Data Cleaning, and Preparation (Basic EDA):

## Operating Room Utilization – 2018 Analysis Report

### Objective

The goal of this analysis was to assess the daily utilization rates of operating rooms throughout 2018, identify patterns and inefficiencies, and recommend areas for operational improvement.

### Data Cleaning and Preprocessing

The raw data presented several challenges and inconsistencies that required careful cleaning and validation before any reliable analysis could be conducted. Below is a summary of the key steps taken:

#### Filtering Real Surgeries

* Rows marked as "No Show" = True were removed, as the patient never underwent surgery.
* We further excluded cases where there was no recorded incision time, closure time, or operating room entry — indicating that no surgery took place, even if the case wasn’t formally marked as "No Show".

#### Verifying Time Fields

* Time-related fields such as Actual Surgery Room Entry Time, Closure Time, and End of Surgery Time (Exit from OR) were often misformatted — showing up as numbers (e.g., 1, 0) instead of HH:MM:SS.
* We explicitly converted all these columns into timedelta format in Python and later formatted them as proper strings for Excel output.

#### Removing Data Artifacts

* Some rows had room numbers such as -999, which were used as placeholders for missing values. These were flagged for special attention during visualization.
* Duplicate or ambiguous rows were filtered based on patient ID and room usage overlap.

#### Calculating Clean Utilization

Utilization rate was calculated per room per day using the formula:

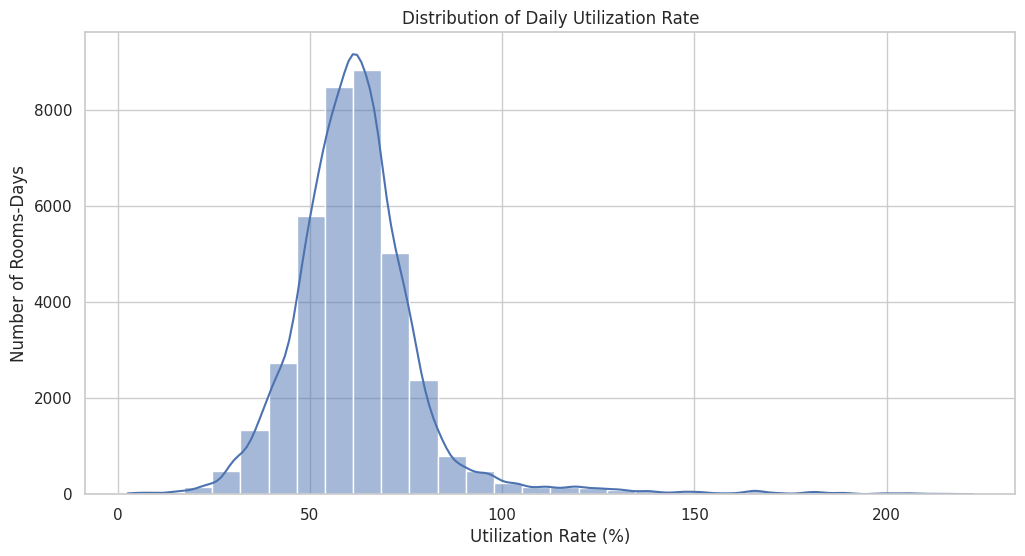
Utilization (%) = (Surgery Time + Setup Time) / Scheduled Shift Time × 100

Only records with complete and valid data were included in this calculation.

### Key Visual Findings

#### Distribution of Daily Utilization

The histogram showed a bell-shaped distribution centered around 60%, suggesting effective room usage on most days. However, a long right tail highlighted multiple days with over-usage, likely due to surgery overruns.



#### 

#### 

#### 

#### 

#### 

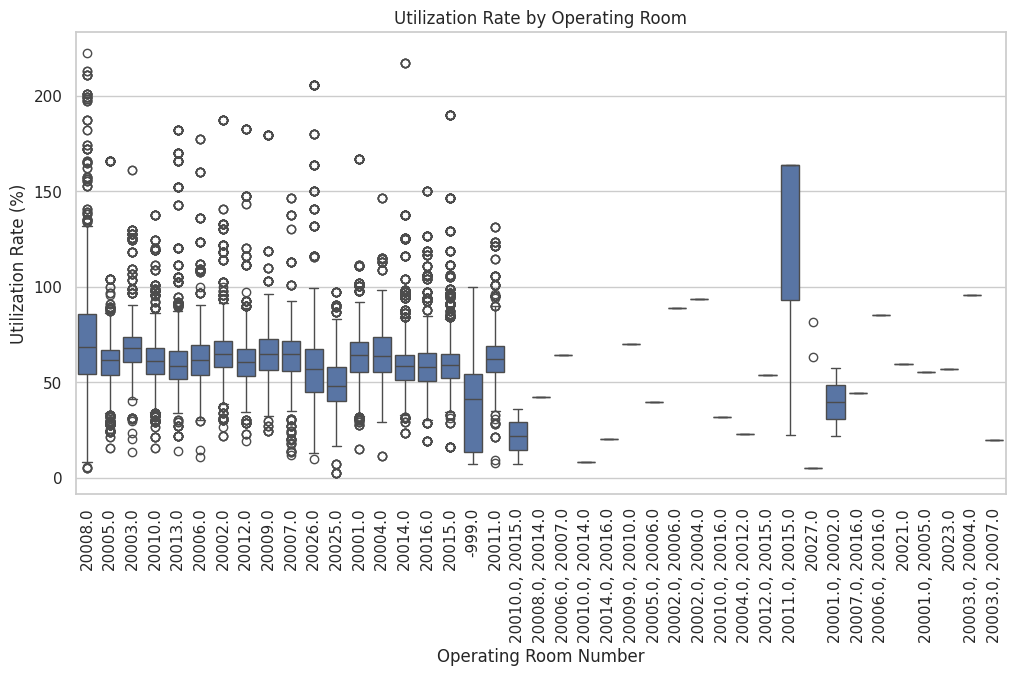
#### 

#### 

#### 

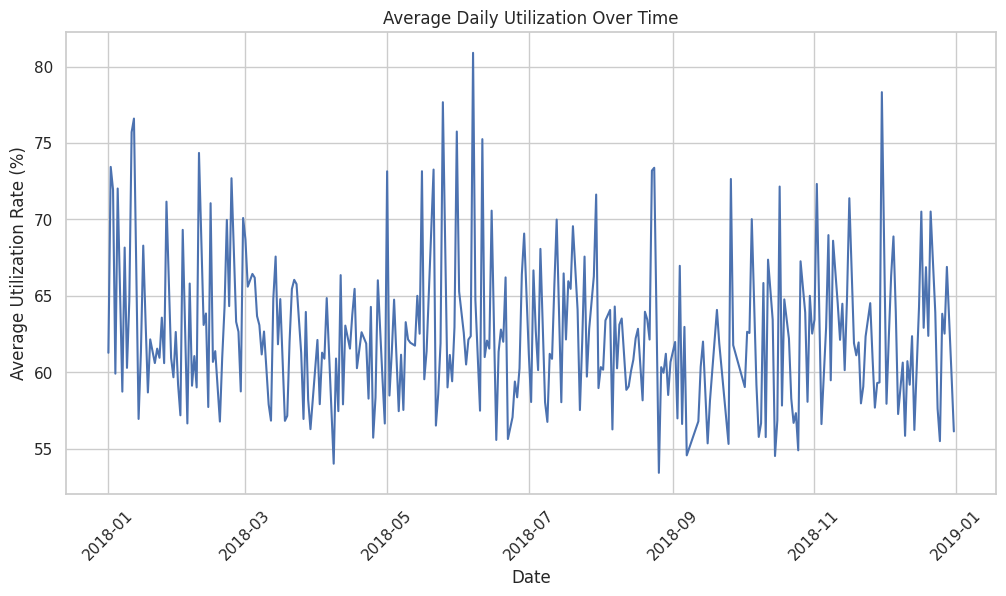
#### Utilization by Room

* Most rooms operate around 60% median usage.
* Some rooms have high variability and frequent overloads.
* Others are underused or inactive.
* Room -999 indicates inconsistent or missing room assignment — requiring cleanup or flagging.



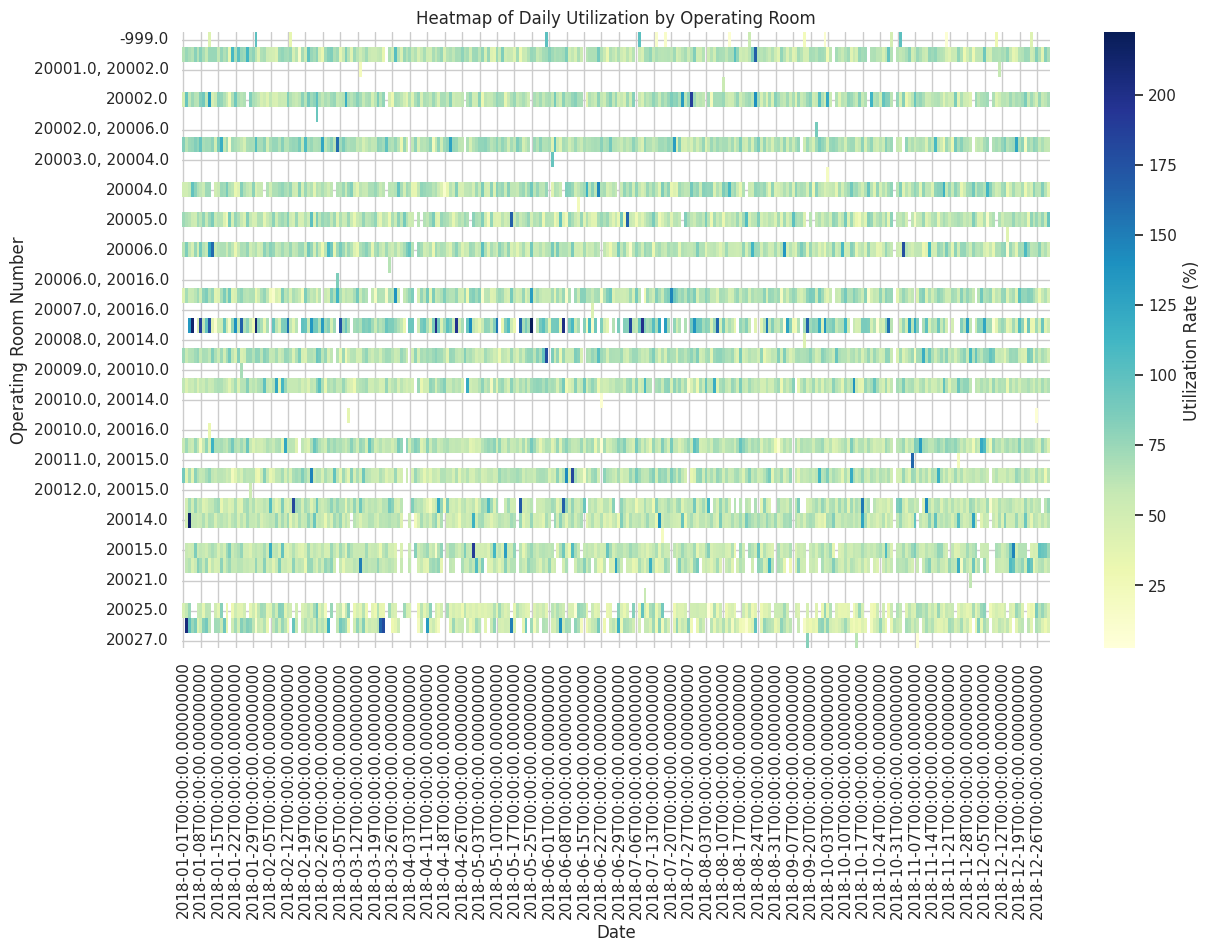
#### Utilization Over Time

The time series plot showed seasonal dips (notably in July–August), and moderate end-of-year volatility. This matches expected hospital dynamics such as vacations and scheduling pressure.



The heatmap visualized room usage over time, revealing:

* Persistent underutilization in certain rooms.
* Clustered over-utilization in others.
* Occasional anomalies suggesting poor time estimates or overbooked schedules.



### Issues Encountered

The time-related columns in the original Excel files were inconsistently formatted, containing a mixture of actual time values, integers, and blank cells. This inconsistency created significant challenges in interpreting and calculating utilization accurately. In several cases, surgeries were documented without complete time stamps, which resulted in utilization rates being incorrectly calculated as 0%, even though the surgery was performed. Additionally, some rows appeared to represent valid surgical cases but, upon closer inspection, lacked any evidence of actual surgical activity.

Another data integrity issue was the occurrence of utilization rates exceeding 100%. While this could occasionally reflect real scenarios such as emergency cases or extended surgeries, in many instances it was due to misaligned timing records — for example, a surgery marked as starting before the previous one had officially ended. This overlap caused double-counting of time and inflated the utilization calculation. After quantifying this discrepancy, we found that the average deviation between recorded surgery times and actual room usage was approximately 11 minutes, across all operating rooms.

All of these issues were resolved through a combination of targeted filtering, validation of key time fields, and the exclusion of incomplete or misleading records from the final analysis.

### Summary

This analysis reveals a mostly efficient use of operating room resources in 2018. However, several challenges were identified that warrant attention. Some operating rooms were consistently over-utilized, raising concerns about staff burnout and potential delays in surgical schedules. At the same time, certain rooms remained underutilized, highlighting opportunities for improved resource distribution and scheduling efficiency. Additionally, data-related issues—such as missing time entries and invalid room numbers—were encountered and should be addressed at the source to ensure greater accuracy in future analyses. Based on these findings, we recommend implementing stricter validation protocols during scheduling and data entry, thoroughly investigating outliers and unusually high-utilization days to uncover underlying causes, and considering a more balanced allocation of operating rooms to enhance overall performance and fairness.