

Property Valuation Model

Predicting the sale price for homes in Ames, Iowa



What is a Property Valuation Model (PVM)?

This is a model used within some type of application-based tool that aims to estimate property value, usually given property characteristics and other factors that relate to market value.

Why is PVM Important?



Knowledge

Provide buyers and sellers with a reasonable estimation for what the home is valued at and/or likely to sell for.



Time

The time it takes to get an estimate via traditional processes vs. AVM cannot be understated.



Processes

While initially used to estimate residential real estate value, these models are now supporting underwriting decisions for consumer loans

You May Know...

01

Freddie Mac & Equifax

(Commercial AVM Providers)

02

Zillow & Trulia

(Free AVM Provider)

Can this PVM Compete With the Most Popular Free AVM (Zillow) on the Market?

5 bd | 3 ba | 2,205 sqft
1326 Lincoln Way, Ames, IA 50010
Sold: \$233,000 | Sold on 06/07/21 | Zestimate®: \$291,800
Est. refi payment: \$1,592/mo [Refinance your loan](#)

[Home value](#) [Owner tools](#) [Home details](#) [Neighborhood details](#)

Home value

Zestimate range

\$199.8K

Home value history
View the history of this home's value, listing price, Rent Zestimate, and more using the graph below.

Home values ▾

1 year 5 years 10 years

This home \$240.1K Apr 2018

\$200K

3 bd | 3 ba | 1,368 sqft
2916 Roxboro Dr, Ames, IA 50010
Sold: \$279,750 | Sold on 07/08/20 | Zestimate®: \$341,700
Est. refi payment: \$2,018/mo [Refinance your loan](#)

[Home value](#) [Owner tools](#) [Home details](#) [Neighborhood details](#)

Home value

Zestimate range

\$279.9K

Home value history
View the history of this home's value, listing price, Rent Zestimate, and more using the graph below.

Home values ▾

1 year 5 years 10 years

This home \$282.5K Jun 2020

\$279.7K

What is a Zestimate?

A Zestimate valuation is a proprietary Zillow tool for getting a free, instant estimate of a home's market value. The Zestimate is calculated through a Zillow algorithm that crunches data from public property records, tax records, recent home sales in the area, and user-submitted information to come up with an approximate market value for a home.*

The Zestimate takes into account:

- Home details like square footage, location and the number of bedrooms and bathrooms.
- On-market data such as listing price, description, comparable homes in the area and days on the market.
- Off-market data such as tax assessments, prior sales and other publicly available records
- Market trends, such as seasonal changes in demand.

A Note on Zestimate Accuracy?*

How accurate is the Zestimate?

The Zestimate's accuracy depends on location and the availability of data in an area. The [Zestimate](#)'s median error rate for on-market homes nationwide is 3.2%, meaning Zestimates for half of all on-market homes are within 3.2% of the ultimate sale price, and half are not. For off-market homes, the median error rate is 7.52%. Check your home value estimate for free by entering your home address. You can improve the accuracy of your Zestimate by claiming the address as your home and updating the home details.

Data Sources

The screenshot shows the City of Ames website. At the top, there's a navigation bar with links for HOME, ABOUT AMES, GOVERNMENT (highlighted in green), DOING BUSINESS, LIVING, VISITING, I WANT TO..., a search bar, and a language selection dropdown. Below the navigation is the City of Ames logo and social media links (Facebook, Twitter, YouTube). A sidebar on the left contains links for City Assessor, Property Owner's Legal Responsibility, Property Search, Credits & Exemptions, Estimate Your Property Tax, Appealing Your Assessment, Conference Board, Sales, Reports, Staff Contacts and Address, and Iowa State Association of Assessors. The main content area has a heading "SALES" and a sub-section "Commercial Sales" with a note about being searchable via Beacon. It also includes a "Residential Sales Questionnaire".

The screenshot shows the FRED Economic Research website. At the top, there's a logo and a header with "ECONOMIC RESEARCH" and "FRED ECONOMIC DATA ST. LOUIS FED". Below the header is a message: "Your trusted data source since 1991." On the right, there are links for "Release Calendar" and "Economic Rese...". The main content area features a section titled "House Price Indexes" under "Categories > Prices". It includes a table of concepts like House Price Index, Price, and Price Index, each with a count (e.g., 672). There are also sections for "Geography Types" (Metropolitan Statistical Area, State, Census Division) and "S&P/Case-Shiller U.S. National Home Price Index" (Index Jan 2000=100, Monthly, with options for Not Seasonally Adjusted and Seasonally Adjusted).

The screenshot shows the Beacon Story County, IA / City of Ames website. At the top, there's a logo and a header with "Beacon Story County, IA / City of Ames". Below the header is a navigation bar with links for Layers, Map, Search, Ames Comp, and Story Comp. The main content area features three search functions: "Comparables Search", "Address Search", and "Parcel ID Search". Each search function has its own input fields and search buttons.



Cleaning & Preprocessing

The dataset has 4,181 observations (Homes sold within the past 18-24 months) and 92 features. The data dictionary provided by the city website was vital to understanding the features.

General Rule: > 10%

Eliminate features that have more than 10% missing values. Here all features that were eliminated by this had > 50% missing values

Augment the Data

Add HPI: A quarterly index that measures the average price change in repeat sales or refinancing on the same properties

Exploratory Data Analysis

The goal is to reduce the feature space via 1. eliminating features that are not useful to predict target, and 2. Combining existing features that relate better to target.

Outliers



Missing Values

Explore the +- 2.5% Target Values

Eliminate observations that fall too far outside of the “middle” based on the distribution of the data (which is fairly normal).

Data Leakage



Train Test Split

Split the data at this point prior to examining correlations and trends.

Feature Engineering



Feature Space

Missing Values: (18 Features)

Cleaning & Preprocessing

The dataset has 4,181 observations (Homes sold within the past 18-24 months) and 92 features. The data dictionary provided by the city website was vital to understanding the features.

General Rule: > 10%

Eliminate features that have more than 10% missing values. Here all features that were eliminated by this had > 50% missing values

Augment the Data

Add HPI: A quarterly index that measures the average price change in repeat sales or refinancing on the same properties

Exploratory Data Analysis

The goal is to reduce the feature space via 1. eliminating features that are not useful to predict target, and 2. Combining existing features that relate better to target.

Outliers



Missing Values

Explore the +- 2.5% Target Values

Eliminate observations that fall too far outside of the “middle” based on the distribution of the data (which is fairly normal).

Data Leakage



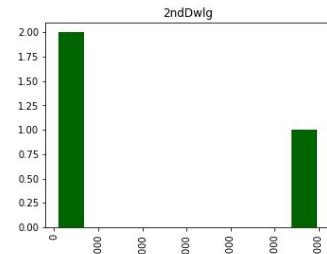
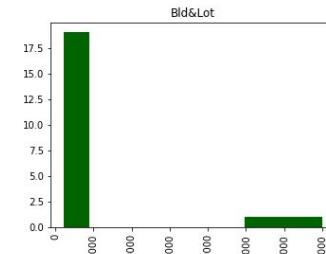
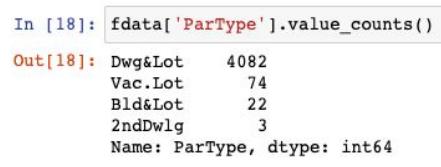
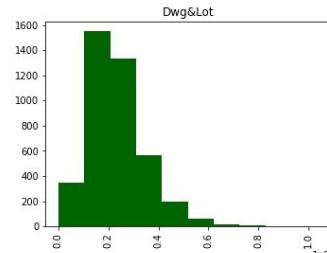
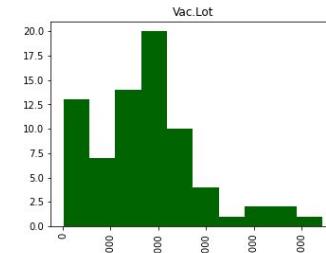
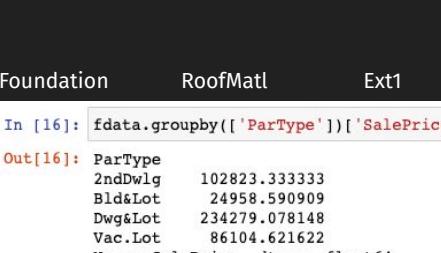
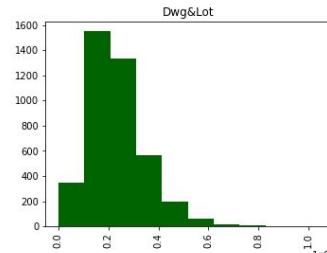
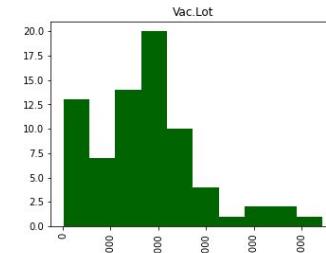
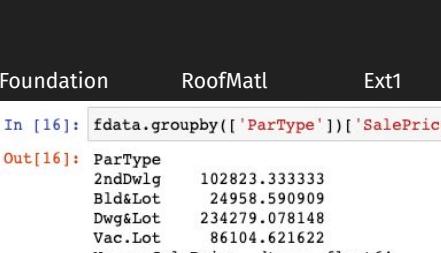
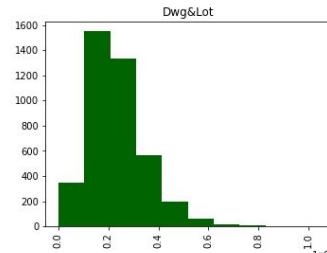
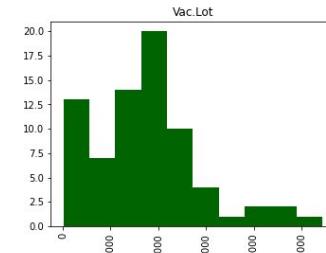
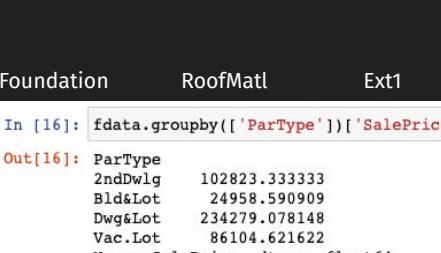
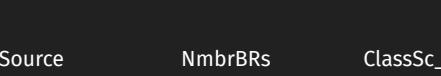
Train Test Split

Split the data at this point prior to examining correlations and trends.

Feature Engineering



Outliers: (99 Observations)

MapRefNo	GeoRefNo	Tier	Range	Prop_Addr	ZngCdPr	ZngCdSe	ZngOLPR	ClassPr_s	Legal_pr
SchD_S	TxD_S	MA_Ownr1							
Rcrd_Yr	Rcrd_Mo	Inst1_No							Yr
ValType	X1TPr_D	X1TSe_D							
ImpAcX1S	ImpAcX2S	HSTtl_D							
LotArea	ParType	BldgNo_S							
Ext2	MasVnrType	Heating							
GarYrBlt	Cars	GarageArea							
PA_Pred	PA_Strt	PA_StSfx							
ZngOLSc									

```
In [18]: fdata['ParType'].value_counts()
Out[18]: Dwg&Lot    4082
          Vac.Lot     74
          Bld&Lot     22
          2ndDwlg      3
          Name: ParType, dtype: int64
```

X2TPr_S X2TSe_S LndAcX1S

AcreGr AcreNt_S Neighborhood

Foundation RoofMatl Ext1

```
In [16]: fdata.groupby(['ParType'])['SalePrice'].mean()
Out[16]: ParType
          2ndDwlg    102823.333333
          Bld&Lot    24958.590909
          Dwg&Lot    234279.078148
          Vac.Lot     86104.621622
          Name: SalePrice, dtype: float64
```

Cleaning & Preprocessing

The dataset has 4,181 observations (Homes sold within the past 18-24 months) and 92 features. The data dictionary provided by the city website was vital to understanding the features.

General Rule: > 10%

Eliminate features that have more than 10% missing values. Here all features that were eliminated by this had > 50% missing values

Augment the Data

Add HPI: A quarterly index that measures the average price change in repeat sales or refinancing on the same properties

Exploratory Data Analysis

The goal is to reduce the feature space via 1. eliminating features that are not useful to predict target, and 2. Combining existing features that relate better to target.

Outliers



Missing Values

Explore the +- 2.5% Target Values

Eliminate observations that fall too far outside of the “middle” based on the distribution of the data (which is fairly normal).

Data Leakage



Train Test Split

Split the data at this point prior to examining correlations and trends.

Feature Engineering



Seasonal Adjustment: (1 New Feature)

Cleaning & Preprocessing

The dataset has 4,181 observations (Homes sold within the past 18-24 months) and 92 features. The data dictionary provided by the city website was vital to understanding the features.

General Rule: > 10%

Eliminate features that have more than 10% missing values. Here all features that were eliminated by this had > 50% missing values

Augment the Data

Add HPI: A quarterly index that measures the average price change in repeat sales or refinancing on the same properties

Exploratory Data Analysis

The goal is to reduce the feature space via 1. eliminating features that are not useful to predict target, and 2. Combining existing features that relate better to target.

Outliers



Missing Values

Explore the +- 2.5% Target Values

Eliminate observations that fall too far outside of the “middle” based on the distribution of the data (which is fairly normal).

Data Leakage



Train Test Split

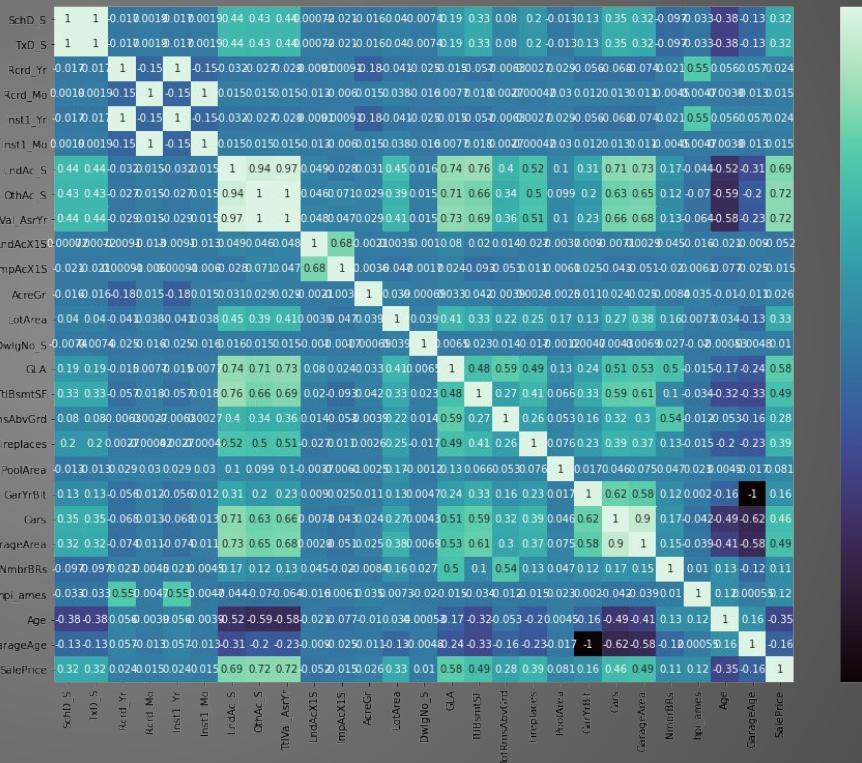
Split the data at this point prior to examining correlations and trends.

Feature Engineering



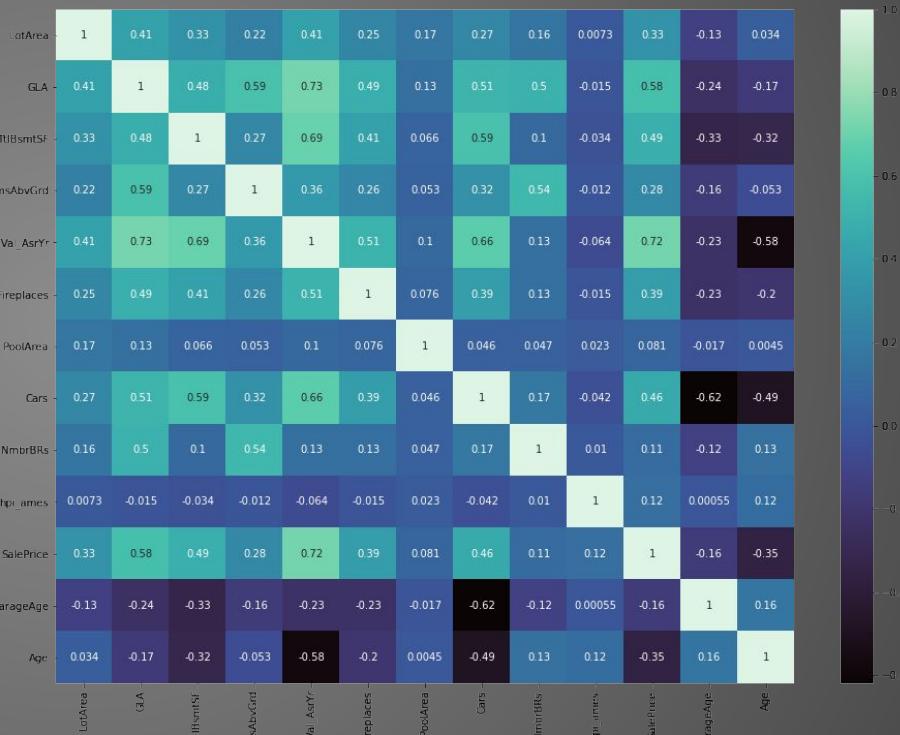
EDA & Feature Engineering: Numerical

- Create Age & Garage Age: (-)
- Multicollinearity: Does not affect performance (XGBoost) but definitely affects importance plot.
 - Cars & Garage Area
 - Assessed Value
 - School vs. Tax District
- Identify Strong Predictors (abs() >.10)
 - HPI (low but necessary)



EDA & Feature Engineering: Numerical

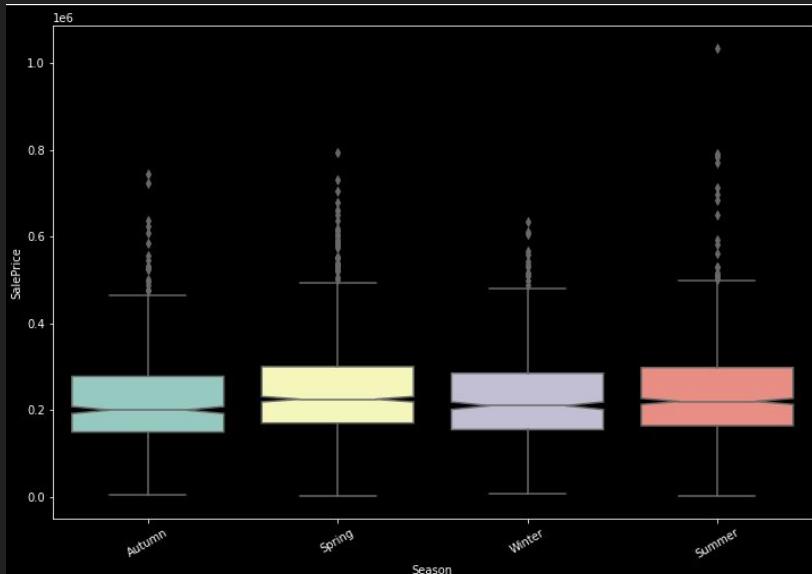
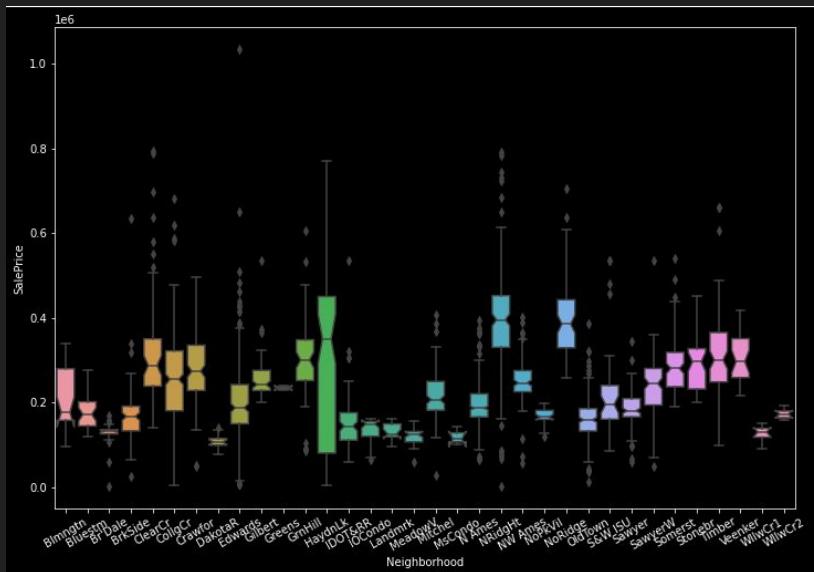
- Create Age & Garage Age: (-)
- Multicollinearity: Does not affect performance (XGBoost) but definitely affects importance plot.
 - Cars & Garage Area
 - Assessed Value
 - School vs. Tax District
- Identify Strong Predictors ($\text{abs}(< \text{value}) > .10$)
 - HPI (low but necessary)



Note: SalePrice column is NOT at the end...

EDA & Feature Engineering: Categorical

Notched box-plot inspection usually serves as a good starting place for me. Eliminate features based on visual inspection or statistical test where the visual alone does not show good separation.

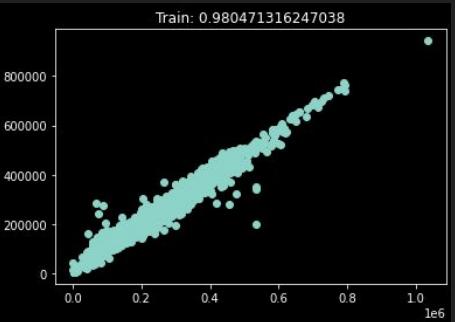


EDA: (28 Features | 3061 Observations (4082 Total))

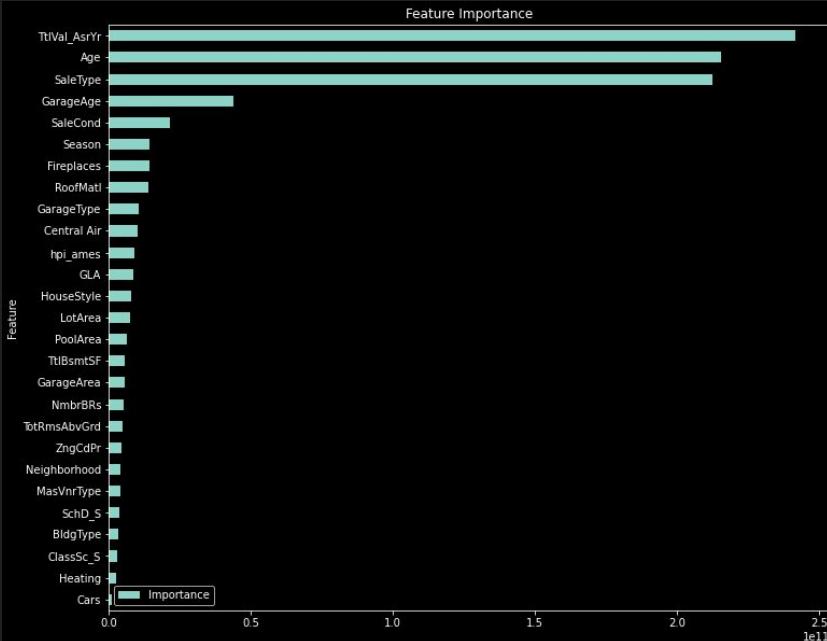
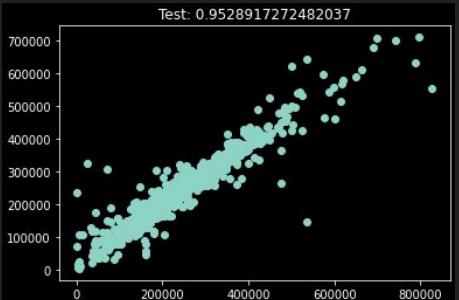
MapRefNo	GeoRefNo	Tier	Range	Prop_Addr	ZngCdPr	ZngCdSe	ZngOLPR	ClassPr_s	Legal_pr
SchD_S	TxD_S	MA_Ownr1	MA_Ownr2	MA_Line1	MA_Line2	MA_City	MA_State	MA_Zip1	MA_Zip2
Rcrd_Yr	Rcrd_Mo	Inst1_No	Inst1_Yr	Inst1_Mo	Inst1Pr	LndAc_S	ImpAc_S	othAc_S	TtlVal_AsrYr
ValType	X1TPr_D	X1TSc_D	X2TPr_D	X2TSc_D	X1TPr_S	X1TSc_S	X2TPr_S	X2TSc_S	LndAcX1S
ImpAcX1S	ImpAcX2S	HSttl_D	MilVal_D	HSttl_S	MilVal_S	AcreX_S1	AcreGr	AcreNt_S	Neighborhood
LotArea	ParType	BldgNo_S	DwlgNo_S	BldgType	YrBuilt	HouseStyle	Foundation	RoofMatl	Ext1
Ext2	MasVnrType	Heating	Central Air	GLA	TtlBsmtSF	TotRmsAbvGrd	Fireplaces	PoolArea	GarageType
GarYrBlt	Cars	GarageArea	YrSold_YYYY	MoSold_MM	SalePrice	SaleType	SaleCond	ParcelRet	PA_Nmbr
PA_PrcD	PA_Stt	PA_StSfx	PA_PostD	PA_UntTyp	PA_UntNo	Date	Source	NmbrBRs	ClassSc_S
ZngOLSe	HPI	Age	GarageAge	Season					

Model: XGBoost: Great Performance

	Train	Test
RMSE	22,641.53	35,060.57
MAE	14,860.41	19,971.07



	CV Results	Notes
eta	.01	Reduces overfitting
Objective	Squared Error	Outlier Penalty
n	500	Best Value



Model: PVM Performance Metrics

	Test	Description
PP10	70.5%	% predictions within 10% of true value
PP20	88.7%	% predictions within 20% of true value
PP30	6.37%	% predictions outside 30% of true value

Active listings accuracy

Top Metro Areas States National

State	Median Error ⓘ	Homes With Zestimates ⓘ	Within 5% of Sales Price ⓘ	Within 10% of Sales Price ⓘ	Within 20% of Sales Price ⓘ
Iowa	2.06%	15.0 K	80.93%	93.65%	97.36%

Off Market

Top Metro Areas States National

State	Median Error ⓘ	Homes With Zestimates ⓘ	Within 5% of Sales Price ⓘ	Within 10% of Sales Price ⓘ	Within 20% of Sales Price ⓘ
Iowa	8.01%	1.14 M	34.73%	56.89%	79.31%

PP30...What happened here?

	Prop_Addr	SalePrice	TtlVal_AsrYr	PredictedPrice	Difference
145	5726 QUARRY DR	70310.0	374000	308380.09375	238070.09375
110	1912 LEOPOLD DR	827581.0	588200	551207.93750	276373.06250
1009	3904 ALDRIN AVE	25000.0	433800	326351.93750	301351.93750
2562	3114 WOODLAND ST	535000.0	120700	145577.93750	389422.06250

Sales

Date	Seller	Buyer	Recording	Sale Condition - NUTC	Type	Multi Parcel	Amount
10/23/2020	CARMER, RICHARD & KATHLEEN	CARMER, RICHARD LEE REVOC TRUST ETAL	2020-13216	No consideration	Warranty		\$0.00
5/31/2017	HUNZIKER CHRISTY SHIRK BUILDERS INC	CARMER, RICHARD & KATHLEEN	2017-06620	Partial Assessment (New construction/incomplete structural changes as of Jan 1)	Corp/Business Entity		\$429,706.00
4/6/2017	HUNZIKER DEVELOPMENT COMPANY LLC	HUNZIKER CHRISTY SHIRK BUILDERS INC	2017-03027	Partial Assessment (New construction/incomplete structural changes as of Jan 1)	Corp/Business Entity		\$25,000.00

 There are other parcels involved in one or more of the above sales:

[Recording: 2017-03027 - Parcel: 0529260170](#)



—
Thank you.



Problems to solve

1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna nec. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

2

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna nec. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

3

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna nec. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

4

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna nec. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Problems to solve

1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna nec. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

2

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna nec. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

3

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna nec. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

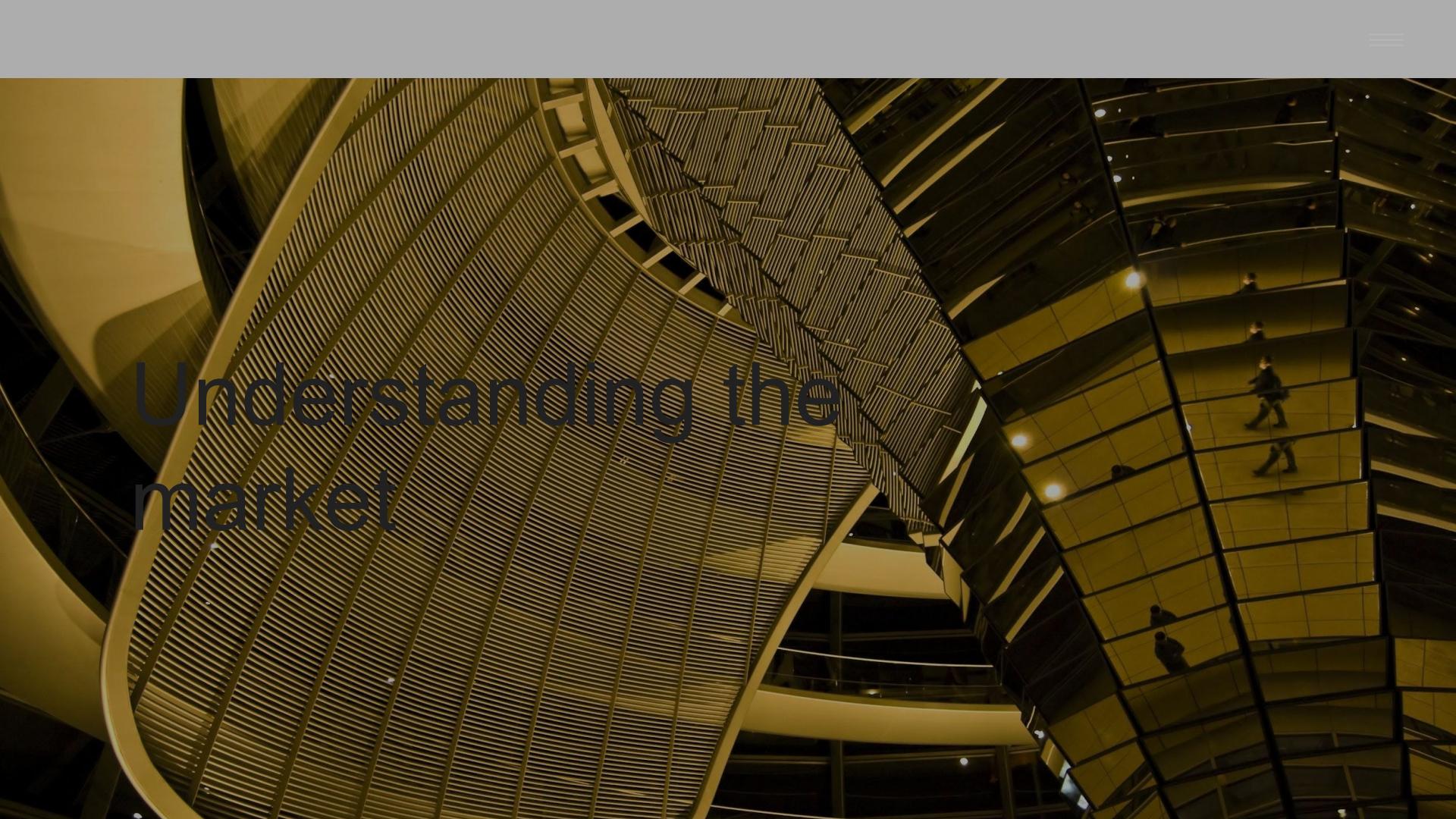
4

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna nec. Lorem ipsum dolor sit amet, consectetur adipiscing elit.



Project objective

**Lorem ipsum dolor sit amet, consectetur
adipiscing elit, sed do eiusmod tempor
incididunt ut labore et dolore magna
aliqua ipsum dolor sit amet.**



Understanding the
market



Market trends

01

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt labore magna aliqua. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Client Implications:

Consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor sit amet, consectetur adipiscing elit tempor incididunt ut labore et dolore magna aliqua.



Market trends

02

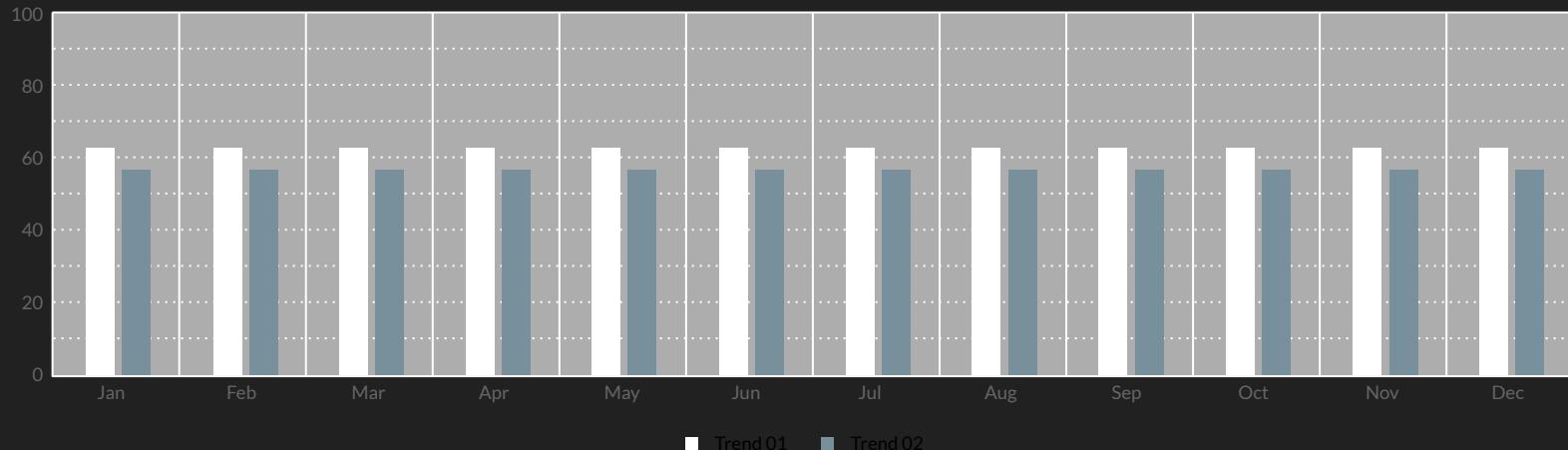
Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt labore dolore magna aliqua. Lorem ipsum dolor sit amet, consectetur adipiscing elit.

Client Implications:

Consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor sit amet, consectetur adipiscing elit tempor incididunt ut labore et dolore magna aliqua.



Trend analysis

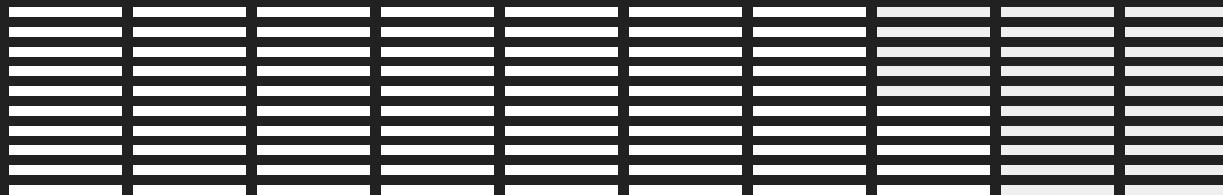


Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt labore dolore magna aliqua. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt labore dolore magna aliqua. Lorem ipsum dolor sit amet, consectetur adipiscing elit.



Consectetur adipisciing elit, sed do eiusmod tempor incididunt labore dolore magna aliqua. Consectetur adipisciing elit. Sed do eiusmod tempor incididunt ut labore et dolore magna nec. Lorem ipsum dolor sit amet, adipisciing elit. Eiusmod tempor incididunt ut labore et dolore magna nec. Lorem ipsum dolor sit amet, consectetur dolor ipsum adipisciing elit.

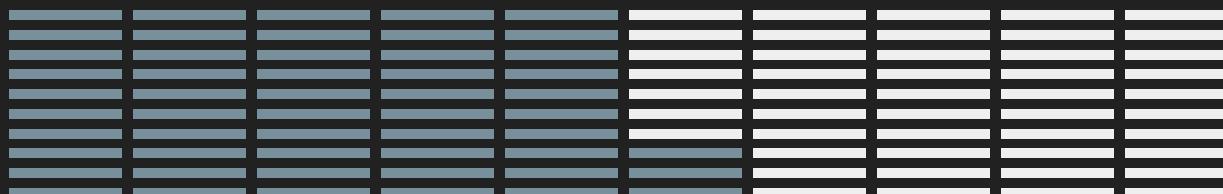
Trend 01



75%

Incididunt ut labore et dolore sed
do eiusmod tempor incididunt ut
labore et dolore magna aliqua.

Trend 02



53%

Tempor incididunt labore et dolore
sed do eiusmod tempor incididunt
ut labore et dolore magna aliqua.

Target audience

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt labore dolore magna aliqua. Lorem ipsum dolor sit amet consectetur adipiscing elit.

- 01 | Consectetur adipiscing elit
- 02 | Sed do eiusmod tempor incididunt ut labore
- 03 | Magna aliqua lorem ipsum dolor sit amet
- 04 | Eiusmod tempor incididunt ut labore et
- 05 | Dolore magna aliqua





Proposed solution

Consectetur adipiscing elit ipsum dolor
sit amet eiusmod tempor, incididunt nec
labore et dolore magna aliqua ipsum
dolor sit amet tempor incididunt .

Process



0
1

Lorem ipsum dolor sit

Lorem ipsum dolor sit



0
2



0
3

Lorem ipsum dolor sit

Deliverables

Consectetur nec labore

45K

**Consectetur
adipiscing
elit sed
tempor
incididunt
ut labore
et dolore magna aliqua.**

**Consectetur
adipiscing
elit sed
tempor
incididunt
ut labore
et dolore magna aliqua.**

Adipiscing elit sed

690K

**Consectetur
adipiscing
elit sed
tempor
incididunt
ut labore
et dolore magna aliqua.**

Sed eiusmod

100K

**Consectetur
adipiscing
elit sed
tempor
incididunt
ut labore
et dolore magna aliqua.**

Vision

Lorem ipsum dolor sit

Lorem ipsum dolor sit amet, consectetur adipiscing
elit, sed do eiusmod tempor incididunt labore dolore
magna aliqua dolor sit amet, consectetur
adipiscing.

20XX

20XX