Capstone Project: Data Analysis

Project Overview:

You will act as a data analyst for a fictitious organization called **HealthConnect**, a company providing innovative healthcare solutions. The project is designed to analyze patient data, discover trends, and build insightful dashboards to aid management in decision-making. You will apply the full data analysis pipeline: cleaning, processing, visualizing, and presenting the data.

Project Steps:

1. **Problem Definition:**
   - **Objective**: Analyze the organization's patient database to discover trends and insights. The company is interested in identifying common health issues, treatment outcomes, and geographical distribution of diseases to better tailor its services.
2. **Dataset**:
   - You will be provided with a **CSV file** containing anonymized patient data, including variables such as age, gender, region, symptoms, diagnosis, treatment plan, and outcomes. **Note:** Feel free to conduct your own research and find a dataset of your choice.
3. **Data Cleaning:**
   - Clean and preprocess the dataset. Handle missing values, outliers, and erroneous data. Document your cleaning steps and justify any decisions made (e.g., removing outliers, imputing missing values).
4. **Data Analysis:**
   - **Explore the Data**: Use descriptive statistics to provide an overview of the data.
   - **Key Questions**:
     - What are the most common diagnoses based on region?
     - What are the success rates of the various treatment plans?
     - What are the demographics (age, gender) of patients with specific health conditions?
5. **Data Visualization (Using Tableau):**
   - Create **interactive dashboards** that visualize the following:
     - Patient demographics distribution across different health conditions.
     - Treatment success rates by diagnosis.
     - Geographical distribution of health conditions.
   - Include at least 3 different types of visualizations: bar charts, heatmaps, and pie charts.
6. **Advanced Analytics:**

- Apply **predictive analytics** to predict patient outcomes based on factors such as age, treatment type, and diagnosis. Use **regression analysis** or **decision trees** to model the outcomes.
7. **Final Report and Presentation:**
   - Prepare a **comprehensive report** summarizing your findings.
   - Create a **5-minute presentation** with key visuals, explaining how the analysis can be used for decision-making.

Submission Requirements:

- Submit your data analysis Tableau dashboards (for the Data Analysis Capstone).