Analysis and Prediction of Patient-Hospital Experience in US Medicare Hospitals

HANIEH MIRZAEE

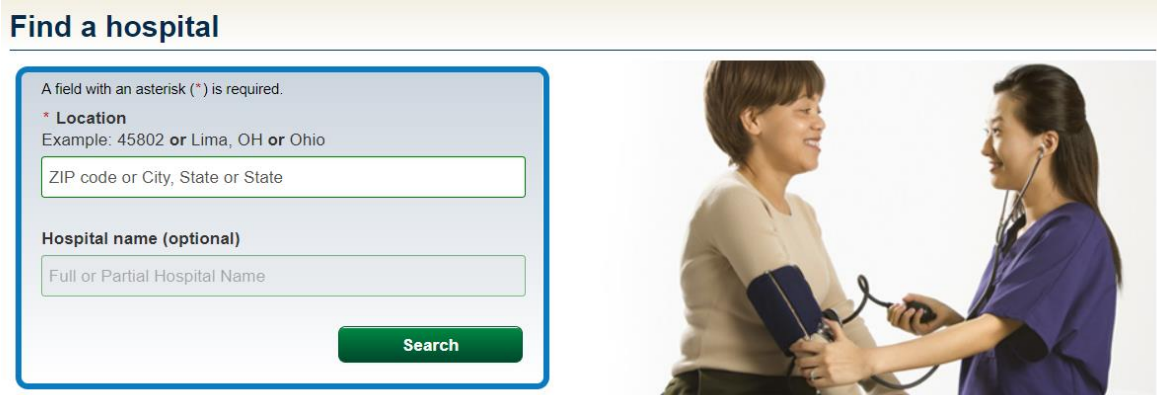
Roadmap

- Problem at hand
- Who cares?
- Project Data
- Data Curation and Wrangling
- Exploratory Data Analysis
- Predictive Modeling
- Conclusion and Remarks

Problem at Hand

Medicare.gov | Hospital Compare
The Official U.S. Government Site for Medicare

- **Hospital Compare a consumer oriented website**
  - Provides information on the hospitals quality of care
  - Helps patients make informed decision on their healthcare plans
- **Consumers can select hospitals and compare performance measures**
- **Medicare additionally provides patient satisfaction 5-star ratings based on patient surveys**



Problem at Hand

- Hospitals must have at least 100 completed surveys before they can be assigned a rating

Questions:

Is there a **relationship between survey ratings and hospital characteristics**?

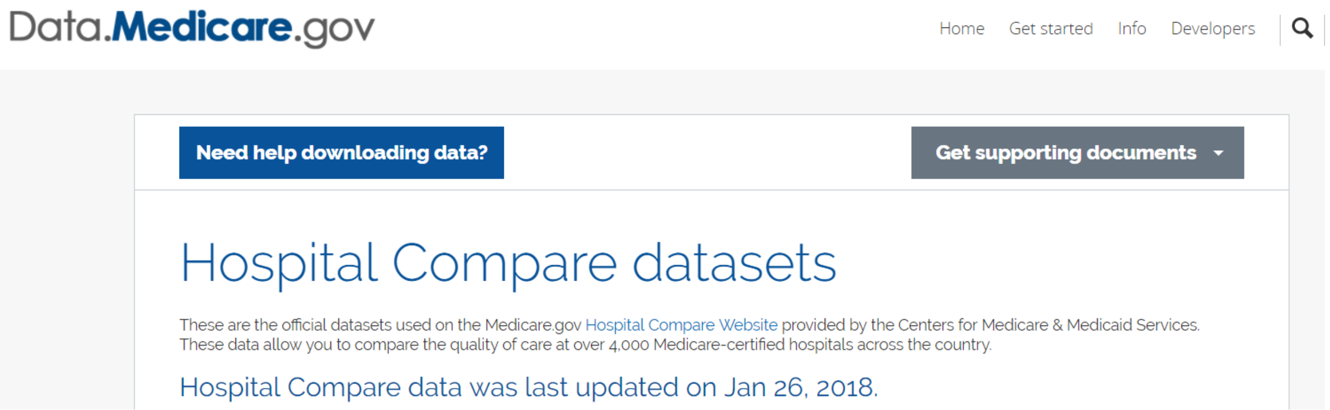
Can we somehow **predict these ratings** with confidence without performing the surveys?

Who Cares?

- **Consumers of the hospital Compare Website**
  - It would be valuable for the consumer to be able to query the website for such survey results
- **Medicare**
  - They'd need this information to estimate the payments to hospitals
- **Hospital owners and local county governments**
  - They'd need this information to estimate the Medicare reimbursements
  - Can exploit the information to improve the quality of care

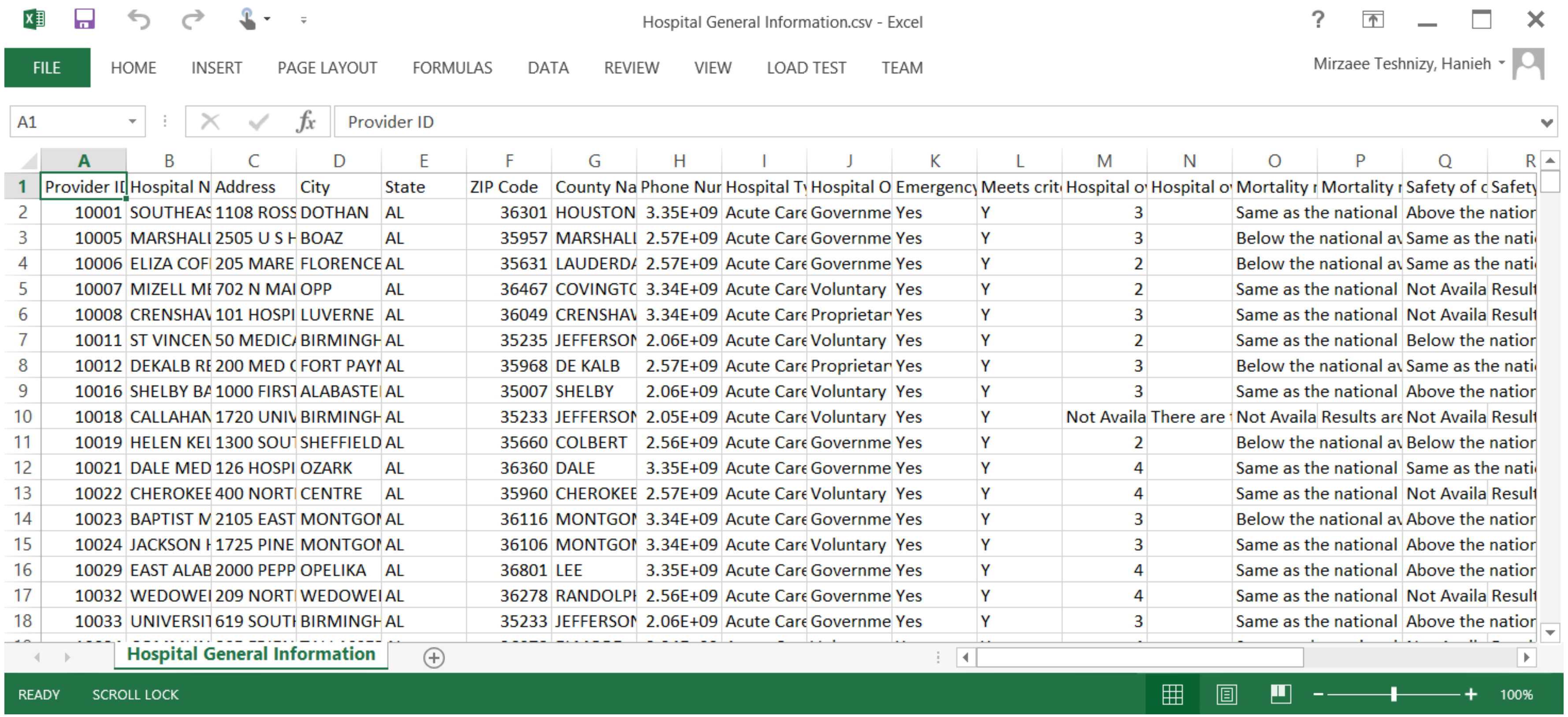
Project Data

- Hospital Compare Datasets
  - Flat files downloadable in .zip format



Project Data

- 67 .csv files with around 100 hospital characteristics (measures)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Provider ID	Hospital N	Address	City	State	ZIP Code	County Na	Phone Nur	Hospital Ty	Hospital O	Emergency	Meets crit	Hospital o	Hospital o	Mortality r	Mortality r	Safety of c	Safety
2	10001	SOUTHEAS	1108 ROSS	DOTHAN	AL	36301	HOUSTON	3.35E+09	Acute Care	Governme	Yes	Y	3		Same as the national		Above the nation	
3	10005	MARSHALL	2505 U S H	BOAZ	AL	35957	MARSHALL	2.57E+09	Acute Care	Governme	Yes	Y	3		Below the national av		Same as the nati	
4	10006	ELIZA COF	205 MARE	FLORENCE	AL	35631	LAUDERDA	2.57E+09	Acute Care	Governme	Yes	Y	2		Below the national av		Same as the nati	
5	10007	MIZELL ME	702 N MAI	OPP	AL	36467	COVINGTO	3.34E+09	Acute Care	Voluntary	Yes	Y	2		Same as the national		Not Availa	Result
6	10008	CRENSHAW	101 HOSPI	LUVERNE	AL	36049	CRENSHAW	3.34E+09	Acute Care	Proprietar	Yes	Y	3		Same as the national		Not Availa	Result
7	10011	ST VINCEN	50 MEDICA	BIRMINGH	AL	35235	JEFFERSON	2.06E+09	Acute Care	Voluntary	Yes	Y	2		Same as the national		Below the nation	
8	10012	DEKALB RE	200 MED C	FORT PAYN	AL	35968	DE KALB	2.57E+09	Acute Care	Proprietar	Yes	Y	3		Below the national av		Same as the nati	
9	10016	SHELBY BA	1000 FIRST	ALABASTE	AL	35007	SHELBY	2.06E+09	Acute Care	Voluntary	Yes	Y	3		Same as the national		Above the nation	
10	10018	CALLAHAN	1720 UNIV	BIRMINGH	AL	35233	JEFFERSON	2.05E+09	Acute Care	Voluntary	Yes	Y	Not Availa	There are t	Not Availa	Results are	Not Availa	Result
11	10019	HELEN KEL	1300 SOUT	SHEFFIELD	AL	35660	COLBERT	2.56E+09	Acute Care	Governme	Yes	Y	2		Below the national av		Below the nation	
12	10021	DALE MED	126 HOSPI	OZARK	AL	36360	DALE	3.35E+09	Acute Care	Governme	Yes	Y	4		Same as the national		Same as the nati	
13	10022	CHEROKEE	400 NORT	CENTRE	AL	35960	CHEROKEE	2.57E+09	Acute Care	Voluntary	Yes	Y	4		Same as the national		Not Availa	Result
14	10023	BAPTIST M	2105 EAST	MONTGOM	AL	36116	MONTGOM	3.34E+09	Acute Care	Governme	Yes	Y	3		Below the national av		Above the nation	
15	10024	JACKSON H	1725 PINE	MONTGOM	AL	36106	MONTGOM	3.34E+09	Acute Care	Voluntary	Yes	Y	3		Same as the national		Above the nation	
16	10029	EAST ALAB	2000 PEPP	OPELIKA	AL	36801	LEE	3.35E+09	Acute Care	Governme	Yes	Y	4		Same as the national		Above the nation	
17	10032	WEDOWEE	209 NORT	WEDOWEE	AL	36278	RANDOLPH	2.56E+09	Acute Care	Governme	Yes	Y	4		Same as the national		Not Availa	Result
18	10033	UNIVERSIT	619 SOUTH	BIRMINGH	AL	35233	JEFFERSON	2.06E+09	Acute Care	Governme	Yes	Y	3		Same as the national		Above the nation	

Data Curation and Wrangling

- Only hospital-level flat files were selected for further analysis:
  1. Hospital General Informations.csv
  2. HCAHPS – Hospital.csv → contains the patient survey information
  3. Complications and Deaths- Hospital.csv
  4. Healthcare associated infections- Hospital.csv
  5. Medicare Hospital Spending per Patient.csv
  6. Outpatient Imaging Efficiency- Hospital.csv
  7. Structural Measures- Hospital.csv
  8. Timely and Effective Care- Hospitals.csv

Data Curation and Wrangling

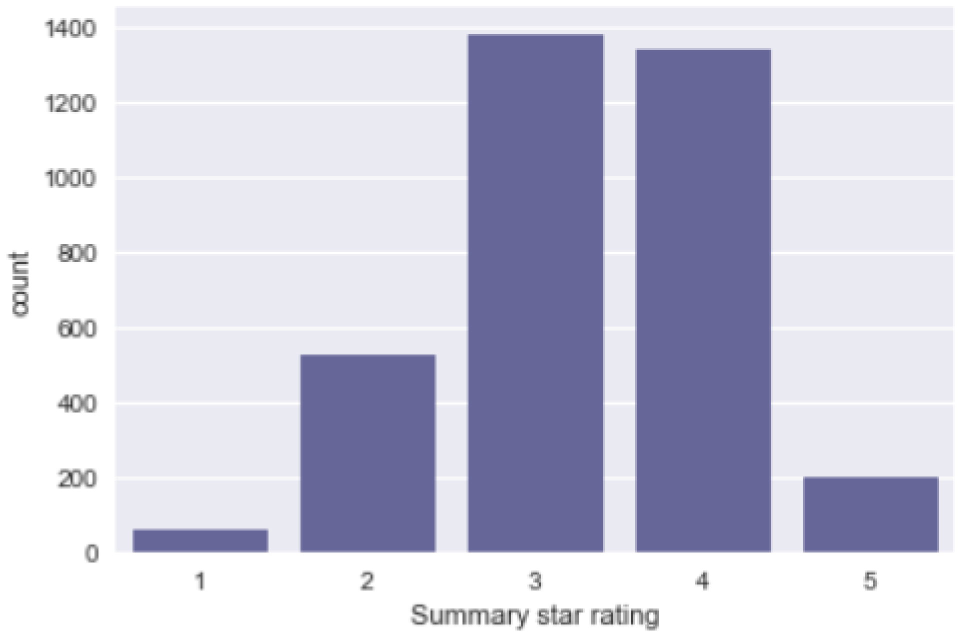
- Star rating distribution:
  - 1322 hospitals were removed since no rating was available
  - All hospital level measures were extracted and combined into one big flat file
  - Each row contains per hospital measures

3	1380
4	1335
Not Available	1322
2	525
5	195
1	55

Exploratory Data Analysis (EDA)

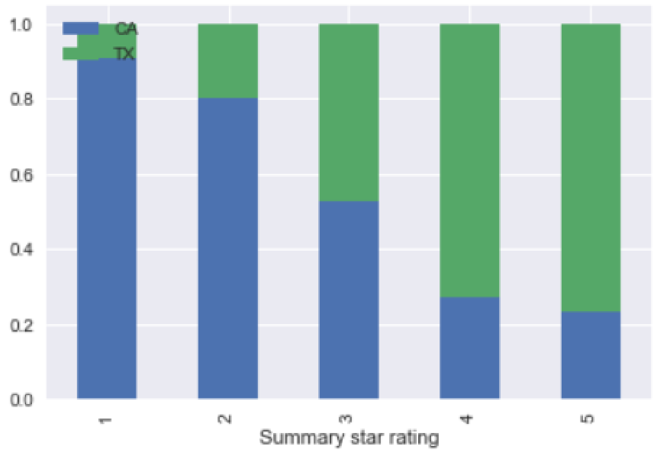
- Total of 3490 hospitals
  - Most hospitals have ratings 3 and 4
  - Average rating: 3.3

```
Number of Hospitals: 3490
3    0.395415
4    0.382521
2    0.150430
5    0.055874
1    0.015759
```

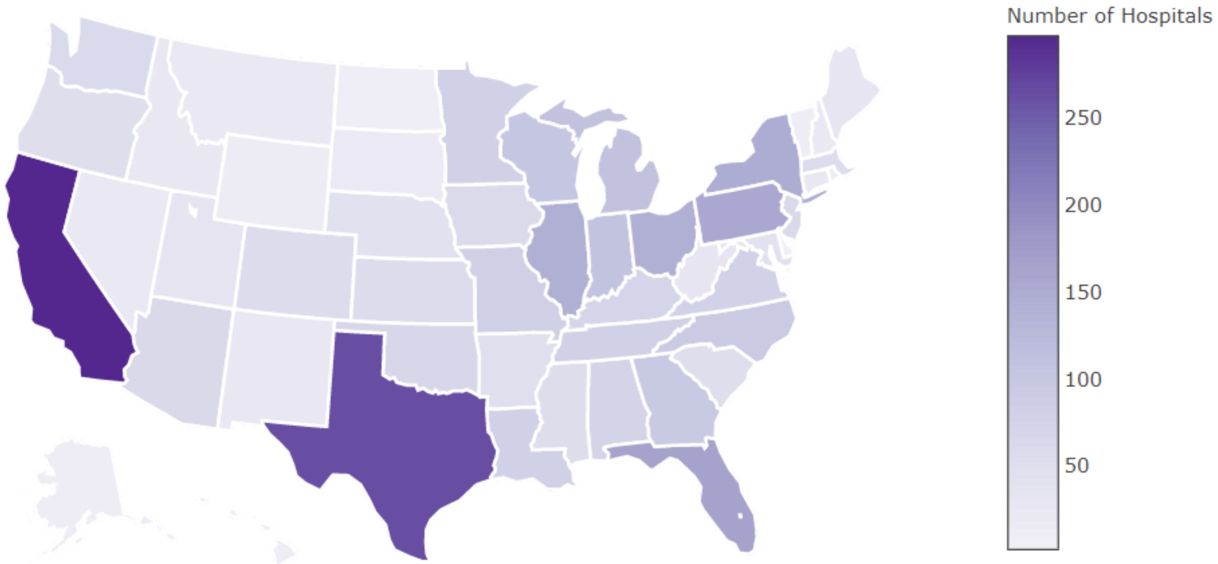


EDA:
How are hospitals distributed in the US?

- Texas (TX) and California (CA) with Most Hospitals
  - TX has more higher rated (4&5) hospitals than CA



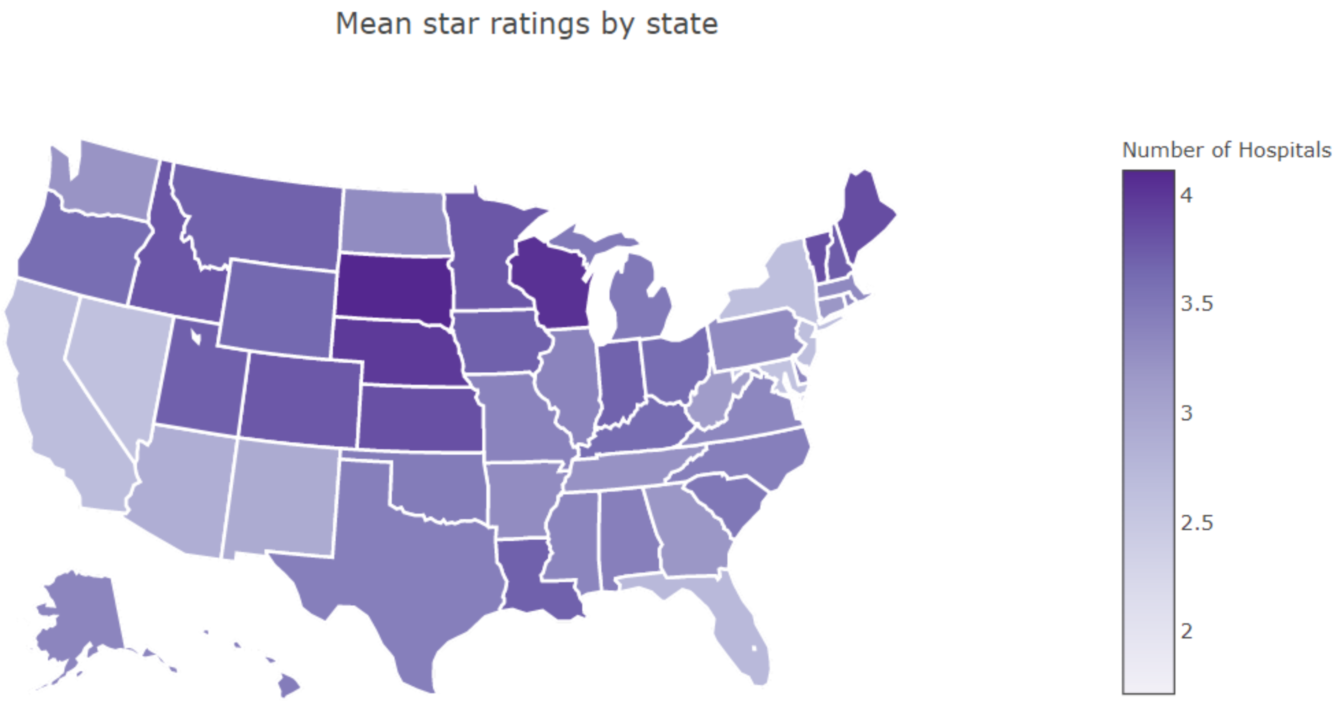
Medicare hospitals distribution



EDA:
What is the average rating per State?

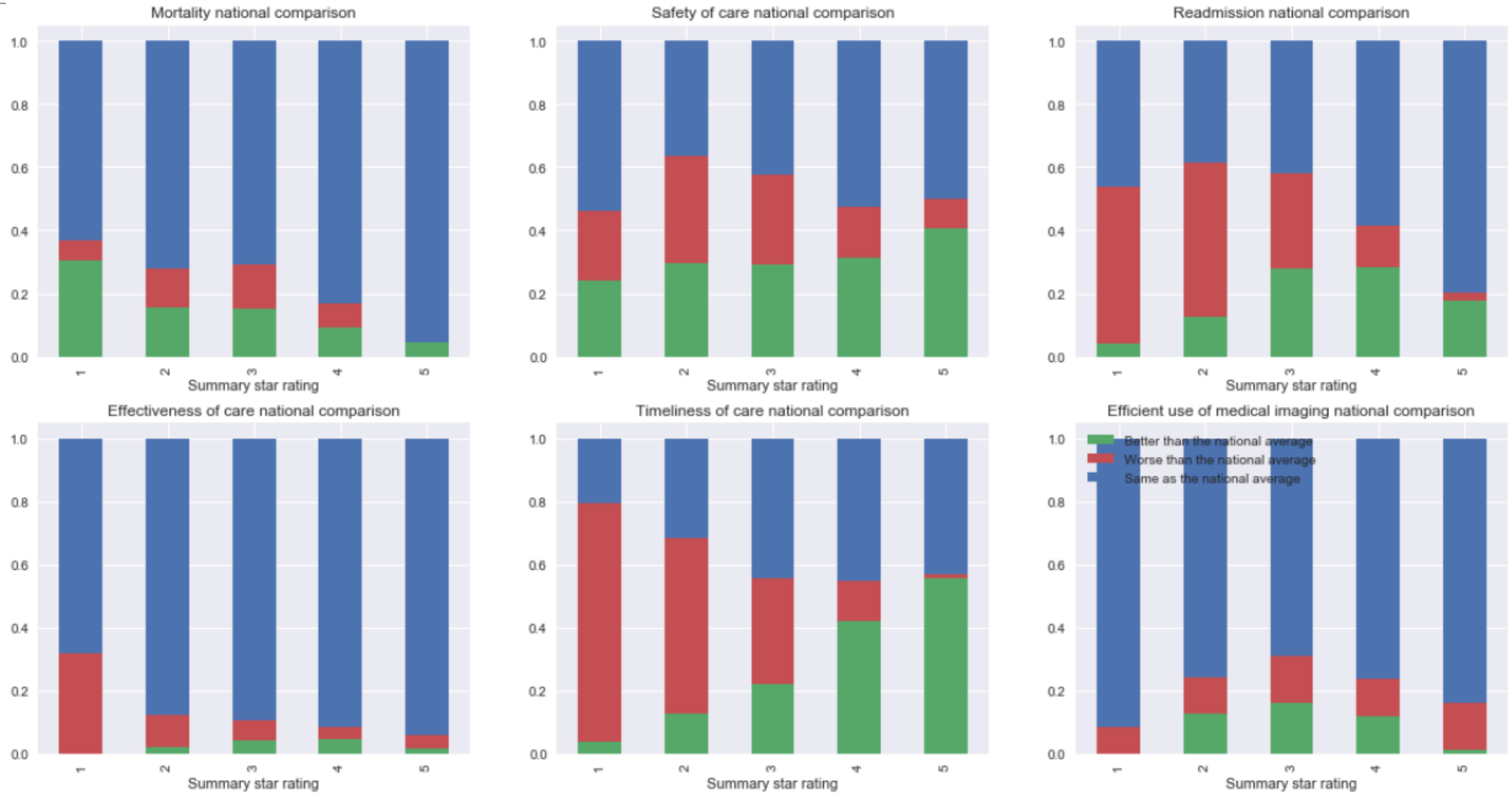
❖ Middle states towards north have generally a higher average rating

- Could this be related to population?
- Less populated states have better hospital ratings?



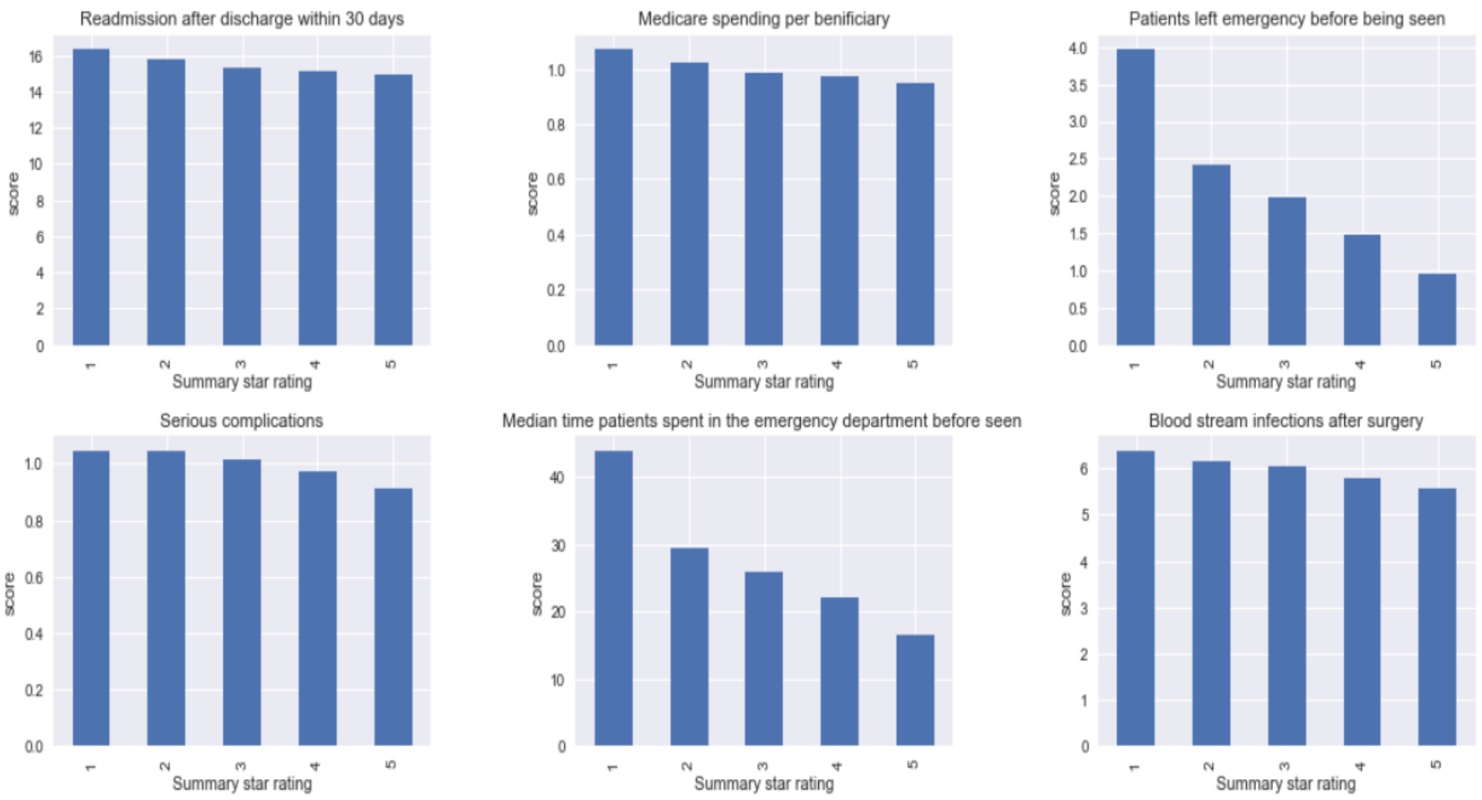
EDA:
How is the national level measures related to star ratings?

- Generally better ranked hospitals performed better in ratings
- We observe a noticeable trend in **Timeliness of care**



EDA:
Some Interesting Measures

- Score should be interpreted according to the context
- When considering **time-related measures**, a lower score is better
- Again we observe higher rated hospitals are performing better



Predictive Modeling: Can we predict the ratings based on hospital measures?

- 3490 hospitals and 121 measures (after accounting for categorical measures)
- Test data constitutes 20% of the original data set (698 hospitals)
  - Ordinal and Logistic regression yield the best performance metrics

Model	Accuracy on Test data	Ave. Precision	Ave. Recall	Average f1-score	Model performance via nested CV
Ridge Classifier	54%	0.55	0.54	0.53	0.55 ± 0.017
SVC	55%	0.55	0.55	0.53	0.56 ± 0.022
Random Forest	55%	0.56	0.55	0.52	0.56 ± 0.018
SVM	58%	0.59	0.58	0.56	0.57 ± 0.02
KNN	53%	0.53	0.53	0.50	0.33 ± 0.023
Gradient Boosting	55%	0.57	0.55	0.53	0.56 ± 0.011
xgboost	57%	0.58	0.57	0.56	0.57 ± 0.016
Ordinal Regression	58%	0.61	0.58	0.57	0.60 ± 0..025
Logistic Regression	58%	0.58	0.58	0.57	0.60 ± 0.022

Predictive Modeling: How would a dummy classifier perform?

- Accuracy on test data: 32%

accuracy: 0.3194842406876791

	precision	recall	f1-score	support
1	0.00	0.00	0.00	11
2	0.13	0.17	0.15	105
3	0.41	0.36	0.38	276
4	0.40	0.40	0.40	267
5	0.00	0.00	0.00	39
avg / total	0.33	0.32	0.33	698

Predictive Modeling: Can we improve the class imbalance?

- As we suffer from huge class imbalance, how about combining the ratings?
  - Still class imbalance, but less dramatic

Summary star rating	Number of hospitals
1	55
2	525
3	1380
4	1335
5	195

New star rating	Number of hospitals
1	580
2	1380
3	1535

- We consider the new ratings, as a notion of average (rate 2), above average (3), and below average (1) classification

Predictive Modeling: Performance using the combined ratings

- Similar and improved performance among the different algorithms
  - Logistic regression classifier still performs slightly better

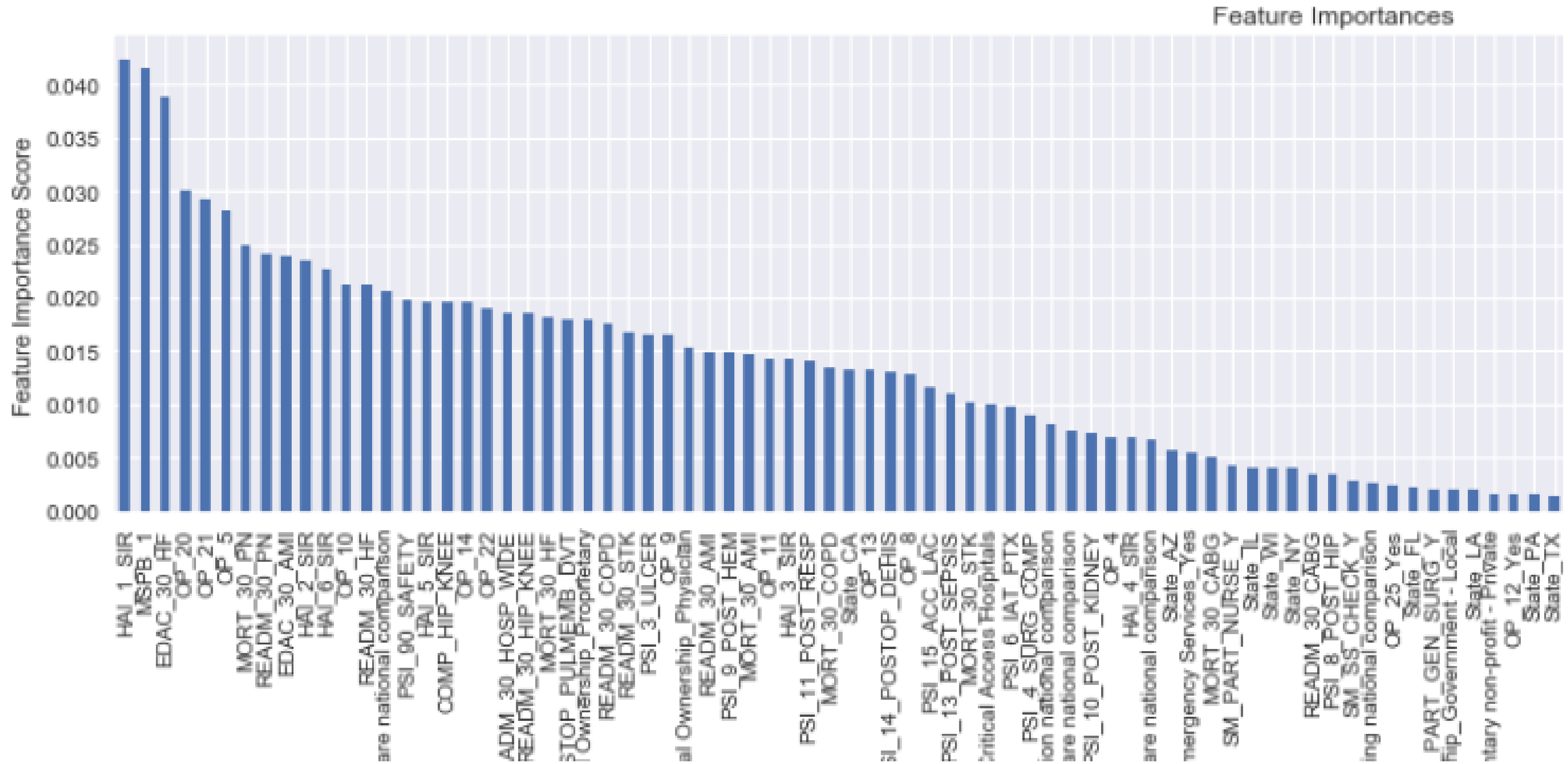
Model	Accuracy on Test data	Model performance via nested CV	Average f1-score on the test data
SVM	61%	0.63 ± 0.020	0.61
Gradient Boosting	65%	0.62 ± 0.009	0.65
xgboost	62%	0.62 ± 0.009	0.61
Ordinal Regression	62%	0.62 ± 0.011	0.61
Logistic Regression	62%	$\boxed{0.63 \pm 0.006}$	0.62

Predictive Modeling:
What if we make up for class imbalance synthetically?

- SMOTE algorithm was used to upsample the underdamped classes
  - Average accuracy on 3 different test datasets is reported
  - Gradient Boosting performs slightly better

Model	seed=12	seed=20	seed=42	Average
SVM	0.61	0.62	0.63	0.62
Gradient Boosting	0.64	0.62	0.65	0.64
Logistic Regression	0.61	0.62	0.62	0.62

Predictive Modeling:
Are there particular hospital measures that are driving these results?



Predictive Modeling: Are there particular hospital measures that are driving these results?

- Top 10 measures with relative higher importance:

1. HAI_1_SIR: Central line-associated blood stream infections in ICUs and select wards
2. MSPB_1: Medicare spending per beneficiary
3. EDAC_30_HF: Hospital return days for heart failure patients
4. OP_20: Average time patients spent in the emergency department before they were seen by a healthcare professional
5. OP_21: Average (median) time patients who came to the emergency department with broken bones had to wait before getting pain medication
6. OP_5: Average (median) number of minutes before outpatients with chest pain or possible heart attack got an ECG
7. MORT_30_PN: Death rate for pneumonia patients
8. READM_30_PN: Rate of readmission for pneumonia patients
9. EDAC_30_AMI: Hospital return days for heart attack patients
10. HAI_2_SIR: Catheter-associated urinary tract infections in ICUs and select wards

- **Infections**, **Medicare spending**, **Readmission**, and **Waiting times** are noticeable

Predictive Modeling: How can we improve the modeling performance?

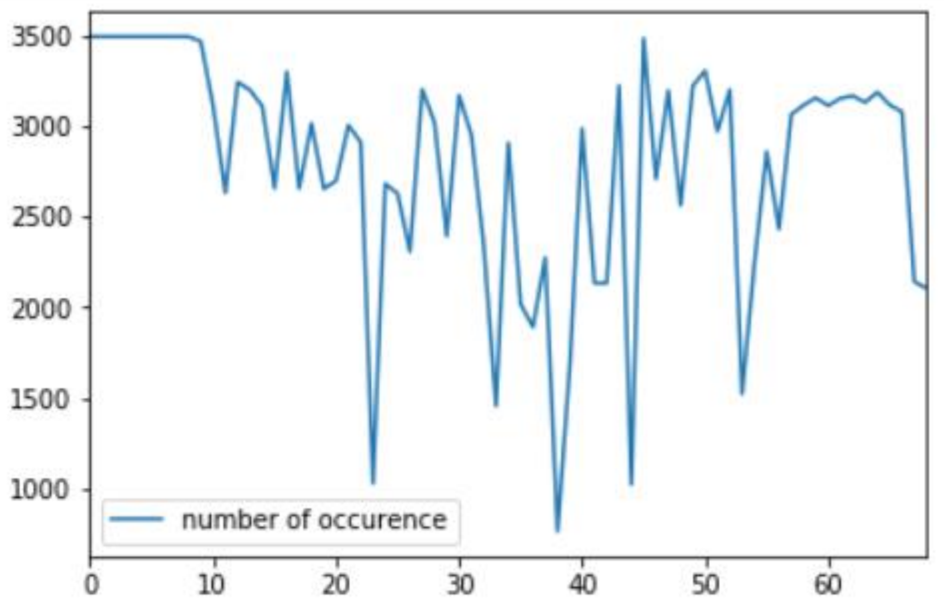
- More data could help
  - The cross-validation score has not converged yet

➢ Learning curve computed based on the training data and the logistic regression classifier

Predictive Modeling:
How can we improve the modeling performance?

- Collect data for the missing values, specially for most important features
  - The imputation strategy might not be accurate enough

- Number of occurrence (not missing)
  - There are many measures with more than 500 missing values



Predictive Modeling:
How can we improve the modeling performance?

- Use additional data sources
  - We observed that time-related measures are among the drivers of the modeling results
  - Timing can be related to denser hospitals
  - Cold we use state/city population data as another measure?

Conclusion and Remarks

- On average our developed model can predict the rating of a new hospital with 60% accuracy
- Improvement on common sense measures such as lowering **Infections** and **waiting times** can potentially improve patient experience and consequently the ratings
- More data would help developing a better model
  - Considering other relevant measures such as population data
  - Better estimation of missing values
  - More hospitals with ratings
- Clustering algorithms are worth trying here to investigate similarity patterns between equally-rated hospitals