Health disparities in Chicago based on Socio-economic and Business Factors

Adithyan Subramanian
Mentors: Mukesh Mithrakumar and Danny Wells

Background and Goals

- Chicago - Extremely diverse and segregated
- Segregation = health disparities
- Which neighborhoods are more healthy?
- How do businesses affect neighborhood health?
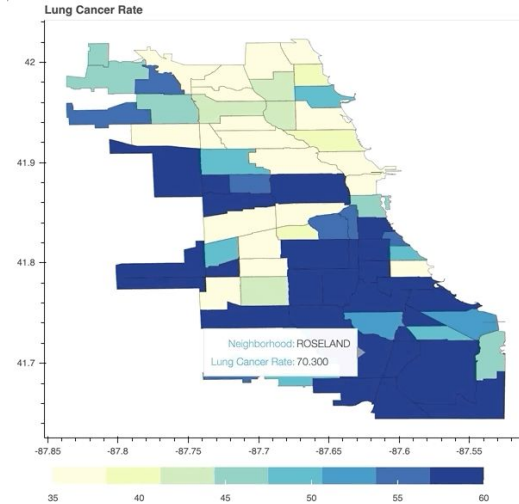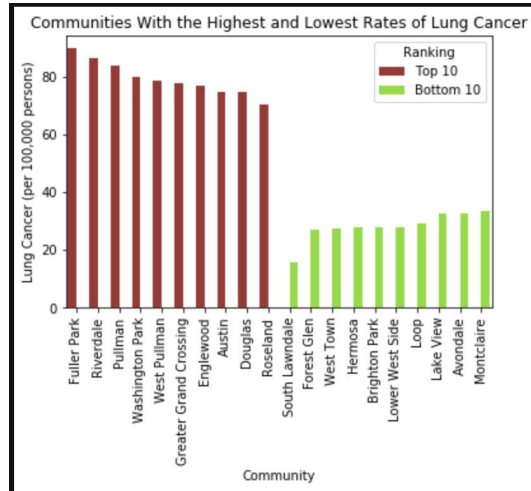- Predict lung cancer and infant mortality rates

Data

- 2 Datasets
  - Health dataset
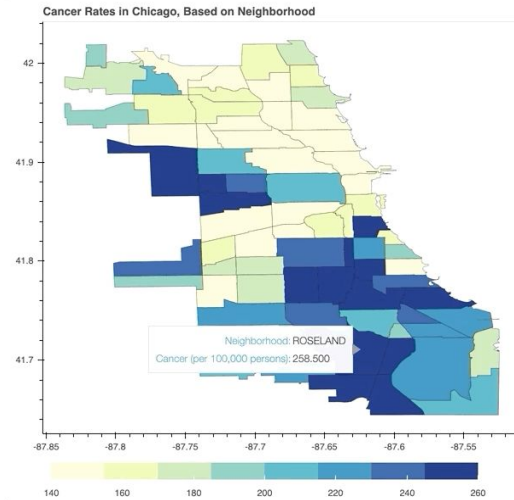 - Business dataset
- Both from the Chicago Data Portal

Exploratory Data Analysis

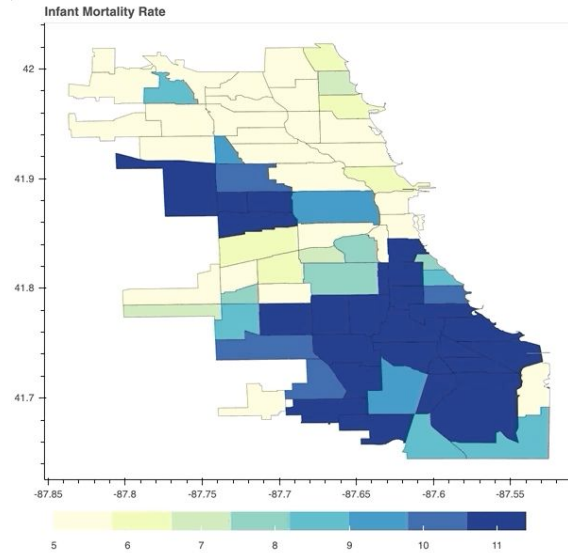# Lung Cancer Disparities: A Case study
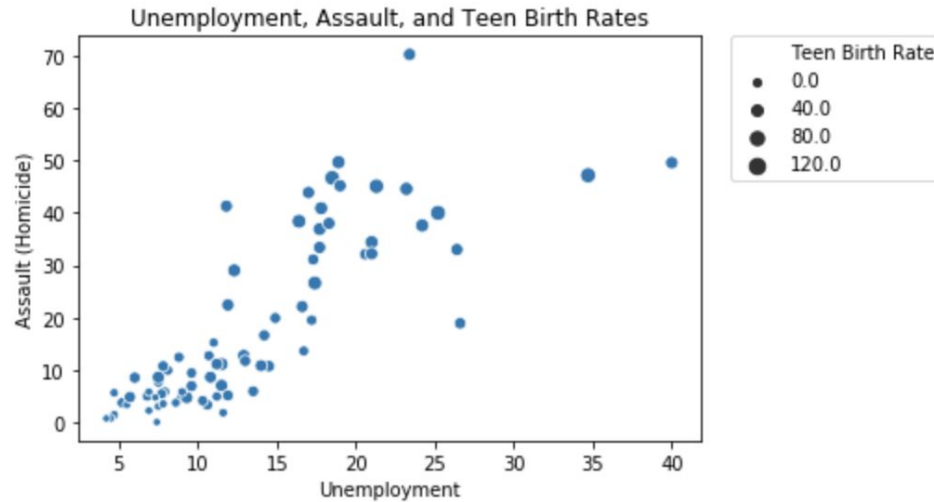
General Cancer Rates

Cancer Rates in Chicago, Based on Neighborhood

Infant Mortality Rates

Infant Mortality Rate

Comparison of Socioeconomic Factors

Unemployment, Assault, and Teen Birth Rates

The Effect of Tobacco
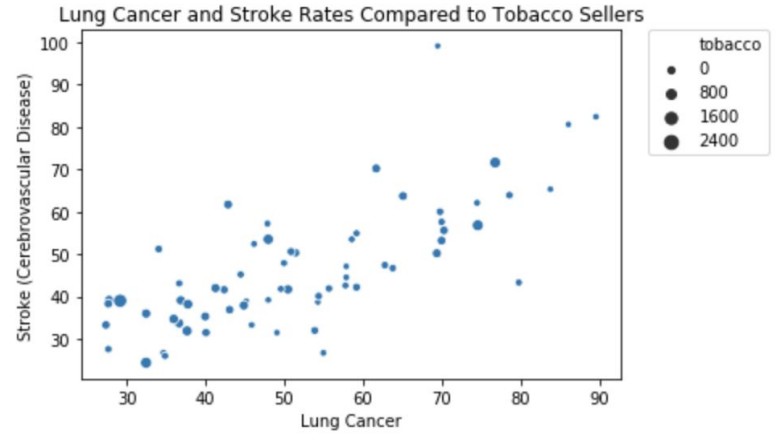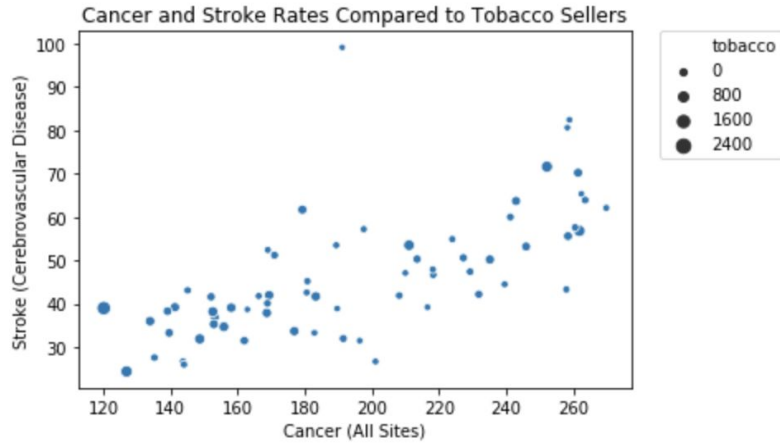
Infant Mortality and Child-Care Businesses

Prenatal Care and Infant Mortality Rates Compared to Children Related Businesses
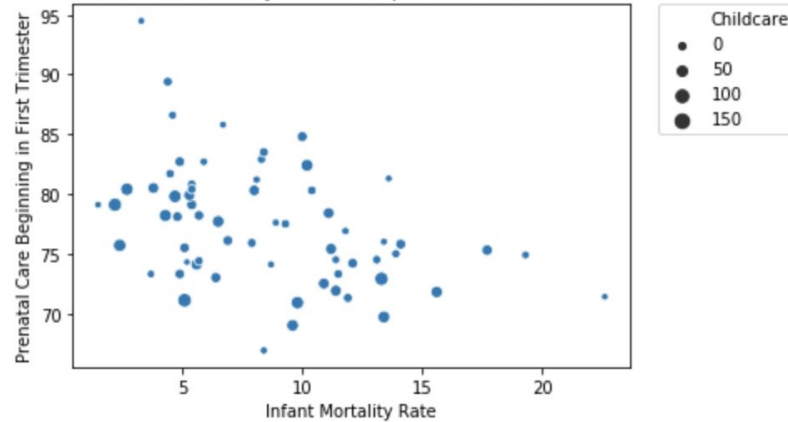
EDA Conclusions

- Significant difference in lung cancer rates in neighborhoods with a high number of tobacco sellers compared to neighborhoods with low number of tobacco sellers.

- No significant difference in infant mortality rate based on the number of child related businesses.
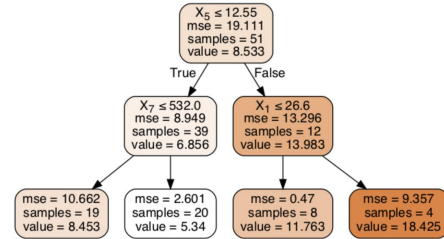
Machine Learning

Infant Mortality

- Used following metrics as features
    - Number of "child-related businesses"
    - Percentage of residents with no high school diploma
    - Per capita income, crowded housing
    - Percentage of households living under the poverty line
    - Percentage of babies born with low birth weight
    - General fertility rate
    - Number of liquor sellers

Infant Mortality Results

- Linear Regression
  - R^2: 0.63
- Non-Optimized Decision Tree
  - MSE: 15.6
  - Test Set Accuracy: 0.22
- Optimized Decision Tree
  - MSE: 13.0
  - Test Set Accuracy: 0.22
  - Only used Percentage of households living under the poverty line, general fertility rate, and number of childhood related businesses

Lung Cancer

- Used Following Metrics as features
    - number of tobacco sellers
    - Percentage of residents with no high school diploma
    - Per capita income
    - Number of liquor sellers
    - Crowded housing
    - Percentage of households living under the poverty line
    - Percentage of adults who are dependent

Lung Cancer Optimized Decision Tree

- Test Set Accuracy: 0.198
- Mean-squared error: 149.85
- Utilized only per capita income, number of liquor sellers, and crowded housing
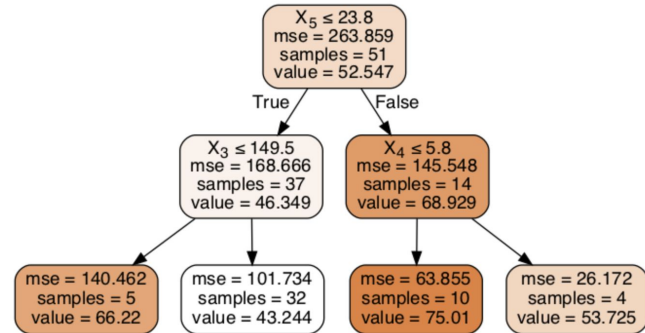
Other Algorithms

Method	MSE	Test Set Accuracy
Non-optimized Decision tree	160.9	0.138
Optimized Decision Tree	149.9	0.198
Non-optimized Random Forest	119.2	0.36
Optimized Random Forest	89.9	0.52
Lasso Regression	154.5	0.17
Ridge Regression	154.4	0.17

What Went Wrong

- Small data set
  - Only 77 neighborhoods
- Not extensive
  - Does not incorporate all business and health care data

Conclusion

Final Thoughts and Recommendations

- Must prioritize South and West sides of the city
- Less tobacco sellers = less lung cancer
- Prioritize solutions where per capita income rises, and the number of liquor sellers and citizens in crowded housing decreases
- Prioritize solutions the number of families decreases and where the general fertility rate, and number of child related businesses increases

Next Steps

- Other businesses?
- Other outcomes?
- How location of businesses affect health outcomes?
- More data?
  - From other cities?
- Models on other metrics