

## Project Vision – EDA on Cancer Dataset

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

Our dataset offers a **360-degree view of cancer patient profiles**, including:

- **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

Using statistical methods, we'll explore and answer critical healthcare questions:

- 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, we will draw clear, evidence-backed inferences
- All findings will be systematically documented for future clinical, academic, and operational use
- Each insight will be linked to real-world implications—to guide decision-making and policy formulation