**📊 1. Análises descritivas (Entendimento geral dos dados)**

* **Distribuição demográfica**: Qual a distribuição por gênero, idade, país e etnia?
* **Fatores de risco**: Quais os fatores de risco mais comuns entre os pacientes (histórico familiar, exposição à radiação, etc.)?
* **Distribuição dos níveis hormonais**: Como estão distribuídos os níveis de TSH, T3 e T4? Há valores extremos?
* **Classificação de risco**: Qual a proporção de pacientes classificados como *baixo*, *médio* e *alto risco*?
* **Diagnósticos**: Quantos casos foram diagnosticados como *benignos* vs. *malignos*?

**📈 2. Análises exploratórias (Procurando padrões e correlações)**

* **Idade e risco**: Existe correlação entre a idade do paciente e o risco de câncer?
* **Fatores combinados**: Pacientes com múltiplos fatores de risco (ex.: histórico familiar + radiação) têm maior chance de um diagnóstico maligno?
* **Nódulo x diagnóstico**: Existe uma relação clara entre o tamanho dos nódulos e o risco ou diagnóstico final?
* **Hormônios e risco**: Como os níveis hormonais (TSH, T3, T4) se comportam entre os diferentes níveis de risco?
* **Distribuição geográfica**: Há países ou regiões com maior incidência de casos graves?

**🔮 3. Análises preditivas (Projeções e modelagem)**

* **Modelos preditivos**: Podemos criar um modelo de classificação para prever o risco de câncer com base nos fatores disponíveis?
* **Importância das variáveis**: Quais fatores têm maior peso ao determinar o risco de câncer?
* **Análise de cluster**: É possível agrupar pacientes com perfis semelhantes (idade, hormônios, histórico) para identificar padrões ocultos?
* **Probabilidade de malignidade**: Dado um conjunto de fatores, qual a probabilidade de um diagnóstico maligno?

📊 1. Descriptive analyses (General understanding of the data)

• Demographic distribution: What is the distribution by gender, age, country and ethnicity?

• Risk factors: What are the most common risk factors among patients (family history, radiation exposure, etc.)?

• Distribution of hormone levels: How are TSH, T3 and T4 levels distributed? Are there extreme values?

• Risk classification: What proportion of patients are classified as low, medium and high risk?

• Diagnoses: How many cases were diagnosed as benign vs. malignant?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

📈 2. Exploratory analyses (Looking for patterns and correlations)

• Age and risk: Is there a correlation between the patient's age and the risk of cancer?

• Combined factors: Do patients with multiple risk factors (e.g. family history + radiation) have a higher chance of a malignant diagnosis? • Nodule x diagnosis: Is there a clear relationship between nodule size and risk or final diagnosis?

• Hormones and risk: How do hormone levels (TSH, T3, T4) behave between different risk levels?

• Geographic distribution: Are there countries or regions with a higher incidence of severe cases?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

🔮 3. Predictive analysis (Projections and modeling)

• Predictive models: Can we create a classification model to predict cancer risk based on the available factors?

• Importance of variables: Which factors have the greatest weight when determining cancer risk?

• Cluster analysis: Is it possible to group patients with similar profiles (age, hormones, history) to identify hidden patterns?

• Probability of malignancy: Given a set of factors, what is the probability of a malignant diagnosis?