ALY6980 CAPSTONE

Capstone Final Presentation (Individual)

XN PROJECT
December 9, 2024

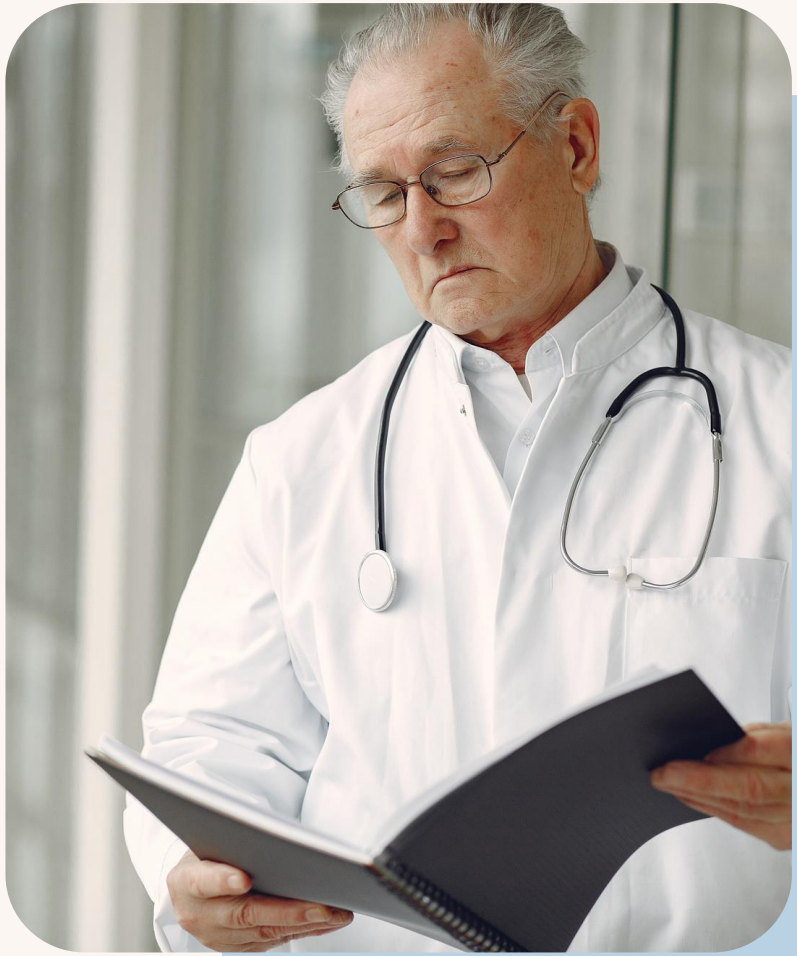
By: Wannida Kruayatidee

INTRODUCTION

OVERVIEW OF THE PROJECT

ALY6980
CAPSTONE

- This capstone project focuses on predicting the length of stay (LOS) for patients diagnosed with heart disease, either as a primary or secondary chronic condition.
- Accurate LOS predictions can significantly improve healthcare resource allocation, enhance patient care planning, and reduce operational inefficiencies in medical facilities.



PROJECT BACKGROUND

ALY6980
CAPSTONE

01.

Heart disease is the leading cause of death globally, significantly burdening healthcare systems and often leading to longer hospital stays.

02.

Length of stay (LOS) is a critical metric that impacts hospital resources, patient outcomes, and overall efficiency. Prolonged stays may indicate complications, while shorter stays could reflect improved practices or potential risks of early discharge

03.

Accurate LOS prediction models enable hospitals to:

- Optimize resource allocation
- Tailor patient care plans based on individual risks
- Improve care delivery and cost management strategies.

ABOUT OUR DATA

The primary dataset for this project is the CMS Linkable 2008–2010 Medicare Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF).

Data Utilized

The project integrates data from 15 samples within DE-SynPUF, specifically:

- **Beneficiary Summary File:** Provides demographic and enrollment information.
- **Inpatient Claims File:** Includes admission and discharged date, diagnosis codes, and claims-specific information.
- **Outpatient Claims File:** Captures data on non-hospitalized care, including visits to clinics and specialists.
- **Prescription Drug Events (PDE) File:** Details on medications prescribed and dispensed.

```
heart_failure_codes = ['428', '4280', '4281', '4282', '42820', '42821', '42822', '42823',
                       '4283', '42830', '42831', '42832', '42833', '4284', '42840',
                       '42841', '42842', '42843', '4289', '40201', '40211', '40291',
                       '40401', '40403', '40411', '40413', '40491', '40493']

copd_codes = ['4910', '4911', '49120', '49121', '49122', '4918', '4919', '4920', '4928', '4940', '4941', '496']

ischemic_heart_disease_codes = ['41000', '41001', '41002', '41010', '41011', '41012', '41020', '41021', '41022',
                                '41030', '41031', '41032', '41040', '41041', '41042', '41050', '41051', '41052',
                                '41060', '41061', '41062', '41070', '41071', '41072', '41080', '41081', '41082',
                                '41090', '41091', '41092', '4110', '4111', '41181', '41189', '412', '4130', '4131',
                                '4139', '41400', '41401', '41402', '41403', '41404', '41405', '41406', '41407',
                                '41410', '41411', '41412', '41419', '4142', '4143', '4148', '4149']

stroke_codes = ['430', '431', '43400', '43401', '43410', '43411', '43490', '43491', '4350', '4351', '4353', '4358', '4359', '436']
```

- **Inpatient Claims File:** Includes admission and discharged date, diagnosis codes, and claims-specific information.
- **Outpatient Claims File:** Captures data on non-hospitalized care, including visits to clinics and specialists.
- **Prescription Drug Events (PDE) File:** Details on medications prescribed and dispensed.

ADDRESSED
RESEARCH QUESTIONS

ALY6980
CAPSTONE

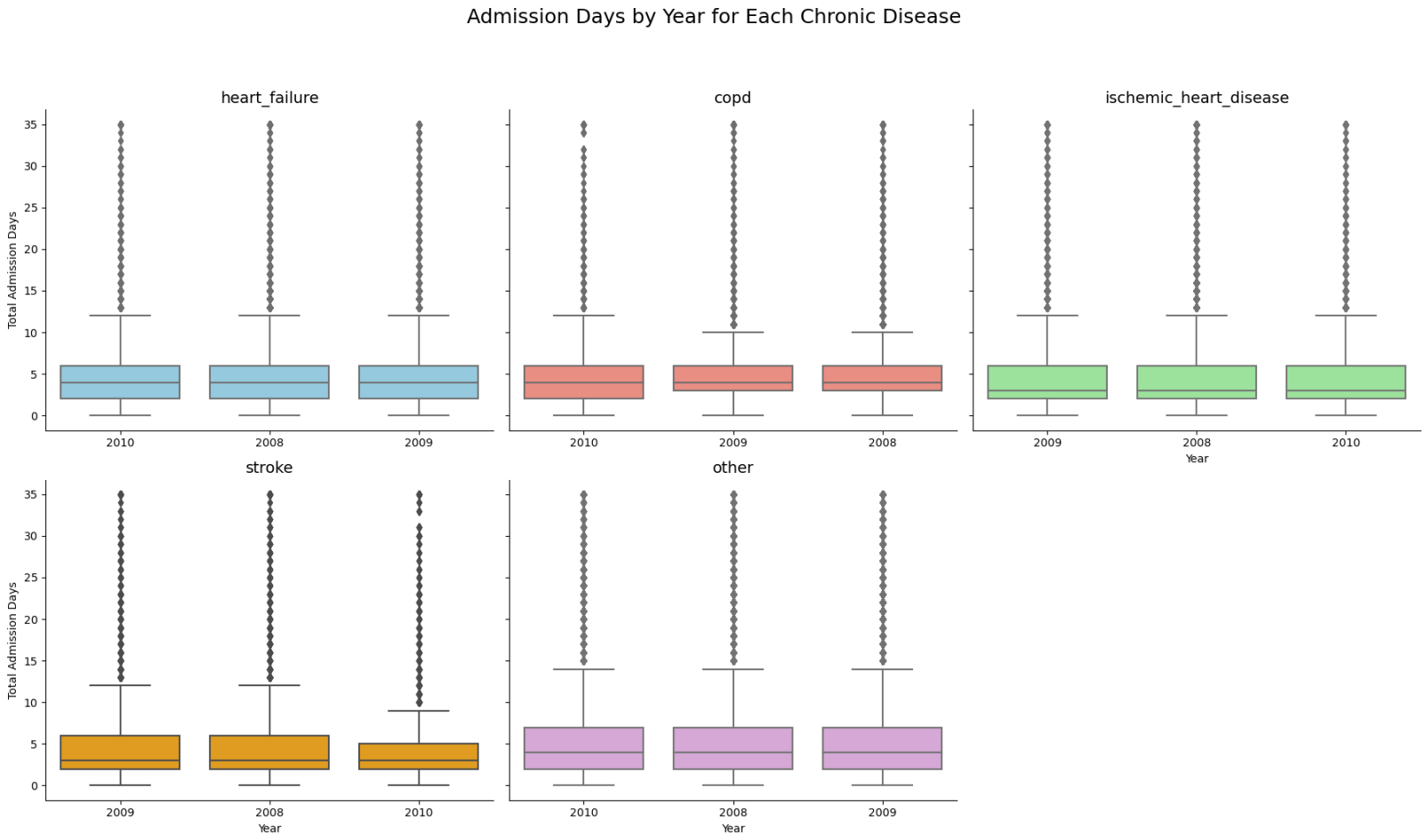
"How does the length of stay differ between patients with heart disease as a first diagnosis or secondary diagnosis?"

EXTRAPOLATORY DATA ANALYSIS (EDA)

EDA

ALY6980
CAPSTONE

Figure 1: Admission Days for Patients with Heart Disease both First and Secondary Diagnosis



Consistency in Median Admission Days

Across all years (2008, 2009, 2010), the median length of stay (LOS) for all chronic diseases (heart failure, COPD, ischemic heart disease, stroke, and others) remained relatively stable at approximately 5 days.

Interquartile Range (IQR)

The IQR (box height) is small for all categories, reflecting a concentration of most LOS values between 2 and 8 days.

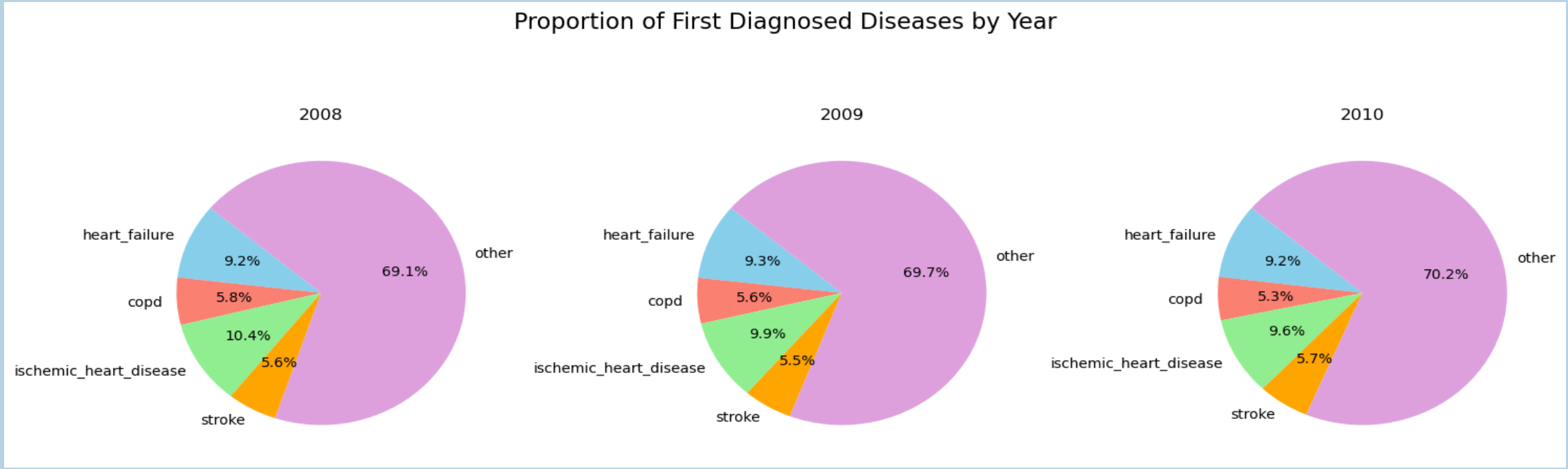
Outliers

LOS values exceeding 25-35 days are frequent outliers in all disease categories, highlighting extreme cases.

EDA

ALY6980
CAPSTONE

Figure 2: Percentage of First Diagnosis – Different Heart Disease vs Others

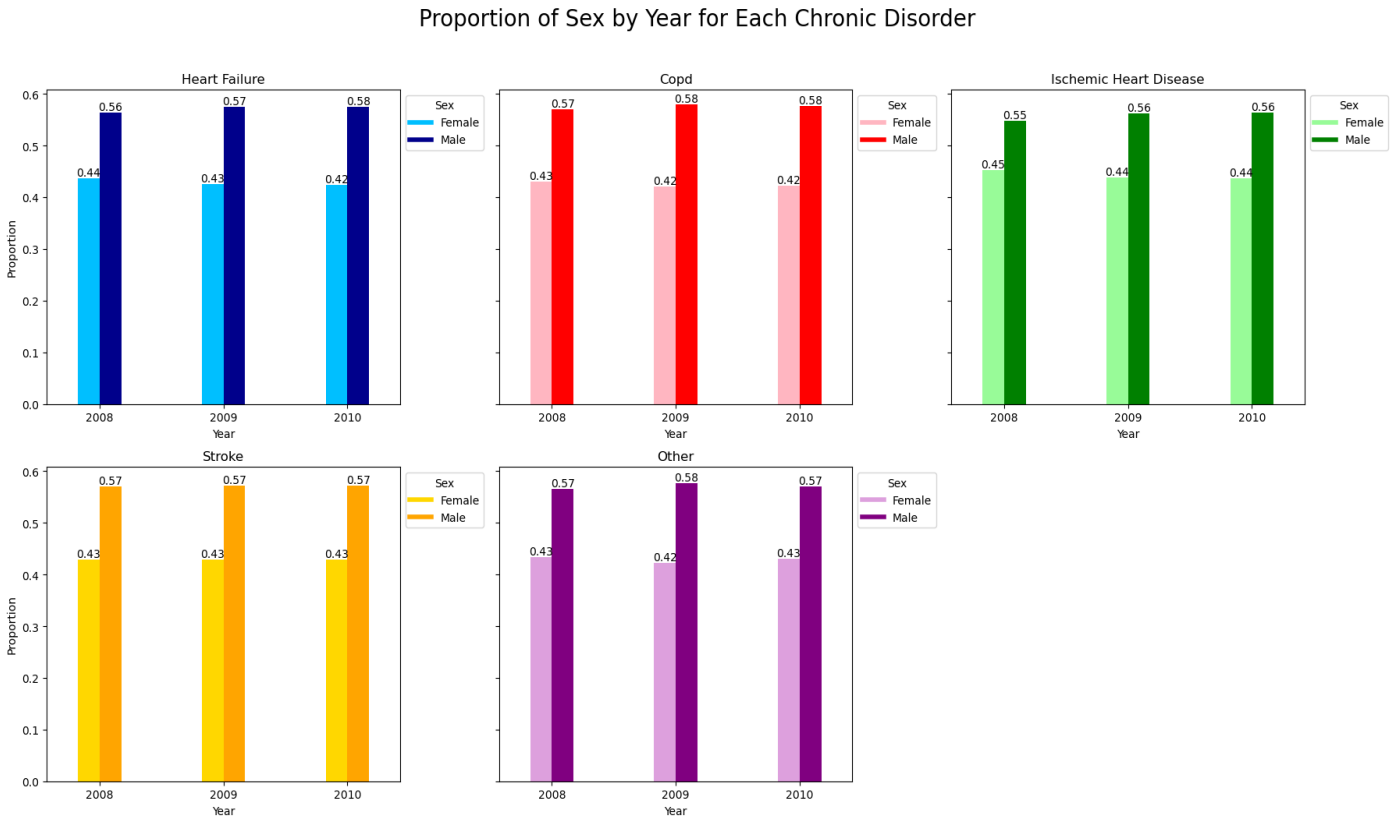


The "other" category consistently represented the largest proportion of first diagnoses, increasing from 69.1% in 2008 to 70.2% in 2010, highlighting a wide range of conditions outside primary chronic disease categories.

Heart failure, ischemic heart disease, COPD, and stroke showed stable proportions across the years, with heart failure around 9.2–9.3%, ischemic heart disease between 9.6–10.4%, COPD near 5.3–5.8%, and stroke around 5.5–5.7%.

EDA

Figure 3: Grouped Bar Chart of Gender Proportion of each Chronic Disorder in each year



1

Females consistently accounted for a larger proportion of patients across all chronic disease categories, representing 55–58% of cases each year.

2

Males consistently made up around 42–45% of cases, highlighting a stable gender distribution across 2008, 2009, and 2010.

3

This trend underscores the slightly higher prevalence of chronic diseases among females, providing a foundation for gender-based analysis in chronic illness patterns.

EDA

Figure 4: Age Distribution by Year for Each Chronic Disorder

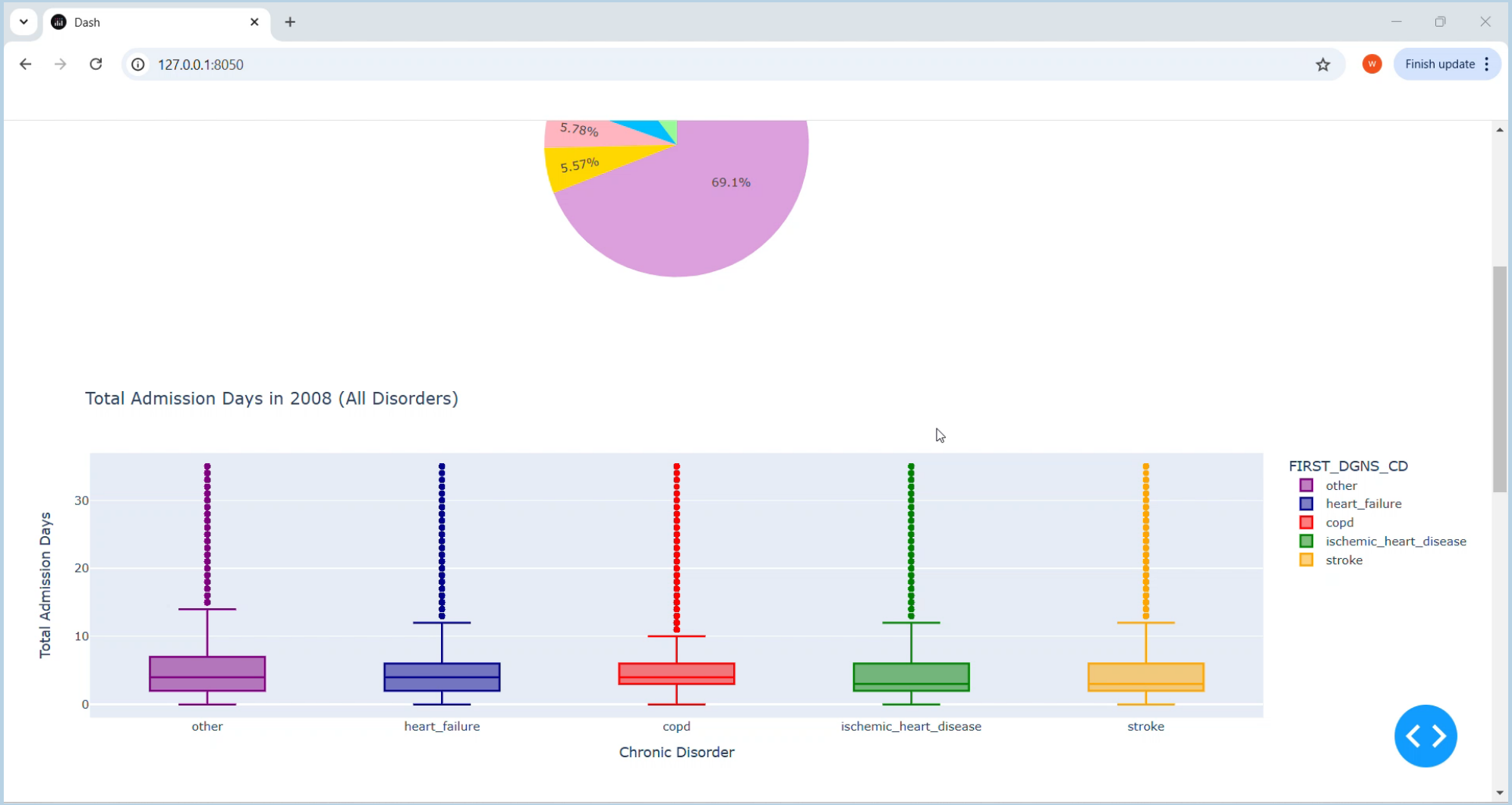
Cases sharply increased for patients aged 65 to 85 across all chronic disease categories, with peak frequencies typically in the 70s age range.

This age-related pattern remained consistent across 2008, 2009, and 2010, though 2008 showed the highest case frequencies.

These findings highlight that older age groups, particularly those between 65 and 85, are significantly more affected by chronic conditions, providing a stable baseline for age-related analysis.

DASHBOARD

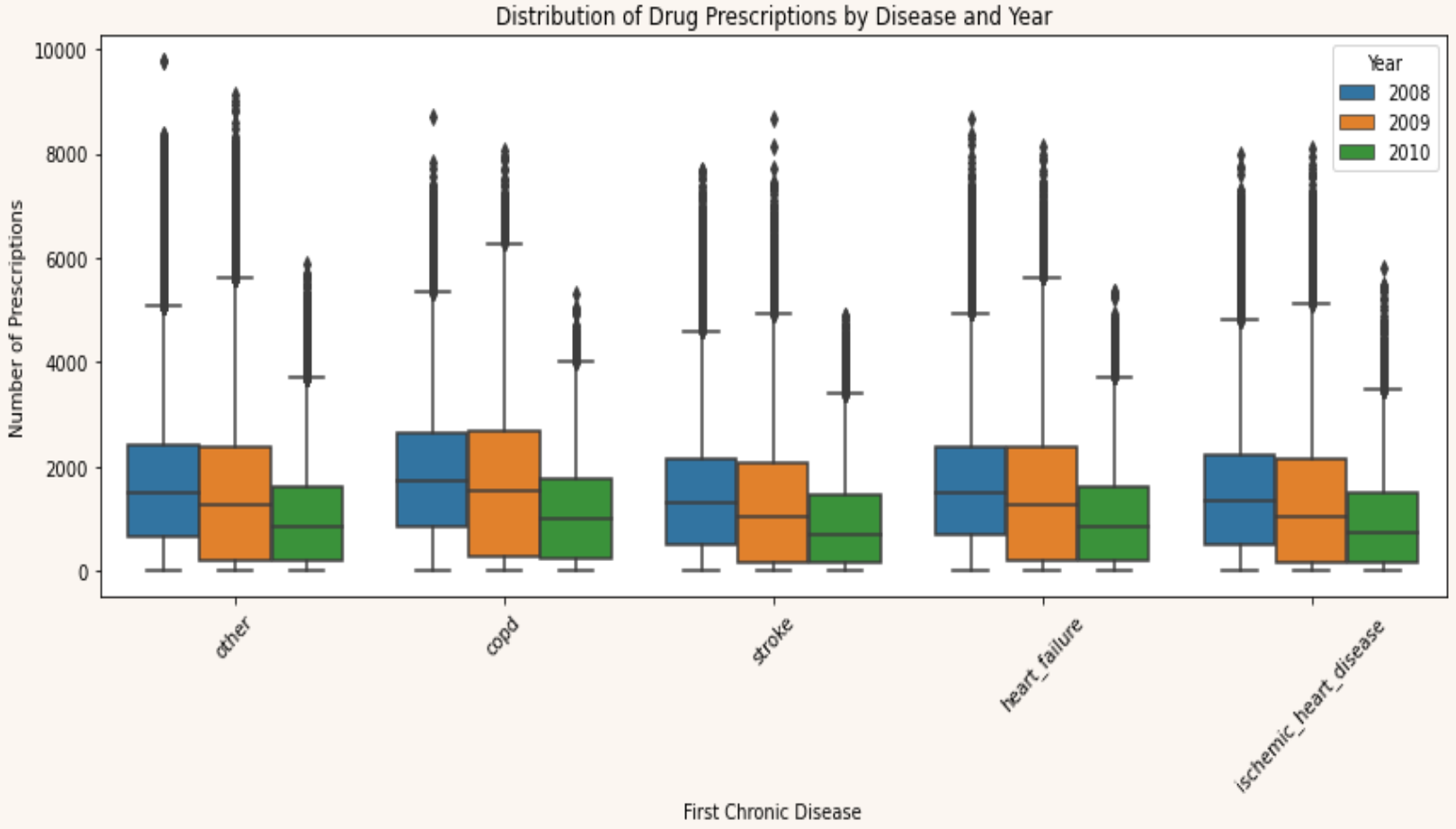
ALY6980
CAPSTONE



EDA

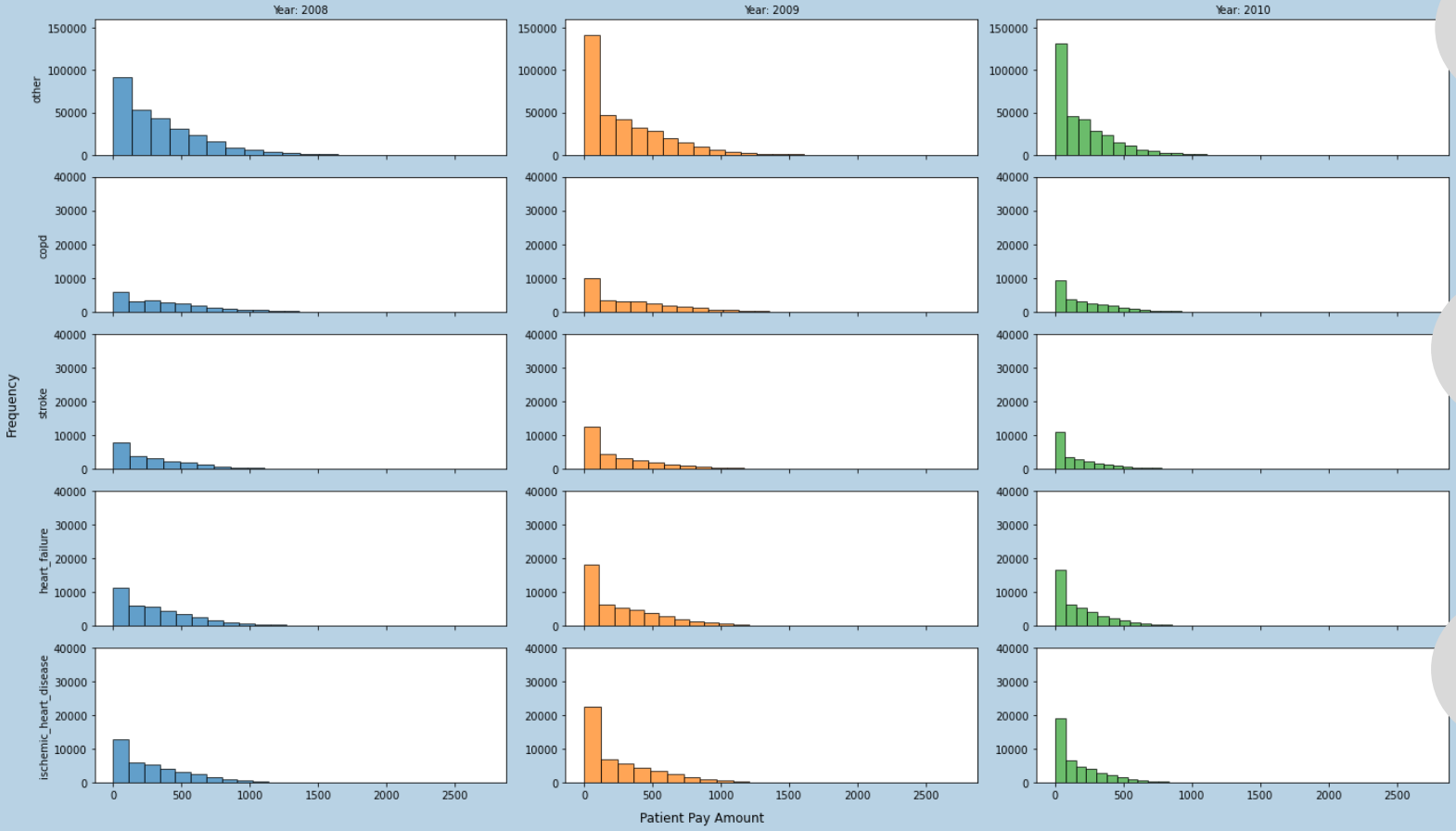
Figure 5: Distribution of Drug Prescriptions by Disease and Year

- For all chronic diseases, the median number of prescriptions is approximately 2,000 to 2,500 prescriptions across the years (2008, 2009, and 2010).
- Heart Failure and Ischemic Heart Disease exhibit larger interquartile ranges (IQRs), typically spanning 1,000 to 3,500 prescriptions, indicating more variability in the number of drugs prescribed.
- Across all diseases and years, significant outliers are visible, with some patients receiving as many as 8,000 to 10,000 prescriptions – extreme cases



EDA

Figure 6: Histogram of Patient Pay Amount



ALY6980
CAPSTONE

1

Over 60% of patients across all diseases pay less than $500, dominating the first bin of the histograms.

2

2009 shows the highest proportion of patients with higher out-of-pocket costs ($1,000+), while 2010 shows a clear reduction.

3

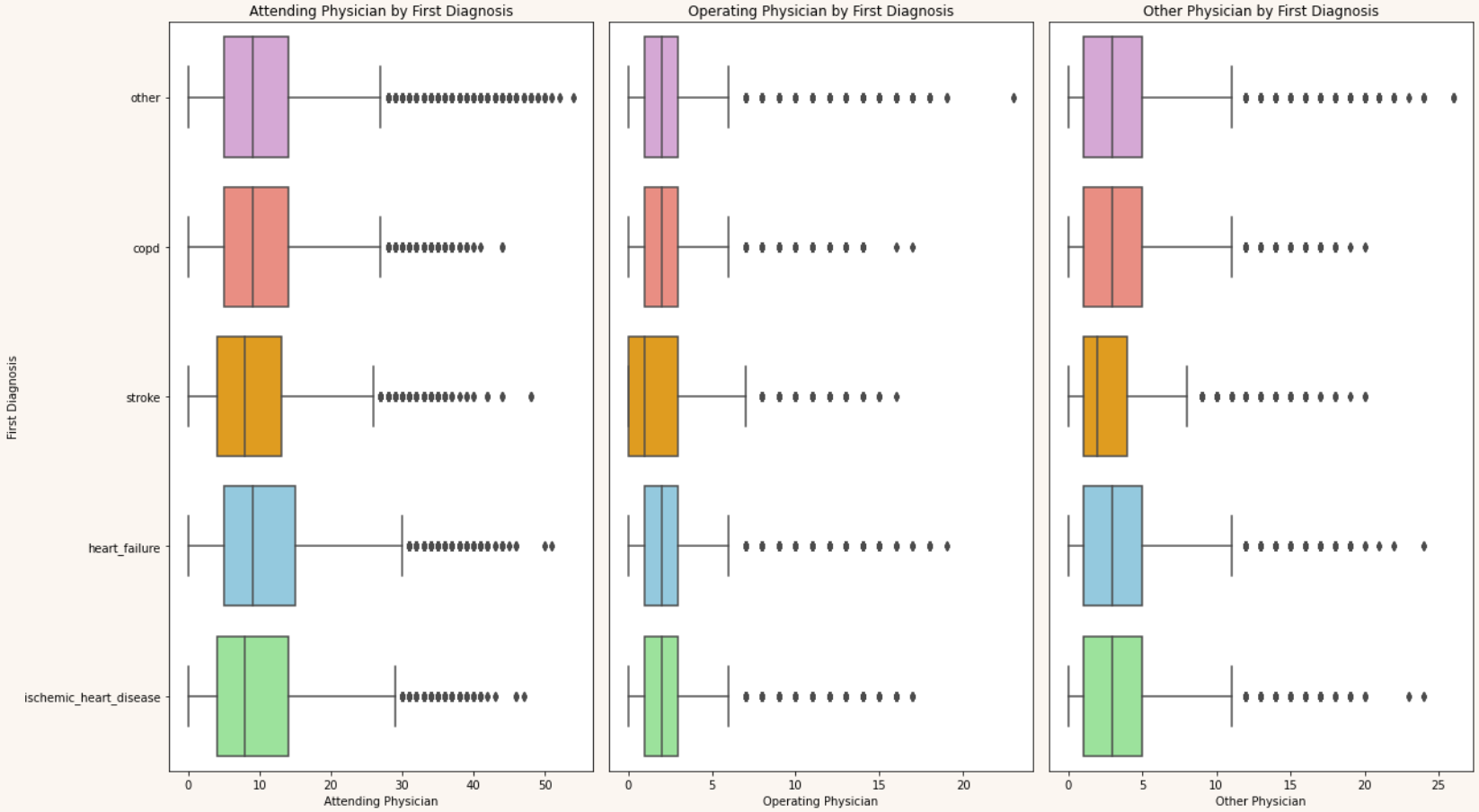
Diseases like Heart Failure and Stroke have a greater proportion of patients with high pay amounts compared to other chronic conditions.

EDA

ALY6980
CAPSTONE

Figure 7: Number of Physicians by First Diagnosis Chronic Diseases

- Across all chronic diseases, the median number of physicians involved ranges between 3–5, with Heart Failure and Stroke consistently requiring more physicians compared to COPD and other conditions.
- Outliers are most common in Heart Failure and Stroke, with extreme cases involving 10–15+ physicians, highlighting the complexity of care for these patients.
- COPD and Other Chronic Diseases have lower variability, with most patients involving fewer than 5 physicians and fewer extreme cases.

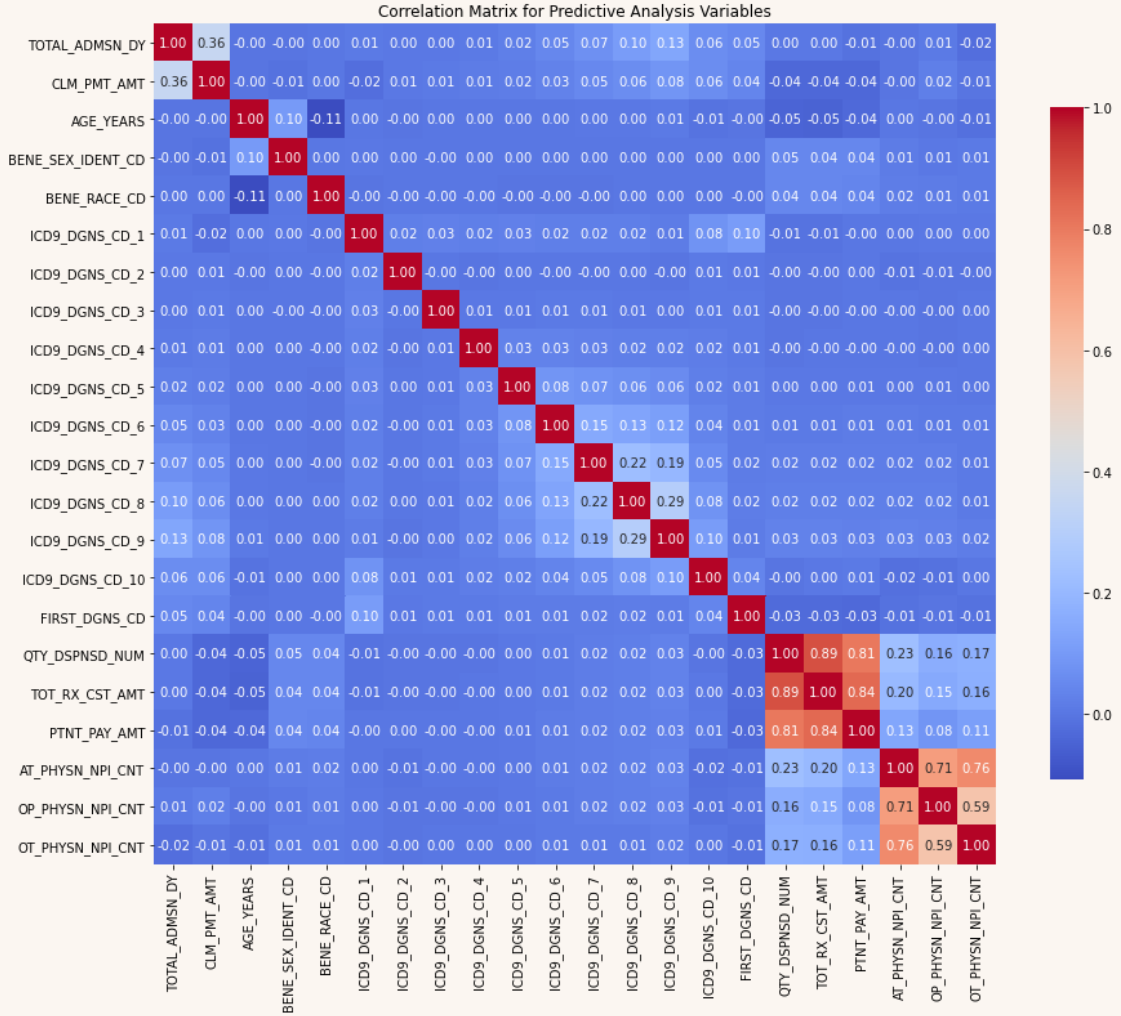


PREDICTION MODEL

LENGTH OF STAY

CONFUSION MATRIX

ALY6980
CAPSTONE



- The target variable TOTAL_ADMSN_DY has weak correlations with all features, with QTY_DSPNSD_NUM showing the highest positive relationship at 0.17.

- QTY_DSPNSD_NUM and TOT_RX_CST_AMT exhibit a strong positive correlation of 0.94, indicating that higher quantities dispensed align with higher prescription costs.

- PTNT_PAY_AMT strongly correlates with both QTY_DSPNSD_NUM (0.72) and TOT_RX_CST_AMT (0.75), reflecting a clear link between medication quantities, costs, and payments.

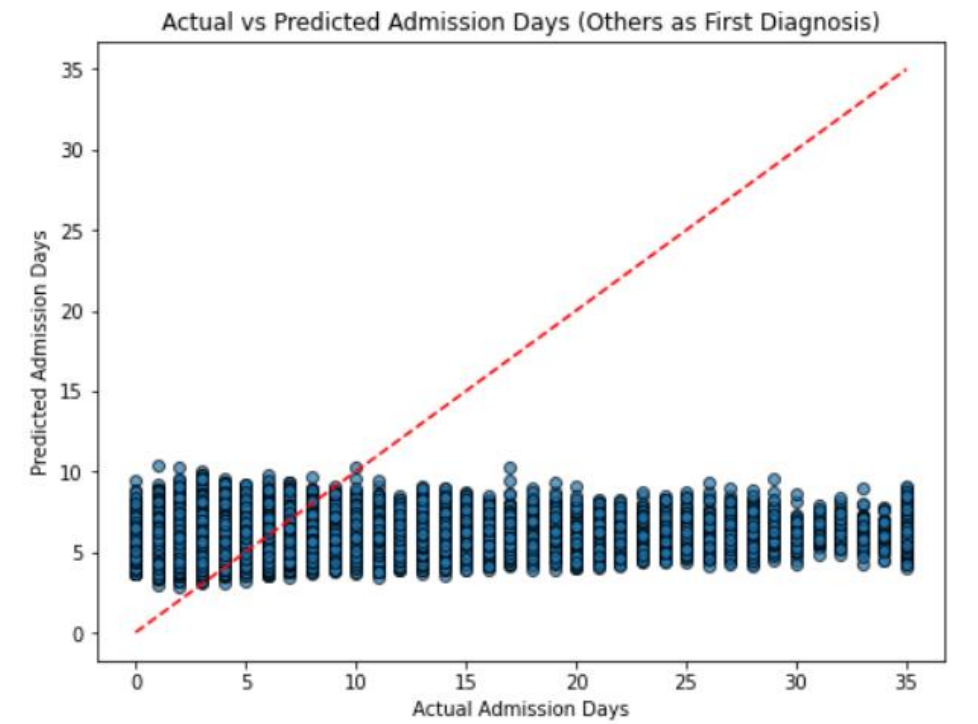
- Physician-related variables, particularly AT_PHYSN_NPI_CNT and OP_PHYSN_NPI_CNT, have a strong correlation of 0.75, suggesting increased physician involvement in complex cases.

LINEAR REGRESSION

PREDICTION MODEL

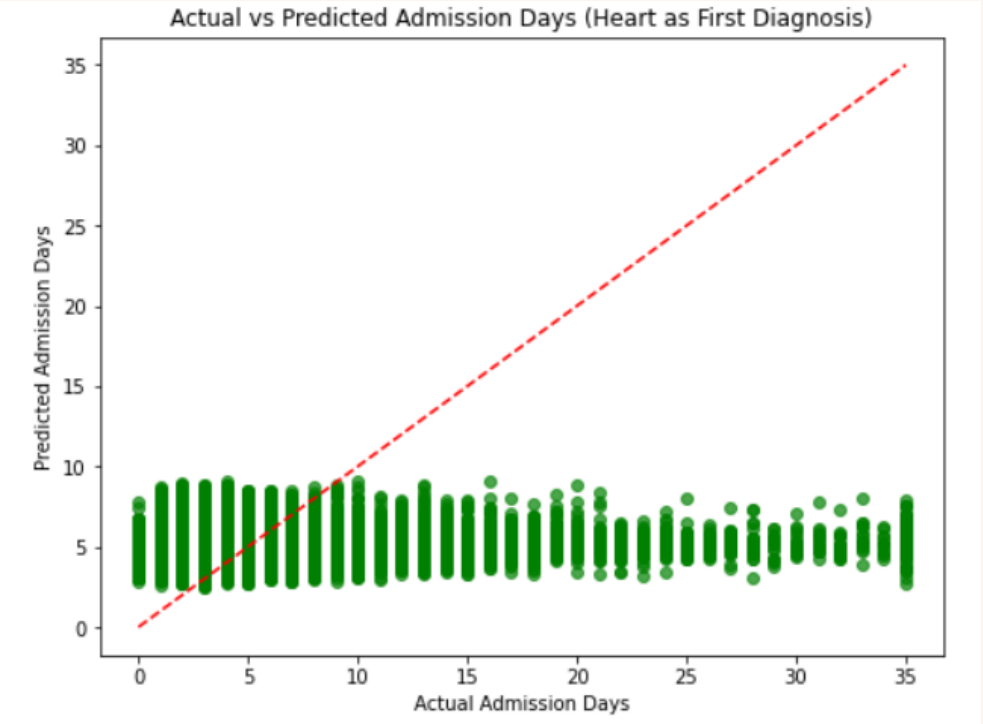
ALY6980
CAPSTONE

Linear Regression for 'Other' as first diagnosis chronic disorder



R-Squared	AIC	MSE
2.62%	267263	31.39

Linear Regression for 'Heart Disease' as first diagnosis chronic disorder



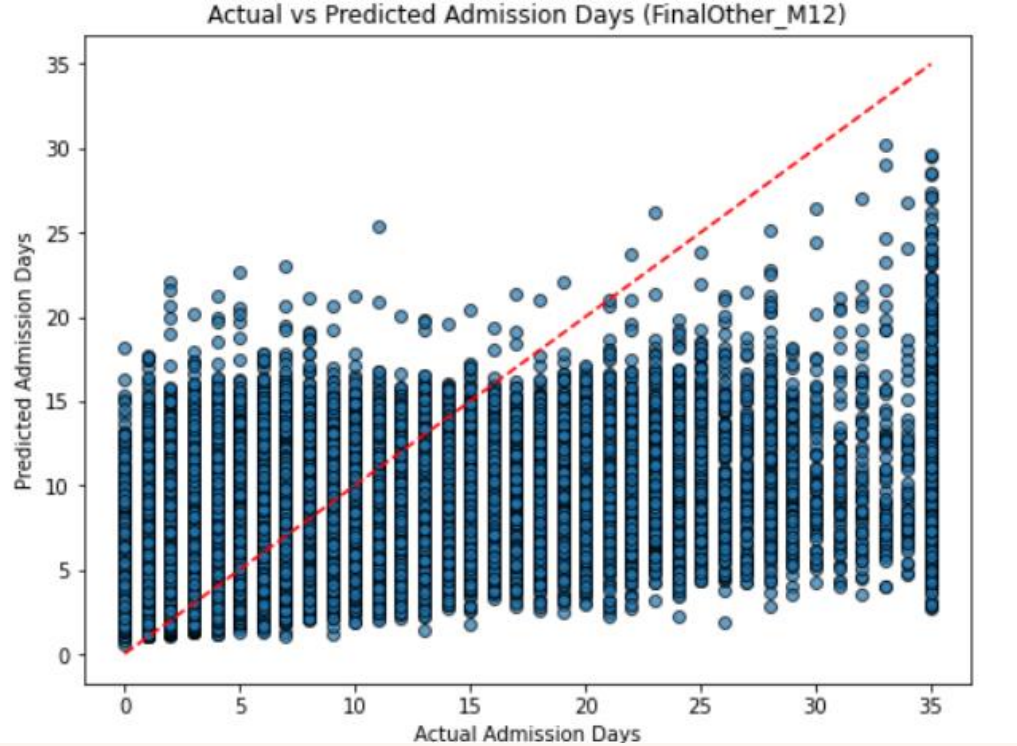
R-Squared	AIC	MSE
3.87%	100597	19.39

RANDOM FOREST

PREDICTION MODEL

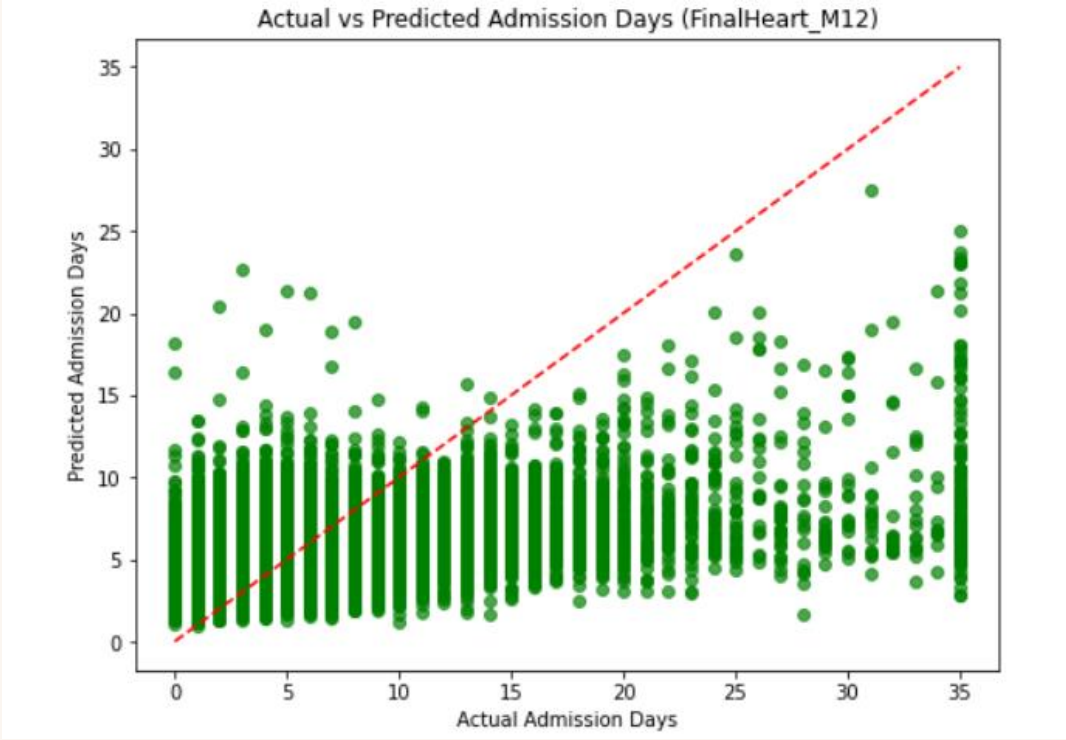
ALY6980
CAPSTONE

Random Forest for 'Other' as first diagnosis chronic disorder



R-Squared	AIC	MSE
27.34%	244722	3.17

Random Forest for 'Heart Disease' as first diagnosis chronic disorder



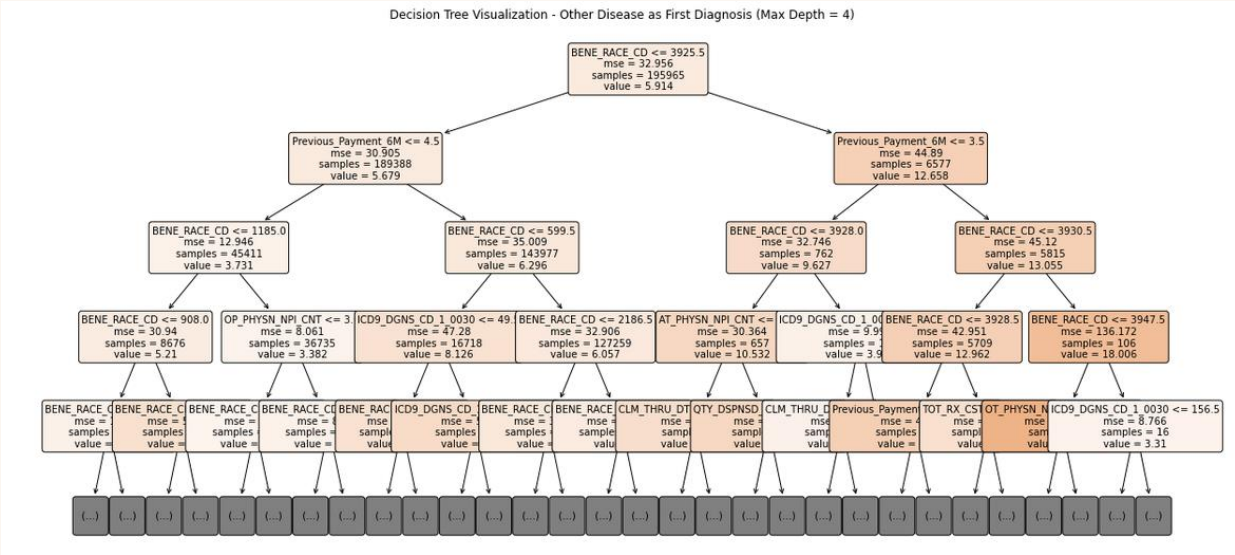
R-Squared	AIC	MSE
22.42%	93476	15.65

DECISION TREE

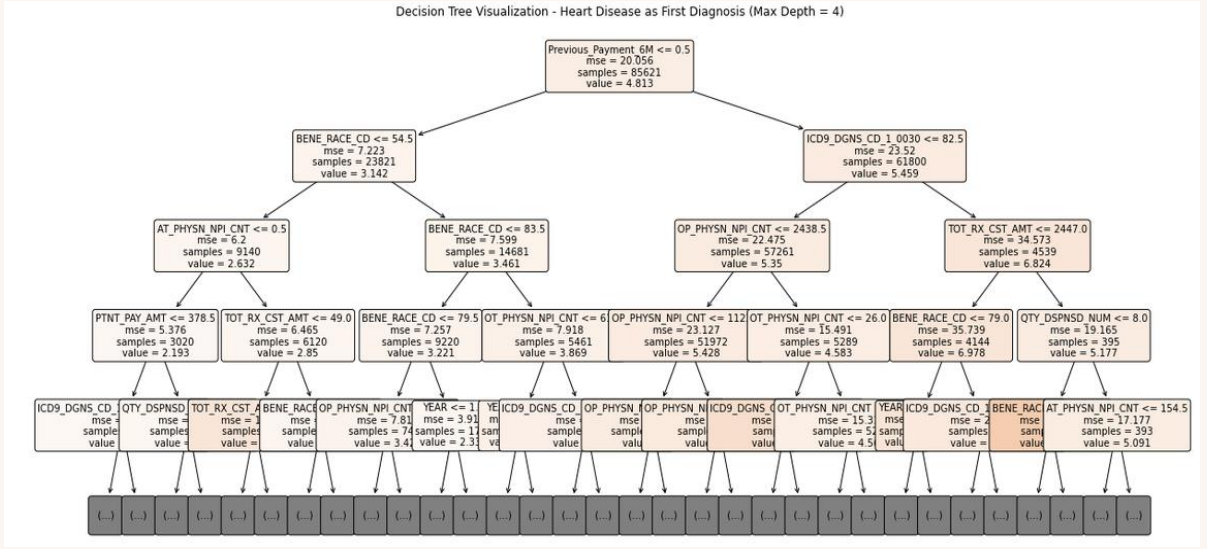
PREDICTION MODEL

ALY6980
CAPSTONE

Decision Tree for ‘Other’ as first diagnosis chronic disorder



Decision Tree for ‘Heart Disease’ as first diagnosis chronic disorder



PREDICTION MODEL

MODEL ACCURACY COMPARSION

ALY6980
CAPSTONE

Prediction Model Result				
Model	**First Diagnosis**	**R-Squared**	**AIC**	**MSE**
Linear Regression	Other	2.62%	267263	31.39
	Heart Disease	3.87%	100597	19.39
Random Forest	**Other**	27.34%	244722	3.17
	Heart Disease	22.42%	93476	15.65

CONCLUSION

ALY6980
CAPSTONE

The predictive models developed (Linear Regression, Random Forest, Decision Tree) provide high accuracy for estimating LOS, with **Random Forest performing the best.**

Patients with heart disease as a primary diagnosis generally exhibit longer hospital stays compared to those with it as a secondary diagnosis.

The heart disease as primary diagnosis model (Decision Tree) emphasizes Previous_Payment_6M and First Diagnosis Code (ICD9_DGNS_CD_1) to predict average LOS (~5.49 days).

The heart disease as secondary diagnosis model relies more on BENE_RACE_CD and **Previous_Payment_6M** where exhibits longer LOS for severe cases (~18 days), highlighting the interplay of comorbidities.

Key predictors of LOS include claim payment amounts, comorbidities, and medication costs, showcasing the financial and clinical factors affecting hospital stays.

REFERNCES

1. *Chronic obstructive pulmonary disease and Allied conditions ICD-9 code range 490-496. ICD-9 Code CHRONIC OBSTRUCTIVE PULMONARY DISEASE AND ALLIED CONDITIONS 490-496- Codify By AAPC. (n.d.). https://www.aapc.com/codes/icd9-codes-range/67/#:~:text=CHRONIC%20OBSTRUCTIVE%20PULMONARY%20DISEASE%20AND%20ALLIED%20CONDITIONS%20ICD%2D9%20Code,World%20Health%20Organization%20(WHO).*

2. Cozzolino, F., Abraha, I., Orso, M., Mengoni, A., Cerasa, M. F., Eusebi, P., Ambrosio, G., & Montedori, A. (2017, March 29). Protocol for validating cardiovascular and Cerebrovascular ICD-9-CM codes in healthcare administrative databases: The Umbria Data Value Project. BMJ open. https://pmc.ncbi.nlm.nih.gov/articles/PMC5372118/

3. *Diseases of the circulatory system ICD-9 code range (390-459)*. ICD-9 Code DISEASES OF THE CIRCULATORY SYSTEM 390-459- Codify By AAPC. (n.d.). https://www.aapc.com/codes/icd9-codes-range/53/?srsltid=AfmBOopevh5Z7GbkMcnK57yvd_9B1xSewMu_CzIKNn8H4LY8Wk_nF7ep

4. *Ischemic heart disease ICD-9 code range 410-414*. ICD-9 Code ISCHEMIC HEART DISEASE 410-414- Codify By AAPC. (n.d.). https://www.aapc.com/codes/icd9-codes-range/57/

5. *Medicare claims synthetic public use files (synpufs)*. CMS.gov. (n.d.). https://www.cms.gov/data-research/statistics-trends-and-reports/medicare-claims-synthetic-public-use-files

6. Other forms of heart disease ICD-9 code range 420-429. ICD-9 Code OTHER FORMS OF HEART DISEASE 420-429- Codify By AAPC. (n.d.). https://www.aapc.com/codes/icd9-codes-range/59/?srsltid=AfmBOoqdVnW6hCCibfhN3EtxbpYB3ygvZpoayXwTjLgdLHxiNrUoZb0t

ALY6980 CAPSTONE

THANK YOU!