Elevating Nhs Hospital Stay Durations With Precision Analytics

Nishtha Singh

1 INTRODUCTION

The National Health Service faces a pressing need to better manage the duration of hospital stays, particularly in the case of back surgery. Extended hospital admissions overburden facilities, drive up costs, and reduce the quality of care provided to patients. The Chief Executive of the NHS is, therefore, interested in implementing predictive analytics for enhanced resource utilization and optimized patient flow. Meanwhile, the NHS Trust channels its effort into the analysis of the data concerning patients who underwent back surgery to identify trends affecting the duration of hospital stays. Leveraging principles and methods of advanced analytics, such as data exploration and data transformation, and predictive methods, including predictive modelling, seeks to elevate the quality of the available data and uncover opportunities for improvement. It is a critical initiative as it will influence the development of the institution's guidelines and policies, as well as the allocation of sources. In particular, by providing evidence-based recommendations and transparent reporting of insights, the initiative will improve efficiency and patient outcomes while allowing the NHS to continue providing the quality healthcare it is committed to.

1.1 AIM

Predictive analytics could be applied to NHS Trust's Health System to analyse the period that each patient stays in the hospital after back surgery. This concept enhances patient care by determining the causes of prolonged stay and providing accurate forecast duration lastly making an evidence-based strategic plan. The initiative also enhances effective resource utilization and economical budget rates that successively influence patient satisfaction levels.

1.2 OBJECTIVE

The objective is to analyze and forecast hospital stays using data exploration, transformation, and predictive modeling methods to determine the hospital stay drivers post-back surgery with the intention of supporting resource allocation and patient care management strategies in the NHS.

- Hospital stay length forecasting to support resource allocation and operational efficiency.
- Studying the factors associated with extended stays for budget planning and patient care optimization.
- Forecasting to support strategic planning and medical system preparedness.

- Targeted interventions to increase patient satisfaction and outcomes.

2 DATA ANALYSIS

The data analysis process is a critical aspect of transforming raw data into valuable insights, which is crucial, particularly in the healthcare sector, considering the significant implications of decision-making on the operational efficiency and patient welfare.

2.1 DATA INPUT

To start with **Data Input**, involving four distinct datasets for the analysis: Patient Information (comprising information like Patient ID, First Name, Last Name, Date of Birth and patients Gender) Surgical Information (includes Patient ID, Hospital Admission Date, ICD-10 Code, Surgery End Date time, First Ambulation and Hospital Release date and time of the patient), ICD-10 Codes (contains the ICD-10 Codes and their Description), and Current Patients (gives us data about the Gender, Age, Surgery Type and Hours till Ambulation). Each data set was crucial in building an in-depth insight into the reasons underpinning the hospital stays subsequent to back surgery.

2.2 DATA PREPARATION AND DATA PARSING

This step helps in improving the reliability of analysis, principally in Patient information file. 'Date of Birth' was converted from string to date format, enhancing data reliability and well placed to facilitate a better analysis.

Surgical Information file involved breaking down data elements to calculate the Pre-Surgery Hours and Ambulation to Hospital Release times to enhance the granularity level of patient hospital stay durations. These elements have been instrumental in making the hospital stay more manageable and analyzable by applying DateTime tool, Select Tool, Formula Tool and Sort tool.

The ICD-10 Code input file was decoded to classify surgery types and procedures, and even the whitespace was removed to clean surgical data using Text to Column and Data Cleansing tools.

2.3 DATA BLENDING AND PRE-PROCESSING

Blending consisted of incorporating the output of the preparation and parsing container into a single harmonized and coherent join, blending patient records across different sources to create an overview of each patient's journey. The union of the datasets was crucial to calculate the Age and using the formula tool to further segregate them into respective Age Groups. Sort tool was used to extract the Total duration of Stay in descending order, which helps us get the idea on the maximum duration of stay.

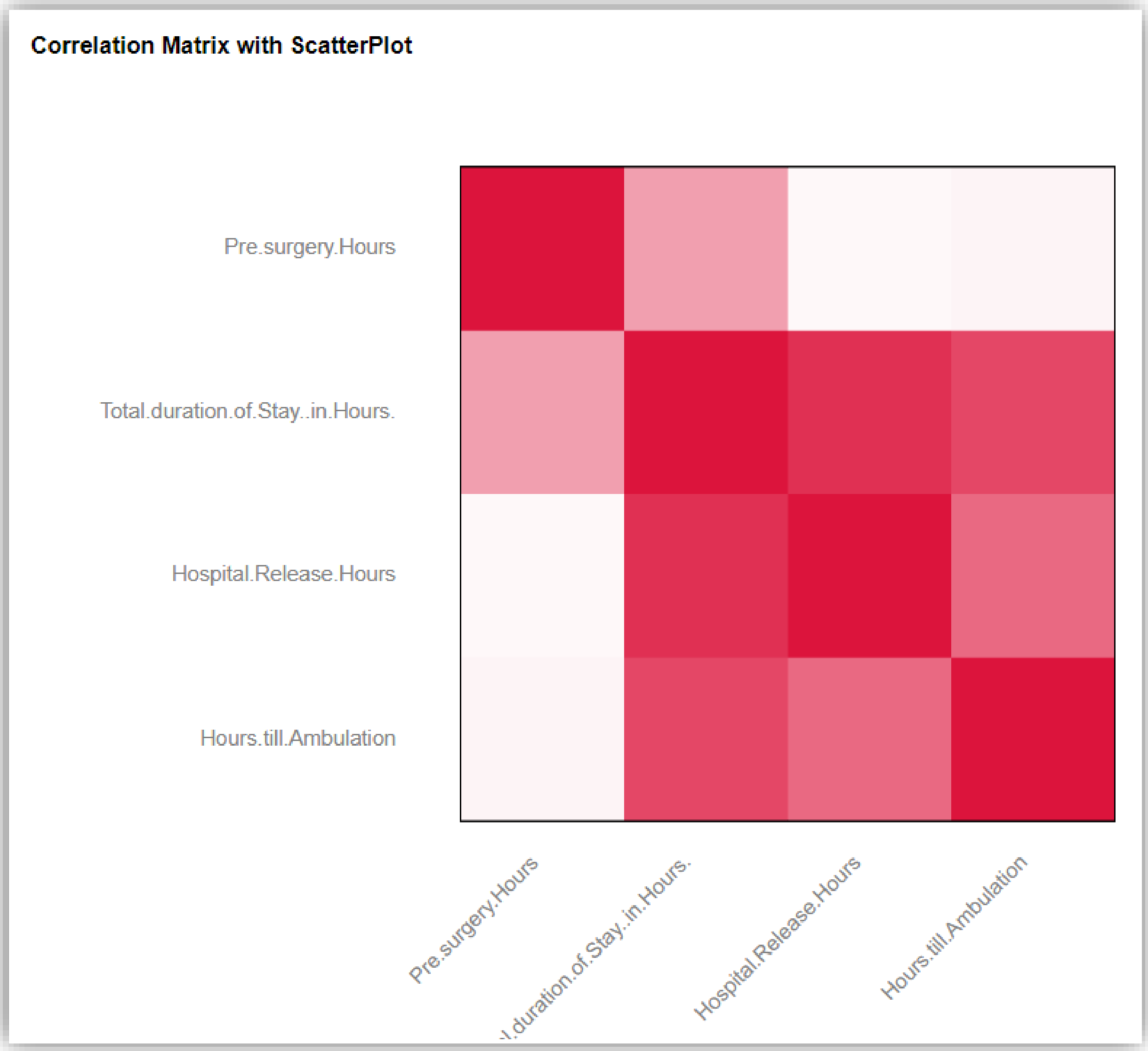
Pre-processing further refined the data set for features relevant to hospital stay durations, encode variable categories, and scale numeric inputs, perfect for predictive modeling. This was done using the Select tool. Data cleansing was implemented to remove trailing white spaces. Essentially, this stage ensures the analysis was grounded in precise, comprehensive, and consistent data.

2.4 DATA INVESTIGATION

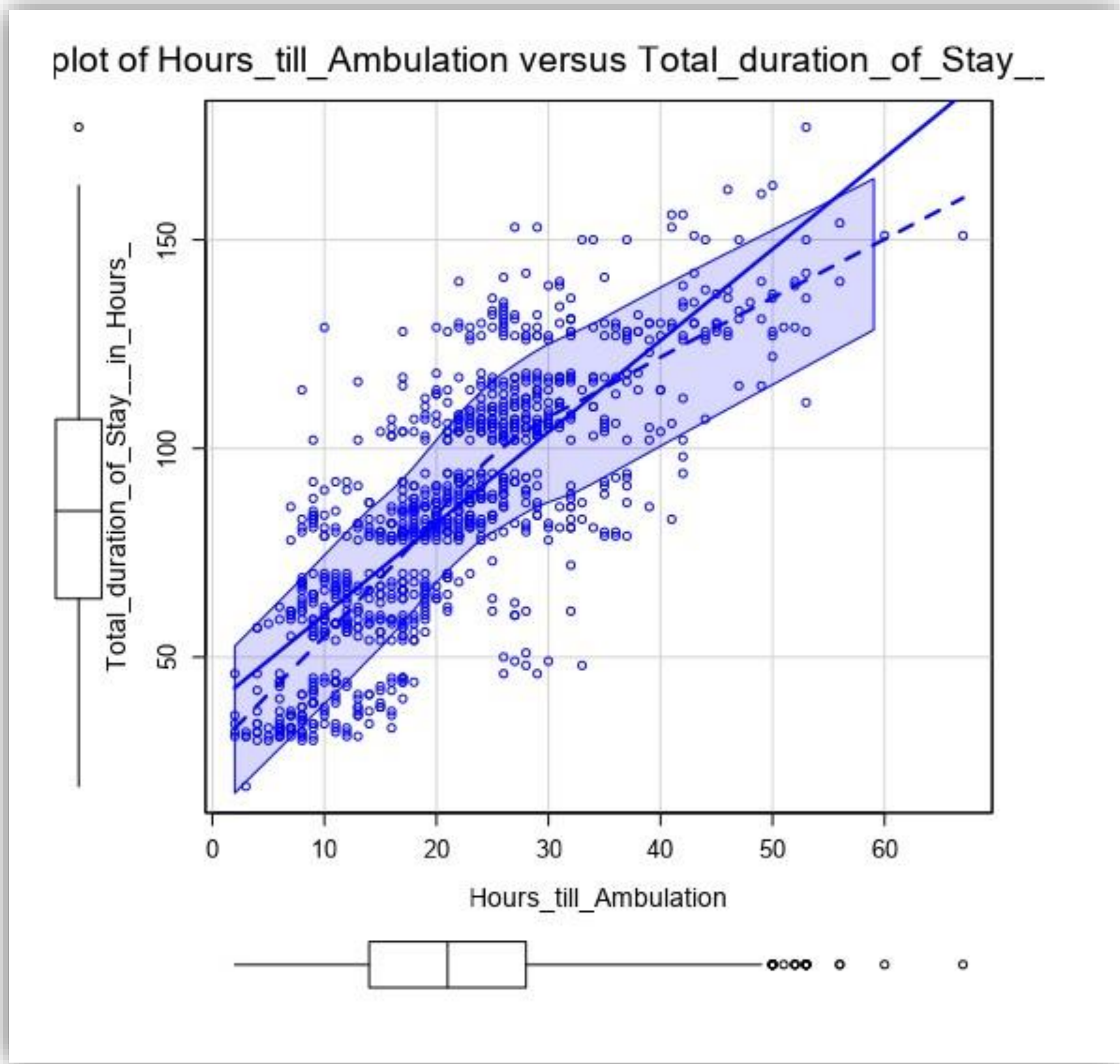
Exploratory Data Analysis (EDA) with a cleansed and consolidated dataset conducted to grasp initial insights and create the modelling process. Visualization tools and statistics were used to understand the distributions, correlations and time-series investigations.

- **Correlation analysis** has been used to find the correlation between different variables. The correlation types have been used from Data Investigation palette.

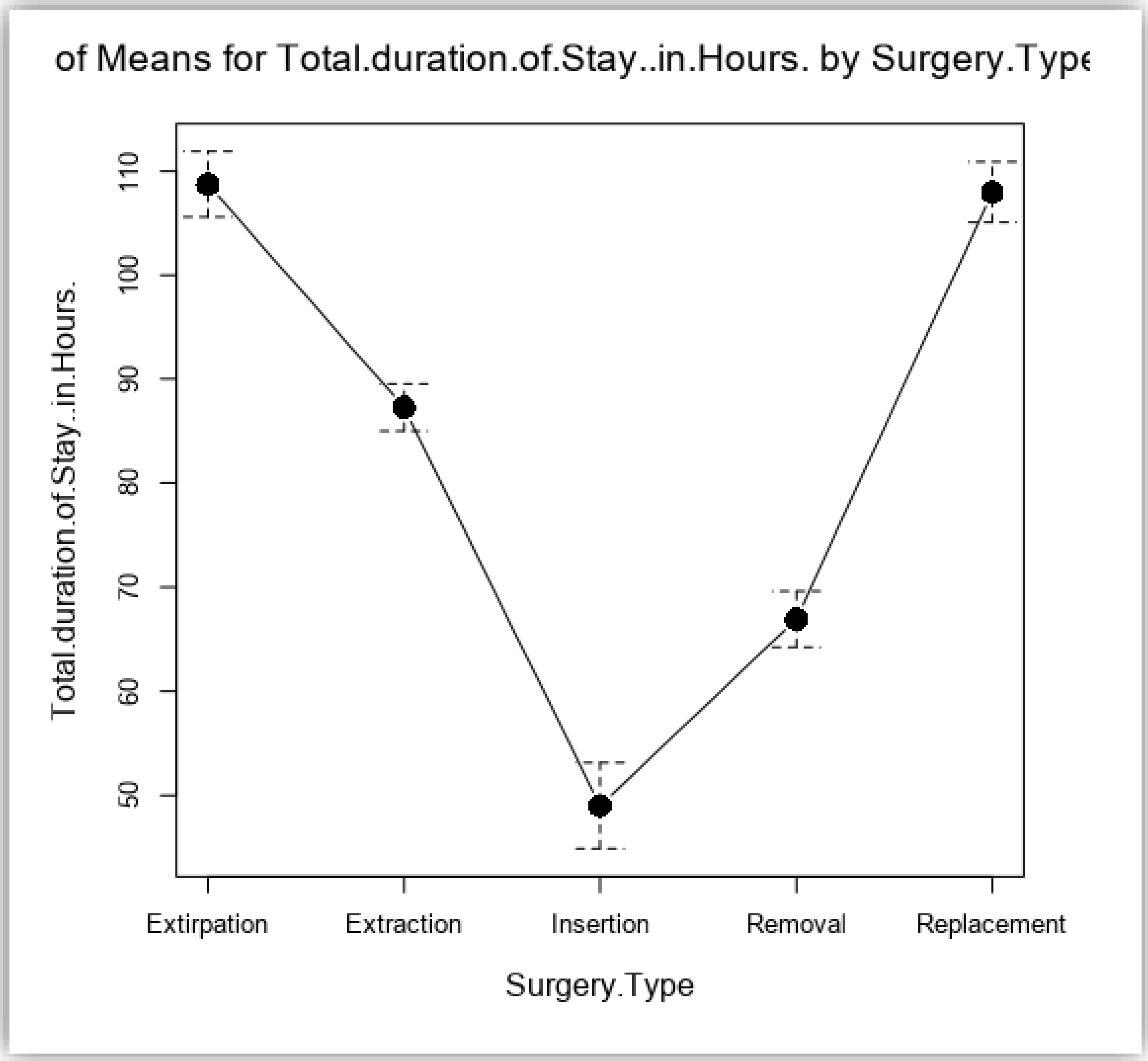
  1. **Association analysis** reveals a strong positive relationship between the total duration of hospitalization and hours til ambulation, demonstrating longer stays are correlated with delayed ambulation.



2. **Plot of Means** shows an upward curve in stay duration with longer ambulation hours, with notable outliers showcasing areas for patient care improvements to lower the time of hospitalization.



- **Length of Stay Analysis**: The average Total duration of Stay (in Hours) for various types of surgical procedures, while different surgery types differ significantly. The extirpation and replacement surgeries have the maximal stays, which averages about 110 hours. The corresponding extraction surgeries have the above-100-hour result. The insertion and removal procedures have almost twofold lower indicators, which is about 60 hours. This trend recurs with the help of a combination of various tools.



2.5 MODEL EVALUATION AND COMPARISON

An array of models, from linear regression to advance ensembles like Forest Model and Decision Tree, were assessed for their appropriateness of hospital stay durations. The comparison was based on a comprehensive assessment utilizing the following performance metrics: Correlation (r), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Percentage Error (MPE), and Mean Absolute Percentage Error (MAPE). These metrics are fundamental to evaluate the consistency, accuracy, and practicality of each model in capturing the intricacies of hospital stay durations.

Linear Regression emerged as the preferred model, described by its higher correlation coefficient of 0.8959. This indicates a robust linear relationship between the predicted hospital stays and the actual stretches, a vital aspect for assuring the reliability of forecasts in a healthcare environment. Moreover, its RMSE and MAE are the lowest among the models evaluated, standing at 12.7033 and 9.7384, respectively. These numbers indicate that Linear Regression's predictions are not only close to the actual values but also unvarying, with smaller deviation across error magnitude predictions.

The Linear Regression model also signified the least bias in projections, as observed by its MPE and MAPE values. With an MPE of -3.3270, it highlights a balanced level to underestimation and overestimation, crucial to maintain a realistic overview of hospital resource requirements. The MAPE value of 13.1445 further proves its superiority in providing stable and realistic predictions across a range of scenarios, which makes this highly applicable to strategic-oriented decisions.

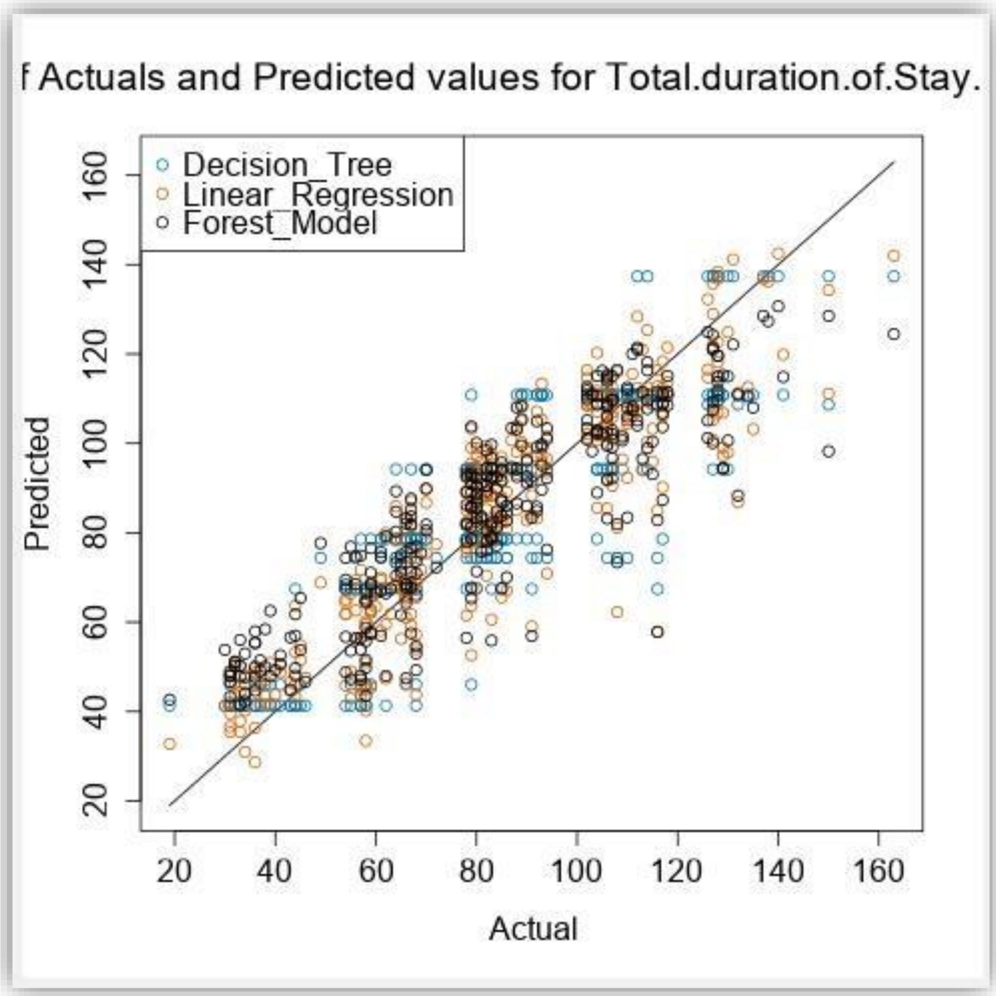
The preference of the Linear Regression model over the Decision Tree and Forest Model was not merely based on quantitative benchmarks but also on the model's efficiency and accountability.

Model Comparison Report

Fit and error measures

Model	Correlation	RMSE	MAE	MPE	MAPE
Decision_Tree	0.8776	13.7333	10.9089	-3.7225	14.6081
Linear_Regression	0.8959	12.7033	9.7384	-3.3270	13.1445
Forest_Model	0.8805	13.8835	10.9138	-6.5605	15.7932

Model: model names in the current comparison.
Correlation: correlation between the predicted values and the actual values.
RMSE: root mean square error.
MAE: mean absolute error.
MPE: mean percentage error. Note: based on its definition, MPE may be positive or negative infinity if the target variable has 0 values. In this case, we return a weighted percentage error (WPE).
MAPE: mean absolute percentage error. Note: based on its definition, MAPE may be positive or negative infinity if the target variable has 0 values. In this case, we return a weighted absolute percentage error (WAPE).



2.6 SCORING

The final stage in the data analysis process was using the selected model to score the current patient dataset, predicting their hospital stay durations. This scoring step retrieved the NHS's anticipated duration forecasts while confirming the model’s real-world application. The derived scores are influenced by patient demographics and clinical information, including gender, age, surgery type, and patient ambulation. Higher scores align with older age groups for invasive surgeries like "Replacement," with extended ambulation periods leading to extended stays. The corresponding scoring, thus, gives a valuable benchmark to healthcare professionals to gauge expected hospital stays, enabling proper patient care management and hospital resource allocation based on each recovery program.

3 CHOICES AND ASSUMPTIONS

3.1 CHOICES

- The preference of Linear Regression for the NHS's predictive analytics is justified by the lucidity and straightforward interpretability, strengthens trust and supports better decisions. With a MAPE of 13.1445, it guarantees precise predictions, necessary for detailed operational planning and resource allocation, where accuracy correlates with patient care.

Report

Report for Linear Model Linear_Regression

Basic Summary

Call:
lm(formula = Total.duration.of.Stay..in.Hours. ~ Gender + Hours.till.Ambulation + Surgery.Type + Age, data = the.data)

Residuals:

Min	1Q	Median	3Q	Max
-28.38	-10.01	-3.25	5.08	51.07

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	63.63416	3.76537	16.89984	< 2.2e-16 ***
GenderGenderqueer	-6.81487	8.69869	-0.78344	0.43395
GenderMale	0.24125	1.66028	0.14531	0.88456
GenderOTH	-0.19379	14.97005	-0.01295	0.98968
GenderTransgender Female	-7.68584	10.66556	-0.72062	0.47166
GenderTransgender Male	-2.28227	7.61443	-0.29973	0.76458
Hours.till.Ambulation	1.57671	0.07848	20.09064	< 2.2e-16 ***
Surgery.TypeExtraction	-10.12812	2.50780	-4.03864	7e-05 ***
Surgery.TypeInsertion	-39.03637	2.99487	-13.03443	< 2.2e-16 ***
Surgery.TypeRemoval	-27.92786	2.59518	-10.76143	< 2.2e-16 ***
Surgery.TypeReplacement	-2.90446	2.52693	-1.14940	0.25124
Age	0.04746	0.04592	1.03372	0.30204

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.829 on 322 degrees of freedom
Multiple R-squared: 0.7753, Adjusted R-Squared: 0.7676
F-statistic: 101 on 11 and 322 degrees of freedom (DF), p-value < 2.2e-16

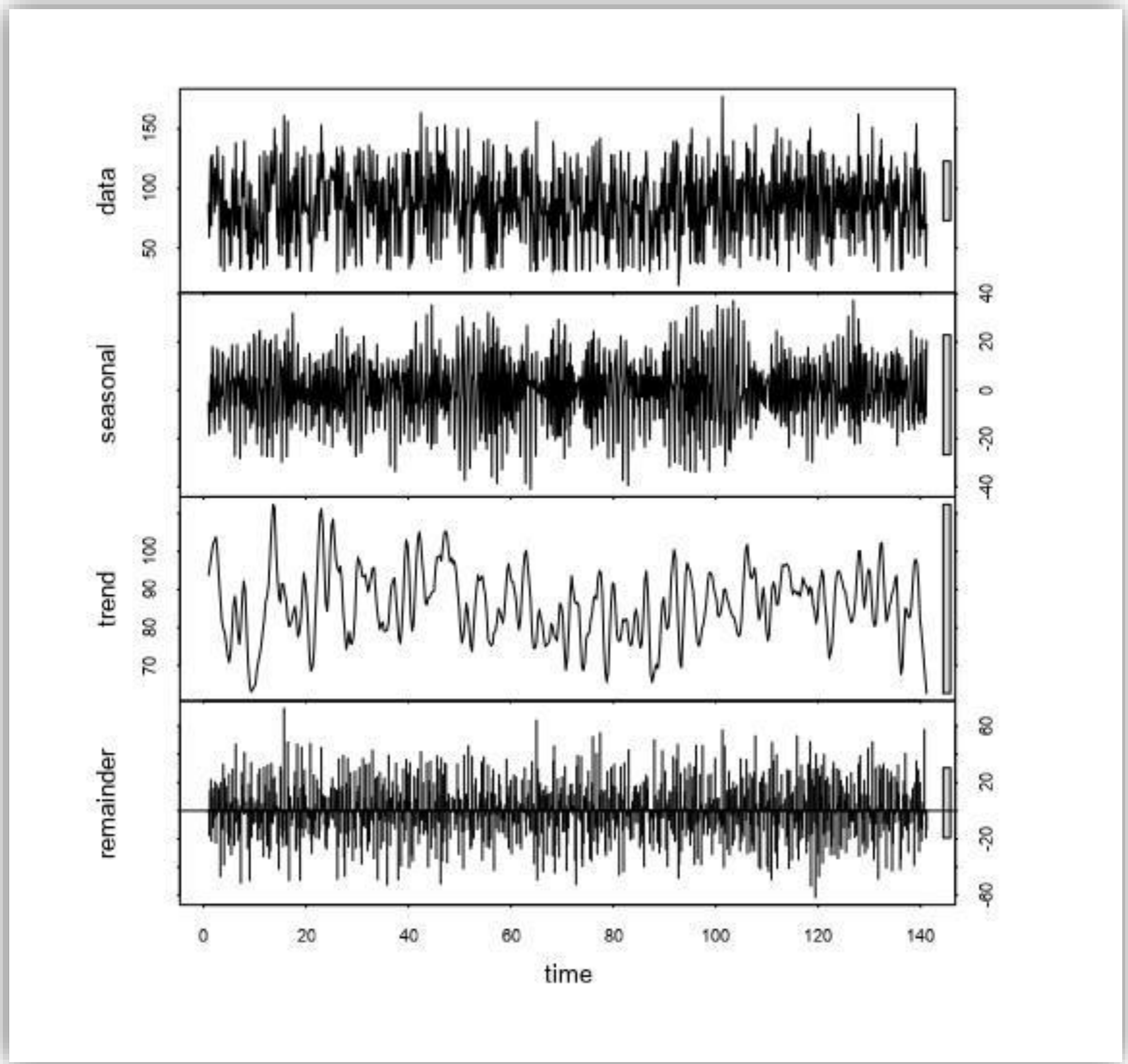
Type II ANOVA Analysis

Response: Total.duration.of.Stay..in.Hours.

	Sum Sq	DF	F value	Pr(>F)
Gender	283.73	5	0.26	0.93559
Hours.till.Ambulation	88762.24	1	403.63	< 2.2e-16 ***
Surgery.Type	54721.05	4	62.21	< 2.2e-16 ***
Age	234.99	1	1.07	0.30204
Residuals	70810.35	322		

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

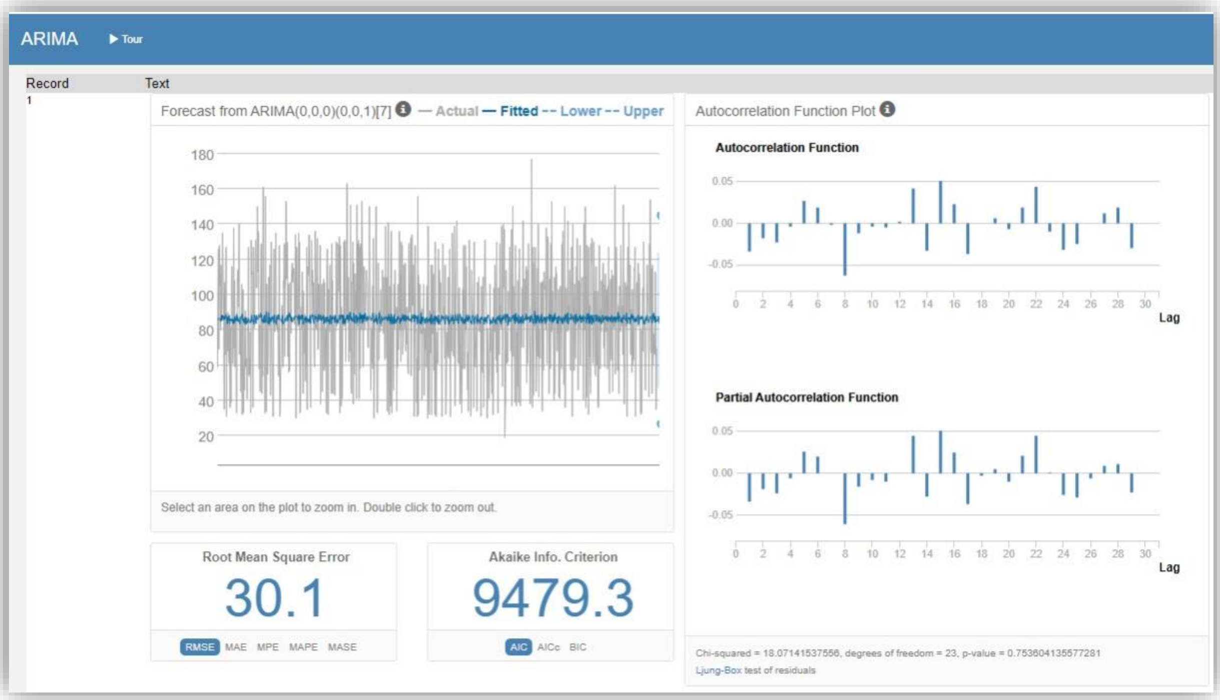
- Time series decomposition of hospital stay durations into trends, seasonality, and residuals offers the NHS clear pattern insights, with a daily analysis aligning with internal administrative cycles. The model chooses additive and multiplicative approaches based on the consistency of seasonal contrasts, ensuring authentic replication of patient stay tendencies.

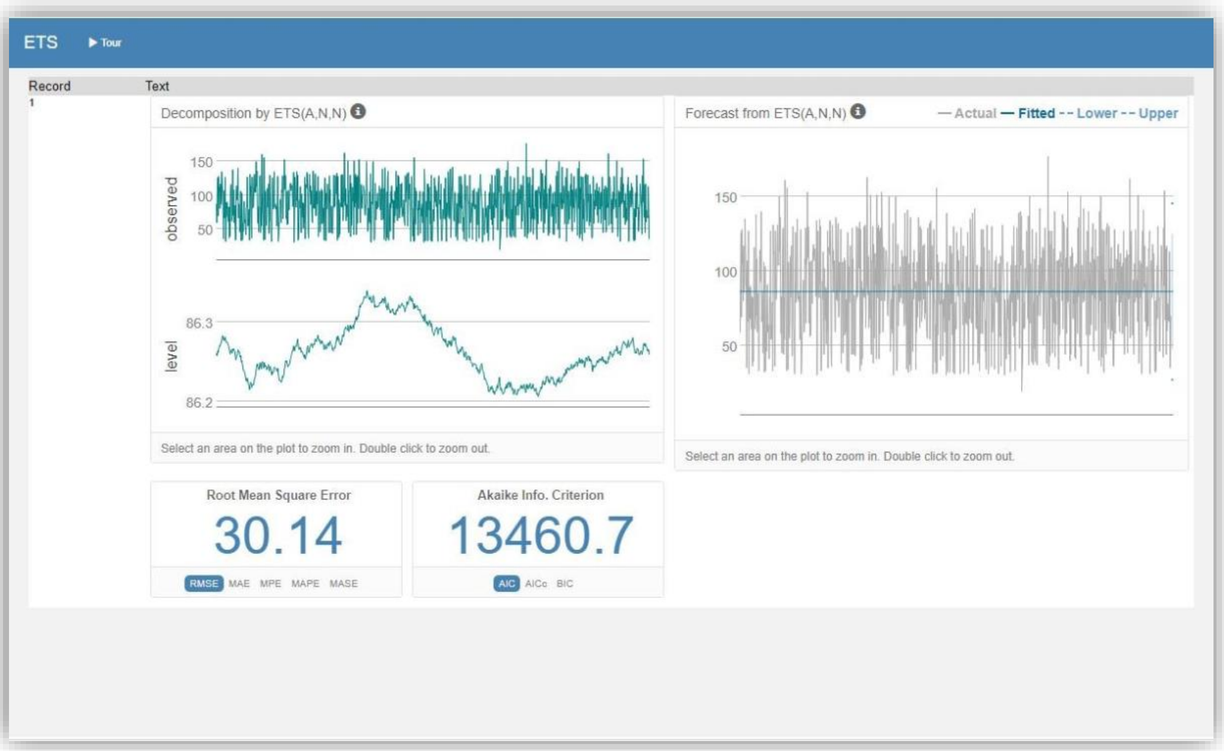


3.2 ASSUMPTIONS

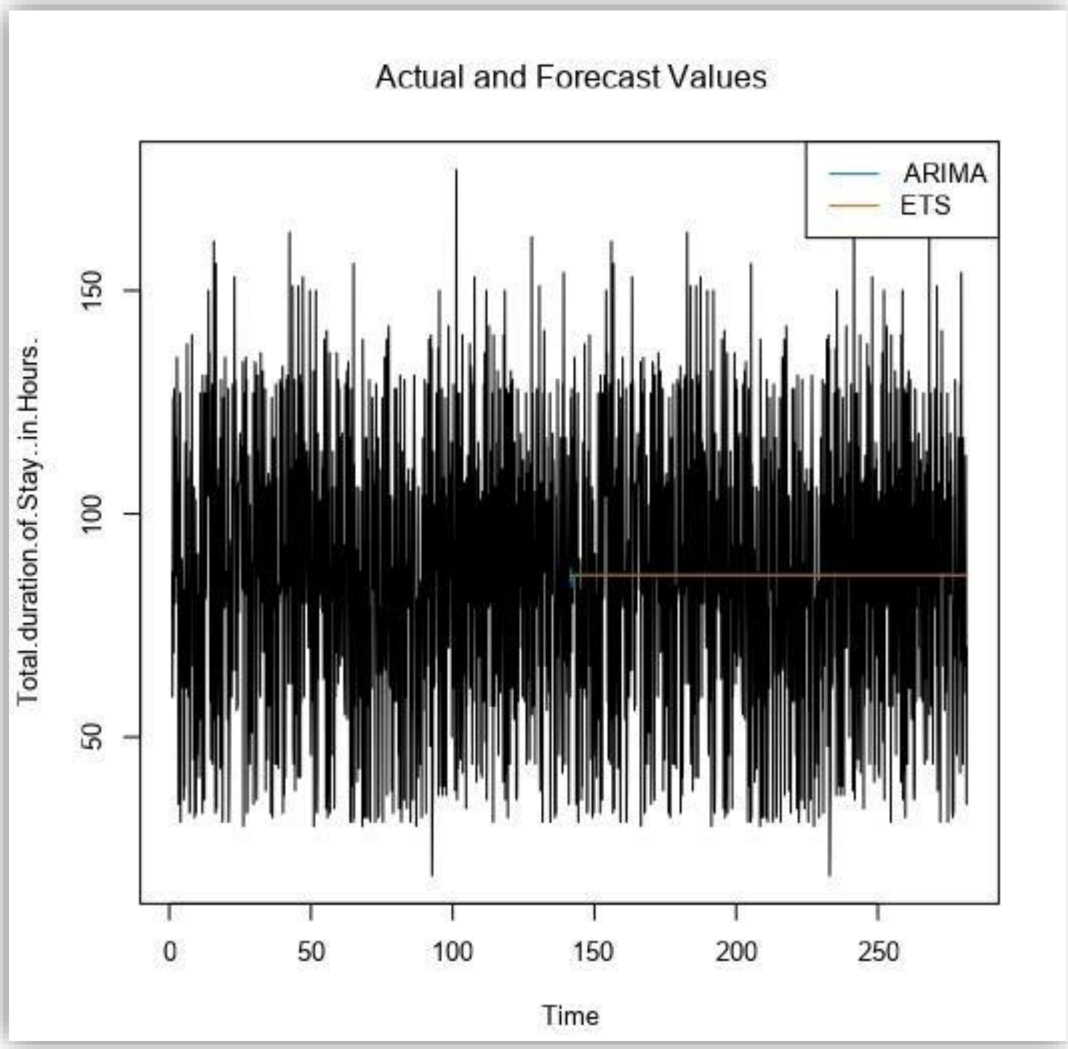
- Unchanging seasonality and consistent data recording for NHS stay durations are assumed in the analysis both in terms of linear trends and captured external factors. Outliers are contemplated as random, with historical patterns educating future forecasts, supporting the study's predictive reliability.
- The ARIMA and ETS models applied to daily hospital stay durations- Assumptions:

ARIMA assumes data stationary, suggesting that trends and variances remain consistent over time, while ETS relies on the time series being decomposable into error, trend, and seasonal elements. The daily data is presumed detailed enough to indicate and predict patterns, with the expectation that historical patterns predicts the future. It is also assumed that there are no policy shifts or outside disruptions affecting the data's trend or seasonality.





ETS
▶ Tour
Record
Text
1
Decomposition by ETS(A,N,N)
150
100
50
observed
86.3
level
86.2
Select an area on the plot to zoom in. Double click to zoom out.
Forecast from ETS(A,N,N)
— Actual — Fitted - - Lower - - Upper
150
100
50
Select an area on the plot to zoom in. Double click to zoom out.
Root Mean Square Error
30.14
RMSE MAE MPE MAPE MASE
Akaike Info. Criterion
13460.7
AIC AICc BIC



- The data collection process is presumed to be uniform and without error, ensuring the forecasts are based on accurate and complete information and therefore, the forecasts are accurate and complete. These assumptions make the predictions relevant for future operational adjustments within the NHS.

4 RECOMMENDATIONS

The recommendation to NHS Trust, based on the analysis, is to improve operational efficiency, patient care, and resource management.

1. **Implementation and Refinement of Analysis**: Develop predictive models recognizing the hospital stay's duration at a high degree of precision. This enables accurate allocation of

resources, like beds and manpower, coordinates patient flow, directly resolving operational efficiency and avoiding wait times.

2. **Pre- and Post-Operative Protocols**: Analyzing key indicators impacting increased hospital stays through data analysis. Targeted pre-operative assessments and optimized post-operative recovery can lead to reduced recovery time and better outcomes.

3. **Improved Surgical Efficiency**: Employing surgical data insights to streamline operations, focusing on procedures identified as outliers in terms of duration and resource use. Implementing best practices and updating older techniques can lead to shorter hospital stays and better resource utilization.

4. **Personalize Patient Plans**: Utilize model predictions to efficiently communicating with the patients about their expected recovery timelines. Tailored recovery plans can also facilitate better post-hospitalization care.

4.1 EVALUATING RECOMMENDATIONS

The effectiveness of the recommendations shall be measured using a systematic review of key performance indicators (KPIs) such as average length of stay, patient throughput, and satisfaction scores before and after implementation. Analyzing trends in these metrics will provide quantifiable ounce of improvement.

4.2 ADDITIONAL DATA RECOMMENDATIONS

- Data enrichment with more detailed identified on patient health behaviours, socio-economic status and post-discharge follow up outcomes. It would provide more valuable data and reviews a set of factors affecting stay durations and measures the actual long-term results of the proposed recommendations.
- Implemented with the critical balance between technology, patient focus with significant room for improvement with focus on predictive analytics and focal interventions are impactful on the gains in driving efficiency and patient satisfaction.

5 DATA COLLECTION STRATEGY

Data collection throughout NHS Trust should be streamlined for enhanced predictive analytics, the following strategy is recommended:

1. **Implement a Unified Data Management System**: One data system across NHS facilities that integrates patient, surgical, and post-operative reports, ensuring standardized formats and real-time results for comprehensive analysis.

2. **Expand Real-Time and Patient-Centered Data Collection**: Utilize electronic health record (EHR) systems for real-time data capture, including patient-reported outcome measures (PROMs) and social determinants of health (SDOH). This approach ensures up-to-date patient data, enriches the dataset with patient experiences and facilitates personalized care planning.

3. **Post-Discharge Parameters**: Collect systematic data on post-discharge metrics such as re-admittance and recovery to evaluate treatment efficacy.

4. **Ensure Data Quality and Governance**: Implement stringent data security with scheduled audits for all staff, and strict adherence to privacy laws to ensure data integrity.

6 CONCLUSION

The analytical implications achieved for the NHS Trust has yielded a strong basis for improving the management of hospital stay durations undergoing back surgery. The scientific approach to data curation, combined with modeling techniques, has shed light to critical determinants of hospitalization, providing an empirical foundation for optimizing resources and improving patient care. Strategizing Linear Regression model, based on its interpretation simplicity and sustainable robustness, stands as a powerful justification amidst the complexity. The proposed recommendations advocate for a strategic focus of refined evaluation, better pre and post-operation care, and better-tailored pathways for patients. The proposed data collection strategy, underpinned by the integration of electronic health records and patient-centric metrics, represents a forward-thinking step towards a data-enriched healthcare environment. Ultimately the goal is readiness to apply the insights from analytics to achieve operational excellence, better patient care, and continued delivery of top-rated healthcare sustaining the NHS Trust's pledge to the highest standards of healthcare delivery.