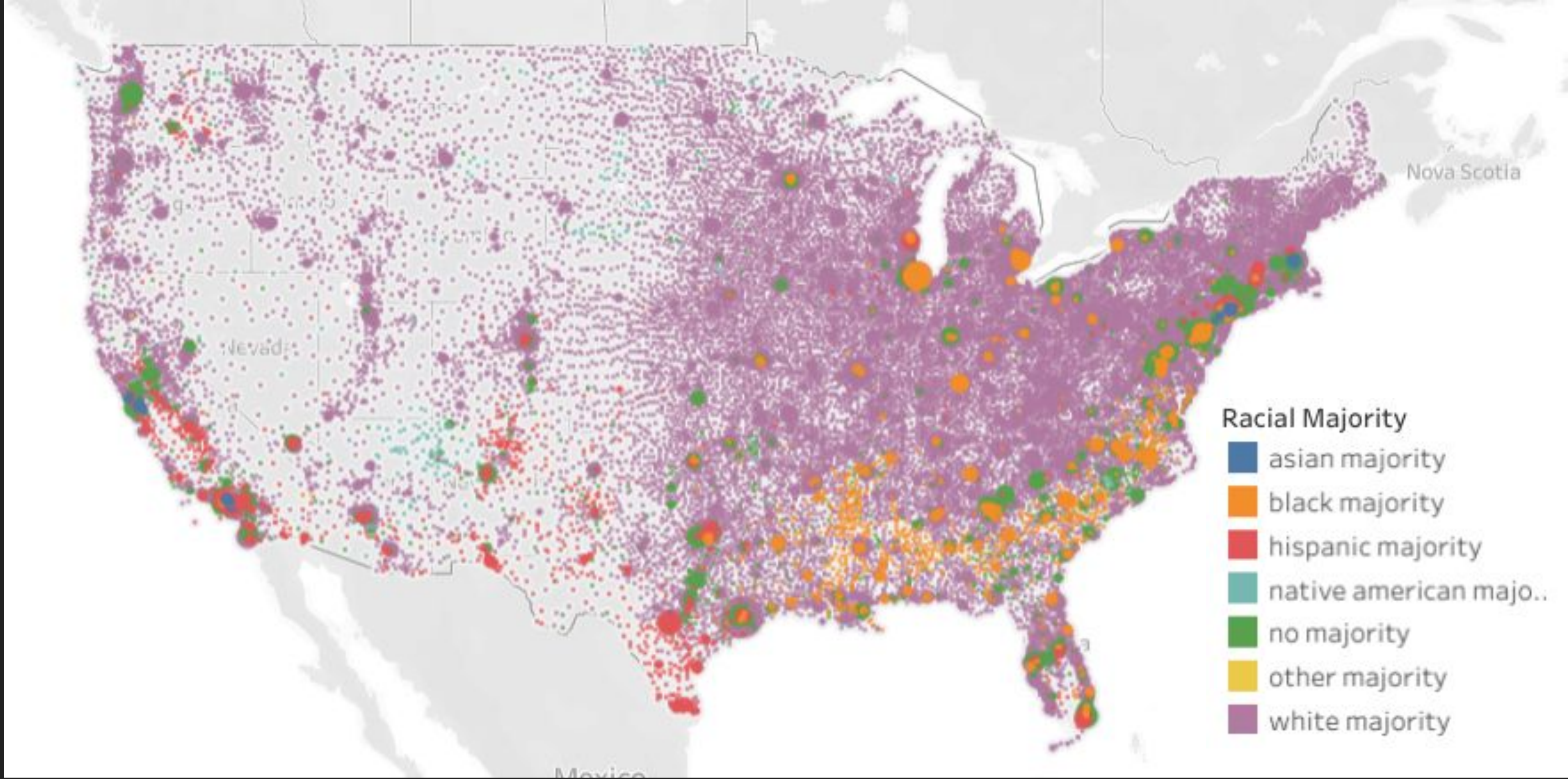
Disparity in Medicare

Investigating unintended bias in Medicare coverage

Alexander Vizzi, Jan 2021

Springboard cohort April 2020



Does Medicare favor patients of providers in white zip codes with higher coverage rates?

How to measure coverage?

Using two features:

‘Total allowed payments’ :

a calculated value for services; this is the total Medicare allows the provider to charge

‘Total Medicare payments’:

amount Medicare pays after patient copay, deductible, and/or third party responsibility

I create a new metric:

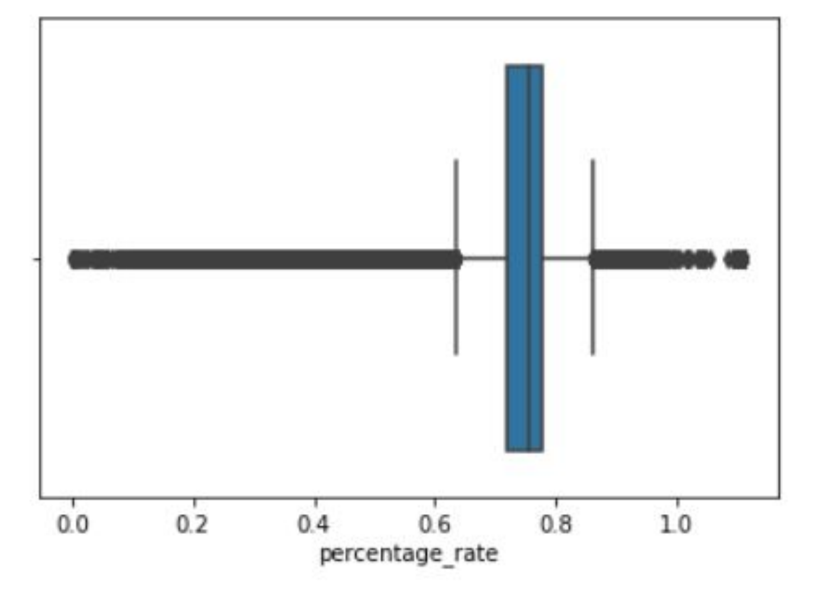
‘coverage_rate’:

the fraction of the actual allowed charges that Medicare pays
(Total Medicare payments / Total allowed payments)

Approach

Pay classes based on distribution plot of coverage rates.

- Very low
- Low
- Medium
- High
- Very high



Using classification models, I will build a best predictor for the payrate class. Then I will examine the feature importances and examine the extreme errors made by the model.

Project limitations

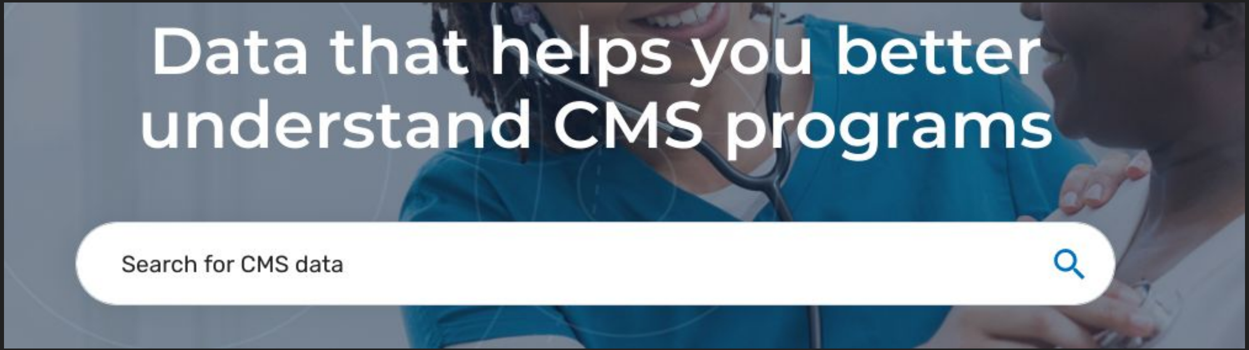
- Patient race data is sparse for each provider; zip-code-level populations were used instead

- “Coverage_rate” has some confounding factors:
  - Adjustments to payments for service intensity and other details
  - Different coverage for different services
  - Patient differences that cause high deductibles or copays

Data Acquisition

- Racial disparities in health outcomes during Covid pandemic highlight research need.
- Few nationwide public sources of health data
- Medicare, a national health program, releases data on its providers

data.cms.gov



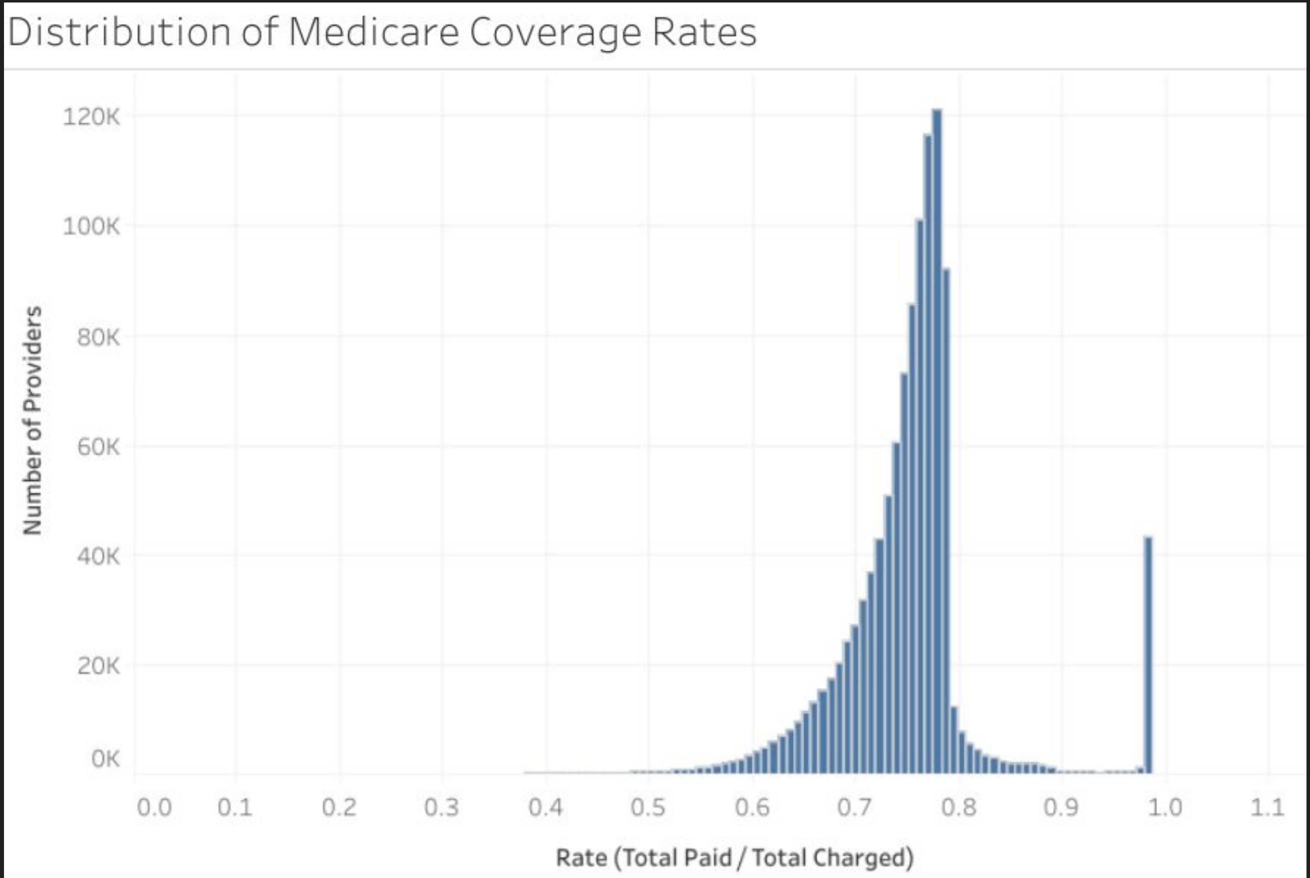
Data Wrangling

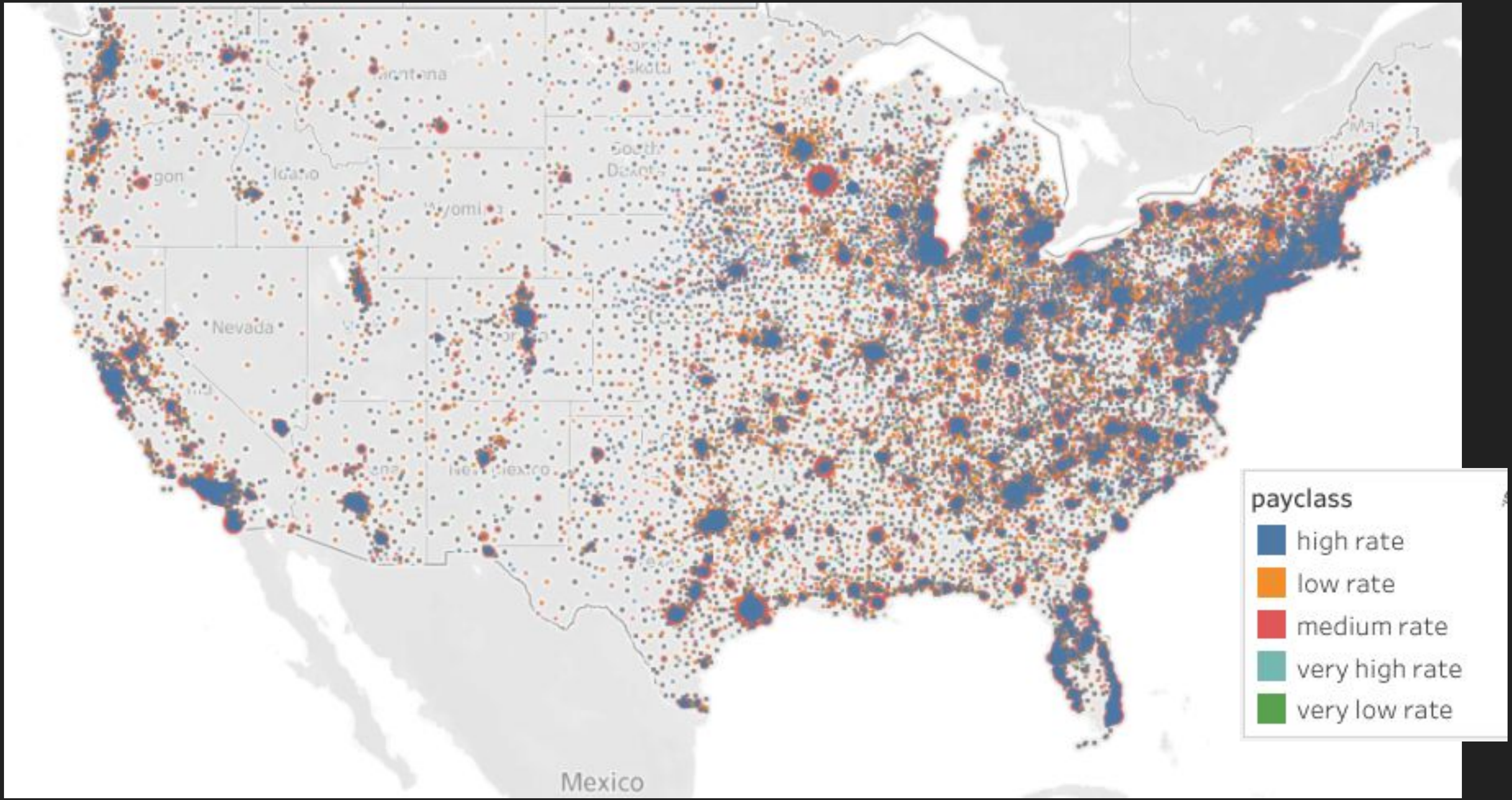
- Lots of nulls - race, age bins, % with depression et al.
- Tried to drop rows with missing values: high info loss
- Dropped columns with missing values instead.
- Merged Census data to supplement missing race data.

Exploratory Data Analysis

Exploring Medicare payrates

- Left skewed
- Lots of high outliers

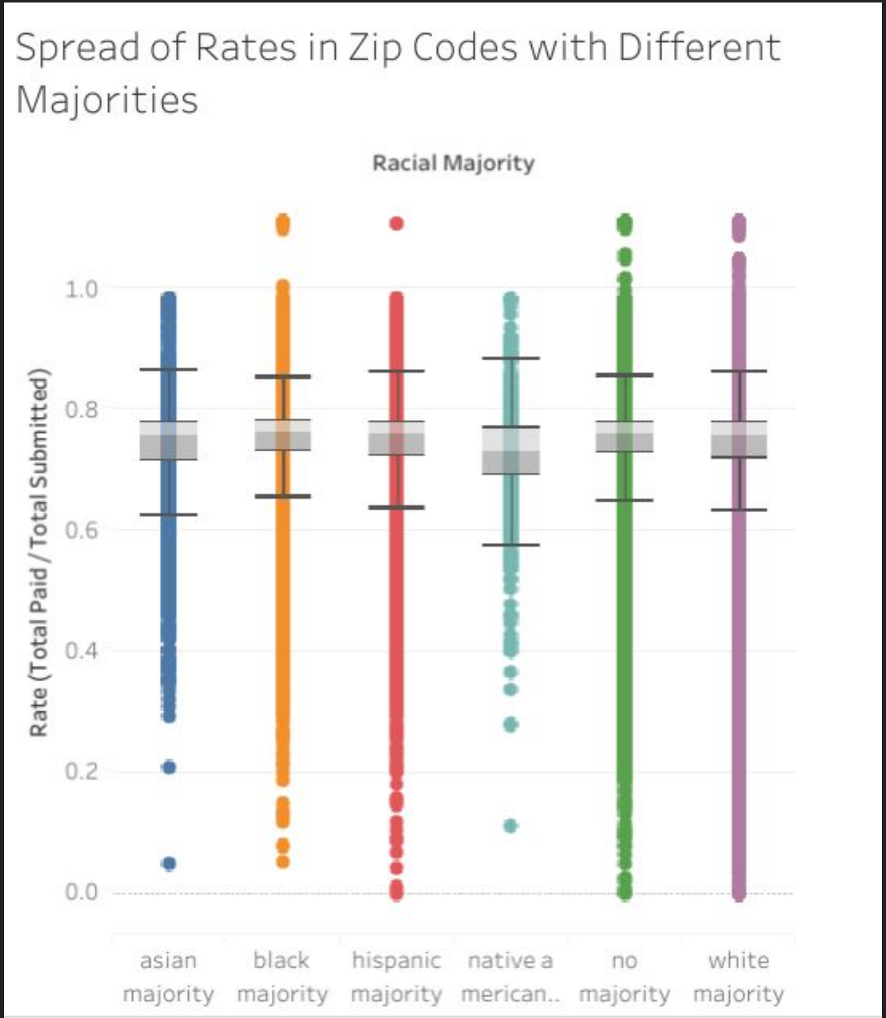




Pay classes somewhat evenly distributed nationally.

Most zip code payrates have similar interquartile ranges.

Some zip codes have much larger spreads and more low outliers.



Other measures besides pay rate could reveal differences.

Aggregate Measures by Zip

lower — higher

Racial Majority	Mean Allowed/person	Mean Payment per Person	Mean Allowed/Service	Mean Payment/service	Mean Services /person	Mean Rate	Median Rate
no majority	360.262	279.361	39.994	31.013	9.008	0.775	0.759
hispanic majority	384.953	298.164	45.097	34.930	8.536	0.775	0.759
black majority	315.762	242.629	46.983	36.101	6.721	0.768	0.761
white majority	338.601	258.737	42.827	32.726	7.906	0.764	0.755
asian majority	464.902	353.090	64.622	49.080	7.194	0.759	0.755
native american majority	269.218	201.544	40.296	30.167	6.681	0.749	0.731
other majority							

Doctors in Asian zip codes have higher charges and higher Medicare payments but the rates are similar to other majorities

Native American zip codes are on the lower end of all measures

Feature Selection

- Chi squared test
- Ridge and Lasso regression using continuous coverage_rate

Cross Validation

- Ran CV scores on basic classification models, using different scale methods and features. Used 100,000 samples due to memory limits
- Attempted but did not complete SVC due to memory limitations

models	CV: Scale1, all feats	CV: Scale1, kbest	CV: Scale1, ridge feats	CV: Scale2, all feats	CV: Scale2, kbest feats	CV: Scale2, ridge feats
Random Forest	0.590946	0.530178	0.474977	0.591320	0.521617	0.474534
Gradient Boosting	0.606442	0.529329	0.466971	0.606779	0.521617	0.459689
Logistic Regression	0.596830	0.528741	0.470371	0.596491	0.521617	0.459697

Extended Modeling

Using F1 Score as metric, models performed similarly in cross-val.

- GBC was best at 0.606 but took five times as long
- RF and LR were close second and third, 0.591 and 0.596
- Feature selection greatly reduced performances

Random Forest and Logistic Regression were used for further tuning

- Trained on ALL data (1,000,000 samples); scores increased by 0.05
- Used ALL features, since reduced features hurt CV scores
- Used GridSearch CV and RandomizedSearch CV on best models (RF)

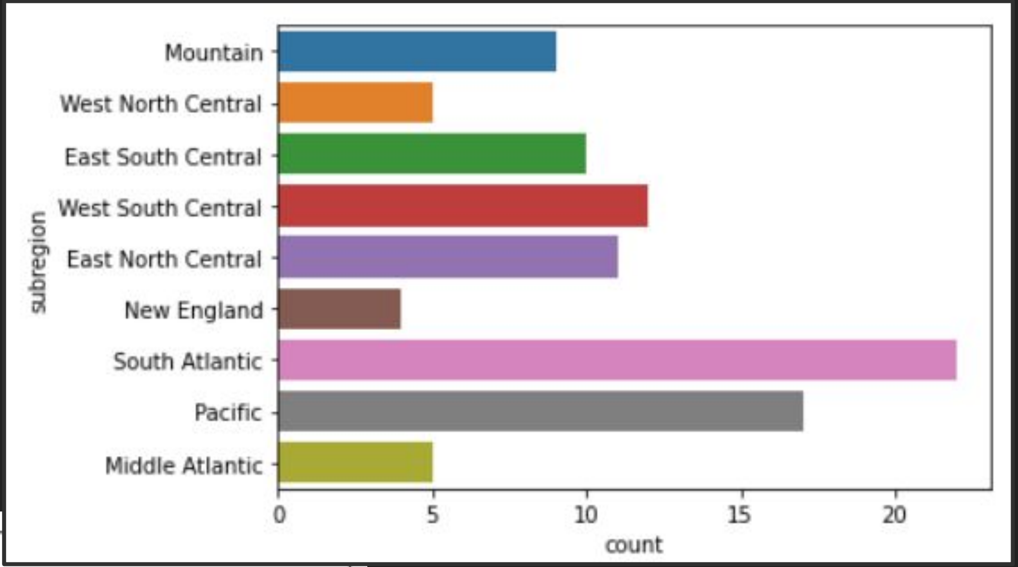
models	F1 Scores, out-of-box	F1 Scores, Tune1	F1 Scores, Tune2	F1 Scores, Tune3	F1 Scores, Tune4	F1 Scores, Tune5
Random Forest	0.656930	0.658595	0.656472	0.655376	0.657754	0.655587
Logistic Regression	0.622273	--	--	--	--	--

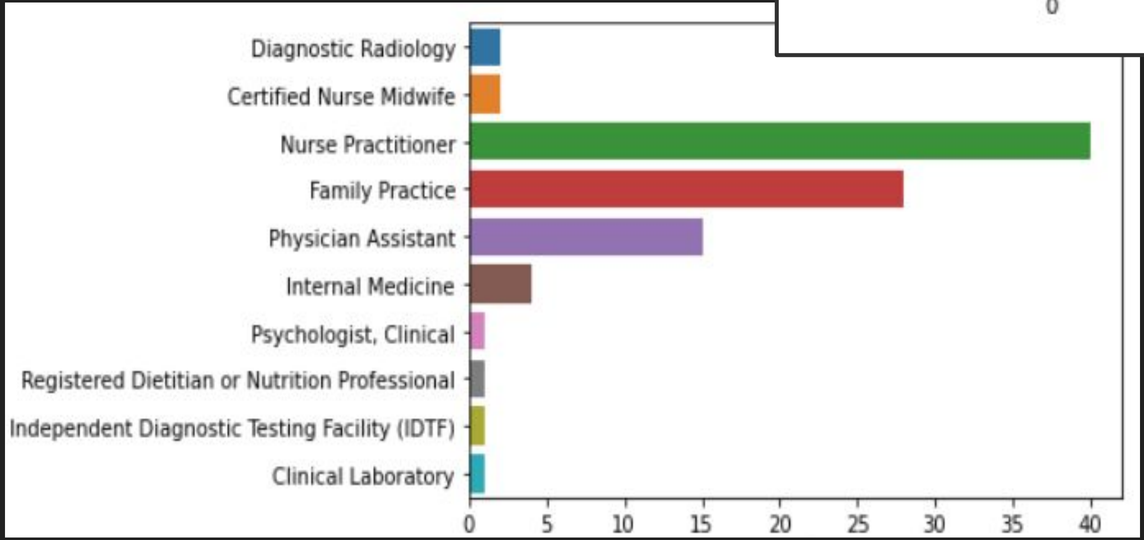
Does the model show bias?

- Generally, not very much. The top features relate to pay amount. The racial majority of each city doesn't appear.
- Some patterning exists: percent_race appears in top 20 features. The ordering of % race pop may be telling (especially '% Asian'). A regression model could investigate this.

	Features	Importances
6	submitted_charges_per_person	0.090531
5	avg_hcc_risk_score	0.073834
8	submitted_charges_per_service	0.061953
3	total_submitted_charges	0.061168
7	services_per_person	0.055767
1	number_of_services	0.049064
2	total_beneficiaries	0.044112
0	number_of_hcpcs	0.043498
15	percent_asian	0.033756
16	percent_two_or_more_races	0.033311
9	total_pop	0.033273
4	avg_beneficiary_age	0.033096
13	percent_hispanic	0.033004
10	percent_black	0.032702
11	percent_white	0.032151
12	percent_native_amer	0.030337
14	percent_other	0.027938
17	entity_code_I	0.019515
18	entity_code_O	0.016897
118	provider_type_Mass Immunizer Roster Biller	0.011026

Whose coverage was predicted “very high” but actually “very low”?





-mostly primary care

-in the South and West

-mostly white or mixed majorities

Further modeling

Searching for undercompensated patients/services

-training a stacked model on the mislabeled “very low” samples to search for unfair deficits in coverage

Looking for overcompensated patients/services

-reducing unnecessary coverage expenses
-training a stacked model on the mislabeled “very low” samples

Using another (continuous) target

-linear regression to predict % Black, white, etc
-linear regression models to predict % coverage
-examining feature importances that may correlate, i.e. % Black/white/etc.