Characteristics of Health Insurance and Access to Care across the United States and Minnesota

Elsa Carlson, Kylee LaPierre, Tarick Mehanna, and Logan Sell

Introduction and Initial Questions

Introduction

- Primary objective: identify trends and patterns of health insurance coverage so that resources can be best allocated to areas or groups where coverage is less common/accessible
- Definitions for our analysis:
  - Sufficient access to care: those who do not delay or avoid seeking healthcare services due to cost or location.
  - Uninsured: people who don't have any type of health insurance.
- 2019 is the primary year for our data
  - To interpret data from immediately pre-pandemic
- Our model used data from 2015 to 2019

Initial Questions

1. How do uninsurance rates vary among demographics such as race, age, sex, population, income, and educational attainment level?
2. How does access to healthcare vary at a national level and by county?
3. How does access to healthcare vary among Minnesota counties?
4. Do rates of uninsurance vary based on offered coverage types?
5. Does the rate of uninsurance vary for individuals based on the number of hospitals in their county?
6. Can we predict the coverage level for a county based on demographics, income or other factors?
7. For those who are covered, is their coverage sufficient for their health needs?
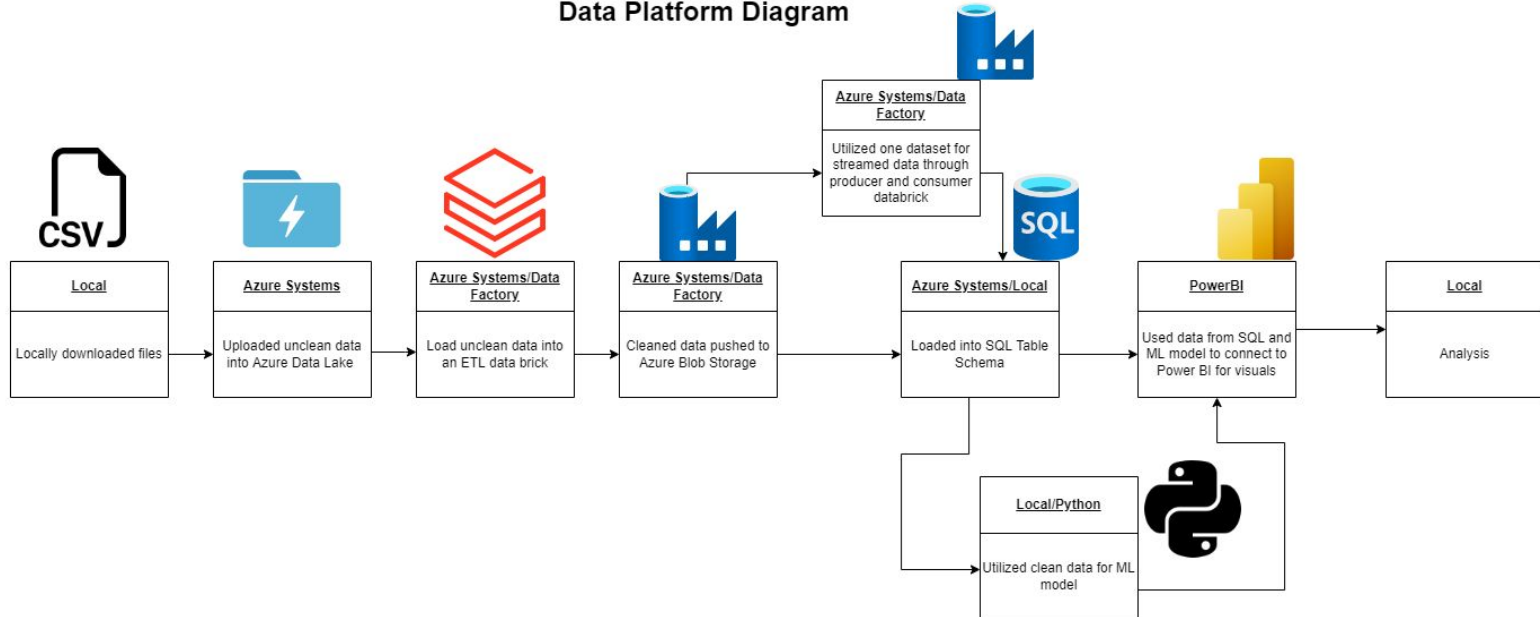
Research Process
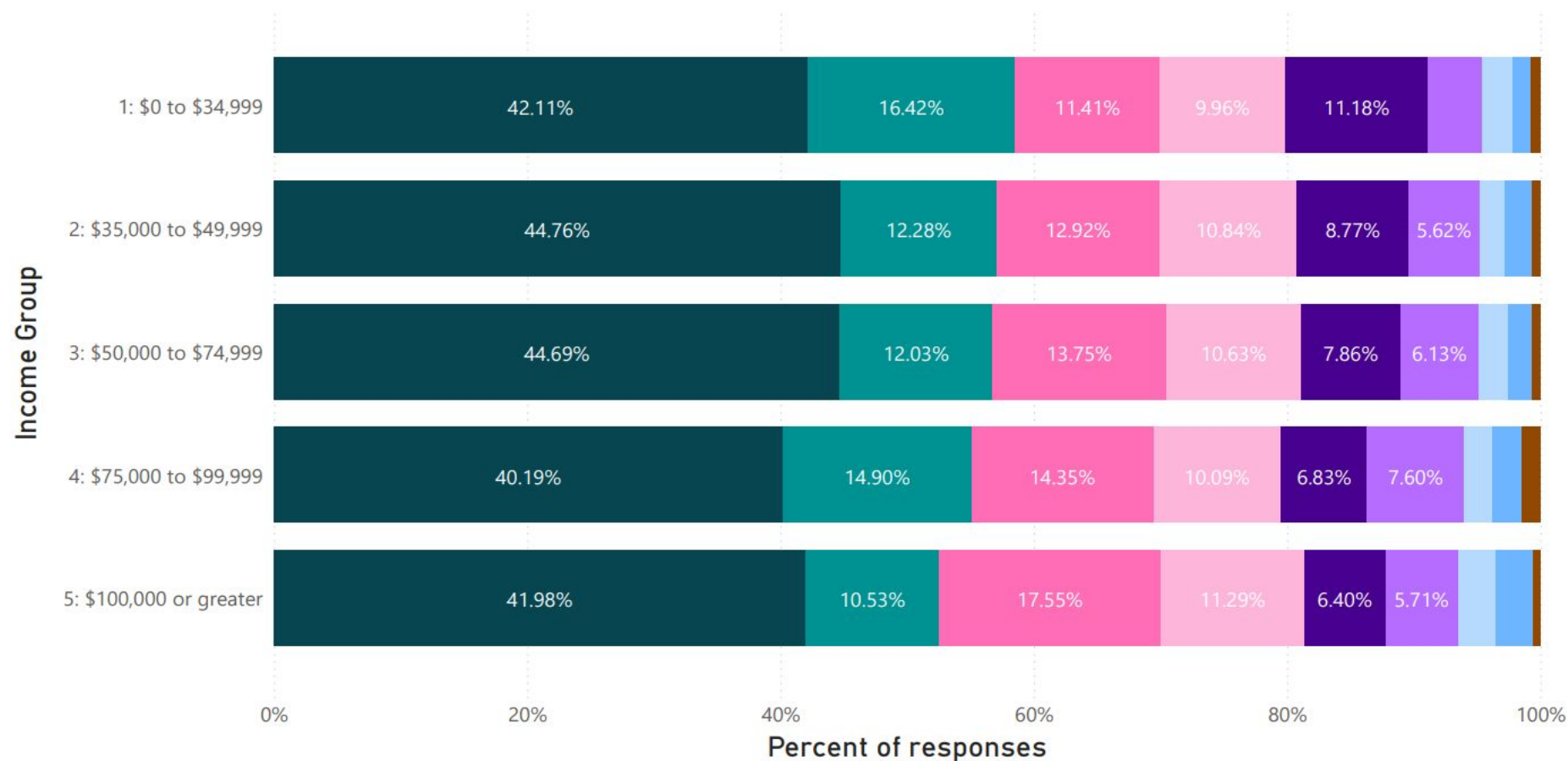
Finding & Processing Data

- Used data mainly from the US Census Bureau and the CDC
- Behavioral Risk Factor Surveillance System (BRFSS)
- Data process: extraction → cleaning & transformation → loading to SQL database

Data Platform Diagram

Azure Systems/Data Factory
Utilized one dataset for streamed data through producer and consumer databrick

Local
Locally downloaded files

Azure Systems
Uploaded unclean data into Azure Data Lake

Azure Systems/Data Factory
Load unclean data into an ETL data brick

Azure Systems/Data Factory
Cleaned data pushed to Azure Blob Storage

Azure Systems/Local
Loaded into SQL Table Schema

PowerBI
Used data from SQL and ML model to connect to Power BI for visuals

Local
Analysis

Local/Python
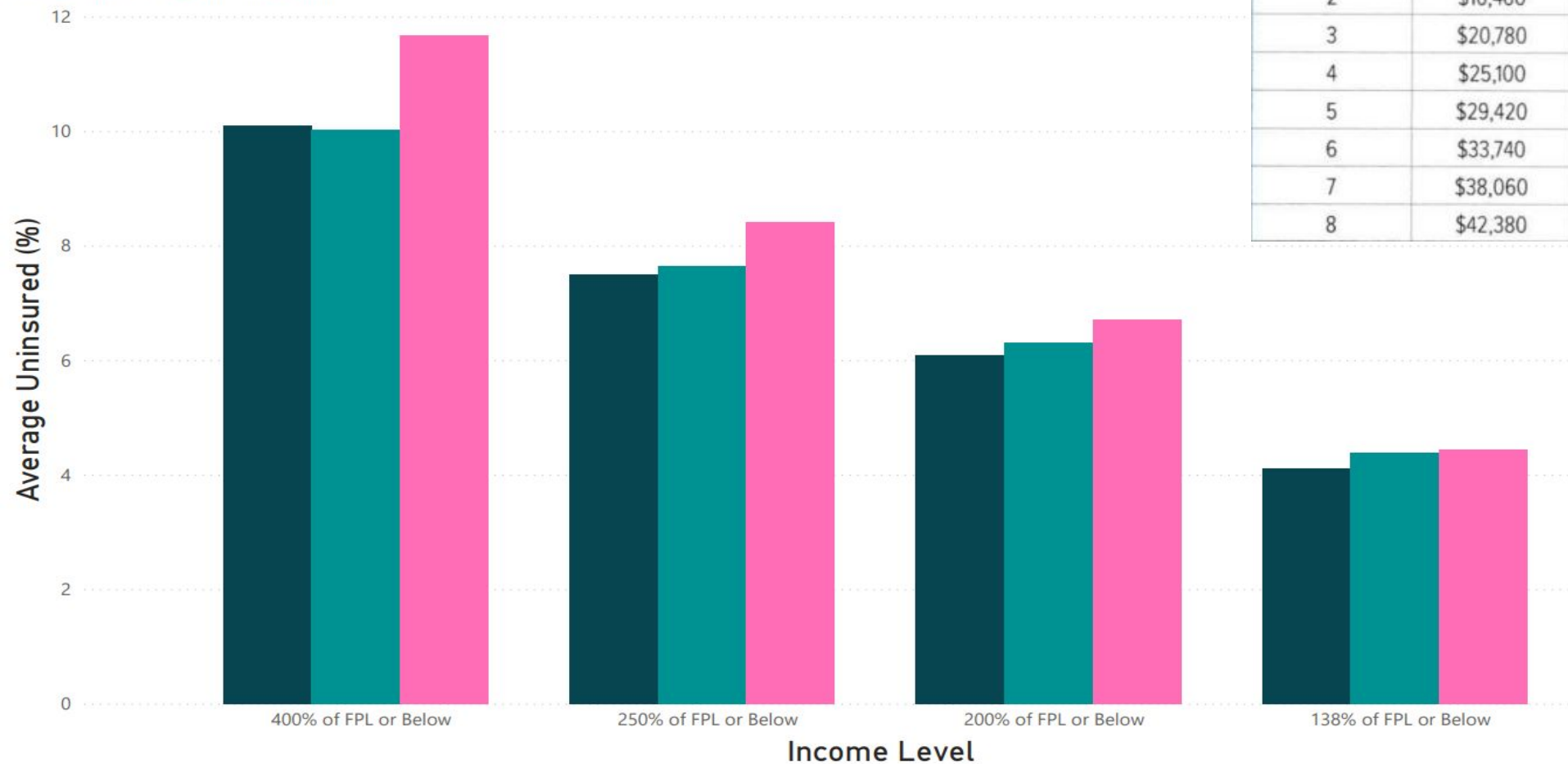Utilized clean data for ML model

Reasons Americans Lack Coverage

● Too expensive ● Not eligible ● Don't want coverage ● Can't find a plan ... ● Process too c... ● Coverage has ... ● Other reason ● Lost job ● Missed de...

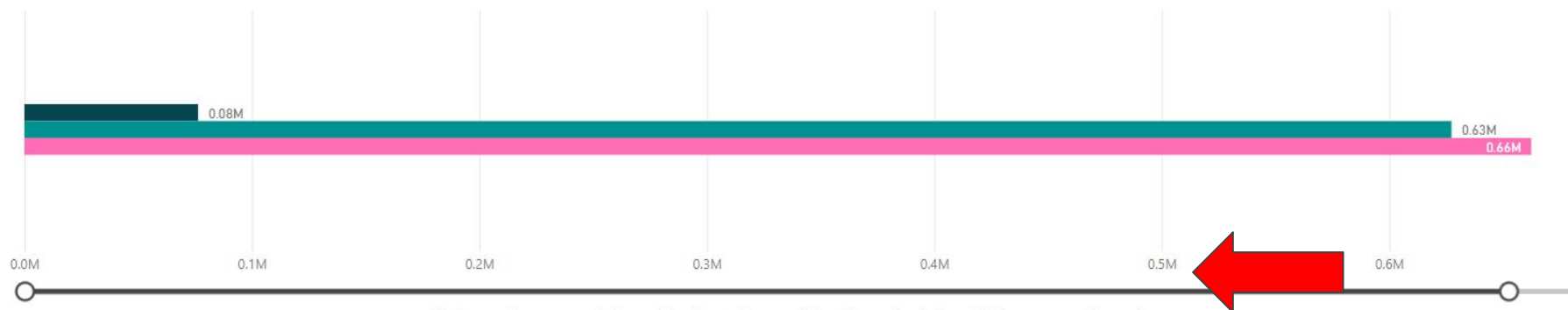| Income Group	Too expensive	Not eligible	Don't want coverage	Can't find a plan	Process too c...	Coverage has ...
1: $0 to $34,999	42.11%	16.42%	11.41%	9.96%	11.18%	
2: $35,000 to $49,999	44.76%	12.28%	12.92%	10.84%	8.77%	5.62%
3: $50,000 to $74,999	44.69%	12.03%	13.75%	10.63%	7.86%	6.13%
4: $75,000 to $99,999	40.19%	14.90%	14.35%	10.09%	6.83%	7.60%
5: $100,000 or greater	41.98%	10.53%	17.55%	11.29%	6.40%	5.71%

Percent of responses

Average Percent Uninsured by Income Level

Sex ● Both Sexes ● Female ● Male

# in Household	100% FPL
1	$12,140
2	$16,460
3	$20,780
4	$25,100
5	$29,420
6	$33,740
7	$38,060
8	$42,380

Average Uninsured (%) (y-axis)

Income Level (x-axis): 400% of FPL or Below, 250% of FPL or Below, 200% of FPL or Below, 138% of FPL or Below

Public Insurance Types by Enrollment Numbers in Select Minnesota Counties

● VA Coverage ● Medicare Coverage ● Medicaid/Means-Based Coverage

Coverage	Value
VA Coverage	0.08M
Medicare Coverage	0.63M
Medicaid/Means-Based Coverage	0.66M

0.0M   0.1M   0.2M   0.3M   0.4M   0.5M   0.6M

Private Insurance Types by Enrollment Numbers in Select Minnesota Counties

● Military Coverage ● Direct Purchase Coverage ● Employer-Based Coverage

Coverage	Value
Military Coverage	0.01M
Direct Purchase Coverage	0.58M
Employer-Based Coverage	2.62M

0.0M   0.5M   1.0M   1.5M   2.0M   2.5M
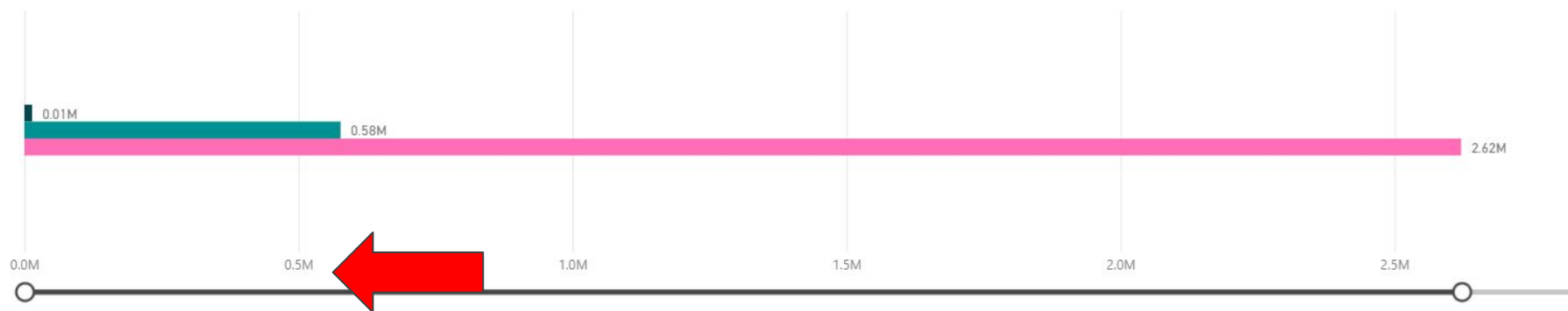
Note: These measures include people with more than one type of coverage.

The select counties consist of Anoka, Blue Earth, Carver, Crow Wing, Dakota, Hennepin, Olmsted, Ramsey, Rice, Scott, Sherburne, St. Louis, Stearns, Washington, and Wright Counties.

Has cost prevented you from seeking medical care anytime in the last 12 months?

Response ● Yes
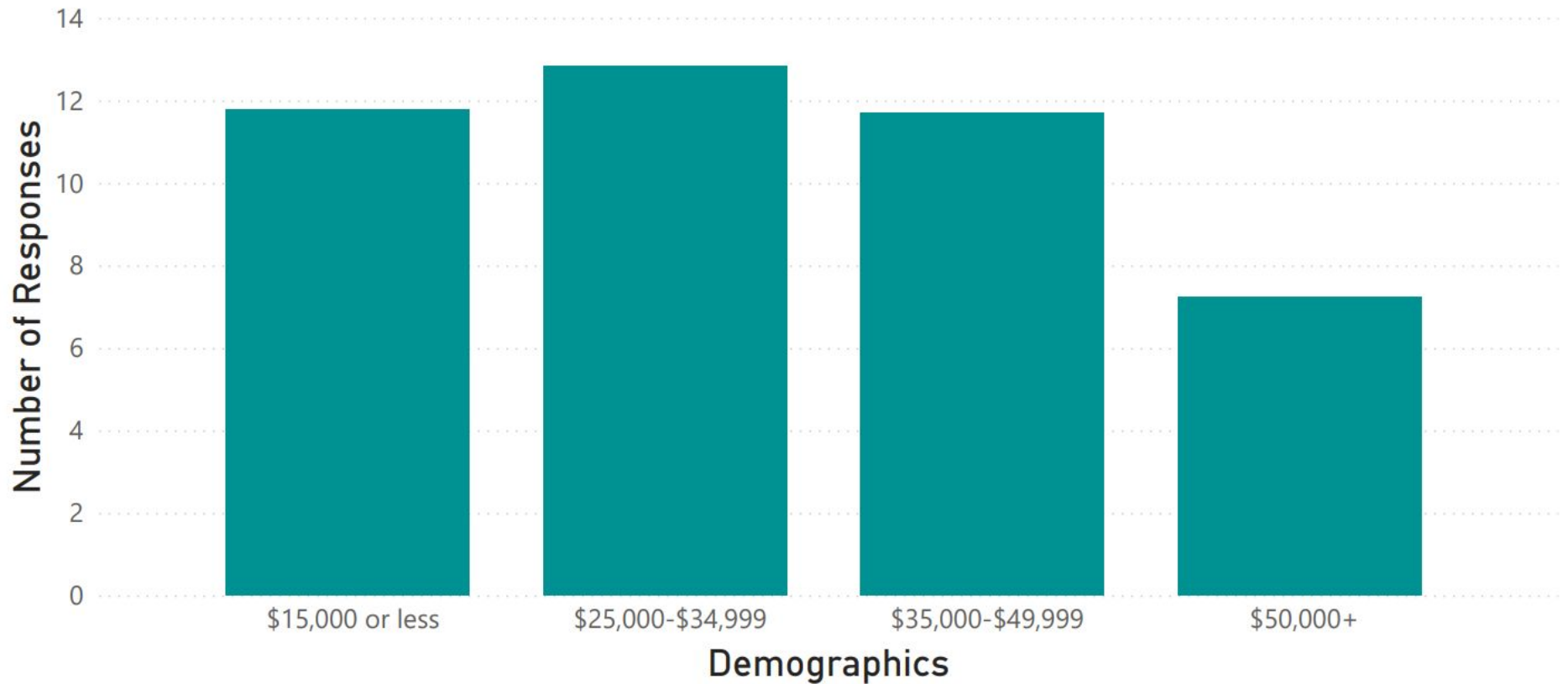
Average Percent Uninsured by County in Minnesota

Cluster of low rates (2%)

Nobles Co. (7.85%)
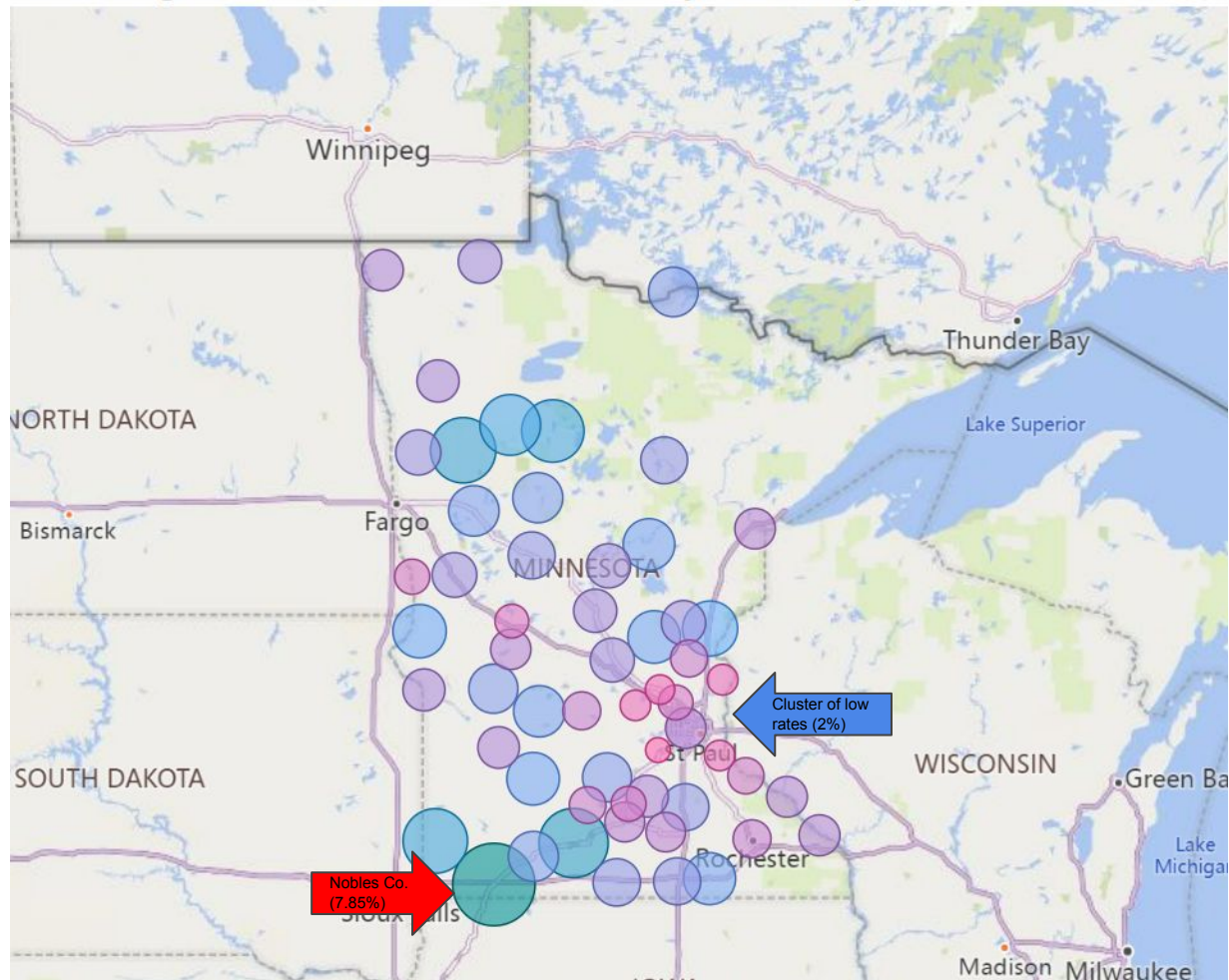
Machine Learning

Machine Learning Model

Columns of interest:

- Age
  - Under 6 years, 6 to 18 years, 19 to 25 years, 26 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 to 74 years, and 75 years and older.
- Race
  - White, African American, Asian, American Indian and Alaska Native, Native Hawaiian and Pacific Islander, Hispanic or Latino, and Some other race not categorized..
- Gender
  - Male or female
- Place of birth
  - Native born, Foreign born, Naturalized, Not a citizen
- Educational Attainment
  - Less than high school, High school or equivalent, Some college or associates degree, and Bachelors degree or higher.
- Annual Salary
  - Under $25,000, $25,000 to $49,999, $50,000 to $74,999, $75,000 to $99,999 and $100,000 and over.

```python
# Creating training and testing datasets
y_test = x_m['Uninsured']
x_test = x_m.drop(columns = 'Uninsured', axis = 1).copy()


Y_train = h_m['Uninsured']
X_train = h_m.drop(columns = 'Uninsured', axis = 1).copy()



# testing all hyperparameters
from sklearn.model_selection import GridSearchCV
algorithm = LogisticRegression(max_iter = 1_000)


p_range = ['l1', 'l2', 'elasticnet', 'none']
s_range = ['liblinear', 'lbgfs', 'sag', 'saga', 'newton-cg']
c_range = list(np.arange(0.0001, 2, 0.0001))
param_grid = dict(solver = s_range, penalty = p_range, C = c_range)


grid = GridSearchCV(algorithm,param_grid,scoring='accuracy')
grid_search = grid.fit(X_train, Y_train)
print(grid_search.best_params_)


reg_log1 = LogisticRegression(C = 1.201, penalty = 'l1', solver = 'liblinear', max_iter = 10_000)
reg_log1.fit(X_train, Y_train)
y_pred = reg_log1.predict(x_test)
```
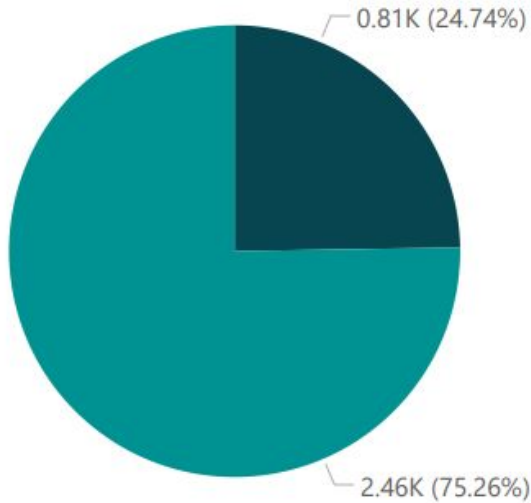
Predicted Insurance Category from ML Model
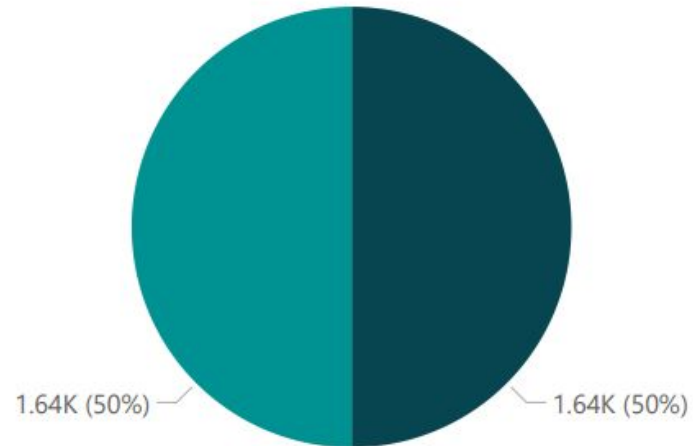
Actual Insurance Category

0.81K (24.74%)

2.46K (75.26%)

Insurance Category
- Insured
- Uninsured

Insurance Category
- Insured
- Uninsured

1.64K (50%)

1.64K (50%)

```python
train = health_model[(health_model['Insurance_Category'] == 'Uninsured')].copy()
train2 = train.drop(columns = ['HealthModelID', 'State', 'County', 'Insurance_Category', 'Year']).copy()

demo = Xdemo.drop(columns = ['TestDemoID', 'State', 'County']).copy()
demo = demo[demo['Insurance_Category'] == 'Uninsured']
demo.drop(columns = 'Insurance_Category', inplace = True)

# Scaling the data
scalar = StandardScaler().fit(demo)
demo_scaled = pd.DataFrame(scalar.transform(demo), columns = demo.columns)

scalar = StandardScaler().fit(train2)
train2_scaled = pd.DataFrame(scalar.transform(train2), columns = train2.columns)


# Creating training and testing datasets
Y_train = train2_scaled['Under_25000S'].copy()
X_train = train2_scaled.drop(columns = ['Under_25000S', 'Total_Population']).copy()

Y_test = demo_scaled['Under_25000S'].copy()
X_test = demo_scaled.drop(columns = ['Under_25000S', 'Total_Population']).copy()

reg_lin = ElasticNetCV(max_iter = 5_000)
reg_lin.fit(X_train, Y_train)
y_pred = reg_lin.predict(X_test)
```
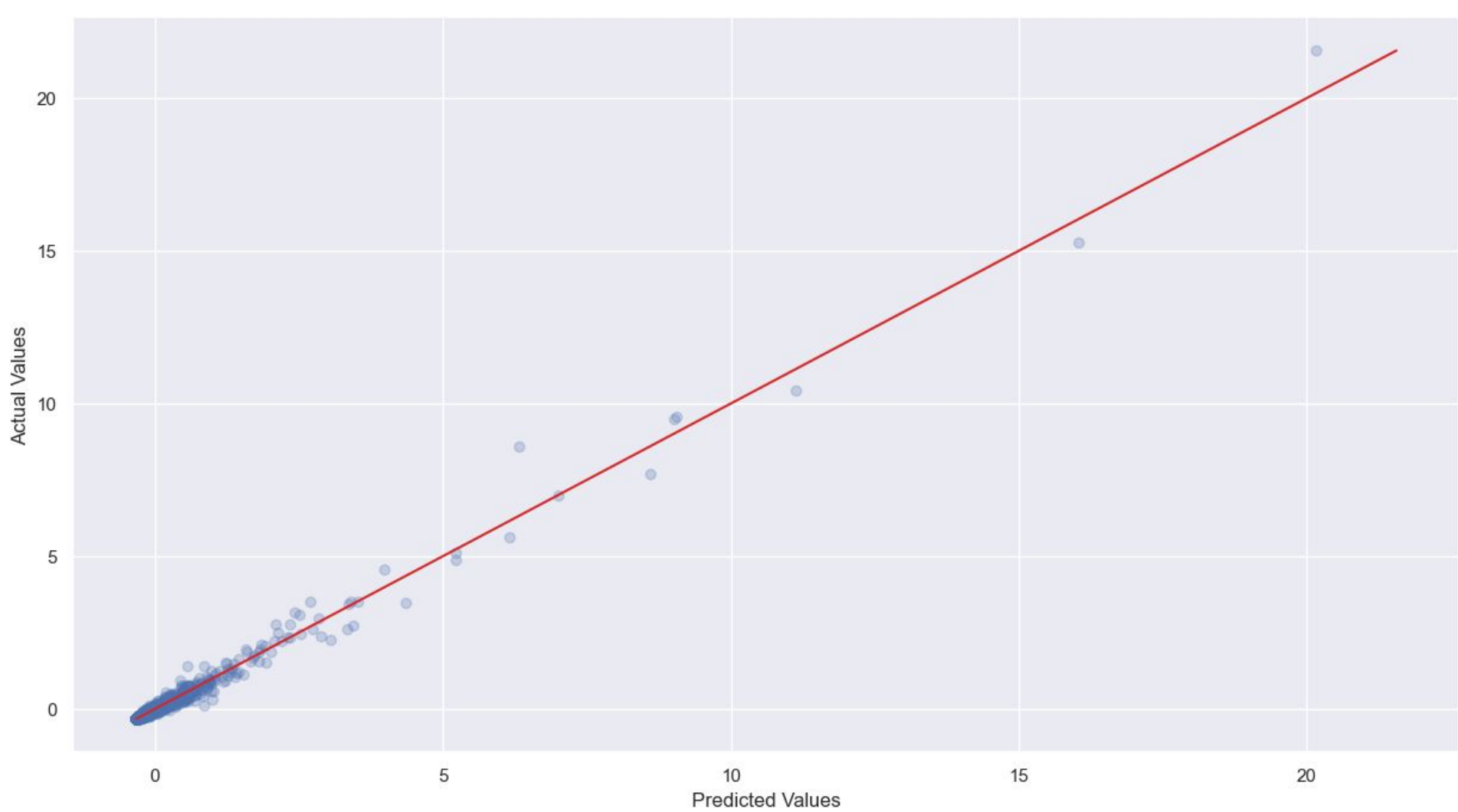
Recommendations

Recommendations for Future Data Collection & Research

- Increase data collection in rural and lower-population counties
- Track reasons for lacking insurance by demographics other than income level
- Study intersection of demographics in relation to avoidance of medical care & access to coverage (including location, race/ethnicity, income, etc)
- Classify public insurance enrollees who work full-time by whether or not their employers offer employer-based insurance
- Track tobacco usage in relation to insurance affordability
- Investigate strategies to increase insurance affordability

Sources

1. Bureau, U. S. C. (2019). CB1900CBP: All Sectors: County Business Patterns by Legal Form of Organization and Employment Size Class for U.S., States, and Selected Geographies: 2019. Explore census data. Retrieved September 21, 2022, from https://data.census.gov/cedsci/table?q=hospital+cb&g=0100000US%240400000_0400000US27&tid=CBP2019.CB1900CBP

2. Bureau, U. S. C. (2019). DP05: ACS Demographic and Housing Estimates. Explore census data. Retrieved September 21, 2022, from https://data.census.gov/cedsci/table?q=DP05&g=0100000US%240400000&tid=ACSDP1Y2019.DP05

3. Bureau, U. S. C. (2019). S2701: Selected Characteristics of Health Insurance Coverage in the United States. Explore census data. Retrieved September 21, 2022, from https://data.census.gov/cedsci/table?q=health+insurance&g=0100000US%240500000_0400000US27&tid=ACSST1Y2019.S2701&moe=false

4. Bureau, U. S. C. (2019). S2703: Private Health Insurance Coverage by Type and Selected Characteristics. Explore census data. Retrieved September 21, 2022, from https://data.census.gov/cedsci/table?q=health+insurance&g=0400000US27%2C27%240500000&tid=ACSST1Y2019.S2703

5. Bureau, U. S. C. (2019). S2704: Public Health Insurance Coverage by Type and Selected Characteristics. Explore census data. Retrieved September 21, 2022, from https://data.census.gov/cedsci/table?q=health+insurance&g=0400000US27%2C27%240500000&tid=ACSST1Y2019.S2704

Sources

6. Bureau, U. S. C. (2022, August 10). 2008 - 2020 small area health insurance estimates (SAHIE) using the American Community Survey (ACS). Census.gov. Retrieved September 21, 2022, from https://www.census.gov/data/datasets/time-series/demo/sahie/estimates-acs.html
Only used the year 2019 from this website.

7. CDC. (2021, April 5). NHIS - 2019 NHIS. Centers for Disease Control and Prevention. Retrieved September 21, 2022, from https://www.cdc.gov/nchs/nhis/2019nhis.htm

8. CDC. (2021, December 16). Behavioral risk factors: Selected metropolitan area risk trends (SMART) MMSA prevalence data (2011 to present). Centers for Disease Control and Prevention. Retrieved September 21, 2022, from https://chronicdata.cdc.gov/Behavioral-Risk-Factors/Behavioral-Risk-Factors-Selected-Metropolitan-Area/j32a-sa6u?category=Behavioral-Risk-Factors&view_name=Behavioral-Risk-Factors-Selected-Metropolitan-Area

9. CDC. (2021, September 14). BRFSS: Table of health care access/coverage. Centers for Disease Control and Prevention. Retrieved September 21, 2022, from https://chronicdata.cdc.gov/Behavioral-Risk-Factors/BRFSS-Table-of-Health-Care-Access-Coverage/f7a2-7inb

10. Prescott, R. (2016, June 22). USA-cities-and-states/US_CITIES_STATES_COUNTIES.CSV at master · Grammakov/USA-Cities-and-States. GitHub. Retrieved September 26, 2022, from https://github.com/grammakov/USA-cities-and-states/blob/master/us_cities_states_counties.csv

Questions?

SCAN ME

QR Code that links to our Github.