Factor
Male Gender	515	211	71	High Risk Factor
Female Gender	292	70	81	High Risk Factor

HOW DOES AGE DISTRIBUTION AFFECTS LESION TYPES ?

	age_group text	diagnostic_type text	total_count bigint	percentage numeric
1	0-19	Benign	20	100
2	20-39	Benign	85	67
3	20-39	Malignant	41	33
4	40-59	Benign	82	21
5	40-59	Malignant	306	79
6	60-79	Benign	79	17
7	60-79	Malignant	381	83
8	80+	Benign	15	16
9	80+	Malignant	79	84

EXAMINING LESION TRAITS TO IDENTIFY UNDERLYING PATTERNS

AGE SEGMENTATION

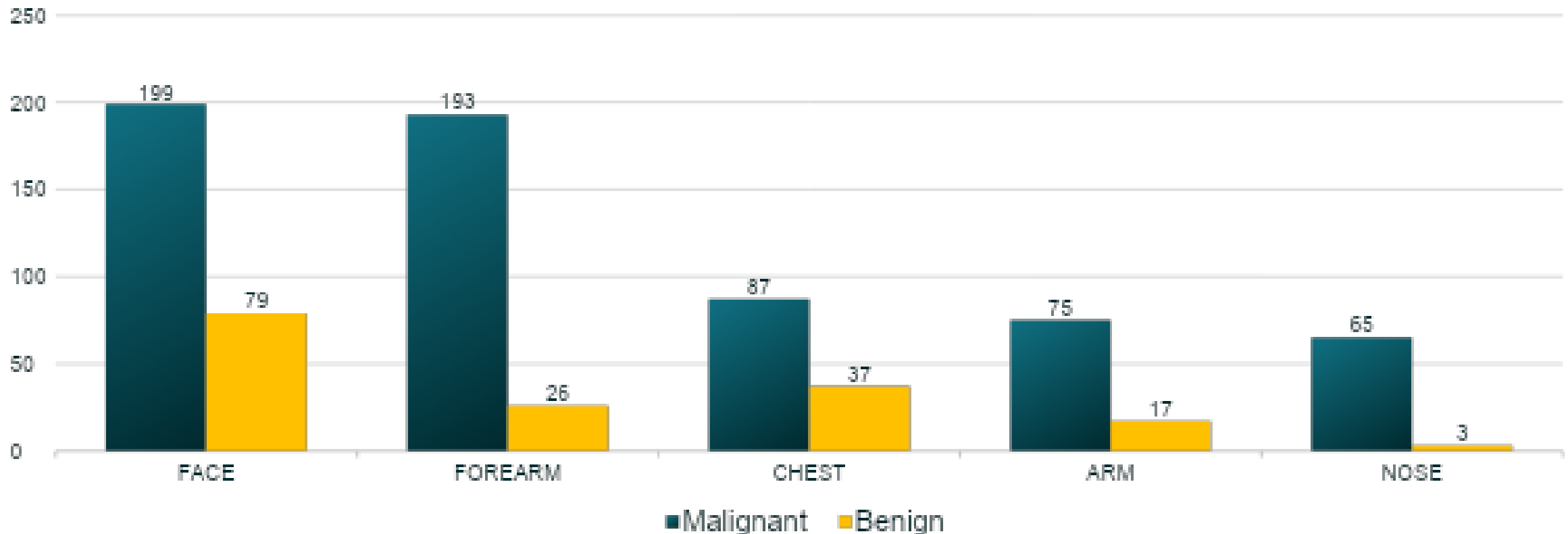


VARIABLES	MALIGNANT	BENIGN	% MALIGNANT	REMARKS
0-19	0	20	0	Low Risk Factor
20-39	41	85	33	Low Risk Factor
40-59	306	82	79	High Risk Factor
60-79	381	79	83	High Risk Factor
80+	79	15	84	High Risk Factor





EXAMINING LESION TRAITS TO IDENTIFY UNDERLYING PATTERNS





REGION SEGMENTATION

The top 5 body regions were analyzed and the table reveals regions that have an higher risk factor of getting a skin legion



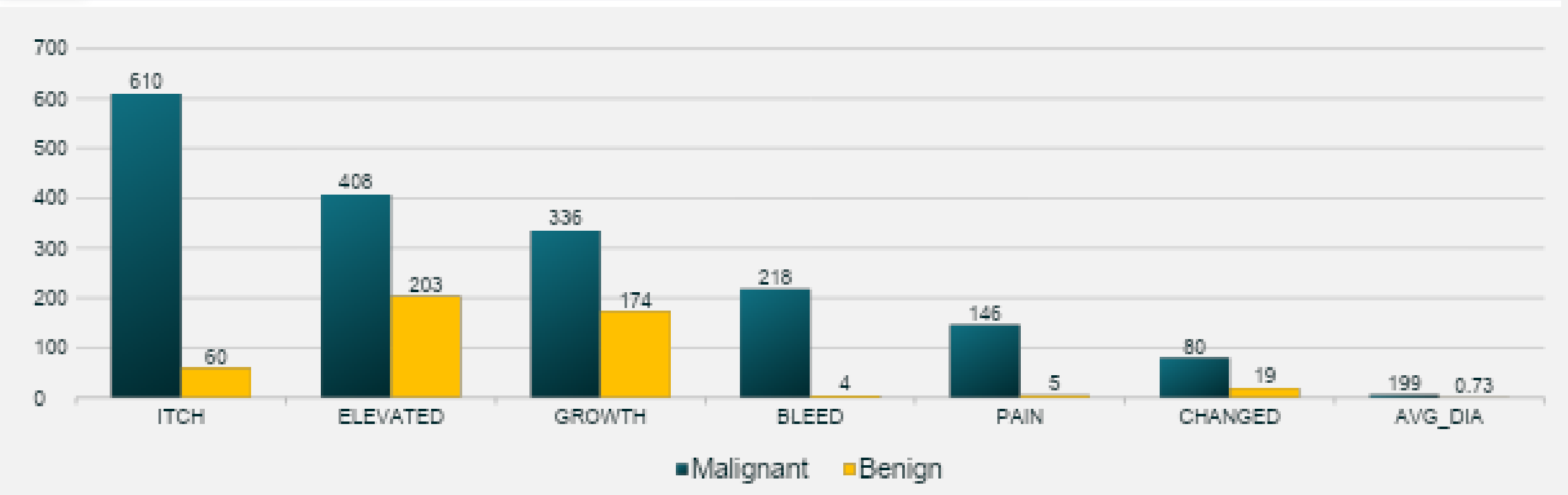
HOW DOES CANCER HISTORY AFFECT LESION DIAGNOSTIC TYPE?

	skin_cancer_history  boolean	diagnostic_type  text	total_count  bigint	percentage  numeric
1	false	Benign	268	31.02
2	false	Malignant	596	68.98
3	true	Malignant	211	94.20
4	true	Benign	13	5.80

	cancer_history  boolean	diagnostic_type  text	total_count  bigint	percentage  numeric
1	false	Benign	263	32.23
2	false	Malignant	553	67.77
3	true	Malignant	254	93.38
4	true	Benign	18	6.62

DISTRIBUTION OF THE LESION CHARACTERISTICS

	diagnostic_type text	total_lesions bigint	Itch bigint	Growth bigint	Pain bigint	Changed bigint	Bleed bigint	Elevated bigint	avg_diameter numeric
1	Malignant	807	610	336	146	80	218	408	5.89
2	Benign	281	60	174	5	19	4	203	0.73



INFLUENCE OF FITZPATRICK SCALE ON CANCER TYPE

	fitzpatrick integer	skin_type text	skin_description text	reaction_to_sun text
1	0	NULL	NULL	NULL
2	1	I	very fair/ pale white skin	Always burn, never tans
3	2	II	fair skin	usually burns, tans poorly
4	3	III	medium white to olive skin	sometimes burns, gradually tans
5	4	IV	olive or light brown skin	rarely burns, tans well
6	5	V	Brown skin	very rarely burns, tans very easily
7	6	VI	dark brown or black skin	never burns, deeply pigmented

	fitzpatrick integer	count bigint	malignant_count bigint	benign_count bigint	malignant_percentage numeric
1	0	579	328	251	56.6
2	1	46	43	3	93.5
3	2	285	273	12	95.8
4	3	142	135	7	95.1
5	4	28	22	6	78.6
6	5	7	6	1	85.7
7	6	1	0	1	0.0

SYMPTOM SCORE TABLE

Using the six lesion symptoms from the dataset, a table of symptom score was created to help patients and dermatologists.

	symptom_score integer	total_cases bigint	malignant_count bigint	benign_count bigint	malignant_percentage numeric
1	6	7	7	0	100.00
2	5	77	75	2	97.40
3	4	114	110	4	96.49
4	3	150	116	34	77.33
5	2	294	168	126	57.14
6	1	342	257	85	75.15
7	0	104	74	30	71.15

TOP 10 PATIENT PROFILES

```
714
715 -- Identifying the top 10 patient profiles (combinations of risk factors) most likely
716 -- to have malignant lesions?
717 -- Using:
718 -- Smoking/drinking
719 -- Skin cancer history
720 -- Cancer history
721 -- Lesion behaviors
722
```

Data Output Messages Notifications

≡

📄

▼

📋

▼

🗑

🗄

⬇

📈

SQL

Showing rows: 1 to 10

Page No: 1 of 1

⏪ ⏩ ⏴ ⏵

	smoke boolean	drink boolean	skin_cancer_history boolean	cancer_history boolean	grew boolean	changed boolean	itch boolean	hurt boolean	bleed boolean	elevation boolean	total_patients bigint	malignant_count bigint	benign_count bigint	malignancy_percentage numeric
1	false	false	true	false	true	false	true	true	true	true	8	8	0	100.00
2	false	false	false	false	true	false	true	true	true	true	8	8	0	100.00
3	false	false	true	true	true	false	true	false	false	true	6	6	0	100.00
4	false	false	false	true	true	false	true	false	false	true	5	5	0	100.00
5	false	false	true	true	true	false	true	true	true	true	6	6	0	100.00
6	false	false	false	false	false	false	true	true	false	true	5	5	0	100.00
7	false	false	false	true	false	false	true	false	false	true	8	8	0	100.00
8	false	false	true	true	false	false	true	true	true	true	5	5	0	100.00
9	false	false	true	true	false	false	true	false	false	false	7	7	0	100.00
10	false	false	true	true	false	false	false	false	false	false	8	8	0	100.00

ADVANCING SKIN HEALTH RESEARCH WITH ORGANIZED CLINICAL DATASETS

Key Insights & Recommendations

1. **Age & Malignancy:** Older patients have higher malignancy rates. Early screening should focus on older age groups, while younger people with symptoms shouldn't be overlooked.
2. **Smoking & Drinking:** Though few, **95–100%** of drinkers/smokers had malignant lesions—showing strong risk ties.
3. **Lesion Size:** Treat larger lesions as potential malignancy indicators and prioritize their examination.
4. **Cancer History:** Include both personal and family cancer histories in patient evaluations to predict malignancy risk better.
5. **Utility Access:** Limited access to piped water/sewage may delay diagnosis or limit preventive care, affecting lesion detection.

Age, habits, lesion traits, cancer history, and living conditions all impact skin cancer risk. These factors should guide targeted screening, early detection, and resource prioritization for better outcomes.

**THANK
YOU**

