Saint Bridget Medical Center

Data Analysis Report

Lava McCoy
Data Analyst

November 29th, 2025

Report# 401994-C

Table of Contents

Introduction/Purpose……………………………………….2

Data Sources & Preparation………………………………..3

Methodology……………………………………………….5

Observations ……………………………………………....7

2023 Observations ………………….………...…7

2024 Observations……………………...……..8

Average Length of Stay per Diagnosis……………9

Geographic Analysis – Heatmap by City……….....10

Summary of Key Observations …………………………..11

Data Quality Assessment ………………………………...12

Appendix…………………………………………….....13

Appendix A…………………………………….13

Appendix B……………………………………....14

Introduction/Purpose

This analysis was conducted in response to a request from Dr. Sarah Johnson, Chief Medical Officer, Saint Bridget Medical Center, to evaluate patient admissions over the past two years (January 1, 2023 – December 31, 2024). The goal of the project is to provide actionable insights that can help improve hospital workflow and patient care by examining admissions patterns, diagnosis trends, and patient length of stay.

Specifically, this report addresses the following stakeholder needs:

1. **Total Admissions:** Analyze the total number of patient admissions per month and year to identify overall trends.
2. **Diagnosis Breakdown:** Provide a detailed view of key diagnoses (Chickenpox, Flu, COVID-19, Cold, Asthma) over time, including seasonal patterns such as flu peaks in winter.
3. **Average Length of Stay:** Determine which diagnoses are associated with longer hospital stays to inform resource allocation.
4. **Data Quality Assessment:** Identify missing or inconsistent patient data (e.g., discharge dates, demographic inconsistencies) that could affect reporting accuracy.
5. **Recommendations:** Suggest improvements for patient record data quality to support future analyses.
6. **Geographic Analysis:** Generate a Texas city-level heatmap of patient admissions to detect regions with unusually high or clustered diagnosis patterns.

Data Sources & Preparation

The analysis utilized multiple data sources and tools to provide a comprehensive view of patient admissions at Saint Bridget Medical Center from January 1, 2023 to December 31, 2024**.** All data was cleaned, processed, and integrated to ensure accuracy and support the stakeholder's request.

1. **Patient Admission Data (Excel / CSV)**
   - The raw dataset consisted of 2,000 patient records with columns: Patient_ID, DOB, Admission_Date, Discharge_Date, Diagnosis, City, and State.
   - A **cleaned dataset** was created in Excel, including handling of missing or inconsistent data:
     - 67 missing Discharge_Date values were updated to "NA".
     - Latitude and longitude columns were added to support GIS mapping.
   - The cleaned Excel sheet served as the primary source for subsequent analysis in SQL, Python, and QGIS.
2. **SQL Database & Queries**
   - Database "SaintBridget" and table "SBMCRecords" were created to store patient records.
   - Key queries included:
     - First15Rows: preview of initial data
     - ColumnCheck: verify all columns present
     - AdmissionsPerMonthAndYear: total admissions by month/year
     - LongestAvgHospitalStay: identify diagnoses with the longest average stay
     - MonthlyDiagnosis: count of key diagnoses (Chickenpox, Flu, COVID-19, Cold, Asthma) per month
3. **Python / Jupyter Notebook**
   - A Jupyter Notebook was used to calculate average length of stay per diagnosis**.**
   - The notebook also helped verify data quality and highlight inconsistencies in the dataset.
   - Screenshots of the calculations were included for transparency.
4. **PowerBI Dashboards**
   - Interactive visualizations were created to summarize admissions and diagnoses:
     - Pie charts showing seasonal diagnosis trends (Spring, Summer, Fall, Winter) for 2023 and 2024
     - Bar charts showing total admissions per month
     - Bar charts showing diagnosis type per month
   - Separate pages were created for 2023 and 2024 data to allow comparison across years.
5. **Geographic Data & QGIS Analysis**
   - Texas state boundary and city shapefiles were downloaded from census.gov**.**
   - CSV points representing patient admission locations were imported and projected using EPSG:4326 (latitude/longitude), then reprojected to EPSG:3083 (Texas Centric Albers) for heatmap generation.
   - A heatmap of patient admissions by city was generated using kernel density estimation (radius = 50 km, pixel size = 1,000–2,000 m).

    - City points were overlayed and labeled to match the admissions data, providing a clear visual of geographic clustering.
6. **File List / Outputs**
    - Heatmap (Texas_Patient_Cities_Heatmap.tif)
    - PowerBI Dashboards (2023 & 2024)
    - Jupyter Notebook screenshots for average length of stay
    - Cleaned CSV dataset

Preparation Summary:
All data was checked for completeness, merged as needed, and prepared for analysis in SQL, Python, PowerBI, and QGIS. The workflow ensured consistency across tools and enabled the generation of accurate visualizations and metrics aligned with the stakeholder's request.

Methodology

The following methodology outlines the steps taken to address the stakeholder's request for patient admission analysis, linking each requirement to the specific tools and processes used.

1. **Total Admissions per Month/Year**
   - **Tool:** SQL & PowerBI
   - **Process:**
     - SQL query AdmissionsPerMonthAndYear extracted the total number of admissions grouped by month and year from the cleaned dataset.
     - PowerBI bar charts were created to visualize admissions trends across 2023 and 2024, enabling quick identification of peaks and overall patient volume patterns.
2. **Detailed Breakdown of Diagnoses Over Time**
   - **Tool:** SQL & PowerBI
   - **Process:**
     - SQL query MonthlyDiagnosis calculated the number of patients diagnosed with Chickenpox, Flu, COVID-19, Cold, and Asthma per month.
     - PowerBI dashboards displayed seasonal trends, dividing the year into Spring (Mar–May), Summer (Jun–Aug), Fall (Sep–Nov), and Winter (Dec–Feb).
     - Separate dashboard pages were created for 2023 and 2024 for comparative analysis.
3. **Average Length of Stay per Diagnosis**
   - **Tool:** Python / Jupyter Notebook
   - **Process:**
     - Patient admission and discharge dates were imported into a Jupyter Notebook.
     - Python code calculated the average length of stay for each diagnosis.
     - Screenshots of the notebook results were included to provide transparency and reproducibility.
4. **Identification of Missing or Inconsistent Patient Data**
   - **Tool:** Excel & Jupyter Notebook
   - **Process:**
     - Audit of the cleaned Excel dataset revealed missing Discharge_Date values (67 records) and other inconsistencies.
     - Data cleaning included updating missing discharge dates to "NA" and verifying latitude/longitude for GIS mapping.
     - Jupyter Notebook checks confirmed no remaining critical inconsistencies affecting the analysis.
5. **Recommendations for Improving Data Quality**
   - **Tool:** Excel audit log & observations from SQL/Python
   - **Process:**
     - Documented missing or invalid data points during cleaning.

        - Recommendations include implementing data validation rules in patient records, mandatory entry of discharge dates, and regular audits to maintain dataset integrity.

6. **Geographic Analysis of Admissions**
    - **Tool:** QGIS
    - **Process:**
        - Cleaned CSV patient records were imported as point layers and projected to EPSG:3083 – Texas Centric Albers.
        - A heatmap was generated using kernel density estimation with a 50 km radius and 1,000–2,000 m pixel size to visualize areas of high patient concentration.
        - Texas state boundaries and city layers (from census.gov) were overlaid to provide geographic context.
        - City points were labeled to highlight locations with patient admissions, allowing stakeholders to identify regional patterns and potential hotspots.

Notes on Workflow Integration

- All steps were documented in the audit log, ensuring reproducibility and transparency.
- Outputs from SQL, Jupyter, PowerBI, and QGIS were integrated into the final report, providing both numerical and visual evidence aligned with stakeholder objectives.

Observations

2023 Observations

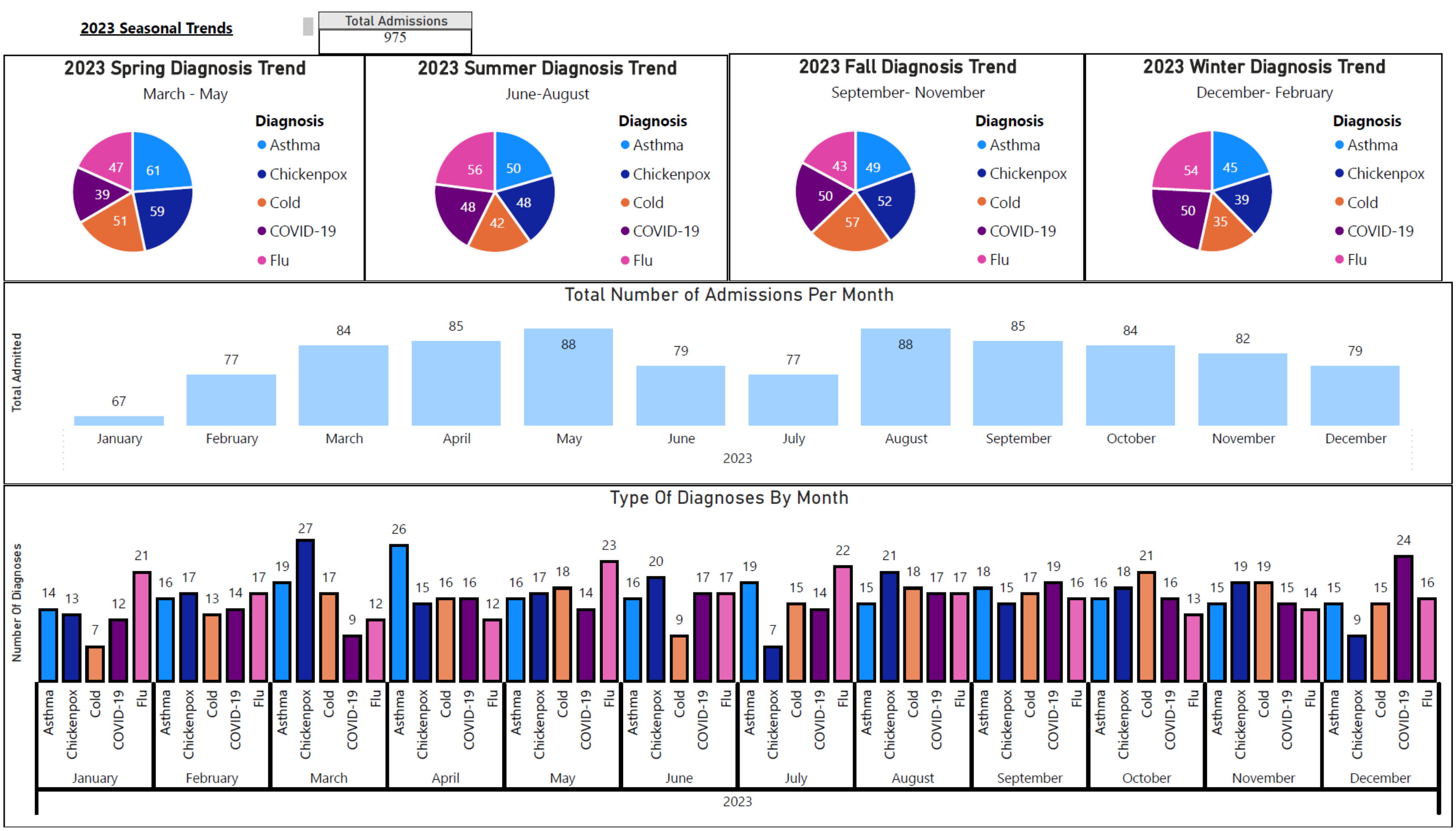
Visualization: Showing total patient admissions, seasonal diagnosis trends for 2023.

- **Source:** SQL query AdmissionsPerMonthAndYear → PowerBI dashboard

  SQL query MonthlyDiagnosis → PowerBI seasonal charts

- **Observation:**
  - Total admissions for 2023 were 975**.**
  - Admissions peaked in ***May*** and ***August***, with ***88*** admissions each month.
  - Asthma peaked in the spring months
  - The flu peaked in the summer and winter months
  - Colds peaked in the fall months

Figure 1: Total Patient Admissions per Month, 2023.



2024 Observations

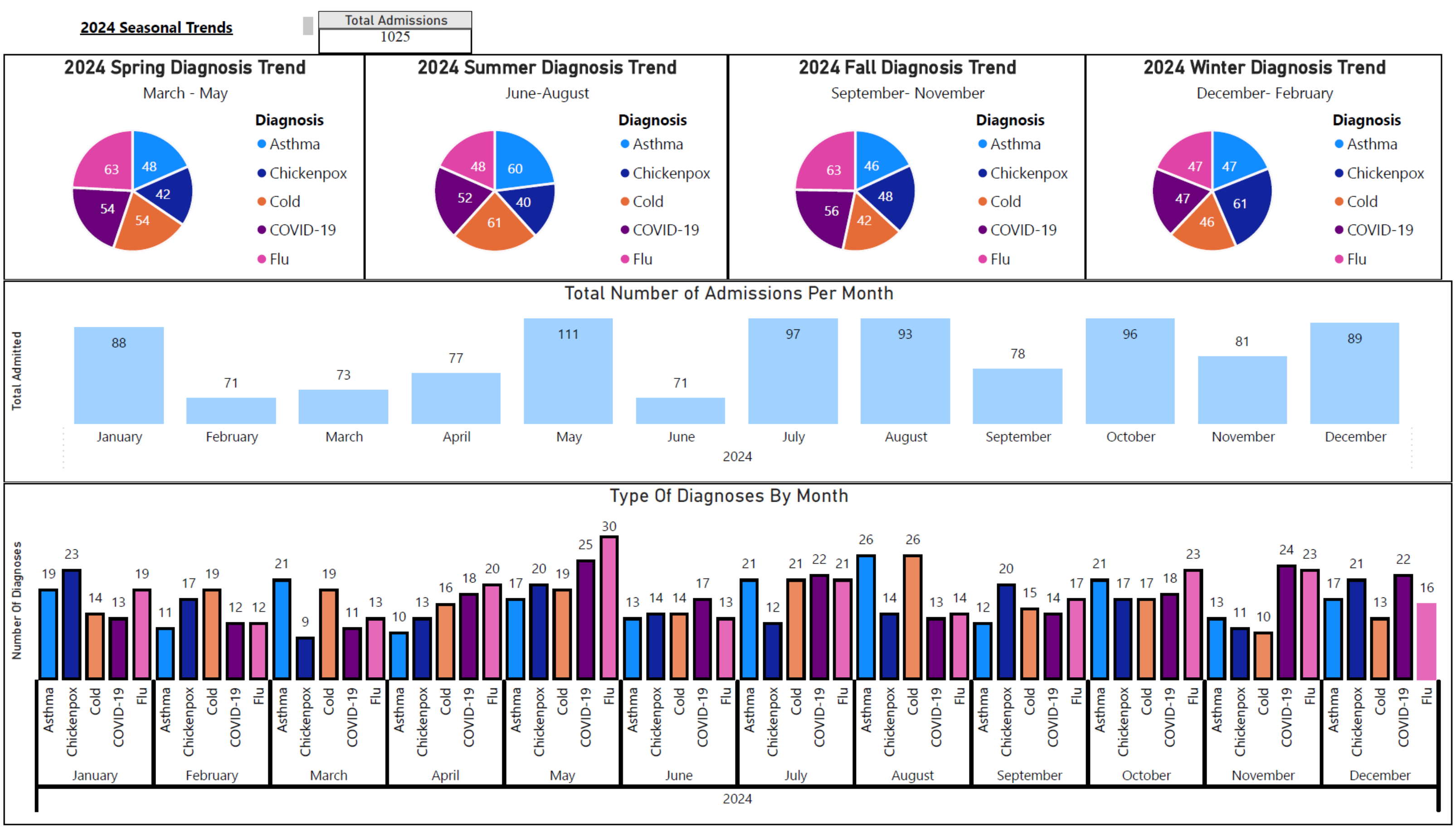
Visualization: Showing total patient admissions, seasonal diagnosis trends for 2024.

- **Source:** SQL query AdmissionsPerMonthAndYear → PowerBI dashboard

  SQL query MonthlyDiagnosis → PowerBI seasonal charts

- **Observation:**
  - Total admissions for 2024 were 1025**.**
  - Admissions peaked in ***May – 111*** admissions, ***July – 97*** admissions and ***October – 97*** admissions.
  - Flu peaked in the spring and fall months.
  - Asthma peaked in the summer months.
  - Chickenpox peaked in the winter months.

Figure 2: Total Patient Admissions per Month, 2024.



Average Length of Stay per Diagnosis

Visualization: Jupyter Notebook outputs calculating average length of stay

- **Source:** Python / Jupyter Notebook
- **Observation:**
  - Diagnoses with the longest average hospital stays include COVID-19 with 6.58 days and Asthma with 3.57 days.
  - Shorter stays observed for the flue with 1.93 days and colds with 1.33 days.

Figure 3: Average Length of Stay per Diagnosis

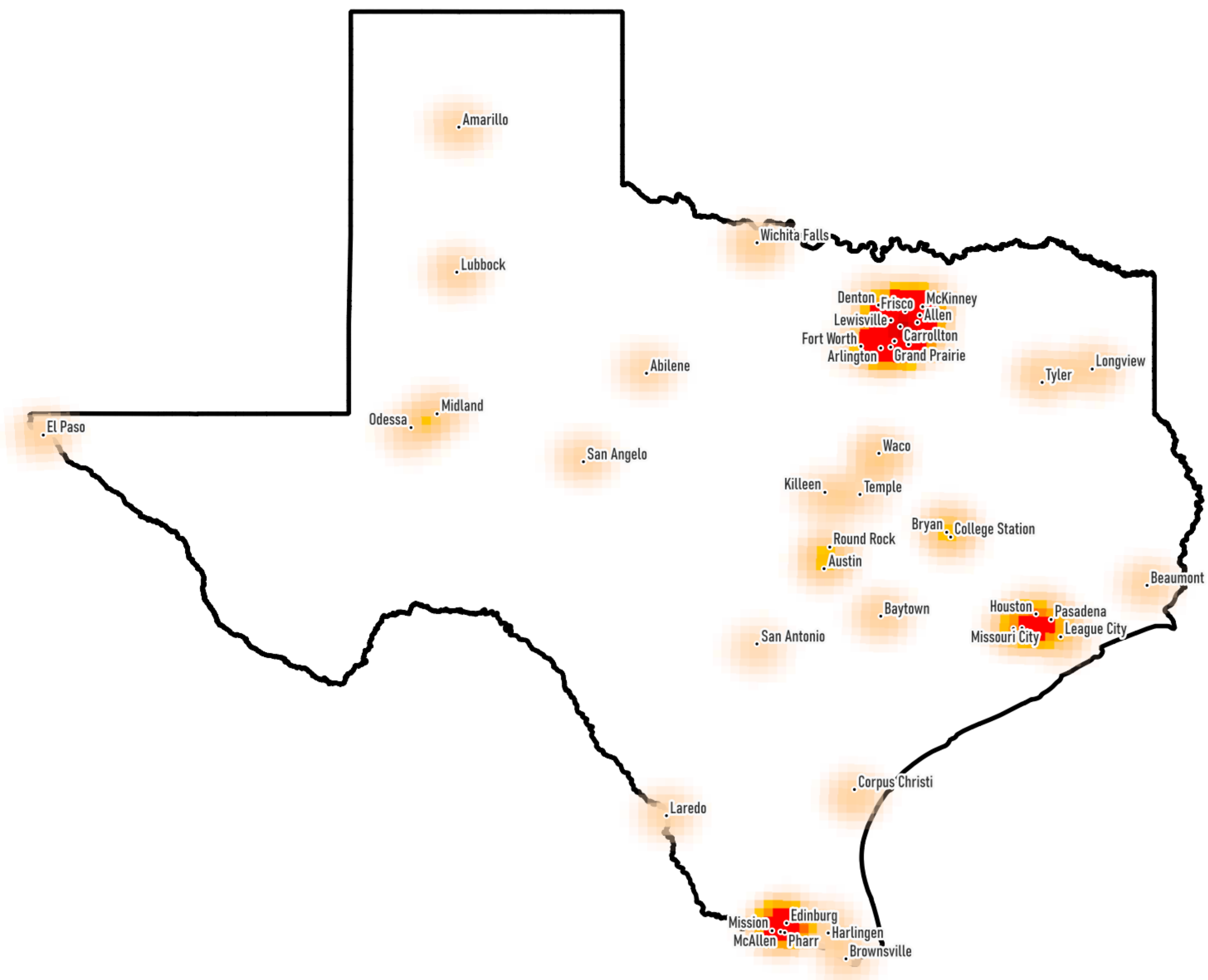
Diagnosis	Length_of_Stay
COVID-19	6.58
Asthma	3.57
Chickenpox	2.54
Flu	1.93
Cold	1.33

Geographic Analysis – Heatmap by City

Visualization: QGIS-generated heatmap of patient admissions with city overlays

- **Source:** CSV patient records → QGIS heatmap → Texas state and city layers
- **Observation:**
  - Heatmap clearly identifies regional clusters and potential hotspots for hospital workflow attention.

Figure 4: Heatmap of Patient Admissions by City in Texas



Summary of Key Observations

1. Admissions fluctuate seasonally, with predictable peaks for flu and other respiratory illnesses.
2. Certain diagnoses require longer hospital stays, which should be factored into workflow and bed management planning.
3. Geographic clustering of admissions in major Texas cities can help target resources efficiently.
4. Data cleaning and quality control are essential to ensure accurate reporting and reliable visualizations.

Data Quality Assessment

As requested by the stakeholder, an in-depth review of patient data was conducted to identify any missing or inconsistent information that could affect reporting accuracy and analysis outcomes.

1. Identification of Missing or Inconsistent Data

- **Discharge Dates:** 67 patient records were missing Discharge_Date values.

2. Actions Taken

- Missing discharge dates were manually or programmatically updated to “NA” to ensure downstream calculations would not fail.

3. Recommendations for Improving Data Quality

To prevent future inconsistencies and ensure reliable reporting:

1. **Mandatory Data Fields:** Require all critical fields such as Discharge_Date and Diagnosis to be entered before record submission.
2. **Regular Audits:** Periodically review the patient records to identify missing or inconsistent data before analysis.
3. **Documentation:** Review the systems audit log for all data cleaning activities to track corrections and modifications.

Summary:
These measures will help maintain dataset integrity, prevent errors in analysis, and ensure that future reports provide accurate and actionable insights for hospital management.

Appendix

Appendix A — SQL Queries Used in the Analysis

A1.Determined Average Length of Stay by Diagnosis

```python
import pandas as pd

# Load CSV
df = pd.read_csv(r"C:\Users\amigo\Desktop\Data Analyst Path\Projects\Project 5- Saint Bridget Medical Center\Working In\SQL READY.csv")

# CAdmission and Discharge to datetime
df["Admission_Date"] = pd.to_datetime(df["Admission_Date"])
df["Discharge_Date"] = pd.to_datetime(df["Discharge_Date"])

# Length of Stay in days
df["Length_of_Stay"] = (df["Discharge_Date"] - df["Admission_Date"]).dt.days

# average Length of Stay per Diagnosis
avg_los = df.groupby("Diagnosis", as_index=False)["Length_of_Stay"].mean().round(2)

# sort by longest stay first
avg_los = avg_los.sort_values("Length_of_Stay", ascending=False)

#result
avg_los
```

Python

	Diagnosis	Length_of_Stay
1	COVID-19	6.58
0	Asthma	3.57
2	Chickenpox	2.54
4	Flu	1.93
3	Cold	1.33

Appendix B — SQL Queries Used in the Analysis

B1. Verified Columns

```sql
SELECT name AS ColumnName
FROM sys.columns
WHERE object_id=
OBJECT_id('dbo.SBMCREcords');
```

100 % No issues found

Results Messages

	ColumnName
1	Patient_ID
2	DOB
3	Admission_Date
4	Discharge_Date
5	Diagnosis
6	City
7	State

B2. Displayed First 15 Rows to Verify Correct File

```sql
SELECT TOP 15*
FROM SBMCRecords
```

100 % No issues found

Results Messages

	Patient_ID	DOB	Admission_Date	Discharge_Date	Diagnosis	City	State
1	P00000	1946-09-20	2023-05-09	2023-05-10	Cold	Grand Prairie	TX
2	P00001	1958-11-28	2024-03-24	2024-03-27	Asthma	Frisco	TX
3	P00002	1950-02-19	2024-09-06	2024-09-08	Flu	Bryan	TX
4	P00003	1999-05-07	2023-03-31	2023-04-01	Flu	Mission	TX
5	P00004	1954-08-06	2023-09-29	2023-10-02	Chickenpox	Arlington	TX
6	P00005	1970-02-24	2023-08-29	2023-08-31	Chickenpox	Brownsville	TX
7	P00006	1991-05-05	2023-05-26	2023-06-02	COVID-19	Dallas	TX
8	P00007	1958-07-28	2024-12-05	2024-12-09	COVID-19	Sugar Land	TX
9	P00008	1999-08-19	2024-03-10	2024-03-14	Asthma	Missouri City	TX
10	P00009	1949-06-29	2024-11-14	2024-11-20	COVID-19	Grand Prairie	TX
11	P00010	2004-07-30	2024-10-27	2024-10-28	Flu	Temple	TX
12	P00011	1982-11-20	2023-04-13	2023-04-14	Cold	McAllen	TX
13	P00012	1957-04-18	2024-09-02	2024-09-03	Cold	Garland	TX
14	P00013	1963-12-21	2023-07-28	2023-07-29	Flu	Harlingen	TX
15	P00014	1935-01-30	2023-09-16	2023-09-20	Asthma	Odessa	TX

B3. Total Admission Per Month

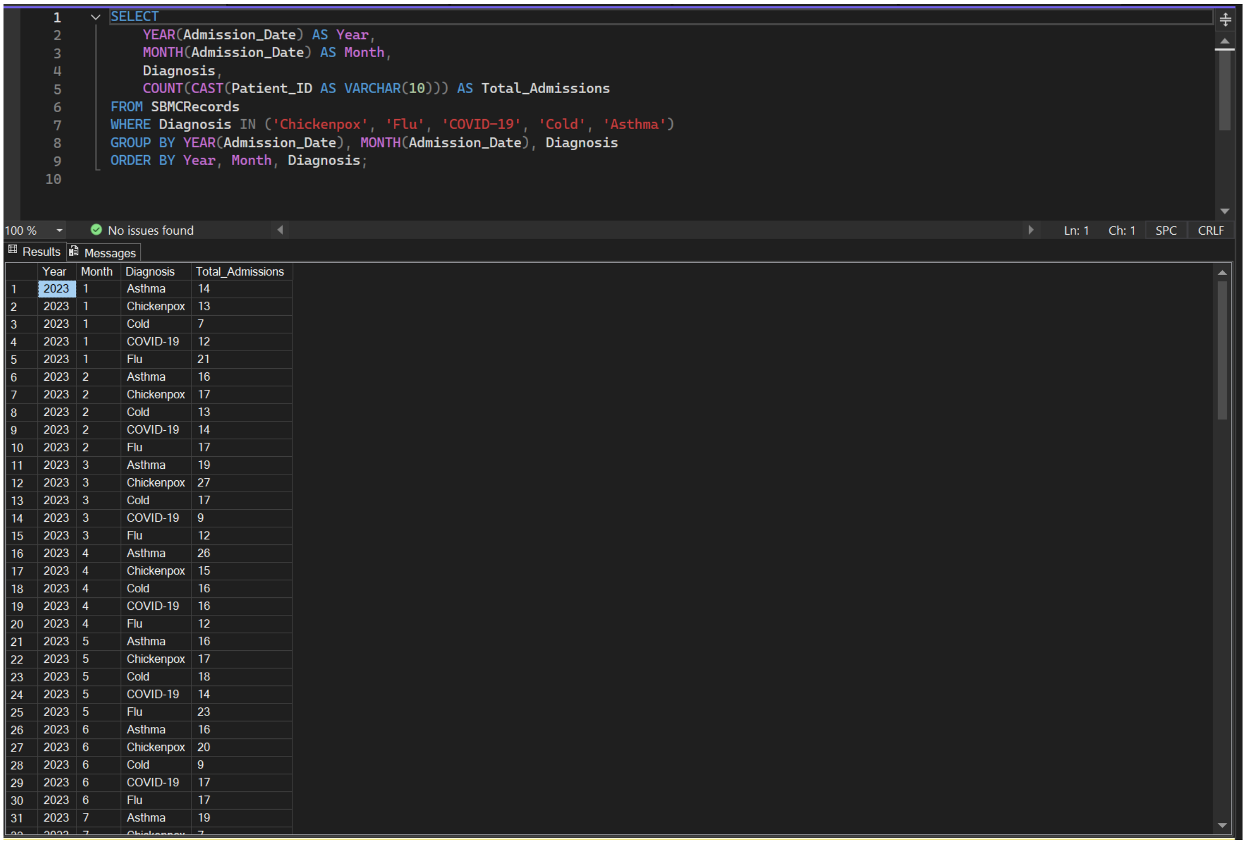
```sql
SELECT
    YEAR(Admission_Date) AS Year,
    MONTH(Admission_Date) AS Month,
    COUNT(CAST(Patient_ID AS VARCHAR(10))) AS Total_Admissions
FROM SBMCRecords
GROUP BY YEAR(Admission_Date), MONTH(Admission_Date)
ORDER BY Year, Month;
```

100 % No issues found

Results Messages

	Year	Month	Total_Admissions
1	2023	1	67
2	2023	2	77
3	2023	3	84
4	2023	4	85
5	2023	5	88
6	2023	6	79
7	2023	7	77
8	2023	8	88
9	2023	9	85
10	2023	10	84
11	2023	11	82
12	2023	12	79
13	2024	1	88
14	2024	2	71
15	2024	3	73
16	2024	4	77
17	2024	5	111
18	2024	6	71
19	2024	7	97
20	2024	8	93
21	2024	9	78
22	2024	10	96
23	2024	11	81
24	2024	12	89

B4. Total Number of Diagnoses Per Month & Year



```sql
SELECT
    YEAR(Admission_Date) AS Year,
    MONTH(Admission_Date) AS Month,
    Diagnosis,
    COUNT(CAST(Patient_ID AS VARCHAR(10))) AS Total_Admissions
FROM SBMCRecords
WHERE Diagnosis IN ('Chickenpox', 'Flu', 'COVID-19', 'Cold', 'Asthma')
GROUP BY YEAR(Admission_Date), MONTH(Admission_Date), Diagnosis
ORDER BY Year, Month, Diagnosis;
```

100 % No issues found Ln: 1 Ch: 1 SPC CRLF

Results Messages

	Year	Month	Diagnosis	Total_Admissions
1	2023	1	Asthma	14
2	2023	1	Chickenpox	13
3	2023	1	Cold	7
4	2023	1	COVID-19	12
5	2023	1	Flu	21
6	2023	2	Asthma	16
7	2023	2	Chickenpox	17
8	2023	2	Cold	13
9	2023	2	COVID-19	14
10	2023	2	Flu	17
11	2023	3	Asthma	19
12	2023	3	Chickenpox	27
13	2023	3	Cold	17
14	2023	3	COVID-19	9
15	2023	3	Flu	12
16	2023	4	Asthma	26
17	2023	4	Chickenpox	15
18	2023	4	Cold	16
19	2023	4	COVID-19	16
20	2023	4	Flu	12
21	2023	5	Asthma	16
22	2023	5	Chickenpox	17
23	2023	5	Cold	18
24	2023	5	COVID-19	14
25	2023	5	Flu	23
26	2023	6	Asthma	16
27	2023	6	Chickenpox	20
28	2023	6	Cold	9
29	2023	6	COVID-19	17
30	2023	6	Flu	17
31	2023	7	Asthma	19