**Final Project**

Knowledge Management Tools and Technologies

INFO 5307

**Predictive Analytics for Hospital Length of Stay**

VARUN KUMAR ATKURI

11694018

**1.Monday.com URL:**

<https://view.monday.com/8004386202-08761f343da0be6c25f223fba153d088?r=use1>

**2. System Architecture:**

**A diagram of a hospital dashboard

Description automatically generated** *System Architecture showcasing the end-to-end data pipeline for the Hospital Dashboard Project.*

**3. Description of the Dataset and Analytical Solution**

**1.Dataset Description:**

* **Source**: The dataset was sourced from **Kaggle** and includes anonymized hospital records. It contains critical fields for analyzing patient length of stay (LOS), discharge outcomes, and facility performance.

**Key Attributes**: The dataset includes fields essential for analyzing hospital performance and patient outcomes:

* eid: Unique patient identifier.
* vdate: Admission date.
* discharged: Discharge date.
* lengthofstay: Number of days spent in the hospital.
* gender: Gender of the patient (Male/Female).
* facid: Facility description where the patient received treatment.
* Various medical conditions, encoded numerically (e.g., asthma, renalendstage).

**Dataset Size**:

* Rows: 100,000
* Columns: 31

**2. Preprocessing Steps**

* **Environment Used**: Data cleaning and preprocessing were conducted in **Jupyter Notebook** using Python with the **Pandas** and **NumPy** libraries. The cleaned dataset was stored locally for upload into Power BI.

**Steps Performed**

**Data Cleaning**:

* Checked for missing values and ensured that key columns like lengthofstay and vdate had no null entries.
* Handled inconsistent or erroneous data (e.g., removing negative or extremely high values in lengthofstay).
* Standardized date formats for vdate and discharged columns using the datetime module

**Derived Columns**:

* Calculated missing lengthofstay values where necessary.

**Exploratory Data Analysis (EDA)**:

* Conducted statistical analysis to identify trends:
  + Average length of stay across facilities.
  + Gender-based and facility-based discharge trends.
* Identified outliers in lengthofstay for further investigation.

**Feature Engineering**:

* Encoded categorical fields (gender, facid) for efficient modeling and analysis.

**3. Analytical Solution**

* **Goal**: Transform the raw Kaggle dataset into a structured and actionable format for analysis and visualization in Power BI.
* **Steps to Implement Solution**:
  1. **Data Cleaning and Preparation**:
     + Jupyter Notebook was used to clean and preprocess the dataset, ensuring data quality and consistency.
  2. **Storage**:
     + Saved the cleaned dataset locally as a CSV file (Cleaned\_LengthOfStayData.csv) for import into Power BI.
  3. **Analysis and Visualization**:
     + Uploaded the cleaned dataset into Power BI for dashboard creation.
     + Created interactive visualizations (e.g., KPIs, bar charts, treemaps) to analyze trends and patterns.

**4. Tools Used**

* **Jupyter Notebook**: For preprocessing and exploratory data analysis.
* **Pandas and NumPy**: For data cleaning, feature engineering, and statistical computations.
* **Matplotlib/Seaborn**: For initial visualizations during EDA.
* **Power BI**: For interactive dashboard creation and visual analytics.

**Code Snippets**

Here are some key Python codes implemented in Jupyter Notebook:

1. **Handling Missing Values**:

data['lengthofstay'] = data['lengthofstay'].fillna(

(data['discharged'] - data['vdate']).dt.days

)

1. **Standardizing Dates**:

data['vdate'] = pd.to\_datetime(data['vdate'])

data['discharged'] = pd.to\_datetime(data['discharged'])

1. **Filtering Outliers**:

data = data[(data['lengthofstay'] > 0) & (data['lengthofstay'] <= 365)]

1. **Encoding Categorical Data**

data['gender\_encoded'] = data['gender'].map({'M': 1, 'F': 0})

**Dashboard**

**1. Dashboard Overview**

* **Title:** Hospital Length of Stay Analysis Dashboard

**Purpose:**

* To analyze hospital patient outcomes, length of stay (LOS), and facility performance.
* Provides actionable insights to hospital administrators for improving patient care and resource allocation.

**2. Public URL**

<https://app.powerbi.com/links/Jj5QdhwNFl?ctid=70de1992-07c6-480f-a318-a1afcba03983&pbi_source=linkShare&bookmarkGuid=68b3e4da-b0b8-4a39-a4b5-d51e3cf79dbc>

**3. Key Components**

**Page 1: Length of Stay Analysis**

* **KPIs:**
  + **Average LOS**: Displays the overall average length of stay for all patients.
  + **Total Patients**: Shows the total number of patients in the dataset.
  + **Total Discharges**: Total number of patient discharges (Home/Expired).
* **Trends Chart:**
  + Displays monthly trends of patient discharges and LOS.
* **Facility Analysis:**
  + Bar chart comparing average LOS across hospital facilities.
* **Health Condition Analysis:**
  + Treemap showing average LOS by medical condition (e.g., Renal Failure, Fibrosis).

**Page 2: Patient Outcomes and Insights**

* **KPIs:**
  + **Total Readmissions**: Total number of patient readmissions.
  + **Critical Cases**: Total number of patients categorized as critical.
* **Gender-Wise Comparison:**
  + Bar chart showing LOS across health conditions for male and female patients.
* **Discharge Status Distribution:**
  + tacked bar chart visualizing the proportion of Home vs. Expired outcomes for each facility.
* **Raw Data Table:**
  + Provides a detailed view of patient records for transparency and drill-down analysis.

**User Guide for Dashboard**

* **Filters and Interaction:**
  + "Use the filters on the left-hand side to adjust the analysis based on Gender, Facility, or Discharge Date."
  + "Hover over data points in charts to view detailed values (e.g., LOS, discharge counts)."
* **Interpreting Visuals:**
  + "The Treemap for Health Conditions shows which conditions are associated with longer stays. Darker colors indicate higher LOS."
  + "The Discharge Status chart highlights facilities with higher expired rates, like Cardiology and Neurology."
* **Actionable Steps:**
  + "If a facility shows high expired rates, consider investigating resources and patient management processes in that department."
  + "Focus on health conditions with high LOS (e.g., Fibrosis) to optimize treatment protocols."

**Value to the Organization**

* Enables hospital administrators to identify underperforming facilities and health conditions needing attention.
* Facilitates data-driven decision-making for optimizing resources and improving patient outcomes.