

A photograph of a wind farm with several white wind turbines on a green, grassy hill under a clear blue sky. The turbines are arranged in a line, with the closest one in the foreground and others receding into the distance. The blades are white and the towers are also white. The ground is covered in green grass, and there are some small trees or bushes in the distance.

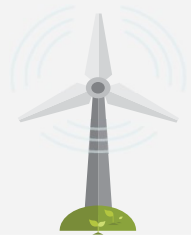
# **Predicting Windfarms Impact on Home Values**

**Final Report**

## Objective

1. **Build a model that can be used to predict the change in value of a home**
2. **Assess if a windfarm in the vicinity of a home impacts the value of the home**

*Home owners, home buyers, & local government officials interested in tax revenue can use this model to predict the likelihood that a wind farm will impact home prices.*



## Contents

1

**About the Data**

2

**Observations**

3

**Machine Learning**

4

**Summary**

## About the Data Sources



### US Wind Turbine database as a csv file

- The United States Wind Turbine Database (USWTDB) provides the locations of land-based and offshore wind turbines in the United States, corresponding wind project information, and turbine technical specifications.



### Zillow Housing Data as a csv file

- Zillow Home Value Index (ZHVI) is a smoothed, seasonally adjusted measure of the median estimated home value across a given region and housing type. Zillow provides data on sold homes, including median sale price for various housing types, and sales volume



### US Zipcode Database as a csv file

- uszipcode is a feature-rich zipcode database with a zipcode search engine that can be customized to the desired search results.

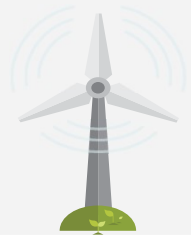
## About the Data

### Features of Highest Interest

Features that appear most significant to impact a home's value  
and  
assess if windfarms impact home values  
are



- **Annual change in median home value** (*as a percentage*) for zip codes with wind farms and zip codes 25 miles away without wind farms
- **Population density** of zip codes with wind farms and adjoining or nearby zip codes without wind farms
- **Median income** of zip codes with wind farms and adjoining or nearby zip codes without wind farms



## Observations Windfarm Locations

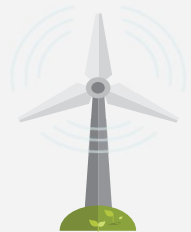
- States with highest and lowest count of zip codes with windfarms...

*Highest 5*

	state	zipcode_count
0	CA	19
4	MA	13
5	MN	6
7	OH	5
9	RI	3

*Lowest 5*

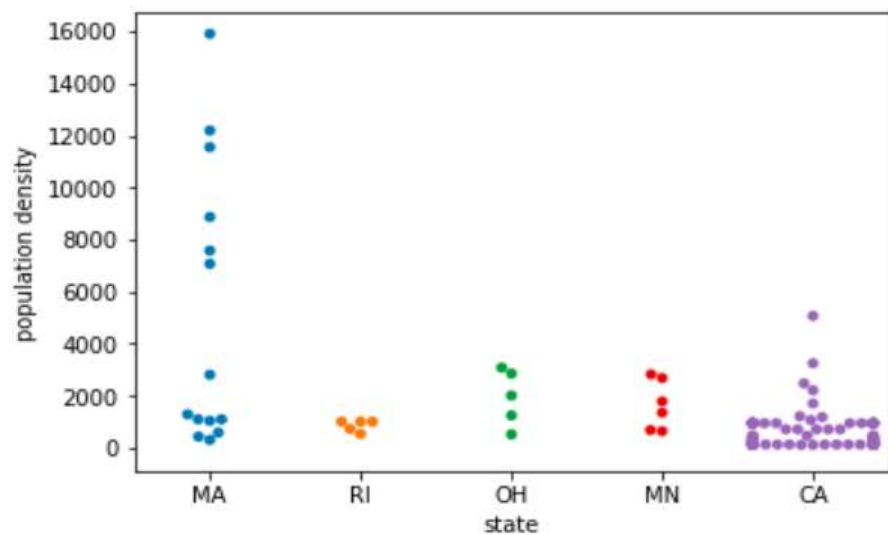
	state	zipcode_count
3	IL	1
6	NY	1
8	OK	1
11	UT	1
12	WA	1



## Observations Population and Income (1 of 2)

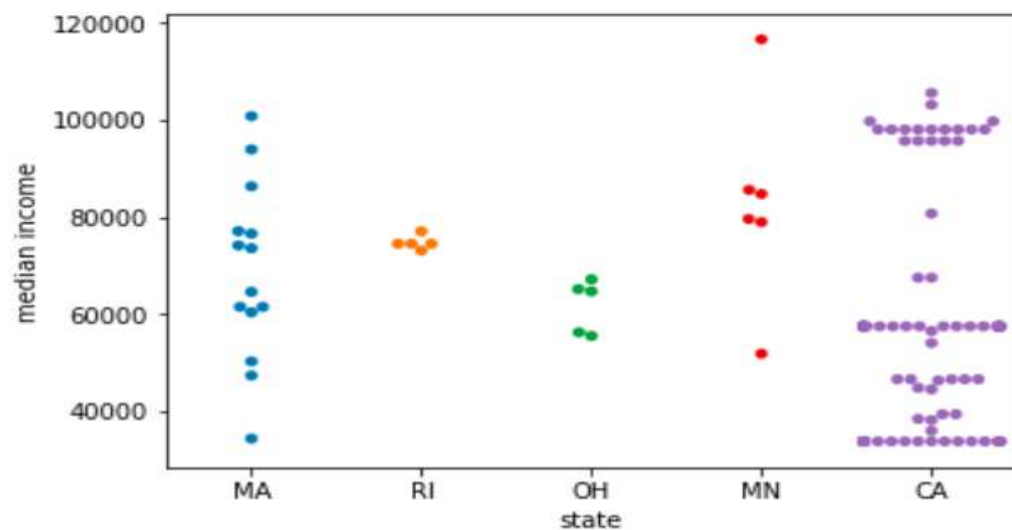
MA has several windfarms in areas w/high population densities

*Population Density*



CA has a number of windfarms in areas with a high median income

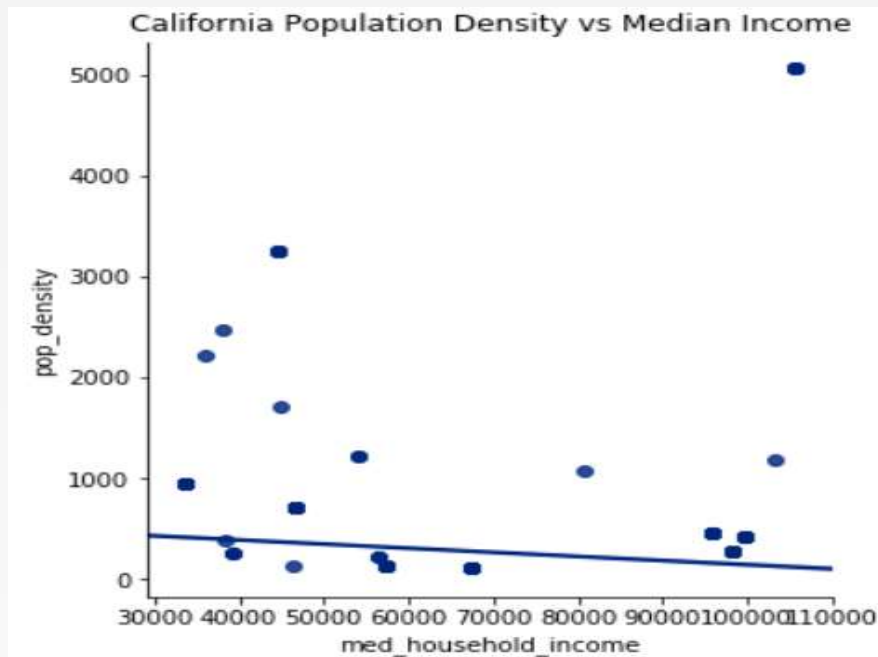
*Median Income*



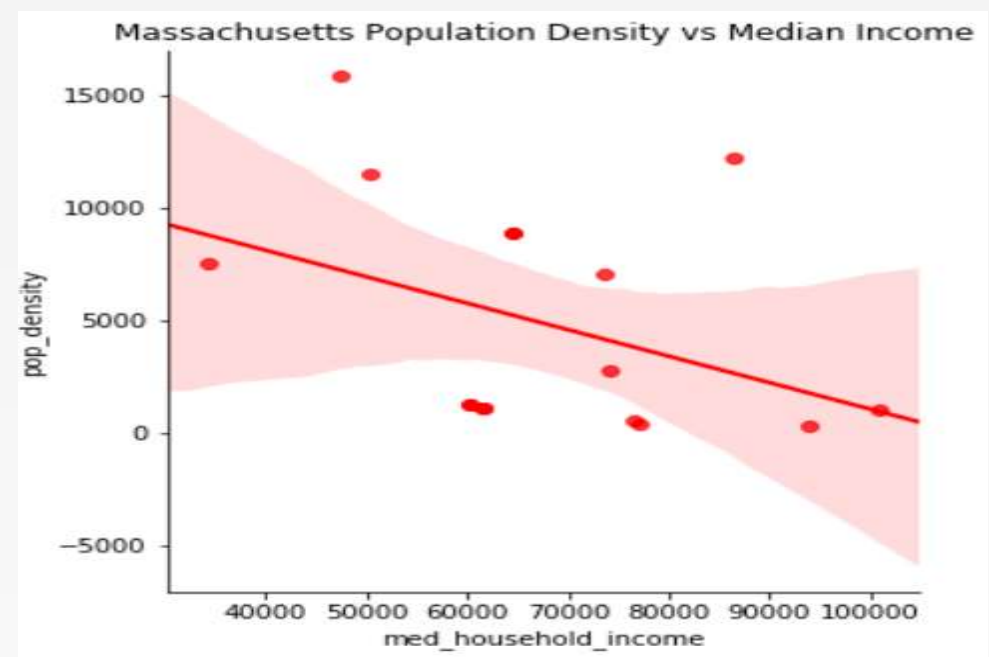


## Observations Population and Income (2 of 2)

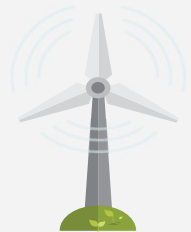
*California*



*Massachusetts*





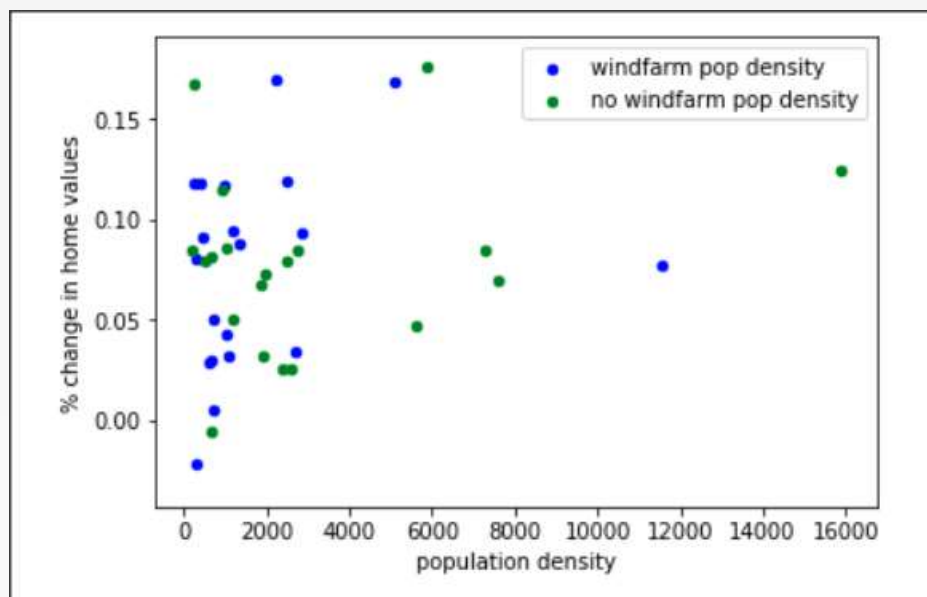


## Observations Correlations (1 of 2)

- The correlation coefficient between **Population Density & Mean % Change in Home Values** differs by only 2%

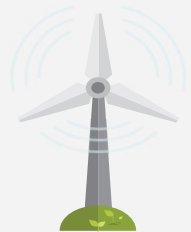
Correlation of zip codes  
with windfarms:

23%



Correlation of zip codes  
without windfarms:

25%

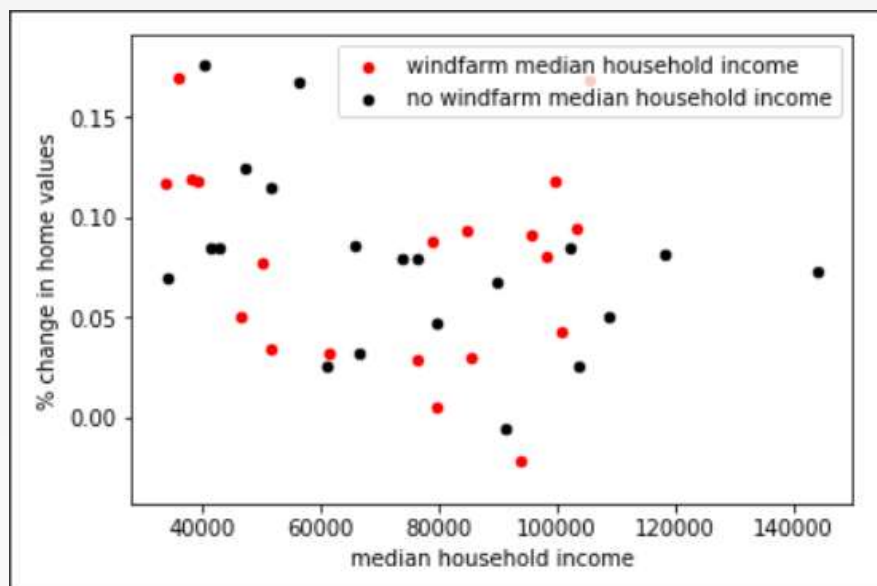


## Observations Correlations (2 of 2)

- The correlation coefficient between **Median Income & Mean % Change in Home Values** differs by 22%

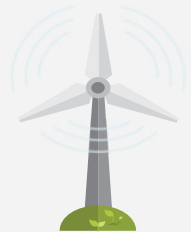
Correlation of zip codes  
with windfarms:

**19%**



Correlation of zip codes  
without windfarms:

**41%**

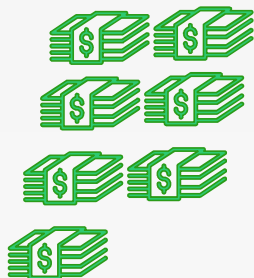


## Observations Significance

- Using 2017 data, there **is not** a statistically significant difference in mean home values for zip codes with versus without windfarms

 = \$50,000

Mean home value for zip codes **with windfarms**



\$360,000

Mean home value for zip codes **without windfarms**



\$400,000

Difference:  
\$40,000

*p-Value*  
**0.99**

*Difference in value 95% confidence Interval*  
**-\$90,000 to \$188,000**

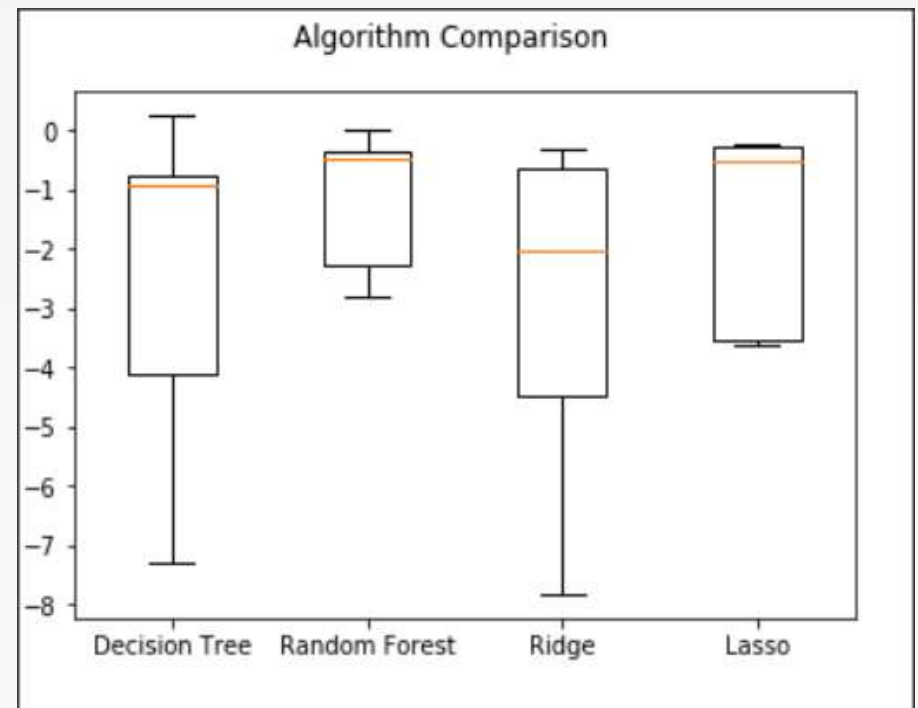


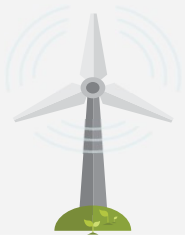
## Machine Learning

- Four algorithms were employed to research a model for predicting home values...

### *Points of interest to learn from the box chart...*

- All the algorithms produce a similar top, but Random Forest and Lasso appear have the highest top and a much tighter range than Ridge or Decision Tree. These appear to be the best models for further analysis.
- Random Forest's plot is short and Lasso has very short whiskers indicating a high level of agreement on home value predictions across zip codes.
- Height of the Decision Tree and Ridge plots are similar, and are much greater than Random Forest and Lasso. This lends further support to not continuing with the Decision Tree and Ridge algorithms.
- The medians of Decision Tree, Random Forest, and Lasso are very similar, but the size of their box plot varies, showing a much higher variance in the Decision Tree results.





## Summary

### Restatement of Objectives...

1. Build a model that can be used to predict the change in value of a home
2. Assess if a windfarm in the vicinity of a home impacts the value of the home

### Key Conclusions...

- Using a Random Forest Algorithm a model predicting home values with 77% accuracy was achieved
- The presence of a windfarm does not appear to have an impact on home values in the surrounding area
- The most significant predictor is population density, followed closely by median income.
- For median income, correlation to home values is much lower for zipcodes with a windfarm than zipcodes without a windfarm (19% versus 41%). A potential reason is windfarms may tend toward more rural areas and in rural areas median income may be less of a factor on home values.
- Not surprisingly, windfarms are primarily in rural areas, though several zip codes with high population density and high median incomes host windfarms