



# HEALTHCARE DISPARITIES

## PRIMARY CARE ACCESS ANALYSIS

Jacqueline Clinesmith, Tony Ferri, and Jenny Johnson

MSU Data Analytics Boot Camp, 2022



# Healthcare Disparities

## Primary Care Access Analysis



WHAT ARE WE  
TRYING TO DO?



DATA EXPLORATION

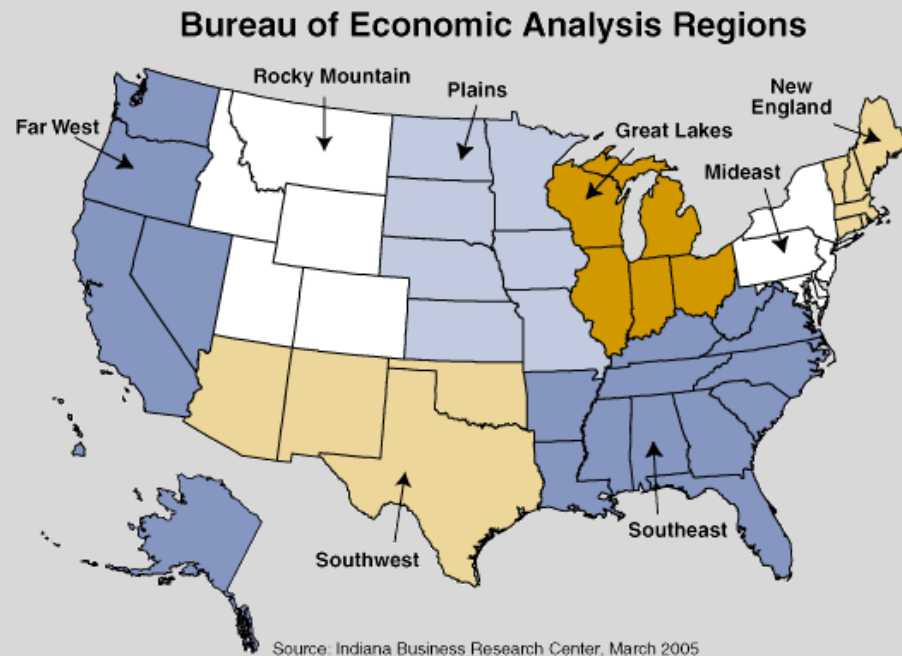


ANALYSIS RESULTS



WHAT ARE WE  
TRYING TO DO?

# Build a machine learning model that can determine primary care provider availability

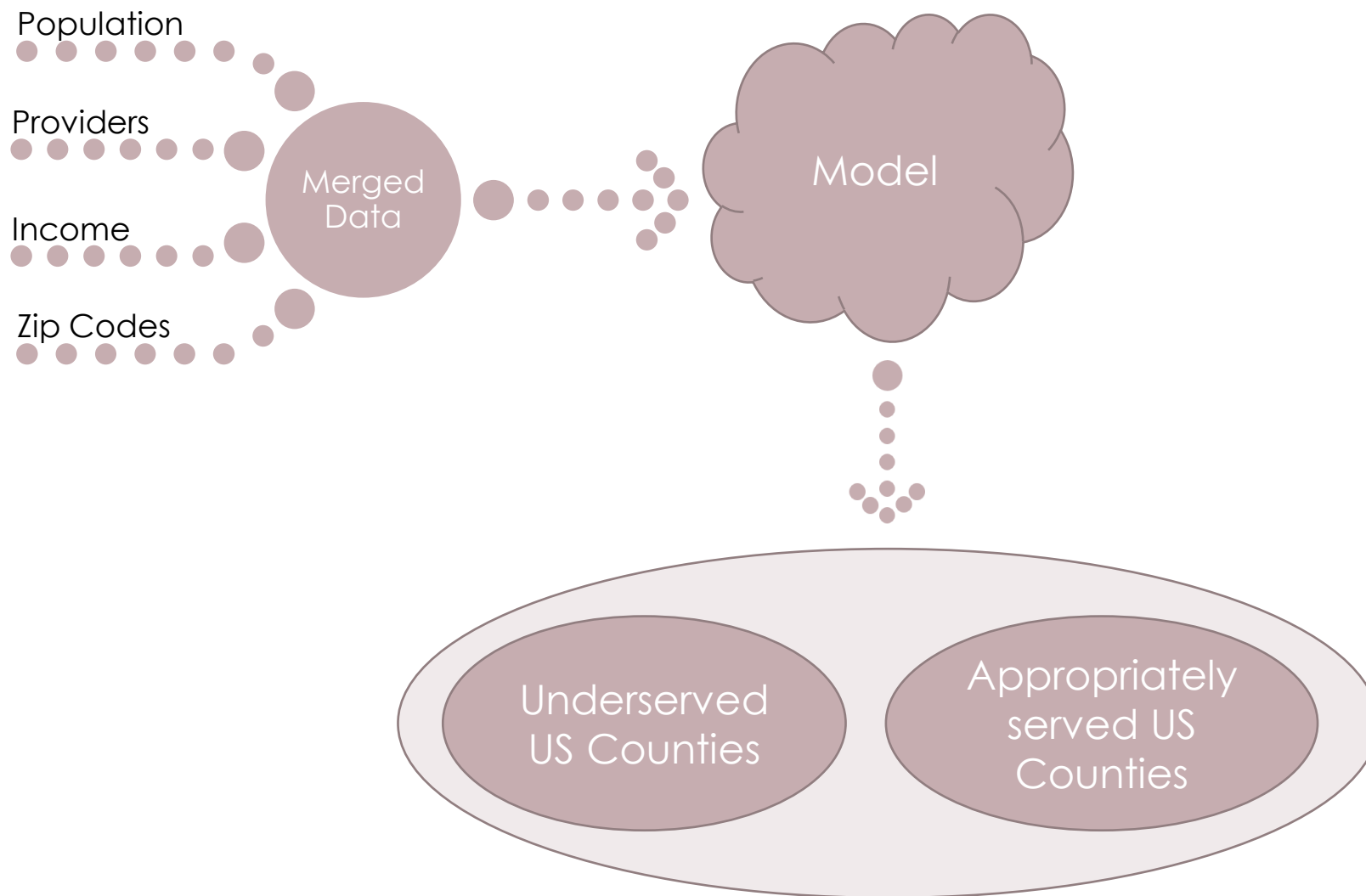


- We're hoping to determine:
  - Are there counties in the United States that will be underserved by primary care providers?
  - Are there any other factors that impact availability, like income, region, or population?

Meaningful access to a primary care providers can help reduce health care disparities.



# DATA EXPLORATION



Population provided by US Census Bureau  
Providers provided by The Centers for Medicare and Medicaid Services  
Income provided by US Bureau of Economic Analysis  
Zip Codes provided by simplemaps

# Data Exploration

- Data from official government sources was given preference due to accuracy and reliability
- Data preprocessing included:
  - Removing irrelevant information
  - Standardizing zip codes
  - Removing headers and footers
  - Standardizing column names
  - Transformed into bins
  - Calculating primary care providers by county
  - Determine csv encoding1

```

Physicians
---
npi varchar(10)
grd_yr varchar(6)
pri_spec varchar(50)
cty varchar(50)
st varchar(2) FK >- Zip.st
zip varchar(5) FK >- Zip.zip
ind_assgn varchar(1)
grp_assgn varchar(1)

```

```

Income
---
GeoFips varchar(5) FK >- Population.GeoFips
per_capita_income numeric

```

```

Population
---
GeoFips varchar(5)
population numeric
pop_density float(20)

```

```

Zip
---
zip varchar(5)
st varchar(2)
GeoFips varchar(5) FK >- Income.GeoFips
county_name varchar(50)

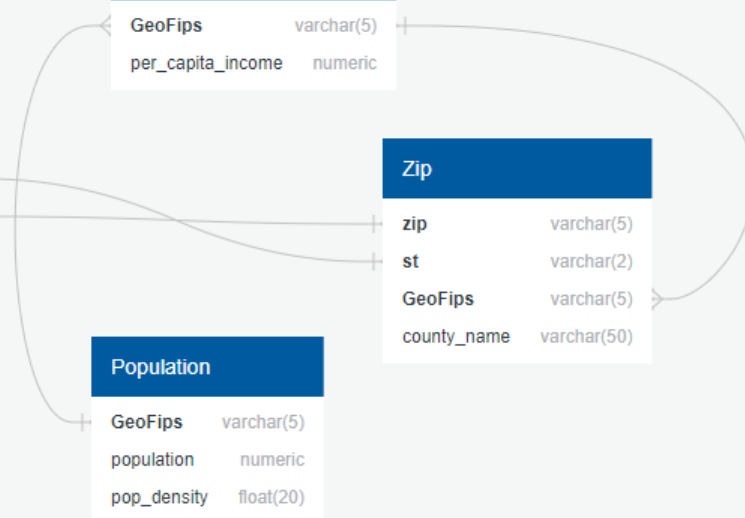
```

Physicians	
npi	varchar(10)
grd_yr	varchar(6)
pri_spec	varchar(50)
cty	varchar(50)
st	varchar(2)
zip	varchar(5)
ind_assgn	varchar(1)
grp_assgn	varchar(1)

Income	
GeoFips	varchar(5)
per_capita_income	numeric

Zip	
zip	varchar(5)
st	varchar(2)
GeoFips	varchar(5)
county_name	varchar(50)

Population	
GeoFips	varchar(5)
population	numeric
pop_density	float(20)



# Database

- Cleaned data sources were loaded into a PostgreSQL database
- Next steps:
  - Join all tables in Postgres
  - Connect merged table to analysis

Tables (4)

income

physicians

population

zip

1SELECT \* FROM income;

Data Output

Explain

Messages

geofipsinteger

per\_capita\_incomeinteger

1100146814

2100350953

3100537850

1SELECT \* FROM physicians;

Data Output

Explain

Messages

Notifications

npiinteger

grd\_yrdouble precision

pr\_speccharacter varying

112152839081998FAMILY MEDICINE

212152576052007GENERAL PRACTICE

312152482732010INTERNAL MEDICINE

1SELECT \* FROM population;

Data Output

Explain

Messages

Notifications

geofipsinteger

populationinteger

pop\_densitydouble precision

110015520035.8534189940189

2100320810750.5415035640023

310052578211.247981205619

1SELECT \* FROM zip;

Data Output

Explain

Messages

Notifications

zipinteger

stcharacter varying

geofipsinteger

county\_namecharacter varying

1601PR72001Adjuntas

2602PR72003Aguada

3603PR72005Aguadilla



# ANALYSIS RESULTS



```
# Calculated the balanced accuracy score
y_pred = easy.predict(X_test)
balanced_accuracy_score(y_test, y_pred)
```

```
0.8957196167984851
```

```
# Display the confusion matrix
y_pred = easy.predict(X_test)
matrix = confusion_matrix(y_test, y_pred)
print(matrix)
```

```
[[334  42]
 [ 37 345]]
```

```
# Print the imbalanced classification report
print(classification_report_imbalanced(y_test, y_pred))
```

	pre	rec	spe	f1	geo	iba	sup
0	0.90	0.89	0.90	0.89	0.90	0.80	376
1	0.89	0.90	0.89	0.90	0.90	0.80	382
avg / total	0.90	0.90	0.90	0.90	0.90	0.80	758

```
# find r-squared score
r2_score(y_test, y_pred)
```

```
0.5830873342987635
```

## Analyze Results

- Logistical Regression
- Accuracy score: 89.6%
- R-squared score: 58.3%

## Story Board Title

Map of US by county or region color coded by underserved and appropriately served

Information on benefits of Primary Care Providers

Counts of underserved counties by region and percentage of counties that is

Interesting graphic/image/etc

Weight of model factors affecting underserved designation

Authors and References



Questions



# APPENDIX

# Data Sources

Source	Information Used	Location
<b>Income data</b> from US Bureau of Economic Analysis	GeoFips and 2020 from bea_income_2020.csv	<a href="https://apps.bea.gov/iTable/iTable.cfm?reqid=70&amp;step=30&amp;isuri=1&amp;major_area=4&amp;area=xx&amp;year=2020&amp;tableid=20&amp;category=720&amp;area_type=4&amp;year_end=-1&amp;classification=non-industry&amp;state=xx&amp;statistic=3&amp;yearbegin=-1&amp;unit_of_measure=levels">https://apps.bea.gov/iTable/iTable.cfm?reqid=70&amp;step=30&amp;isuri=1&amp;major_area=4&amp;area=xx&amp;year=2020&amp;tableid=20&amp;category=720&amp;area_type=4&amp;year_end=-1&amp;classification=non-industry&amp;state=xx&amp;statistic=3&amp;yearbegin=-1&amp;unit_of_measure=levels</a>
<b>Region data</b> from US Bureau of Economic Analysis	Regions by State from website	<a href="https://www.bea.gov/news/2015/gross-domestic-product-state-advance-2014-and-revised-1997-2013/regional-maps">https://www.bea.gov/news/2015/gross-domestic-product-state-advance-2014-and-revised-1997-2013/regional-maps</a>
<b>Population data</b> from US Census Bureau	GEOID, B01001_001E, and B01001_calc_PopDensity from population_census.csv	<a href="https://covid19.census.gov/datasets/average-household-size-and-population-density-county/explore?location=15.251650%2C0.315550%2C3.67&amp;showTable=true">https://covid19.census.gov/datasets/average-household-size-and-population-density-county/explore?location=15.251650%2C0.315550%2C3.67&amp;showTable=true</a>
<b>Zip code data</b> from simplemaps Basic Download	zip, stae_id, county_fips, and county_name from uszips.csv	<a href="https://simplemaps.com/data/us-zips">https://simplemaps.com/data/us-zips</a>
<b>Physician data</b> from The Centers for Medicare and Medicaid Services	NPI, grd_yr, pri_spec, city, st, zip, ind_assgn, and grp_assgn from physician_data.csv	<a href="https://data.cms.gov/provider-data/dataset/mj5m-pzi6">https://data.cms.gov/provider-data/dataset/mj5m-pzi6</a>