PREDICTIVE MODELS IN ICU DATABASES

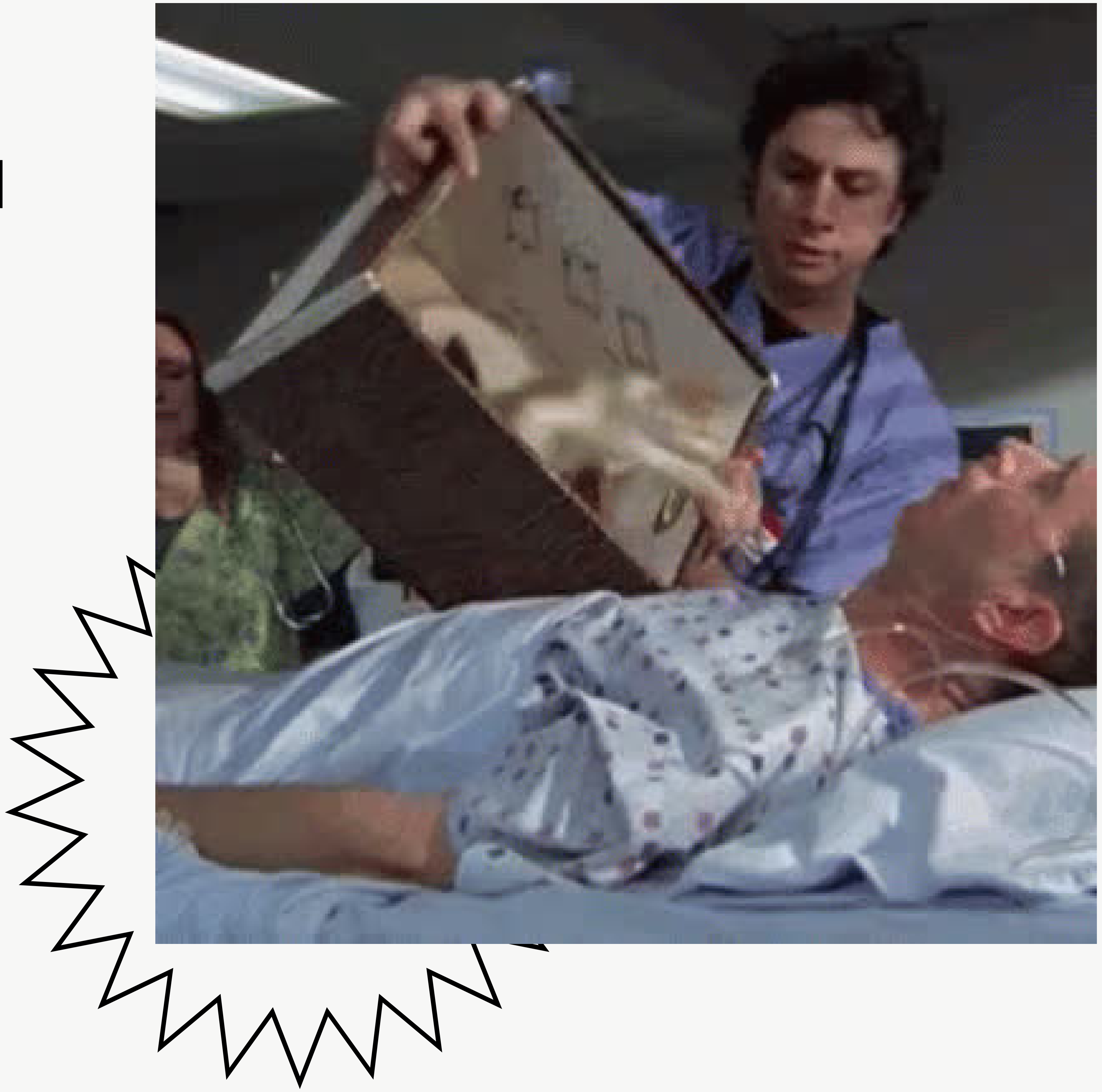
ALAIN, DARLENE, MARIAH

DATA 4381: DATA CAPSTONE

THE UNIVERSITY OF TEXAS AT ARLINGTON
COLLEGE OF SCIENCE - DIVISION OF DATA SCIENCE

BACKGROUND: DISPARITIES BEHIND THE DATA

- ICU - Intensive Care Unit
- Healthcare disparities hidden behind the data, for example:
  - Frequency of vitals
  - High false positive diagnoses in minorities
- Is everyone getting the same care?
- Difference in care = difference in outcome



WHAT'S MISSING? WHERE ARE THE GAPS?

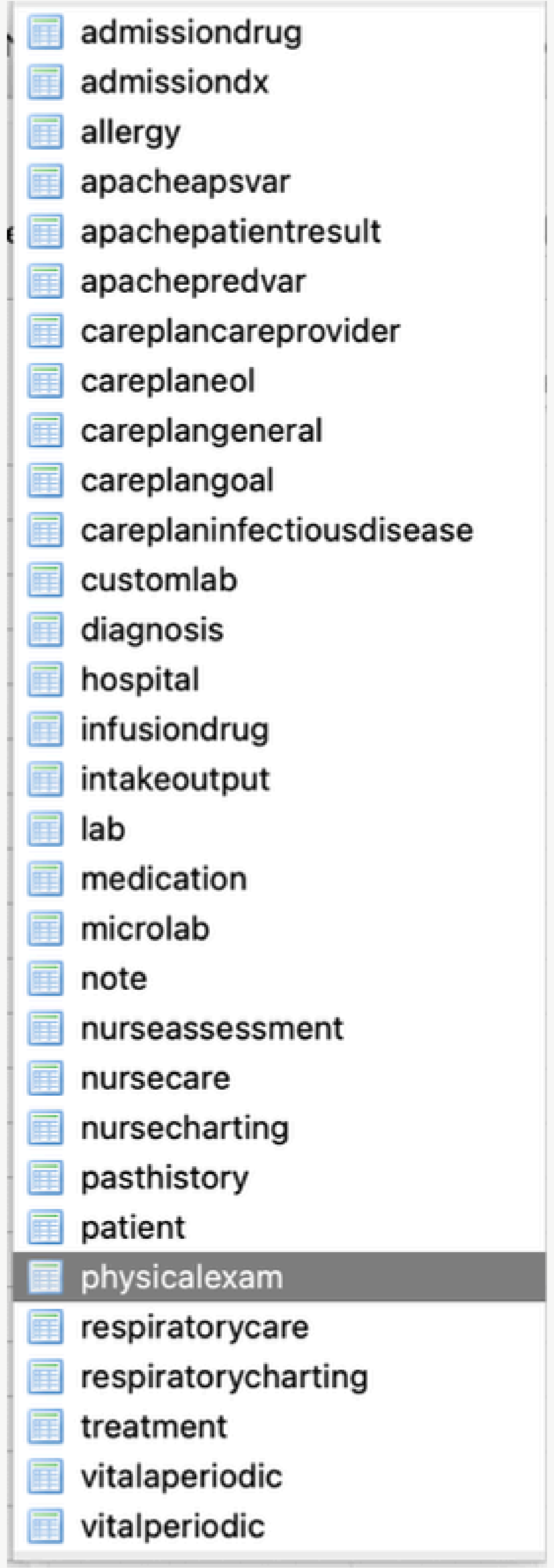
- What factors should be used against bias?
  - Several factors are missed, including charting and site
  - Not robust enough
- Performance vs Bias?
  - Imbalanced vs Rebalanced?
- Outcome Distribution?

OUR OBJECTIVE

- Objective: To use machine learning methods to evaluate ICU mortality or readmission risk based on demographic and physiological variables
  - To use these models to explore any bias (using fairness metrics) that may be present and create solutions from the start
  - Aiming for 70% Recall across both classes
- **Research Question: What patterns of performance disparity or bias are present in ICU mortality and readmission models when comparing patients of different demographics or hospital contexts, and how might these nuances influence clinical decision-making?**
- Expected outcomes:
  - Nuanced, robust, contextual models
  - Better clinical decision-making
  - Foundation for inclusive healthcare policies
  - Driving change in standard of care in hospitals

DATASET OVERVIEW

- **Source:** eICU Collaborative Research Database (Demo v2.0.1, PhysioNet)
- **Scope:** Multi-hospital ICU dataset, 31 relational tables
- **Focus for this project:** 11 key tables covering:
  - Demographics (Patient)
  - Mortality outcomes (APACHE IVa predictions & observed)
  - Clinical data (Diagnoses, Admission reasons)
  - Time-series (Labs, Vitals, Nurse notes)
  - System-level context (Hospital info, Fluid balance)
- **Scale:** Thousands to millions of rows depending on table



KEY TABLES

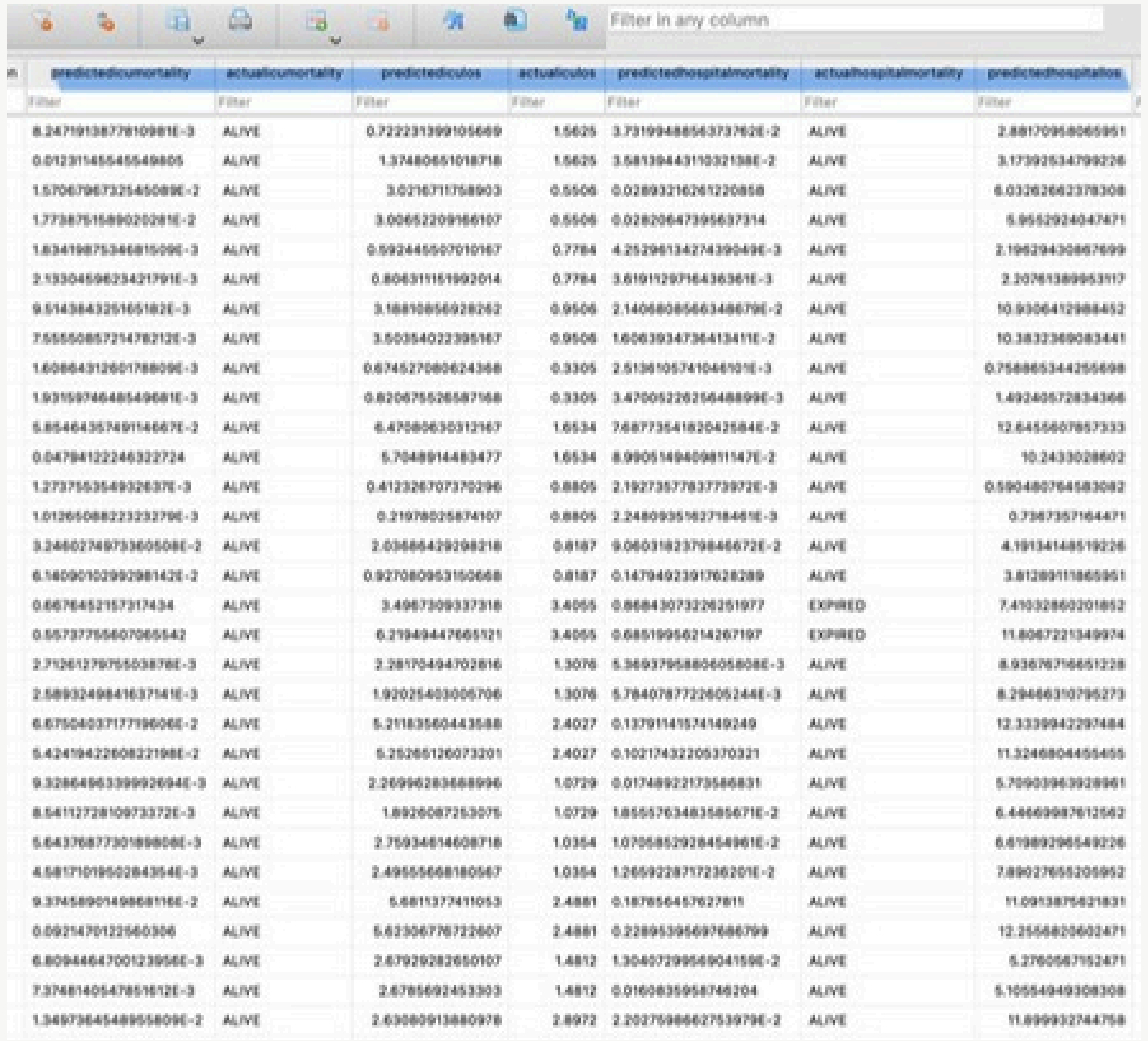
```python
tables = {
    "patient": patient,
    "apachePatientResult": apachePatientRes,
    "apachePredVar": apachePredVar,
    "apacheApsVar": apacheApsVar,
    "diagnosis": diagnosis,
    "admissionDx": admissionDx,
    "lab": lab,
    "vitalPeriodic": vitalPeriodic,
    "nurseCharting": nurseCharting,
    "intakeOutput": intakeOutput,
    "hospital": hospital,
}
```

LOOKING IN TO FOCUS ON THE FOLLOWING 11 TABLES:

- **Patient** 👉 demographics (age, sex, ethnicity), admission/discharge times, discharge status
- **apachePatientResult** 👉 predicted ICU mortality (APACHE IVa) and observed outcomes
- **apacheApsVar & apachePredVar** 👉 physiological and scoring variables used in predictions
- **Diagnosis & AdmissionDx** 👉 admission reasons, illness severity
- **Lab, VitaPeriodic, NurseCharting** 👉 labs (glucose, creatinint, BUN), vitals (heart rate, SaO2) and nurse recorded notes
- **IntakeOutput** 👉 fluid balance tracking
- **Hospital** 👉 metadata (region, type, size)

WHAT IS APACHE IVa?

- It is the standard ICU severity scoring system
- Combines age, vital signs, and lab results
- Estimates probability of death during ICU stay



What the apachePatient table provides:

- Predicted mortality risk (APACHE IVa)
- Observed outcome (Alive / Expired)
- Cross checked with patient table discharge status

TABLE SIZES & SCOPE

- There is a wide range in size amongst the tables
- Tables range from **hundreds** 👉 **millions of rows**
- Mix of **categorical** (ethnicity, diagnosis) **and numeric** (labs, vitals) features

	table	rows	cols
0	vitalPeriodic	1634960	19
1	nurseCharting	1477163	8
2	lab	434660	10
3	intakeOutput	100466	12
4	diagnosis	24978	7
5	admissionDx	7578	6
6	apachePatientResult	3676	23
7	patient	2520	29
8	apachePredVar	2205	51
9	apacheApsVar	2205	26
10	hospital	186	4

PATIENT TABLE CHARACTERISTICS

Major role in project:

- Foundation for startifying outcomes (by age, sex, race, site)
- Links with hospital, diagnosis, and APACHE results

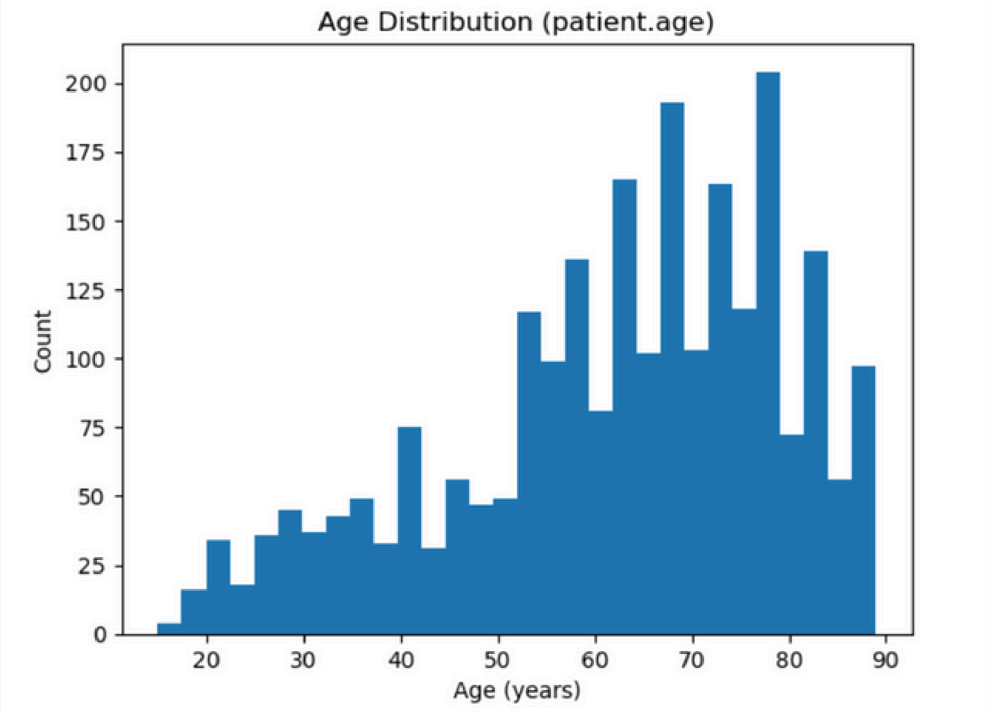
Table: patient

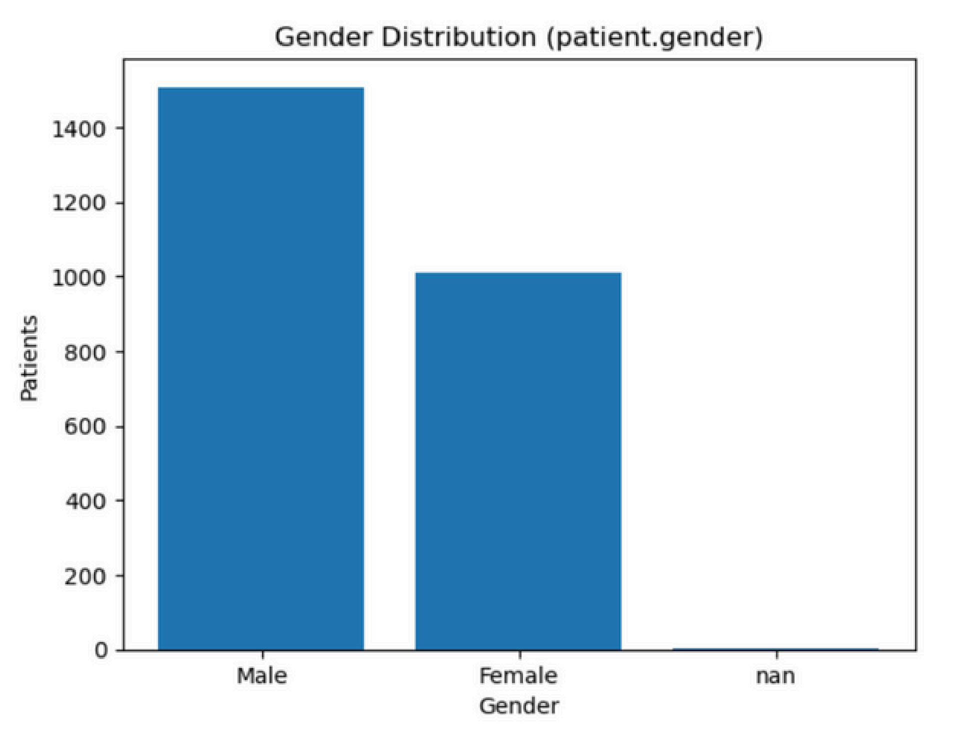
	gender	age	ethnicity	hospitalid	wardid	apacheadmissiondx	admissionheight	hospitaladmittime24	hospitaladmitoffset	hospitaladmitsource	hospitaldischargeyear	hospitaldischargetime24
1	Female	87	Caucasian	59	91		157.5	23:36:00	-2258		2015	19:20:00
2	Female	87	Caucasian	59	91	Rhythm disturbance (atrial, supraventricular)	157.5	23:36:00	-8		2015	19:20:00
3	Male	76	Caucasian	68	103	Endarterectomy, carotid	167	20:46:00	-1	Operating Room	2014	17:05:00
4	Female	34	Caucasian	56	82	Overdose, other toxin, poison or drug	172.7	01:44:00	-23	Emergency Department	2015	21:05:00
5	Male	61	Caucasian	68	103	GI perforation/rupture, surgery for	177.8	23:48:00	-10	Emergency Department	2014	15:41:00
6	Female	55	Caucasian	63	95		157.5	23:23:55	-495	Operating Room	2015	17:48:00
7	Female	55	Caucasian	63	95	Endarterectomy, carotid	157.5	23:23:55	-19	Operating Room	2015	17:48:00
8	Female	60	Hispanic	67	109	Coma/change in level of consciousness (for hepatic see G...	154.9	05:06:00	0		2015	23:08:00
9	Male	28	Caucasian	61	120	Overdose, other toxin, poison or drug	182.9	18:02:00	-1	Emergency Department	2015	15:15:00
10	Female	34	Caucasian	68	103		165.1	05:37:41	-121	Emergency Department	2015	14:05:00
11	Female	34	Caucasian	68	103		165.1	05:37:41	-114	Emergency Department	2015	14:05:00
12	Female	> 89	Caucasian	60	83	Infarction, acute myocardial (MI)	157.5	19:05:53	-8	Emergency Department	2015	20:50:00
13	Female	> 89	Caucasian	60	83		157.5	19:05:53	-1187	Emergency Department	2015	20:50:00
14	Female	59	Caucasian	66	90	Sepsis, cutaneous/soft tissue	149.9	17:05:00	-1	Emergency Department	2014	03:38:00
15	Male	44	Caucasian	68	103	GI perforation/rupture, surgery for	172.7	18:43:00	-252	Operating Room	2014	19:10:00
16	Female	66	Caucasian	73	97	Sepsis, pulmonary	165.1	10:02:00	-23	Emergency Department	2014	20:20:00
17	Female	41	Caucasian	71	87	Respiratory - medical, other	170.2	01:56:00	0	Emergency Department	2014	20:07:00
18	Male	63	Caucasian	59	91	Bleeding, lower GI	172.7	16:43:00	-36	Emergency Department	2014	19:51:00
19	Female	57	Caucasian	56	82	Knee replacement, total (non-traumatic)	157.5	17:40:00	-8	Operating Room	2014	18:38:00
20	Male	87	Caucasian	60	83	Sepsis, pulmonary	172.7	08:46:00	-1	Emergency Department	2014	20:30:00
21	Female	52	Caucasian	58	108	Emphysema/bronchitis	160	03:05:00	-1	Emergency Department	2015	15:30:00
22	Female	23	Caucasian	63	95	GI medical, other	162.6	14:36:00	-3331	Floor	2015	19:39:00
23	Male	73	Caucasian	56	82	Rhythm disturbance (atrial, supraventricular)	180.3	20:41:00	0	Emergency Department	2015	21:30:00
24	Male	39	Caucasian	60	83	Embolus, pulmonary	193	20:24:00	-1117	Floor	2015	18:06:00
25	Male	39	Caucasian	60	83		193	20:24:00	-3177	Floor	2015	18:06:00
26	Female	57	Caucasian	63	95			13:18:00	-196		2015	17:00:00
27	Female	57	Caucasian	63	95			13:18:00	-197		2015	17:00:00
28	Female	20	Caucasian	61	120	Overdose, antidepressants (cyclic, lithium)	162.6	22:58:00	-202	Emergency Department	2014	15:48:00
29	Male	29	Caucasian	60	83		188	04:49:00	-334		2014	17:55:00
30	Male	29	Caucasian	60	83	Overdose, sedatives, hypnotics, antipsychotics, ...	188	04:49:00	-226		2014	17:55:00

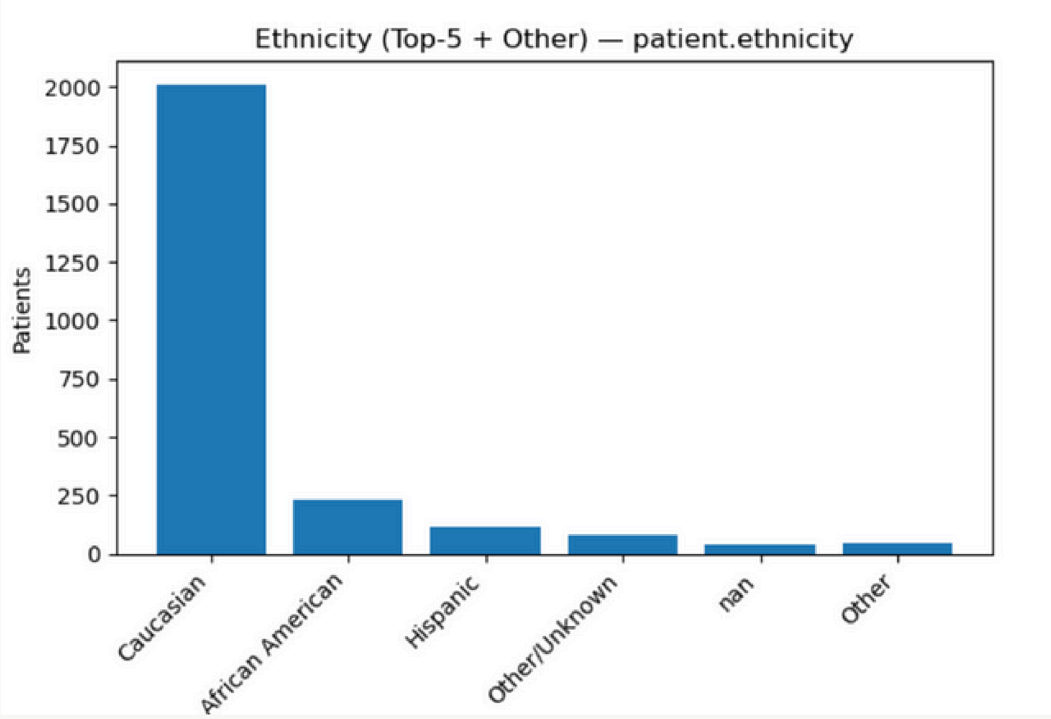
Table: patient

	et	hospitaldischargelocation	hospitaldischargestatus	unittype	unitadmittime24	unitadmitsource	unitvisitnumber	unitstaytype	admissionweight	dischargeweight	unitdischargetime24	unitdischargeoffset	unitdischargelocation	unitdischarg
1	66	Home	Alive	Med-Surg ICU	13:14:00	ICU to SDU	2	stepdown/other			18:58:00	344	Home	Alive
2	16	Home	Alive	Med-Surg ICU	23:44:00	Emergency Department	1	admit	46.5	45	13:14:00	2250	Step-Down Unit (SDU)	Alive
3	18	Home	Alive	SICU	20:47:00	Operating Room	1	admit	77.5	79.4	10:00:00	793	Floor	Alive
4	38	Other Hospital	Alive	Med-Surg ICU	02:07:00	Emergency Department	1	admit	60.3	60.7	20:48:00	1121	Other External	Alive
5	63	Home	Alive	SICU	23:58:00	Operating Room	1	admit	91.7	93.1	22:47:00	1369	Floor	Alive
6	10	Home	Alive	Med-Surg ICU	07:38:00	ICU to SDU	2	stepdown/other			17:48:00	610	Home	Alive
7	86	Home	Alive	Med-Surg ICU	23:42:00	Operating Room	1	admit	72.5	72.5	07:38:00	476	Step-Down Unit (SDU)	Alive
8	62	Home	Alive	Med-Surg ICU	05:06:00	Emergency Department	1	admit	95.6	97.6	20:47:00	2381	Floor	Alive
9	72	Other	Alive	Med-Surg ICU	18:03:00	Emergency Department	1	admit	91.8	91.9	15:11:00	1268	Floor	Alive
10	87	Other	Alive	SICU	07:38:00	ICU to SDU	2	stepdown/other		61	14:05:00	387	Other	Alive
11	94	Other	Alive	SICU	07:31:00	Emergency Department	1	admit	60.7	60.8	07:38:00	7	Step-Down Unit (SDU)	Alive
12	77	Skilled Nursing Facility	Alive	Med-Surg ICU	19:13:00	Emergency Department	1	admit	58.5	58.6	14:52:00	1179	Step-Down Unit (SDU)	Alive
13	98	Skilled Nursing Facility	Alive	Med-Surg ICU	14:52:00	ICU to SDU	2	stepdown/other		57.1	02:15:00	683	Floor	Alive
14	72	Death	Expired	Med-Surg ICU	17:06:00	Emergency Department	1	admit		73	02:50:00	4904	Floor	Alive
15	75	Home	Alive	SICU	22:55:00	Operating Room	1	admit		84.1	06:18:00	1883	Floor	Alive
16	55	Skilled Nursing Facility	Alive	MICU	10:25:00	Emergency Department	1	admit	86.8	79	20:05:00	3460	Floor	Alive
17	11	Home	Alive	Med-Surg ICU	01:56:00	Emergency Department	1	admit	81	80.6	00:20:00	4224	Floor	Alive
18	92	Home	Alive	Med-Surg ICU	17:19:00	Emergency Department	1	admit		101.4	19:04:00	1545	Home	Alive
19	70	Home	Alive	Med-Surg ICU	17:48:00	Operating Room	1	admit	120.1	123.4	18:39:00	1491	Floor	Alive
20	83	Home	Alive	Med-Surg ICU	08:47:00	Emergency Department	1	admit	86.18		20:30:00	3583	Home	Alive
21	84	Home	Alive	Med-Surg ICU	03:06:00	Emergency Department	1	admit	115.2	115.6	14:39:00	2133	Home	Alive
22	72	Home	Alive	Med-Surg ICU	22:07:00	Floor	1	admit	63.5	63.5	19:39:00	4172	Home	Alive
23	29	Home	Alive	Med-Surg ICU	20:41:00	Emergency Department	1	admit	86.2	85.9	21:30:00	2929	Home	Alive
24	24	Home	Alive	Med-Surg ICU	15:01:00	Floor	1	admit	163.8	164.6	01:21:00	2060	Step-Down Unit (SDU)	Alive
25	64	Home	Alive	Med-Surg ICU	01:21:00	ICU to SDU	2	stepdown/other		161.7	04:15:00	174	Floor	Alive
26	66	Home	Alive	Med-Surg ICU	16:34:00		1	admit		71.3	16:35:00	1	Step-Down Unit (SDU)	Alive
27	65	Home	Alive	Med-Surg ICU	16:35:00	ICU to SDU	2	stepdown/other		71.4	16:51:00	1456	Floor	Alive
28	08	Home	Alive	Med-Surg ICU	02:20:00	Emergency Department	1	admit	49.2	51.7	15:48:00	808	Home	Alive
29	52	Home	Alive	Med-Surg ICU	10:23:00	ICU to SDU	2	stepdown/other			15:53:00	330	Floor	Alive
30	60	Home	Alive	Med-Surg ICU	08:35:00	Emergency Department	1	admit	75	74.9	10:23:00	108	Step-Down Unit (SDU)	Alive
31	62	Home	Alive	Med-Surg ICU	18:33:00	Emergency Department	1	admit	111.1	111.1	15:14:00	1241	Home	Alive
32	60	Other Hospital	Alive	Med-Surg ICU	00:08:00	Emergency Department	1	admit	55.8	56	17:23:00	1035	Floor	Alive

PATIENT TABLE CHARACTERISTICS







Age Distribution:

- Broad adult coverage (20-90)
- Most patients range from 50-80

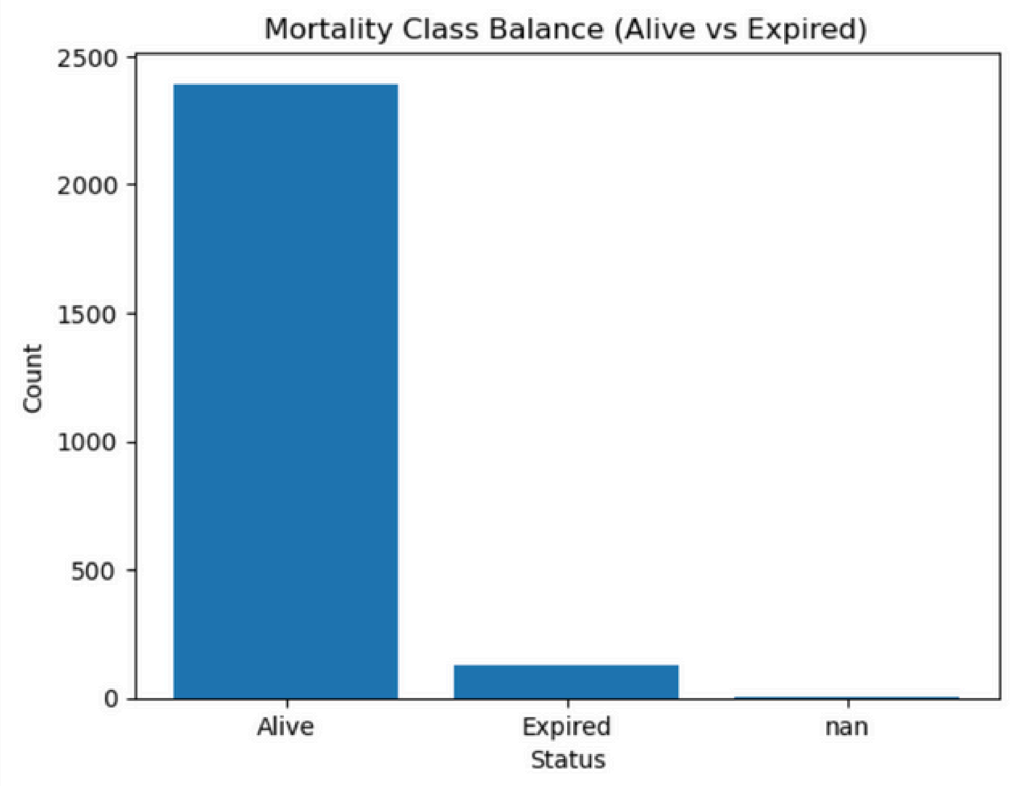
Gender Split:

- Slightly more males than females

Ethnicity:

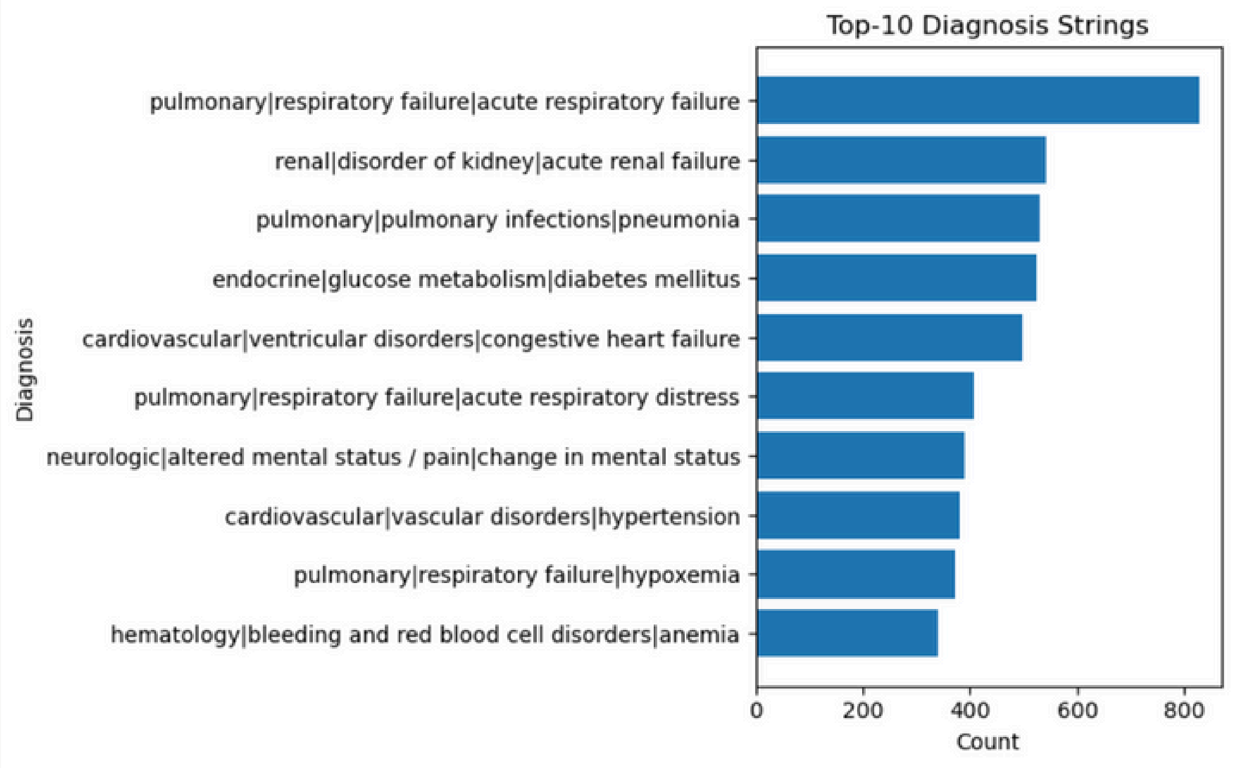
- Skewed (majority Caucasian, small minoriy groups)

MORTALITY, DIAGNOSIS, ADMISSION



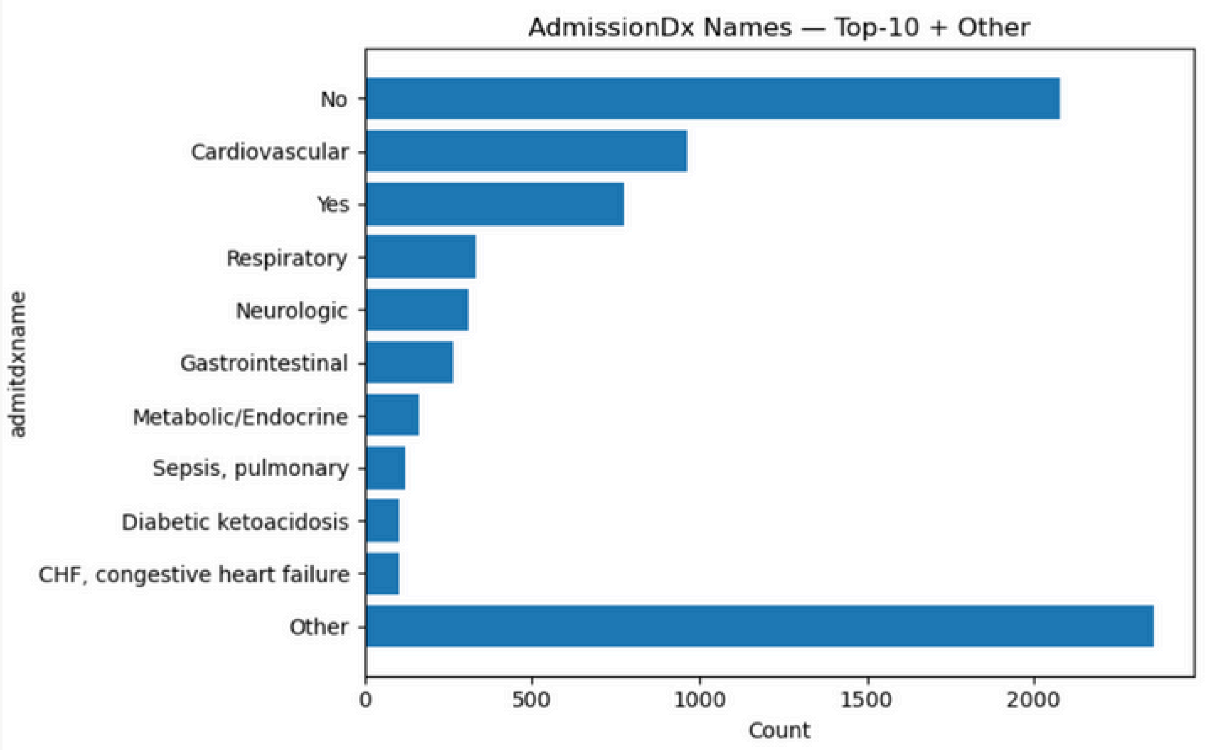
Mortality Outcomes:

- Strong class imbalance
- From apachePatientResult (predicted vs. observed)
- Cross-checked with patient discharge status



Diagnosis Reasons:

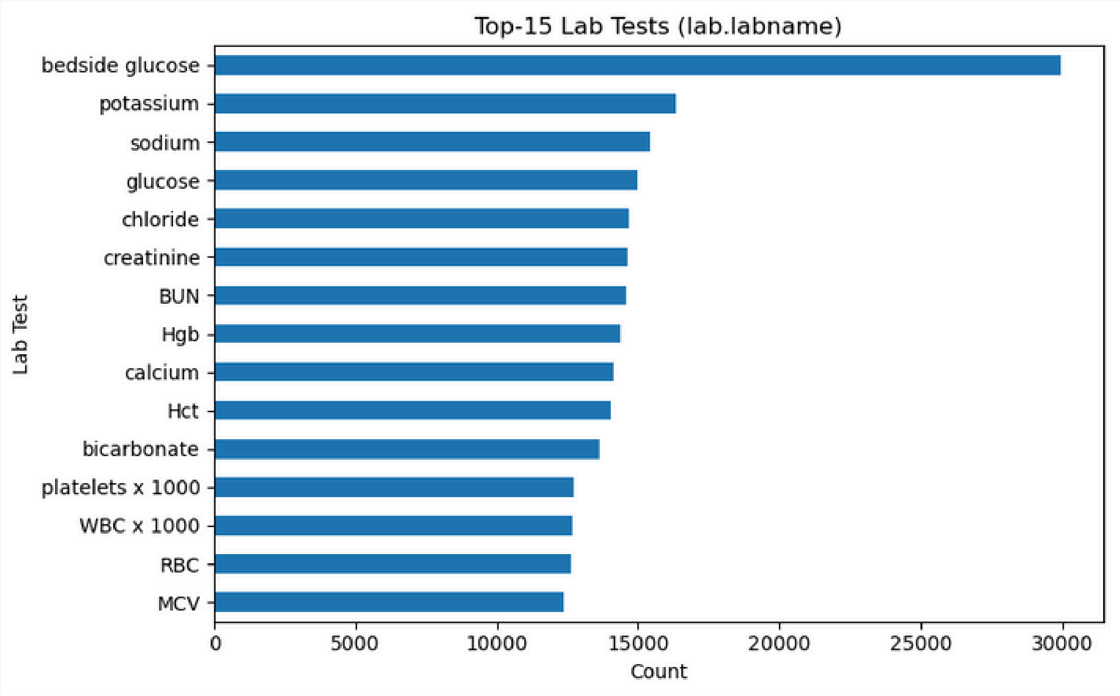
- Over 25k entries and 1000+ unique categories
  - High-cardinality → many rare categories
- Top conditions: respiratory failure, renal failure, infections, heart failure, diabetes

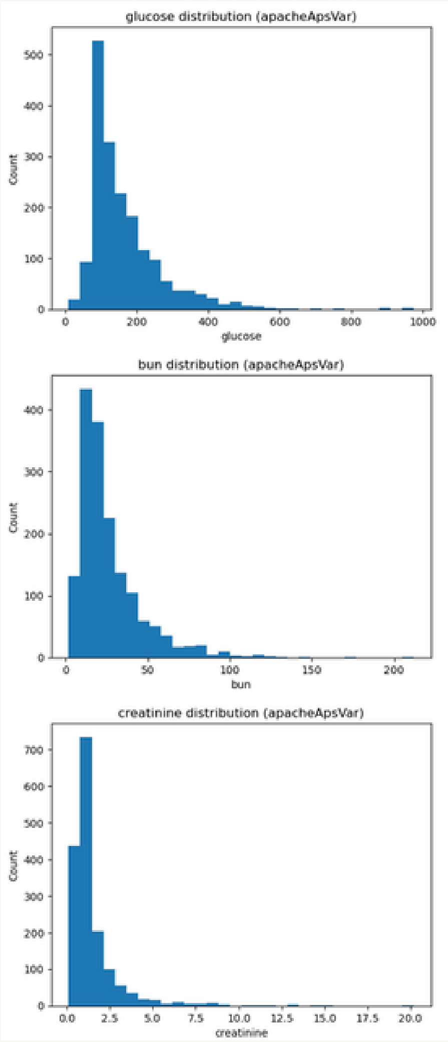


AdmissionDx:

- Top AdmissionDx groups: Cardiovascular, Respiratory, Neurologic, Gastrointestinal
- “Other” category very large → many rare diagnoses grouped
- Cleaning issue: “Yes/No” codes not clinically meaningful

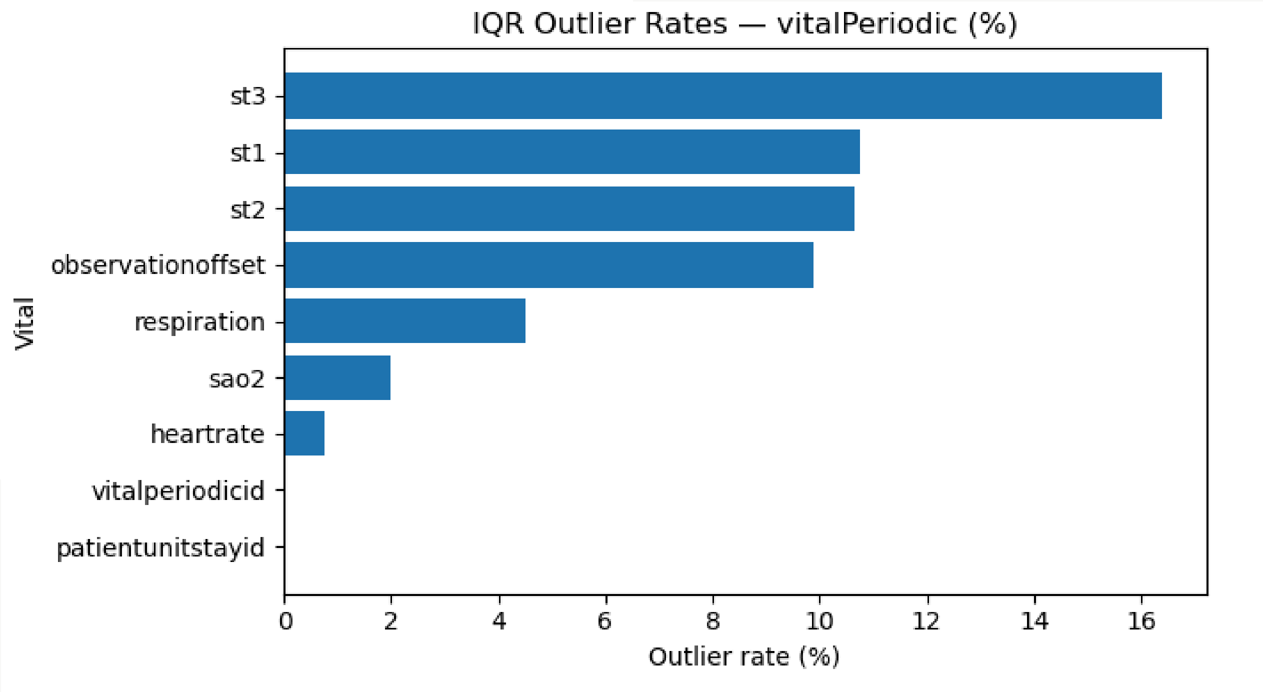
LABS & VITALS






```python
# check out nans
vital_missing = vitalPeriodic.isna().mean().sort_values(ascending=False)
display(vital_missing.head(15))
```

```
icp                   0.991931
padiastolic           0.980379
pasystolic            0.980377
pamean                0.980041
etco2                 0.955485
temperature           0.930995
cvp                   0.876276
systemicdiastolic     0.861482
systemicsystolic      0.861480
systemicmean          0.860302
st3                   0.635238
st1                   0.624904
st2                   0.598178
respiration           0.155116
sao2                  0.118205
```



LABS:

- Top tests: bedside glucose, sodium, potassium, creatinine, BUN, Hct, etc.
- Represent core ICU monitoring markers (renal, metabolic, hematologic).
- Distribution shows skew (e.g., glucose, creatinine)

VITALS:

- High missing values across many columns (more than 80% for some).
- The rest show expected ICU patterns but with extreme values.

CHALLENGES & CONSIDERATIONS

🚨 **CLASS IMBALANCE –** Mortality outcomes are heavily skewed (most survived)

🚨 **HIGH CARDINALITY CATEGORICAL FEATURES –** diagnosis & ethnicity have many rare categories

🚨 **LARGE "OTHER" BUCKETS –** AdmissionDx and diagnosis groupings aren't always clinically precise

🚨 **MISSING VALUES IN TIME-SERIES –** Vitals (esp. blood pressure, SaO_2) have over 70% missing

🚨 **OUTLIERS –** Extreme values in labs/vitals (glucose, creatinine, etc.)

METHODOLOGY

QUESTION: WHAT PATTERNS OF BIAS EXIST IN ICU **MORTALITY** AND **READMISSION** MODELS ACROSS PATIENT DEMOGRAPHICS AND HOSPITAL SETTING, AND HOW MIGHT THESE NUANCES INFLUENCE CLINICAL DECISION-MAKING?

GENERAL PLAN FOR CAPSTONE 1

- We have 2 target variables
  - Mortality: hospitaldischargestatus
  - Readmission: patientstayid
- Run model pipeline twice
  - Classification
- Evaluate metrics (recall focus)
- Basic bias/fairness check

Capstone 1 Pipeline

Raw data → EDA → Data preprocessing & cleaning → Modeling plan → Training strategy → Evaluation → Capstone 2

RAW DATA

- ☐ Choose which out of 31 tables to use
- ☐ Decide which features are most important
- ☐ Merge on “patientunitstayid”

EDA

- ☐ Inspect dataset shape
- ☐ Check for null/missing values
- ☐ Stats for numerical features
- ☐ Distributions for categorical/numerical
- ☐ Check class balance for both targets
- ☐ Check multicollinearity using VIF

PREPROCESSING

□ Verify data types
□ Impute missing values
□ Normalize/standardize features
□ Encode categorical features (one-hot or label)

MODEL PLAN

□ Create both target variables
□ Logistic regression for baseline
□ Random Forest/XGBoost (non-linear patterns)
□ Neural Networks (good with non-linear)
□ Clustering/KNN
  - Unsupervised, target variables not needed
  - Experiment to see if patients cluster based on features
□ LLM possibly

TRAINING

- ◻ 70/15/15 Train-Validation-Test Split
- ◻ 5-Fold Cross Validation
- ◻ Hyperparameter Tuning (grid search)
- ◻ Batch size for NN 32-64
- ◻ Epoch size for NN ~20-30

EVALUATION

- ◻ Primary metric is Recall
- ◻ Precision, F1, AUROC, Calibration plots
- ◻ Basic bias check (stratify metrics by demographics/hospital)
- ◻ Error analysis (SHAP influence)

CAPSTONE 2

- ☐ Full fairness assessment
- ☐ Feature engineering/model stacking
- ☐ Demonstrate usability
- ☐ Iterate if needed
- ☐ Poster/paper on findings

Baseline Model Results

Target: Mortality
0 if survived, 1 if expired

	Precision	Recall	F1
0	0.92	1	0.96
1	0	0	0
Accuracy			0.92
Macro Avg	0.46	0.50	0.48

Features: age, admission weight, discharge weight, gender, ethnicity, where the patient was admitted

CHALLENGES

- A LOT of data (31 tables with thousands of rows)
    - Merging the files together
        - Most relevant columns
        - Many features to choose from
- Some variables were not their correct type
    - Age → object
- Class imbalance
    - Medical data must keep integrity
- Did not use all the features we wanted to for baseline but it's okay

EXPECTED OUTCOMES

APPLICATIONS

- Can be used in throughout the entirety of healthcare collaboration (physicians, nurses, healthcare administrators, etc)
  - Targets where specifically our healthcare delivery systems are lacking
- Insights for policymaking
- More pertaining to training and etiquette

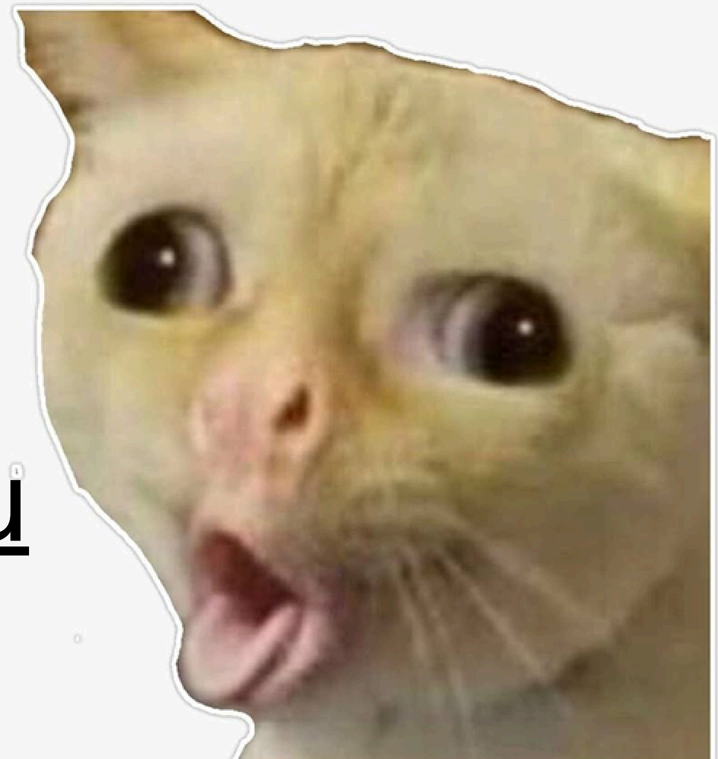
DELIVERABLES

- Poster Presentations
  - College of Science Discover Symposium 2026
  - UT System 2026 AI In Healthcare Symposium
- Publications
  - Stimulus: UTA Medical Humanities Journal
  - Leads for new research!

CONTACT US!

Alain Areeba Siddiqui
axs6903@mavs.uta.edu

Darlene Eligado
dae9134@mavs.uta.edu



Mariah Noelle Cornelio
mnc3287@mavs.uta.edu