

A medical professional, likely a dermatologist or researcher, is shown in a clinical or laboratory setting. They are wearing light blue scrubs and a name tag that reads "AMERICAN UNIVERSITY CASPER". They are pointing their right index finger at a large digital screen that displays a complex, blue-toned image of skin tissue, possibly a histological section or a 3D model of a skin lesion. The background is slightly blurred, showing what appears to be a window with blinds and some office furniture. The overall lighting is cool and professional.

ANALYSIS OF DERM AI DIAGNOSTICS --- SKIN CANCER DATASET

ABEF Data Hub

SQL CAPSTONE PROJECT

Analysis Of DermAi Diagnostics Skin Cancer Dataset

GROUP MEMBERS

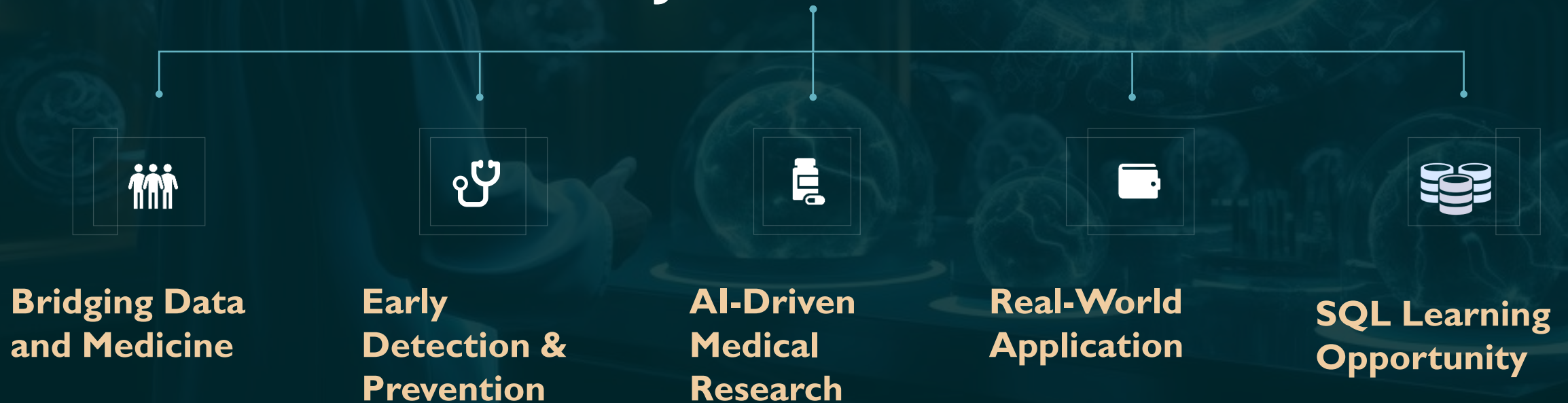
Abimbola Orekoya
Fanta Conta
Elo Merit
Babajide Adesanya



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



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

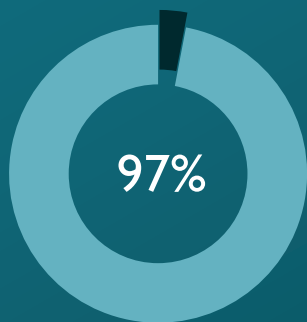
	smoke boolean	diagnostic_type text	total_count bigint	percentage numeric
1	false	Benign	279	27
2	false	Malignant	747	73
3	true	Malignant	60	97
4	true	Benign	2	3

	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

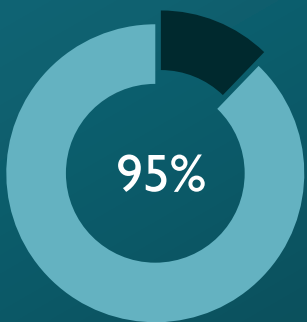
	gender character varying (10)	diagnostic_type text	total_count bigint	percentage numeric
1	FEMALE	Benign	70	19
2	FEMALE	Malignant	292	81
3	MALE	Malignant	515	71
4	MALE	Benign	211	29

EXAMINING LESION TRAITS TO IDENTIFY UNDERLYING PATTERNS

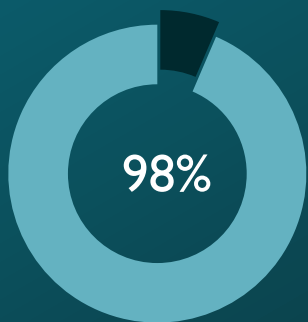
MULTIPLE VARIABLE SEGMENTATION



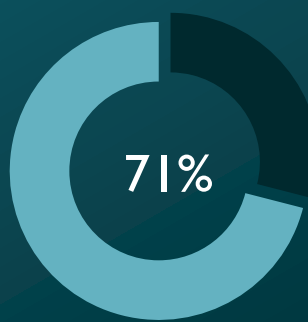
SMOKE



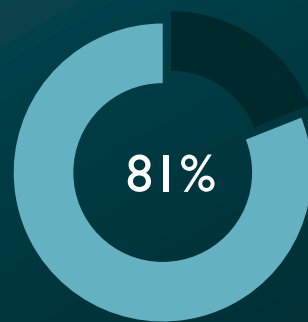
DRINK



PESTICIDE



MALE



FEMALE

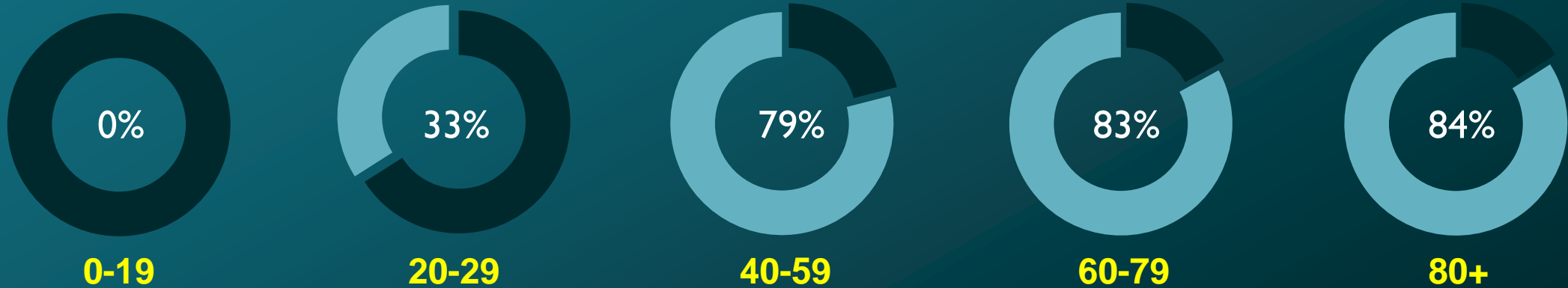
VARIABLES	MALIGNANT	BENIGN	% MALIGNANT	REMARKS
Smoke	60	2	97	High Risk Factor
Drink	131	7	95	High Risk Factor
Pesticide	218	5	98	High Risk Factor
Male Gender	515	211	71	High Risk Factor
Female Gender	292	70	81	High Risk Factor

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

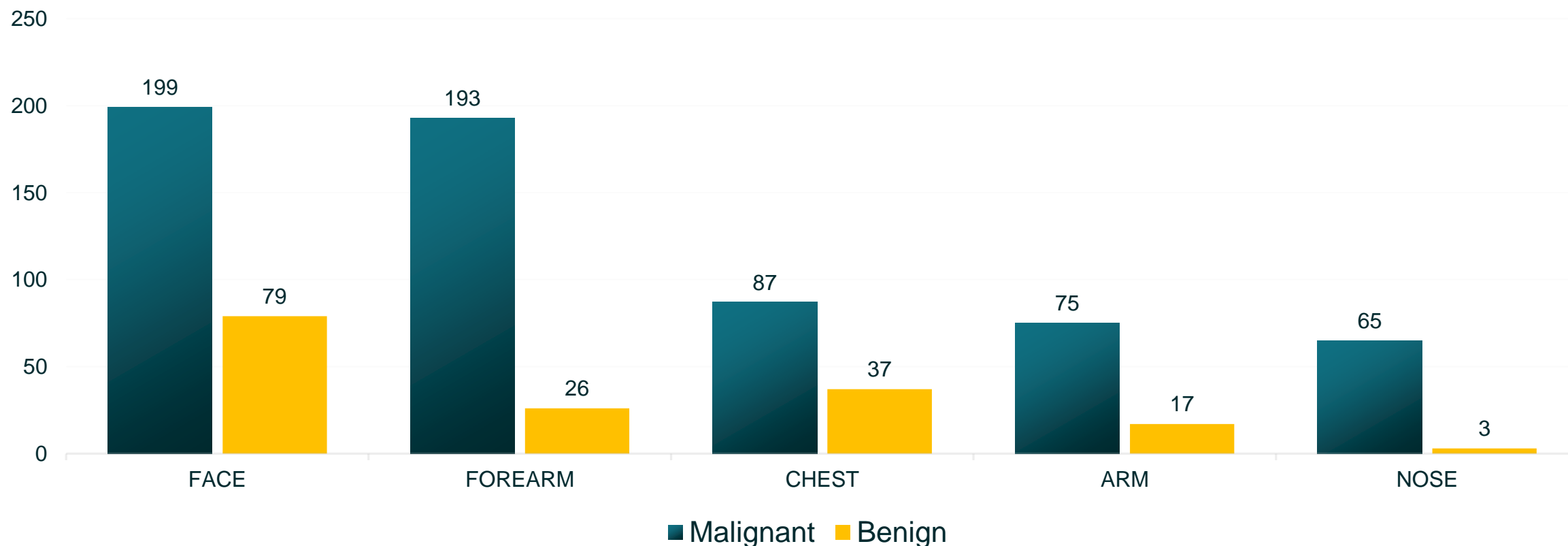
HOW DOES REGION AFFECT THE LESION DIAGNOSTIC TYPE ?

	region character varying (255) 	total_cases bigint 	benign_count bigint 	malignant_count bigint 
	FACE	278	79	199
	FOREARM	219	26	193
	CHEST	124	37	87
	ARM	92	17	75
	NOSE	68	3	65
	HAND	55	3	52
	BACK	105	59	46
	NECK	37	6	31
	EAR	32	9	23
0	THIGH	35	14	21
1	LIP	7	0	7
2	ABDOMEN	20	16	4
3	FOOT	5	3	2
4	SCALP	11	0	0





EXAMINING LESION TRAITS TO IDENTIFY UNDERLYING PATTERNS





REGION SEGMENTATION

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



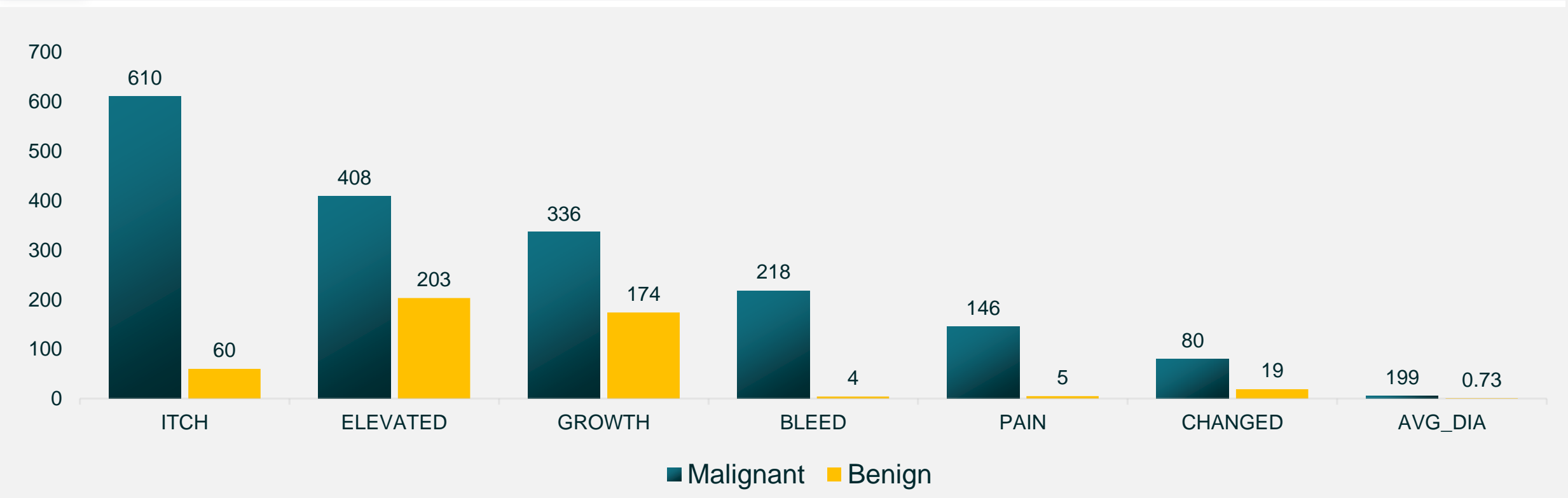
HOW DOES CANCER HISTORY AFFECT LESION DIAGNOSTIC TYPE?

	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



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**

