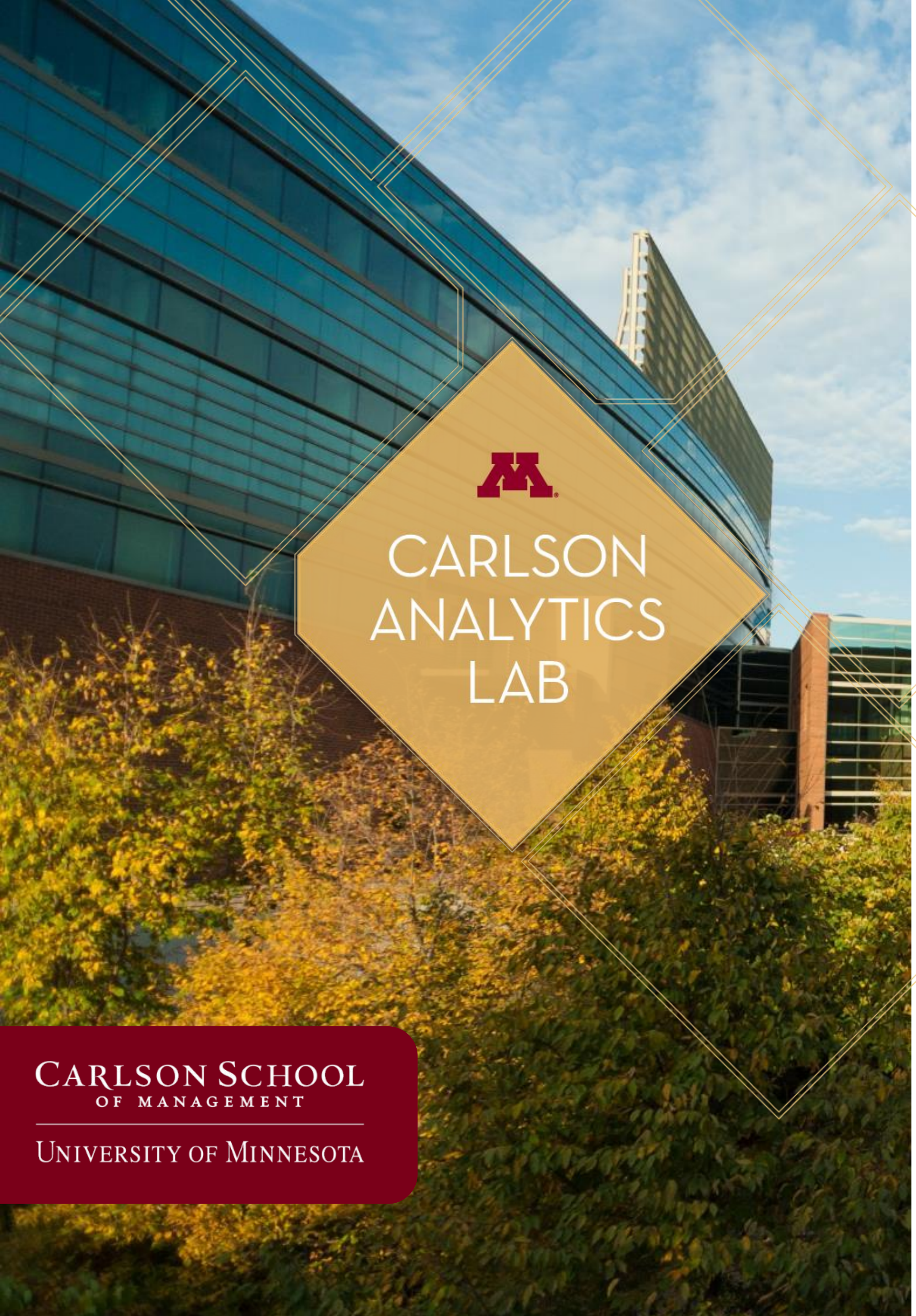


Predicting Private Market Twin Cities Rents

Final Presentation

HousingLink

April 25, 2019

Henrik Kowalkowski, Robert Lu, Will Thomas, Xiyang Xu, Xander Yang

Today's Discussion

- **Final Project Scope**
  - Brief overview of the project definition process
- **Dashboard**
  - A walkthrough, use case of the rental prediction dashboard
- **Model Review**
  - Data quality, model selection, model validation
- **Deployment**
  - Hosting the dashboard and transferring the tool
- **Potential Future Work**
  - Refreshing the data, adding features
- **Appendix**

How project scope has evolved over time

Kickoff
2/1/2019

- ❑ Project proposal and kickoff meeting
- ❑ Model impact of investments in predicting rents

Follow-Up
2/7/2019

- ❑ Requirements gathering meeting
- ❑ Model demographic changes based on changes to number of housing units in a location for next 3 years
- ❑ Dashboard for what-if scenarios; line graph for each prediction

Discussion
3/20/2019

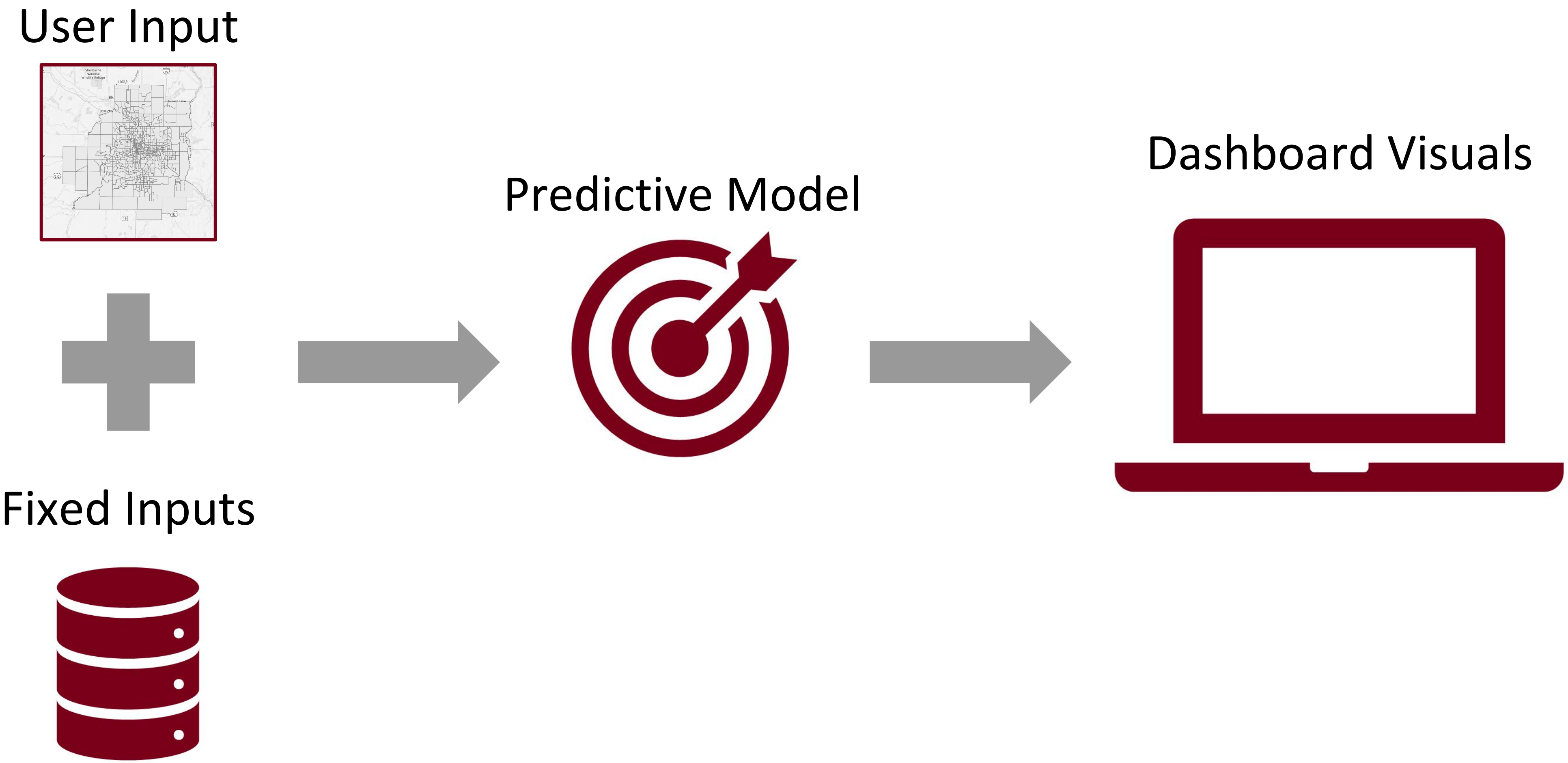
- ❑ Concerns with misinterpretation of dashboard; causal vs. predictive
- ❑ Model rent prices in a location for next 3 years
- ❑ Dashboard displaying 3 year rent predictions; line graph displaying multiple locations and heat map animating rent changes over time

How we’re helping HousingLink and advocates achieve their mission

Solution

- Aggregation of multiple private and public data sources to create:
  - **Historical rents** for location of interest
  - **Predicted rents** for next 3-year period
- **Interactive dashboard** with multiple visualizations for different analyses and geographies
- Focus on maintenance (e.g., updating with future data)

Predictions are built from user and fixed inputs



How we’re helping HousingLink and advocates achieve their mission

Solution

- Aggregation of multiple private and public data sources to create:
    - **Historical rents** for location of interest
    - **Predicted rents** for next 3-year period
- **Interactive dashboard** with multiple visualizations for different analyses and geographies
- Focus on maintenance (e.g., updating with future data)

Benefits

- Rent prices can **inform policy discussions and investment decisions**
- Robust & accurate 3-year predictions
- A centralized dataset to explore further

Dashboard Demonstration

CARLSON SCHOOL OF MANAGEMENT
UNIVERSITY OF MINNESOTA

CARLSON ANALYTICS LAB

Understanding the Process


Research
Data
Model
Dashboard

The process is sequential as each stage relies on the last to produce robust results

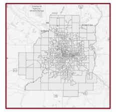
Research

Data

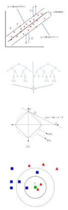
Model

Dashboard

Variable Name	Source
Average Annual Household Income	American Community Survey
Percent Bachelor's Attained	American Community Survey
Population Growth Rate	American Community Survey
Weighted Median Rent	HousingLink
Subsidized Units at Different AMI Levels	HousingLink
Annual Residential Permit Value ($)	Metropolitan Council
Annual Non Residential Permit Value ($)	Metropolitan Council
Small Area Housing Estimates	Metropolitan Council
U3 Unemployment Rate	St. Louis Federal Reserve
S&P 500 Index Value	St. Louis Federal Reserve

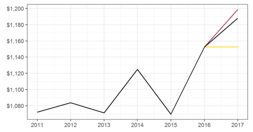


Twin Cities
County
City
Community
Planning District
Ward



= 28

Learning Model	Error
Baseline	143.88
K-Nearest Neighbors Regression	129.53
Lasso Regression	138.13
Random Forest Regression	121.51
Support Vector Regression	110.22




Research
Data
Model
Dashboard

We hypothesized education, population, macroeconomics and investment drive rents

Of those influential factors we collected these variables as proxies

Variable Name	Source
Average Annual Household Income	American Community Survey
Percent Bachelor's Attained	American Community Survey
Population Growth Rate	American Community Survey
Weighted Median Rent	HousingLink
Subsidized Units at Different AMI Levels	HousingLink
Annual Residential Permit Value ($)	Metropolitan Council
Annual Non Residential Permit Value ($)	Metropolitan Council
Small Area Housing Estimates	Metropolitan Council
U3 Unemployment Rate	St. Louis Federal Reserve
S&P 500 Index Value	St. Louis Federal Reserve


Research
Data
Model
Dashboard

We collected, cleaned and then aggregated the data from the tract level

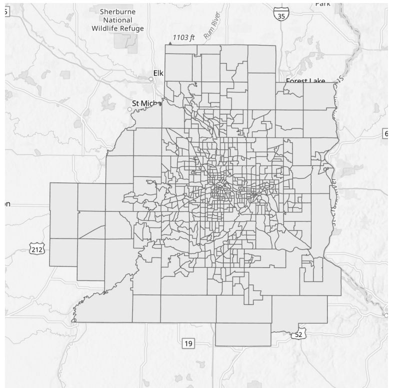
HousingLink

MINNESOTA GEOSPATIAL COMMONS

FRED ECONOMIC DATA | ST. LOUIS FED

AMERICAN COMMUNITY SURVEY U.S. CENSUS BUREAU

METROPOLITAN COUNCIL



Twin Cities

County

City

Community

Planning District

Ward

Tract Level Data Collection → **Tract Level Data Cleaning** → **Data Aggregation**

Research

Data

Model

Dashboard

We tested 4 different learning models across 7 different geographies to find the best model

Planning District

Twin Cities

Tract

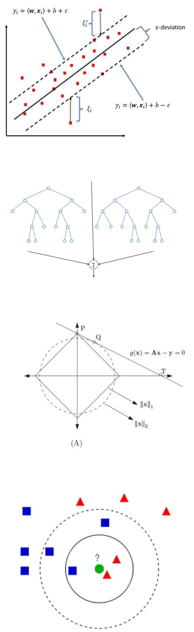
City

Ward

Community

County

×



=

28

Geographic Levels

Supervised Learning Models

Geo-Specific Models

Support vector regression minimizes error across different geographic levels

Supervised Learning Model	Avg. Difference from Actual
Baseline	$143.88
Lasso Regression	$138.13
K-Nearest Neighbors Regression	$129.53
Random Forest Regression	$121.51
Support Vector Regression	$110.22

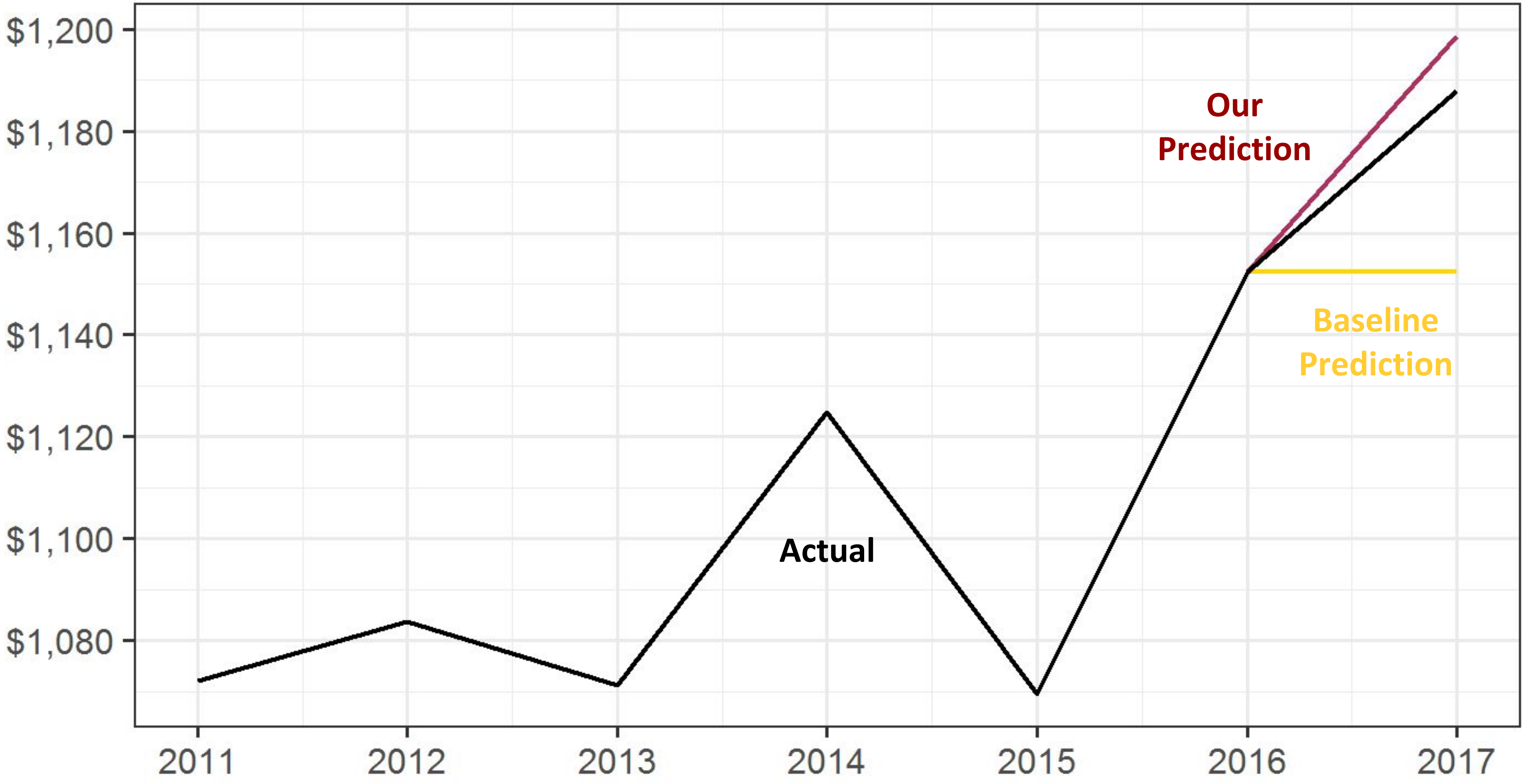
24% better than baseline

10% better than next best learning model

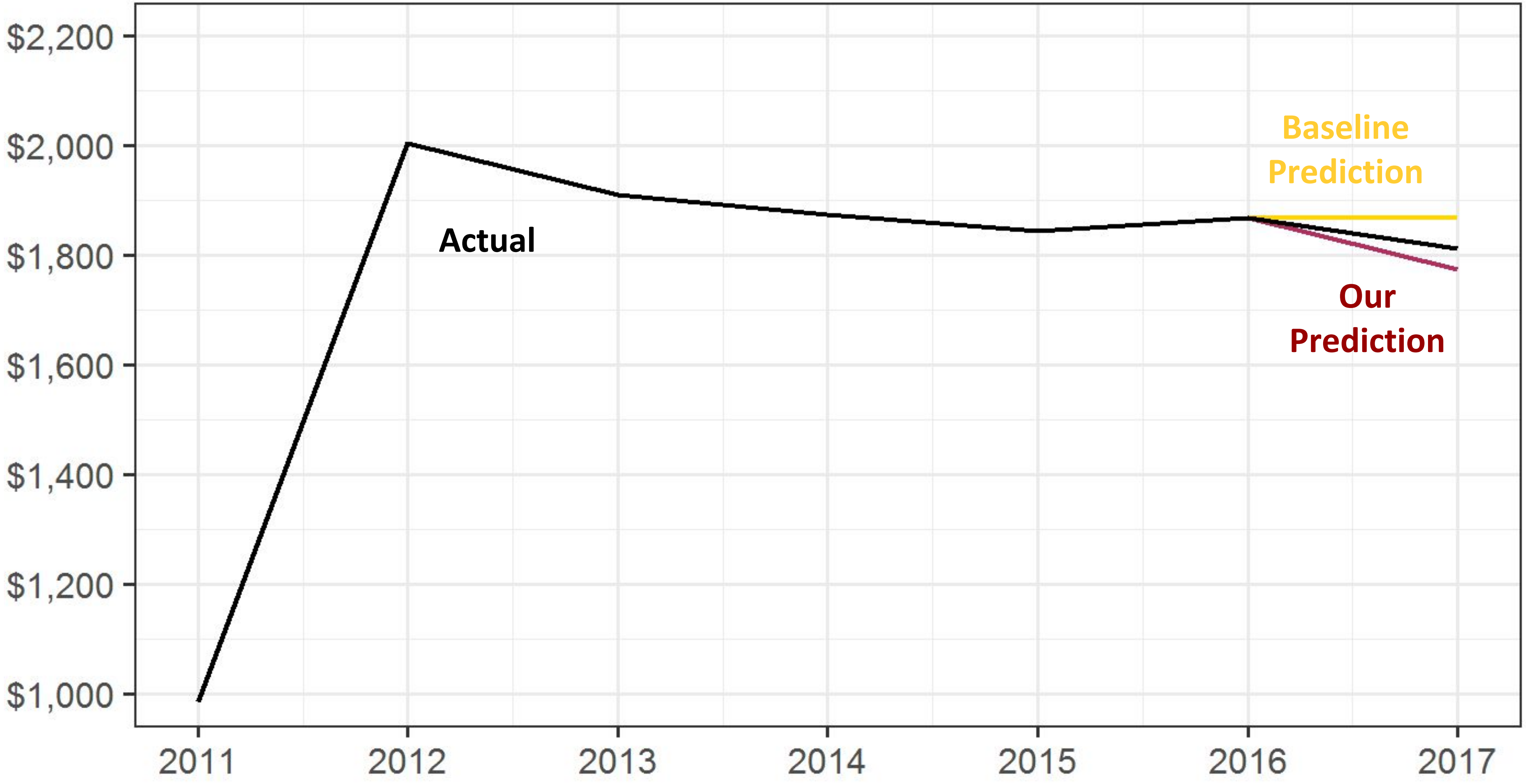
Rent and education give the most predictive power while keeping the model simple

Variable Name	Source
Average Annual Household Income	American Community Survey
Percent Bachelor's Attained	**American Community Survey**
Population Growth Rate	American Community Survey
Weighted Median Rent	**HousingLink**
Subsidized Units at Different AMI Levels	HousingLink
Annual Residential Permit Value ($)	Metropolitan Council
Annual Non Residential Permit Value ($)	Metropolitan Council
Small Area Housing Estimates	Metropolitan Council
U3 Unemployment Rate	St. Louis Federal Reserve
S&P 500 Index Value	St. Louis Federal Reserve

Model performance in Longfellow illustrates potential future accuracy



Downtown illustrates that predictions are not linear extrapolation




Research
Data
Model
Dashboard

Our dashboard is driven by R Studio and hosted by R Shiny

Benefits
Free tier, 25 hours per month
Fast, responsive performance
Refresh without altering code
User metrics and system up time
The opportunity to increase resources and available hours at any time

The tool works because each step is carefully built off the last

Research

Data

Model

Dashboard

Variable Name	Source
Average Annual Household Income	American Community Survey
Percent Bachelor's Attained	American Community Survey
Population Growth Rate	American Community Survey
Weighted Median Rent	HousingLink
Subsidized Units at Different AMI Levels	HousingLink
Annual Residential Permit Value ($)	Metropolitan Council
Annual Non Residential Permit Value ($)	Metropolitan Council
Small Area Housing Estimates	Metropolitan Council
U3 Unemployment Rate	St. Louis Federal Reserve
S&P 500 Index Value	St. Louis Federal Reserve

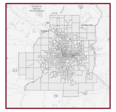
HousingLink

MINNESOTA GEOSPATIAL COMMONS

FRED ECONOMIC DATA | ST. LOUIS FED

AMERICAN COMMUNITY SURVEY U.S. CENSUS BUREAU

METROPOLITAN COUNCIL



Twin Cities

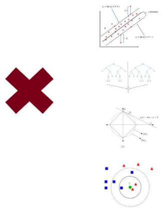
County

City

Community

Planning District

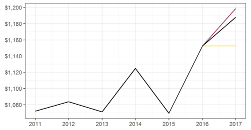
Ward



= 28

Learning Model	Error
Baseline	143.88
K-Nearest Neighbors Regression	129.53
Lasso Regression	138.13
Random Forest Regression	121.51
Support Vector Regression	110.22

Variable Name	Source
Average Annual Household Income	American Community Survey
Percent Bachelor's Attained	**American Community Survey**
Population Growth Rate	American Community Survey
Weighted Median Rent	**HousingLink**
Subsidized Units at Different AMI Levels	HousingLink
Annual Residential Permit Value ($)	Metropolitan Council
Annual Non Residential Permit Value ($)	Metropolitan Council
Small Area Housing Estimates	Metropolitan Council
U3 Unemployment Rate	St. Louis Federal Reserve
S&P 500 Index Value	St. Louis Federal Reserve



RStudio

Shiny by RStudio

Final Logistics and Next Steps

CARLSON ANALYTICS LAB

CARLSON SCHOOL OF MANAGEMENT
UNIVERSITY OF MINNESOTA

Deployment Plan

HL & CAL Deploy Dashboard

Week of 4/22

- ❑ Host Shiny App from HousingLink
- ❑ Link dashboard from HousingLink site

HL Communicate Launch

Week of 4/29

- ❑ HousingLink releases relevant information regarding app availability

CAL Knowledge Transfer

5/2

- ❑ CAL transfers full documentation to HousingLink

CAL Post Production Support Ends

5/15

- ❑ HousingLink and partners maintain the app going forward

Possible future enhancements

Enhancement	Difficulty
Refresh data with 2018 values	Moderate
Handle 2020 census tract redefinitions	High
Allow users to export data from dashboard	Moderate
Embed the app within a website	Moderate
Explore causal relationship between affordable housing and market units in relation to rent further	High

CARLSON SCHOOL OF MANAGEMENT
UNIVERSITY OF MINNESOTA

Thank You

Appendix

CARLSON ANALYTICS LAB

CARLSON SCHOOL OF MANAGEMENT
UNIVERSITY OF MINNESOTA

How does a simple predictive model work?

We fit a line on the values of education and rent that we have observed

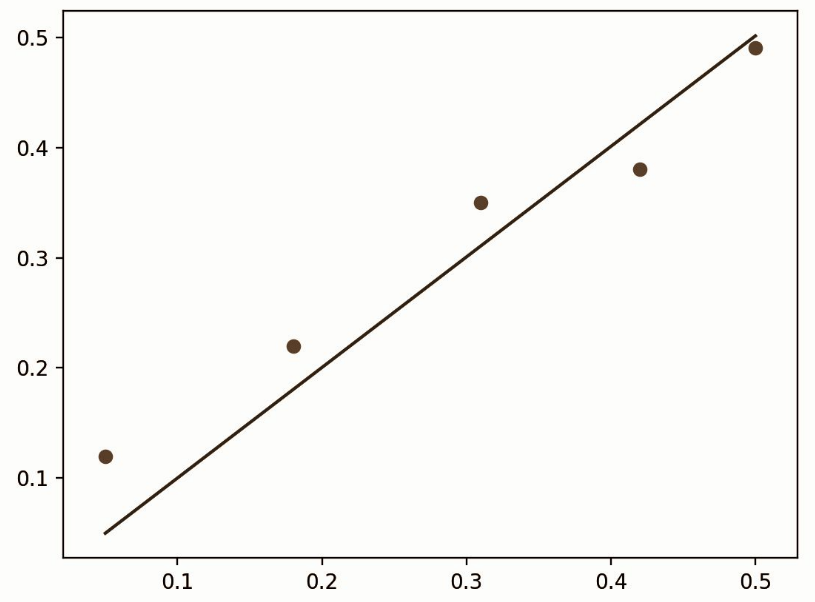
This gives us slopes or “effects” for the values

*education = $200 “for every 1% increase in education rent increases by 200 * 1% dollars”*

rent = $1.2 “for every $1 increase in previous rent, future rent increases by $1.2”

When we get an unseen data point e.g. education = 90% and rent = $800 we just multiply each value by its slope and add them together:

200 * (0.9) + 1.2 * (800) = $1,140 (prediction)



Dashboard User Guide [Video]

A brief walkthrough of the dashboard and its features

https://youtu.be/9RixO5p2n9c