Analyzing and Optimizing Surgical Wait Times in Nova Scotia

Table of Contents

Analyzing and Optimizing Surgical Wait Times in Nova Scotia

Chukwuemeka Onyeukwu

ABSTRACT

This project leverages data-driven insights to address critical issues in healthcare by analyzing surgical wait times in Nova Scotia and predicting the surgical wait times for the next four quarters. By examining historical patterns, regional disparities, and procedure-specific wait times—such as those for hip replacements and cancer treatments **-** we aim to identify inefficiencies, suggest optimization strategies, and support more effective, data-informed decision-making in healthcare administration.

The project involved extensive data preprocessing, visualization, and the application of some machine learning models for time series, which are **SES/Holts, FB Prophet, and Hybrid (combination of two models mentioned before).** The final model selects the best from both models, which makes this hybrid a good choice. This approach aligns with the course's emphasis on data analysis, machine learning, and decision-making, allowing us to apply theoretical knowledge to real-world healthcare challenges.

I. INTRODUCTION

Surgical wait times significantly impact the quality of healthcare delivery and patient outcomes. In Nova Scotia, procedures such as hip replacements and cancer surgeries are often associated with extended wait times, which can lead to worsened health conditions and increased system costs. This project seeks to apply data-driven methods to analyze surgical wait time patterns, identify regional disparities, and predict wait times based on demographic and health-related attributes.

By leveraging historical data, we aim to uncover inefficiencies in the system and suggest actionable strategies for optimizing healthcare efficiency in the province. Through this project, we utilize machine learning techniques to build predictive models that inform better decision-making in healthcare administration.

Ultimately, our goal is to provide data-backed insights that can support the enhancement of healthcare delivery and resource management in Nova Scotia.

II. RELATED WORK

1. Carandang et al. (2018) utilized multiple linear regression to analyze 2014–2016 surgical wait times in Nova Scotia. The study focused on the 90th percentile wait times across specialties, identifying significant predictors in general surgery, dental, ophthalmology, orthopedics, and ENT.

2. Fraser Institute (2024) reported that Nova Scotia had the longest median wait time in Canada at 56.7 weeks, with nearly equal durations between referral to specialist and specialist to treatment phases.

3. **Leong et al. (2023) proposed** ARIMA- machine learning hybrid models, specifically ARIMA-ANN, for forecasting elective surgery demand. The hybrid model demonstrated superior performance over standalone ARIMA and other machine learning models.

III. METHODOLOGY

The surgical wait time dataset is obtained from the Nova Scotia government's official website (Government of Nova Scotia, n.d.).

- data cleaning and preprocessing,
- exploratory data analysis (EDA),
- model building,
- data visualization,
- report.

Data Cleaning and Preprocessing

One of the major challenges was addressing missing values in critical columns: **'Specialty' (72%)**, **'Provider' (72%)**, and **'Procedure' (0.4%)**, which significantly affected data integrity. Our cleaning strategy included the following steps:

1. **Mapping 'Procedure' to 'Specialty'**: We created a manual mapping to assign missing 'Specialty' values based on the 'Procedure' column.
2. **Filling Missing 'Specialty' Values**: Using the mapping, we populated missing specialties where the procedure was known.
3. **Manual Assignments**: For procedures with ambiguous or unmatched specialties, we assigned 'General Surgery' based on research on those procedures
4. **Filling Missing 'Procedure' Values**: Missing procedures were imputed using the most frequent procedure associated with their respective specialty.
5. **Imputing Median Wait Times**: Missing values in **'Surgery_Median'** and **'Consult_Median'** columns were filled using the median wait time grouped by procedure. The median was chosen as the primary measure of central tendency because it is less susceptible to outliers and skewed data, offering a more robust representation of typical wait times.

Exploratory Data Analysis (EDA)

We used horizontal bar charts and boxplots to explore distributions and identify patterns in surgical wait times and specialties

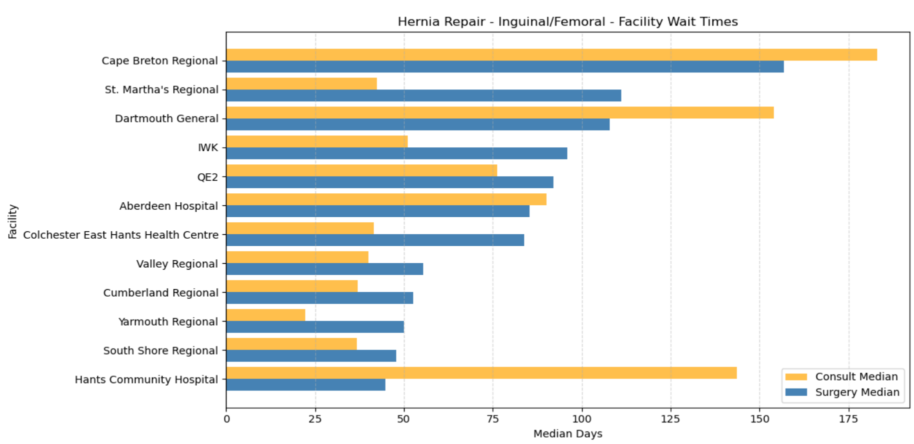

Figure 1: Hernia Repair – Inguinal/Femoral – Facility Wait Times

This **figure (1)** shows the number of waiting times for the procedure “Hernia Repair – Inguinal/Femoral”.

1. Longest Total Wait Times:

- Cape Breton Regional and Hants Community Hospital have the longest combined wait times, with consultation and surgery both taking over 150 days.

- At Cape Breton Regional, surgery wait times are extremely long, even after consultation.

2. Shorter Wait Facilities:

- South Shore Regional and Yarmouth Regional have some of the shortest wait times overall.

- These facilities are more efficient in both consultation and surgeries.

3. Surgery Wait vs. Consult Wait:

- In many cases, the surgery wait is significantly longer than consultation, especially in Cape Breton and Dartmouth General.

- At IWK and Aberdeen Hospital, consult and surgery wait times are more balanced.

4. Facility Variation:

There’s a huge disparity between facilities. For example:

- One hospital may complete both consultation and surgery in under 50 days, while others may take over 150 days just for surgery.

In other words:

- Wait time efficiency is inconsistent across hospitals.
- Facilities like South Shore Regional could be models of efficiency.
- Facilities with long waiting times (e.g., Cape Breton Regional) may need resource reallocation or process improvements.
- Patients may face delays that are heavily location-dependent, impacting on health outcomes and quality of care.

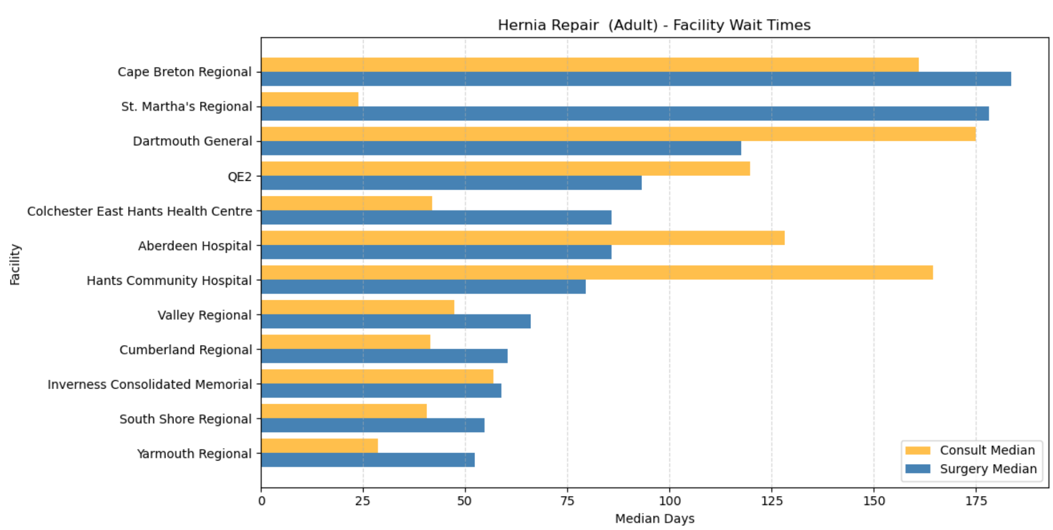

Figure 2: Hernia Repair (Adult) – Facility Wait Times

This horizontal grouped bar chart (**Figure 2**) shows “Hernia Repair (Adult) - Facility Wait Times”, as we can see:

- Cape Breton Regional and St. Martha's Regional have some of the longest total wait times, with both consult and surgery durations being high (around or over 150 days total).

- Dartmouth General also has high surgery wait times (~150 days), although its consult time is shorter.

- Hants Community Hospital has a notably long consult wait time—almost equal to or longer than its surgery time, which may indicate scheduling or resource constraints.

- QE2, Valley Regional, and South Shore Regional appear to offer relatively shorter overall wait times, especially in consultations.

- Yarmouth Regional has the shortest consult wait time among all listed facilities.

<u>In other words:</u>

- There's **significant variability** in how quickly patients can receive care depending on the facility.
- In some facilities, **consultation is the bottleneck** (e.g., Hants Community Hospital).
- In others, **surgery scheduling is the bigger delay** (e.g., Cape Breton Regional, Dartmouth General).

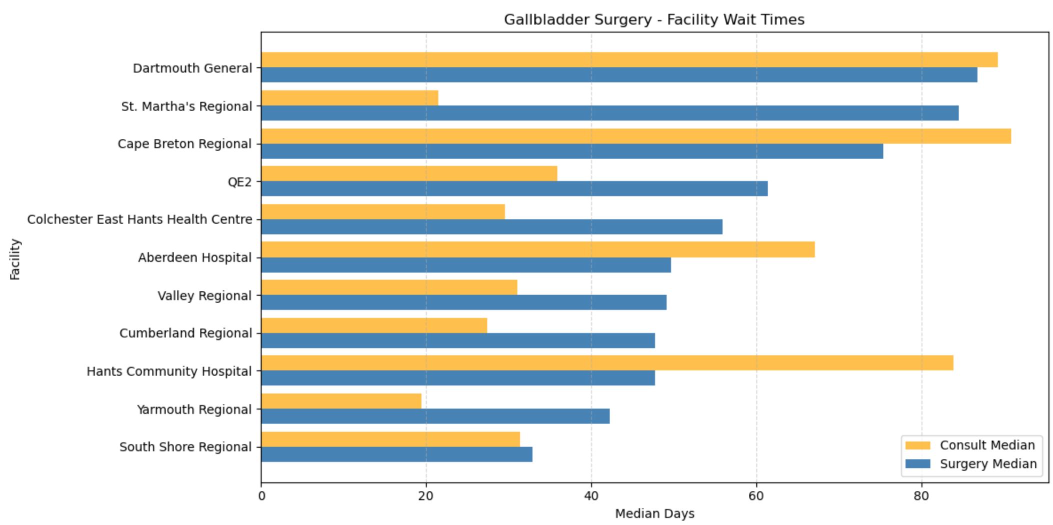

Figure 3: Gallbladder Surgery - Facility Wait Times

This visualization **(Figure 3)** presents the **median wait times** (in days) for **consultation** and **Gallbladder surgery** across different medical facilities.

- Dartmouth General and St. Martha’s Regional have the longest surgery wait times, with both nearing 90 days, even though St. Martha’s has a very short consultation wait.

- Cape Breton Regional shows balanced but high wait times for both consult and surgery, each around 80–90 days.

- Hants Community Hospital has an unusually long consultation wait (~85 days), significantly higher than its surgery wait.

- Aberdeen Hospital stands out with a very long consult time (~75 days), while its surgery wait is shorter in comparison.

- South Shore Regional and Yarmouth Regional offer the shortest overall wait times—both consult and surgery durations are below 50 days.

<u>In other words:</u>

- Facilities like **St. Martha's Regional** are efficient with consultation but experience **surgical backlogs**.
- Others like **Hants Community Hospital** and **Aberdeen Hospital** suffer from **delayed consultations**, which might delay the entire care process.
- Facilities such as **South Shore Regional** could serve as **models of efficiency**, showing consistently shorter wait times.

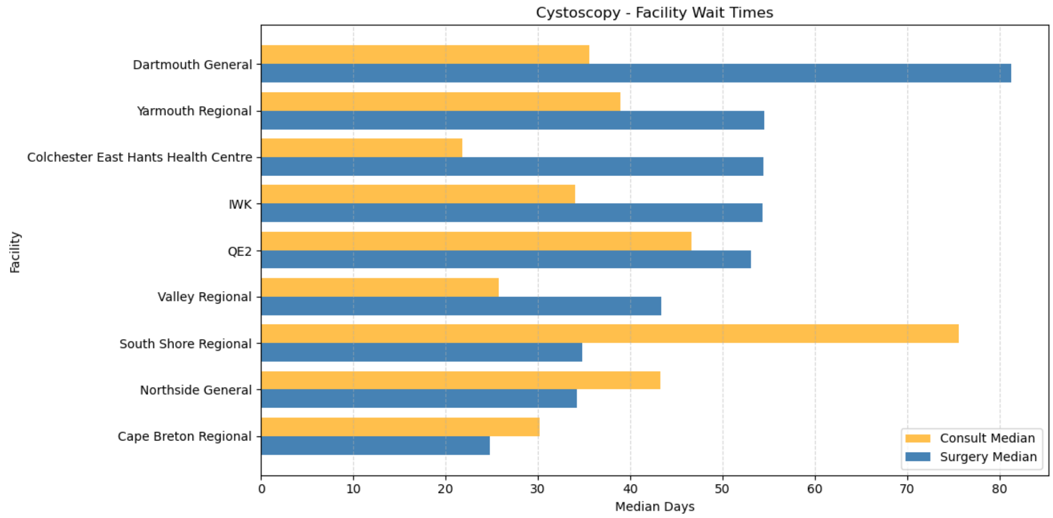

Figure 4: Cystoscopy - Facility Wait Times

This graph **(Figure 4)** shows the median wait times (in days) for consultation and Cystoscopy procedures across different medical facilities.

- Dartmouth General has the longest surgery wait time, exceeding 80 days, despite a moderate consultation wait (~35 days).
- South Shore Regional shows an unusual pattern, with an extremely long consultation wait (~78 days) compared to a much shorter surgery wait (~35 days).
- Colchester East Hants Health Centre and Valley Regional report some of the shortest consult wait times, both around 20–25 days.
- Cape Breton Regional and Yarmouth Regional have relatively balanced wait times for both consult and surgery, with each sitting around 30–40 days.
- QE2, IWK, and Northside General show moderate consult wait times (30–50 days) and slightly longer surgery waits in the 50–60 day range.

In other words:

- Facilities like **Dartmouth General** experience significant surgical delays, indicating potential procedure backlogs.
- Others like **South Shore Regional** suffer from delayed consultations, which may stall patient progression through the care pathway.
- Facilities such as **Colchester East Hants** and **Valley Regional** demonstrate more efficient front-end access, with shorter consult waits that could improve patient flow.

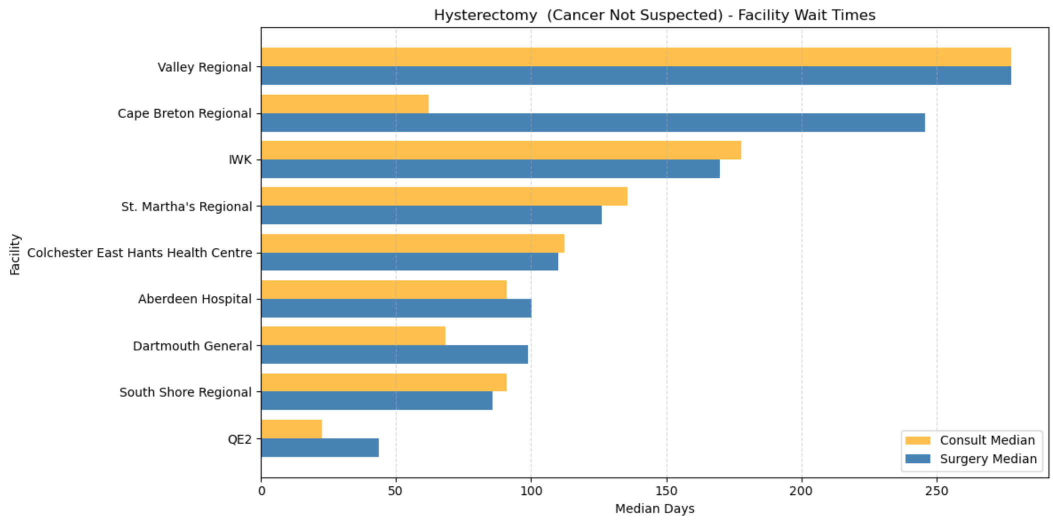

Figure 5: Hysterectomy (Cancer not Suspected) - Facility Wait Times

The horizontal grouped bar chart **(Figure 5)** presents the median wait times (in days) for consultation and Hysterectomy procedures (when cancer is not suspected) across different medical facilities.

- Valley Regional has the longest overall wait times, with both consultation and surgery exceeding 250 days, indicating a substantial delay at every stage of care.
- Cape Breton Regional also shows a significant surgical wait (~250 days), although its consultation wait is much shorter (~60 days).
- IWK and St. Martha’s Regional report long waits for both consult and surgery, with times ranging from ~150 to 180 days.
- Colchester East Hants Health Centre and Aberdeen Hospital have mid-range wait times, with both stages hovering around 90–120 days.
- Dartmouth General has a relatively short consult wait (~60 days) but a longer surgery wait (~100 days).
- South Shore Regional shows slightly higher consult wait times (~90 days) compared to its surgical wait (~80 days).
- QE2 has the shortest wait times across the board, with consultation under 30 days and surgery under 50 days.

In other words:

- Facilities like **Valley Regional** face systemic delays across both consultation and surgery stages.
- Others like **Cape Breton Regional** are efficient in consultation but have long surgical backlogs.
- Facilities such as **QE2** represent high efficiency and could serve as models for improving access to care.

The boxplot below **(Figure 6)** shows:

Table 1

Gallbladder Surgery	Hernia Repair (Adult)	Gastrointestinal Tract Surgery	Cystoscopy	Hysterectomy (Cancer not Suspected)	Hernia Repair—Inguinal/Femoral
- Median wait time: Around 50 days. - IQR: Relatively tight, suggesting consistent scheduling. - Outliers: Some extreme values (>100 days) hint at occasional delays.	- Median wait time: Around 75 days. - Spread: Wide distribution with a long upper tail *(many patients waited over 150+ days).* - Skew: Positive, indicating a backlog for some cases.	Shortest median wait: Around 35–40 days. - Tight distribution: Suggests high scheduling efficiency. - Few outliers: Minimal disruption or bottlenecks.	- Median: Around 40–45 days. - Tight IQR, but slightly higher outlier presence than GI surgery. Reliable delivery, though some cases exceed 100 days.	- Median wait: Around 110–120 days. - Highest variability: Extremely widespread, from 30 to 360 days. Implication: Possible triaging issues or resource constraints.	Median: Around 70 days. - Similar to adult hernia repair, with significant upper outliers. Suggestion: Similar systemic pressures affecting both hernia categories.
- Hysterectomy (Cancer Not Suspected): Red flag for delays. Needs closer policy or operational review. - GI Tract Surgery: Most efficiently managed. - Hernia Repairs: Frequently delayed, possibly due to lower urgency classification, but may require a triage rethink.					

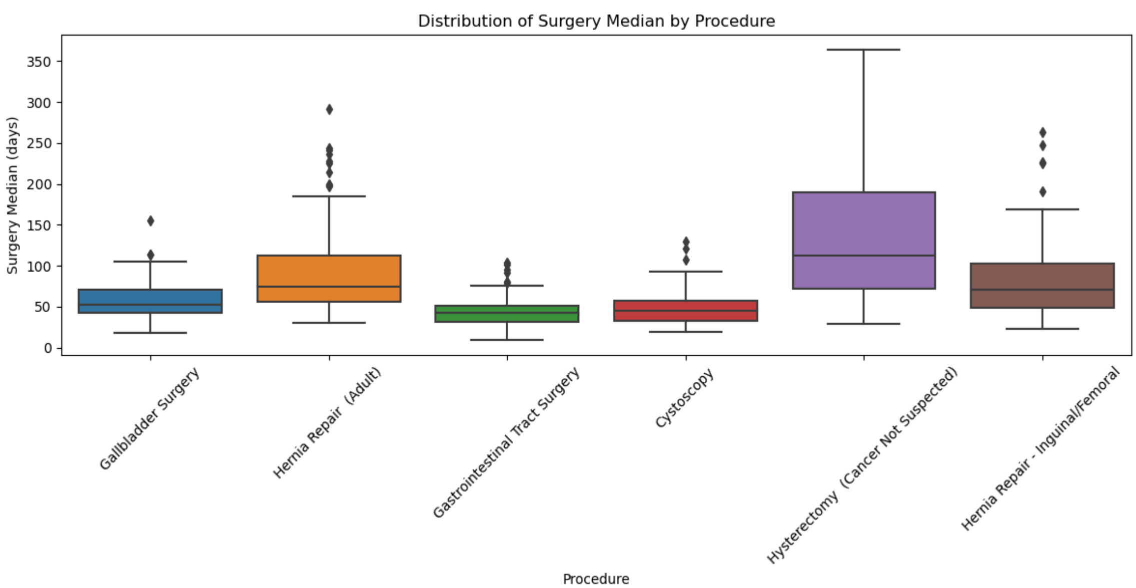

Figure 6: Distribution of Surgery Median by Procedure

Model Building

- **Feature Selection:** Given the extensive number of procedures available, we identified the top six most commonly performed procedures within each specialty. These were selected as representative features for our model development and analysis

  Three forecasting models were trained:

- SES/Holt
- Fb Prophet
- Hybrid Model (combination of SES/Holt and Prophet)

Model performance was evaluated using Mean Absolute Error (MAE).

Tools Used

- Python: Programming and analysis
- Jupyter Notebooks: Interactive coding environment
- Pandas: Data manipulation
- Matplotlib & Seaborn: Data visualization
- Scikit-learn: Machine learning

IV. RESULTS

We evaluated three models: SES/Holt, FB Prophet, and Hybrid model (combination of SES/Holt, and FB Prophet). The dataset was split into training and testing sets. Evaluation metrics used were Mean Absolute Error (MAE).

Main steps:

Step 1: Standardized and Convert the period format to fit the model:

Standardize period formats: Ensure that periods like "q1_2022" or "2022_q1" are consistent in format (e.g., "2022_Q1"). Sort periods chronologically: Sort the periods correctly, handling different period formats.

Convert periods to datetime: Transform periods into a timestamp that allows time series analysis (e.g., "2022_Q1" to 2022-01-01).

Step 2: Building the Time Series Matrix:

2.1. SES/Holt Model

Main steps:

Step 1: Test for Stationarity:
Use the Augmented Dickey-Fuller (ADF) test to test for stationarity.

- Null hypothesis: the series is non-stationary.
- If p-value < 0.05, the series is stationary.

Table 2. Stationarity Test

	Procedure	ADF Statistic	p-value	Stationary	Used Lag	N Obs
0	Cystoscopy	-4.697411	0.000085	True	0	8
1	Gallbladder Surgery	-1.80427	0.378344	False	2	6
2	Gastrointestinal Tract Surgery	-2.473817	0.121976	False	1	7
3	Hernia Repair (Adult)	-3.750837	0.003452	True	0	8
4	Hernia Repair - Inguinal/Femoral	-1.890102	0.336732	False	0	8
5	Hysterectomy (Cancer Not Suspected)	-1.728697	0.416391	False	0	8

Step 2: Model Selection:

- For stationary series, use **Simple Exponential Smoothing (SES)**.
- For non-stationary series, use **Holt's Linear Trend**.

Step 3: Forecasting:

- Fit the selected model to each procedure and generate forecasts for the next 4 quarters.

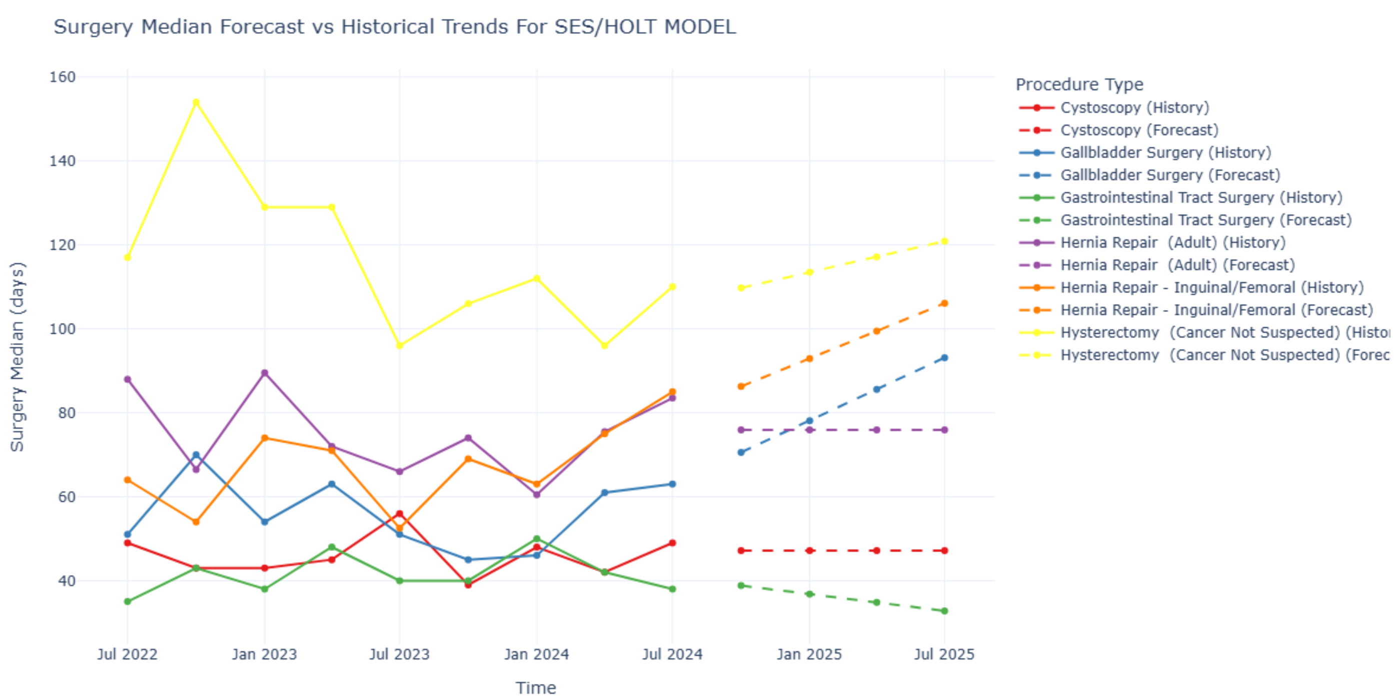

Figure 7. Surgery Median Forecast vs Historical Trends For SES/Holt

The figure above presents the forecast for each procedure's surgery wait time for the next four quarters (2024-2025)

Step 4 : Model Evaluation

Table 3. SES/Holt Evaluation

	Procedure	MAE
0	Cystoscopy	3.500000
2	Gastrointestinal Tract Surgery	8.614188
5	Hysterectomy (Cancer Not Suspected)	9.783690
3	Hernia Repair (Adult)	11.973862
4	Hernia Repair - Inguinal/Femoral	15.016780
1	Gallbladder Surgery	23.817850

Cystoscopy has the **lowest MAE at 3.5**, indicating highly accurate predictions with minimal deviation from actual values.

Gastrointestinal Tract Surgery (MAE = 8.61) and **Hysterectomy (Cancer Not Suspected)** (MAE = 9.78) show **moderate error**, suggesting acceptable forecast reliability.

2.2. FB Prophet Model

Main steps:

Step 1. Model Fitting and Forecasting (Prophet):

- Fit the Prophet model to the data.
- Use Prophet to predict the next 4 quarters by creating a future data frame and generating forecasts.

Table 4. Prophet Prediction

	Forecast_Quarter	Cystoscopy	Gallbladder Surgery	Gastrointestinal Tract Surgery	Hernia Repair (Adult)	Hernia Repair - Inguinal/Femoral	Hysterectomy (Cancer Not Suspected)
0	2024-09-30	37.640417	18.623766	37.439833	70.986496	84.006054	16.598589
1	2024-12-31	54.613586	36.973845	61.931916	26.690086	52.072131	67.817481
2	2025-03-31	61.320497	70.222147	62.755605	11.230516	69.581143	17.995870
3	2025-06-30	76.065738	50.141006	53.583615	-13.745850	31.911361	-69.945825

- Insight: Hernia Repair (Adult) shows a sharp decline across quarters, signaling a potential drop in demand or data/modeling error, while Cystoscopy consistently rises, indicating increasing wait times and possible capacity strain.

Step 2. Model Evaluation (MAE):

- Calculate the Mean Absolute Error (MAE) for Prophet's forecast by comparing predicted values with actual test values.

Table 5. MAE Prophet

	Procedure	MAE
0	Cystoscopy	3.461212
2	Gastrointestinal Tract Surgery	8.577686
5	Hysterectomy (Cancer Not Suspected)	9.566783
4	Hernia Repair - Inguinal/Femoral	15.167972
1	Gallbladder Surgery	18.502166
3	Hernia Repair (Adult)	20.264359

Insight: Cystoscopy has the lowest MAE (3.46), suggesting that the model provides highly accurate forecasts for this procedure. Hysterectomy (Cancer Not Suspected) have MAE o 9.57. These indicate that the forecast model performs moderately well but could benefit from refinement, especially for procedures with more variability. Hernia Repair (Adult) has the highest (20.26), revealing poor model performance

2.3. Hybrid model (SES/Holt and FB Prophet)

- **Forecasting with Two Models (SES/Holt and Prophet)**:

- The function applies two forecasting models (Simple Exponential Smoothing/Holt and Prophet) to predict future values for each procedure.
- Then it selects the best model for the procedure based on the best mae

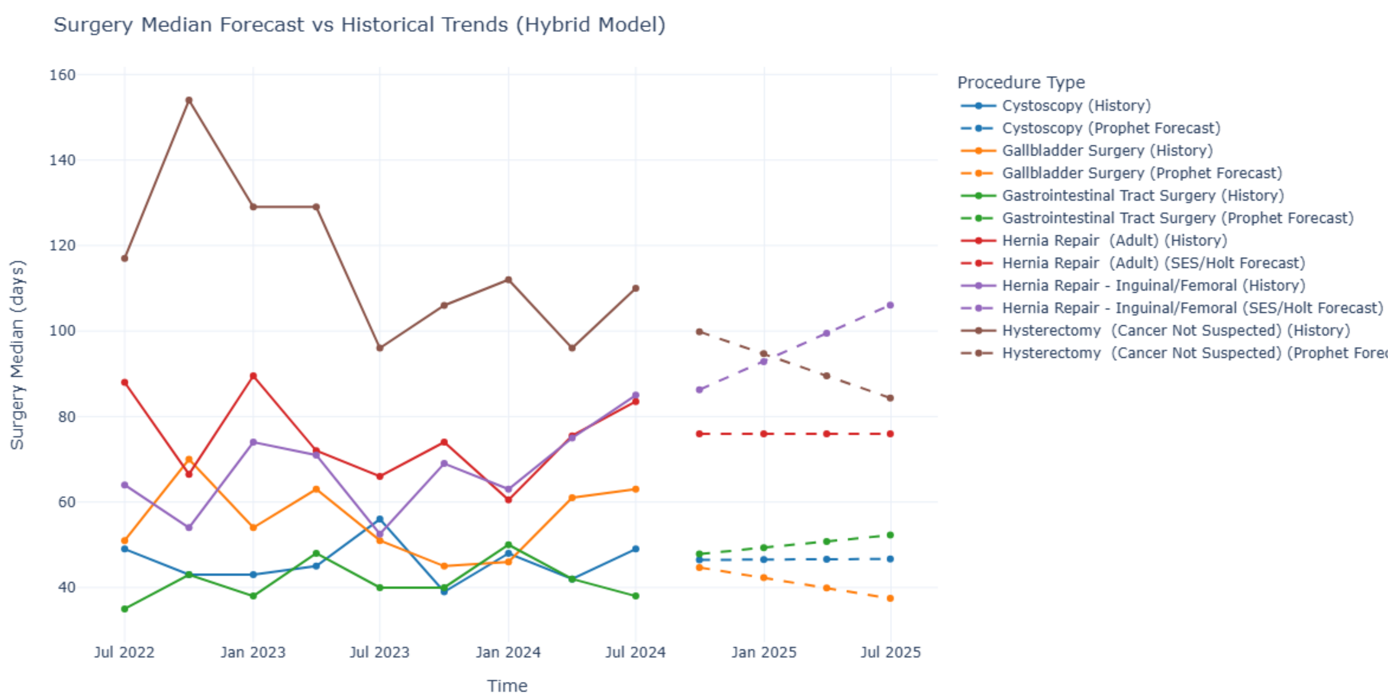

Figure 8 Hybrid Model

<u>MAE Comparison:</u>

Table 6. MAE Comparison

Procedure	**SES/Holt MAE**	**Prophet MAE**	Hybrid MAE	**Best Model**
Cystoscopy	3.50	3.46	3.46	Prophet
Gastrointestinal Tract Surgery	8.61	8.58	8.58	Prophet
Hysterectomy (Cancer Not Suspected)	9.78	9.57	9.57	Prophet
Hernia Repair (Adult)	11.97	20.26	11.97	SES/Holt
Hernia Repair - Inguinal/Femoral	15.02	15.17	15.02	SES/Holt
Gallbladder Surgery	23.81	18.50	18.50	Prophet

The table above presents the model performance:

- **Prophet** works best for Cytoscopy, Gallbladder Surgery and Hysterectomy (Cancer Not Suspected) with low MAE, capturing long-term trends well.

- **SES/Holt** is more effective for Gastrointestinal Tract Surgery, Hernia Repair (Adult), and Hernia Repair - Inguinal/Femoral, showing consistent, linear trends.

V. DISCUSSION

In this study, we evaluated three forecasting models—SES/Holt, FB Prophet, and a hybrid model (combination of SES/Holt and Prophet)—to predict surgery wait times across various medical procedures. Each model's performance was evaluated based on the Mean Absolute Error (MAE), which is a key metric for measuring forecasting accuracy.

Significance of the results shows that : **Prophet is the best overall model** for 4 out of 6 procedures, offering the lowest MAE in most cases. This suggests it handles seasonality and trend components well in surgery wait time forecasting. **SES/Holt performs better for Hernia Repairs**, especially for the *Adult* and *Inguinal/Femoral* categories. This implies these procedures may follow more linear or short-term patterns where exponential smoothing is more effective. **Cystoscopy shows excellent predictability**, with the lowest MAE across all models (3.46), indicating stable and consistent patterns in wait times.

VI. CONCLUSION

This analysis demonstrates the utility of both SES/Holt and Prophet for forecasting surgery wait times in the healthcare context.

1. **Model Selection is Crucial**: The choice of forecasting model should be tailored to the nature of the data. For procedures with stable, linear trends, SES/Holt is effective, whereas Prophet excels in forecasting procedures with complex, non-linear trends or seasonal components. **In summary,**Model performance varies by procedure. Prophet dominates overall but SES/Holt is crucial for certain trend-stable surgeries. Custom model selection per procedure enhances predictive precision and supports better healthcare planning
2. **Implications for Capacity Planning**: The findings emphasize the need for capacity planning and system-level interventions in the healthcare sector. Given the projected stability or worsening of wait times for certain procedures, hospitals and healthcare providers should prioritize resource allocation to improve wait times and patient care. Specifically, interventions should be focused on high-demand procedures that exhibit a trend toward longer wait times.

   Recommendations for Future Work:

   - Incorporate external data (e.g., healthcare policies, seasonal fluctuations, or hospital capacity) into the forecasting models to enhance their predictive power.
   - Investigate more advanced hybrid forecasting techniques, combining multiple models and external variables, to further improve forecasting accuracy.

In conclusion, while forecasting surgical wait times is complex, the use of advanced time series models like SES/Holt and Prophet can provide valuable insights. These insights can be leveraged to inform healthcare decision-making, optimize resource allocation, and enhance patient care. By continuously improving these models and incorporating external factors, we can further enhance the accuracy and reliability of the predictions, leading to better-informed capacity planning and improved healthcare delivery.

VII. REFERENCES

Carandang, C., Horne, G., Wells, W., & Stokes, C. (2018). Analysis of surgical wait times in Nova Scotia. *ResearchGate*. https://www.researchgate.net/publication/323837798_Analysis_of_Surgical_Wait_Times_in_Nova_Scotia

Leong, X. Y., Jajo, N., Peiris, S., & Khadra, M. (2023). Forecasting elective surgery demand using ARIMA-machine learning hybrid model. *European Journal of Artificial Intelligence and Machine Learning*, 2(3), 1–7. https://doi.org/10.24018/ejai.2023.2.3.19

Moir, M., & Barua, B. (2023). *Waiting your turn: Wait times for health care in Canada, 2023*. Fraser Institute. https://www.fraserinstitute.org/studies/waiting-your-turn-wait-times-for-health-care-in-canada-2023

Government of Nova Scotia. (n.d.). *Surgical wait times*. https://waittimes.novascotia.ca/

VIII. Appendices

Libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import re from sklearn.metrics

import mean_absolute_error from statsmodels.tsa.stattools import adfuller

df = pd.read_csv("Surgical_Wait_Times.csv")

df.describe()

df.isnull().sum()

Data Cleaning

procedure_to_specialty = (df[~df['Specialty'].isna()] .groupby('Procedure')['Specialty'] .agg(lambda x: x.mode()[0]) # use mode in case of duplicates)

def fill_specialty(row): if pd.isna(row['Specialty']): return procedure_to_specialty.get(row['Procedure'], None) return row['Specialty']

df['Specialty'] = df.apply(fill_specialty, axis=1)

procedure_to_specialty = { 'Cardiac Valve Replacement': 'Cardiac Surgery', 'Congenital Heart Surgery': 'Cardiac Surgery', 'Coronary Artery Bypass Graft': 'Cardiac Surgery', 'Pacemaker Insertion': 'Cardiac Surgery', 'Internal Defib Insertion / Loop Recorder': 'Cardiac Surgery', 'Cardioversion': 'Cardiac Surgery',

'Hip Replacement': 'Orthopaedic Surgery',
'Knee Replacement': 'Orthopaedic Surgery',
'Knee Replacement Revision': 'Orthopaedic Surgery',
'Hip Replacement Revision': 'Orthopaedic Surgery',
'Knee Replacement - Partial': 'Orthopaedic Surgery',
'Meniscectomy': 'Orthopaedic Surgery',
'Shoulder Arthroscopy': 'Orthopaedic Surgery',
'Knee Scope': 'Orthopaedic Surgery',
'Knee Scope with ACL Repair': 'Orthopaedic Surgery',
'Shoulder Surgery': 'Orthopaedic Surgery',
'Shoulder Arthroplasty': 'Orthopaedic Surgery',
'Foot - Bunionectomy': 'Orthopaedic Surgery',
'Foot - Osteotomy': 'Orthopaedic Surgery',
'Ankle - Arthrodesis': 'Orthopaedic Surgery',
'Ankle - Arthroscopy': 'Orthopaedic Surgery',
'Ankle - Arthroplasty': 'Orthopaedic Surgery',
'Patella Surgery': 'Orthopaedic Surgery',

```
'Hip Arthroscopy': 'Orthopaedic Surgery',
'Osteotomy': 'Orthopaedic Surgery',
'Hand / Upper Extremity': 'Orthopaedic Surgery',
'Tendon / Ligament Repair': 'Orthopaedic Surgery',
'Foot - Arthrodesis': 'Orthopaedic Surgery',
'Foot - Reconstruction': 'Orthopaedic Surgery',
'Resurfacing - Knee': 'Orthopaedic Surgery',
'Manipulation of Extremity': 'Orthopaedic Surgery',
'Orthopaedic Trauma': 'Orthopaedic Surgery',

'Esophagectomy / Esophageal Surgery': 'General Surgery',
'Bowel Resection - Open': 'General Surgery',
'Bowel Resection - Laparoscopic': 'General Surgery',
'Bowel Resection': 'General Surgery',
'Pancreas Surgery': 'General Surgery',
'Liver Surgery': 'General Surgery',
'Gallbladder Surgery': 'General Surgery',
'Sternotomy / Thoracotomy': 'Thoracic Surgery',
'Peritoneal Lavage or Catherization': 'General Surgery',
'Appendectomy': 'General Surgery',  # If exists
'Nissen Fundoplication - Laparoscopic': 'General Surgery',
'Nissen Fundoplication - Open': 'General Surgery',
'Gastric Surgery': 'General Surgery',
'Gastric Surgery - Laparoscopic': 'General Surgery',
'Whipple Procedures': 'General Surgery',
'Splenic Surgery': 'General Surgery',

'Prostate Surgery': 'Urology',
'Prostatectomy': 'Urology',
'Male Circumcision  (Adult)': 'Urology',
'Male Circumcision  (Pediatric)': 'Urology',
'Male Reproductive System Surgery': 'Urology',
'Bladder Cancer Surgery': 'Urology',
'Kidney / Upper Urinary Tract': 'Urology',
'Kidney Stone Surgery': 'Urology',
'Kidney Removal': 'Urology',
'Bladder Surgery': 'Urology',
'Prostate Cancer Surgery': 'Urology',
'AV Fistula Creation/Closure': 'Urology',  # Can vary; assume Urology
'Male Incontinence Surgery': 'Urology',

'Cataract Surgery': 'Ophthalmology',
'Vitreoretinal Diseases': 'Ophthalmology',
'Glaucoma (Eye Pressure Lowering Surgery)': 'Ophthalmology',
'Cornea and External Disease': 'Ophthalmology',
'Oculo Plastic Surgery': 'Ophthalmology',
'Tear Duct Surgery': 'Ophthalmology',
'Lacrimal Duct Probing': 'Ophthalmology',
'Strabismus Surgery  (Adult)': 'Ophthalmology',
'Strabismus Surgery  (Pediatric)': 'Ophthalmology',

'Ear Tubes  (Pediatric)': 'Otolaryngology (ENT)',
'Ear Tubes  (Adult)': 'Otolaryngology (ENT)',
'Tonsillectomy and/or Adenoidectomy  (Pediatric)': 'Otolaryngology (ENT)',
'Tonsillectomy and/or Adenoidectomy  (Adult)': 'Otolaryngology (ENT)',
```

'Surgery for Hearing or Balance Problems of the Ear': 'Otolaryngology (ENT)',
'Nose Surgery - Reconstructive or Aesthetic': 'Otolaryngology (ENT)',
'Sinus Surgery': 'Otolaryngology (ENT)',
'Septoplasty': 'Otolaryngology (ENT)',
'Functional Endoscopic Sinus Surgery (FESS)': 'Otolaryngology (ENT)',
'Laryngology and Vocal Cord Surgery': 'Otolaryngology (ENT)',

'Craniotomy or Craniectomy': 'Neurosurgery',
'Brain Surgery (Adult)': 'Neurosurgery',
'Brain Surgery (Pediatric)': 'Neurosurgery',
'Spinal Cord Stimulation Surgery': 'Neurosurgery',
'Nerve or Brain Stimulation': 'Neurosurgery',
'Cervical Spine (Neck) Surgery': 'Neurosurgery',
'Nerve Surgery': 'Neurosurgery',

'Oral and Maxillofacial': 'Oral Maxillofacial',
'Maxillofacial Deformity Surgery': 'Oral Maxillofacial',
'Temperomandibular Joint Surgery (TMJ)': 'Oral Maxillofacial',
'Dental Extractions and Restorations (Adult)': 'Dental',
'Dental Extractions and Restorations (Pediatric)': 'Dental',

'Breast Cancer Surgery': 'General Surgery',
'Breast Reconstruction': 'Plastic Surgery',
'Breast Augmentation': 'Plastic Surgery',
'Breast Reduction Surgery': 'Plastic Surgery',
'Aesthetic Surgery': 'Plastic Surgery',
'Otoplasty': 'Plastic Surgery',
'Cleft Lip/Palate (Pediatric)': 'Plastic Surgery',
"Palmar Fascia Excision for Dupuytren's Disease": 'Plastic Surgery',

'Lung Cancer Surgery': 'Thoracic Surgery',
'Lung Surgery': 'Thoracic Surgery',
'Video-Assisted Thoracic Surgery': 'Thoracic Surgery',
'Thoroscopy / Pleuroscopy': 'Thoracic Surgery',

'Aneurysm Repair': 'Vascular Surgery',
'Carotid Endarterectomy': 'Vascular Surgery',
'Endarterectomy': 'Vascular Surgery',
'Angioplasty': 'Vascular Surgery',
'Vascular Bypass Surgery': 'Vascular Surgery',
'Varicose Veins': 'Vascular Surgery',

}

df['Specialty'] = df.apply(lambda row: procedure_to_specialty.get(row['Procedure'], row['Specialty']) if pd.isnull(row['Specialty']) else row['Specialty'], axis=1) df

fill missing values in the surgery_median column based on the median for each procedure

rocedure_medians = df.groupby('Procedure')['Surgery_Median'].median()

def fill_missing_median(row): if pd.isnull(row['Surgery_Median']): return procedure_medians.get(row['Procedure'], row['Surgery_Median']) return row['Surgery_Median']

df['Surgery_Median'] = df.apply(fill_missing_median, axis=1)

Grouped median imputation by 'procedure' (only if enough data points)

data['Consult_Median'] = data.groupby('Procedure')['Consult_Median'].transform(lambda x: x.fillna(x.median()))

data['Surgery_Median'] = data.groupby('Procedure')['Surgery_Median'].transform(lambda x: x.fillna(x.median()))

Exploratory Data Analysis

top_procedures = data['Procedure'].value_counts().head(6).index.tolist()

df_selected = data[data['Procedure'].isin(top_procedures)]

facility_grouped = df_selected.groupby(['Procedure', 'Facility'])[['Consult_Median', 'Surgery_Median']].mean().reset_index()

Plot for each procedure in top_procedures

for proc in top_procedures: # Filter and sort data proc_data = facility_grouped[facility_grouped['Procedure'] == proc].sort_values(by='Surgery_Median', ascending=False)

```
# Setup
x_labels = proc_data['Facility']
x_pos = np.arange(len(proc_data))
bar_height = 0.4

# Create figure
plt.figure(figsize=(12, 6))

# Side-by-side bars
plt.barh(x_pos - bar_height/2, proc_data['Consult_Median'], height=bar_height,
         color='orange', alpha=0.7, label='Consult Median')
plt.barh(x_pos + bar_height/2, proc_data['Surgery_Median'], height=bar_height,
         color='steelblue', label='Surgery Median')

# Add y-axis labels at the correct positions
plt.yticks(x_pos, x_labels)

# Titles and labels
plt.title(f'{proc} - Facility Wait Times')
plt.xlabel('Median Days')
plt.ylabel('Facility')
plt.legend()
plt.grid(axis='x', linestyle='--', alpha=0.5)
plt.gca().invert_yaxis()  # Highest at top
plt.tight_layout()
plt.show()
```

df_trend = (df_selected.groupby(['Period', 'Procedure'])['Surgery_Median'] .median() .reset_index())

Step 2: Standardize inconsistent period formats (e.g., q1_2022, 2022-q1, etc.)

Goal: Make sure periods like "q1_2022" or "2022-q1" are reformatted consistently to "2022_Q1" def standardize_period(p): p = str(p) match = re.match(r"(q[1-4])[-*]?(\d{4})", p, re.IGNORECASE) if match: return f"{match.group(2)}*{match.group(1).upper()}" match = re.match(r"(\d{4})[-*]?(q[1-4])", p, re.IGNORECASE) if match: return f"{match.group(1)}*{match.group(2).upper()}" return p # fallback if not matched

df_trend['Period'] = df_trend['Period'].apply(standardize_period)

Step 3: Sort periods chronologically using a smart key

def smart_sort_key(period_str): year, quarter = period_str.split('_') return (int(year), int(quarter.replace('Q', '').replace('q', '')))

ordered_periods = sorted(df_trend['Period'].unique(), key=smart_sort_key) df_trend['Period'] = pd.Categorical(df_trend['Period'], categories=ordered_periods, ordered=True)

Step 4: Plot line chart of surgery median trend

plt.figure(figsize=(14, 6))

sns.lineplot(data=df_trend, x='Period', y='Surgery_Median', hue='Procedure', marker='o')

plt.title('Surgery Median Wait Time Over Time by Procedure')

plt.xlabel('Period')

plt.ylabel('Surgery Median (days)')

plt.xticks(rotation=45)

plt.tight_layout()

plt.show()

Convert standardized periods to Pandas quarterly timestamps

def period_to_datetime(p): parts = p.lower().split('_') if 'q' in parts[0]: quarter = int(parts[0].replace('q', '')) year = int(parts[1]) else: year = int(parts[0]) quarter = int(parts[1].replace('q', '')) return pd.Timestamp(f'{year}-Q{quarter}')

Apply transformation and aggregate median values

df_ts = (df_selected.copy() .assign(period_dt=lambda x: x['Period'].apply(standardize_period).apply(period_to_datetime)) .groupby(['period_dt', 'Procedure'])['Surgery_Median'] .median() .reset_index() .sort_values(by='period_dt'))

Pivot: rows = periods, columns = procedures, values = median surgery wait time

ts_pivot = df_ts.pivot(index='period_dt', columns='Procedure', values='Surgery_Median') #print(ts_pivot)

from IPython.display import display display(ts_pivot)

Check for Stationarity

def adf_summary(series): series = series.dropna() result = adfuller(series) return { "ADF Statistic": result[0], "p-value": result[1], "Stationary": result[1] < 0.05, "Used Lag": result[2], "N Obs": result[3] }

Apply ADF test to each procedure's series

adf_results = {col: adf_summary(ts_pivot[col]) for col in ts_pivot.columns}

Format results as a DataFrame

adf_df = pd.DataFrame(adf_results).T adf_df.index.name = "Procedure" adf_df.reset_index(inplace=True)

from IPython.display import display display(adf_df)

SES model

from statsmodels.tsa.holtwinters import SimpleExpSmoothing, Holt

Forecast horizon

forecast_periods = 4

Store forecasts

forecast_results = {}

Loop over each procedure

for proc in ts_pivot.columns: series = ts_pivot[proc].dropna()

```
#if proc in ['Cystoscopy', 'Gastrointestinal Tract Surgery', 'Hernia Repair  (Adult)']:
if proc in ['Cystoscopy','Hernia Repair  (Adult)']:
    # Stationary → SES
    model = SimpleExpSmoothing(series).fit()
else:
    # Non-stationary → Holt's Linear
    model = Holt(series).fit()

forecast = model.forecast(forecast_periods)
forecast_results[proc] = forecast
```

Combine forecasts into a DataFrame

forecast_df = pd.DataFrame(forecast_results)

from IPython.display import display

display(forecast_df)

Confirm forecast periods

forecast_df.index.name = 'Forecast_Quarter' forecast_df.reset_index(inplace=True) forecast_df.head()

FB Prophet

from prophet import Prophet from pandas.tseries.offsets import QuarterEnd

Forecast horizon

forecast_periods = 4 forecast_results = {}

Loop over each procedure column

for proc in ts_pivot.columns: series = ts_pivot[proc].dropna().reset_index() series.columns = ['ds', 'y'] # Prophet requires this format

```
# Initialize and fit the Prophet model
model = Prophet()
model.fit(series)

# Create future dataframe (next 4 quarters)
last_date = series['ds'].max()
future = model.make_future_dataframe(periods=forecast_periods, freq='Q')

# Forecast
forecast = model.predict(future)

# Extract only the forecasted yhat values for the future periods
forecast_tail = forecast[['ds', 'yhat']].tail(forecast_periods).copy()
forecast_tail.set_index('ds', inplace=True)
forecast_results[proc] = forecast_tail['yhat']
```

Combine into single DataFrame

forecast_df = pd.DataFrame(forecast_results) forecast_df.index.name = 'Forecast_Quarter' forecast_df.reset_index(inplace=True)

Display forecast

from IPython.display import display display(forecast_df)