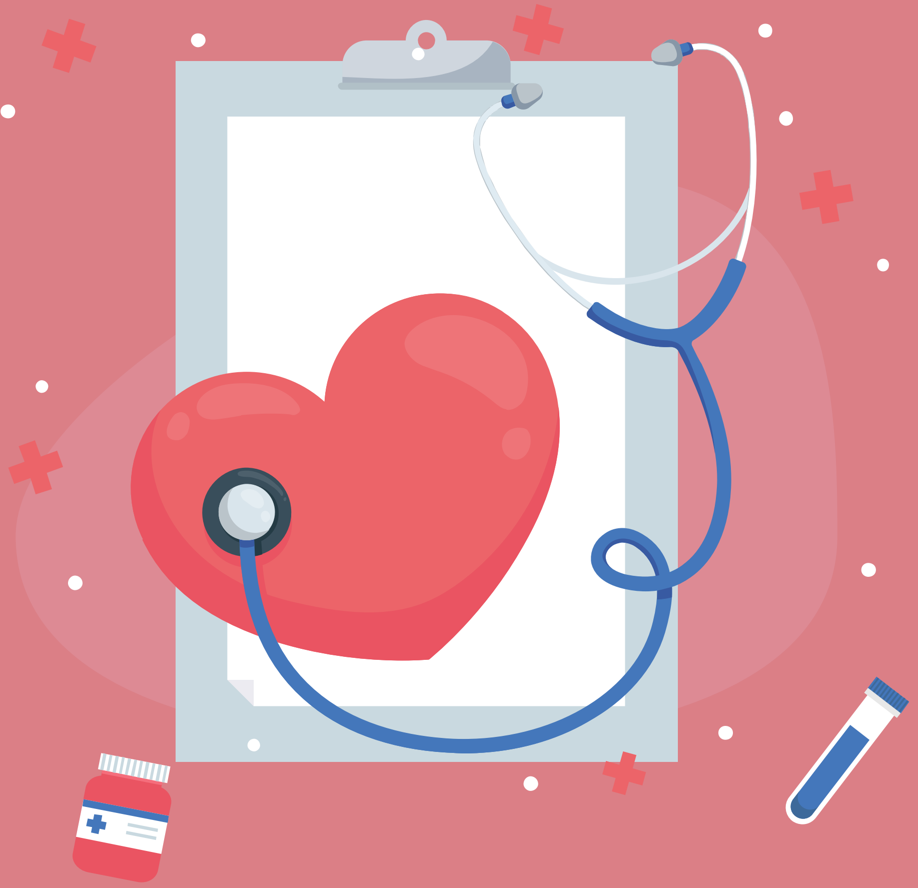
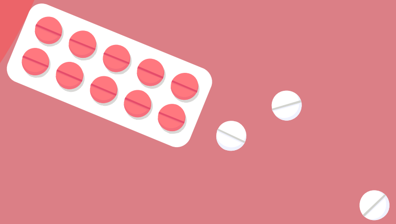
PREVENTING HOSPITAL READMISSIONS

By predicting which patients are most at risk





Reasons for this evaluation

Healthcare costs have been rising. There are many reasons this is concerning.

1. Rising costs for individuals and families reduces their ability to make other purchases, both large (a home, for example) and small (sports equipment for a child, or a new refrigerator).
2. As much of the medical care in the United States is paid by either government programs (like Medicare) or medical insurance, this is also big business.
3. Hospitalizations are one of the highest costs. While emergencies are hard to predict or avoid, re-admissions can be reduced.

Problem Statement

It is vital that patients who are most likely to be readmitted be identified as early as possible, so that the reasons for that increased risk of returning to the hospital can be identified and mitigated. This will help avoid increased expense for the patient, disruption of their life, increased cost for the healthcare system, and the potential of occupancy and staffing issues for the facility.

Steps

Prioritization

Labeling a patient as at-risk when they are safe is a result **preferred over** labeing a patient safe when they are actually at risk

Model Creation

After much data cleaning, multiple models will be tested to find the one that best meets the priority

Classification

The model can then be run on additional patient data to classify them as safe or at-risk

Data Included

Data set provided by
the University of California Irvine Machine Learning Repository

- Encounter_id
- Patient_nbr
- Race
- Gender
- Age
- Weight
- Admission_type_id
- Discharge_disposition_id
- Admission_source_id
- Time_in_hospital
- Payer_code
- Medical_specialty
- Num_lab_procedures
- Num_procedures
- Num_medications
- Number_outpatient
- Number_emergency
- Number_inpatient
- Diag_1
- Diag_2
- Diag_3
- Number_diagnoses
- Max_glu_serum
- A1Cresult
- *Information on 23 diabetes medicines*
- Change
- DiabetesMed
- Readmitted

Data Cleaning

What to modify

- For all medications, change to binary
  - 1 – if patient taking it
  - 0 – if patient not taking it
- Change the readmission status (target variable)
  - 1 – if readmitted within 30 days
  - 0 – otherwise

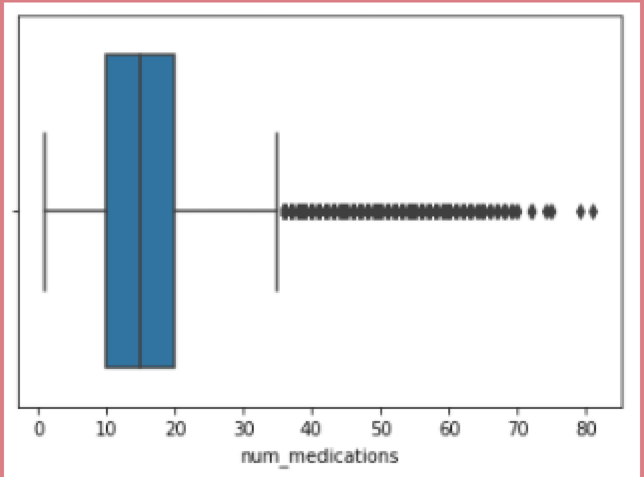
What to eliminate

- Patients sent to other medical facilities
- Patients who died
- Patients with no diagnosis code
- Max_glu_serum and A1Cresult (both over 80% missing)
- All admission data is dropped as many of the observations were missing this information

Data Analysis

Number of medications had a high variance from minimum to maximum (1 to 81)
A quick view via boxplot shows that there don't appear to be any true outliers.

This feature was not adjusted.

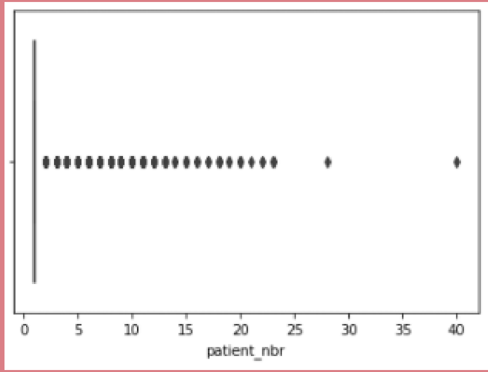


Data Analysis

Any observations where patient numbers matched represent people
who were hospitalized more than once during the decade covered by the data.

It's important that patients who were "regulars" don't skew the data.
After review, it was determined that only 97 patients were hospitalized more than once a year on average.

This feature was not adjusted.



Data Visualization

Once data cleaning was done,
there were 91,397 observations left to work with.

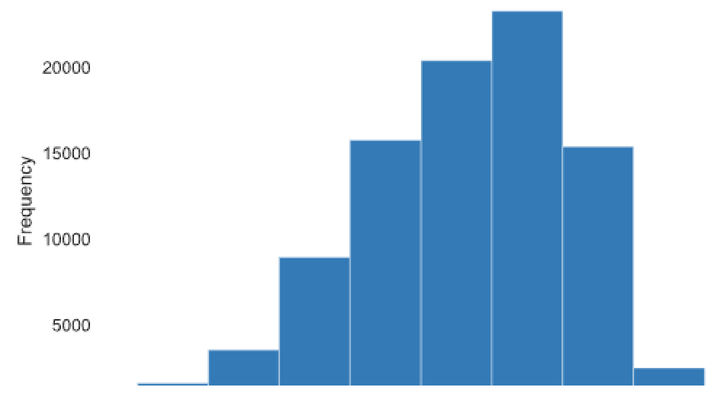
We can see that 81,215 (or 88.86%) were not readmitted.

readmitted
Categorical

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	714.2 KiB

Data Visualization

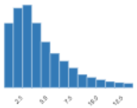
age_group Real number ($\mathbb{R}_{\geq 0}$)				
	Distinct	8	Mean	61.17925096
	Distinct (%)	< 0.1%	Minimum	20
	Missing	0	Maximum	90
	Missing (%)	0.0%	Zeros	0
	Infinite	0	Zeros (%)	0.0%
	Infinite (%)	0.0%	Memory size	714.2 KiB



Data Visualization

time_in_hospital
Real number ($\mathbb{R}_{\geq 0}$)

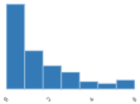
Distinct	14	Mean	4.35053667
Distinct (%)	< 0.1%	Minimum	1
Missing	0	Maximum	14
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	714.2 KiB



num_procedures
Real number ($\mathbb{R}_{\geq 0}$)

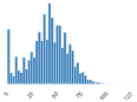
ZEROS

Distinct	7	Mean	1.33701325
Distinct (%)	< 0.1%	Minimum	0
Missing	0	Maximum	6
Missing (%)	0.0%	Zeros	41781
Infinite	0	Zeros (%)	45.7%
Infinite (%)	0.0%	Memory size	714.2 KiB



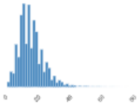
num_lab_proc…
Real number ($\mathbb{R}_{\geq 0}$)

Distinct	117	Mean	42.82208388
Distinct (%)	0.1%	Minimum	1
Missing	0	Maximum	132
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory size	714.2 KiB



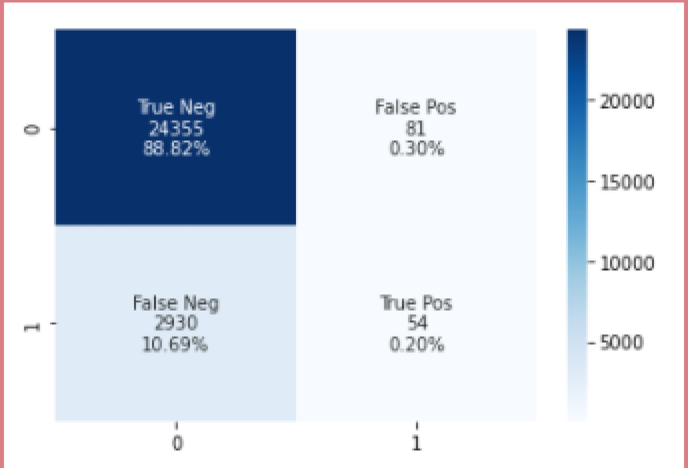
num_medicati…
Real number ($\mathbb{R}_{\geq 0}$)

Distinct	75	Mean	16.05129271
Distinct (%)	0.1%	Minimum	1
Missing	0	Maximum	81
Missing (%)	0.0%	Zeros	0
Infinite	0	Zeros (%)	0.0%
Infinite (%)	0.0%	Memory	714.2 KiB



Results

Results from Random Forest with h2o in Python



THANK YOU FOR YOUR TIME

Debbie Hunton
Bellevue University
dhunton@my365.bellevue.edu