

About data:

The dataset contains information about 318438 patients who were admitted to a hospital between 2012 and 2017. The dataset also includes the details of the hospital, such as the hospital code, the hospital type, the city code, the region code, and the number of available extra rooms. The dataset has a target variable called Stay, which indicates the length of stay of the patient in the hospital in days.

The dataset has several features that describe the patient's demographics, such as age, and city code. Some of these features have missing values, which need to be imputed or handled appropriately. The dataset also has some anonymized variables, such as Admission_Deposit, which are numerical and have different ranges and distributions.

The dataset has some features that describe the patient's medical history, such as the type of admission, the severity of illness, and the number of visitors. These features can be used to measure the patient's condition and needs.

The dataset has some features that describe the patient's department and ward details, such as the department code, the ward type, the ward facility code, and the bed grade. These features can be used to measure the patient's treatment and comfort.

Cleaning Data:

This is how i worked with data in Power Bi:

Explore data via Query Editor

1. I removed the missing data in the column Bed Grade,City code of patient
2. Column 'Stay' and column 'Age' display ranges, but some rows contain dates instead of ranges. I replaced these dates with the intended ranges; for example, replacing '11-20' instead of '20- Nov'
3. I also adjusted some data types that were not compatible with the columns. I changed the 'Stay' and 'Age' columns to text, as well as the city codes and hospital codes.

This is how i worked with data in Python(Additional steps):

Explore data via (data.head())to get a sample of data ,data.info to get a brief on my data ,data.isnull().sum() to check Nulls

1. I handled missing data by taking the median for the "Bed Grade" column and mode value for the "City Patient Code" column.

```
#Handling Missing values for Bed Grad
```

```
# Impute with the median
```

```
data['Bed Grade'].fillna(data['Bed Grade'].median(), inplace=True)
```

```
#Handling Missing values for city code patient
```

```
# Impute with the mode
```

```
data['City_Code_Patient'].fillna(data['City_Code_Patient'].mode()[0],  
inplace=True)
```

2. Column 'Stay' and column 'Age' display ranges, but some rows contain dates instead of ranges. I replaced these dates with the intended ranges; for example, replacing '11-20' instead of '20- Nov'

```
data['Stay'] = data['Stay'].apply(lambda x: x.replace('20-Nov', '11-20') if  
'20-Nov' in x else x)
```

```
data['Age'] = data['Age'].apply(lambda x: x.replace('20-Nov', '11-20') if '20-Nov'  
in x else x)
```

```
data['Stay'] = data['Stay'].apply(lambda x: x.replace('More than 100 Days',  
'>100') if 'More than 100 Days' in x else x)
```

Data Exploration:

1. Factors Influencing Length of Stay:

- Objective: Understand the various factors that contribute to the length of stay for patients in the hospital.
- Approach: Explore relationships between the length of stay and different variables present in the dataset.

2. Effect of Hospital Type and Region:

- Objective: Investigate how hospital type and region impact both the length of stay and the availability of extra rooms.
- Approach: Analyze the correlation between hospital characteristics (type and region) and patient outcomes.

3. Impact of Patient Demographics and Medical History:

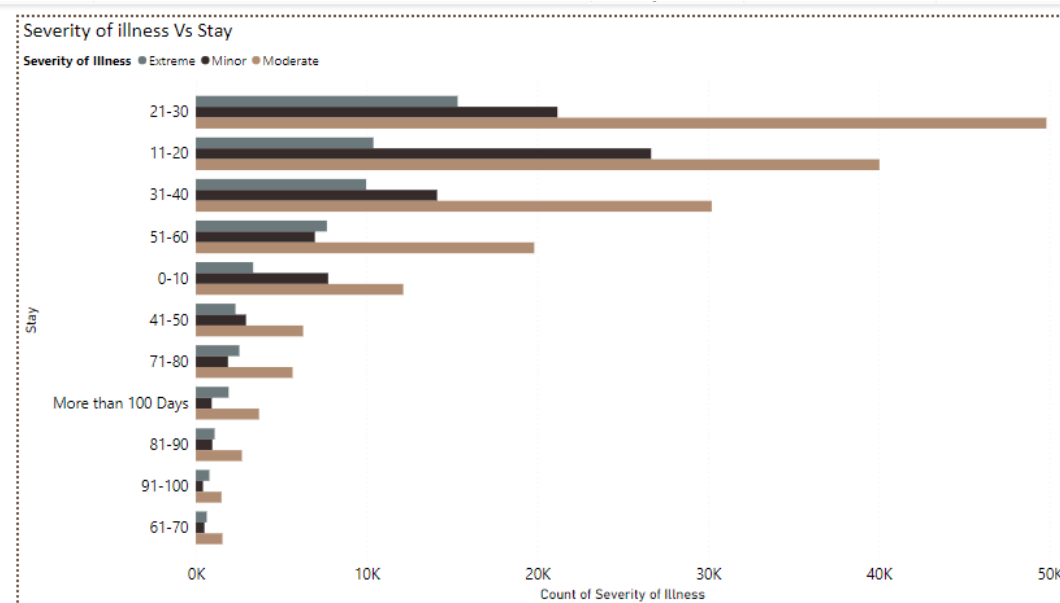
- Objective: Examine the influence of patient demographics and medical history on the length of stay and admission deposit.
- Approach: Identify patterns and relationships between individual patient characteristics and their hospital experience.

Data Visualization:

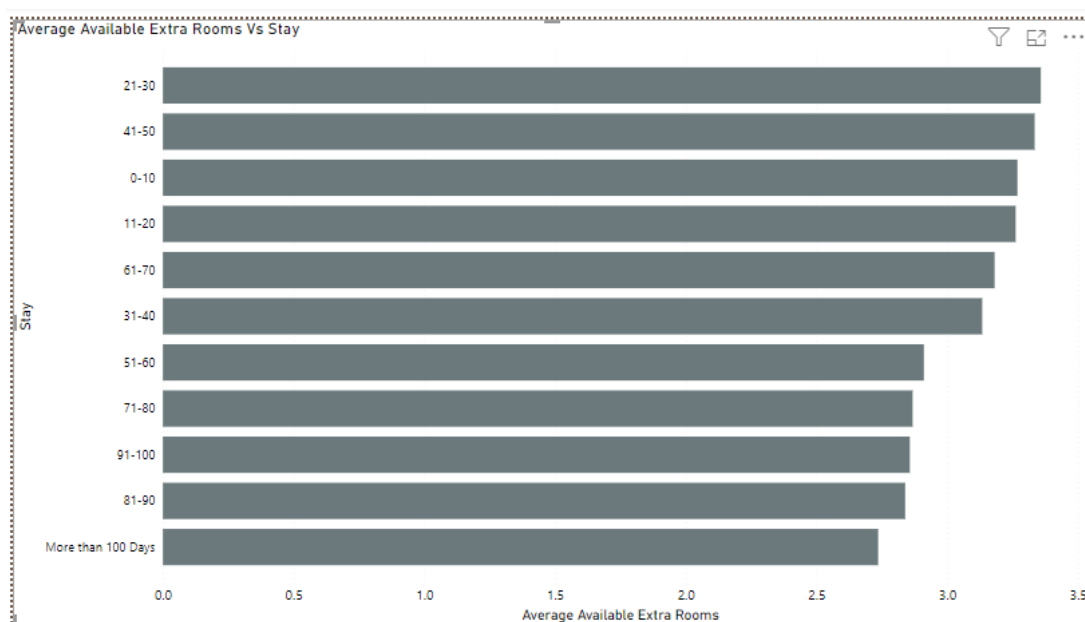
Working with Power Bi

1. Factors Influencing Length of Stay:

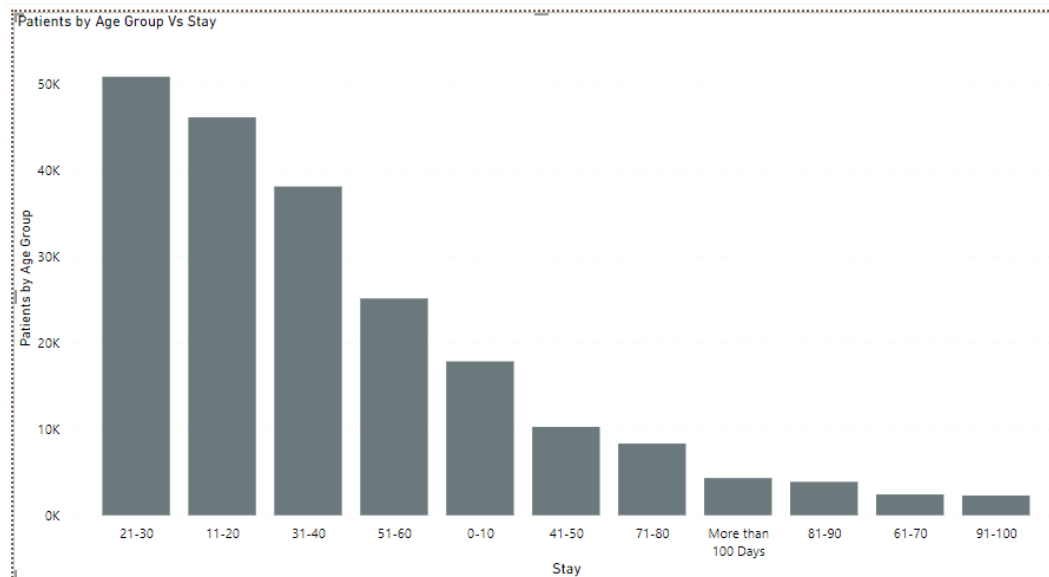
Severity of illness



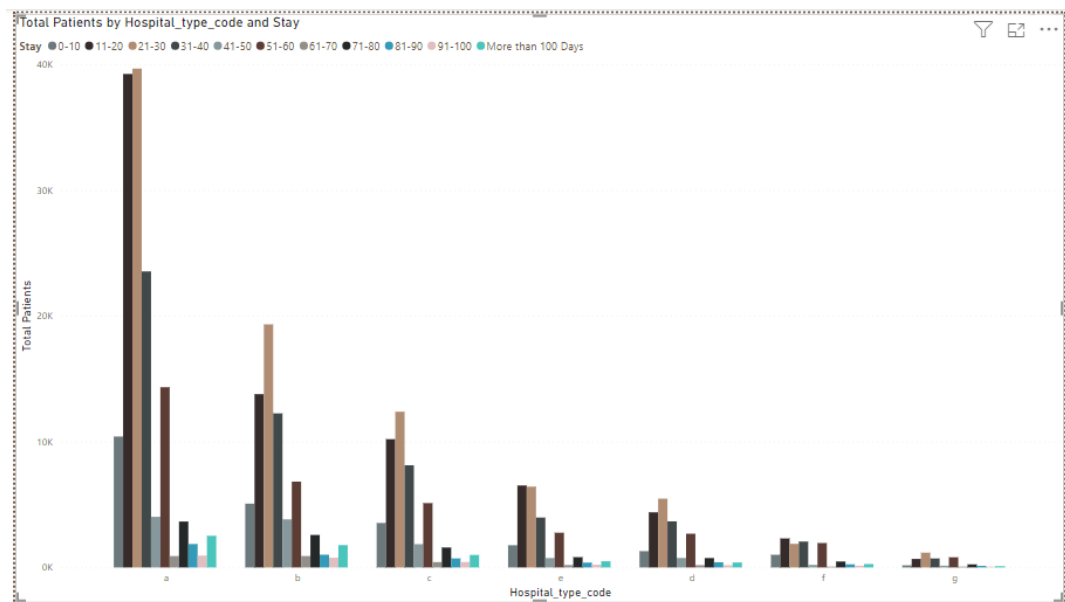
Calculated Measure Average of available extra rooms Vs Stay



Calculated measure patient by age group Vs Stay

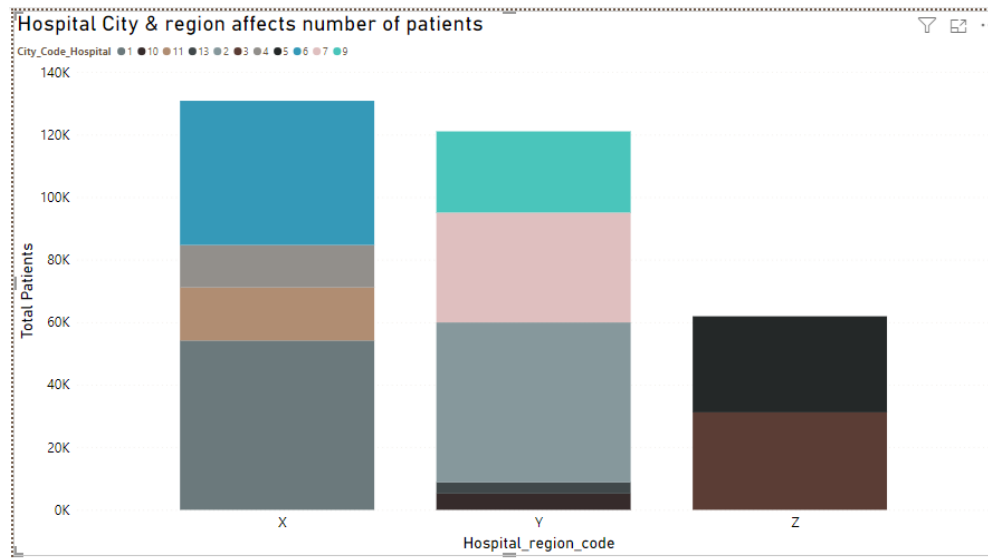


Calculated measure Total Patients ,Hospital type , Stay

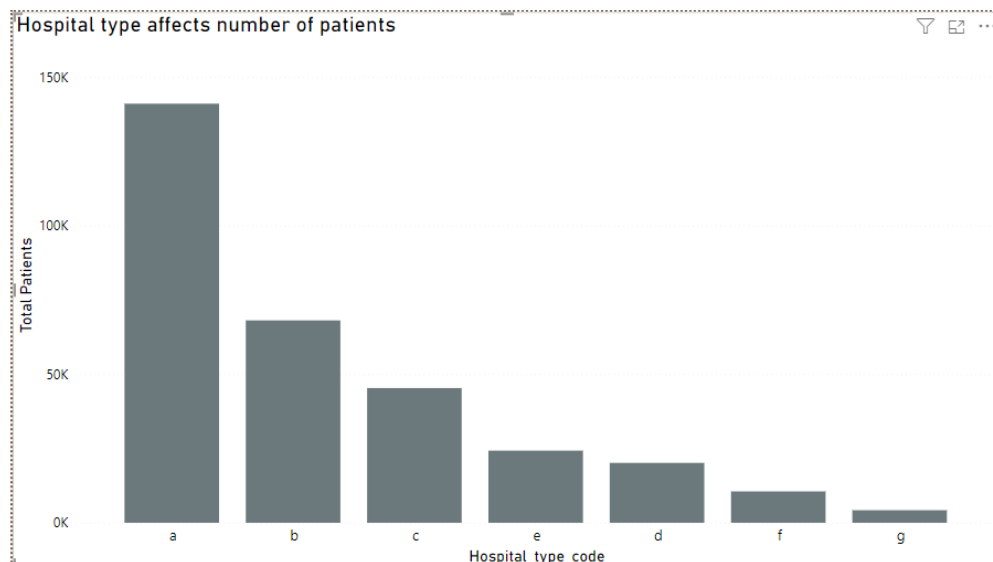


2. Effect of Hospital Type and Region:

Hospital City and region affects number of patients



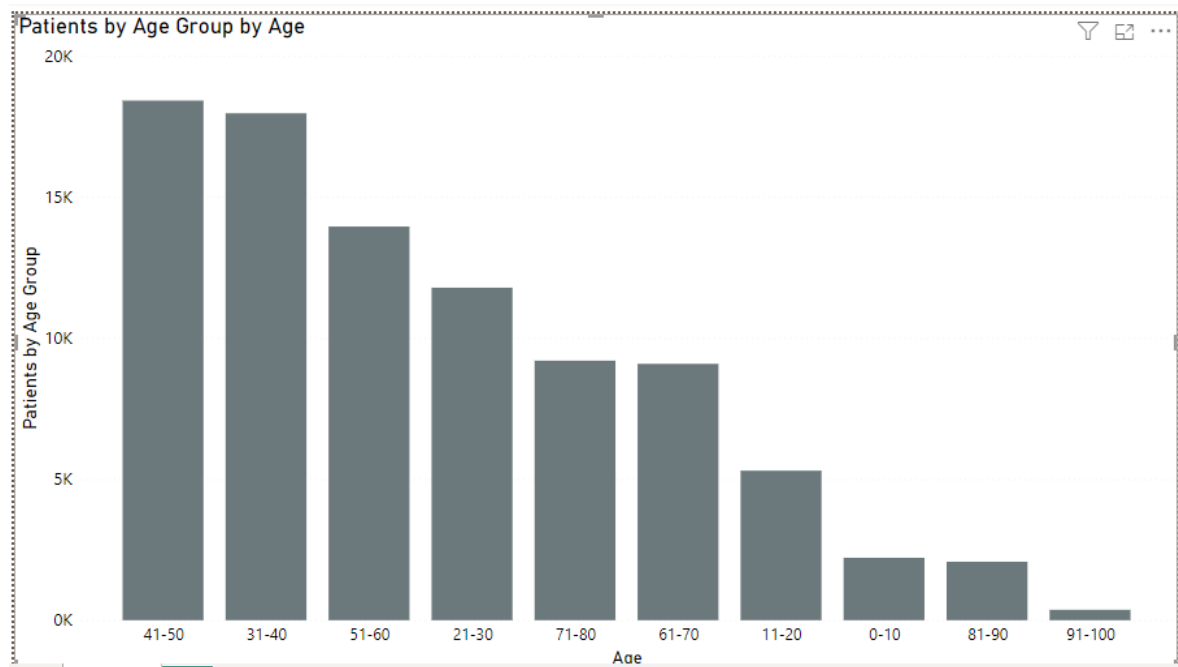
Also Hospital type affects number of patients



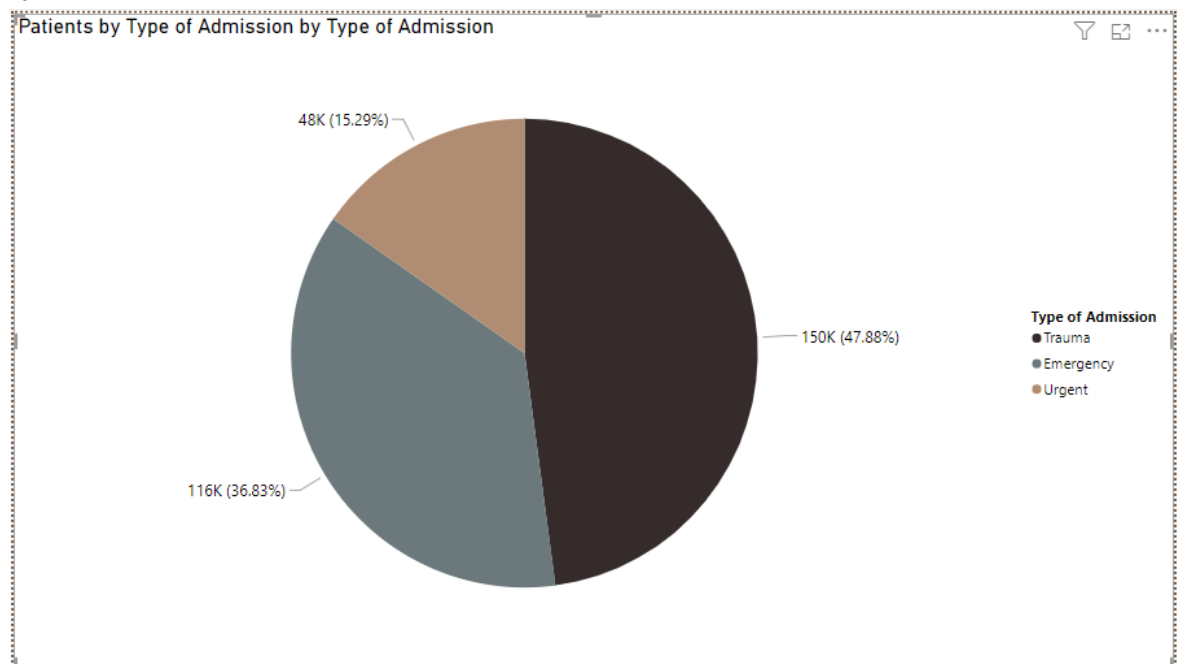
i mentioned the relation between Hospital type and length of stay And the relation between number of Extra rooms above.

3. Impact of Patient Demographics and Medical History

Number of patients per each Age period



Type of admission



Tools:

Power Bi to visualize data