HINF 5102

Data Management in Healthcare

Final Project

Resource Allocation Optimization in Healthcare

Anugrah Suresh Kumar Nair

Shagun

Akshay Venkatesh

Astha Gungun Patel

Introduction

This project addressed a critical challenge in healthcare operations: optimizing patient flow and ensuring the efficient use of physicians to reduce operational costs while maintaining high-quality care. Through data-driven analysis, we explored patterns in patient visits, identified bottlenecks in resource allocation, and developed forecasting tools to support strategic decision-making.

To complement this task, we designed an interactive Tableau dashboard that captures doctor utilization and revenue by time and services. We built a forecasting model from the ARIMA algorithm using Python to forecast peak-demand times. All these findings were encapsulated in a comprehensive report and presentation that included key findings, visualization, and actionable recommendations. We named our system **Smartstaff**, which transformed raw data into meaningful and final recommendations.

The technologies used in this project included Tableau for visualization, Excel for aggregating data and summarizing statistics, and Python for predictive analysis and time-series modeling. The technologies were well-suited to complement one another in developing a robust decision-support system for optimizing resources in the hospital.

Data Cleaning and Standardization

- Raw data was obtained from Kaggle, which includes patient visit data, consultation duration, and doctor availability.
- The database has timestamped entries for the progression of patients from entry to the end of consultation.
- Variables include Entry Time, Post-Consultation Time, Completion Time, Doctor Type, and Financial Class.

Data preprocessing is the key to any analysis task. Post-consultation time, patient entry time, and completion time in this study are converted to datetime type from Python Pandas library. Measures that are derived from others, like consultation time and post-consultation time, are extracted to quantify consultation efficiency.

Data cleaning techniques include searching and addressing missing values. Incorrect data points are addressed through proper imputation or removal. Time features like hours, days, and months are calculated for further analysis based on time-series.

The following steps were taken to prepare the dataset for analysis:

- Date Formatting: Made the Date column uniform in MM-DD-YY format to facilitate proper chronological ordering.
- Numeric Data Validation: Ensured columns Medication Revenue, Lab Cost, and Consultation Revenue are labeled as numeric, removing any non-numeric characters.
- Converted Entry Time, Post-Consultation Time, and Completion Time Excel Time (hh:mm:ss) format.
- Generated Colums Consultation Duration and Post-Consultation Duration by finding time differences between comparable columns.
- Consistency of Categorical Data: Normalized values in Doctor Type, Financial Class, and Patient Type to ensure consistency or avoiding errors.
- Verification of Patient ID: Confirmed that Patient ID is unique and correctly mapped for each patient visit.

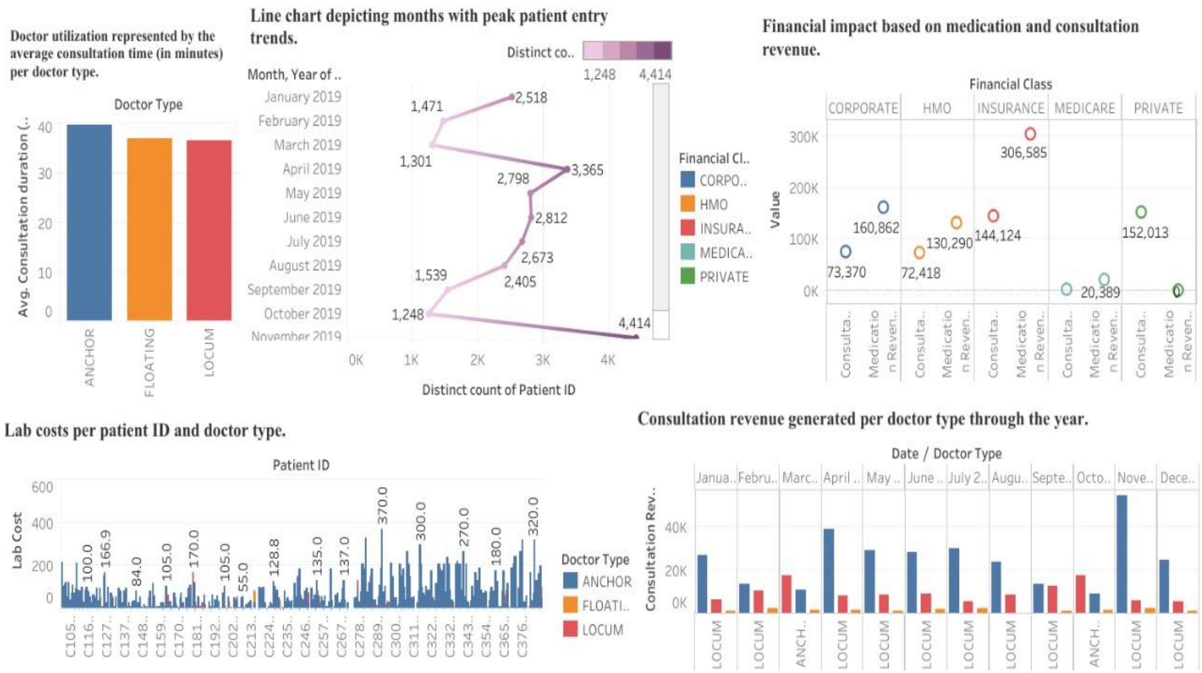
- Missing Data Handling: Replaced missing values with 0 or left blank based on context, ensuring gaps in crucial data are not created.

These processes ensured that the dataset is clean and ready for further analysis.

Exploratory Data Analysis (EDA)

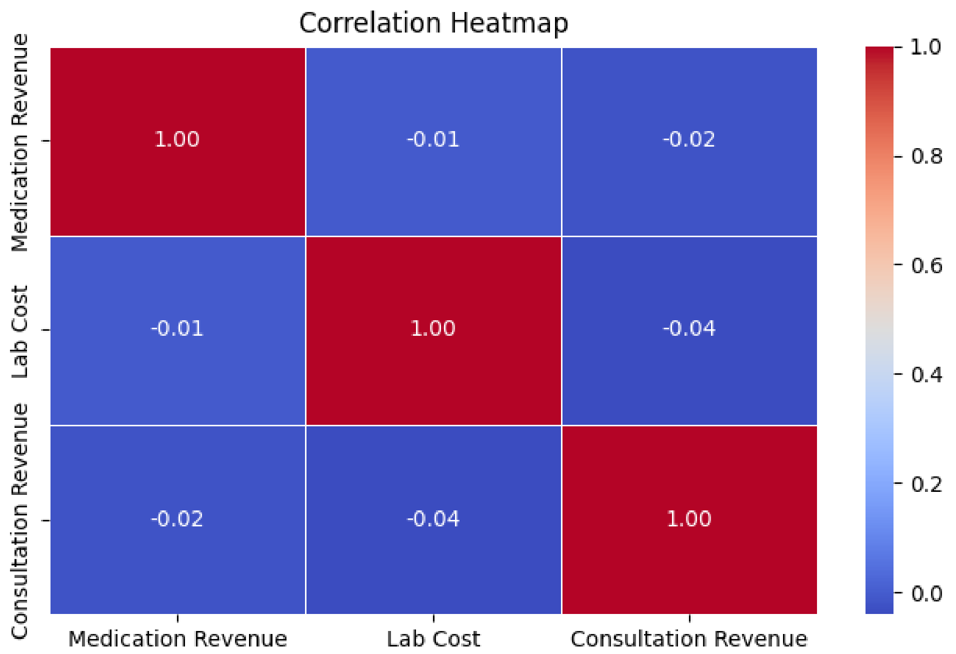
Exploratory Data Analysis (EDA) was conducted to detect patterns of patient flow, physician workload, and revenue distribution. The aim was to achieve actionable insights that would inform resource optimization strategies. Visualization software such as Tableau was used to create an interactive dashboard that provided a unified view of operational performance metrics.

Some of the key components of the Tableau dashboard included doctor usage in the form of average consultation time by type of doctor, which allowed us to view variations in workload and identify areas of over or underutilization. Another key visualization was the financial impact in terms of revenue from medication and consultations, which allowed us to ascertain which services generated the most revenue and where inefficiencies may be present. Further, the dashboard included a timeline view of consultant earnings by type of doctor on a yearly basis, giving insight into seasonal or temporal patterns of care provision and enabling more accurate forecasting and planning of resources.

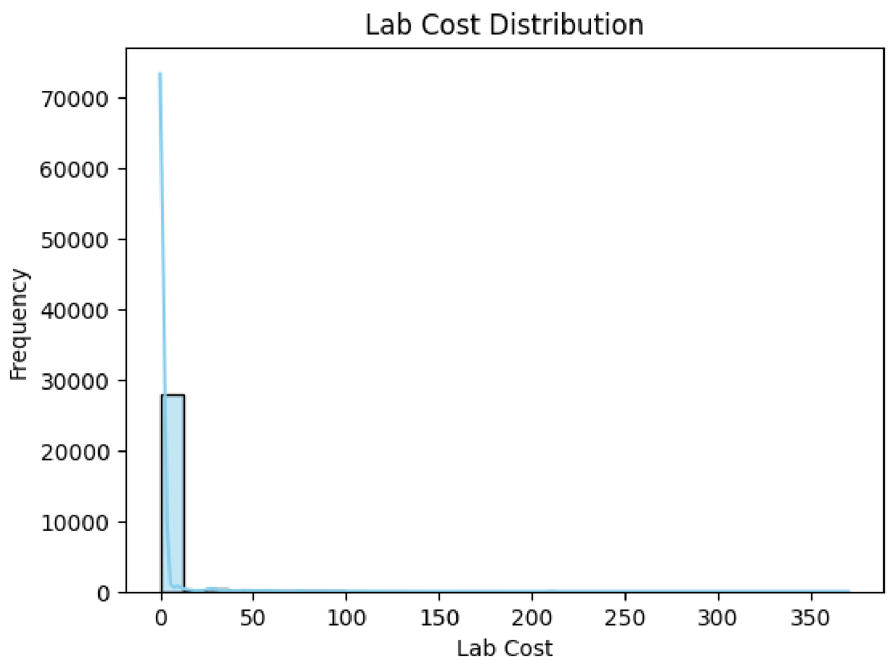


In parallel, Python-based EDA provided deeper statistical insights into the dataset.

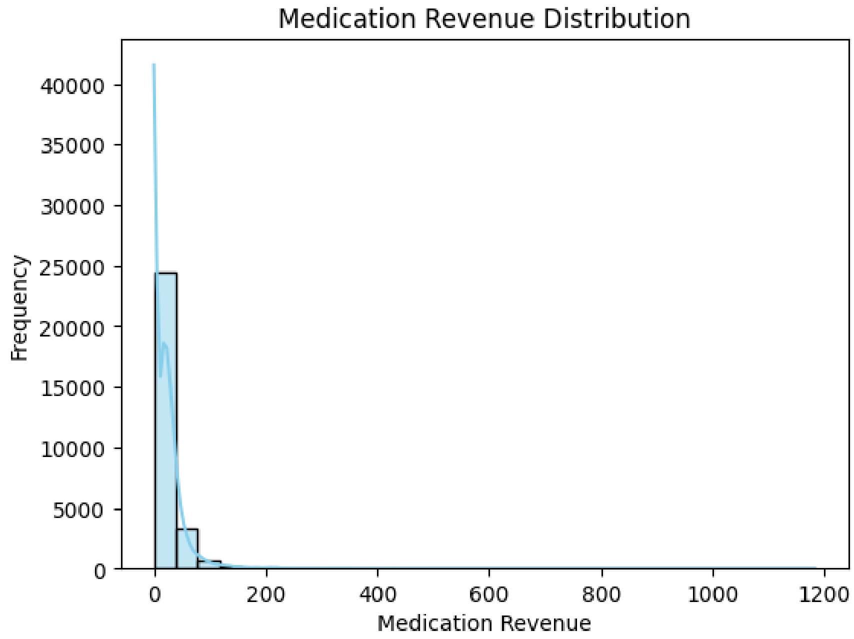
A correlation heatmap was created to evaluate the relationships between key variables such as lab cost, consultation revenue, and medication revenue. This analysis helped identify which services had the strongest financial impact and guided the prioritization of optimization efforts.



A lab cost distribution plot helped to determine the frequency of different lab cost ranges. Understanding this distribution was important to detect any skewed or outlier spending, potentially highlighting inefficiencies or overutilization of certain tests.



The medication revenue distribution plot provided insight into how frequently medications were prescribed at various revenue levels. This helped identify high-revenue medications and supported better inventory and cost control planning in the context of resource optimization.

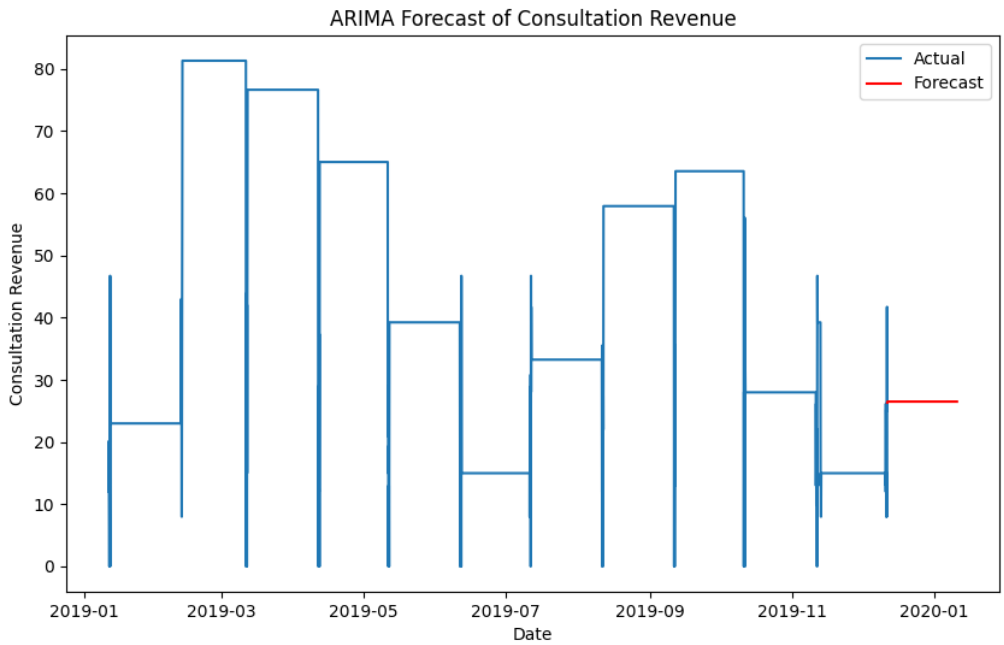


Time-Series Analysis for Demand Prediction

To predict high-demand periods, the data is combined into appropriate time periods (hourly, daily, or weekly). Seasonal patterns and trends are plotted in time-series plots.

Time series decomposition divides the data into trend, seasonality, and residuals. Forecasting models such as ARIMA is utilized in an attempt to predict future patient flow.

Our model was developed on Python based on the existing data to predict for a period of 30 days into the future.



Recommendations

Our system Smartstaff recommended the following measures:

1. **High-Demand Months: Increase staffing and consultation capacity during February, March, and April to address the higher projected consultation revenue.**
   These months experienced the highest revenue peaks indicating higher patient flow and necessitating more physician availability and operational cover to avert congestion and latency.
2. **Low-Demand Months: Scale down non-essential operations or shift resources in June, July, November and December.**
   These months have the lowest forecasted consultation revenue showing fewer patients attending. Resource allocation to minimize unnecessary operating expenses during these months lowers duplicative operating costs.
3. **Next 30-Day Forecast: Maintain baseline levels of staff and watch short-term trends closely.**
   The ARIMA model predicted a low-key recovery in consultation income in the 30 days following December. While not high, this suggests a gentle rising trend with no clear long-term indicators. More data will be needed before extended forecasting can be done with confidence.
4. **Implement preventive care or wellness programs in historically low-demand months to improve revenue stability.**
   According to the expected dip in June, July, and again in November, December, active involvement through wellness checkups or preventive screenings can help boost patient visits and revenue while utilizing underused capacity.
5. **Increase visibility and availability of low-utilization but high average consultation revenue physicians.**
   Year-to-date consultation revenue by doctor type revealed that some doctors earned high revenue per consultation but had relatively low utilization. Their financial performance and care delivery could be improved by promoting their expertise or adjusting their appointment slots.
6. **Prioritize stock monitoring and ordering for drugs with consistently higher revenue.**
   Plots of revenue distribution of drugs revealed a small cluster of drugs that generated a disproportionate share of total revenue. Keeping these well-supplied can prevent supply gaps and improve financial performance.

Conclusion

The aim of this project was to enhance operational efficiency in a healthcare setting by modeling patient flow, physician utilization, and financial trends to inform data-driven resource and staffing planning decisions. Through the use of an integrated solution that featured Tableau visualizations, Python-based statistical modeling, and Excel aggregation, it was possible to highlight important trends in consultation revenue, doctor workload, and service-level expenditure. The use of the ARIMA time-series model also facilitated short-term forecasting of consultation revenue, highlighting seasonal peaks and troughs across the calendar year.

The results of this analysis support the focused management of hospital resources. Recommendations such as increasing personnel in high-demand months, cutting back during low-demand months, and promoting preventive care initiatives during low-demand months are specific to optimize patient flow without compromising operational cost control. Even better, focus on underutilized but high-performing physicians and stockpiling top-revenue medications can have immediate payoffs for financial health and quality of patient care.

In summary, this project demonstrates the way data analytics can be applied to create beneficial strategies for boosting hospital efficiency. By aligning resources to demand patterns and reinforcing evidence-based planning, health institutions can significantly decrease service delivery and operating performance.

References

Akbasli, I. T., Birbilen, A. Z., & Teksam, O. (2025). Artificial intelligence-driven forecasting and shift optimization for pediatric emergency department crowding. JAMIA Open, 8(2). https://doi.org/10.1093/jamiaopen/ooae138

Borhade, R. R. (2024). AI-enhanced predictive analytics for proactive healthcare management: Leveraging machine learning to optimize patient care. Panamerican Mathematical Journal, 34(2), 123–138. https://doi.org/10.52783/pmj.v35.i1s.2096

Choudhury, A., & Urena, E. (2020). Forecasting hourly emergency department arrival using time series analysis. British Journal of Healthcare Management, 26(1), 34–43. https://doi.org/10.12968/bjhc.2019.0067

Erkamp, N. S., van Dalen, D. H., & de Vries, E. (2021). Predicting emergency department visits in a large teaching hospital. International Journal of Emergency Medicine, 14(1). https://doi.org/10.1186/s12245-021-00357-6

Ingole, P., Baviskar, P. R., Ghuge, M. S., Dakhare, B. S., Dongre, Y., & Dubey, P. (2025). Optimizing resource allocation in hospitals using predictive analytics and information systems. Journal of Information Systems Engineering & Management, 10(1s), 400–415. https://doi.org/10.52783/jisem.v10i1s.224

Kim, Y.-H., & Lee, J. (2024). Forecasting hospital room and ward occupancy using static and dynamic variables. JMIR Medical Informatics, 12, e53400. https://doi.org/10.2196/53400

Ko, I., & Chang, H. (2017). Interactive visualization of healthcare data using Tableau. Healthcare Informatics Research, 23(4), 349–354. https://doi.org/10.4258/hir.2017.23.4.349

Nwaimo, C. S., Adegbola, A. E., Adegbola, M. D., & Adeusi, K. B. (2024). Forecasting HR expenses: A review of predictive analytics in financial planning for HR. International Journal of Management & Entrepreneurship Research, 6(6), 1842–1853. https://doi.org/10.51594/ijmer.v6i6.1169

Pereira da Veiga, C., Pereira da Veiga, C. R., Girotto, F. M., Marconatto, D. A. B., & Su, Z. (2024). Implementation of the ARIMA model for prediction of economic variables: Evidence from the health sector in Brazil. Humanities and Social Sciences Communications, 11(1). https://doi.org/10.1057/s41599-024-03023-3

Porto, B. M., & Fogliatto, F. S. (2024). Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning. BMC Medical Informatics and Decision Making, 24, Article 377. https://doi.org/10.1186/s12911-024-02788-6

Rasouli, H. R., Esfahani, A. A., & Farajzadeh, M. A. (2019). Challenges, consequences, and lessons for way–outs to emergencies at hospitals: A systematic review study. BMC Emergency Medicine, 19(62). https://doi.org/10.1186/s12873-019-0275-9

Yinusa, A., & Faezipour, M. (2023). Optimizing healthcare delivery: A model for staffing, patient assignment, and resource allocation. Applied System Innovation, 6(5), 78. https://doi.org/10.3390/asi6050078

Yucesan, M., Gul, M., & Celik, E. (2018). A multi-method patient arrival forecasting outline for hospital emergency departments. International Journal of Healthcare Management, 13(sup1), 283–295. https://doi.org/10.1080/20479700.2018.1531608