Sfundo nondwatyu

Coding and Programming Capstone Project 2024

Group Name

|  |
| --- |
| Individual Work |

Group Members

|  |  |
| --- | --- |
| **Student Number** | **Name and Surname** |
| 220036624 | Sfundo Nondwatyu |

Data Set Used URL

|  |
| --- |
| <https://www.kaggle.com/datasets/devildyno/hospital-patient-records-jan-2021-july-2024/data> |

# Project Description

**Title: Hospital Patient Records Data Analysis and Visualization**

**Problem Statement:**

Hospitals generate vast amounts of data daily, encompassing patient demographics, medical conditions, treatments, billing, and more. This data holds valuable insights that can enhance hospital operations, improve patient care, and streamline financial management. However, without proper analysis and visualization, these insights remain hidden, leading to inefficiencies and missed opportunities. Our project aims to address this challenge by creating a comprehensive tool for analysing and visualizing hospital patient records.

**Proposed Solution:**

My solution involves developing a Python-based application that reads hospital patient records from a CSV file, performs data analysis, and generates insightful visualizations. The application will leverage libraries such as Pandas for data manipulation, Matplotlib and Seaborn for plotting. The key functionalities will include:

* Data cleaning and preprocessing to handle missing values and inconsistencies.
* Statistical analysis to summarize key metrics such as median bill amounts and common medical conditions (mode).
* Visualizations to display patient demographics, bill distributions, and medical condition frequencies.
* Insights to help hospital administrators make informed decisions and improve operational efficiency.

**Additional Technologies:**

In addition to Python and its libraries, I used:

* Kaggle for Datasets - [***Link***](https://www.kaggle.com/datasets/devildyno/hospital-patient-records-jan-2021-july-2024/data)
* GitHub for version control.
* MS Office Word to compile this document.

## Objectives

Data Preprocessing:

* Handled missing values and cleaned data.
* Ensured data types are correctly set for analysis.

Data Analysis:

* Computed summary statistics such as median, and mode for relevant columns/features.
* Identified the most common medical conditions and their distributions.

Data Visualization:

* Created bar graphs, pie chart, and histogram to visualize patient demographics and bill amounts.
* Used count plots to show the distribution of medical conditions by gender.

## Project Plan and Timeline

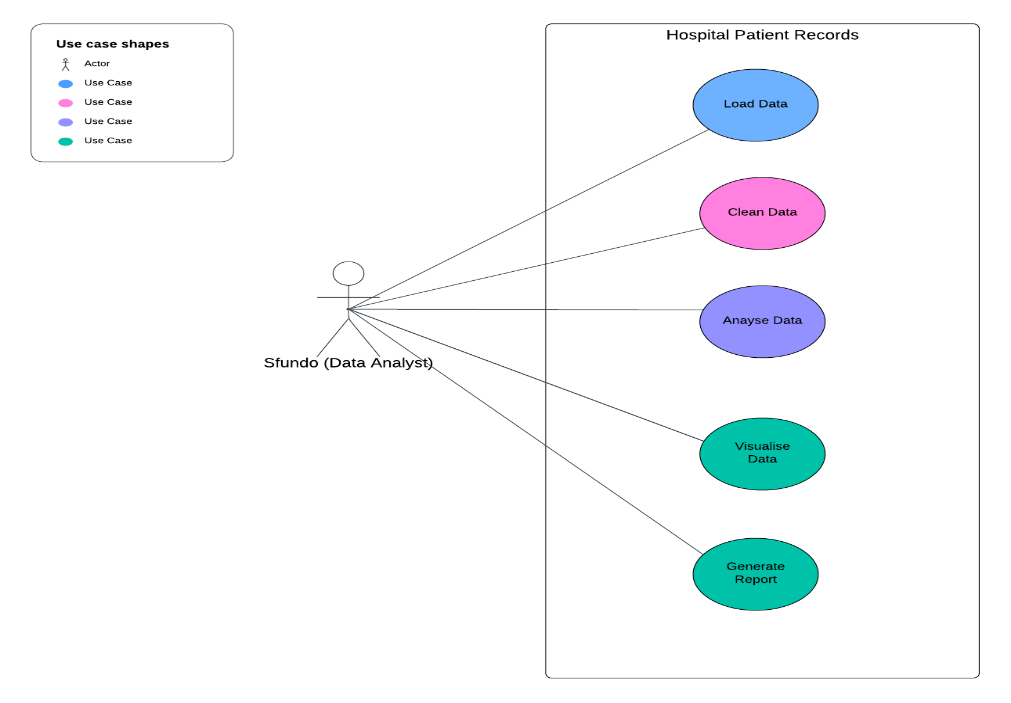
Phase 1: Project Initialization (Week 1)

* Define project scope and objectives.
* Gather and review hospital patient records data.
* Set up the development environment and tools.

Phase 2: Data Preprocessing, Analysis and Visualisation (Week 2)

* Load data (CSV) into a Pandas DataFrame.
* Handle missing values and data inconsistencies.
* Validate and clean data types.
* Calculate summary statistics for key columns (Bill Amount).
* Identify top medical conditions and their frequency.
* Create visualizations for patient demographics (gender distribution).
* Visualize bill amounts using histogram and pie chart.
* Use count plots for medical conditions by gender.

## Use Case Diagram:



## Findings

**Patient Demographics:** The bar chart visualization indicates that there is a slightly higher representation of females in the dataset compared to males. This understanding of patient demographics can assist in resource allocation and the development of targeted healthcare strategies.

**Total Bill Amounts:** The pie chart visualisation highlights the total bill amounts categorized by gender. It shows that females (51%) contribute to a slightly higher portion of the total bill amounts compared to males (49%). This insight can help financial management and resource planning within the hospital.

**Top Medical Conditions:** By analysing the medical conditions, the analysis identified the top medical conditions present in the dataset. The bar chart visualization presents the distribution of patients per medical condition, segmented by gender. The top medical conditions include conditions like Skin Infection, Common Cold, etc. Understanding the prevalence of such medical conditions can aid in resource allocation, treatment planning, and the development of targeted healthcare programs.

**Bill Amount Distribution:** The histogram visualization provides an overview of the distribution of bill amounts in the dataset. It indicates that the majority of bill amounts fall within a certain range, with a peak around a specific value. This understanding of the bill amount distribution can assist in financial management, budgeting, and billing processes within the hospital.