

# Comprehensive EDA and Inferential Analysis on Large-Scale Patient Dataset

## Project Vision

This project harnesses the power of advanced analytics and global health data to uncover actionable insights into cancer care, outcomes, and disparities. Using a robust dataset of 50,000 cancer patient records collected from multiple countries between 2015 and 2024, we aim to bridge the gap between raw data and life-saving decisions.

## What the Data Covers

- Demographics: Age, gender, country, year of diagnosis
- Genetic & Lifestyle Risks: Genetic predisposition, smoking, alcohol use, obesity
- Environmental Exposure: Air pollution
- Clinical & Economic Variables: Cancer type, stage, treatment cost
- Patient Outcomes: Survival years, severity scores

## Core Objectives

### 1. Exploratory Data Analysis (EDA)

- Identify key trends, hidden patterns, and relationships
- Visualize disparities in diagnosis, lifestyle, treatment, and outcomes
- Highlight variations across countries, age groups, and cancer stages

### 2. Inferential & Predictive Analytics

- Determine the relationship between risk factors and cancer severity
- Analyze the proportion of early-stage diagnoses by cancer type
- Identify key predictors of cancer severity and survival years
- Explore the economic burden of cancer treatment across different demographics and countries
- Assess whether higher treatment cost is associated with longer survival
- Evaluate if higher cancer stages lead to greater treatment costs and reduced survival years
- Examine whether higher genetic risk amplifies the negative effects of smoking on cancer severity and survival outcomes

### 3. Insight Extraction and Documentation

- For each major analysis, draw clear, evidence-backed inferences
- Systematically document all findings for clinical, academic, and operational use
- Link each insight to real-world implications to guide decision-making and policy formulation