Saudi Arabia hospital admissions analysis

وزارة الصحة

Ministry of Health

INTRODUCTION

In our project we aim to make an interactive dashboard . It will be useful for hospital operational managers .So that we investigate a dataset in hospital admission. From it we will know the total admissions in the hospital , the average stay , ratio of readmission , the season with the highest visits and the largest admitted age group. All of that will help the hospital operational managers doing their job consist of ensuring that the hospital works smoothly and efficiently.

DATA EXPLORATION

How The Data Was Collected :

The data was collected from Kaggle Datasets. This data contains hospital admission records from Saudi Arabia (2020 - 2024). It contains 41,544 admission records across 8 hospitals and 5 cities (Riyadh, Jeddah, Dammam, Mecca, Medina). It contains 4 medical conditions tracked , Complete patient demographics , Treatment outcome and readmission data . It was downloaded with the api provided from kaggle which loaded into pandas dataframes for analysis.

Features Identified for Analysis :

The features identified for analysis are year , month , day and day of week . Actually they are extracted from an existing feature called admission_date. The reason for choosing these features is to improve operations like grouping ,filtering and make temporal patterns explicit.

Screenshots of Data cleaning and pre-processing:

Install & Import

```
!pip install kagglehub -q

import kagglehub
import pandas as pd
import os
```

Download & Load Data

```
path = kagglehub.dataset_download("datasetengineer/riyadh-hospital-admissions-dataset-20202024")

files = os.listdir(path)
csv_file = [f for f in files if f.endswith('.csv')][0]
df = pd. read_csv(f"{path}/{csv_file}")
```

```
Using Colab cache for faster access to the 'riyadh-hospital-admissions-dataset-20202024' dataset.
```

Fig.1 Adding the dataset into the colab and pandas dataframe.

```
array([[<Axes: title={'center': 'admission_count'}>,
        <Axes: title={'center': 'readmission_count'}>],
       [<Axes: title={'center': 'length_of_stay_avg'}>,
        <Axes: title={'center': 'comorbid_conditions_count'}>],
       [<Axes: title={'center': 'daily_medication_dosage'}>,
        <Axes: title={'center': 'emergency_visit_count'}>]], dtype=object)
```

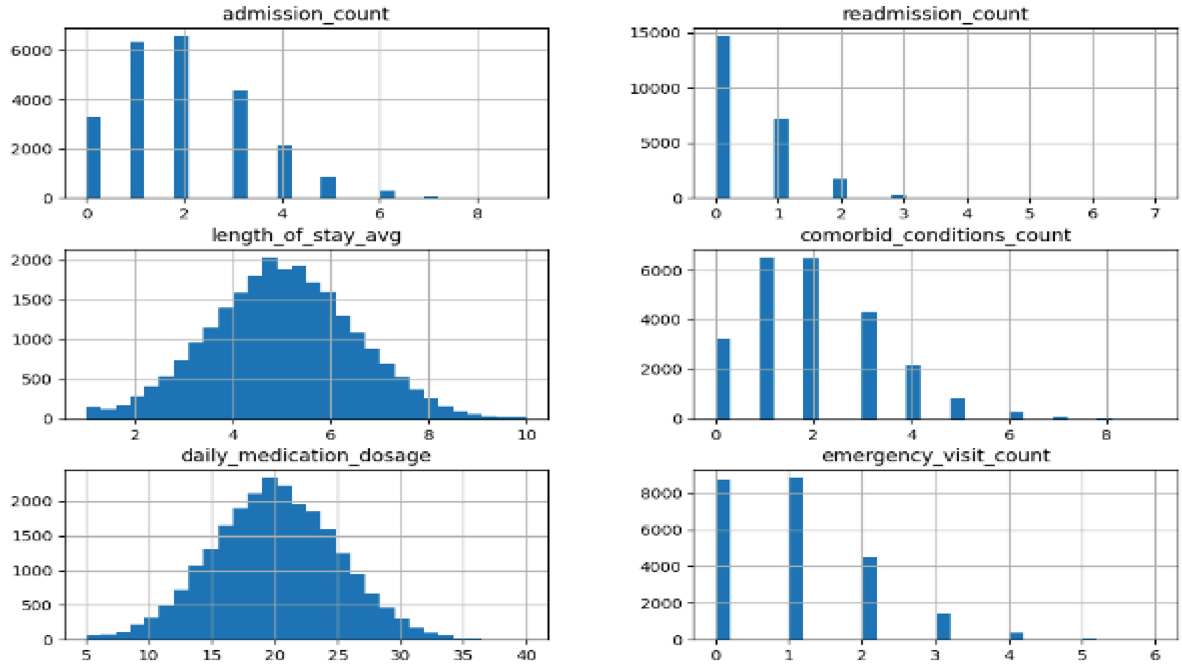

Fig.2 Visualizing the data to see its distributions.

```
df.skew(numeric_only=True)
```

	0
admission_count	0.677736
readmission_count	1.438615
length_of_stay_avg	0.030101
comorbid_conditions_count	0.715130
daily_medication_dosage	-0.001096
emergency_visit_count	0.941482

dtype: float64

Fig.3 Check skewness of data .

Clean the Data

```
df_clean = df.copy()

df_clean.drop_duplicates(inplace=True)
df_clean.reset_index(drop=True, inplace=True)
```

Fill Missing Values Based on Distribution

```
# Normal distribution → use mean()
df_clean['length_of_stay_avg'] = df_clean['length_of_stay_avg'].fillna(df_clean['length_of_stay_avg'].mean())
df_clean['daily_medication_dosage'] = df_clean['daily_medication_dosage'].fillna(df_clean['daily_medication_dosage'].mean())

# Skewed distribution → use median()
df_clean['admission_count'] = df_clean['admission_count'].fillna(df_clean['admission_count'].median())
df_clean['readmission_count'] = df_clean['readmission_count'].fillna(df_clean['readmission_count'].median())
df_clean['comorbid_conditions_count'] = df_clean['comorbid_conditions_count'].fillna(df_clean['comorbid_conditions_count'].median())
df_clean['emergency_visit_count'] = df_clean['emergency_visit_count'].fillna(df_clean['emergency_visit_count'].median())

# Categorical → use mode()
for col in df_clean.select_dtypes(include=['object']).columns:
    df_clean[col] = df_clean[col].fillna(df_clean[col]. mode()[0])
```

Fig.4 Clean the data and fill miss values .

Convert Date Column and Extract Date Features

```
df_clean['admission_date'] = pd.to_datetime(df_clean['admission_date'])
```

```
df_clean['admission_date_only'] = df_clean['admission_date'].dt.date
df_clean['admission_time'] = df_clean['admission_date'].dt.time
```

```
df_clean['year'] = df_clean['admission_date'].dt. year
df_clean['month'] = df_clean['admission_date'].dt.month
df_clean['day'] = df_clean['admission_date'].dt.day
df_clean['dayofweek'] = df_clean['admission_date'].dt.dayofweek
```

Fig.5 Identify new features .

Methods

Pre-Processing Techniques Used:

Loading the Dataset: This Operation made by downloading the dataset with api provided in kaggle in colab not in our local device. And importing the related libraries like kagglehub. And then importing python libraries like pandas which help us in cleaning later. (Figure 1)

Understanding the Dataset: This was made by knowing the features . Understand the columns and what each column represents which will be needed to make the dashboard .

Dataset Cleaning: In this step we first see if there are any null values in the data. In our data there is no null value . So the next step we take is to go and remove duplicates if they exist . First we remove duplicates with little code .Then we implement code that visualizes the distribution of each column (Figure 2) to determine which way to fill the null data that occurs after removing the duplicates. If the data look like bell shape we fill null values with the mean of that column otherwise when it looks skewed we fill null values with the median of the values in the column. The next step is to make sure that all text operations will be safe by making sure that little things like spaces will not affect the operations.

Role & objectives :

Role: Hospital operational manager

Objectives:

1-Monitor hospital load .

2-Establish healthcare programs for high risk patient segments.

3-Asses operational efficiency of the hospital.

4-Monitor competitor hospitals .

5-Track which diseases have the biggest impact on patients.

6-Increase hospital resources such as beds , staff and medical equipment during the high workload and emergency visits periods for efficient utilization.

Dashboard :

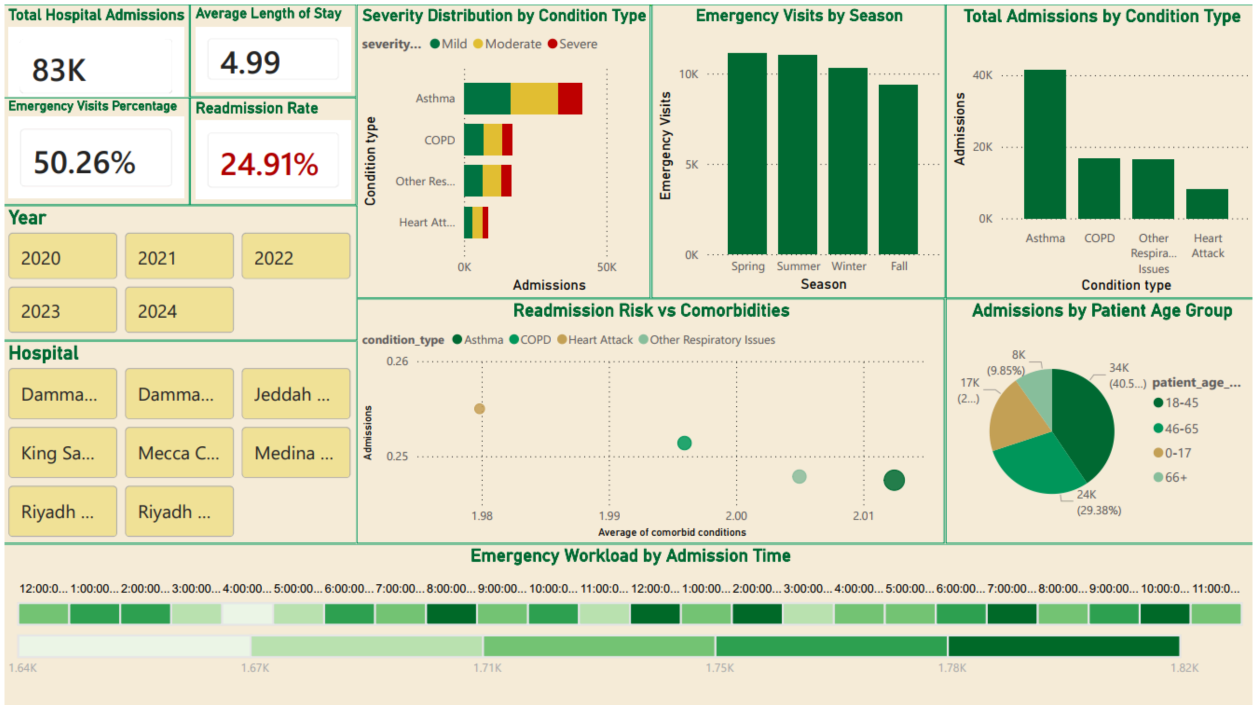

Fig.6 The interactive dashboard.

Explanation of the insights :

1- KPIs cards:

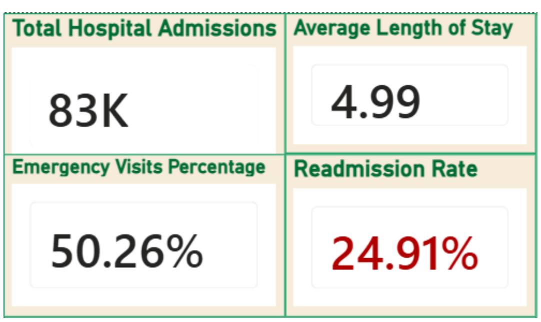

Fig.7 KPIs cards.

-Total admissions that have been done during a selected year and in a specific hospital.
-Average time that the patient spends in a single admission.
-Emergency visits ratio all over the visits.
-The rate of patients that have readmissioned in the hospital.

2-Severity Distribution by Condition Type

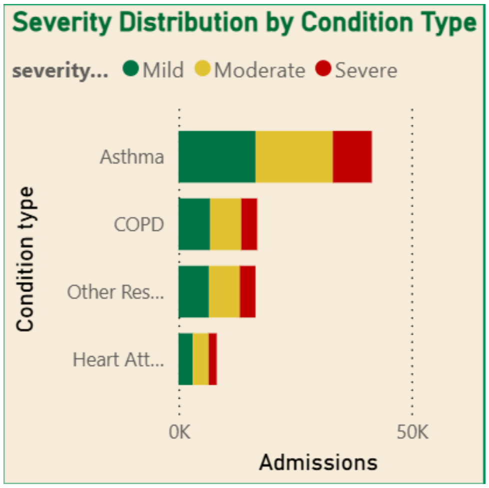

Fig.8 Stacked bar chart

-Explains the number of admissions at each disease and for each disease the portions of severity level.

Example: in 2024 , Asthma was the highest disease that has admissions (as shown in fig.8), inside the cases of asthma a big portion of the cases was moderate cases.

Insights: which disease has the most admissions ?

-For each disease , how much is the portion of severe , moderate and mild cases ?

Benefits: increase critical resources in the hospital for the severe cases , increase the staff that is related to the disease that has the most admissions.

3-Emergency visits by season

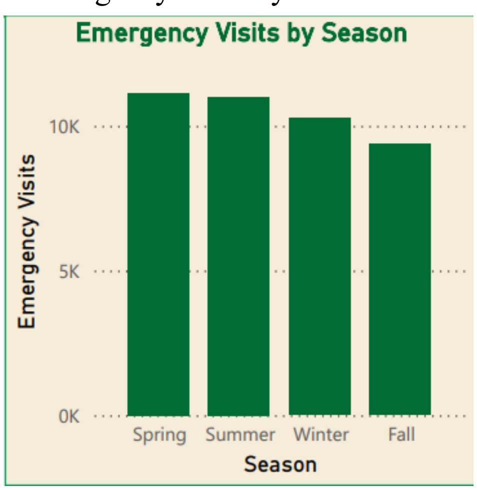

Fig.9 column chart

-Explain the number of emergency visits at each season for a specific year or hospital .

Example: in 2024 , Summer was the most season with emergency visits (as shown in fig.9) .

Insights: which season has the most emergency visits ?

Benefits: increase staff and operational resources at the workload seasons.

4-Admissions by patient Age group

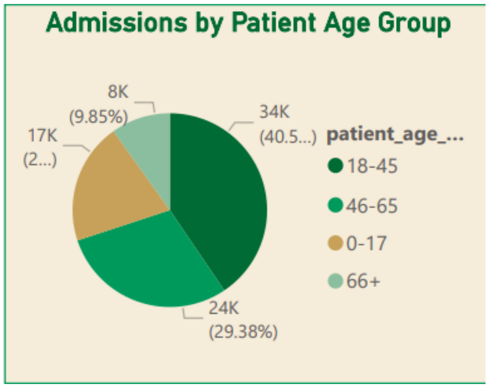

Fig.10 pie chart

-Explains the number of admissions within each age group.
Example: in 2022 , (18-45) group was the group with the most admissions (as shown in fig.10).
Insights: Determine the age profiles of the patients .
Benefits: Create healthcare programs and monitoring programs for the most age profile.
-Provides facilities that suit each age group such as entertainment tools for teens or wheelchairs for old.

5-Readmission Risk vs Comorbidities

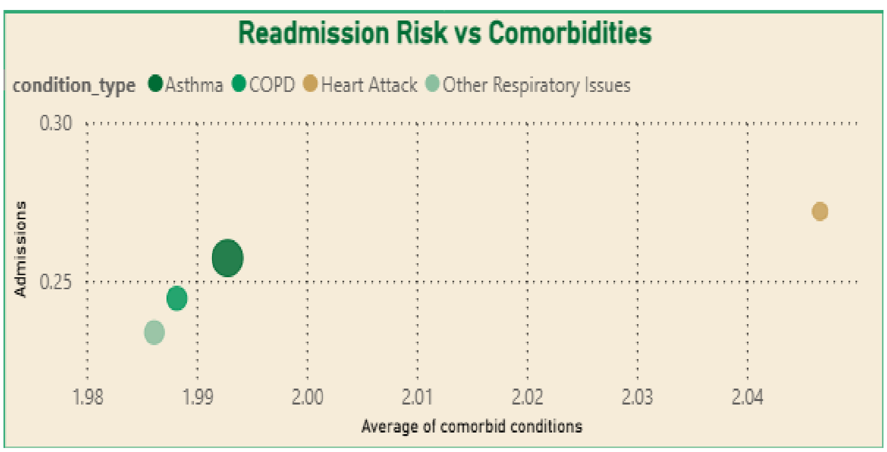

Fig.11 Scatter plot

-Each node representing a disease has two values (x,y) x is the readmission rate and y is the average of the comorbid disease that the patient has with the disease.
Example: Patients with heart attack have the most comorbid diseases and most readmission rate (as shown in fig.11).
Insights: which disease has the most comorbid disease ?
-which disease falls at the High-Risk segment?

-which disease that its patients come back to hospital more often.
Benefits: Make special care management and monitoring with the patients at the High-Risk segment.

6.Emergency Workload Admission Time

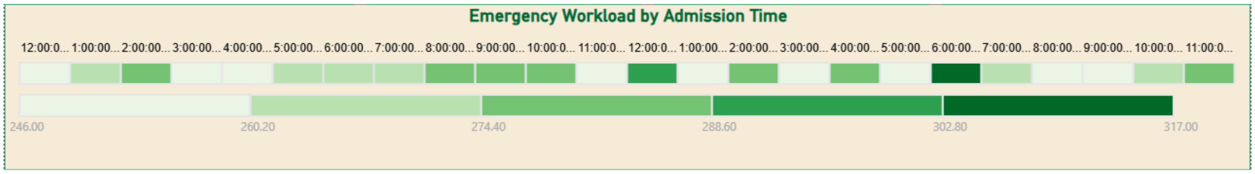

Fig.12 Heatmap plot

-Describes the number of emergency visits at each hour of the day .
Example: 6:00 PM is the most hour that has emergency visits.(as shown in fig.12).
Insights: what hour of the day is a workload ?
Benefits: Increases doctors and nurses at peak hours .

7.Total admissions condition type

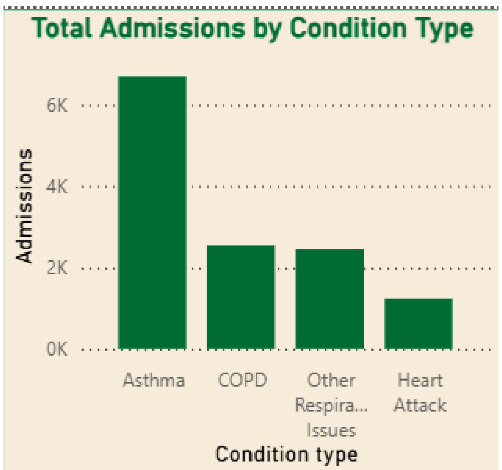

Fig.13 column charts

-Describes the number of admissions for each disease .
Example: Asthma has the most admissions.(as shown in fig.13).
Insights: what diseases require more caring and monitoring ?
Benefits: Identifies the most disease that will affect the resources type and staff .

8. Slicers

Fig.14 Slicer

In some small words, slicers are on-canvas-visual-filters that allow the user to interact with the dashboard . In this figure there are two slicers, one to select the year and one to select the city . Both simulate the way that button works , with only clicking you filter all figures with what you click.

CONCLUSION

After all of that talk, we reach some ratios such as that the number of visits is 83,000 , with an average stay of 4.99 days . The rate of readmission is 24.91% . Half of visits come from the emergency section . Asthma and COPD could be considered as a main reason for visits .
And the most visits come in summer and spring . The largest Admitted age group is between 46 and 65. The busiest period is mid-afternoon to late evening hours. From that we can suggest to improve the speed of readmission procedures , follow up with the patient after leaving the

hospital and distribute the work so that the greatest amount of attention is given during the peak hours . This will improve the work and balance the workload .

REFERENCES

1. Alzeer, N. (2024). *Saudi Arabia hospital admissions (2020–2024)* [Data set]. Kaggle. https://www.kaggle.com/datasets/nawafalzeer/saudi-arabia-hospital-admissions-2020-2024