**CANCER DATASET ANALYSIS**

Description:  
This dataset is a large-scale, structured collection of clinical and demographic records for **890,000 cancer patients** from different countries, designed to support research in oncology, epidemiology, predictive modeling, and health policy evaluation. This dataset captures a wide range of variables related to cancer diagnosis, treatment, comorbidities, and survival outcomes.

Each entry represents an individual cancer patient and includes the following information:

* It is having factors like age, gender, and country of residence.
* **Clinical details:** Cancer stage at diagnosis, family history of cancer, smoking status, body mass index (BMI), and cholesterol level.
* **Comorbidities**: Binary indicators for hypertension, asthma, cirrhosis, and presence of other cancers.
* **Treatments**: Type of treatment received (e.g., chemotherapy, surgery, radiation, or combined therapies), date of diagnosis, and date of treatment completion.
* **Survival Outcome**: Whether the patient survived post-treatment in the form of binary indicators.

Objectives:

**1. Patient Demographics & Geographic Analysis**

* What is the age distribution of patients across different cancer stages?
* How many male and female patients are there, and what is the average age for each gender?
* Are there significant gender-based differences in survival rates, treatment types, or cancer stage at diagnosis?
* Which countries have the most patients and show the best survival rates?
* Does age at diagnosis vary across different countries or stages, and how does this affect treatment decisions?

**2. Cancer Stage & Survival Analysis**

* What are the survival rates across different cancer stages (Stage I to IV)?
* Which cancer stage has the highest death rate?
* How does survival probability change across different age groups?
* What is the impact of early-stage vs late-stage diagnosis on survival outcomes?
* How is cancer stage distributed among patients at the time of diagnosis?

**3. Treatment Effectiveness & Patterns**

* Which treatment types (Chemotherapy, Surgery, Radiation, Combined) yield the highest survival rates for each cancer stage?
* What treatment method is most commonly used, and does it vary by cancer stage?
* What is the survival rate for patients receiving combination treatment vs. single treatment methods?
* What is the average treatment duration from diagnosis to completion, and how does this impact survival?
* Are treatment durations shorter in patient profiles or countries with higher survival rates?

**4. Lifestyle Factors & Risk Assessment**

* How does smoking status correlate with survival rates and cancer stage severity?
* How do different BMI levels affect survival rates and post-treatment outcomes?

**5. Comorbidities & Health Conditions**

* How do comorbidities like asthma, hypertension, or liver cirrhosis affect cancer stage and survival rates?
* What are the survival rates for patients with multiple comorbidities compared to those with single conditions?

**6. Family History & Genetic Factors**

* Does family history of cancer lead to diagnosis at earlier ages, and what are their survival rates?
* Do patients with family history tend to be diagnosed at earlier stages compared to those without?

**7. Multiple Cancer Cases**

* How does the presence of additional cancers impact treatment effectiveness and survival probability?
* Does having multiple cancers significantly decrease overall survival rates?

**8. Time-Based Analysis & Trends**

* How many patients have both diagnosis and end treatment dates, and what is their recovery time span?
* How does the time from diagnosis to treatment completion affect survival outcomes?

**9. Treatment Completion Analysis**

* What is the relationship between cancer stage and treatment completion rates?
* Which patients successfully complete treatment, and what factors contribute to treatment completion?