# Case Study: Analysis of DermAI Diagnostics Skin Cancer Dataset

## Business Introduction

Skin cancer remains one of the most common and life-threatening diseases worldwide. However, **early detection** significantly improves survival rates.
**DermAI Diagnostics** is a health-tech company leveraging **data analysis** and **clinical dermatology research** to enhance early diagnosis and treatment outcomes.

By analyzing **patient demographics**, **environmental factors**, and **lesion characteristics**, DermAI aims to generate **data-driven insights** that support dermatologists in early-stage detection and informed decision-making.

Through structured datasets and SQL-based research, the company bridges the gap between **medical practitioners** and **data analysts**, supporting:

- Medical research
- Epidemiological studies
- Evidence-based diagnostic decision-making

Ultimately, DermAI's goal is to **improve public health outcomes** through data-informed prevention and early detection.

## ❗ Problem Statement

Skin cancer detection is often **delayed** due to:

- Misdiagnosis or lack of clinical expertise
- Limited access to dermatologists
- Poor understanding of environmental and genetic risk factors

With a dataset containing **1,089 instances of skin lesions**, this project explores how **demographic**, **environmental**, and **clinical** attributes interact and contribute to different skin cancer types.

The analysis seeks to uncover key patterns that can assist in **identifying risk factors** and improving awareness of conditions that lead to skin cancer.

## Data Description

### Table 1: `Patient_Info`

| Column Name | Description |
|---|---|
| patient_id | Unique identifier for each patient |
| smoke | Patient smokes (TRUE/FALSE) |
| drink | Patient drinks alcohol (TRUE/FALSE) |
| background_father | Patient's paternal ethnicity |
| background_mother | Patient's maternal ethnicity |
| age | Age of patient |
| 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) |

### Table 2: `Lesion_Info`

| Column Name | Description |
|---|---|
| lesion_id | Unique identifier for each lesion |
| patient_id | Foreign key linking to `Patient_Info` |
| fitspatrick | Fitzpatrick skin type (1–6) |
| region | Body region of the lesion |
| diameter_1 | First diameter of lesion (mm) |
| diameter_2 | Second diameter measurement (mm) |
| diagnostic | Type of skin lesion (e.g., BCC, MEL, NEV, etc.) |
| itch | Lesion causes itching (TRUE/FALSE) |

| | |
|---|---|
| grew | Lesion has grown (TRUE/FALSE) |
| hurt | Lesion causes pain (TRUE/FALSE) |
| changed | Lesion changed in color/size (TRUE/FALSE) |
| bleed | Lesion bleeds (TRUE/FALSE) |
| elevation | Lesion is raised (TRUE/FALSE) |
| img_id | Associated lesion image filename |
| biopsed | Lesion biopsy-confirmed (TRUE/FALSE) |

## 🎯 Project Aim

- Develop an SQL database for analyzing clinical and lesion data.
- Identify **environmental** and **demographic** risk factors related to specific lesion types.
- Examine lesion characteristics to distinguish **cancerous vs. non-cancerous** lesions.
- Support dermatological and epidemiological research through structured, queryable data.