ZILLOW& MACHINE LEARNING

Contents:

- About me
- Dataset
- Descriptive Analytics
- Summary
- Conclusion
- How we can help improve the model

ABOUT ME

- Researcher - Construction Journal - interview and compile construction updates
- Executive Chef - data-driven goals for $4 million food outlets
- Journalist - print and broadcast features

LINKED IN

Scan the QR code or visit LinkedIn.com/in/marcy-misner
to connect, or leave feedback.

Rurchased: $35K
Reno $22K
Sold: $99K

Purchased: $70K
Reno:$65K
Sold: $247K

Two of my renos

ON GOOGLE, MORE PEOPLE ENTER THE SEARCH TERM "ZILLOW" THAN "REAL ESTATE"

29% OF REAL ESTATE WEB TRAFFIC IS ON ZILLOW

80% OF US HOMES SEARCHED

REGARDLESS OF LISTING STATUS

LARGEST PURSE EVER OFFERED FOR ML COMPETITION

May 2017 - Zillow announces $1.2 million competition

3,800 teams from 91 countries eliminated

Jan. 2019 - $1.2 million awarded

HOMEBUYING PROCESS SHOULD HAVE ALLOWED ZILLOW TO CLEAN UP

Feb. 2020 - Zillow kicks off iBuying houses

June 2020 - COVID

Nov. 2020 - Zillow closes the home-buying venture

Q4 Zillow loses ~$500 million, stuck with 7,000 homes

.

7,000 HOMES

How did Zillow end up in the doghouse?

- The ML algorithm failed to adjust to external changes
- Bots overpaid for houses
- Material and labor shortage caused backlog.
- Zillow was aggressively purchasing when market dipped. KeyBanc analyst Edward Yruma found 2/3 of the homes Zillow listed for sale display an asking price ~ 4.5% below what Zillow paid.
- “Zillow may have leaned into home acquisition at the wrong time,” he said.

KAGGLE & CENSUS DATA

1,460 HOMES IN AMES, IOWA

80 columns - 79 characteristics (independent variables) and the sale price of each home.

THE VARIABLES

37 independent variables were numeric.
42 categorical , ordinal and were recoded.

PYTHON, R, TABLEAU, CANVA, TRELLO, EXCEL

Exploratory Factor Analysis, Data Cleaning and Wrangling, Data Visualization, Feature Importance, Splitting data into training/testing sets.

The Dataset

OverallQual: Rates the overall material and finish of the house
OverallCond: Rates the overall condition of the house
1- 10 Very Poor to Very Excellent

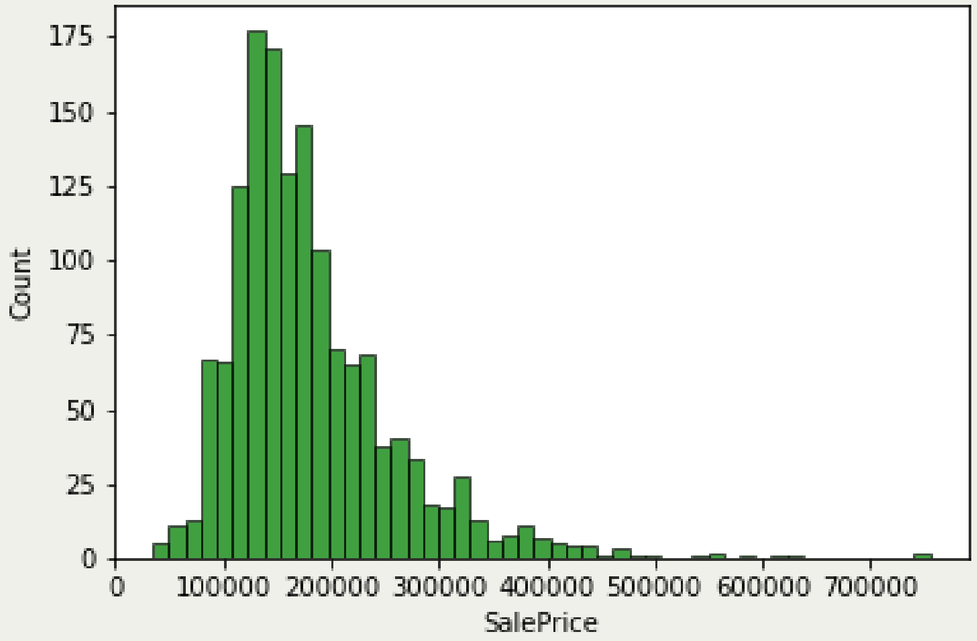
GrLiveArea: SF of house above grade
OpenPorchSF: Open porch area in square feet SF

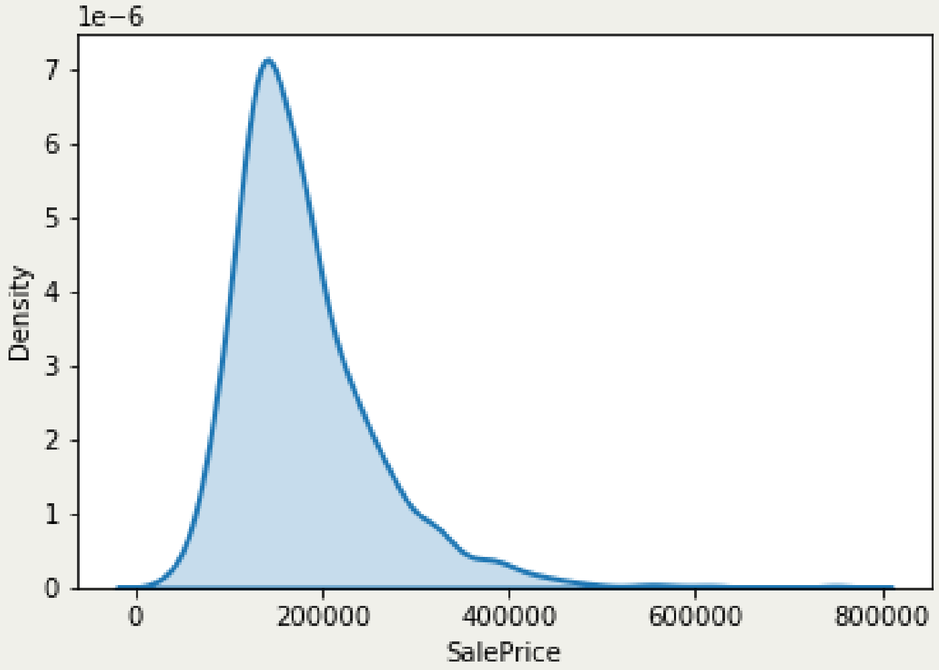
Exterior1st: Exterior covering on house: Asbestos Shingles, Asphalt Shingles, Brick Common, Brick Face, Cinder Block, Cement Board, hard Board, Imitation Stucco, Metal Siding, Other, Plywood, ,Precast, Stone, Stucco, Vinyl Siding, Wood Siding, Wood Shingles.

DESCRIPTIVE ANALYTICS

Ranges make it easier to assess distribution of home prices

25000-50000	5
50000-100000	109
100000-150000	492
150000-200000	406
200000-300000	312
300000-400000	87
400000-500000	19
501000	9

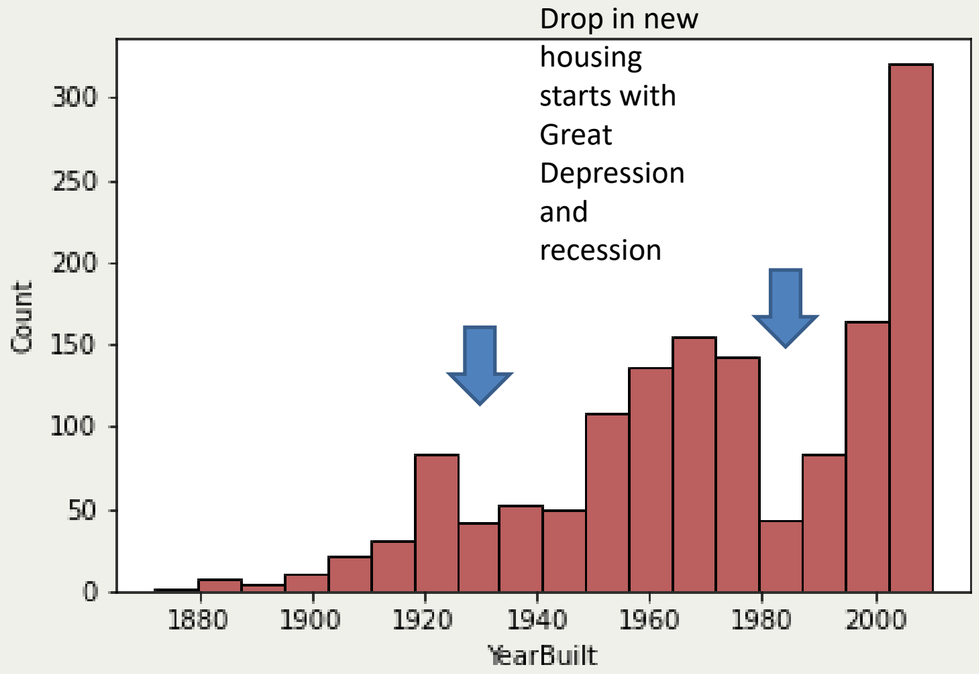


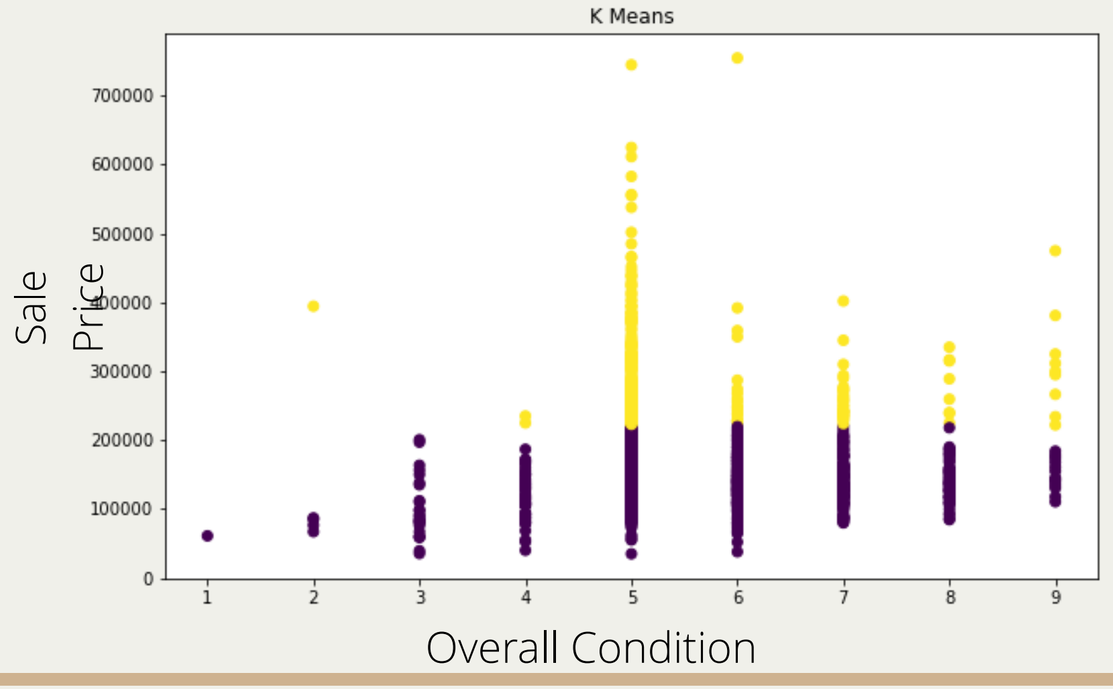


Year Built

DESCRIPTIVE ANALYTICS

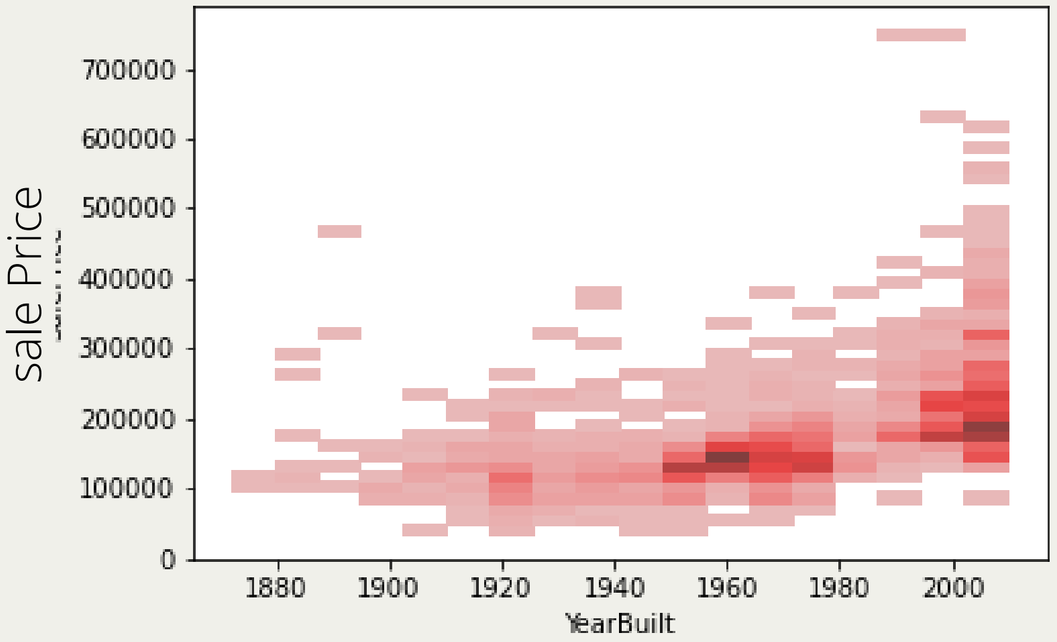
The plot below is really interesting to me because it represents opportunity!
The houses on the lower end of overall condition score are selling at lower prices. Given a subsequent graph where remodeled homes sell for higher prices, identifying these homes as investment opportunities would be KEY.

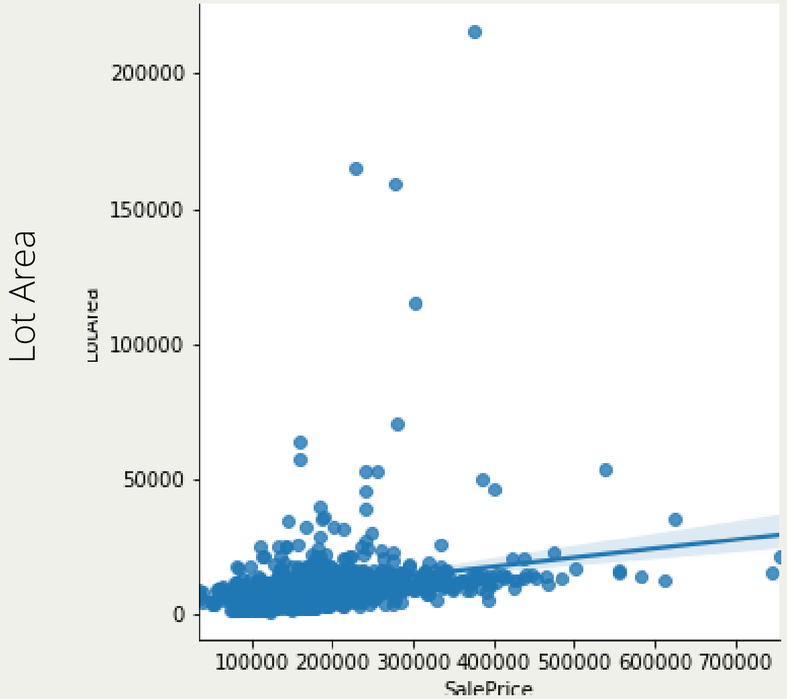




DESCRIPTIVE ANALYTICS

During a recession in the early 1980s, it's easy to see there were few houses being built. Conversely, during a boom after 2000, new housing starts took off.

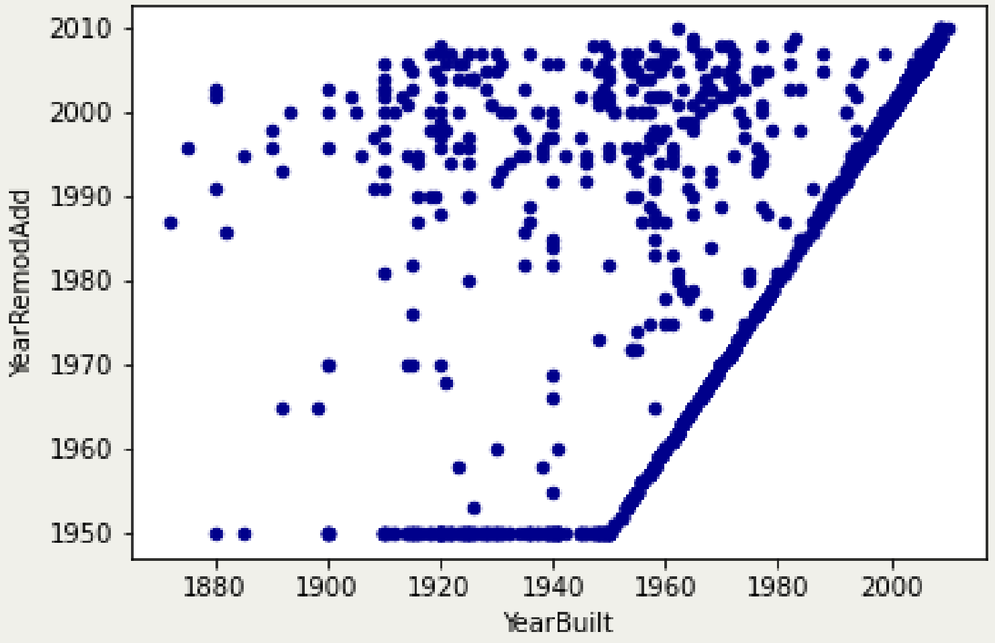


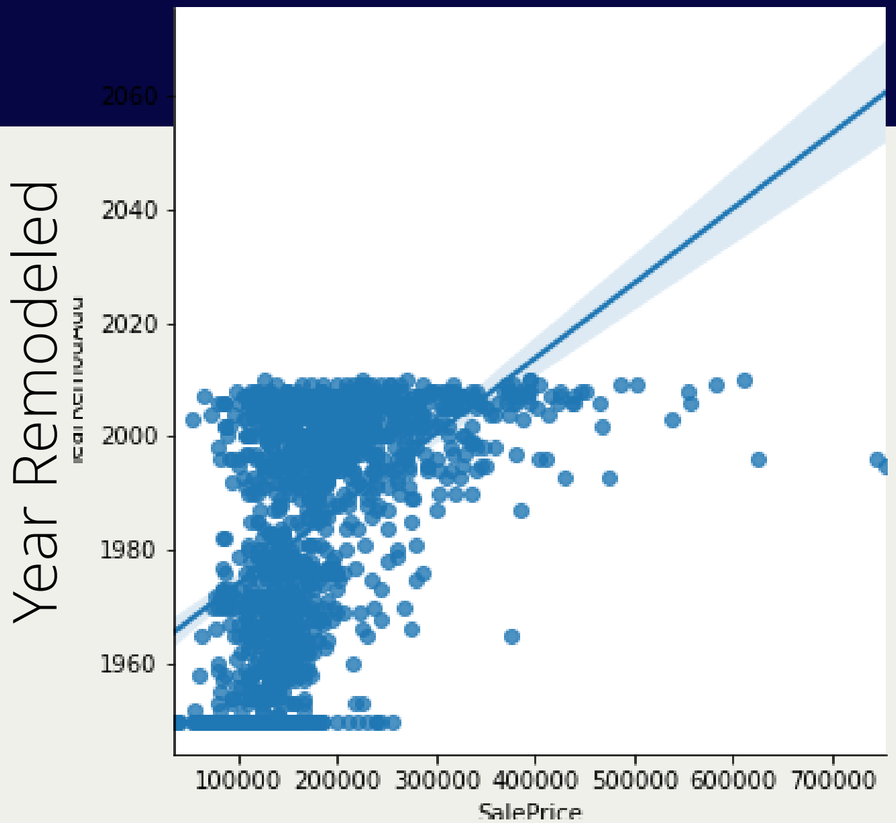


A large yard doesn’t equate to higher selling price, aside from a few outliers.

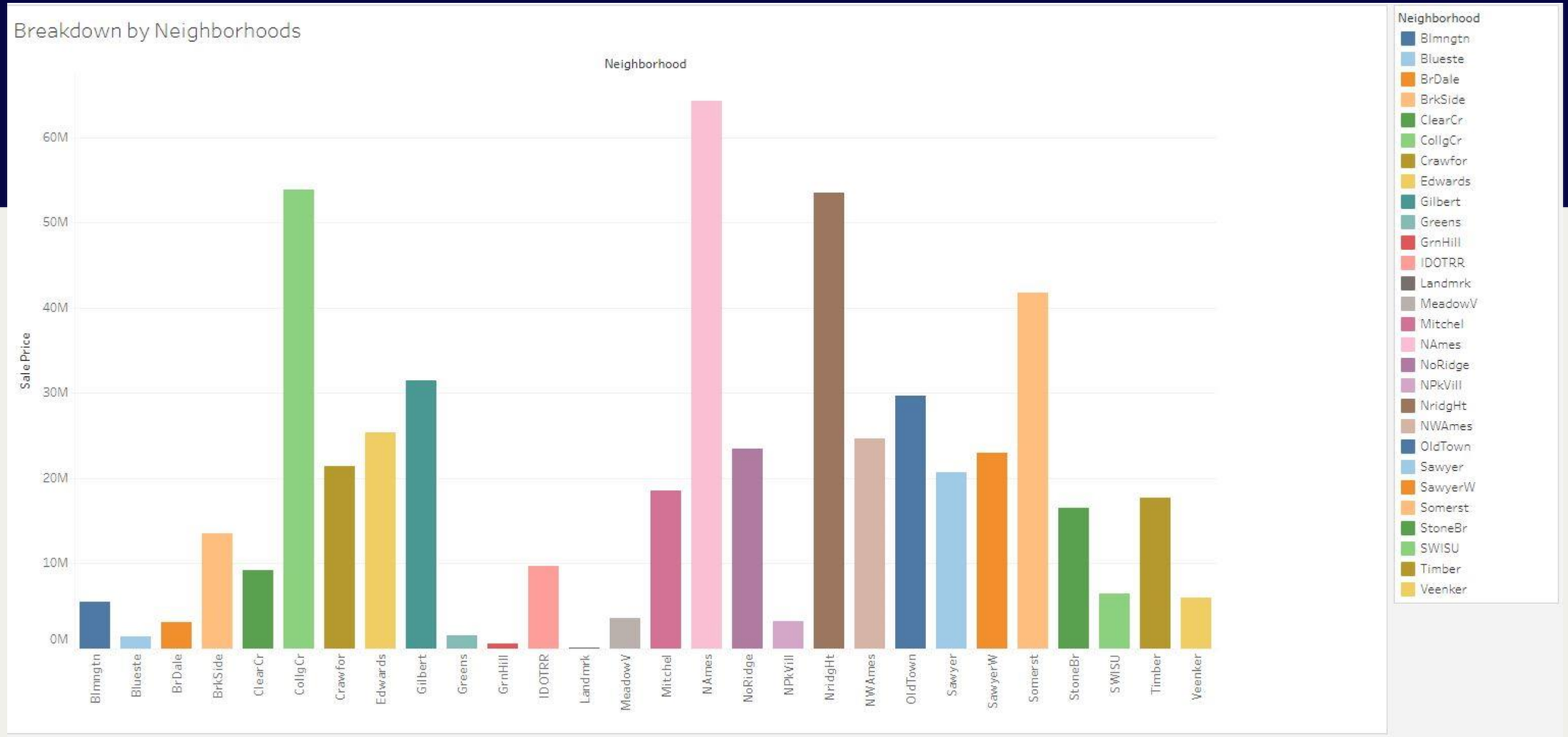
DESCRIPTIVE ANALYTICS

Homes here are remodeled an average of 13.5 years after they're built. .

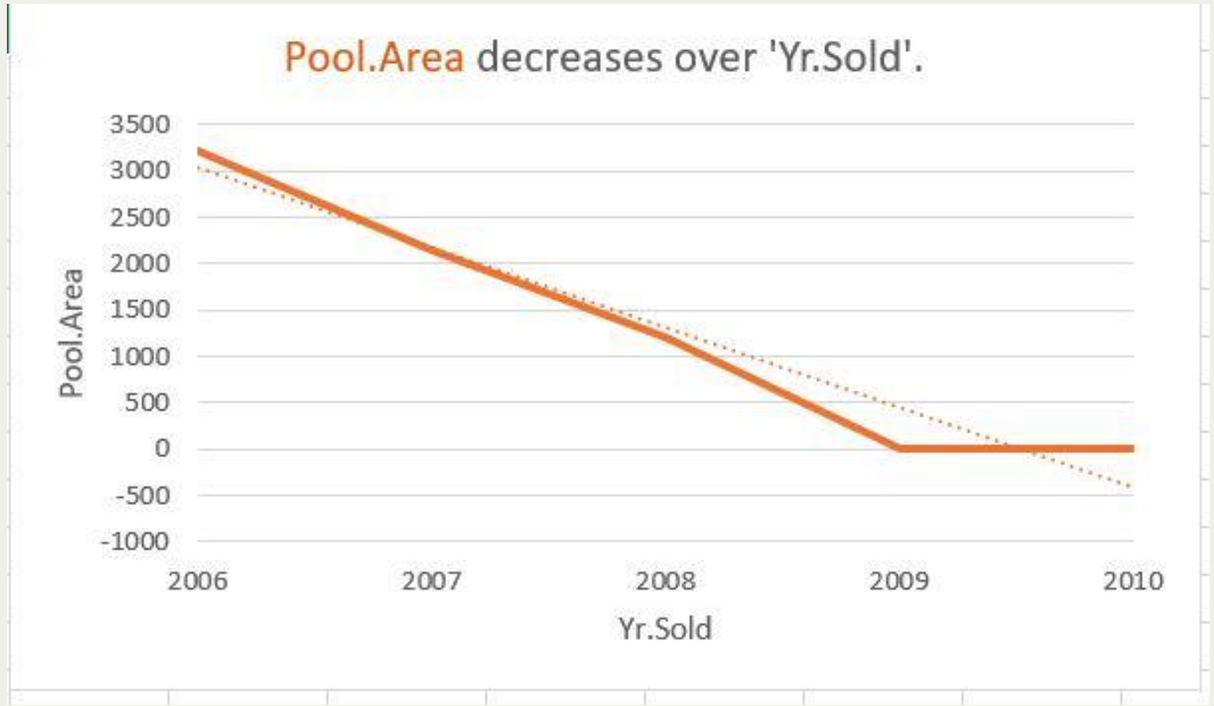


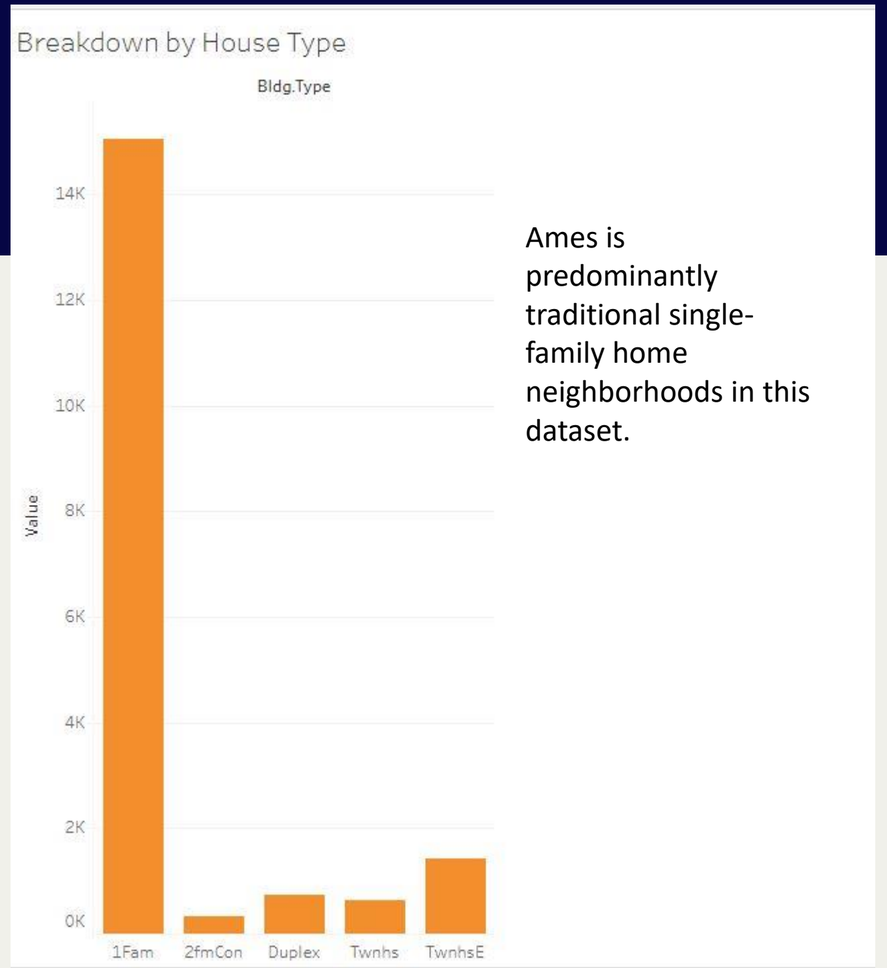


A peek into the data with Tableau. .



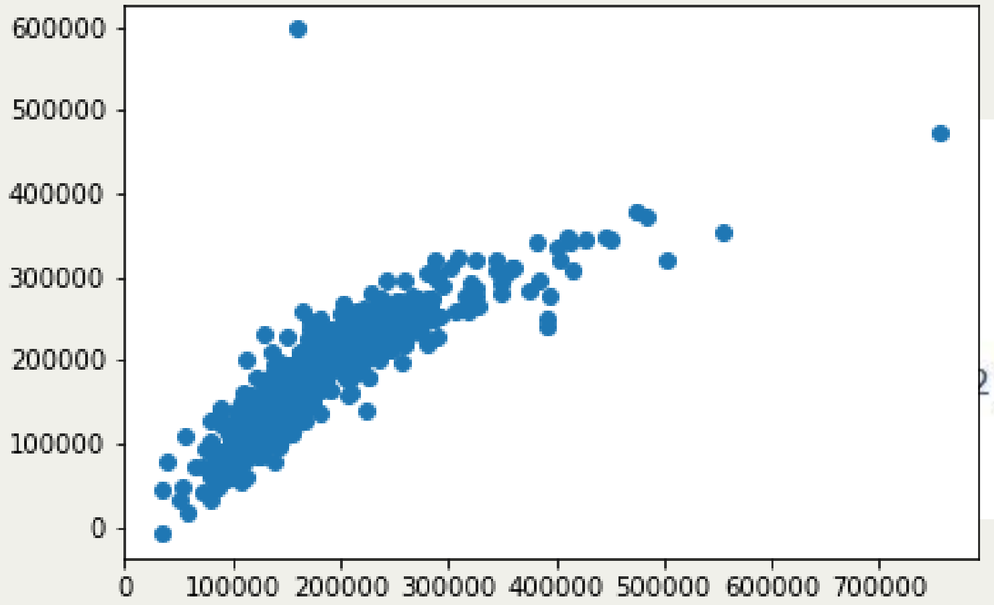
A peek into the data with Tableau & Excel. .





Ames is predominantly traditional single-family home neighborhoods in this dataset.

OVERALL ACCURACY OF THE MACHINE LEARNING MODELS



Another way to gauge accuracy

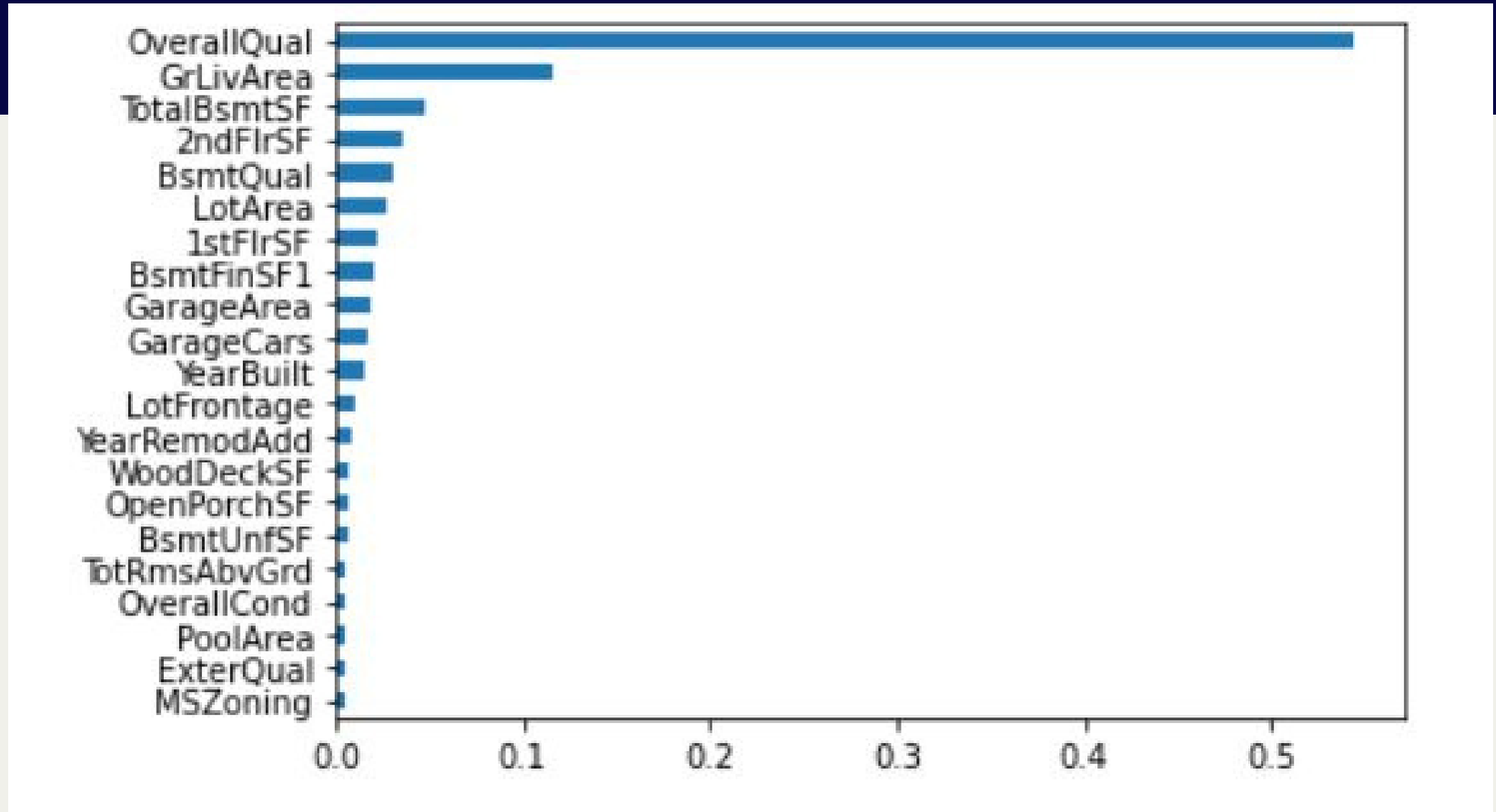
is to print an accuracy score for this model.

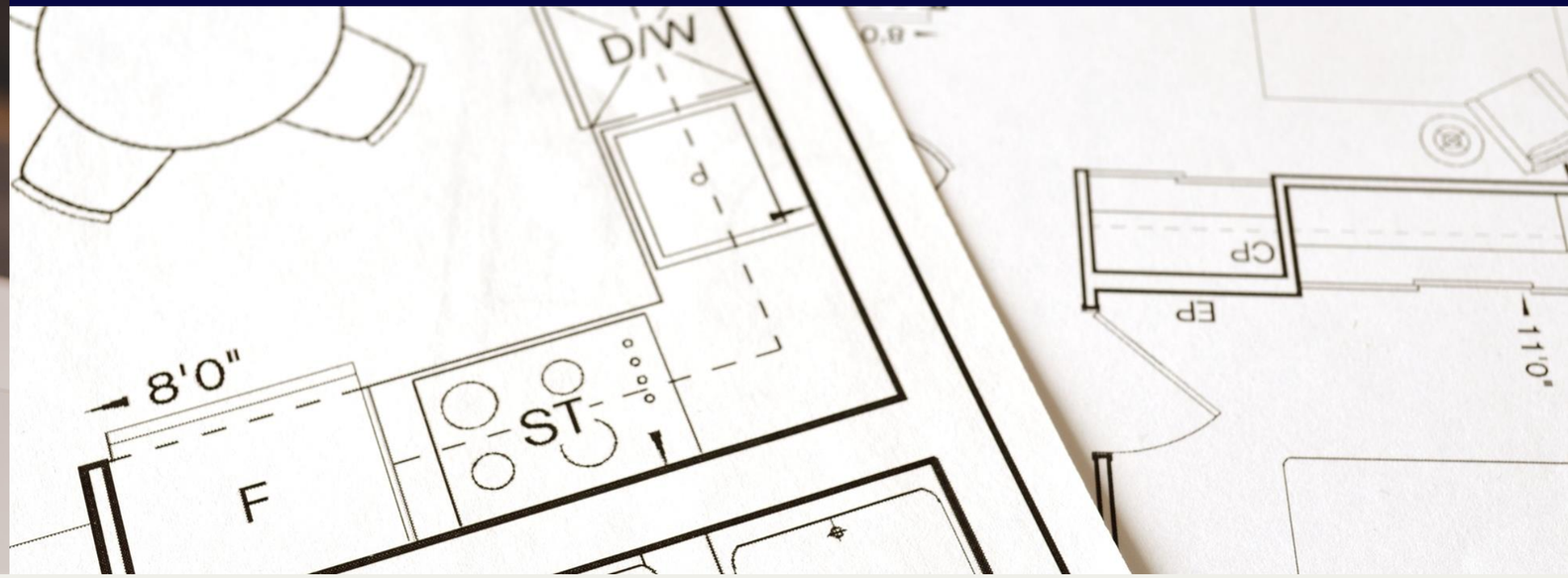
```
print("Score:", lm.score(x_test, y_test))
```

```
Score: 0.7423243890145809
```

TOP 25 FEATURES

After doublechecking with Feature Importance and K Nearest Neighbors, I'm comfortable identifying top features. Out of 61 columns, these are the top 25 most important features tied to sale price of a home. Overall quality and overall condition are similar, as are garage features.





TO SEE THE CODE

https://github.com/M-arcy/Feb2022FinalProject

WHAT CAN WE ADD TO THESE OBSERVATIONS

unemployment numbers - lenders want stable jobs

defaults on mortgages

price & availability of lumber and building supplies

number of contractors who are hiring

US CENSUS BUREAU

WHEN CONTRACTORS WORK

CONSTRUCTION WORKERS EMPLOYED - TRACKING YEAR OVER YEAR, Q1, Q2, Q3, Q4 FOR HIRING TRENDS

Private Primary Jobs for All Workers by NAICS Industry Sector in 2019

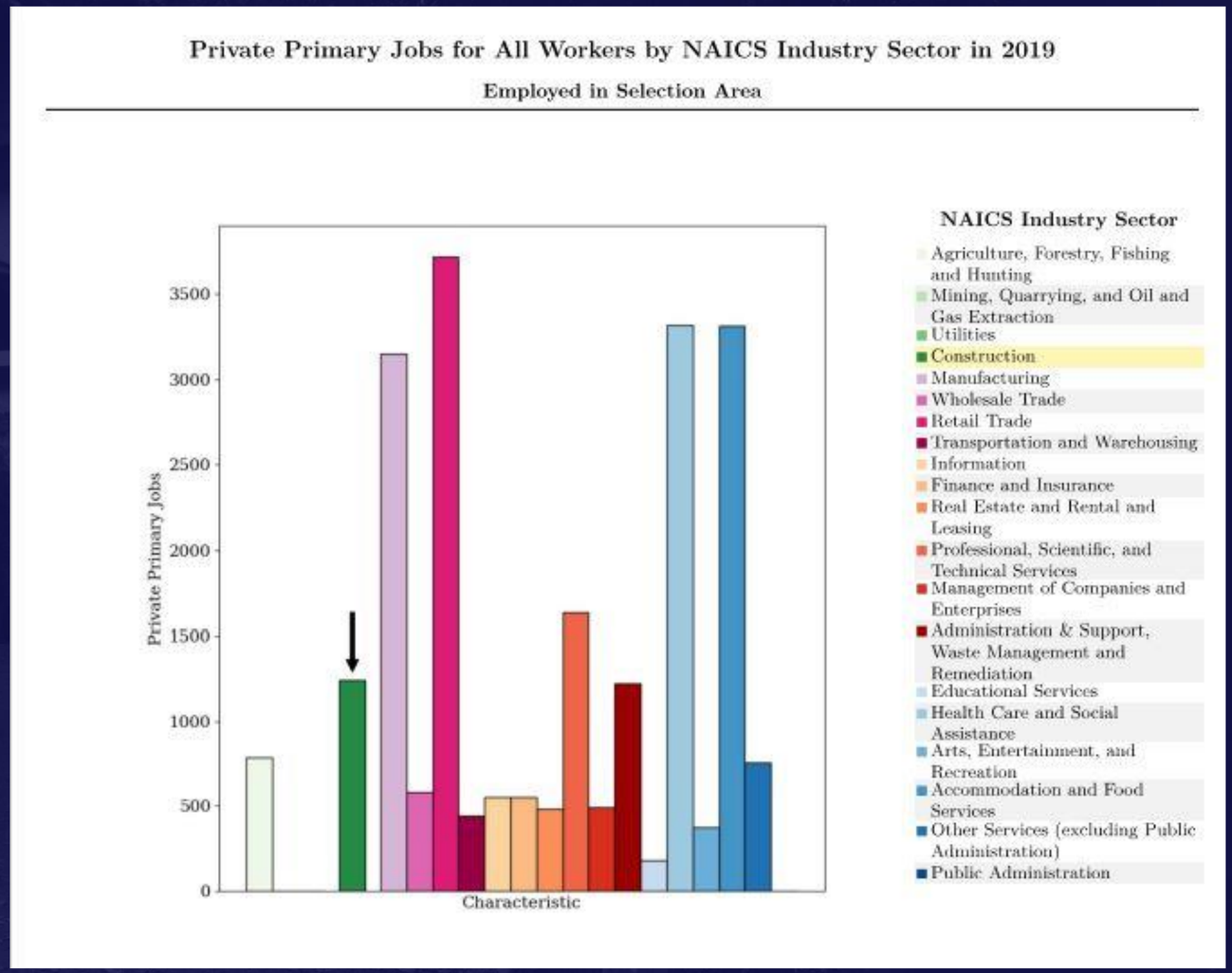
Employed in Selection Area

NAICS Industry Sector	2019 Count	2019 Share
Total Private Primary Jobs	22,787	100.0
Agriculture, Forestry, Fishing and Hunting	782	3.4
Mining, Quarrying, and Oil and Gas Extraction	0	0.0
Utilities	0	0.0
Construction	1,243	5.5
Manufacturing	3,148	13.8
Wholesale Trade	581	2.5
Retail Trade	3,714	16.3
Transportation and Warehousing	441	1.9
Information	553	2.4
Finance and Insurance	549	2.4
Real Estate and Rental and Leasing	480	2.1

Page 2 of 4

United States Census Bureau

NAICS Industry Sector	2019 Count	2019 Share
Professional, Scientific, and Technical Services	1,636	7.2
Management of Companies and Enterprises	490	2.2
Administration & Support, Waste Management and Remediation	1,221	5.4
Educational Services	190	0.8
Health Care and Social Assistance	3,314	14.5
Arts, Entertainment, and Recreation	383	1.7
Accommodation and Food Services	3,310	14.5
Other Services (excluding Public Administration)	752	3.3
Public Administration	0	0.0

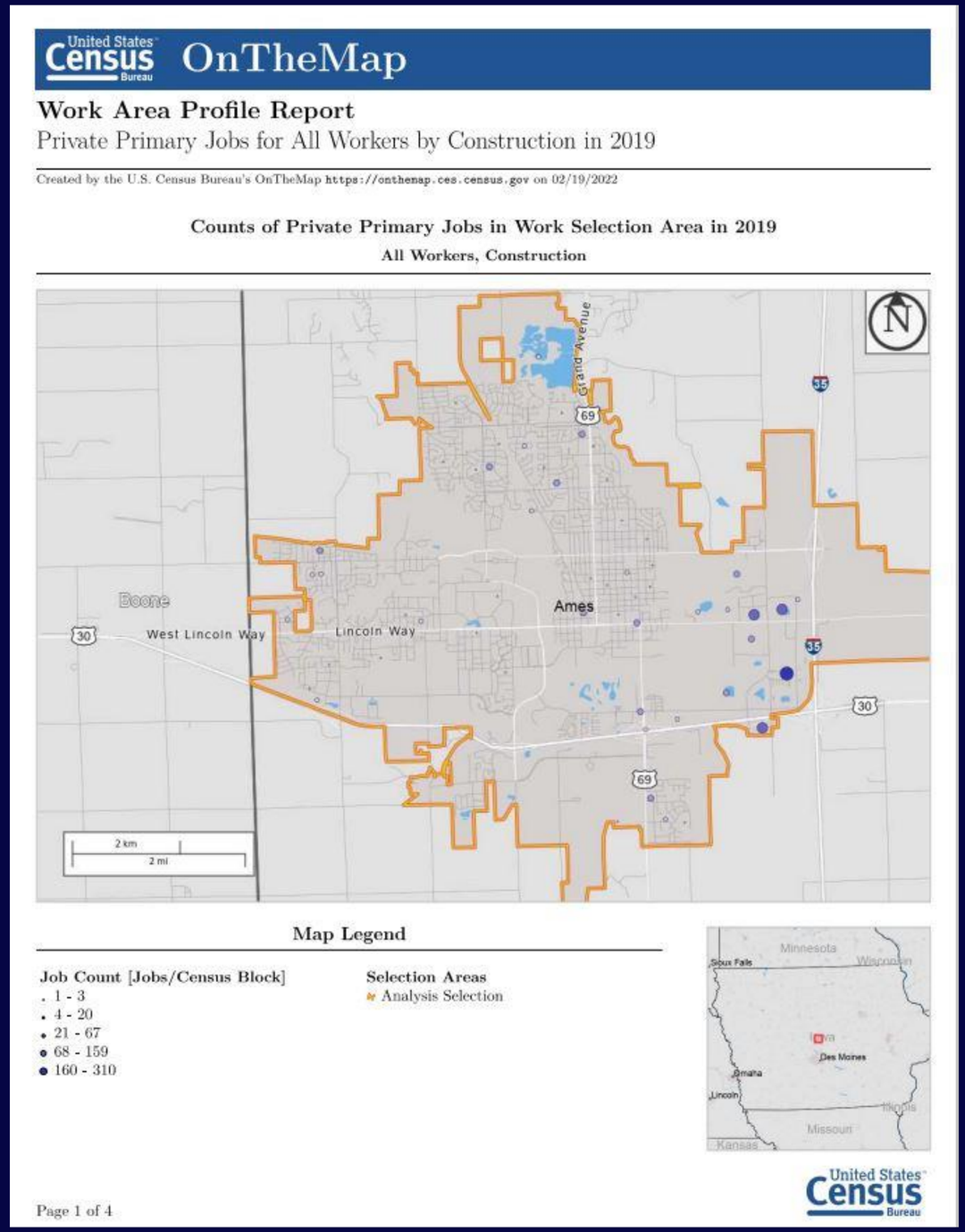


US CENSUS BUREAU

CENSUS.GOV

Leverage government data tools

WORK SMARTER, NOT HARDER



United States Census Bureau OnTheMap

Work Area Profile Report

Private Primary Jobs for All Workers by Construction in 2019

Created by the U.S. Census Bureau's OnTheMap https://onthemap.ces.census.gov on 02/19/2022

Counts of Private Primary Jobs in Work Selection Area in 2019

All Workers, Construction

Map Legend

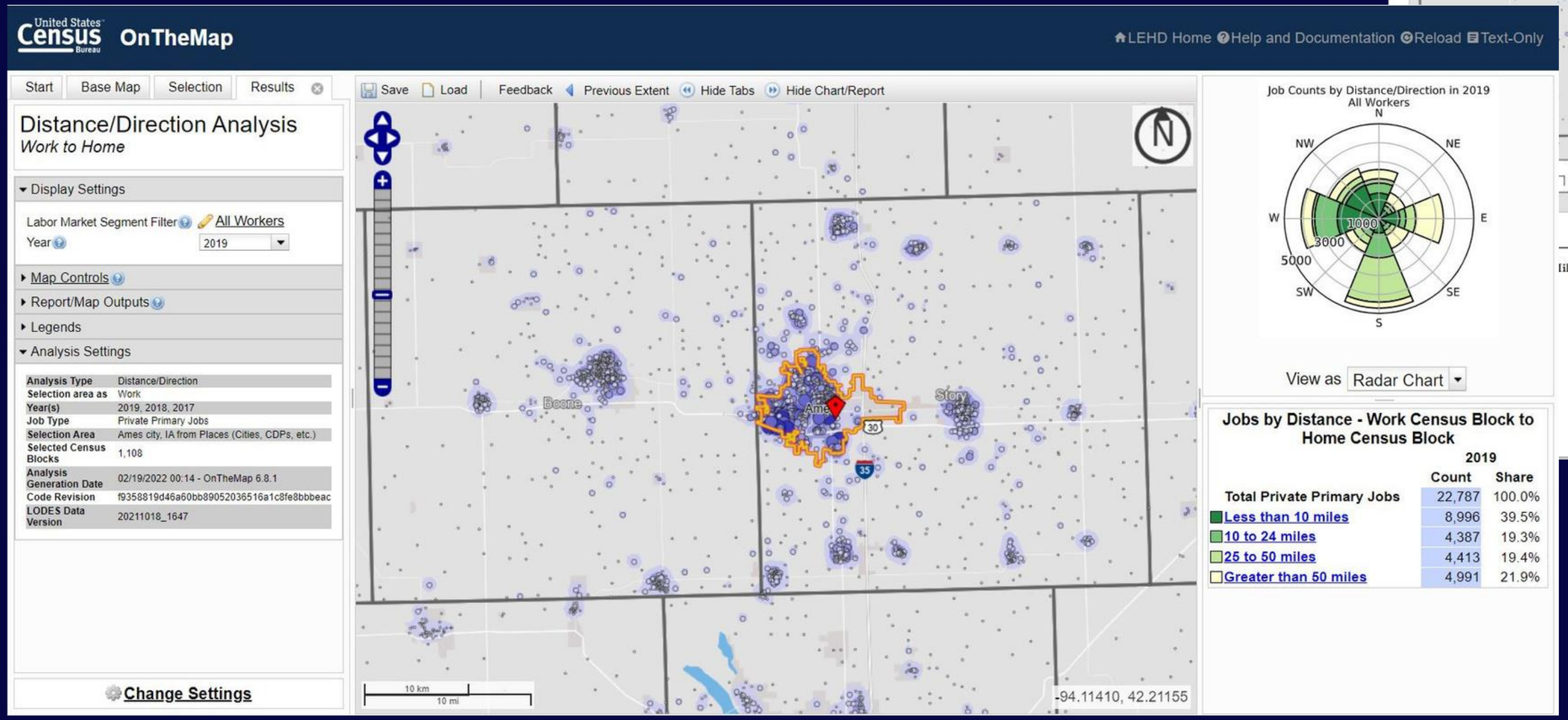
Job Count [Jobs/Census Block]
- 1 - 3
- 4 - 20
- 21 - 67
- 68 - 159
- 160 - 310

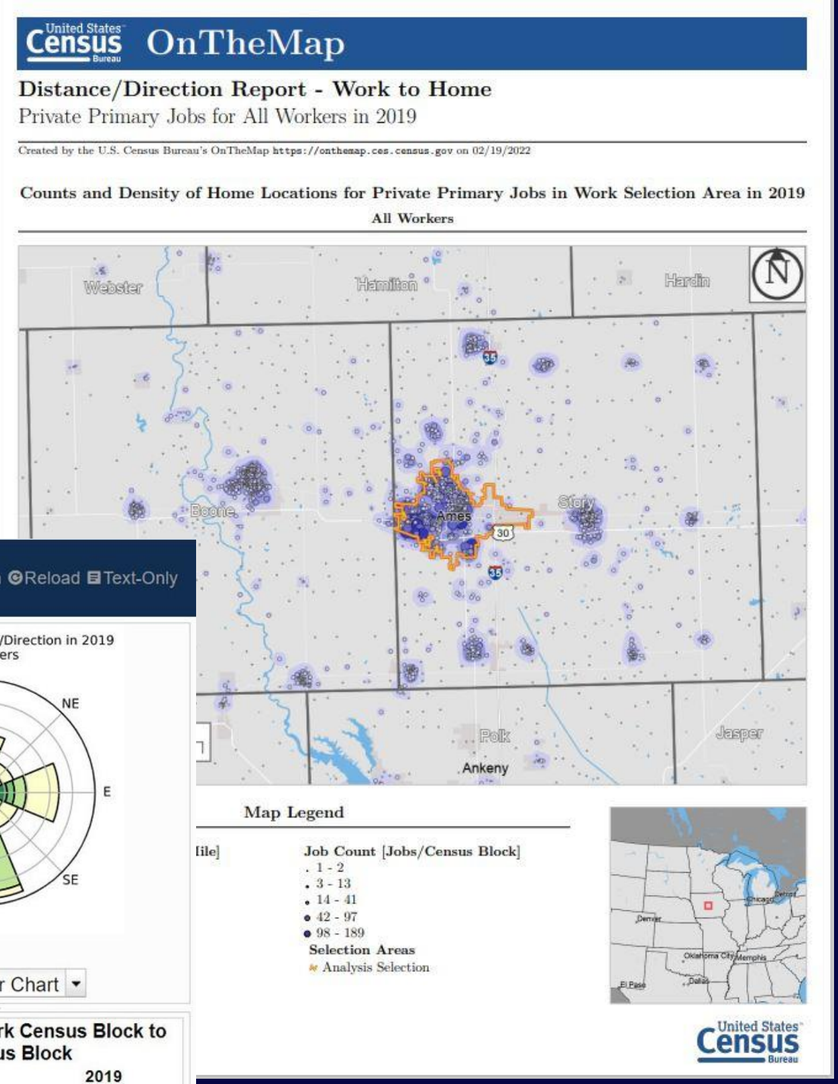
Selection Areas
- Analysis Selection

Page 1 of 4

United States Census Bureau

This analysis shows those who work in Ames, Iowa, and where they live, with a radar graph of which direction most of them travel from.





US CENSUS BUREAU MAY OFFER NEW RESOURCE

MODERNIZING CENSUS CONSTRUCTION INDICATORS

AIDEN SMITH, ASST DIVISION CHIEF, CONSTRUCTION INDICATOR PROGRAMS

Redesigning Building Permits Survey - alternative data sources
Survey of Construction (SOC) - using satellite imagery and machine learning to measure state of construction

US CENSUS BUREAU

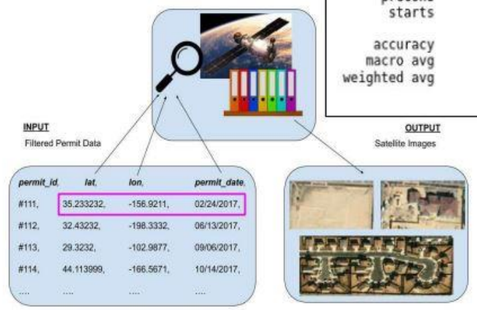
CENSUS.GOV

Leverage government data tools

POC Model – IC1

- **Images**
  - Auto Labeling based on Permit date
  - 3000 total
  - Fixed crop size
- **Methodology**
  - Image Classification FastAI CNN Training (ic1)
  - Results

	precision	recall	f1-score	support
completions	0.93	0.99	0.96	346
precons	0.93	0.92	0.92	273
starts	0.89	0.80	0.84	196
accuracy			0.92	815
macro avg	0.92	0.90	0.91	815
weighted avg	0.92	0.92	0.92	815



Pre-constructions
Ground untouched and no major delimitations or excavations.

3 months before permit authorization date or earlier

Construction Starts
Visible excavation or foundation

Between 1 month before permit authorization date and 4 months after it.

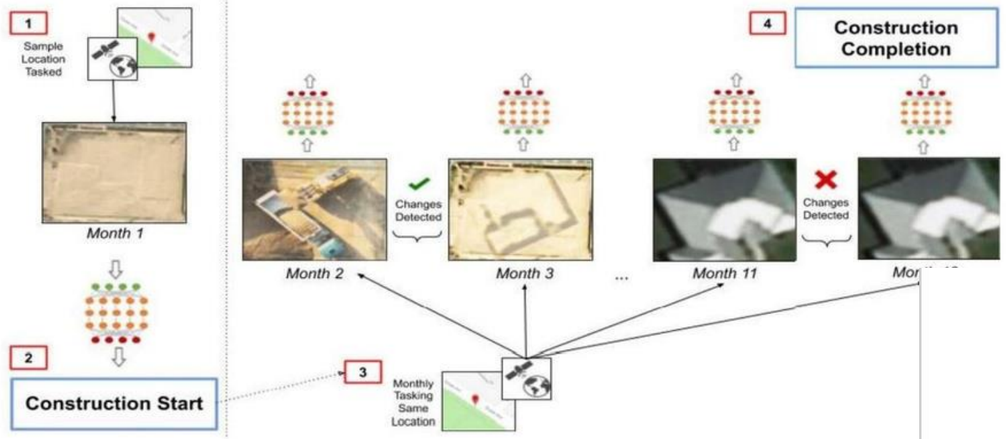
Construction Completions
Completed roof

9 months after permit authorization date.

United States Census Bureau

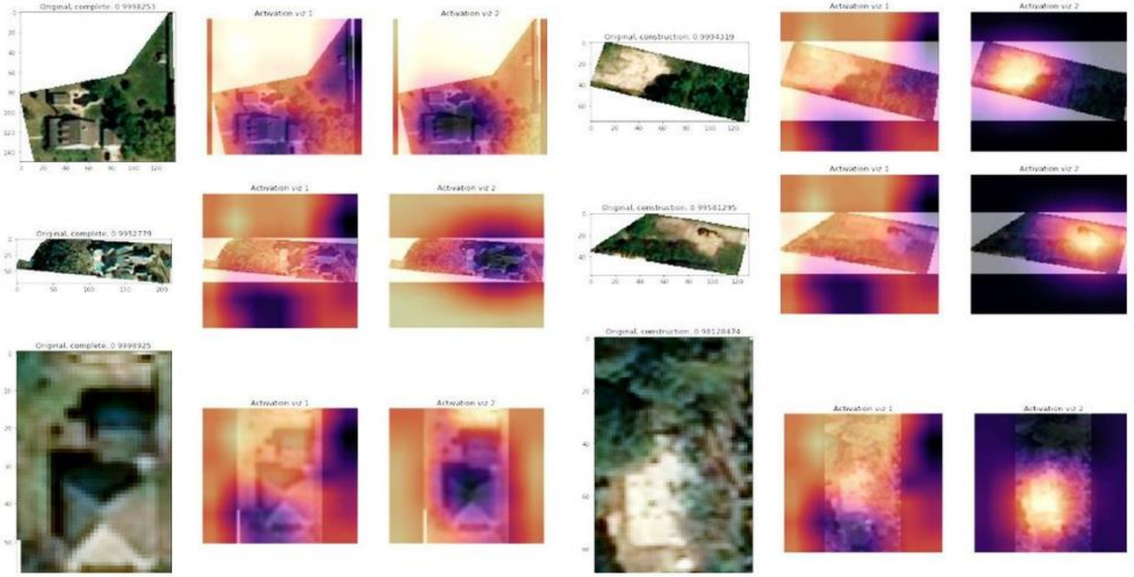
Tracking Mode – Multiple Months

Once a construction start has been detected in an image, we will track that location on a monthly basis for changes. We will apply change detection techniques to track construction progress and tag the construction activity as complete based on the detection of pre-established features such as finished roof, driveway, landscaping, absence of construction vehicles, etc.



United States® Census Bureau

Model Visual Validation

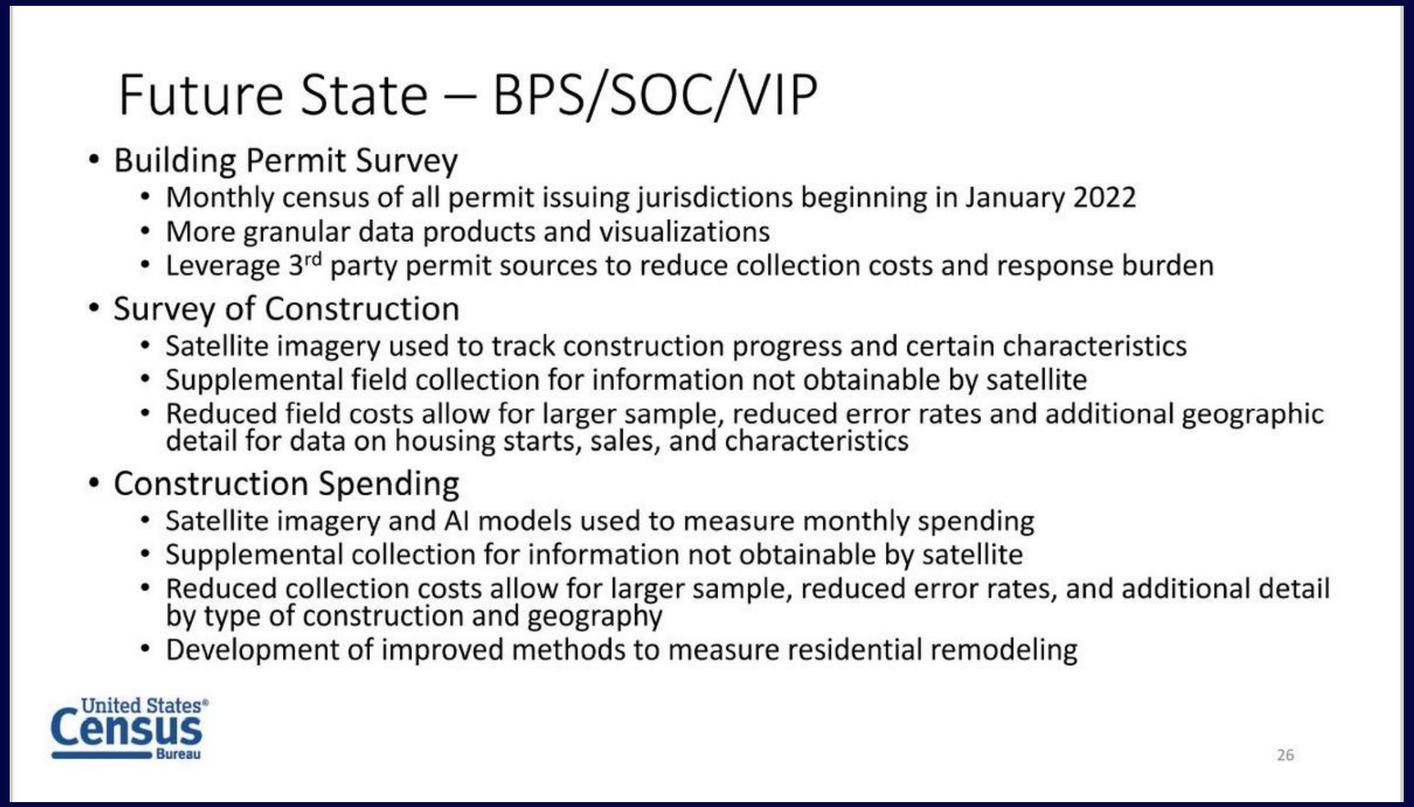


United States® Census Bureau

US CENSUS BUREAU

CENSUS.GOV

Leverage government data tools



WHAT DOES IT MEAN FOR US?

RELIABLE DATA

Zillow has an error rate of +/- 1.9% to 6.9%. This could mean several thousand dollars in difference. Having another source to cross-reference could lower the error rate. Can we find new sources?

THOROUGH RESEARCH

New industry standards for housing prices, new housing starts. Here we found OverallQuality was top feature. Would that be the same with other data? Leveraging new technology to build a better model – will Census Bureau be a part of that?

MORE INFORMED AI

Agile models can respond to external changes in the market which could affect pricing. Unemployment, contractors, lumber availability

WHAT DOES IT MEAN FOR US?

DATA GATHERING

With better data-gathering efforts, it's possible to feed real-time data to tip off a ML algorithm that variables are changing. The hiring levels of contractors in an area, like the Census Bureau rolled out, unemployment numbers, price of lumber - can tip off monitors.

COMPETITION IN THE MARKETPLACE

Other companies like OpenDoor and Redfin, realtor.com can improve selling prices for us through healthy competition with Zillow. We want Zillow to do well and come back strong. They'll learn from this.

CONCLUSIONS

DEFINE YOUR CRITERIA	KNOW WHAT MATTERS	NEURAL NETWORKS & THE HUMAN BRAIN	NEWTON'S 2ND LAW
Using descriptive and predictive analytics to identify important features in the dataset is key to knowing what to track.	While buying a home is subjective and personal, there are features common to most house sales that matter.	AI still has a way to go understanding abstract concepts of desire and market volatility, (ie Covid). Humans still need to monitor.	F=ma ... Zillow was buying hard and fast, Quick changes in the market pulled the rug out: led to "start fast, fail fast."

THANK YOU

THANKS TO JOE RAETANO, DEVIN MOYA, MATTHEW GERARDINO FOR YOUR HELP

BACKGROUND

Supplemental instruction with machine learning concepts.

TRAINING AND EXPERTISE

Coding assistance, direction.


Questions?