Prepared For: Health Data Analysis
Subject: Project Charter & Requirements for Health Dataset Analysis using SQL Server Management Studio (SSMS)

---

1. Project Overview & Objective

The client possesses a comprehensive healthcare dataset (`Health.csv`) containing records of patient admissions, treatments, billing, and outcomes. The primary objective is to leverage SQL Server Management Studio to import, clean, analyze, and derive actionable insights from this data. The goal is to answer key business questions related to patient demographics, hospital performance, financials, and treatment efficacy to support strategic decision-making.

2. Source Data Description

- **Patient Details:** `Name`, `Age`, `Gender`
- **Medical Information:** `Blood Type`, `Medical Condition`, `Medication`, `Test Results`
- **Admission Details:** `Date of Admission`, `Admission Type`, `Discharge Date`, `Room Number`
- **Hospital & Staff:** `Hospital`, `Doctor`
- **Billing & Insurance:** `Insurance Provider`, `Billing Amount`

3. Key Business Questions & Client Requirements

The analysis will be structured to answer the following critical business questions:

A. Patient Demographics & Epidemiology:
1. What is the distribution of patients by `Gender` and `Age` groups (e.g., 0-18, 19-35, 36-50, 51-65, 65+)?
2. What are the top 5 most common `Medical Condition`s among admitted patients?
3. Is there a correlation between `Blood Type` and specific `Medical Condition`s?

B. Hospital & Operational Analysis:
1. Which `Hospital` has the highest number of admissions? Which has the highest average `Billing Amount`?
2. Who are the top 5 `Doctor`s by number of patients treated?
3. What is the average length of stay (`Discharge Date` - `Date of Admission`) for each `Admission Type` (Urgent, Emergency, Elective)?

C. Financial & Insurance Analysis:
1. What is the total and average `Billing Amount` by `Insurance Provider`?
2. Which `Medical Condition` has the highest average billing amount?
3. What is the distribution of `Admission Type` across different `Insurance Provider`s?

D. Treatment & Outcomes Analysis:
1. What is the most common `Medication` prescribed for each `Medical Condition`?
2. How does the `Test Results` (Normal, Abnormal, Inconclusive) breakdown look for each condition?
3. Is there a relationship between the `Length of Stay` and the `Test Results`?

4. Technical Approach in SSMS

1. **Data Import:** The CSV file will be imported into a new SQL Server database table named `HealthData` using the SSMS Import Flat File wizard or a custom `BULK INSERT/OPENROWSET` script.
2. **Data Cleaning & Preparation:**
    - Standardize text fields (e.g., correct `Name`, `Gender` capitalization inconsistencies like 'Male' vs. 'MALE').
    - Ensure `Date of Admission` and `Discharge Date` are in a proper `DATE` or `DATETIME` data type.
    - Check for and handle any missing or invalid values (e.g., negative `Billing Amount`).
    - Create a new calculated column for `LengthOfStay`.
3. **Data Analysis:** Write and execute a series of SQL `SELECT` queries, utilizing `GROUP BY`, `JOIN` (if necessary after normalization), aggregate functions (`COUNT`, `SUM`, `AVG`), and `ORDER BY` to answer the business questions.
4. **Reporting:** The results of these queries will be presented to the client. Options include:
    - Exporting result sets to Excel for charting.
    - Creating SQL Views for frequently accessed data summaries.
    - Building a simple SSRS (SQL Server Reporting Services) report dashboard.

5. Deliverables

1. A fully populated and cleaned `HealthData` table within a SQL Server database.
2. A SQL script file (`.sql`) containing all queries used for the analysis.