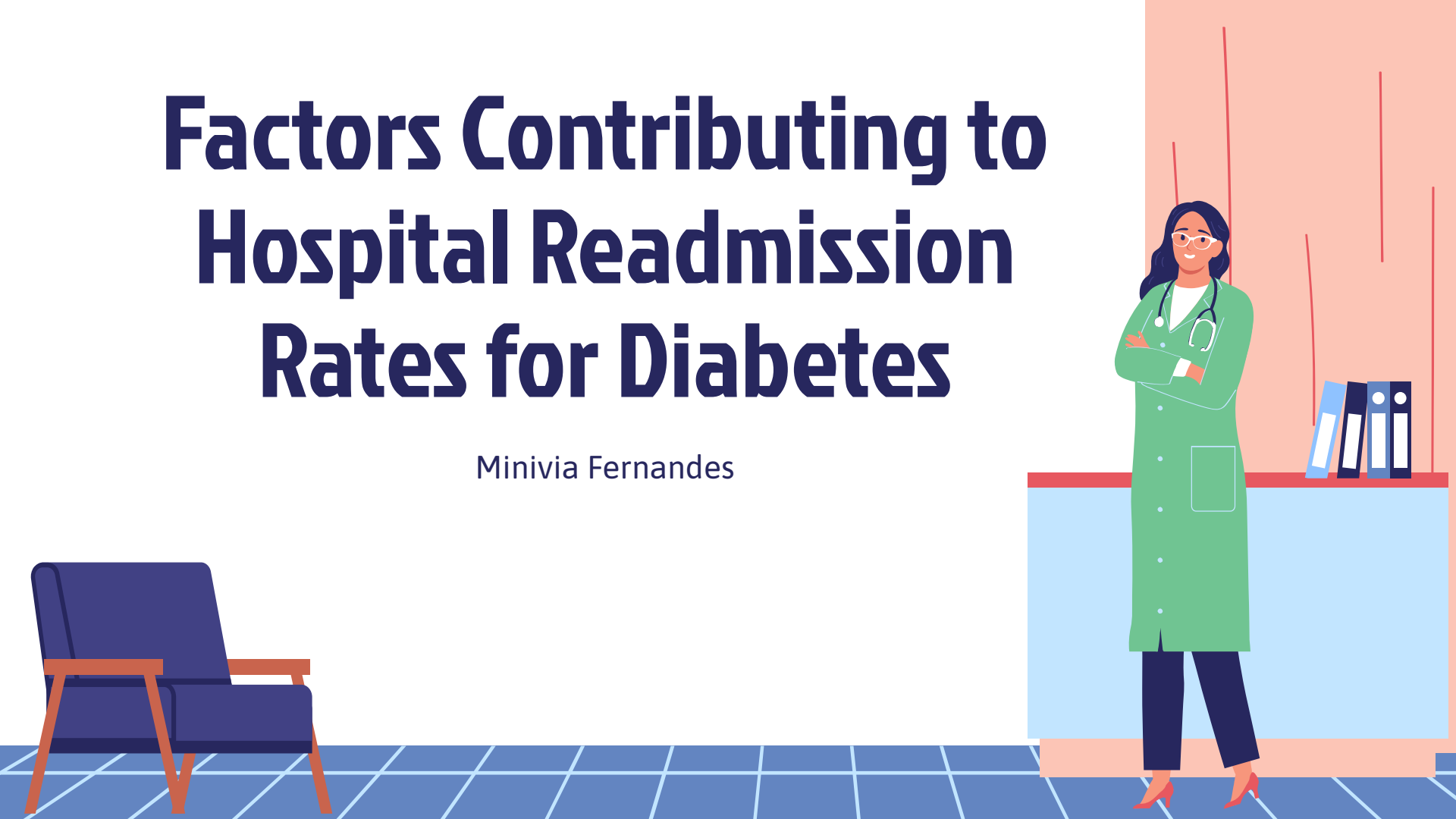
Factors Contributing to Hospital Readmission Rates for Diabetes

Minivia Fernandes



The Problem

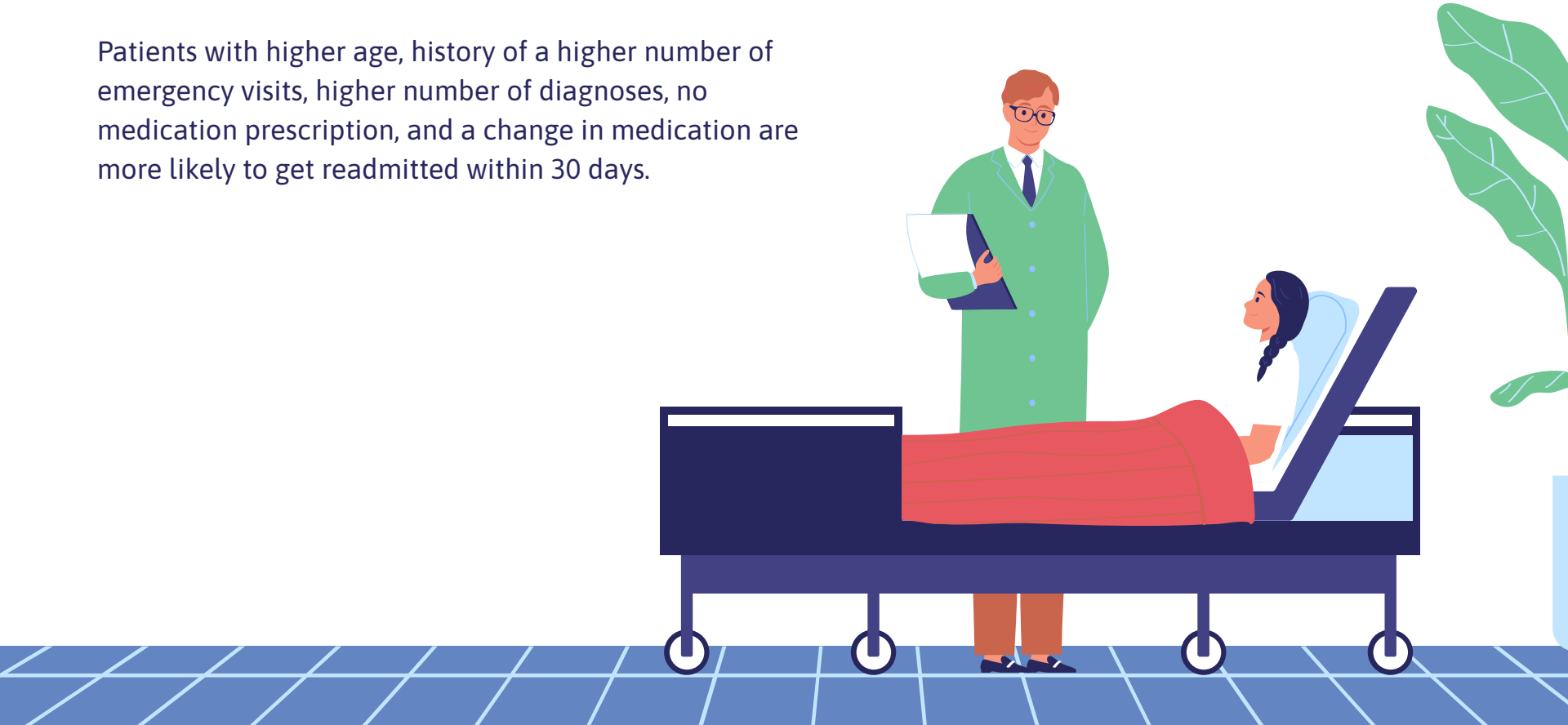
- Diabetes is the 7th leading cause of death and affects about 37.3 million people in the U.S.
- Cost for hospital readmissions <30 days is ~$25 billion per year in the U.S.
- <30 day readmission rates for diabetes are reported to be between 14.4% and 22.7%, much higher than the rate for all hospitalized patients (8.5%–13.5%)

The Opportunity

Early identification of patients facing a high risk of readmission can allow healthcare providers to conduct additional investigations and possibly prevent future readmissions.

Hypothesis

Patients with higher age, history of a higher number of emergency visits, higher number of diagnoses, no medication prescription, and a change in medication are more likely to get readmitted within 30 days.



Dataset

- **Represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks**
- **Includes 101,766 entries & 53 features representing patient and hospital outcomes**
- **Extracted information of encounters meeting the following criteria:**

a.	It is an inpatient encounter (a hospital admission)
b.	It is a diabetic encounter (one during which any kind of diabetes was entered to the system as a diagnosis)
c.	The length of stay was at least 1 day and at most 14 days
d.	Laboratory tests were performed during the encounter
e.	Medications were administered during the encounter

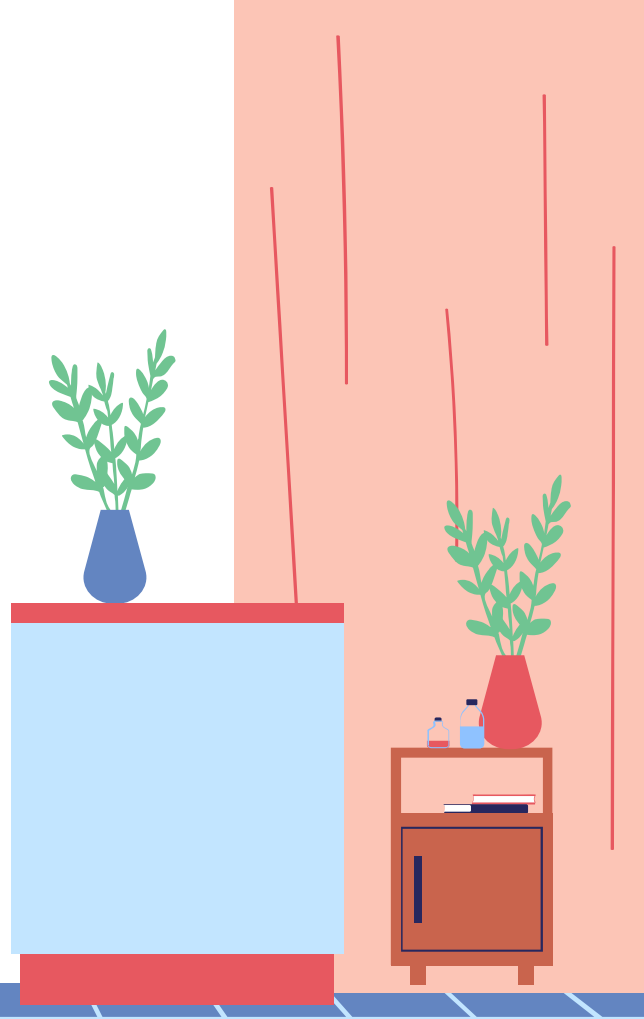
Dataset

Mix of categorical and numeric data:

- encounter_id and patient_nbr are unique identifiers
- age and weight are categorical in this data set
- admission_type_id, discharge_disposition_id, admission_source_id are numerical here, but are IDs. They should be considered categorical.

Processing the Data

- Mismatched data types
- Inconsistent strings
- Misspelled words, inconsistent capitalization
- Extra spaces and characters
- Nulls



Nulls

Null values are not directly present in the data. It is subbed with '?' character.

High % of missing values and should be dropped:

- Weight
- Medical Specialty
- Payer Code

Minor % of missing values and can be used:

- Race
- Diag_1
- Diag_2
- Diag_3

```
(((diabetes_df == '?').sum(axis=0)/len(diabetes_df))*100).sort_values(ascending=False).head(7)
```

```
weight                96.858479
medical_specialty     49.082208
payer_code            39.557416
race                   2.233555
diag_3                 1.398306
diag_2                 0.351787
diag_1                 0.020636
dtype: float64
```

```
=COUNTIF(F2:F101767, "?")/COUNTA(F2:F101767)
```

Exploring the Data

Readmission

- About 11,357 patients were readmitted within 30 days
- 35,545 patients were readmitted after 30 days
- Majority of patients were not readmitted

Exploring the Data

How many encounters by patient?

- The average number of encounters by patient was approximately 1.42 encounters by patient
- There were some outliers (ie: 1 patient having 40 encounters)

```
query = """Select
     patient_nbr,
     count(distinct encounter_id) as encounters
     from diabetes_df
     group by patient_nbr
     order by encounters desc"""

# Run the query
patient_encounters = sqldf(query)
print(patient_encounters)
```

	patient_nbr	encounters
0	88785891	40
1	43140906	28
2	88227540	23
3	23199021	23
4	1660293	23
...	...	...
71513	1305	1
71514	927	1
71515	774	1
71516	729	1
71517	378	1

```
query ="""Select AVG(encounters) as average_encounters
    from (Select
     patient_nbr,
     count(distinct encounter_id) as encounters
     from diabetes_df
     group by patient_nbr
     order by encounters desc)"""

avg_patient_encounters = sqldf(query)
print(avg_patient_encounters)
```

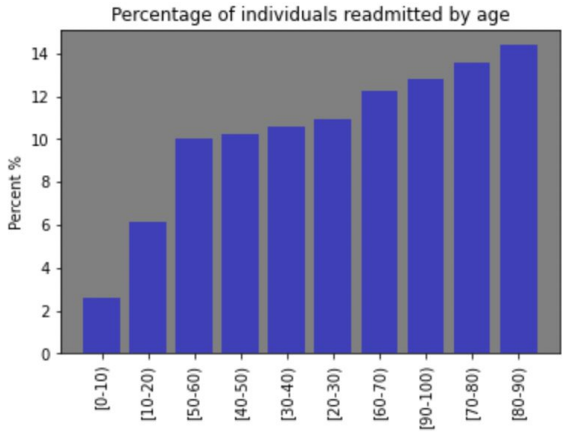
	average_encounters
0	1.422942

Exploring the Data

Age

```
# Age distribution of patients against readmission within 30 days
query3 = """Select age,
        count (distinct patient_nbr) as total_individuals,
        count(distinct (case when readmitted="<30"then patient_nbr else 0 end)) as readmitted_individuals
       from diabetes_df group by 1
       order by readmitted_individuals desc"""
age_readmitted_patients = sqldf(query3)
age_readmitted_patients["percentage_of_individuals_readmitted"]= age_readmitted_patients['readmitted_individuals']/age_
age_readmitted_patients= age_readmitted_patients.sort_values("percentage_of_individuals_readmitted")
print(age_readmitted_patients)
```

- Significant increase in percentage at 50
- More likely to be readmitted within 30 days as age increases

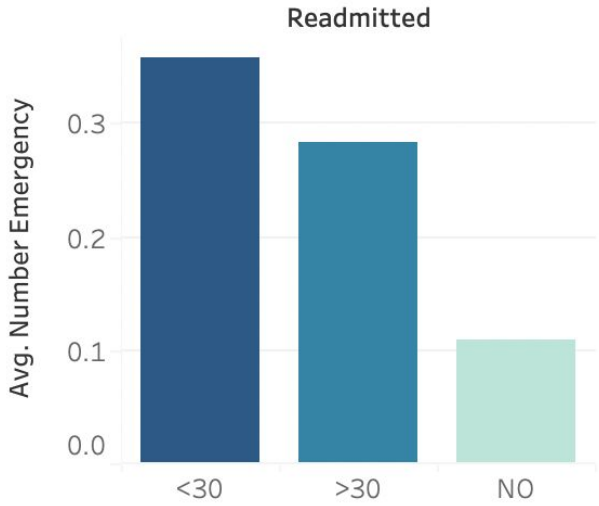


	age	total_individuals	readmitted_individuals	\
9	[0-10)	154	4	
8	[10-20)	536	33	
3	[50-60)	12666	1266	
4	[40-50)	6956	713	
5	[30-40)	2727	289	
7	[20-30)	1138	124	
1	[60-70)	16281	1989	
6	[90-100)	2042	261	
0	[70-80)	18584	2516	
2	[80-90)	12008	1726	

	percentage_of_individuals_readmitted
9	2.597403
8	6.156716
3	9.995263
4	10.250144
5	10.597726
7	10.896309
1	12.216694
6	12.781587
0	13.538528
2	14.373751

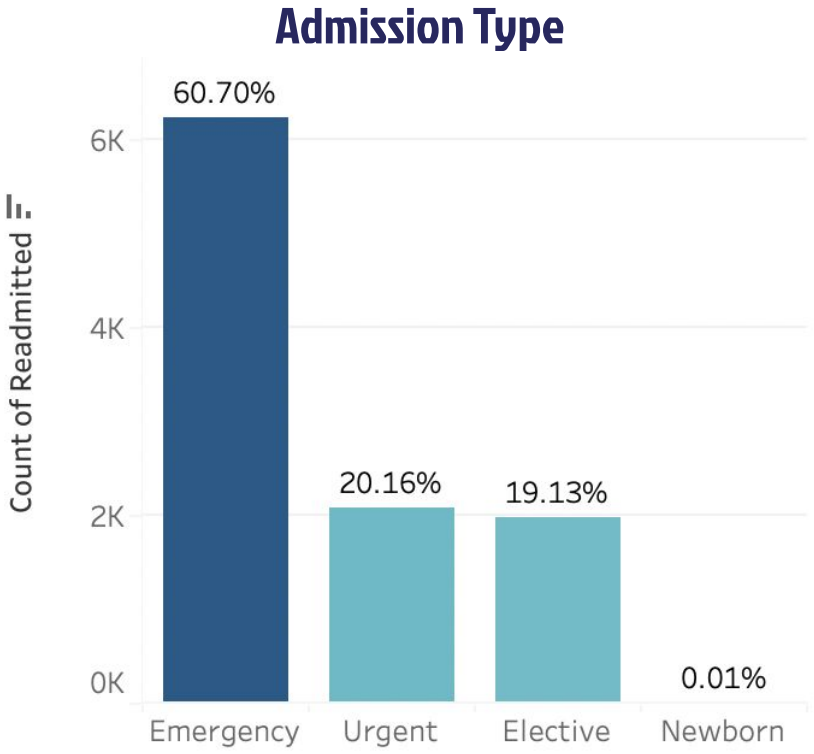
Exploring the Data

Mean Number of Emergency Visits



- Patients readmitted within 30 days had a higher average of emergency visits

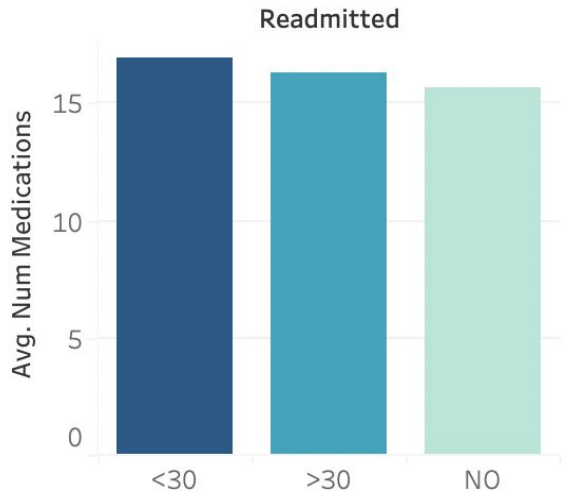
Which admission type is the most prevalent in readmission <30 days?



- Emergency admissions account for 60% of readmission <30 days encounters

Exploring the Data

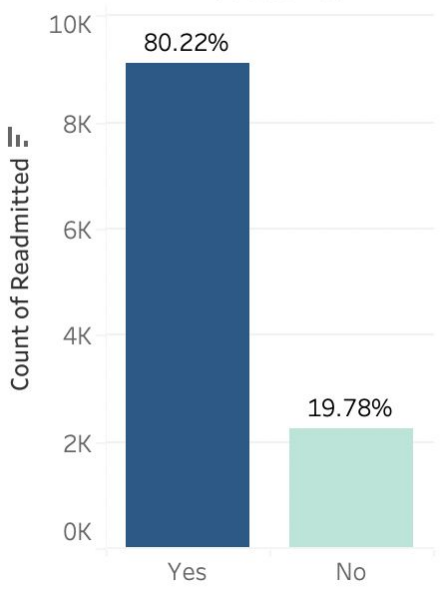
Mean Number of Medications



- Readmissions have higher averages of medications prescribed

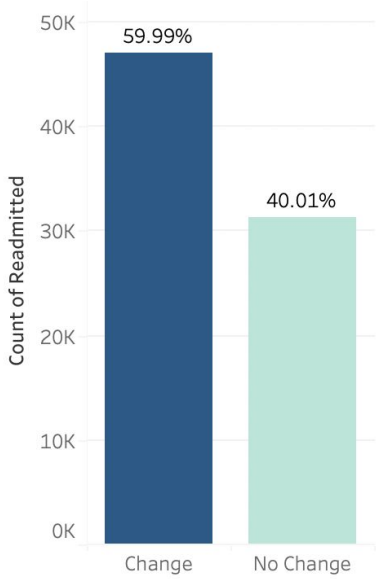
Effects of Medication

Prescribed Diabetes Medication



➢ *Patients readmitted within 30 days were more likely to have been prescribed medication

Change in Diabetes Medication



➢ Change in medication accounts for ~60% of readmission <30 days encounters

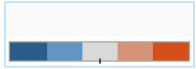
Exploring the Data

Average Number of Medications by Race and Age

Race	[0-10)	[10-20)	[20-30)	[30-40)	[40-50)	[50-60)	[60-70)	[70-80)	[80-90)	[90-100)
African American	6.38	7.65	12.21	14.04	15.26	16.09	16.33	15.50	14.64	12.73
Asian	8.00	5.00	10.33	10.60	11.81	13.85	13.91	13.49	12.82	13.55
Caucasian	6.17	8.60	11.95	14.38	15.61	16.94	17.50	16.63	15.44	13.93
Hispanic	4.00	8.91	11.38	13.02	14.03	13.62	14.51	14.97	14.50	13.00
Other	6.50	7.10	11.52	12.15	14.58	15.26	16.04	15.98	14.97	13.79

Average Number of Diagnoses by Race and Age

Race	[0-10)	[10-20)	[20-30)	[30-40)	[40-50)	[50-60)	[60-70)	[70-80)	[80-90)	[90-100)
African American	2.938	3.823	5.817	6.348	6.861	7.154	7.329	7.380	7.514	7.603
Asian	2.500	2.000	4.667	4.733	6.058	6.854	7.000	7.561	7.633	7.727
Caucasian	2.669	4.005	5.967	6.576	7.065	7.321	7.586	7.730	7.957	7.947
Hispanic	3.500	4.478	5.857	6.178	6.531	6.633	7.274	7.383	7.929	8.000
Other	2.750	3.900	5.519	6.476	6.266	7.045	7.382	7.592	7.920	8.083



Conclusion

- Readmission status within 30 days indicates that something went wrong or was missed in the patient's first visit
- Features we explored: age, number of emergency visits, prescribed medication, and change in medication
- Major features correlating with higher readmission rates within 30 days...
  - Higher age (80-90 years)
  - Higher number of emergency visits
  - Admission type (emergency visits)
  - Taking medication
  - Change in medications

Questions?