

Let Business Thrive!

William Baum
Project Luther

Influence of Location on Business Success

“Location, Location, Location.”

–The Realtor’s Mantra

Goals

Predict areas where
businesses will thrive

Identify key location and
workforce characteristics



Market Saturation

Consumer purchasing power

Competition

Pricing

Demand



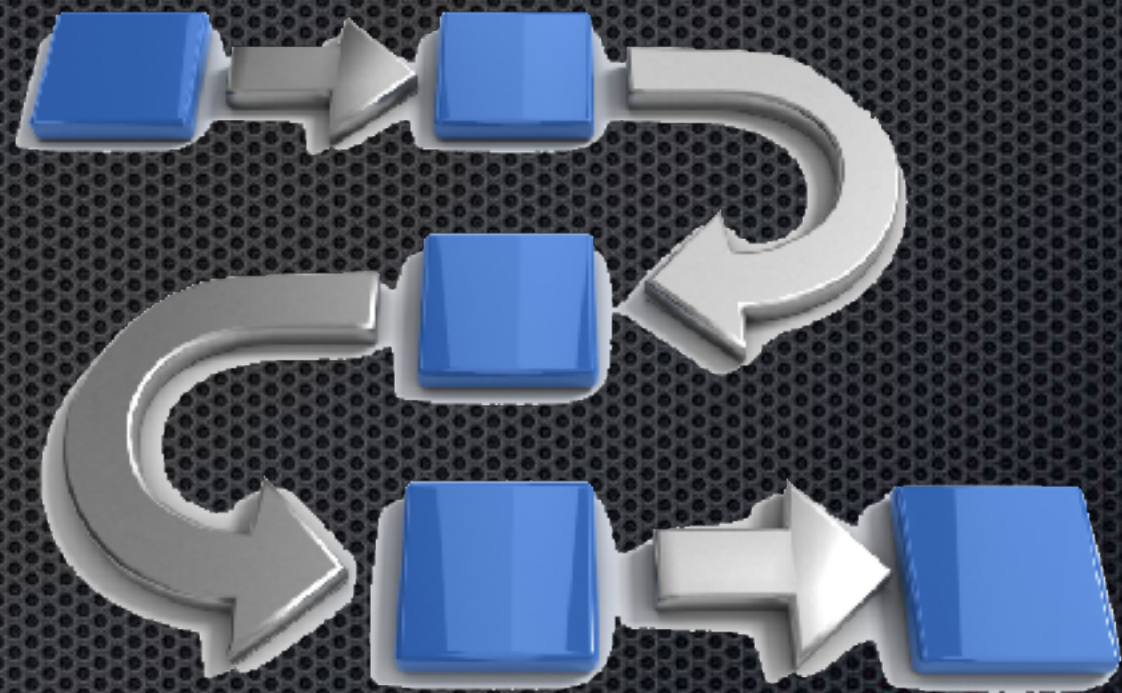
Process

Harvest the Web

Discover Relationships

Feature Development

Build, compare, select models



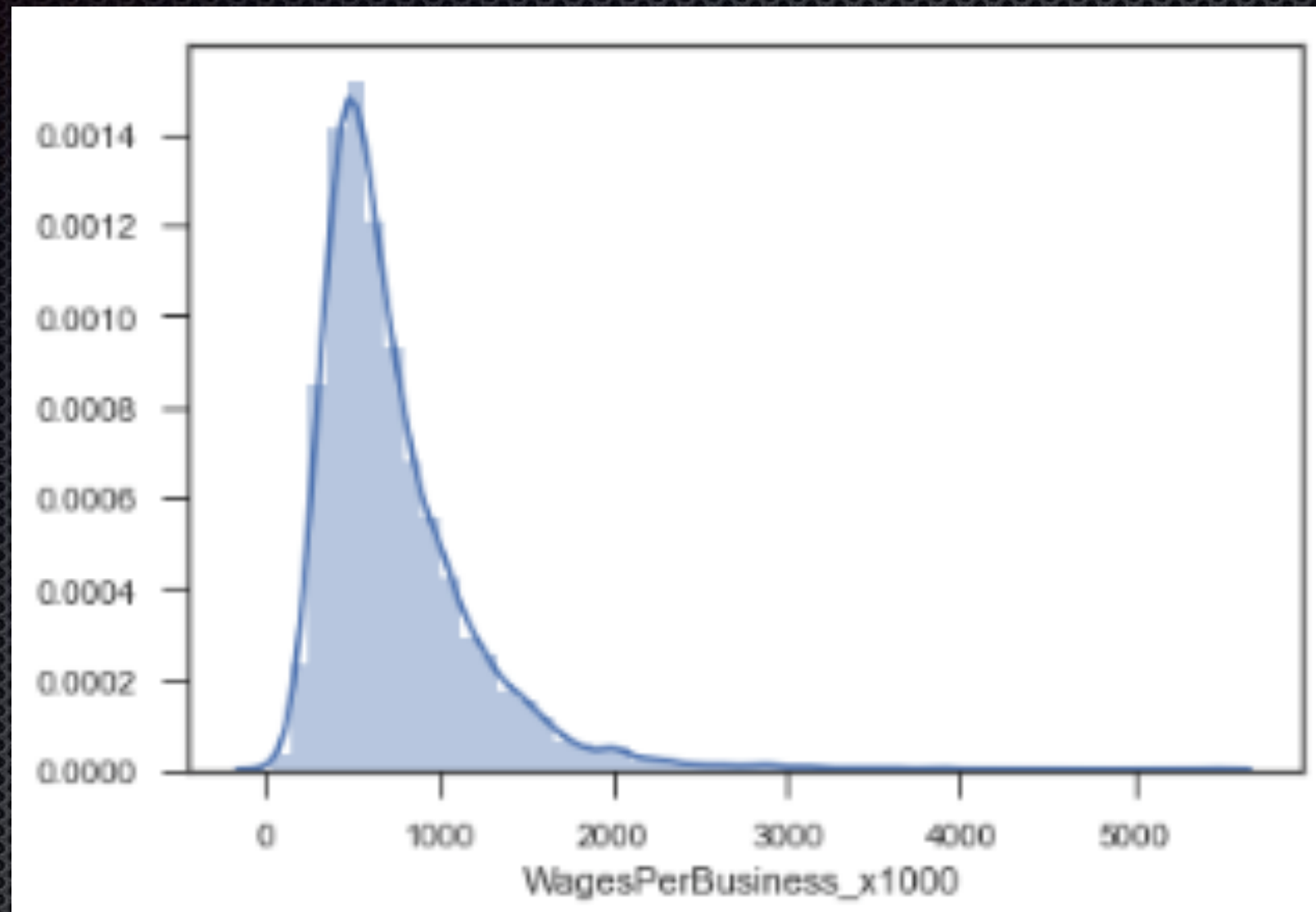
Public Data

- ✦ Business Registrations
- ✦ Tax Returns
- ✦ Demographics
- ✦ Elevation & Location
 - ✦ By Zip Code
 - ✦ Average 90 sq. mi.



Sources

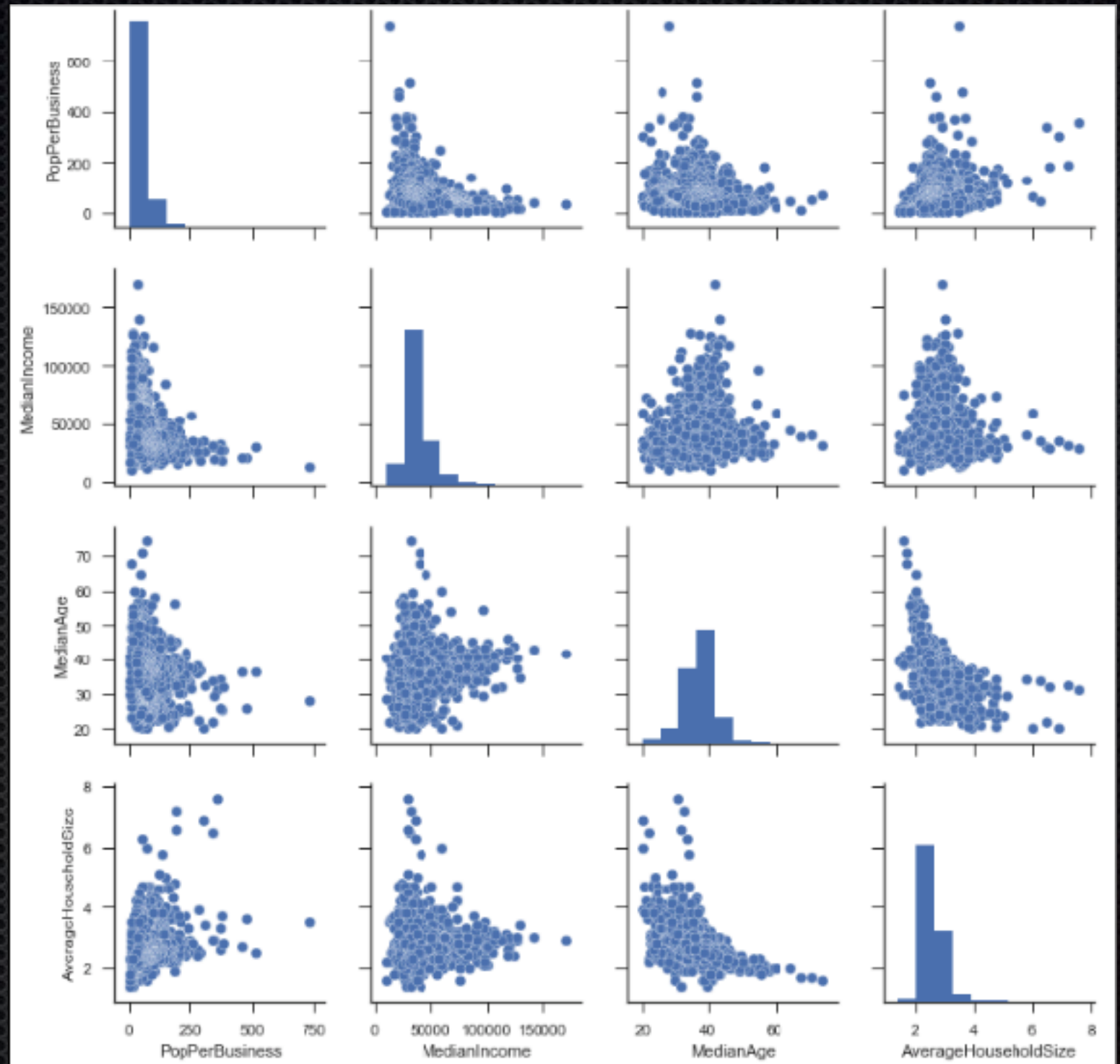
- Websites
 - Beautiful Soup
- National Weather Service
- U.S. Census
 - Annual Reports
 - Individual Taxes



Target Outcome = Wages per Business
(a surrogate outcome measure for Market Share)

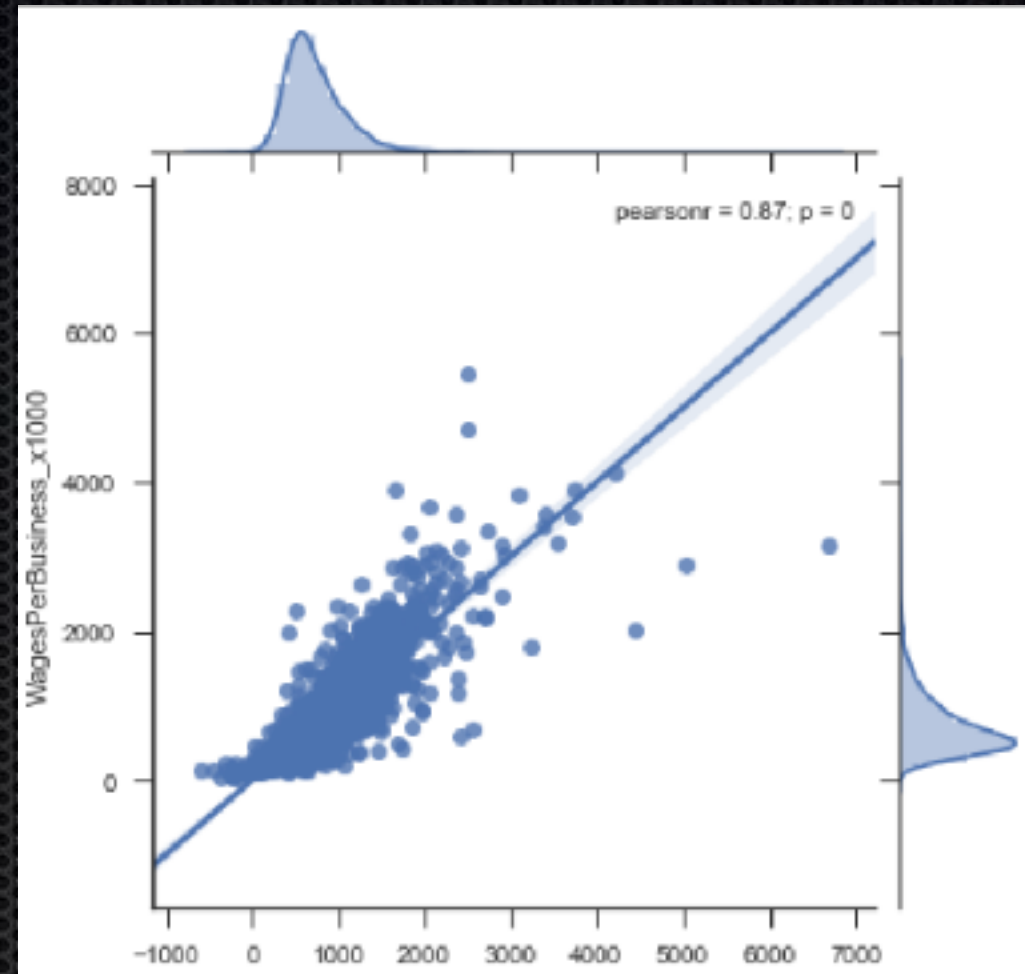
Data Exploration

- ✧ Distributions
 - ✧ Histograms
- ✧ Correlations
 - ✧ Scatterplots

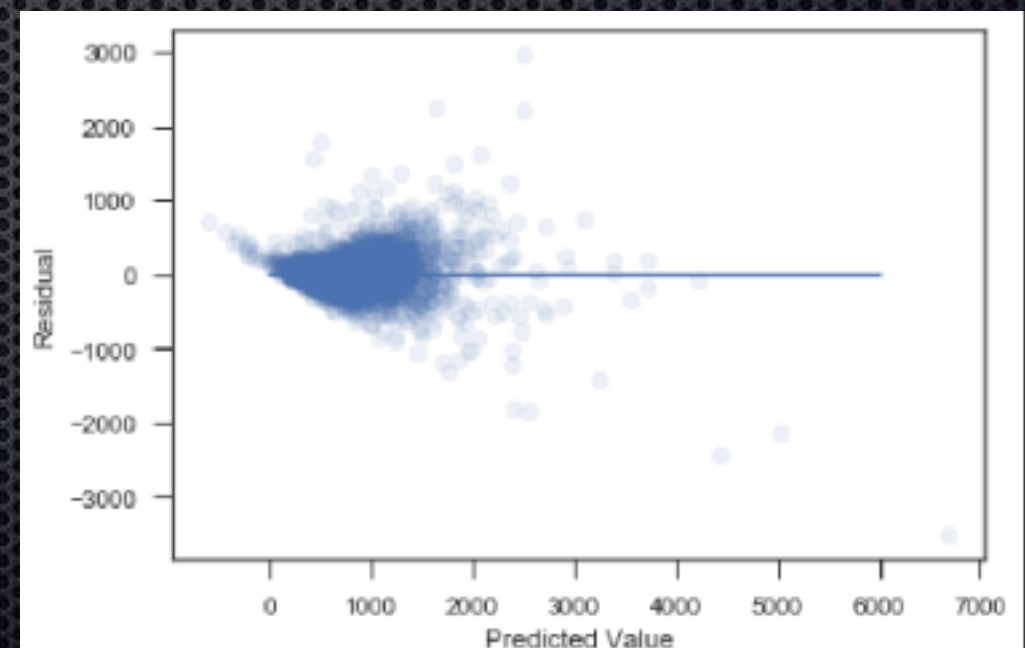


Regression Model

- ✧ OLS
- ✧ Holdback Validation
- ✧ 11 terms
- ✧ Backward Selection
- ✧ Validation / Test scores - 0.76 / 0.74 R-squared
- ✧ RMSE \$221,000 (30%)



Actual by Predicted



Residual by Predicted

Model Evaluations

- ✦ No Interaction or Polynomial terms
- ✦ Regularization did not greatly improve validation scores
- ✦ No Improvement in R-squared by removing outliers
- ✦ Strongest Factors: + Population per Business,
+ Median Income, - Median Age,
- Average Household Size
- ✦ Other Factors: Education, Latitude, Cost of Living Index

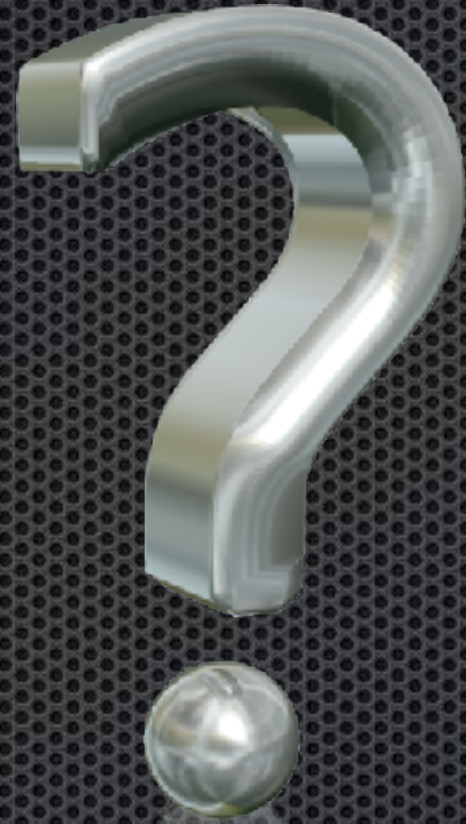
Results

- ✦ Best applied to service industry sectors
- ✦ Highest Locations for Predicted Response
 - ✦ Locations > 10,000 population per zip code
 - ✦ Dallas, TX
 - ✦ Germantown, TN. (Memphis)
 - ✦ Tucson, AZ
- ✦ Interesting outliers
 - ✦ High = small towns with long-established, large companies
 - ✦ Low = small towns with low Cost of Living, Tourism
 - ✦ Teec Nos Pos, AZ - 4 Corners Monument

Next Steps

- ✦ Repeat Analysis by Business Sector
- ✦ Availability of Business Services & Technology
- ✦ Find source of data for Profits
- ✦ Find source of data for Age of business
- ✦ Combine Labor Costs, Cost of Living, Wages/Business
 - ✦ To make a final decision of where to start a new business
 - ✦ One method: Multiple Response Optimization

Questions



References

- ✦ <https://www.thebalance.com/what-location-means-in-real-estate-1798766>
- ✦ https://en.wikipedia.org/wiki/Market_saturation
- ✦ <https://www.dataquest.io/blog/web-scraping-beautifulsoup/>
- ✦ <http://www.gutenberg.org/browse/categories/8>
- ✦ <https://census.gov/data/datasets/2016/econ/cbp/2016-cbp.html>
- ✦ <https://www1.ncdc.noaa.gov/pub/data/normals/1981-2010/>