



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

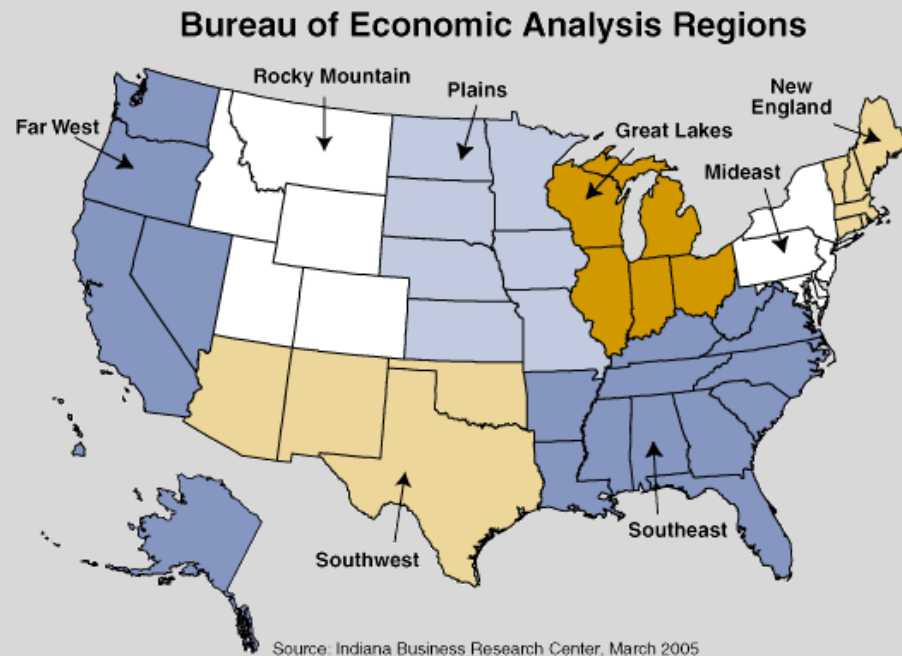


ANALYSIS RESULTS



WHAT ARE WE
TRYING TO DO?

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

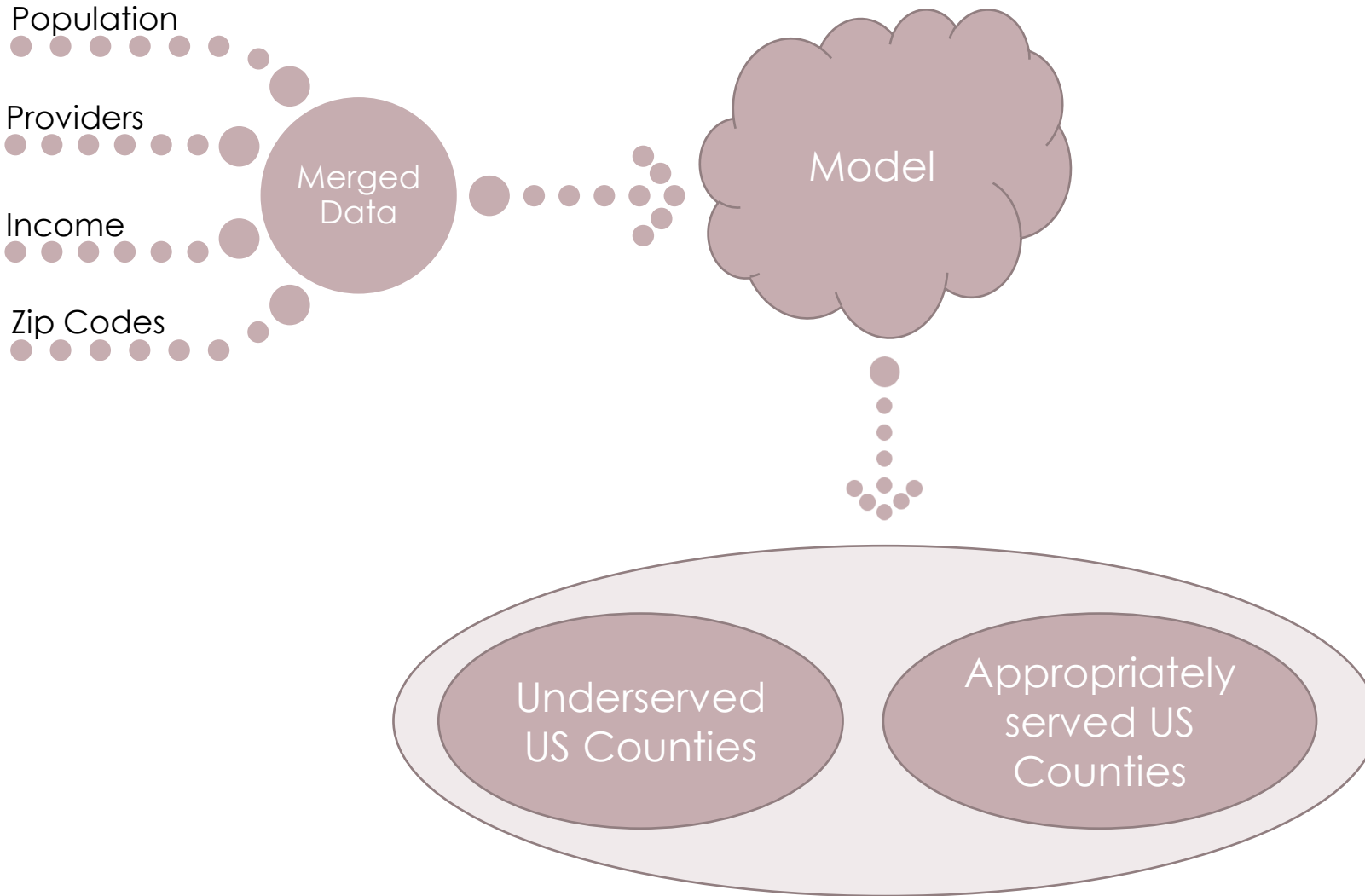


- We're hoping to determine:
 - Are there counties in the United States that will be underserved by primary care providers?
 - Are there any other factors that impact availability, like income, region, or population?

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

```

```

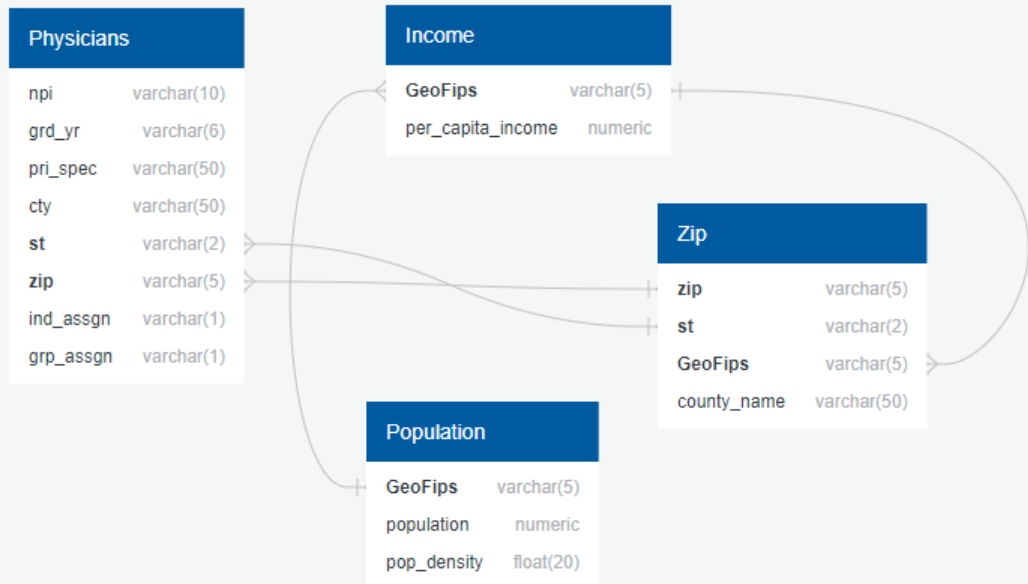
Population
---
GeoFips varchar(5)
population numeric
pop_density float(20)

```

```

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

```



Database

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

<div> <div>Tables (4)</div> <div> <div>income</div> <div>physicians</div> <div>population</div> <div>zip</div> </div> </div>	<div> <div>1</div> <div>SELECT * FROM income;</div> </div> <table> <tr> <th>Data Output</th><th>Explain</th><th>Messages</th></tr> <tr> <td> <div> <div></div> <div>geofips</div> <div>integer</div> <div></div> </div> </td><td> <div> <div></div> <div>per_capita_income</div> <div>integer</div> <div></div> </div> </td><td></td></tr> <tr> <td>1</td><td>1001</td><td>46814</td></tr> <tr> <td>2</td><td>1003</td><td>50953</td></tr> <tr> <td>3</td><td>1005</td><td>37850</td></tr> </table>	Data Output	Explain	Messages	<div> <div></div> <div>geofips</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>per_capita_income</div> <div>integer</div> <div></div> </div>		1	1001	46814	2	1003	50953	3	1005	37850	<div> <div>1</div> <div>SELECT * FROM physicians;</div> </div> <table> <tr> <th>Data Output</th><th>Explain</th><th>Messages</th><th>Notifications</th></tr> <tr> <td> <div> <div></div> <div>npi</div> <div>integer</div> <div></div> </div> </td><td> <div> <div></div> <div>grd_yr</div> <div>double precision</div> <div></div> </div> </td><td> <div> <div></div> <div>pri_spec</div> <div>character varying</div> <div></div> </div> </td><td></td></tr> <tr> <td>1</td><td>1215283908</td><td>1998</td><td>FAMILY MEDICINE</td></tr> <tr> <td>2</td><td>1215257605</td><td>2007</td><td>GENERAL PRACTICE</td></tr> <tr> <td>3</td><td>1215248273</td><td>2010</td><td>INTERNAL MEDICINE</td></tr> </table>	Data Output	Explain	Messages	Notifications	<div> <div></div> <div>npi</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>grd_yr</div> <div>double precision</div> <div></div> </div>	<div> <div></div> <div>pri_spec</div> <div>character varying</div> <div></div> </div>		1	1215283908	1998	FAMILY MEDICINE	2	1215257605	2007	GENERAL PRACTICE	3	1215248273	2010	INTERNAL MEDICINE							
Data Output	Explain	Messages																																										
<div> <div></div> <div>geofips</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>per_capita_income</div> <div>integer</div> <div></div> </div>																																											
1	1001	46814																																										
2	1003	50953																																										
3	1005	37850																																										
Data Output	Explain	Messages	Notifications																																									
<div> <div></div> <div>npi</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>grd_yr</div> <div>double precision</div> <div></div> </div>	<div> <div></div> <div>pri_spec</div> <div>character varying</div> <div></div> </div>																																										
1	1215283908	1998	FAMILY MEDICINE																																									
2	1215257605	2007	GENERAL PRACTICE																																									
3	1215248273	2010	INTERNAL MEDICINE																																									
<div> <div>1</div> <div>SELECT * FROM population;</div> </div> <table> <tr> <th>Data Output</th><th>Explain</th><th>Messages</th><th>Notifications</th></tr> <tr> <td> <div> <div></div> <div>geofips</div> <div>integer</div> <div></div> </div> </td><td> <div> <div></div> <div>population</div> <div>integer</div> <div></div> </div> </td><td> <div> <div></div> <div>pop_density</div> <div>double precision</div> <div></div> </div> </td><td></td></tr> <tr> <td>1</td><td>1001</td><td>55200</td><td>35.8534189940189</td></tr> <tr> <td>2</td><td>1003</td><td>208107</td><td>50.5415035640023</td></tr> <tr> <td>3</td><td>1005</td><td>25782</td><td>11.247981205619</td></tr> </table>	Data Output	Explain	Messages	Notifications	<div> <div></div> <div>geofips</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>population</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>pop_density</div> <div>double precision</div> <div></div> </div>		1	1001	55200	35.8534189940189	2	1003	208107	50.5415035640023	3	1005	25782	11.247981205619	<div> <div>1</div> <div>SELECT * FROM zip;</div> </div> <table> <tr> <th>Data Output</th><th>Explain</th><th>Messages</th><th>Notifications</th></tr> <tr> <td> <div> <div></div> <div>zip</div> <div>integer</div> <div></div> </div> </td><td> <div> <div></div> <div>st</div> <div>character varying</div> <div></div> </div> </td><td> <div> <div></div> <div>geofips</div> <div>integer</div> <div></div> </div> </td><td> <div> <div></div> <div>county_name</div> <div>character varying</div> <div></div> </div> </td></tr> <tr> <td>1</td><td>601</td><td>PR</td><td>72001</td><td>Adjuntas</td></tr> <tr> <td>2</td><td>602</td><td>PR</td><td>72003</td><td>Aguada</td></tr> <tr> <td>3</td><td>603</td><td>PR</td><td>72005</td><td>Aguadilla</td></tr> </table>	Data Output	Explain	Messages	Notifications	<div> <div></div> <div>zip</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>st</div> <div>character varying</div> <div></div> </div>	<div> <div></div> <div>geofips</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>county_name</div> <div>character varying</div> <div></div> </div>	1	601	PR	72001	Adjuntas	2	602	PR	72003	Aguada	3	603	PR	72005	Aguadilla
Data Output	Explain	Messages	Notifications																																									
<div> <div></div> <div>geofips</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>population</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>pop_density</div> <div>double precision</div> <div></div> </div>																																										
1	1001	55200	35.8534189940189																																									
2	1003	208107	50.5415035640023																																									
3	1005	25782	11.247981205619																																									
Data Output	Explain	Messages	Notifications																																									
<div> <div></div> <div>zip</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>st</div> <div>character varying</div> <div></div> </div>	<div> <div></div> <div>geofips</div> <div>integer</div> <div></div> </div>	<div> <div></div> <div>county_name</div> <div>character varying</div> <div></div> </div>																																									
1	601	PR	72001	Adjuntas																																								
2	602	PR	72003	Aguada																																								
3	603	PR	72005	Aguadilla																																								



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