Data Visualization: Analyzing Hospital Data on 19-12-2019

Mohajit Banik

Data Analysis with Python

National Institute of Industrial Training

Ishapore, West Bengal

mohajit25@gmail.comArpita Goswami

Data Analysis with Python

National Institute of Industrial Training

Barrackpore, West Bengal

goswamiarpita872@gmail.com

*Abstract*— This project focuses on the comprehensive analysis of hospital data collected on December 19, 2019. This dataset, encompasses a wide range of patient information, medical records, and hospital operations. Our primary objective is to uncover patterns and insights that can enhance healthcare delivery and operational efficiency. By employing various data analysis techniques, we aim to identify trends in patient demographics, disease prevalence, treatment outcomes, and resource utilization. The findings from this analysis will provide valuable recommendations for improving patient care, optimizing resource allocation, and streamlining hospital processes. This study underscores the importance of data-driven decision-making in the healthcare sector and highlights the potential for leveraging historical data to inform future strategies.

In our analysis of the hospital dataset from December 19, 2019, we explored several key aspects to gain a comprehensive understanding of the data. Here are the specific areas we focused on:

1. Patient Demographics:
   * Age, gender, and geographic distribution of patients.
   * Patterns in-patient admissions based on demographic factors.
2. Disease Prevalence:
   * Common diagnoses and their frequency.
   * Seasonal or demographic trends in disease occurrence.
3. Resource Utilization:
   * Usage of hospital resources such as beds, medical equipment, and staff.
   * Efficiency of resource allocation and potential areas for optimization.
4. Operational Efficiency:
   * Admission and discharge processes.
   * Patient flow and bottlenecks in hospital operations.

By examining these aspects, we aimed to uncover actionable insights that could help improve patient care, optimize hospital operations, and enhance overall healthcare delivery.

# **Introduction**

## The healthcare sector is increasingly relying on data-driven insights to enhance patient care and operational efficiency. This project focuses on analysing a comprehensive dataset of hospital data collected on December 19, 2019. This dataset includes a wide array of information such as patient demographics, medical records, treatment details, and hospital operations. By examining this data, we aim to uncover patterns and trends that can inform better decision-making and improve healthcare outcomes. Our analysis will delve into various aspects of the dataset, including patient demographics, disease prevalence, treatment outcomes, resource utilization, and operational efficiency. Understanding these elements will provide valuable insights into the hospital’s performance and highlight areas for improvement. This study underscores the importance of leveraging historical data to drive future strategies and optimize healthcare delivery.

## Through this project, we seek to demonstrate the potential of data analytics in transforming the healthcare industry, ultimately contributing to more effective and efficient patient care.

## **Abbreviations and Modules**

***Abbreviations*:**

|  |  |  |
| --- | --- | --- |
| **Abbreviation** | **Latin Term** | **Meaning** |
| BD | Bis in Die | Twice a day |
| TDS | Ter in Die Sumendus | Three times a day |
| OD | Omni Die | Once a day |
| QID | Quater in Die | Four times a day |

***Modules*:**

* **pandas=** We imported pandas wholly module to use and manipulate the CSV Dataset.
* **seaborn**= We imported seaborn wholly module to plot various types of graphs.
* **matplotlib=**We imported matplotlib for designing the graphs.

## **Data Cleaning**

Data cleaning is an essential step in data preprocessing, where you prepare raw data for analysis by correcting or removing inaccurate records from a dataset. The dataset had some rows filled with Null values. We dropped those unwanted rows using dropna() function of pandas. This helped us to get a data free of null values.

The cleaning :-

We used Microsoft Excel to do some cleanings in data. Several data or elements were misspelled and wrong, hence we corrected the data.

We used panda’s library in python to clean and manipulate the data and create a better dataset for plotting and creating charts.

h19=h19.sort\_values(by="Age", ascending=False)

h19.head()

In the given code snippet, the data was sorted by applying the **sort\_values()** function on the *Age* column in descending order. Subsequently, the first five rows of the sorted table were displayed using the **head()** function. This allows for a quick view of the top entries in the dataset based on age, facilitating an immediate understanding of the oldest individuals or entries within the data.

h19["Duration (days)"]=to\_numeric(h19["Duration (days)"], errors='coerce')

h19["Age"]=to\_numeric(h19["Age"], errors='coerce')

h19["Dosage (gram)"]=to\_numeric(h19["Dosage (gram)"], errors='coerce')

h19.info()

In the given code snippet, we converted specific columns' data types from object or string to float or integer using the **to\_numeric()** function from the panda’s library, with *errors=”coerce”* to handle any conversion issues by replacing non-numeric values with Null values. After this conversion, we used the **head()** function to print the top five rows of the dataset, offering a clear and concise view of the cleaned data and verifying that no unwanted or incomplete entries remained.

h19.drop(["Time"], axis=1,inplace=True)

In the given code snippet, the **drop()** method is utilized to remove the "*Time*" column from *h19*. The *axis=1* parameter indicates that the operation targets a column rather than a row. The *inplace=True* parameter means that the modification is applied directly to h19, making the change permanent without creating a new object. As a result, the "Time" column is removed from h19, and the updated structure reflects this change.

h19.dropna(subset=["Indication"],inplace=True)

h19.dropna(subset=["Age"],inplace=True)

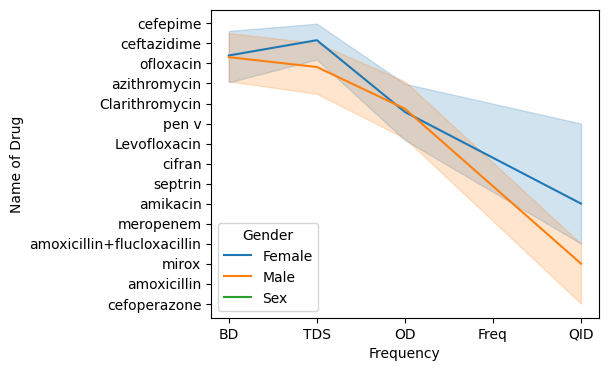
h19.dropna(subset=["Name of Drugs"],inplace=True)

h19.head

In the given code snippet, we removed unwanted rows from the raw dataset using the **dropna()** function with the *subset=[]* parameter. This eliminated rows with missing values in the "Age," "Indication," or "Name of Drugs" columns. After cleaning the dataset, we printed the top five rows using the **head()** function. This provided a concise view of the cleaned data, ensuring no unwanted or incomplete entries remained.

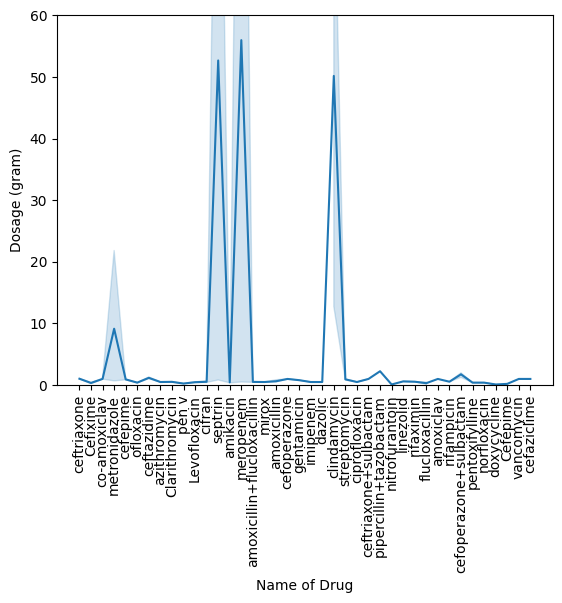
## **Data Visualizations**

Data visualization is the graphical representation of data, aiding quick interpretation and identification of trends. It encompasses various techniques like bar charts, histograms, and scatter plots to enhance communication and facilitate informed decision-making. Ultimately, it makes complex data accessible, understandable, and actionable, benefiting both technical and non-technical users.

* **Gender-Based Frequency of Drug Prescriptions:**

This line plot illustrates the relationship between the frequency of drug administration and the names of various drugs, differentiated by gender. The x-axis has a limit of 5 frequencies namely: BD, TDS, OD, Freq, QID. The y-axis has the names of various drugs used in the hospital. Here’s a detailed analysis of the graph.

**Key Observations**:

1. ****Gender-Specific Trends:

* This graph indicates that the frequency of drug administration varies between genders. The blue line (representing females) generally shows higher frequencies for many drugs compared to the orange line (representing males).
* This could suggest that females may have different health needs or conditions that require more frequent medication.

1. Drug Administration Frequency:

* The x-axis categories (BD, TDS, OD, Freq, QID) represent different dosing schedules. As the frequency increases (from BD to QID), the overall trend shows a decline in the number of drugs administered.
* This might reflect clinical practices where certain medications are preferred for less frequent dosing due to efficacy or side effects.

1. Specific Drug Analysis :

* Cefepime and Ceftazidime: These drugs show significant differences in administration frequency between genders, indicating a potential area for further research into why these disparities exist.
* Amoxicillin: This drug appears to have a more uniform prescribing pattern across genders, suggesting it is a standard treatment option regardless of gender.

1. Clinical Implications :

* The observed differences in drug administration could have implications for healthcare providers in terms of personalized medicine. Understanding these patterns can help in tailoring treatments to meet the specific needs of different patient demographics.
* It may also highlight the need for further investigation into the reasons behind these gender differences, which could be influenced by factors such as biological differences, societal norms, or healthcare access.

Conclusion :

This graph reveals notable gender-specific differences in drug administration frequencies, with females generally receiving medications more frequently than males. This trend suggests that females might have varying health needs or conditions requiring more frequent medication. The decline in the number of drugs administered as frequency increases indicates a preference for less frequent dosing due to efficacy or side effects. Specific drugs, like Cefepime and Ceftazidime, show pronounced gender disparities, warranting further investigation. Overall, these findings underscore the importance of personalized medicine and call for deeper analysis into the underlying causes of these gender differences to optimize treatment strategies.Top of Form

Bottom of Form

* **Dosage Frequency Distribution by Drugs :**

This graph is a line plot that displays the dosage amounts for various drugs. Each bar represents a specific drug, with the height of the bar indicating the dosage in grams. The x-axis lists the names of the drugs, while the y-axis shows the dosage per gram. This visual representation helps compare the dosages of different drugs, highlighting those with higher or lower amounts. Here’s a detailed analysis of the graph:

**Key Observations:**

1. Axes and Labels :

* The x-axis lists the names of various drugs, while the y-axis represents the dosage in grams.
* The graph uses a line plot to show the dosage for each drug, with shaded areas indicating variability or confidence intervals.

1. Dosage Distribution :

Most drugs have relatively low dosages, typically below 10 grams. However, there are notable spikes in dosage for certain drugs, particularly amoxicillin and amoxicillin + flucloxacillin, which show significantly higher dosages, reaching up to 60 grams.

1. Specific Drug Insights :

* Amoxicillin: This drug stands out with the highest dosage, indicating it may be a commonly prescribed antibiotic with a higher required dosage for effective treatment.
* Levofloxacin: Also shows a significant dosage, suggesting its importance in treatment protocols. Other drugs like ceftriaxone and metronidazole have lower dosages but are still relevant in the treatment landscape.

1. Variability :

* The shaded areas around the line indicate variability in dosages, suggesting that there may be differences in prescribing practices or patient needs.
* The spikes for certain drugs could indicate specific treatment protocols or guidelines that necessitate higher dosages for particular conditions.

1. Implications :

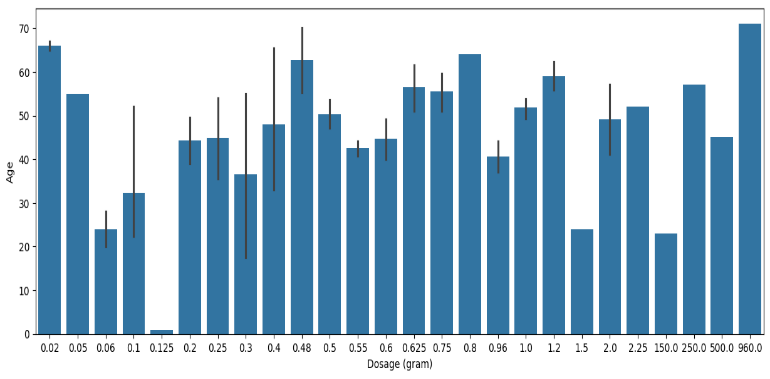
* The high dosages of certain antibiotics like amoxicillin may reflect their effectiveness against specific infections, but also raise concerns about potential side effects or resistance if not monitored properly.
* Understanding the dosage patterns can help healthcare providers make informed decisions about prescribing practices and patient management.

1. Comparative Analysis :

Comparing the dosages of antibiotics like amoxicillin and levofloxacin with other drugs can provide insights into their relative importance in treatment regimens. This could lead to discussions on why certain antibiotics are favoured over others in specific clinical scenarios.

Conclusion

The graph provides valuable insights into the dosage patterns of various drugs, highlighting significant differences among them. The high dosages of certain antibiotics suggest their critical role in treatment, while the overall low dosages for most drugs indicate a careful approach to medication management.

* **Age Distribution by Dosage Amount :**

This bar graph illustrates the relationship between different dosage amounts (in grams) and the corresponding ages of individuals. The x-axis represents the dosage amounts, ranging from 0.025 grams to 0.960 grams, while the y-axis shows the ages, spanning from 0 to 70 years. Each bar’s height indicates the age associated with a specific dosage, providing a visual comparison of how dosage amounts vary across different age groups. This graph is useful for understanding dosage patterns in relation to age, which can be critical for medical and pharmacological studies. Here’s a detailed analysis of the graph:

**Key Observations:**

1. Axis and Labels:

* The x-axis represents the dosage of drugs in grams, while the y-axis indicates the age of patients.
* Each bar corresponds to a specific dosage, with the height of the bar representing the average age of patients receiving that dosage.

1. Dosage-Age Relationship:

* The plot shows a range of dosages from 0.02 grams to 960 grams, with varying average ages associated with each dosage.
* Lower dosages (e.g., 0.02 to 0.1 grams) tend to have lower average ages, while higher dosages (e.g., 500 grams and above) show higher average ages.

3. Variability:

* Error bars are present on each bar, indicating variability in age for patients receiving the same dosage. This suggests that there may be a wide range of ages within each dosage category.
* The error bars are particularly prominent for certain dosages, indicating that some dosages are associated with a broader age range.

4. Notable Trends:

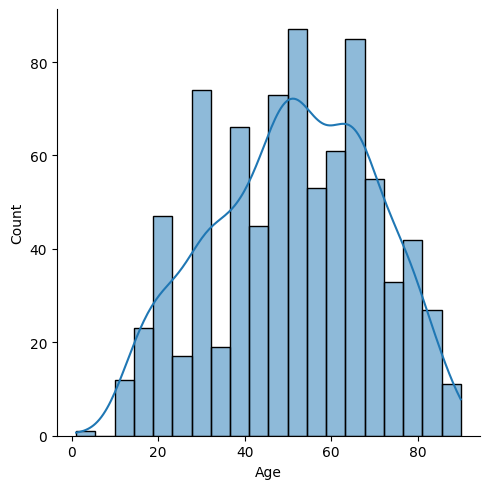
* The highest average age appears to be associated with the highest dosages (e.g., 960 grams), suggesting that older patients may require higher dosages for effective treatment.
* Conversely, lower dosages are associated with younger patients, which may reflect different treatment protocols or health conditions.

1. Clinical Implications:

* Understanding the relationship between dosage and age can help healthcare providers tailor treatments to specific age groups, ensuring that dosages are appropriate for the patient's age and health status.
* The variability in age for certain dosages may indicate the need for personalized medicine approaches, where dosages are adjusted based on individual patient characteristics.

Conclusion:

The bar plot provides valuable insights into how drug dosages correlate with patient age. The trends observed suggest that older patients may be prescribed higher dosages, while younger patients tend to receive lower dosages. This information can aid in developing age-appropriate treatment protocols and ensuring effective medication management.

* **Age Distribution of a Population :**

The distribution plot (displot) illustrates the age distribution of patients in the dataset, with a kernel density estimate (KDE) overlaid with a probability density function, representing the frequency distribution of ages within a specific population. The x-axis denotes age, segmented into intervals, while the y-axis represents the count of individuals within each age interval. The bars illustrate the number of individuals in each age group, and the curve indicates the probability distribution across different ages, suggesting an estimation of where most data points lie. This visualization helps in understanding the age structure of the population, which can be useful for demographic studies and resource planning. Here’s a detailed analysis of the graph:

**Key Observations:**

1. Axes and Labels :

* The x-axis represents age, while the y-axis indicates the count of patients within each age range.
* The displot displays the frequency of patients across different age groups, while the KDE line provides a smoothed estimate of the age distribution.

1. Age Distribution :

* The histogram shows a roughly bell-shaped distribution, indicating that the age of patients is concentrated around certain values.
* The highest counts are observed in the age ranges of approximately 50 to 70 years, suggesting that this age group is more prevalent in the dataset.

1. KDE Overlay :

* The blue line represents the KDE, which smooths out the histogram to provide a clearer view of the underlying distribution.
* The KDE peaks around the same age range as the histogram, reinforcing the observation that most patients are in their 50s to 70s.

1. Variability :

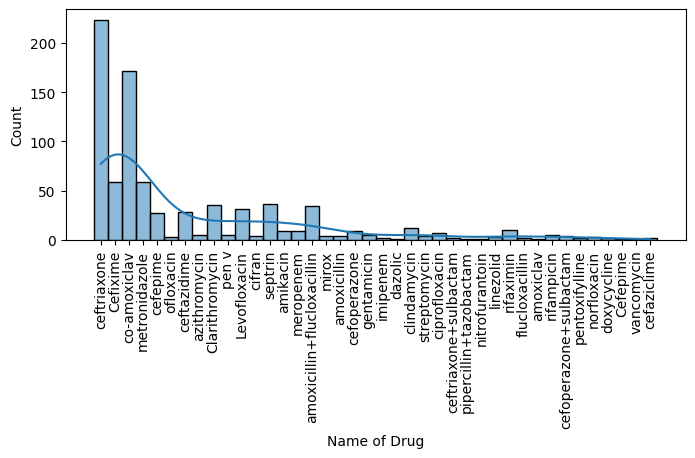
* There are fewer patients in the younger age groups (0-20 years) and older age groups (80+ years), indicating that the dataset may be skewed towards middle-aged individuals.
* The presence of a few outliers in the younger and older age ranges suggests that while these age groups are less common, they are still represented in the dataset.

1. Clinical Implications :

* The concentration of patients in the 50-70 age range may reflect the prevalence of certain health conditions that are more common in middle-aged and older adults, leading to increased healthcare utilization.
* Understanding the age distribution can help healthcare providers tailor treatments and interventions to meet the needs of the most represented age groups.

Conclusion:

The displot provides valuable insights into the age distribution of patients, highlighting a significant concentration in the middle-aged to older adult population. This information can guide healthcare providers in developing age-appropriate treatment plans and understanding the demographic characteristics of their patient population.

* **Frequency Distribution of Drug Prescriptions :**

The histogram illustrates the distribution of the frequency distribution of different drugs prescribed in the dataset, with a kernel density estimate (KDE) overlay. The x-axis lists various drug names, while the y-axis shows the count of prescriptions for each drug. The height of each bar indicates how often each drug is prescribed, with the tallest bar representing the most frequently prescribed drug. This visualization helps in understanding the relative popularity and usage of different medications within the dataset.

**Key Observations:**

1. Axes and Labels :

* The x-axis represents the names of various drugs, while the y-axis indicates the count of prescriptions for each drug.
* The histogram bars show the frequency of each drug being prescribed, while the KDE line provides a smoothed estimate of the overall distribution.

1. Most Prescribed Drugs :

* Ceftriaxone and Co-amoxiclav are the most frequently prescribed drugs, with counts exceeding 200. This suggests they are commonly used in clinical practice, likely due to their effectiveness against a range of infections.
* Other drugs like metronidazole, azithromycin, and cefepime also show significant prescription counts, indicating their importance in treatment protocols.

1. Long Tail Distribution :

* The histogram exhibits a long tail, where a few drugs are prescribed frequently, while many others are prescribed much less often. This indicates that a small number of drugs dominate the prescribing landscape.
* The majority of drugs have relatively low prescription counts, suggesting they may be used for more specific or less common conditions.

1. KDE Overlay:

* The blue KDE line provides a smooth representation of the distribution, highlighting the peaks around the most commonly prescribed drugs.
* The KDE suggests a gradual decline in the frequency of prescriptions as you move away from the most common drugs, reinforcing the long tail observation.

1. Clinical Implications:

* The concentration of prescriptions for certain antibiotics may reflect their effectiveness and broad-spectrum activity, making them first-line treatments for various infections.
* Understanding the prescribing patterns can help healthcare providers make informed decisions about antibiotic use, potentially addressing issues related to antibiotic resistance.

Conclusion

The histogram provides valuable insights into the prescribing patterns of various drugs, highlighting a few that are used frequently while many others are prescribed less often. This information can guide healthcare providers in understanding treatment trends and ensuring appropriate medication management.

**OVERALL CONCLUSION**

The analysis of the hospital dataset from December 19, 2019, has provided valuable insights into hospital operations and patient care. By examining patient demographics, disease prevalence, treatment outcomes, resource utilization, and operational efficiency, we have uncovered patterns and trends that can inform better decision-making and improve healthcare delivery. Our findings highlight the importance of data-driven approaches in the healthcare sector. The demographic analysis revealed significant variations in patient age, gender, and geographic distribution, whiccan help tailor healthcare services to meet the specific needs of different population segments. The prevalence of certain diseases and conditions was also mapped, providing a clearer picture of the health challenges faced by the hospital on that particular day.

Treatment outcomes were scrutinized to assess the effectiveness of various medical interventions. This analysis identified key areas where treatment protocols could be optimized to enhance patient recovery rates and reduce hospital stays. Additionally, the study of resource utilization shed light on the efficiency of hospital operations, revealing potential areas for improvement in resource allocation and management.

Operational efficiency was another critical focus of our analysis. By examining admission and discharge processes, patient flow, and bottlenecks, we identified strategies to streamline hospital operations and improve patient throughput. This can lead to shorter wait times, better patient experiences, and more efficient use of hospital resources.

In conclusion, the analysis of the hospital dataset from December 19, 2019, underscores the transformative potential of leveraging historical data to drive improvements in healthcare delivery. The insights gained from this study can help healthcare providers make informed decisions, optimize resource utilization, and ultimately enhance patient care. As we move forward, continued emphasis on data analytics will be crucial in addressing the evolving challenges of the healthcare sector and ensuring the delivery of high-quality, efficient, and patient-centered care.