Final Project

1. Introduction

This project investigates the factors influencing hospital readmission rates among diabetic patients using a dataset spanning ten years (1999-2008) from 130 US hospitals. Each record reflects a patient diagnosed with diabetes who underwent laboratory tests, received medications, and stayed in the hospital for up to 14 days. The primary objective is to assess the likelihood of early readmission within 30 days post-discharge. Understanding these factors enables healthcare providers to enhance discharge planning and reduce readmission rates, ultimately improving patient outcomes and managing healthcare costs.

2. Data Collection & Cleaning

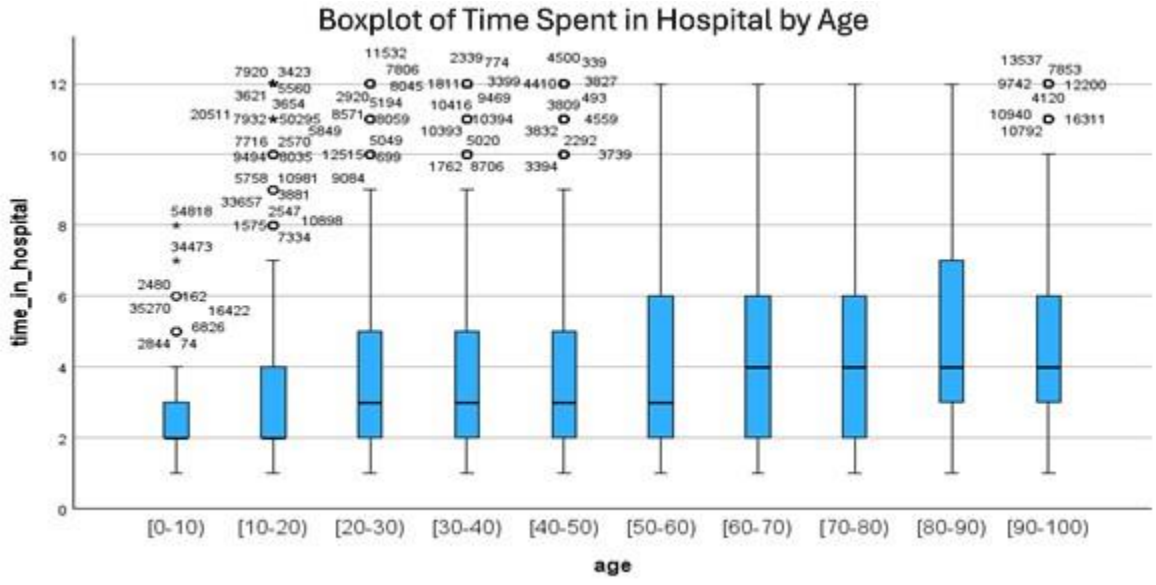
I extracted and cleaned data from hospital records, focusing on over 50 features that detail patient demographics and clinical outcomes. The data cleaning process involved several steps to ensure the dataset's reliability:

- Addressed Missing Values: Instead of deleting rows with missing values, I replaced missing values for numeric data with the average value and used "unknown" for categorical data, preserving the dataset's integrity.
- Eliminated Irrelevant Columns: I removed columns that did not contribute to the analysis, streamlining the dataset.
- Applied Binary Coding: I encoded categorical variables such as gender, race, and medical specialties into binary format, facilitating logistic regression modeling.

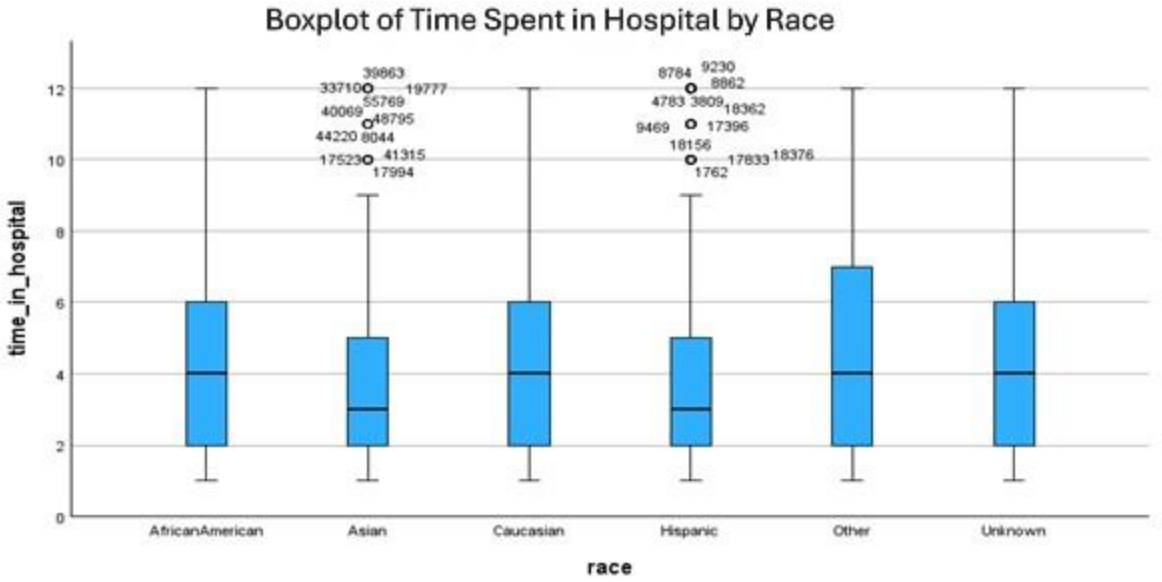
These steps prepared the dataset for an accurate analysis and interpretation, allowing us to uncover more meaningful insights.

3. Exploratory Data Analysis (EDA)

Data analysis key trends:



- Age and Length of Stay: I observed an association between age and length of hospital stay. Younger individuals (0-10 years) average 2.6 days, while those aged 80-90 average 4.95 days, highlighting the increased healthcare needs of older populations.



- Race and Hospital Stay Duration: The analysis revealed that race impacts hospital stay durations, with Asian patients exhibiting the shortest average stays and the "Other" racial group showing the longest. Notably, Caucasian and Hispanic patients display similar median stays but with greater variability. These findings emphasize the need to explore additional factors influencing time in the hospital beyond racial categories.

4. Modeling (Logistic Regression Results)

I developed two logistic regression models to assess the impact of various predictors on hospital readmission rates:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	age			164.320	9	<.001	
	0-10	-1.058	.216	24.100	1	<.001	.347
	10-20	-.002	.101	.000	1	.984	.998
	20-30	.309	.079	15.179	1	<.001	1.362
	30-40	.193	.065	8.843	1	.003	1.213
	40-50	.281	.057	24.095	1	<.001	1.325
	50-60	.248	.055	20.394	1	<.001	1.281
	60-70	.356	.054	43.327	1	<.001	1.428
	70-80	.400	.054	55.752	1	<.001	1.492
	80-90	.384	.055	48.716	1	<.001	1.468
	gender			11.777	2	.003	
	Female	20.684	24673.230	.000	1	.999	961584989.49
	Male	20.629	24673.230	.000	1	.999	910436104.72
	race			207.728	5	<.001	
	AfricanAmerican	.820	.063	168.457	1	<.001	2.270
	Asian	.590	.129	20.999	1	<.001	1.804
	Caucasian	.847	.062	189.091	1	<.001	2.332
	Hispanic	.904	.083	117.347	1	<.001	2.469
	Other	.572	.094	36.855	1	<.001	1.772
	Constant	-21.918	24673.230	.000	1	.999	.000

a. Variable(s) entered on step 1: age, gender, race.

Model 1 emphasized the roles of age and race. The results revealed that:

- Age: Patients aged 60 and older face significantly higher odds of readmission.
- Race: White patients experience an 11.1% higher likelihood of readmission (Exp(B) = 1.111).

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	medical_specialty			784.122	42	<.001	
	Allergya	2.305	1.084	4.520	1	.033	10.024
	Anesthes	-.413	.408	1.023	1	.312	.662
	Cardiolo	.286	.091	9.823	1	.002	1.332
	Dentistr	1.592	1.159	1.888	1	.169	4.914
	Dermatol	21.773	40192.969	.000	1	1.000	2856481870.6
	Emergenc	.733	.093	62.197	1	<.001	2.080
	Endocrin	-.101	.211	.229	1	.632	.904
	Family/G	.446	.090	24.686	1	<.001	1.562
	Gastroen	.319	.133	5.757	1	.016	1.375
	Gynecolo	-1.617	.616	6.882	1	.009	.199
	Hematolo	.575	.156	13.595	1	<.001	1.776
	Hospital	1.275	.696	3.355	1	.067	3.579
	Infectio	.959	.421	5.183	1	.023	2.609
	Internal	.278	.088	9.941	1	.002	1.320
	Nephrolo	.778	.102	58.443	1	<.001	2.177
	Neurolog	-.321	.202	2.527	1	.112	.725
	Obsteric	-.434	.455	.911	1	.340	.648
	Obstetri	-.716	.132	29.310	1	<.001	.489
	Oncology	.406	.143	8.103	1	.004	1.502
	Ophthalm	.127	.466	.074	1	.786	1.135
	Orthoped	-.135	.098	1.903	1	.168	.874
	Osteopat	1.310	.431	9.244	1	.002	3.706
	Otolaryn	-.608	.242	6.311	1	.012	.545
	Outreach	-.223	.871	.066	1	.798	.800
	Patholog	1.161	.516	5.062	1	.024	3.192
	Pediatri	-.280	.128	4.785	1	.029	.756
	Perinato	-20.593	40192.970	.000	1	1.000	.000
	Physical	.124	.136	.835	1	.361	1.132
	Physicia	1.698	1.161	2.138	1	.144	5.462
	Podiatry	.727	.319	5.207	1	.023	2.069
	Proctolo	-20.829	40192.970	.000	1	1.000	.000
	Psychiat	.301	.114	7.009	1	.008	1.352
	Psycholo	.316	.271	1.353	1	.245	1.371
	Pulmonol	.451	.113	15.826	1	<.001	1.569
	Radiolog	.587	.141	17.265	1	<.001	1.798
	Rheumato	.849	.593	2.052	1	.152	2.337
	Speech	-20.861	40192.970	.000	1	1.000	.000
	SportsMe	21.576	40192.969	.000	1	1.000	2347127553.7
	Surgeon	-.233	.394	.348	1	.555	.793
	Surgery-	.103	.092	1.252	1	.263	1.109
	Surgical	.937	.918	1.042	1	.307	2.552
	Unknown	.508	.087	34.069	1	<.001	1.662
	Insulin_Coded	.196	.016	151.150	1	<.001	1.217
	race			192.062	5	<.001	
	AfricanAmerican	.769	.064	146.664	1	<.001	2.158
	Asian	.561	.130	18.630	1	<.001	1.753
	Caucasian	.809	.062	170.046	1	<.001	2.247
	Hispanic	.841	.084	99.843	1	<.001	2.319
	Other	.494	.095	27.003	1	<.001	1.638
	Constant	-1.379	.106	170.707	1	<.001	.252

a. Variable(s) entered on step 1: medical_specialty, Insulin_Coded, race.

Model 2 added medical specialties and insulin usage as predictors. Key findings included:

• Medical Specialties: Specialties such as Allergy and Emergency Medicine significantly elevate the odds of readmission, while Gynecology decreases them.
• Insulin Usage: Insulin use raises the odds of readmission by 21.7% (Exp(B) = 1.217).

Both models consistently highlight race as a critical factor, showing that demographic variables significantly impact patient outcomes.

5. Conclusion & Recommendations

The analysis reveals that the race and length of hospital stay are significant predictors of hospital readmission. Notably, White patients exhibit a higher risk of readmission, and extended hospital stays correlate with an increased likelihood of returning for care. Based on these insights, I recommend the following actions for healthcare providers:

- Enhance Discharge Planning: Implement targeted follow-up care strategies for patients with longer hospital stays, focusing particularly on those identified as high-risk based on race and age.
- Conduct Further Research: Explore whether similar patterns emerge across various racial and ethnic groups to deepen our understanding of these findings and refine healthcare strategies accordingly.

In conclusion, while this analysis highlights critical predictors of hospital remission, it underscores the necessity to research a more diverse patient populations to improve care strategies. Specifically, we must acknowledge that age and race significantly influence readmission rates among diabetic patients. We can conclude that patients over 60 years old face a greater risk of readmission and longer hospital stays, with Caucasian and Hispanic patients experiencing median readmission rates, while Asian patients show the lowest rates.