Duy Nguyen
dcnguyen060899@gmail.com

UC Berkeley ML and AI
Professional Certificate

Healthcare Management Data Analysis: Executive Report

UC Berkeley ML/AI, Capstone Executive Report, 2024

Outline

Executive Summary

In response to challenges exposed by the COVID-19 pandemic, our project focused on optimizing hospital resource management by predicting patient length of stay. Using advanced data analysis, machine learning, and deep learning techniques, we've identified key factors affecting stay duration and developed models to enhance patient care and resource allocation.

Project Goals

- **Enhance Patient Care:** Understand social, financial, and demographic factors impacting recovery times.
- **Optimize Hospital Resources:** Identify operational bottlenecks and resource allocation inefficiencies.
- **Develop Targeted Interventions:** Create data-driven strategies to address factors contributing to prolonged stays.

Patient Demographics and Hospital Characteristics

- Middle-aged patients (31-60 years) form the largest group requiring hospital care.
- Very young (0-10) and very old (91-100) patients are least common.
- Significant variations in patient volumes across hospital types and regions:
  - Hospital codes 8 and 28 handle notably higher volumes.
  - Type 'e' hospitals are most prevalent, followed by type 'b'.
  - Region X has the highest number of cases, followed by Y and Z.

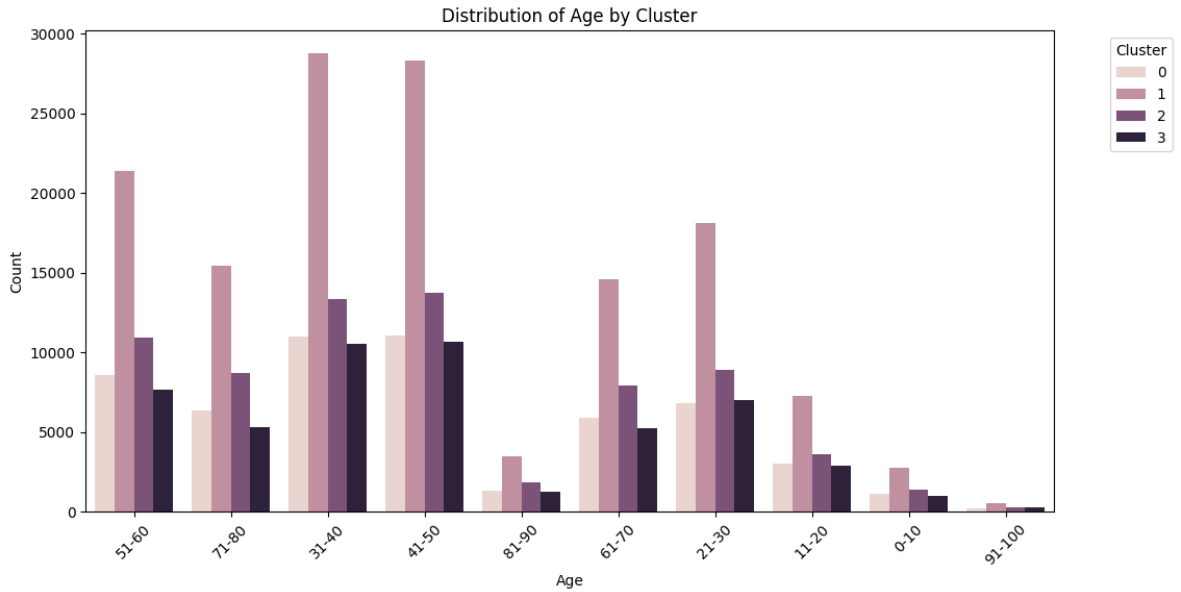

Figure: Age Distribution by Cluster

Length of Stay Patterns

- Multiple peaks in stay duration, with highest frequencies around 10, 20, and 30 days.
- Factors associated with longer stays:
  - Trauma admissions
  - Extreme severity of illness
  - Departments: Surgery, Radiotherapy
  - Ward types: T and U

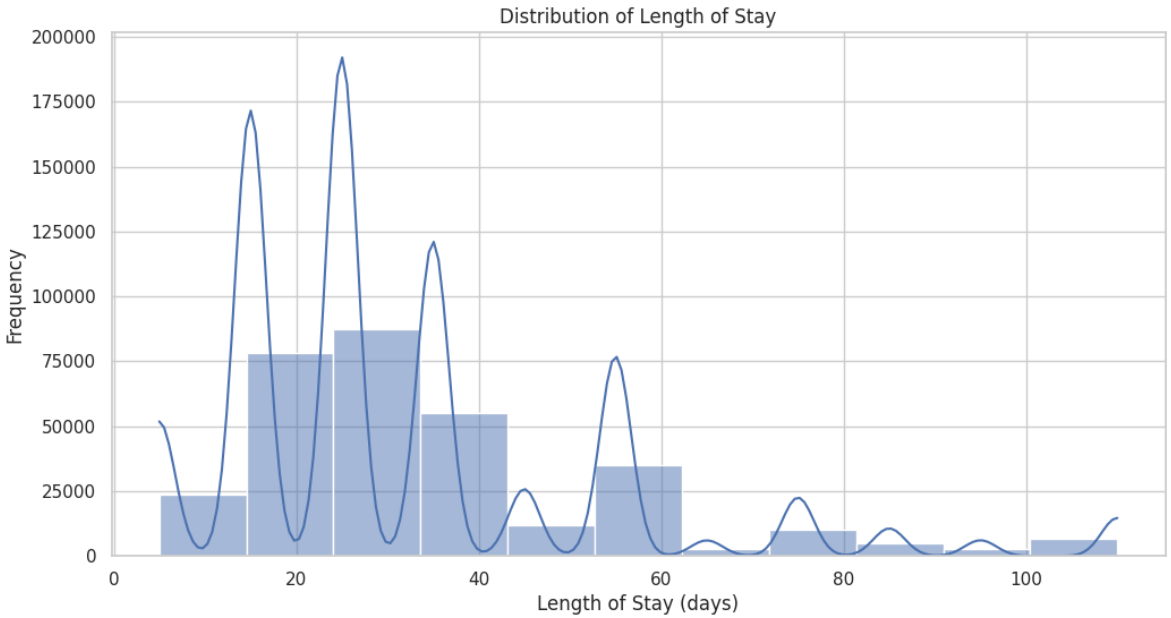

Figure: Length of Stay (LOS) Distribution

Readmission Insights

- Higher readmission rates observed in:
  - Trauma cases
  - Moderate to extreme illness severity
  - Gynecology department
  - Hospital type 'a'
- Patients with more than 10 readmissions tend to have significantly longer average stays.
- Age groups 31-40 and 41-50 have the highest number of readmissions.

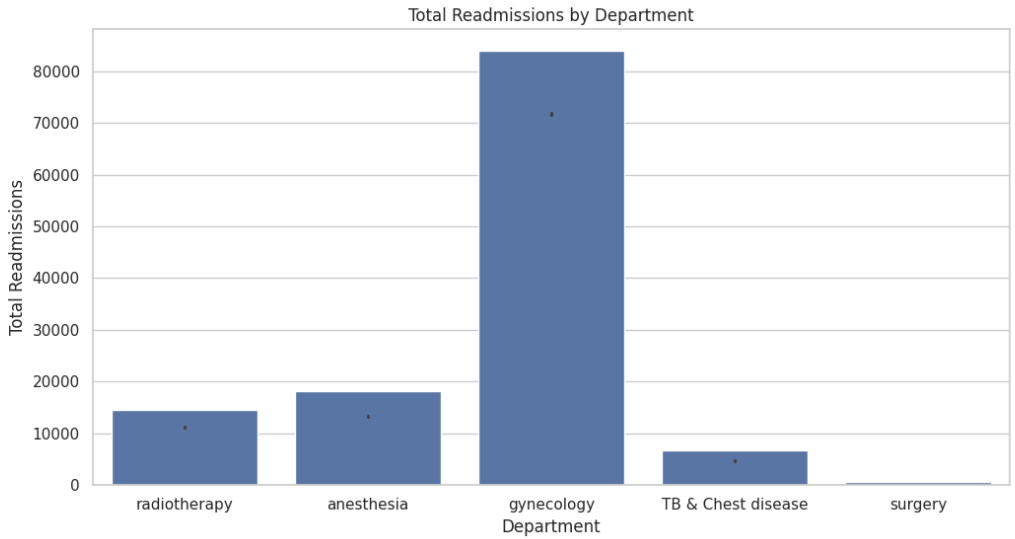

Figure: Total Readmissions by Department

Facility Quality Analysis

- Cluster 0 (Balanced): Efficient resource utilization with moderate bed grades, admission deposits, and extra rooms.
- Cluster 1 (High-End): Premium hospitals with highest admission deposits and extra rooms, most visitors.
- Cluster 2 (Budget): Possible lower-income areas with lowest bed grades, deposits, and extra rooms.
- Cluster 3 (Mixed): High-quality with moderate capacity, highest bed grades, and second-highest deposits.

Facility Quality Analysis

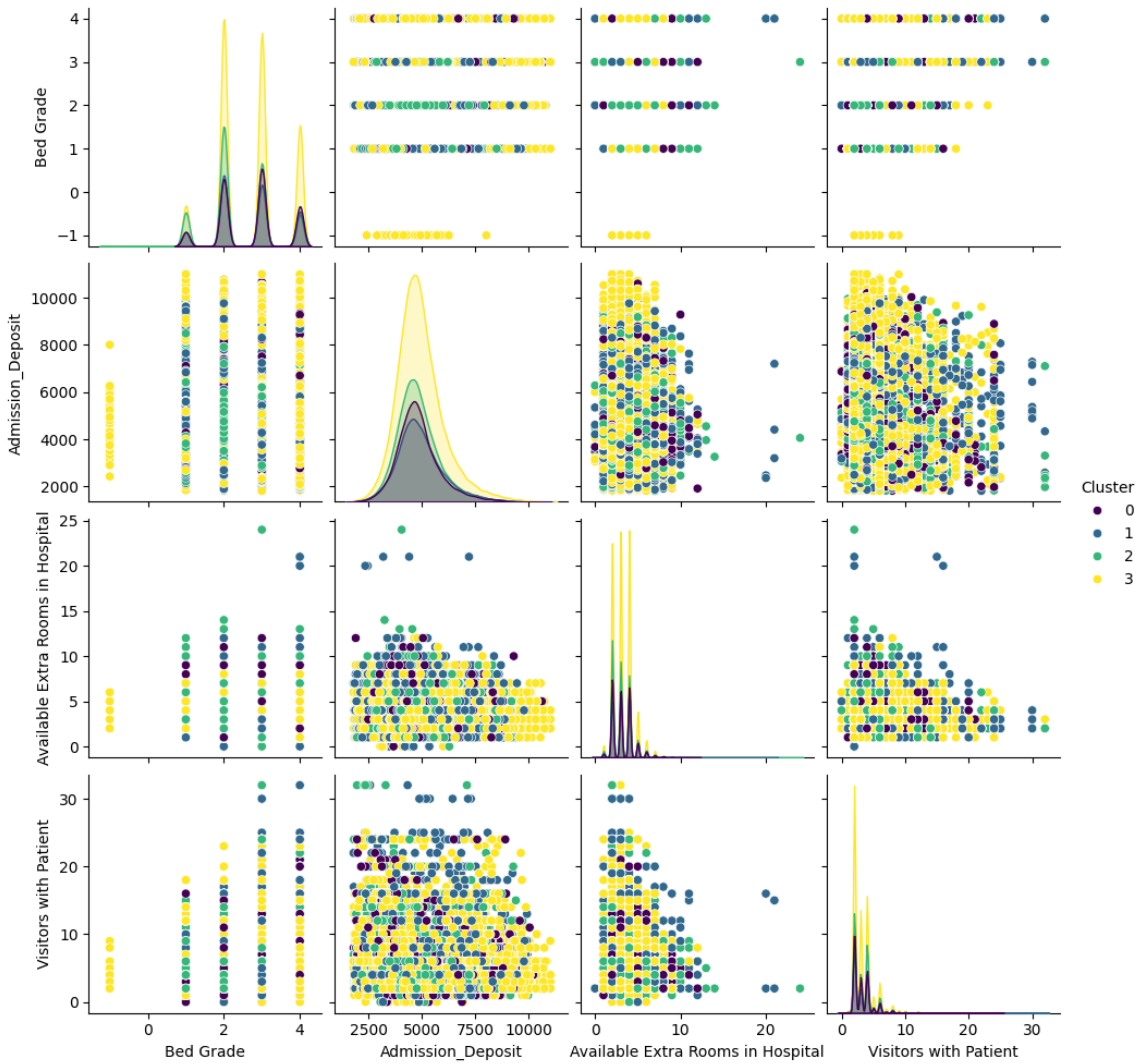

Figure: Numerical Cluster Pair Plot

Key Predictive Factors

- Number of visitors with the patient
- Admission deposit amount
- Available extra rooms in the hospital
- Specific ward types (especially Q, P, S)
- Type of admission (Emergency, Trauma)
- Severity of illness
- Specific hospital and city codes

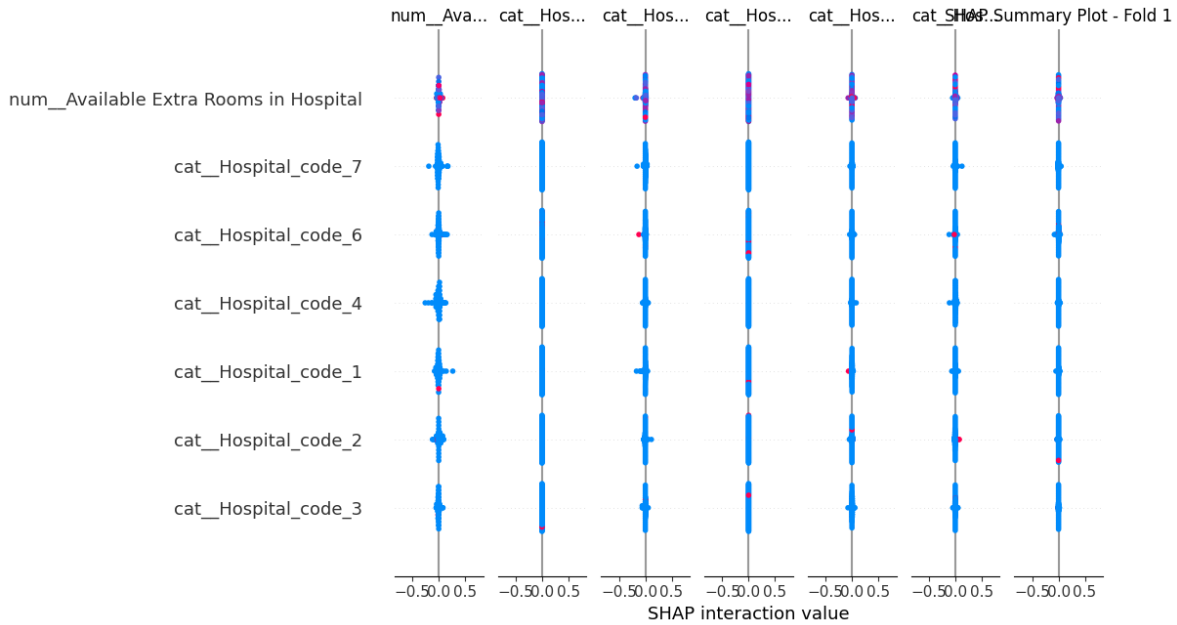

Figure: Neural Network Feature Importances (Fold 1)

Model Performance

Model	Train Accuracy	Test Accuracy
Neural Network	65.24%	80.42%
CatBoost	46.23%	42.84%
XGBoost	45.80%	42.41%
Random Forest	49.68%	42.19%
Gradient Boosting	41.93%	41.62%
Logistic Regression	39.92%	40.10%
Dummies Classifier	27.43%	27.64%

Table: Model Performance Summary

Model Performance

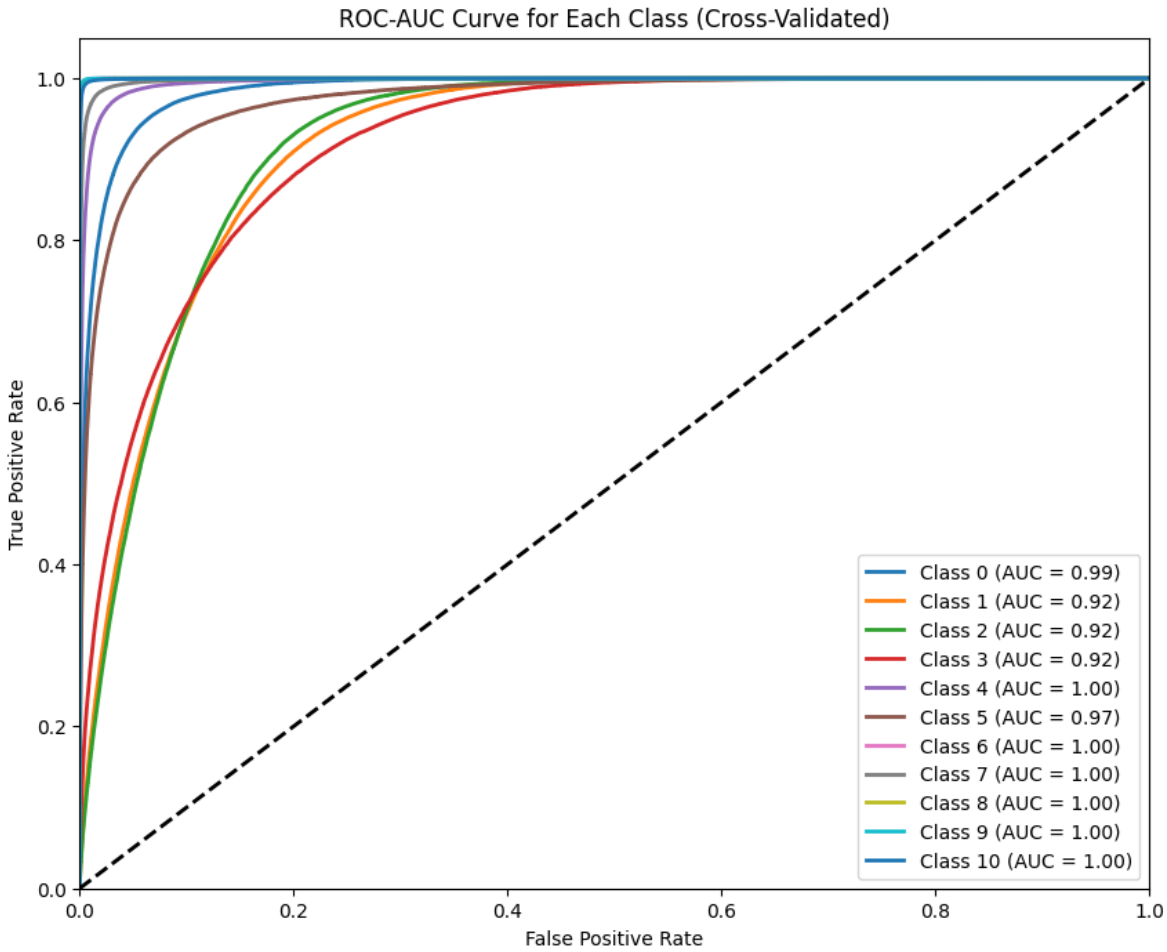

Figure: ROC-AUC Neural Network

Business Impact

Model	FP Cost	FN Cost	Total Cost	Savings
Baseline	$4,689,900	$36,698,500	$41,388,400	-
Deep Learning	$1,000,000	$10,000,000	$11,000,000	$30,388,400
CatBoost	$1,500,000	$17,500,000	$19,000,000	$22,388,400
XGBoost	$1,300,000	$18,500,000	$19,800,000	$21,588,400

Table: Business Cost Analysis for Models

Business Recommendations

- **Resource Allocation:** Optimize based on ward types and illness severity.
- **Visitor Management:** Develop structured visitor programs balancing patient support with operational efficiency.
- **Financial Planning:** Adjust admission deposit strategies to manage patient turnover more effectively.
- **Tailored Care Plans:** Personalize care based on admission type and illness severity.
- **Facility Improvements:** Invest in upgrading bed grades and ensuring adequate extra rooms.
- **Hospital-Specific Strategies:** Implement best practices from high-performing hospitals.
- **Enhanced Follow-Up Care:** Implement targeted post-discharge support for high-risk patients.
- **Department Focus:** Prioritize resources in gynecology and anesthesia.
- **Age-Specific Care:** Develop specialized programs for working-age adults.
- **Continuous Model Improvement:** Regularly retrain and fine-tune models with new data.

Technical Recommendations

- **Enhanced Traditional Modeling:** Implement SMOTE and apply Stratified K-Fold cross-validation.
- **Ensemble Method Exploration:** Develop a voting classifier combining the best-performing traditional models.
- **Comparative Analysis:** Conduct a thorough comparison of all models and assess trade-offs between model complexity, interpretability, and performance.