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

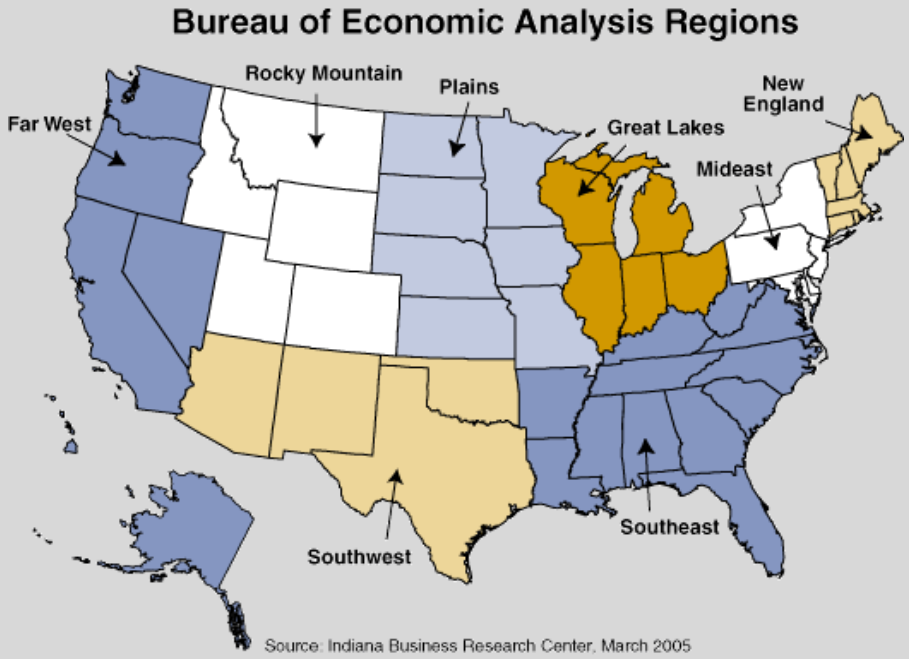
WHAT ARE WE TRYING TO DO?

DATA EXPLORATION

ANALYSIS RESULTS

WHAT ARE WE TRYING TO DO?

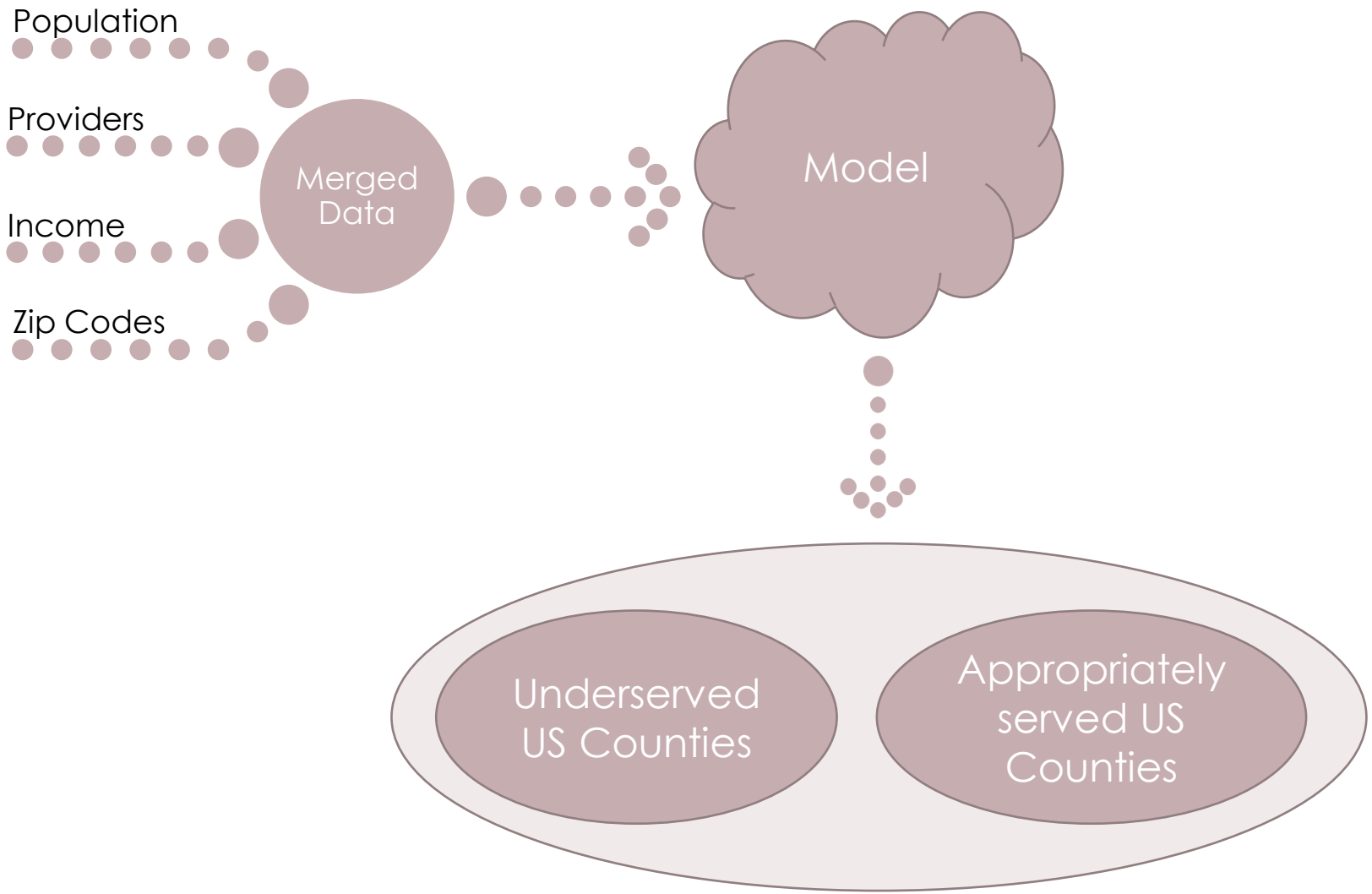
Build a machine learning model that can determine primary care provider availability



- 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.

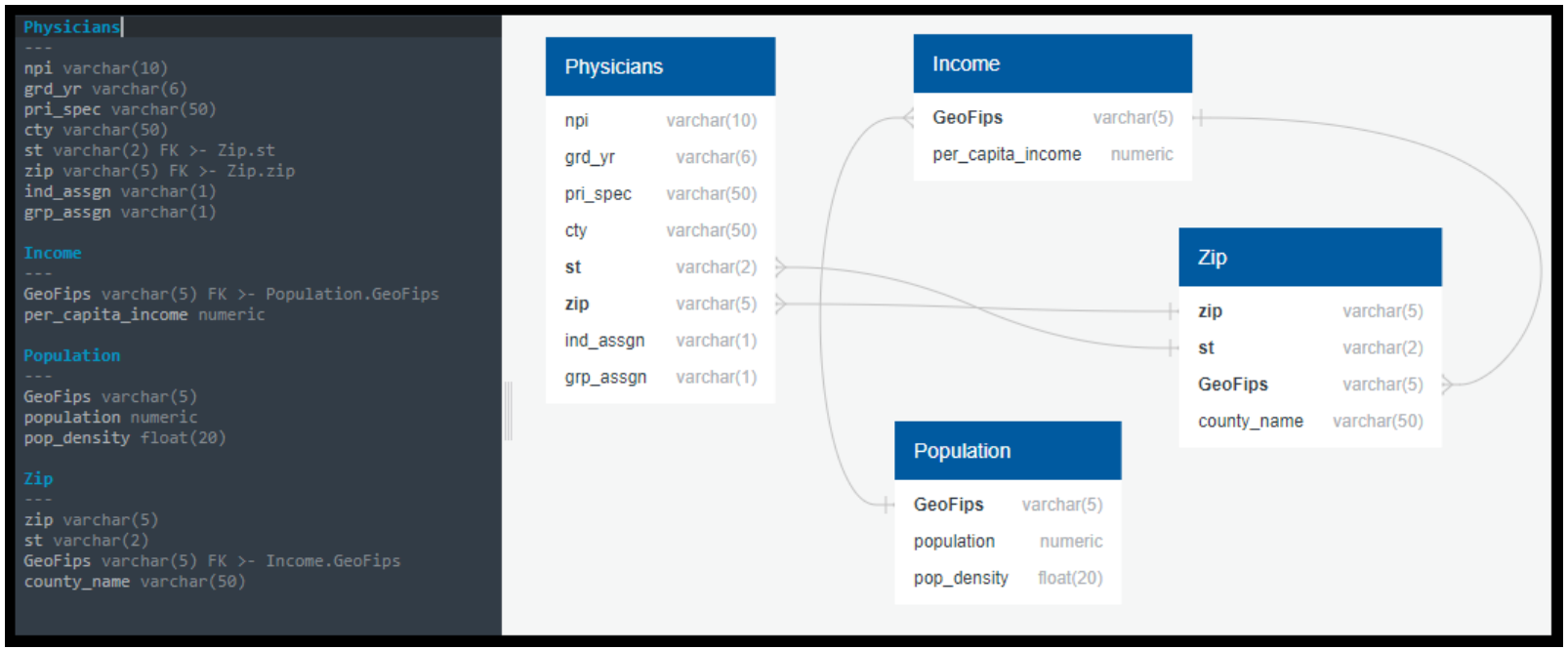
DATA EXPLORATION

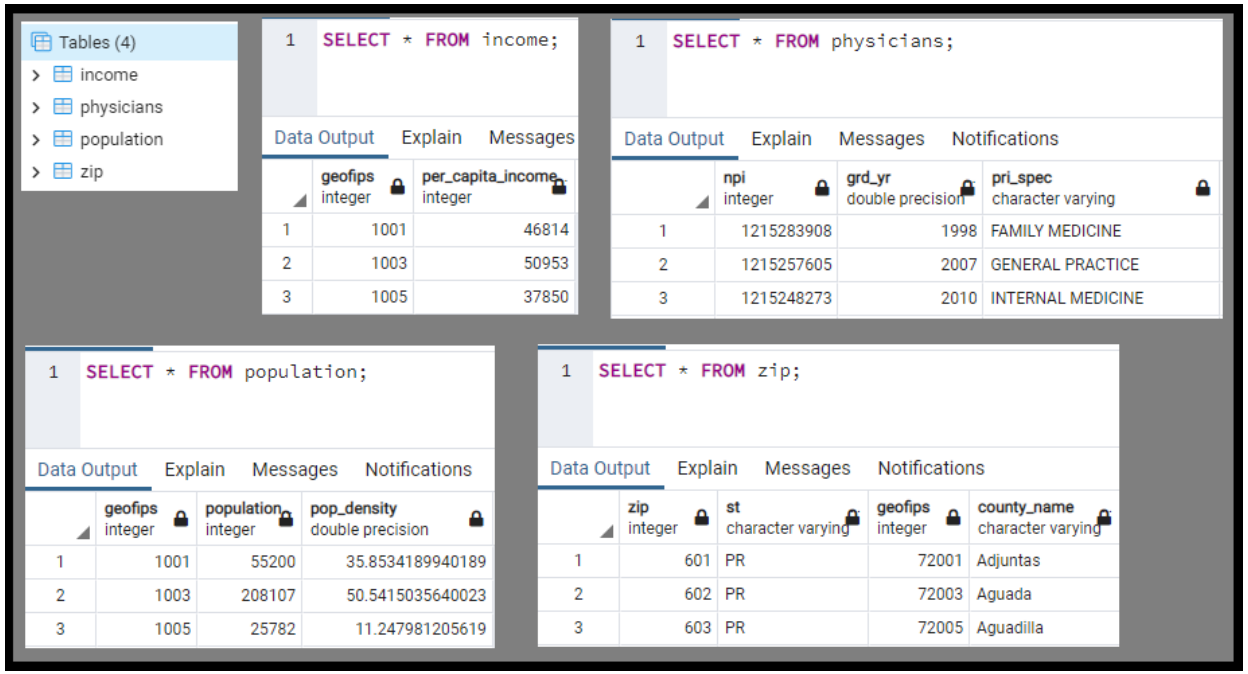


Data Exploration

- 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

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





Database

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

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
Income data from 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
Region data from 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
Population data from 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
Zip code data from simplemaps Basic Download	zip, stae_id, county_fips, and county_name from uszips.csv	https://simplemaps.com/data/us-zips
Physician data 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	https://data.cms.gov/provider-data/dataset/mj5m-pzi6