PREDICTION OF LENGTH OF STAY (LOS) IN POST-SURGICAL PATIENT

Kusumahartani Dwi Estikasari

OVERVIEW

Definition

Length of stay (LOS) in post-surgical patient

The period of time a patient spends in the hospital after undergoing a surgical procedure.

Issues

- Post-surgery patients tend to have a longer LOS. It is necessary to identify risk factors associated with LOS.
- Planning is required to ensure that existing hospital's resources are utilized optimally.

Objectives

- To assist hospitals in more **efficiently planning resources** such as wards, beds, medical personnel, and equipment, ultimately impacting **cost efficiency**.
- To assist hospitals identifying patients who may require additional care or preventive measures, thereby **enhancing patient satisfaction** and the **hospital's reputation**.

BUSINESS UNDERSTANDING

Identify business objectives

Optimizing hospital resource planning and management (LOS, room and bed availability, healthcare resources, and equipment)

Define projects goals

Analyzing the factors influencing LOS and predicting LOS in post-surgical patients.

Understand business process

Understanding of post-surgical patient care steps from registration to discharge. Collaboration with the medical team for data collection.

Align with business strategy

Improving operational efficiency, reducing treatment costs, enhancing patient satisfaction, and bolstering hospital's reputation

Define success criteria

Increased precision of LOS predictions

Reduction in operational costs

Improvement in patient satisfaction

SMART

SPECIFIC

MEASUREABLE

ATTAINABLE

RELEVANT

TIME-BASED

The Project

THE DATA

Resource

https://mover.ics.uci.edu/index.html

Content

Patient information of those who underwent surgery at The University California, Irvine Medical Center from 2017-2023

Rows and columns

```
df.shape
```

```
(65728, 23)
```

Data's information

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 64364 entries, 0 to 65727
Data columns (total 23 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   LOG_ID                64364 non-null  object
 1   MRN                   64364 non-null  object
 2   DISCH_DISP_C          64357 non-null  float64
 3   DISCH_DISP            64357 non-null  object
 4   HOSP_ADMSN_TIME       64364 non-null  object
 5   HOSP_DISCH_TIME       64350 non-null  object
 6   LOS                   64350 non-null  float64
 7   ICU_ADMIN_FLAG        64364 non-null  object
 8   SURGERY_DATE          64364 non-null  object
 9   BIRTH_DATE            64364 non-null  int64
 10  HEIGHT                51638 non-null  object
 11  WEIGHT                62001 non-null  float64
 12  SEX                   64364 non-null  object
 13  PRIMARY_ANES_TYPE_NM  64328 non-null  object
 14  ASA_RATING_C          57553 non-null  float64
 15  ASA_RATING            57553 non-null  object
 16  PATIENT_CLASS_GROUP   64364 non-null  object
 17  PATIENT_CLASS_NM      64364 non-null  object
 18  PRIMARY_PROCEDURE_NM  64358 non-null  object
 19  IN_OR_DTTM            57961 non-null  object
 20  OUT_OR_DTTM           57895 non-null  object
 21  AN_START_DATETIME     57051 non-null  object
 22  AN_STOP_DATETIME      57037 non-null  object
dtypes: float64(4), int64(1), object(18)
```

DATA CLEANING

Rows and columns

```
(46316, 13)
```

Data's information

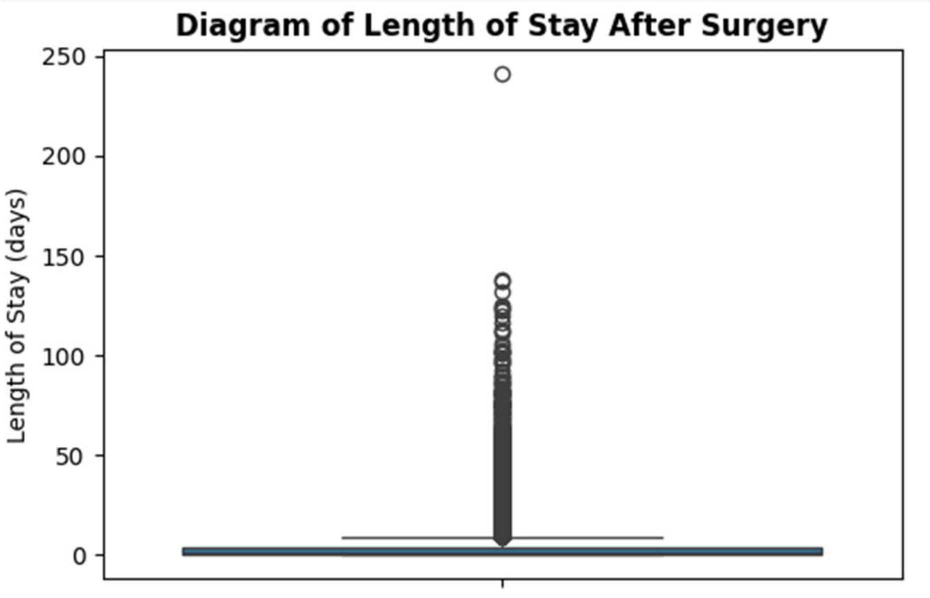
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 46316 entries, 990 to 18208
Data columns (total 13 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   ICU_ADMIN_FLAG      46316 non-null  object
 1   SURGERY_DATE        46316 non-null  datetime64[ns]
 2   BIRTH_DATE          46316 non-null  int64
 3   SEX                 46316 non-null  object
 4   ASA_RATING_C        46316 non-null  float64
 5   PATIENT_CLASS_NM    46316 non-null  object
 6   LOS_AS              46316 non-null  float64
 7   LOS_AS_CAT          46316 non-null  category
 8   LOR                 46316 non-null  float64
 9   SURGERY_MONTH       46316 non-null  int64
 10  SURGERY_YEAR        46316 non-null  int64
 11  ANEST               46316 non-null  object
 12  SURGERY_MONTH_YEAR  46316 non-null  period[M]
```

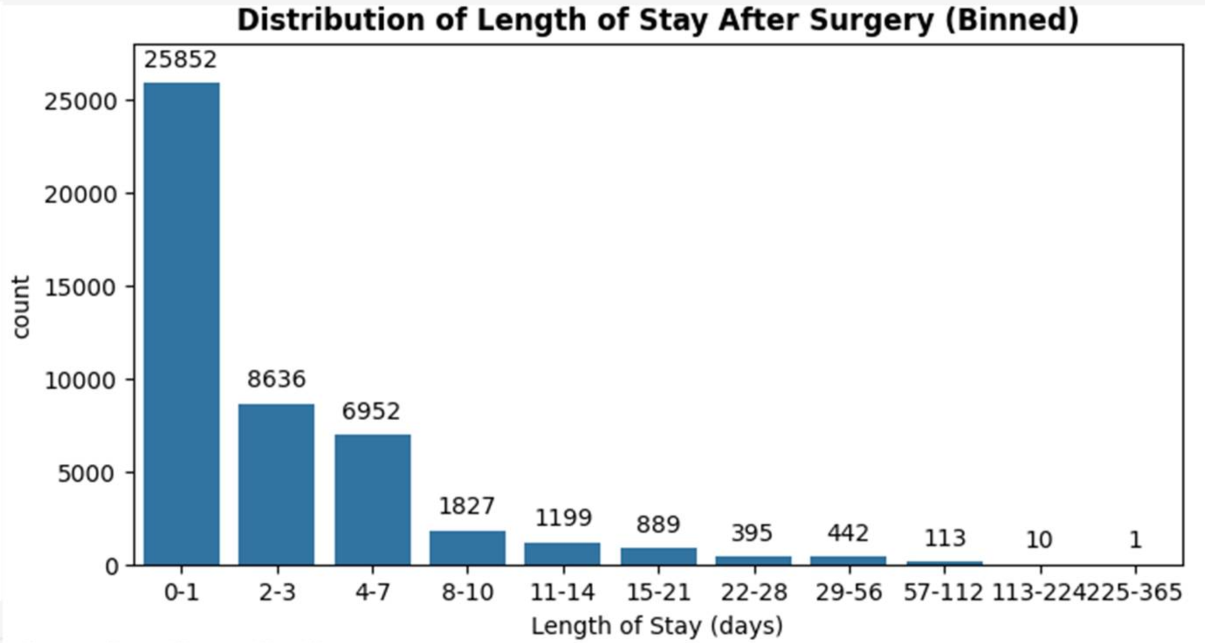
Metadata

Column	Value	Type
ICU_ADMIN_FLAG Patient's ICU admission history	Yes No	Categorical
BIRTH_DATE Age (years)	17-90	Continuous numerical
SEX Gender	Female Male	Categorical
ASA_RATING_C ASA score on patient condition --> Best to worst	1 2 3 4 5 6	Categorical ordinal
PATIENT_CLASS_NM Type of admission	Outpatient surgery Inpatient surgery Inpatient admission	Categorical
LOS_AS Length of stay after surgery (days)	0-241	Continuous numerical
LOS_AS_CAT Binned LOS_AS	0 1 2 3 4 5 6 7 8 9 10	Categorical ordinal
LOR Length of time in the operating room (minutes)	6 -1675	Continuous numerical
ANEST Type of anesthesia	General Non general	Categorical

Exploratory Data Analysis

Distribution of Length of Stay

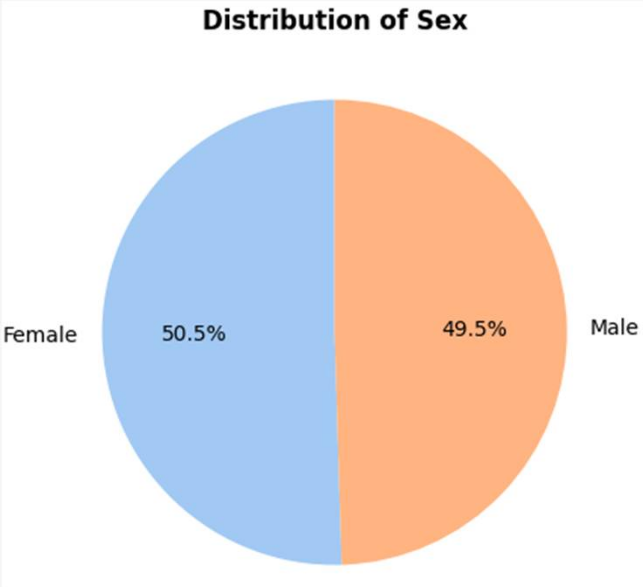


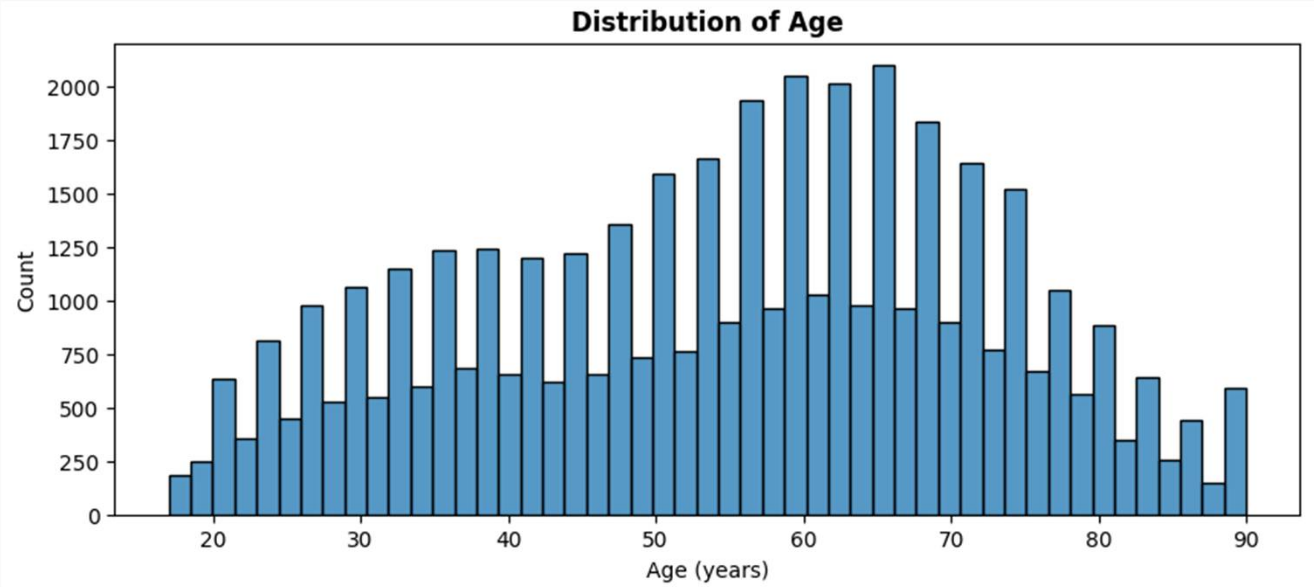


Interpretation

The length of stay is highest in the 0-1 day group and decreases as the days increase. Dominated by stays of less than 7 days.

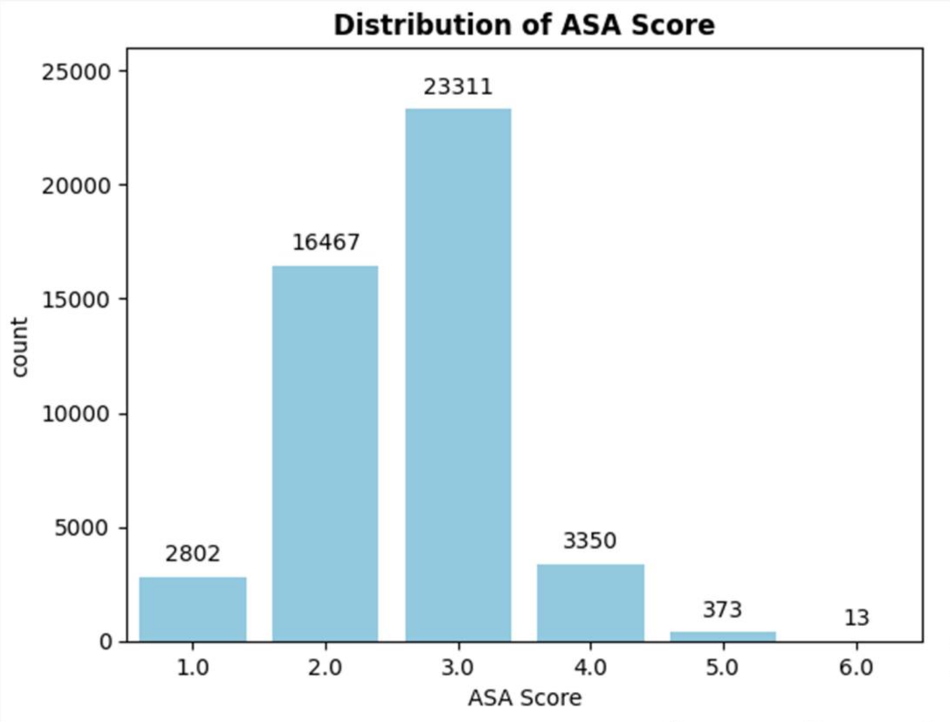
Distribution of Sex and Age

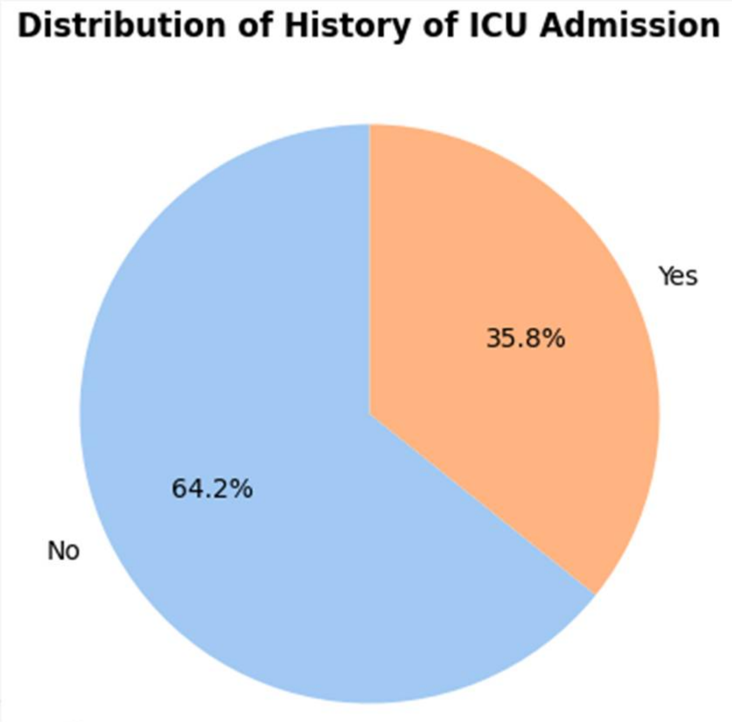




Interpretation
Gender between men and women appears balanced, and age seems to be dominated around the age of 60s.

Distribution of ASA Score and History of ICU Admission

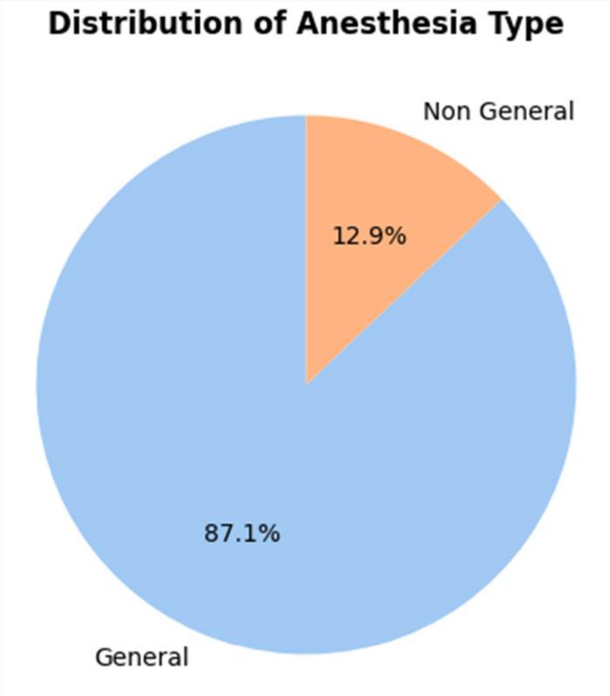


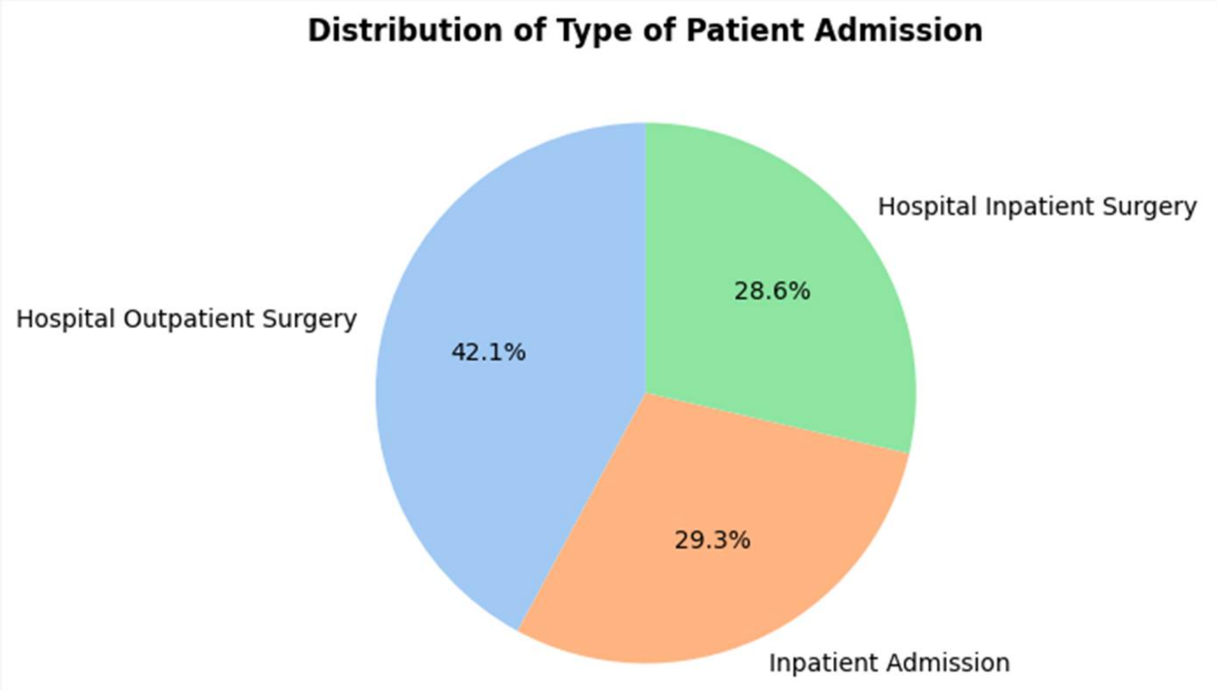


Interpretation

ASA score is dominated by scores 2 and 3, and the majority of patients do not have a history of being admitted to the ICU.

Distribution of Type of Anesthesia and Patient Admission





Interpretation

The majority of patients undergo general anesthesia, and admission from outpatient surgery is the most dominant.

Relationship of Length of Stay and Other Variables

Variable	Hypothesis test	Result	P-value	Interpretation	Value	Median (Mean) LOS
ASA Score	Kruskal-Wallis	4636,18	0,0	Significant relationship	1 2 3 4 5 6	0 1 2 5 10 0
Sex	Mann-Whitney U	2,55 e8	4,34 e-87	Significant relationship	Female Male	1 (2,69) 1 (3,83)
History of ICU Admission	Mann-Whitney U	6,88 e7	0,0	Significant relationship	No Yes	0 4
Age	Pearson	0,011	0,021	Significant relationship Weak correlation	-	-

Interpretation

Hypothesis testing indicates that LOS is significantly associated with ASA score, history of ICU admission, gender, and age, but age shows a weak correlation.

Deep Dive Exploratory Data Analysis

How does the difference in gender affect LOS ?

Distribution Table of Frequency and Percentage of LOS_AS based on SEX :

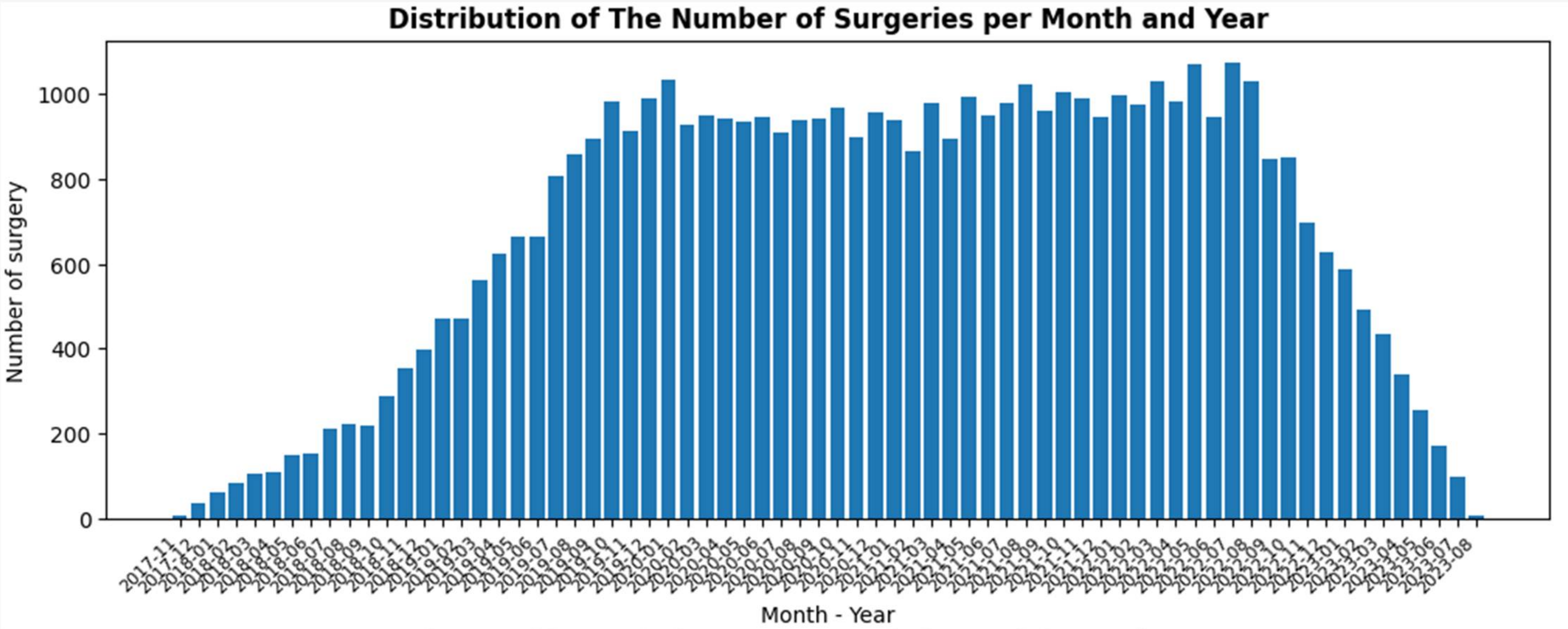
	LOS_AS_CAT	0-1	2-3	4-7	8-10	11-14	15-21	22-28	29-56	57-112	113-224	225-365	Total
	SEX												
Frequency	Female	14098.000	4651.000	3245.000	764.000	486.000	344.000	131.000	143.000	40.000	0.000	0.000	23902.0
	Male	12756.000	4105.000	3801.000	1091.000	755.000	578.000	271.000	308.000	75.000	11.000	1.000	23752.0
	Total	26854.000	8756.000	7046.000	1855.000	1241.000	922.000	402.000	451.000	115.000	11.000	1.000	47654.0
Percentage	Female	58.983	19.459	13.576	3.196	2.033	1.439	0.548	0.598	0.167	0.000	0.000	100.0
	Male	53.705	17.283	16.003	4.593	3.179	2.433	1.141	1.297	0.316	0.046	0.004	100.0
	Total	56.352	18.374	14.786	3.893	2.604	1.935	0.844	0.946	0.241	0.023	0.002	100.0

Fact
In LOS less than 4 days, it is dominated by females, while LOS above 4 days is more dominated by males.

Insight
Males tend to experience a longer LOS after surgery.

Recommendation
The hospital should allocate more for the provision of inpatient rooms for males.

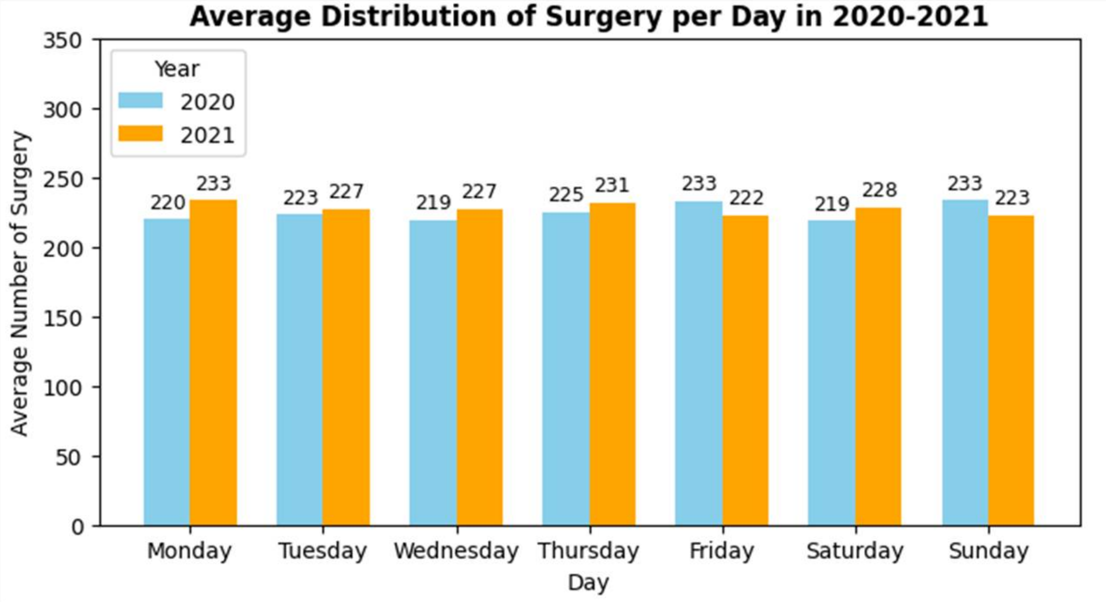
Is there seasonality in the number of surgeries?



Interpretation

It seems that the complete dataset for the entire day is only available for the years 2020-2021. There doesn't seem to be a distinctive pattern of the number of surgeries each month within the range of 2020-2021.

Is there seasonality in the number of surgeries?



Fact
The data indicates that the number of surgeries is almost the same for each day.

Insight
There is no significant difference in the number of surgeries based on the month and day.

Recommendation
In organizing healthcare personnel resources, hospitals should allocate the same amount every day, regardless of holidays or special occasions.

In outpatient surgery patients, are there who need to stay in the hospital?

Distribution of LOS in Hospital Outpatient Surgery

	0	1	2	3	4	5	6	7	8	9	11	12	13	14	15	16	10	17	20	21	22	28	33	23	25	26	34	36	51	61	102
Count	13209	4867	730	284	144	97	41	36	25	13	9	9	9	7	5	4	3	2	2	2	2	2	2	1	1	1	1	1	1	1	1
Percentage	67	24	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fact

Out of outpatient surgery patients, 33% were found to eventually require hospitalization. Approximately 24% stayed for at least 1 day, and the remaining stayed for more than 1 day.

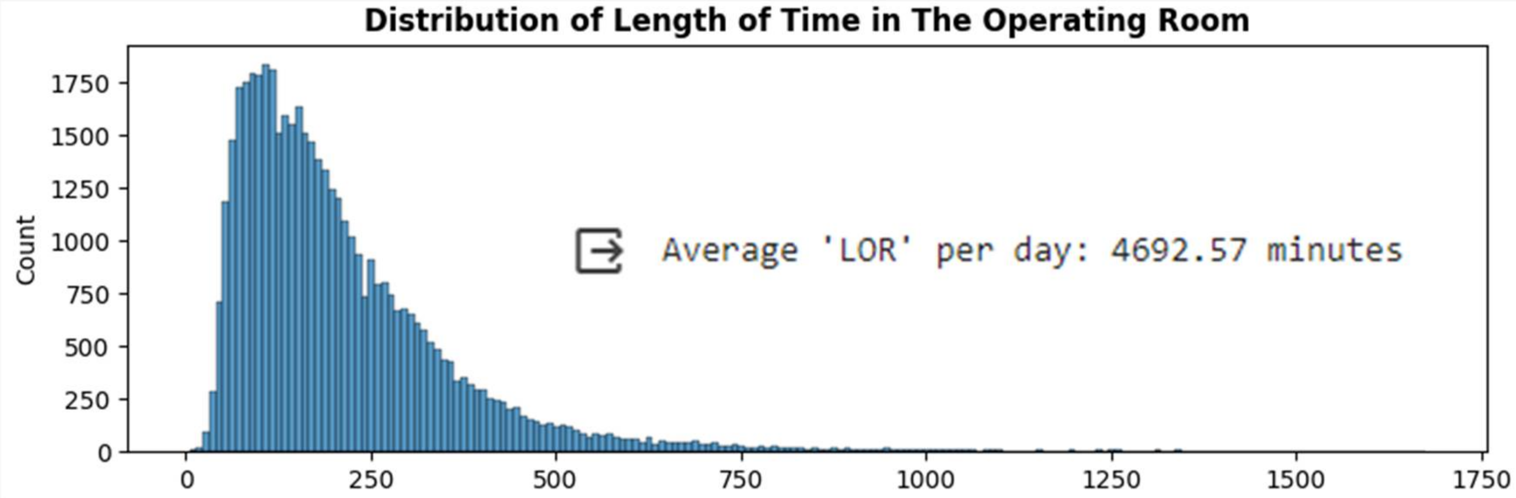
Insight

There is a possibility that one-third of outpatient surgery patients end up becoming inpatients

Recommendation

Hospitals should consider allocating ward space or beds for outpatient surgery patients who were initially predicted not to be hospitalized.

What is the average usage of the operating room per day?



Fact
The average usage of the operating room per day is 4,692.57 minutes, or approximately 78 hours. The duration of operating room usage is dominated by sessions of 100 minutes each.

Insight
The need for operating room usage is around 78 hours per day, with each operation requiring 100 minutes, serving as a reference for operating room preparation and resource allocation.

Recommendation
Hospitals can allocate and manage the required number of operating rooms, along with personnel and other resources, based on this data.

Does the duration of operating room usage depend on the type of anesthesia?

ANEST	General	Non General
count	40344.000000	5972.000000
mean	215.853782	154.962324
std	142.467618	119.562916
min	6.000000	11.000000
25%	115.000000	73.000000
50%	179.000000	118.000000
75%	278.000000	199.000000
max	1675.000000	1188.000000

Fact
For patients undergoing general anesthesia, the median operating room usage duration is 179 minutes, whereas for non-general anesthesia, it is 118 minutes.

Insight
Patients undergoing general anesthesia tend to use the operating room longer, for approximately 3 hours, compared to around 2 hours for non-general anesthesia.

Recommendation
Hospitals can manage operating room utilization by considering time allocation based on the type of anesthesia administered.

The Modelling

Regression

Evaluation

Metrics	Data	Ridge	Lasso	Random Forest Regressor	XGB Regressor	LightGBM	CatBoost	Keras	Ordinary Regression
MAE	Train	2.889	2.890	2.662	2.662	2.665	2.662	2.680	2.890
MAPE	Train	5.527	5.523	4.763	5.020	5.036	4.788	294.988	552.293
RMSE	Train	6.095	6.095	5.912	5.891	5.898	5.905	5.945	6.096
R-squared	Train	19.0%	19.0%	23.8%	24,3%	24.1%	24.0%	22.9%	19.0%
MAE	Test	2.918	2.918	2.725	2.723	2.729	2.721	2.729	2.918
MAPE	Test	5.381	5.378	4.671	4.930	4.940	4.695	291.848	537.758
RMSE	Test	6.357	6.357	6.279	6.264	6.276	6.271	6.263	6.357
R-squared	Test	17.3%	17.3%	19.3%	19.7%	19.4%	19.5%	19.7%	17.3%

All modeling was done using Grid Search CV.

Interpretation

Regression modeling seems to provide less satisfactory results. The R2 score values on the training and test data indicate that the model can only explain the variability of the data by 17.3-24.3%. A lower R2 score on the test data compared to the training data may indicate underfitting.

Regression (Binned LOS 0-7 days)

Evaluation

Metrics	Data	Ridge	Random Forest Regressor
MAE	Train	1.164	1.148
MAPE	Train	3.851	3.781
RMSE	Train	1.557	1.542
R-squared	Train	29.1%	30.5%
MAE	Test	1.167	1.154
MAPE	Test	3.717	3.675
RMSE	Test	1.565	1.554
R-squared	Test	29.3%	30.2%

All modeling was done using Grid Search CV.

Interpretation

By binning the target values into 0-7 days, there appears to be an improvement in the results of regression modeling. The R2 score increases, with the best result achieved by the Random Forest Regressor at 30.2% (on the test data), and the results appear to be a good fit compared to regression modeling without binning the target. However, this value is still considered less satisfactory.

Classification

Evaluation (without binning the target)

Metrics	Random Forest Classification	KNN	Keras	Logistic Regression
Precision weighted avg.	**0.55**	0.48	0.33	0.54
Recall weighted avg.	0.53	0.56	0.31	0.61
F1 weighted avg.	0.50	0.51	0.31	0.56

All modeling was done using RandomOverSampler and Grid Search CV.

Binned target 0-7 days

Metrics	Random Forest Classification	Logistic Regression
Precision weighted avg.	**0.64**	0.56
Recall weighted avg.	0.68	0.64
F1 weighted avg.	0.64	0.55

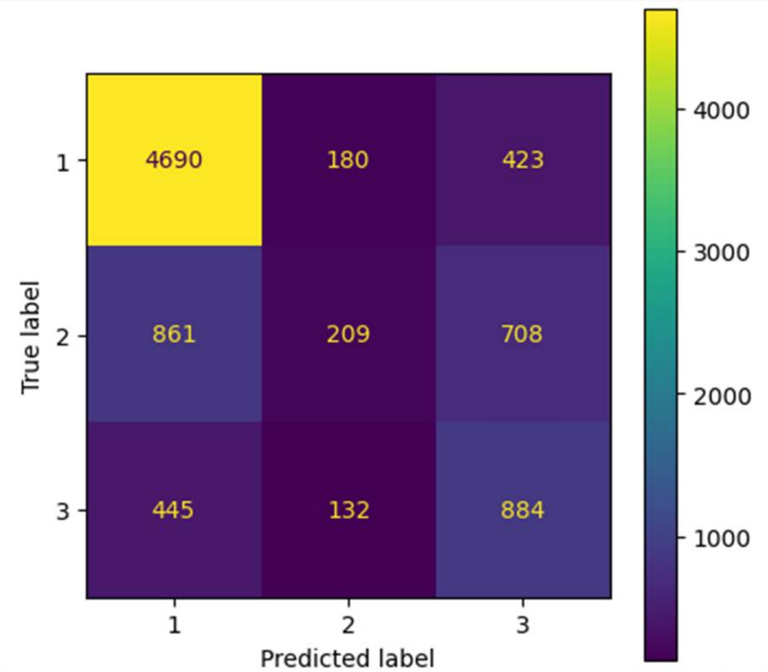
All modeling was done using RandomOverSampler and Grid Search CV.

Interpretation

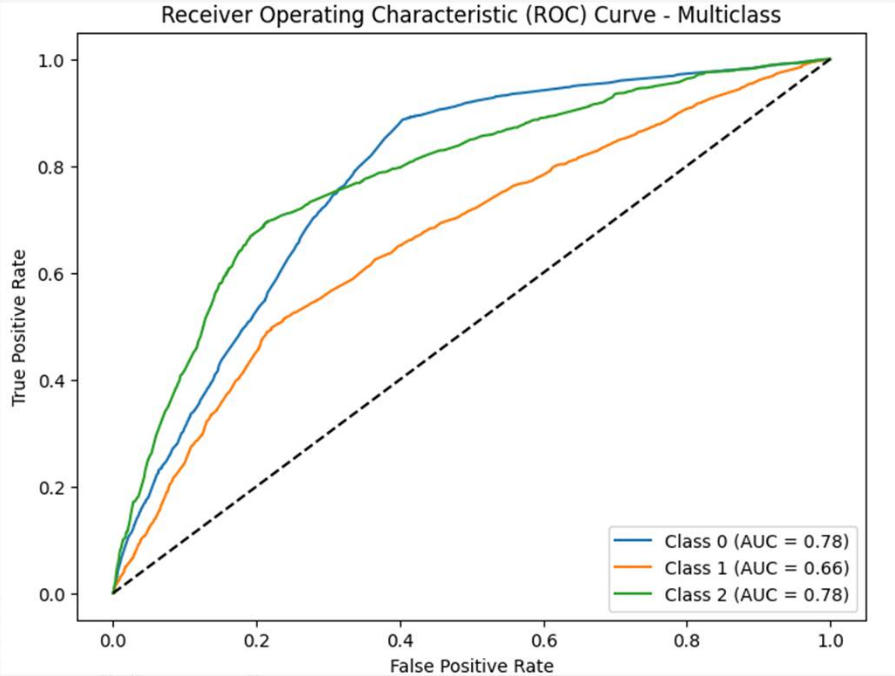
The primary metric chosen is precision, and the most favorable outcomes are observed in the modeling process when the target data is binned into 0-7 days, resulting in a quite satisfactory precision of 64%.

Best model (Random Forest Classification on LOS 0-7 days)

Confusion matrix



AUC/ ROC curve



Bin 1 (Class 0) = 0-1 days
Bin 2 (Class 1) = 2-3 days
Bin 3 (Class 2) = 4-7 days

Interpretation

The model appears to perform well in predicting LOS for 0-1 days and 4-7 days, but it faces challenges in predicting LOS for 2-3 days.

Best model (Random Forest Classification on LOS 0-7 days)

Feature importance

	Feature	Importance
0	ICU_ADMIN_FLAG	0.857450
3	ASA_RATING_C	0.115132
1	BIRTH_DATE	0.021530
2	SEX	0.005888

Interpretation

The feature importance analysis reveals that the history of ICU admission plays a pivotal role in predicting the target outcome, as indicated by its highest importance value of 0.86. ASA score also contributes significantly to the model, while age and sex exhibit lower importance but still provide valuable contributions to the classification process.

RECOMMENDATION

1. The predictive modeling of Length of Stay (LOS) can offer valuable insights for optimizing post-surgical bed utilization over a 7-day period.

2. Based on the analysis findings, the hospital management is advised to strategically organize ward composition, particularly for patients with an increased likelihood of longer stays. This includes prioritizing accommodation for male patients and those with a history of ICU admission, especially those with high ASA scores.

By tailoring ward arrangements to accommodate these risk factors, the hospital can enhance its capacity planning and resource allocation, ultimately improving patient care and operational efficiency.

FOLLOW UP

- Enhance the model performance by :
  - Implement feature engineering techniques, including data splitting based on specific variables
  - Optimization in hyperparameter tuning.
- Collaborate with clinical experts and data analysts to continuously improve the model based on feedback and clinical understanding.

References

Samad M, Rinehart J, Angel M, Kanomata Y, Baldi P, Cannesson M. MOVER: Medical Informatics Operating Room Vitals and Events Repository. medRxiv [Preprint]. 2023 Mar 12:2023.03.03.23286777. doi: 10.1101/2023.03.03.23286777. Update in: JAMIA Open. 2023 Oct 17;6(4):ooad084. PMID: 36945552; PMCID: PMC10029016.

thank you

Contact:
estikasari@yahoo.com