



Data Normalization — Healthcare Data

Project Overview

In real-world engineering, tasks are often defined by high-level requirements rather than detailed, step-by-step instructions. This project mimics such scenarios, challenging to design and implement software applications by applying your knowledge, creativity, and problem-solving skills.

Project Description

I've worked with legacy flat structured healthcare system data. The data contains key information around doctor appointments.

The task is to transform this raw data into a structured database schema. Specifically, to process and normalize the data into a **snowflake schema**, creating an optimized structure that supports efficient querying and data management.

Requirements

Note: Pandas package was not used for processing data

Data Normalization Requirements

- Generate CSV files for all 11 tables in the schema

- Column sequence and name must match the schema provided below
- The primary key of different tables can either be string or integer (even if they have different types in the diagram)
- [Diagram PNG file can be found here](#) for better view
- Patient status:
 - Active
 - Inactive (have not visited since end of 2021 — last day in 2021)

