

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 of the turbines are white and have a three-bladed design.

Predicting Windfarms Impact on Home Values

Milestone Report

Objective

Build an algorithm that can be used to predict the change in value of a home based on proximity to windmills.

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

Dataset Description

2

Exploring the Data

3

Statistics

4

Summary

Dataset Description



US Wind Turbine database as a csv file

- Created a new csv file of the 10 desired fields
- Visual analysis shows that empty string fields in this file are populated with the word, “missing”. However, an analysis of the data in the 10 columns needed comes back with no missing data.
- This study relies on each windfarm having a zip code. Windfarms with geocodes that failed the zip code lookup were assigned a value of ‘99999’.



Zillow Housing Data as a csv file

- Noted constraint of only median-sale price data is available (desired average sales price as well)
- Data initially looks clean, but that is because missing data is populated with NaN values
- Analysis of NaN values shows that the years 2003 – 2013 have very high (over 50%) missing values. This may cause shift in analysis years from 2000 – 2010 to something more recent, i.e. 2014 – 2017



US Zipcode Database as a csv file

- Used each windfarm’s geocode/s to identify the zipcode the windfarm resided in.
- Merged US Zipcode database with Wind Turbine and Zillow Housing databases
- This provided the additional fields of population density and median household income for each zip code.
- The “expanded data file” now consists of windfarm, housing values, and zip code demographic data, keyed by zipcode.



Exploring the Data (1 of 4)

- Out of **1,269** zipcodes w/windfarms in the Windturbine database, only **20** zip codes meet all criteria for this study

Zillow only has median sales price data for 3,797 zipcodes out of 42,000 zipcodes in the US

Out of 1,269 zipcodes w/windfarms only 58 of these zip codes have Zillow housing data

Of these 58 zip codes, selected the top 5 states with the most windfarms (46 zip codes)

Of these 46 zip codes, there are only 20 zip codes twenty-five miles away without a windfarm that have Zillow housing data.

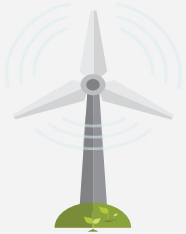
- States with highest and lowest count of zip codes with windfarms...

Highest 5

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

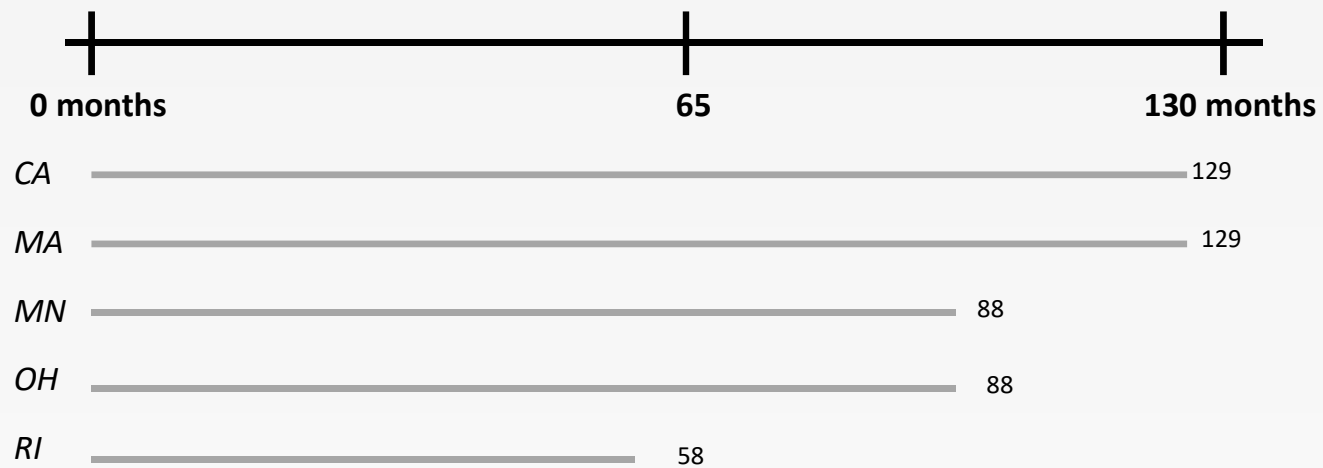
Lowest 5

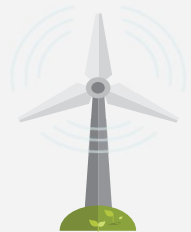
	state	zipcode_count
3	IL	1
6	NY	1
8	OK	1
11	UT	1
12	WA	1



Exploring the Data (2 of 4)

- Out of a possible **130** months (Mar-2008 thru Aug-2018) of housing data, at least **5** states have sufficient housing data...

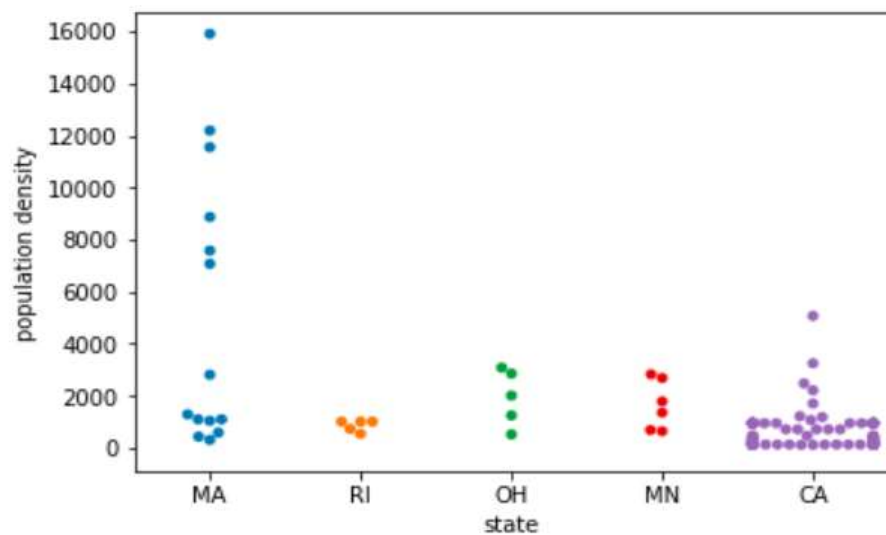




Exploring the Data (3 of 4)

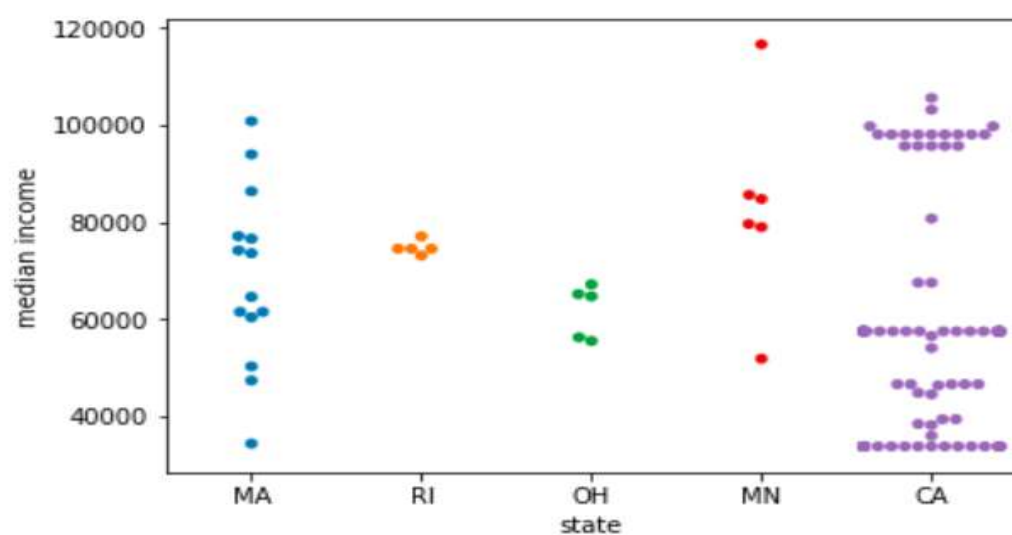
MA has several windfarms in areas w/high population densities

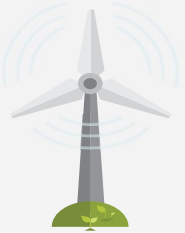
Population Density



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

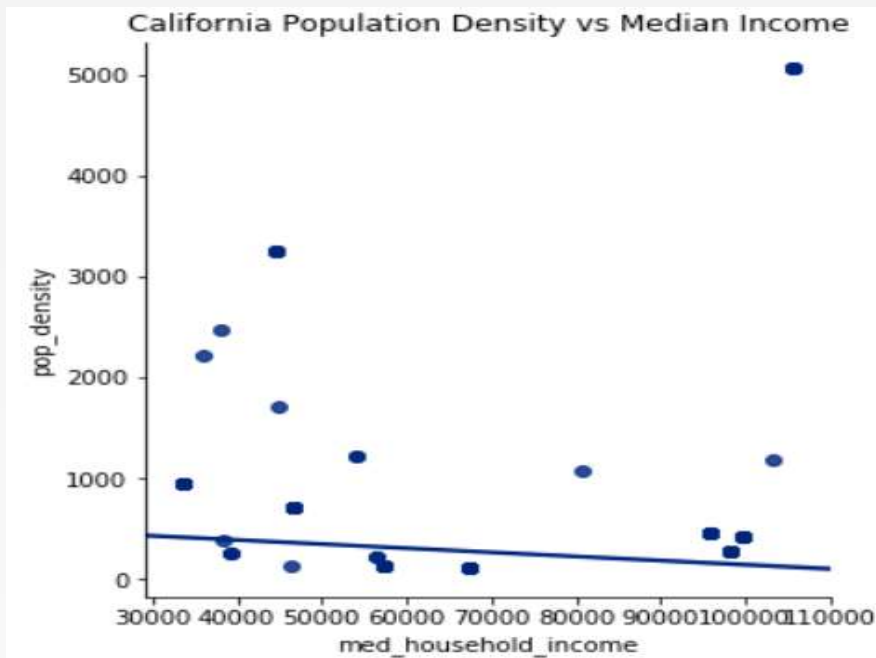
Median Income



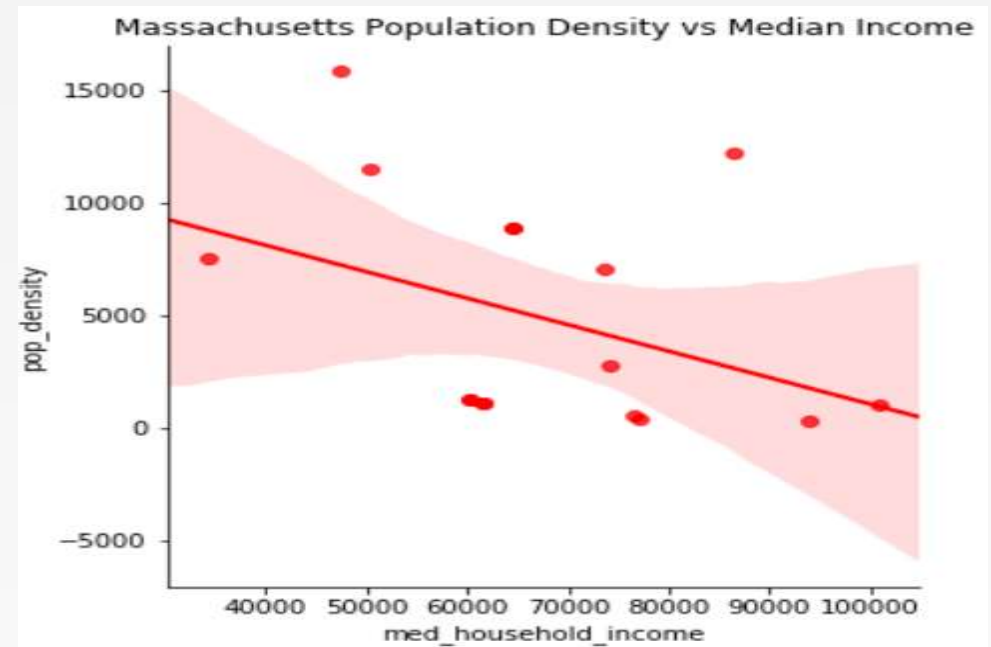


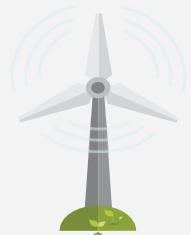
Exploring the Data (4 of 4)

California



Massachusetts

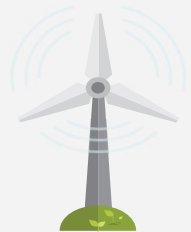




Statistics - Variables

Variables that appear most significant to answer this project's question of...
"Do windfarms impact the value of home prices?"

- *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

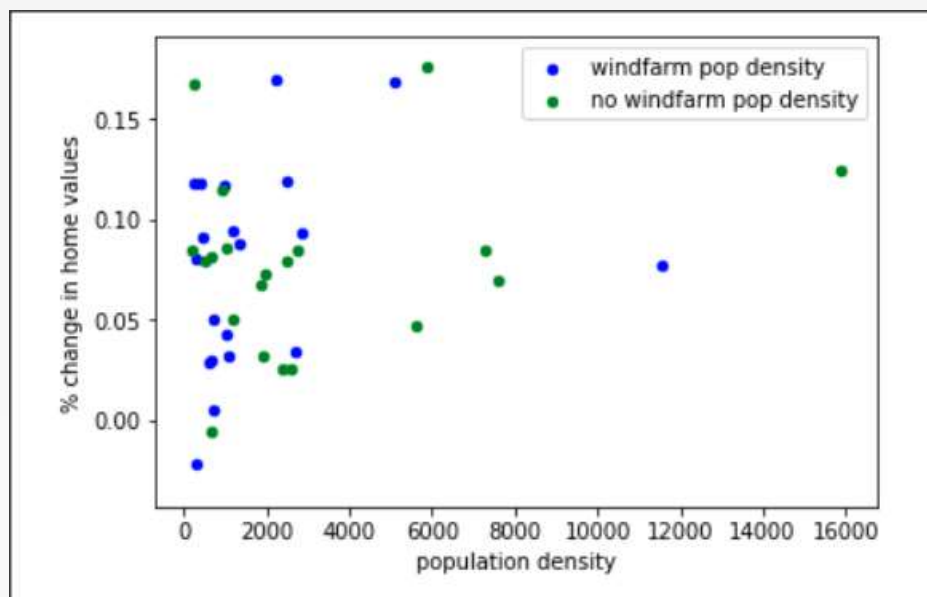


Statistics – Correlations (1 of 2)

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

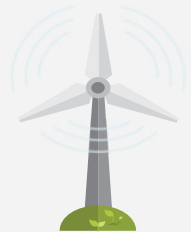
Correlation of zip codes
with windfarms:

23%



Correlation of zip codes
without windfarms:

25%

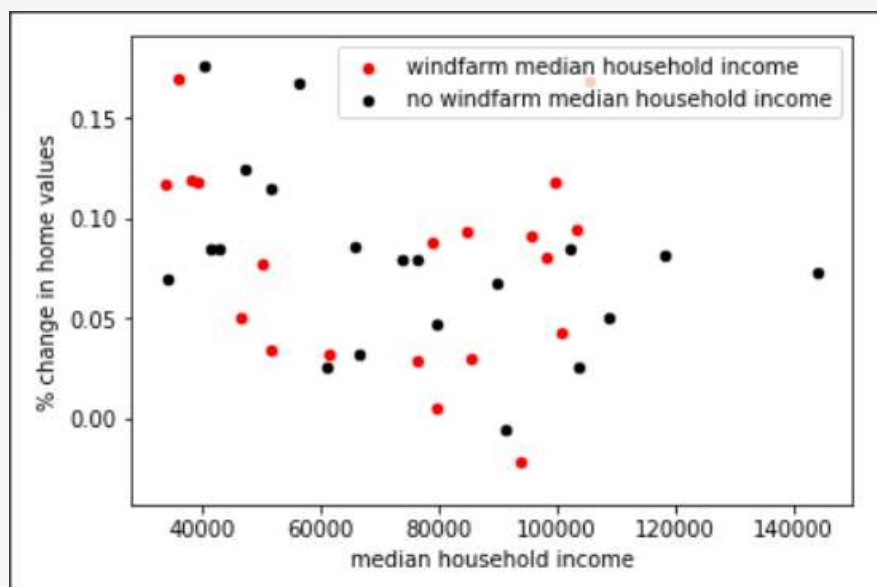


Statistics – Correlations (2 of 2)

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

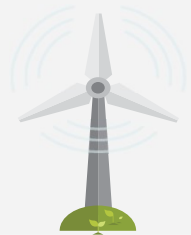
Correlation of zip codes
with windfarms:

19%



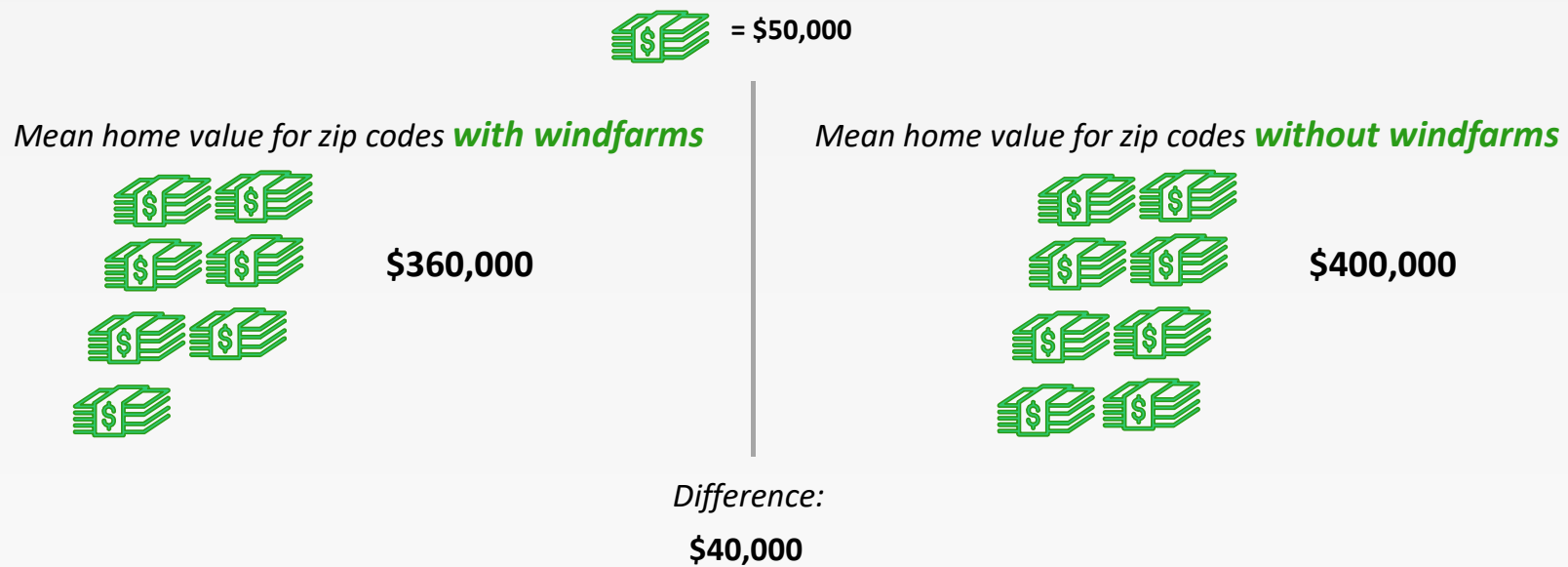
Correlation of zip codes
without windfarms:

41%



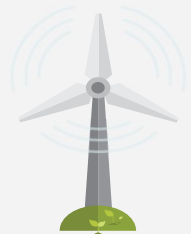
Statistics - Significance

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



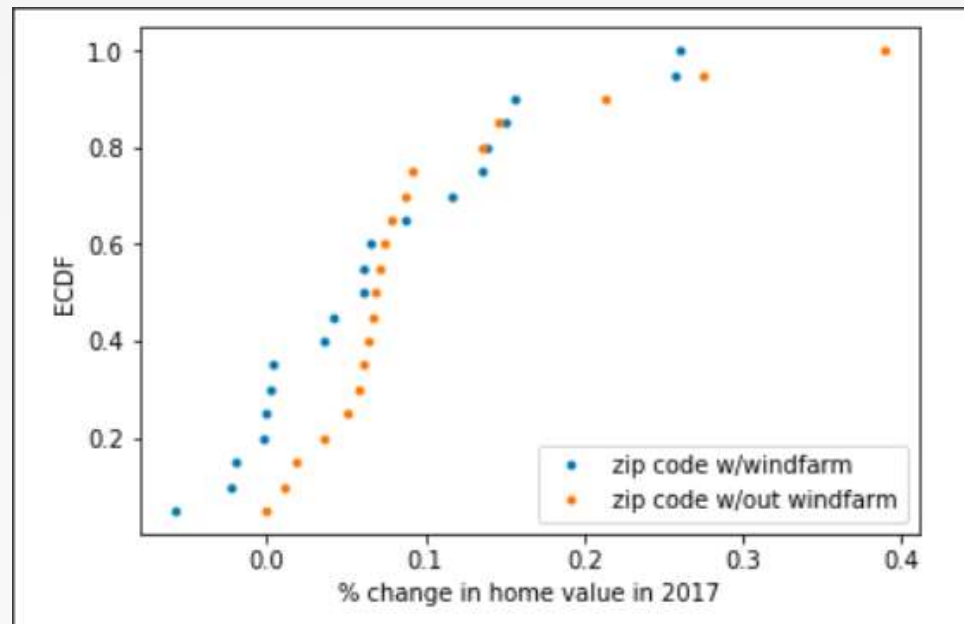
p-Value
0.99

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



Statistics – ECDF's

- Continuing the use of 2017 data and creating ECDF's **reveals that outliers** may cause the previously computed p-value and confidence interval **to be misleading**
- To address this, when applying machine learning algorithms additional years beside 2017 should be used and removing outliers from the training data considered.





Summary

Objective Build an algorithm that can be used to predict the change in value of a home based on proximity to windmills.

Observations:

- Only around 2% (20/1269) of the zip codes with windfarms meet all criteria for this study. Though a small number this is sufficient sample data
- The largest contributor to the drop in eligible zip codes is due to Zillow only having median-sale-price data for 3,297 zip codes out of 42,000 US zip codes
- For the 20 zip codes, there is a very high percentage (75%) of housing data available for each month (2008–2018)
- Population density does not appear to have a high impact on home values
- Not surprisingly...
 - Windfarms are primarily in rural areas, though several zip codes with high populations and high median incomes host windfarms
 - Median income appears to have a significant impact on home values