A&E Inpatient Clustering Using K-Prototypes for Length of Stay Analysis: A Comparative Study Before and After COVID-19

Dino Huang

School of Computing and Communications, Lancaster University, Lancaster, United Kingdom, LA1 4YW

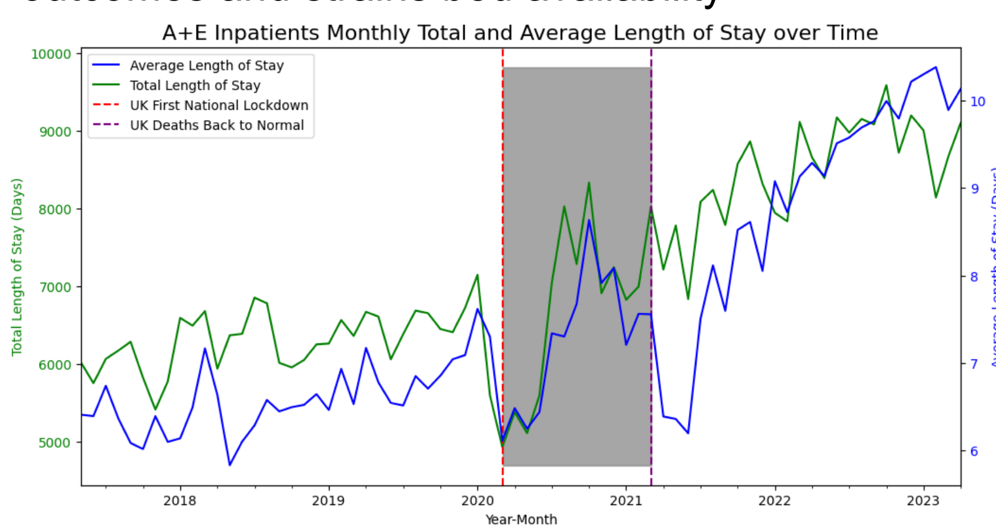
Introduction

Company Background

- Wrightington, Wigan and Leigh Teaching Hospitals NHS Foundation Trust (WWL): A part of the UK's NHS, WWL serves over 300,000 local residents in north-west England.
- Facilities: Includes Wigan Infirmary (General Medicine, A&E), Wrightington Hospital (Orthopaedic Surgery), Leigh Infirmary (Elderly Medicine), and more.
- Distinctive Status: As an NHS Foundation Trust, WWL enjoys more financial and operational autonomy than conventional NHS hospitals.

Research Background

- Demographic Shift: Patients admitted to WWL hospitals since 2016 are generally older and have longer stays.
- Length of Stay (LoS): Increased across all age groups, especially post-COVID-19.
- Impact: Extended LoS negatively affects patient outcomes and strains bed availability



Aim

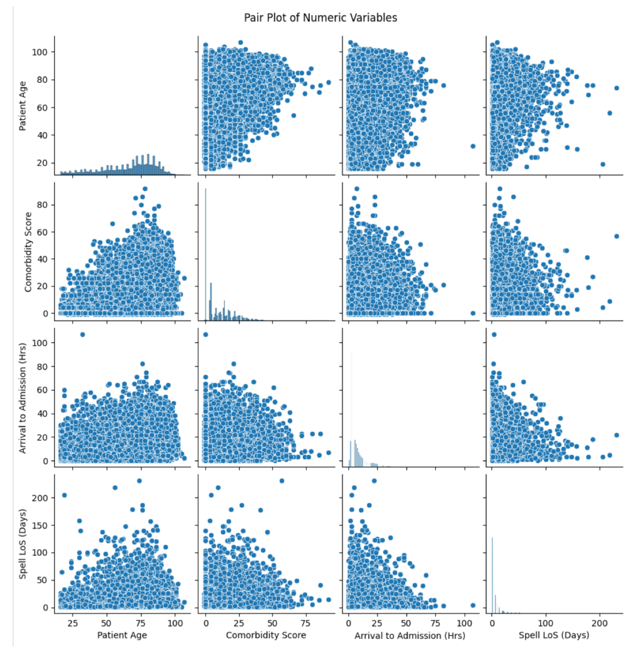
- Analyze Patient Trends: Study patient characteristics and demographics to identify changes in LoS over time.
- COVID-19 Impact: Assess how the pandemic specifically affected patient LoS.
- Advanced Clustering: Use multi-variable approaches to create nuanced patient profiles.
- Operational Optimization: Use insights to develop strategies for improving bed capacity and patient outcomes.

Through longitudinal studies, this research aims to offer valuable insights into patient LoS, ultimately informing better healthcare management strategies.

Method

Statistical Analysis

- ANOVA:
  Used to test the impact of categorical variables on Length of Stay (LoS). Significance level at 0.05.
- Correlation Analysis:
  Employed Pearson correlation to study linear relationships between continuous variables and LoS.



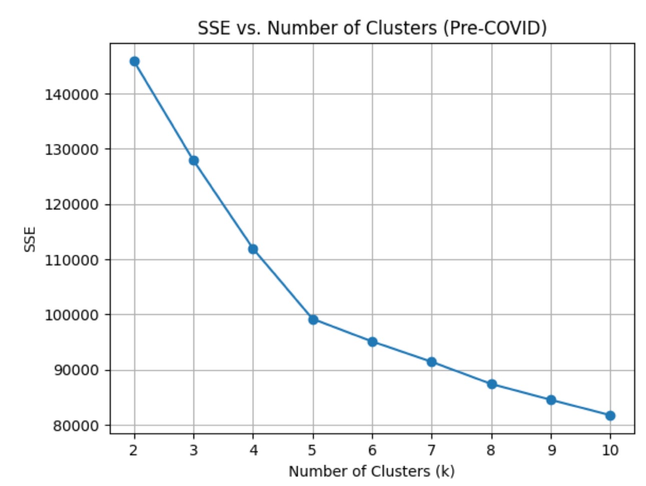
Pairplot of Continuous Variables for Correlation Analysis: This plot visualizes the pairwise relationships and distributions of key continuous variables.

Clustering Technology

- K-Prototypes Algorithm:
  Combines K-Means and K-Modes, optimized for both numeric and categorical attributes.

$$J = \sum_{i=1}^{n} \sum_{k=1}^{K} w_{ik} \left(\alpha \sum_{j=1}^{p} (x_{ij} - z_{kj})^2 + (1-\alpha)\delta\left(g_{ij}, h_{kj}\right) \right)$$

- Pre- and Post-Pandemic Data Clustering:
  Initial baseline model created using pre-pandemic data. Comparison made with post-pandemic patient data.
- Evaluation Metric: SSE:
  Used to assess cluster quality. Optimum number of clusters determined through the elbow method.



Elbow Plot Demonstrating Optimal Number of Clusters: The Sum of Squared Errors (SSE) reaches a noticeable elbow point at 5 clusters, indicating this as the optimal number for well-defined patient clustering.

Results

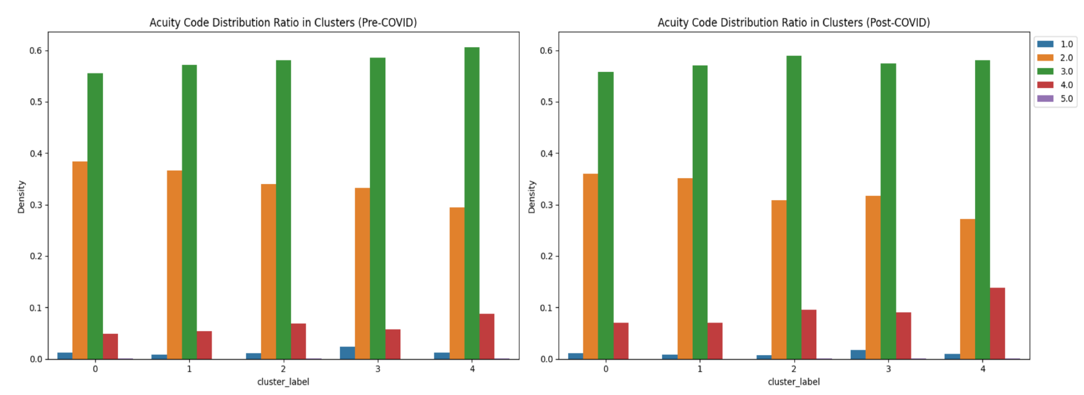
Cluster Characteristics

Table of Cluster Centroids Identifying Five Distinct Patient Subgroups. Each cluster is characterized by key clinical metrics such as patient age, comorbidity score, time from arrival to admission, length of stay (LoS), and referral source. Cluster labels provide an intuitive categorization: 'Elderly', 'Extended A&E stays', 'Normal Majority', 'High LoS', and 'Young Adults.'

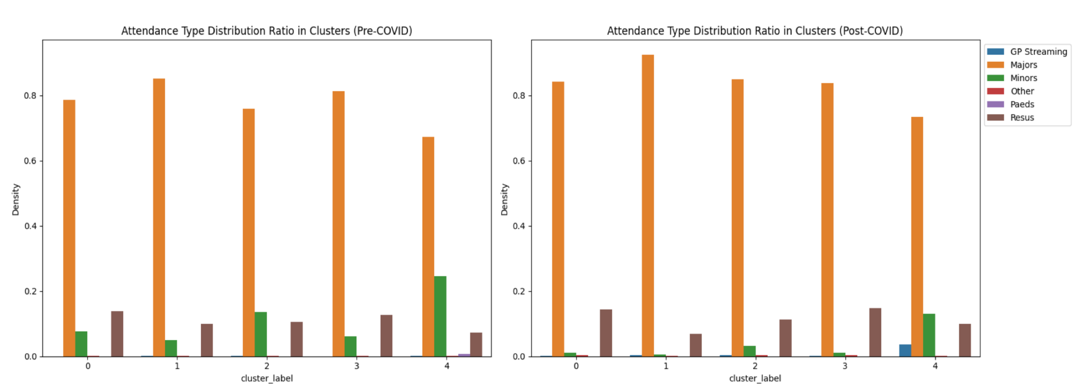
Table 1: Cluster Centroids

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Patient Age	77.24	67.17	74.58	74.11	38.21
Comorbidity Score	23.86	7.50	5.04	15.45	1.38
Arrival To Admission (hrs)	4.79	12.33	3.89	5.65	4.16
LoS (days)	7.01	5.37	4.62	34.19	3.25
Referral Source	Direct	Direct	Direct	Direct	Self referral
Label	Elderly	Extended A&E stays	Normal Majority	High LoS	Young Adults

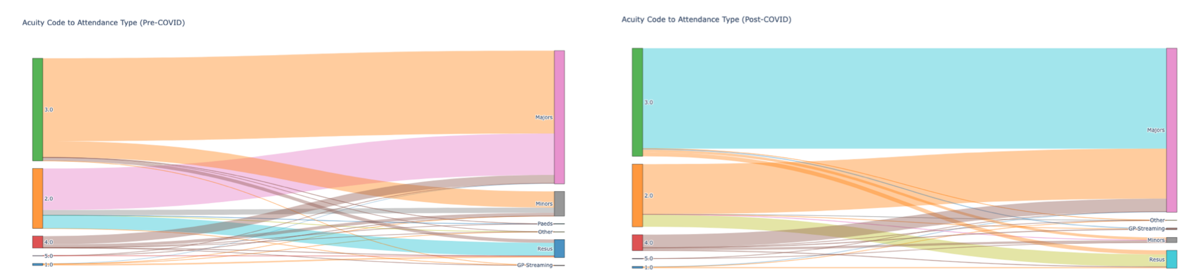
Key Observations



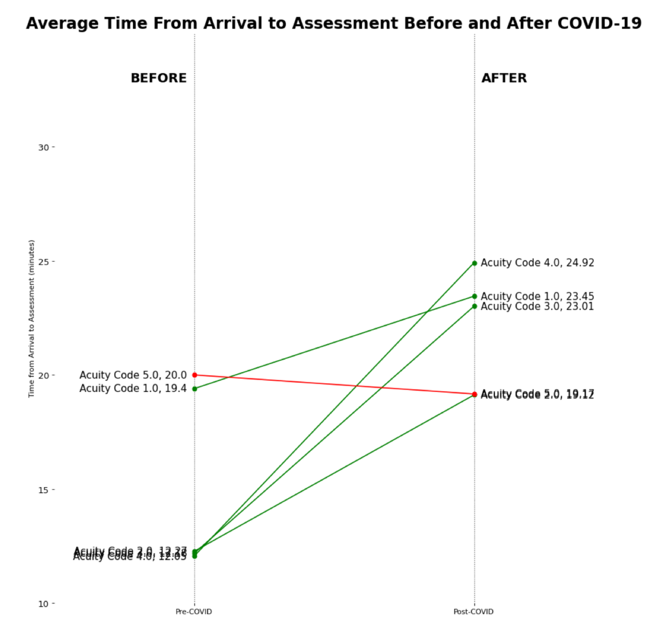
Stable Acuity Code Distribution: Intriguingly, the distribution of acuity codes remained stable before and after the onset of the pandemic. This suggests that the nature and severity of cases presenting to the ED have not significantly changed as a direct result of the pandemic.



Shift in Majors and Minors: Despite the stable distribution of acuity codes before and after the onset of the COVID-19 pandemic, we observed a noticeable shift in the distribution of patients between the 'Majors' and 'Minors' areas. The 'Majors' area typically receives more medical resources and handles more complex cases, while the 'Minors' area has fewer resources and deals with less complicated conditions.



These two Sankey plots illustrate how patients were allocated to different areas after being assigned an acuity code, both before and after the onset of the COVID-19 pandemic. It can be observed that after the pandemic, a higher proportion of patients with an acuity code 4 (Standard level emergency care) were also allocated to the 'Majors' area.



Increased Wait Time for Assessment: Another significant change was observed in the wait time for initial clinical assessments, which has elongated after the pandemic's onset. This may indicate increased strain on resources or changes in patient triage processes.

Conclusion

After COVID-19's onset, there was a noticeable extension in both LoS and initial evaluation waiting times at WWL's A&E. The patient mix shifted towards older individuals presenting with greater comorbidity, notably including more musculoskeletal issues among those with prolonged stays. Importantly, a higher number of patients were directed to the 'Majors' area, which typically necessitates more resources, even though patient severity didn't significantly change. Urgent recommendations involve forming a Rapid Assessment Team, revisiting acuity code assignments, and investing in specialized care for musculoskeletal disorders.

Acknowledgements

Special thanks to WWL healthcare staff for their exceptional service. Appreciation to T. Ingram and B. Wood for data science support, and to Dr. N. Rennie for academic guidance. Gratitude to friends and family for unwavering support.