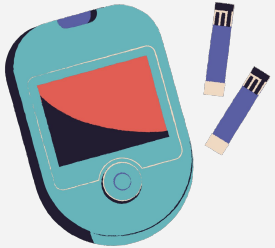
Prevalence of Diabetes in United States Counties

presented by Lindsey Hunnicutt, Data Scientist

Background

Diabetes: a disorder characterized by elevated blood glucose (> 125 mg/dL)(1)

- Complications: damage to the nerves, heart, and/or kidneys, oral health, vision & more…(2)
- Prevalence of diabetes in U.S.(3)
  - 11.3 % Diagnosed
  - 23.0% Adults - Undiagnosed
  - 38.0 % Adults - Prediabetic



Type II Diabetes:

- 9/10 cases can be prevented(4)
- Personal Risk Factors (examples)
  - Weight
  - Poor Diet
  - Low Exercise
- *Also related…*
  - *Chronic stress(5)*
  - *Depression & Anxiety(6)*
  - *Emotional Eating(7)*
  - *Eating Disorders(8)*
- Community Risk Factors
  - Vary by state, county
  - “Food Environment”

1. Diabetes Tests. CDC. https://www.cdc.gov/diabetes/basics/getting-tested.html#:~:text=A%20fasting%20blood%20sugar%20level,higher%20indicates%20you%20have%20diabetes.

2. Prevent Diabetes Complications. CDC. https://www.cdc.gov/diabetes/managing/problems.html#:~:text=Common%20diabetes%20health%20complications%20include,how%20to%20improve%20overall%20health.

3. National Diabetic Statistics Report. https://www.cdc.gov/diabetes/data/statistics-report/index.html.

4. Simple Steps to Preventing Diabetes. Harvard TH Chan School of Public Health. https://www.hsph.harvard.edu/nutritionsource/disease-prevention/diabetes-prevention/preventing-diabetes-full-story/

5. Blood Sugar & Stress. Diabetes Teaching Center at the University of California San Francisco. https://dtc.ucsf.edu/types-of-diabetes/type2/understanding-type-2-diabetes/how-the-body-processes-sugar/blood-sugar-stress/

6. The Association between Diabetes Mellitus and Depression. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4863499/#R13

7. Stress Eating and Health: findings from MIDUS, a National Study of U.S. Adults. Tsekenova et al. (2013). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3733123/

8. Binge Eating Disorder in Patients with Type 2 Diabetes: Diagnostic & Management Challenges. Chevinsky et al. (2019). https://www.dovepress.com/binge-eating-disorder-in-patients-with-type-2-diabetes-diagnostic-and--peer-reviewed-fulltext-article-DMSO

Research Question

How well can I predict prevalence of diabetes in U.S. counties based on food environmental features?

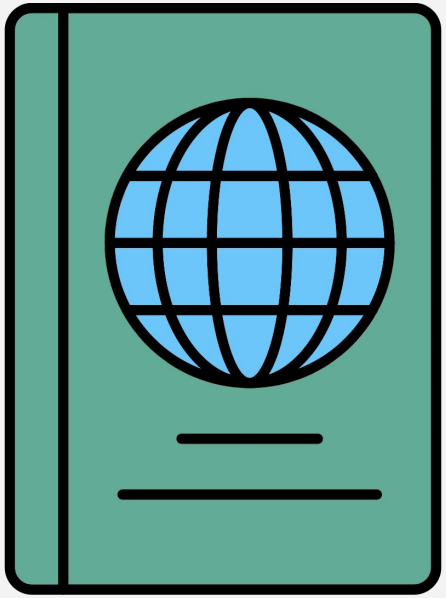
Data Description

Food Environment Atlas

3,142 U.S. counties* assessed

281 Variables assessed 2008 - 2018

1. **Health**
2. **Access**
3. **Assistance**
4. **Food Insecurity**
5. **Local (Farms)**
6. **Restaurants**
7. **Socioeconomics**
8. **Stores**
9. **Taxes**



Diabetes Atlas (Target)

- **Prevalence of Diabetes in Adults**
- **Collected During US Surveillance System**
- **Year 2018**

Mental Health Providers (2017-2018)

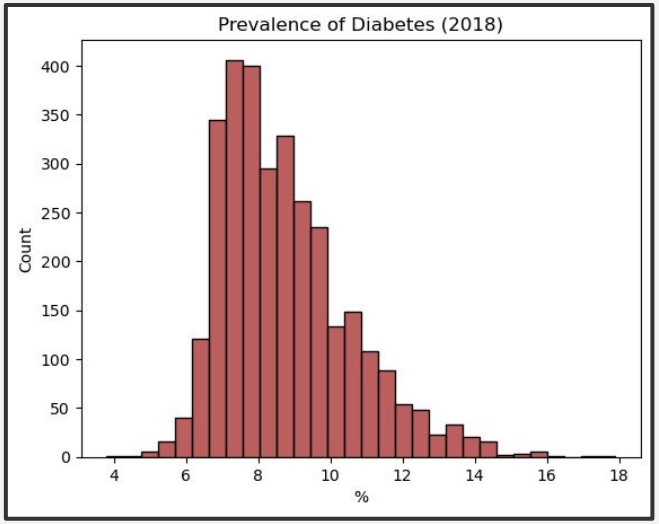
- ***# of MH providers per 100k pop.***
  - ***State-level***

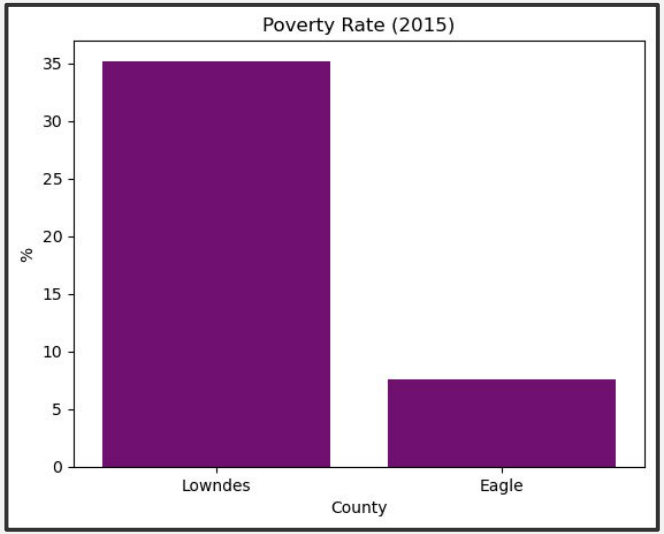
****Adults Receiving MH Services in the Last Year (2015-2016)***

- ***Percentage***
  - ***State-level***

Data: Collection, Cleaning & EDA

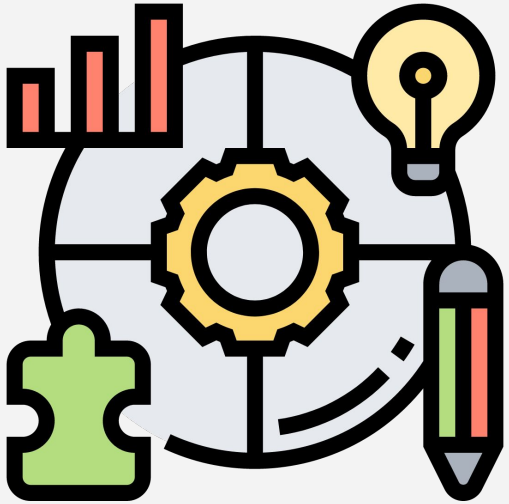
- Data Collection
  - Websites (USDA, CDC, SAMHSA, UHF)
  - Spreadsheets, Tables & CSV's
- Data Cleaning
  - Feature Selection
  - Null values
  - Reconciling counties
- EDA
  - Part I - SQL
    - Cross-comparisons
    - ORDER BY
    - ANTI-JOIN
  - Part II - Pandas
    - Correlations
    - Visualizations
    - Case study
      - Eagle County, CO (3.3%)
      - Lowndes County, AL (23.5%)





Methods

- Feature Selection & Modification
  - MH information into Health dataframe
  - Target Values into FEA datasets
  - Creating Classes: High v. Low Prevalence
    - Mean as separator (8.72%)
    - Accounts for change in % year-to-year
- Modeling
  - Baseline = majority class (0 - Low - ~0.58)
  - KNN
  - Random Forest Classification
    - Highest scoring datasets
  - F1 Score as Metric
    - Models predicted negatives well
    - Positives predicted less well
    - Takes TP, FP, FN into account



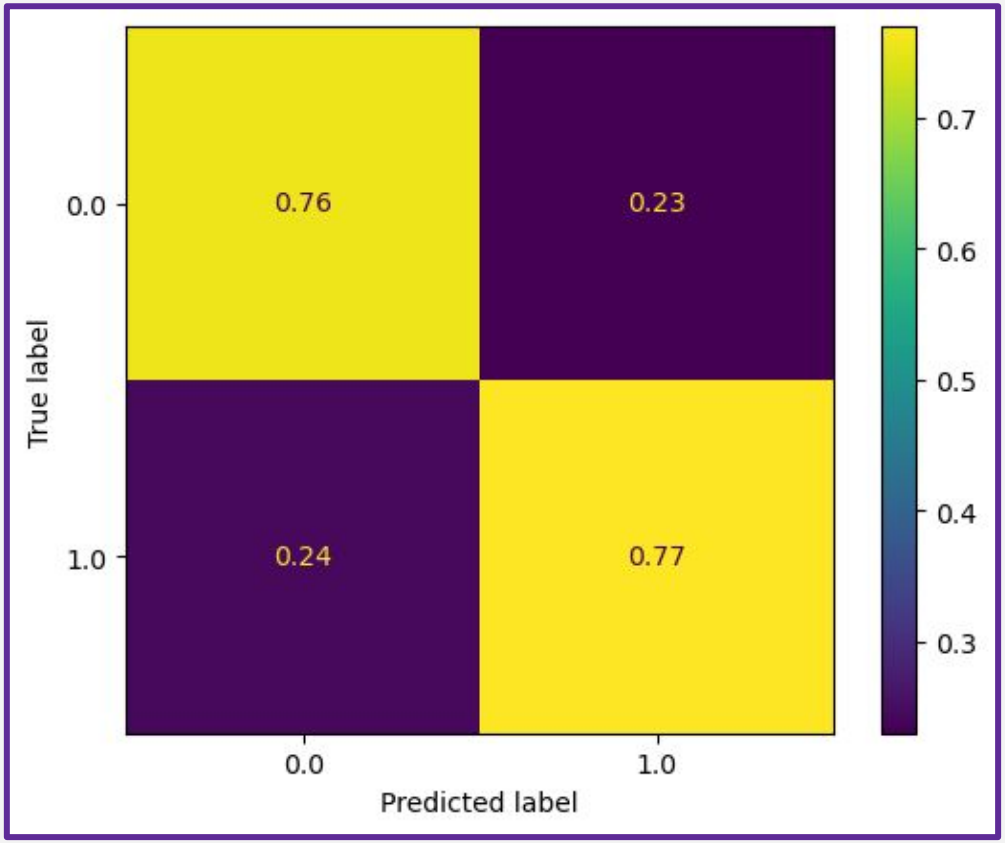
Models & Evaluation

Model	Score (F1)
Health & Restaurants	0.713
Health	0.706
Food Environment Atlas	0.679
Restaurants	0.654
Food Insecurity	*0.626 (KNN)*
Assistance	*0.626 (KNN)*

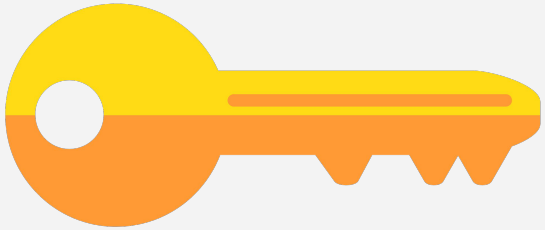
Health & Restaurants (Random Forest Classifier)

Baseline: 0.583 Accuracy: Specificity:

Recall: Precision: F1: 0.713



Key Features



Health

- # of Recreational Facilities (2011, 2016)
  - Total number
  - Per 1000 population
- % Adults receiving MH services in last yr.* (‘15- ‘16)
  - Age 18+
  - Age 18-26
  - Age 26+
- # of MH providers* (2017, 2018)
  - Per 100,000 population
  - State Rank

*State-level variables

Restaurants

- Fast Food Restaurants (2011, 2016)
  - Total number
  - Per 1000 population
- Full-Service Restaurants (2011, 2016)
  - Total number
  - Per 1000 population
- Per capita Fast Food Restaurant Sales (‘07, ‘12)
- Per capita Full-Service Restaurant Sales (‘07, ‘12)

Conclusion

Answer)

- **Up to 76% of low prevalence predictions correct**
- **Up to 77% of high prevalence predictions correct**

We can predict prevalence of diabetes based on:

- **Health**
- **Restaurants**
- ***Assistance***
- ***Food Insecurity***

Demonstrates key areas of focus for community leaders

  - **Interventions targeted to specific communities**

<u>Limitations</u>

- **Datasets are older**
- **Models prone to overfit**
  - **More useful for inference**

Next Steps & Recommendations

Next Steps

- Research the connection between mental health & physical disease states
- Study impact of plant based diets
- Research cost of preventative versus other types of healthcare
    - Total cost of diagnosed diabetes in 2017 (US):
        - $327 billion[1]

Recommendations:

- Education: how to eat healthier on a low budget
- Address portion size & nutrient content
    - Restaurants
    - Products
- Increase minimum wage[2]
- Preventative healthcare

1. Statistics about Diabetes. https://diabetes.org/about-us/statistics/about-diabetes.

2. A Calculation of the Living Wage. https://livingwage.mit.edu/articles/99-a-calculation-of-the-living-wage

Thank you!