Presentation for Project 2 *Linear Regression with Ames Housing Data*

Matt Burke, GA-DSI-DEN-11
April 2020

Presentation Context

For project 2, we were tasked with building a linear regression model to predict sale prices of residential homes using housing data for Ames, IA. As someone with no real estate experience and, prior to starting General Assembly's Data Science Immersive, no data science experience, I was pleasantly surprised with my model's performance. That said, it's not one that I'd recommend using to predict sale prices in the "real world".

To that end, in an effort to deliver a presentation that was both realistic and that also encapsulated the principles of project 2, what follows is a pitch to Zillow for a new program. In this imagined scenario, I am a Zillow employee on the strategic business development team, pitching in-house to Zillow executives.

Zillow Premier Analyst

Finding Alpha with Our Agents

agenda

- staying competitive
- introduce Zillow Premier Analyst (ZPAn)
- how ZPAn works
- case study and POC: Ames Housing Data test group
- what’s next

online real estate
an increasingly
competitive market

Staying competitive

Image sources:
realtor.com
Redfin
Neighborhood Scout
Homesnap

“Buy land, AJ, cause God ain’t makin anymore of it”

-Tony Soprano

(Sopranos, S04E07)

Image source

Zillow Premier Analyst
what’s that?

Zillow Premier Agents

Zillow™ PREMIER AGENT

Image sources

Zillow Premier Agent

is designed to

connect agents with active buyers and sellers

to help them grow their business, as well as providing them with tools and resources to help them convert those buyers and sellers into lasting clients.

Zillow Premier Analyst

is designed to

connect data professionals with Zillow Premier Agents

in order to identify and exploit alpha in the market, as well as provide them with tools and resources to help them refine their technical and analytical skills.

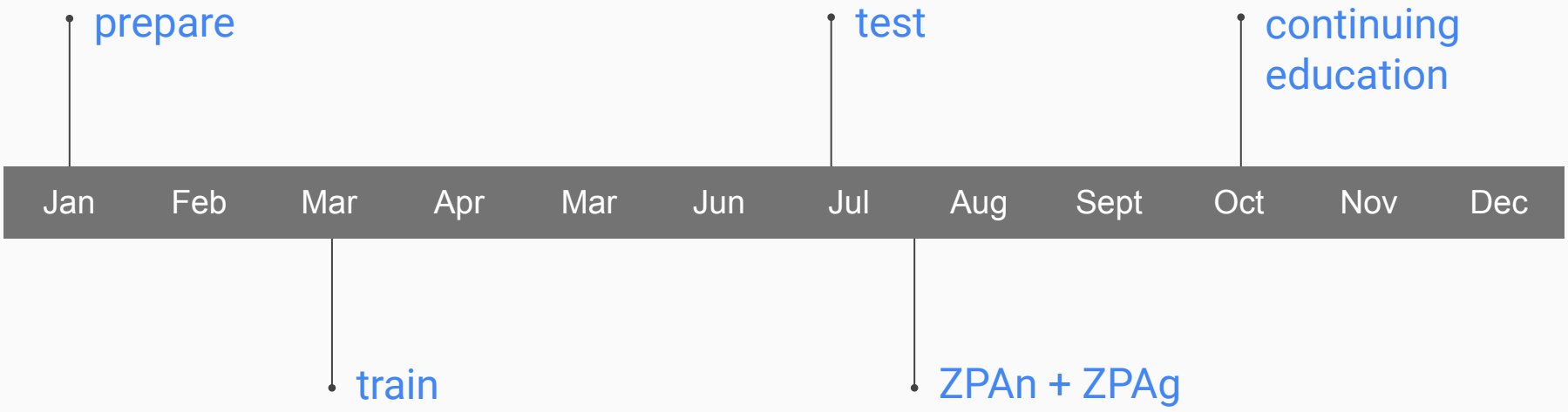
Strengthening the ecosystem

Zillow™ PREMIER AGENT

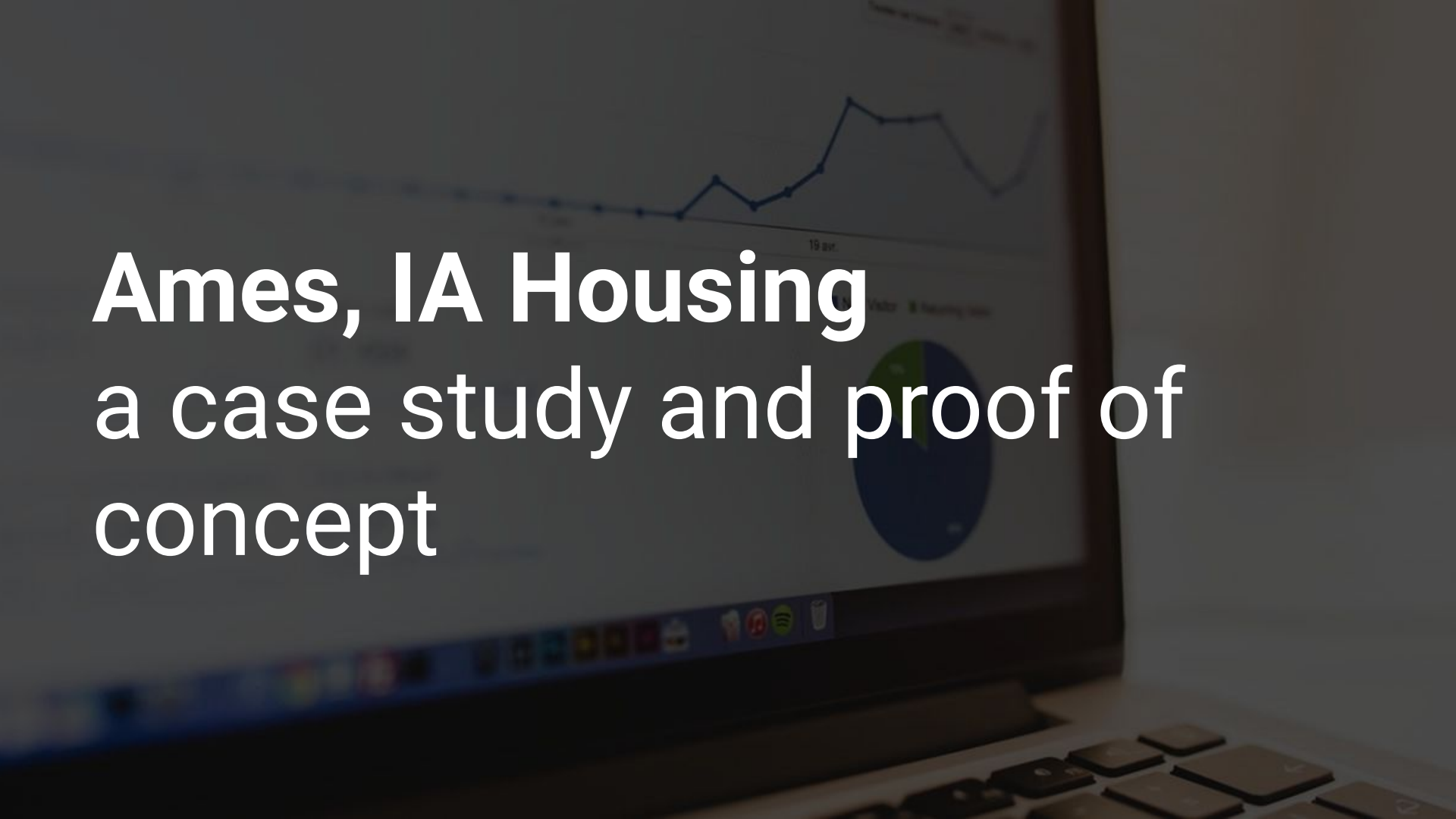
Zillow®

Image sources:
Zillow
Data analyst
Housing chart

ZPAn lifecycle



AGENTS
+
ANALYSTS
=
α



Ames, IA Housing
a case study and proof of concept

proof of concept

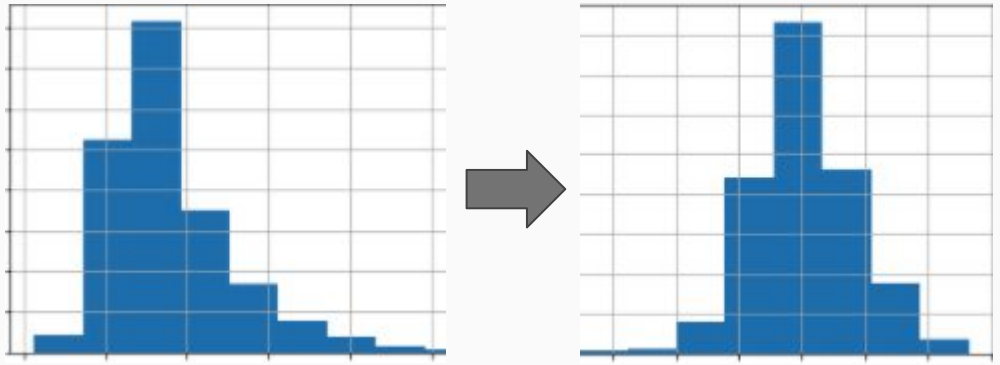
- Case study
  - 70 subjects
  - Subject: Matt
  - *Prior to program…*
  - Real Estate experience: 0
  - Data Science experience: 0

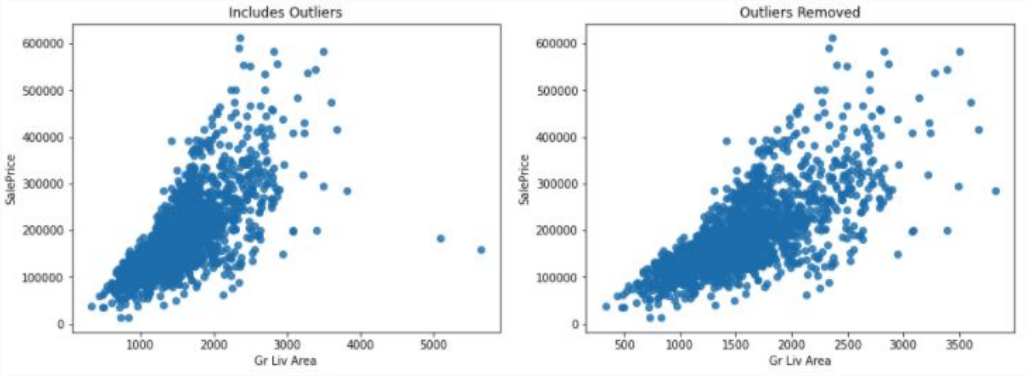
- The program
  - 4 weeks
  - 40h / week
  - 75/25 split - DS/RE
  - *the task...*
  - Ames Housing Data
  - Predict sale prices
  - MAE

the data training...

...is intense...

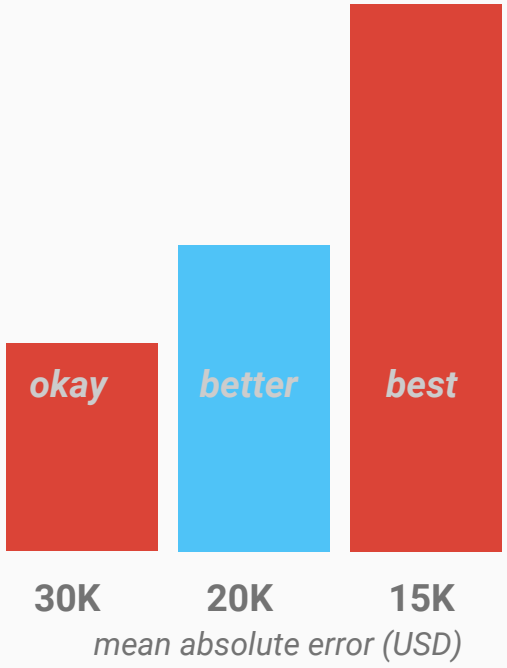
- statistics
- modeling techniques
- programming
- normalizing data
- predicting
- feature engineering
- bias and variance
- data visualization
- measuring model performance
- confidence intervals
- hypothesis testing
- distributions
- correlations
- ...
- ...and more!





the results

- basic modeling techniques
- consistent model performance
- 50% decrease in MAE*
- avg. R2 score of .89**



*MAE (Mean Absolute Error) is the average amount +/- predictions were off by

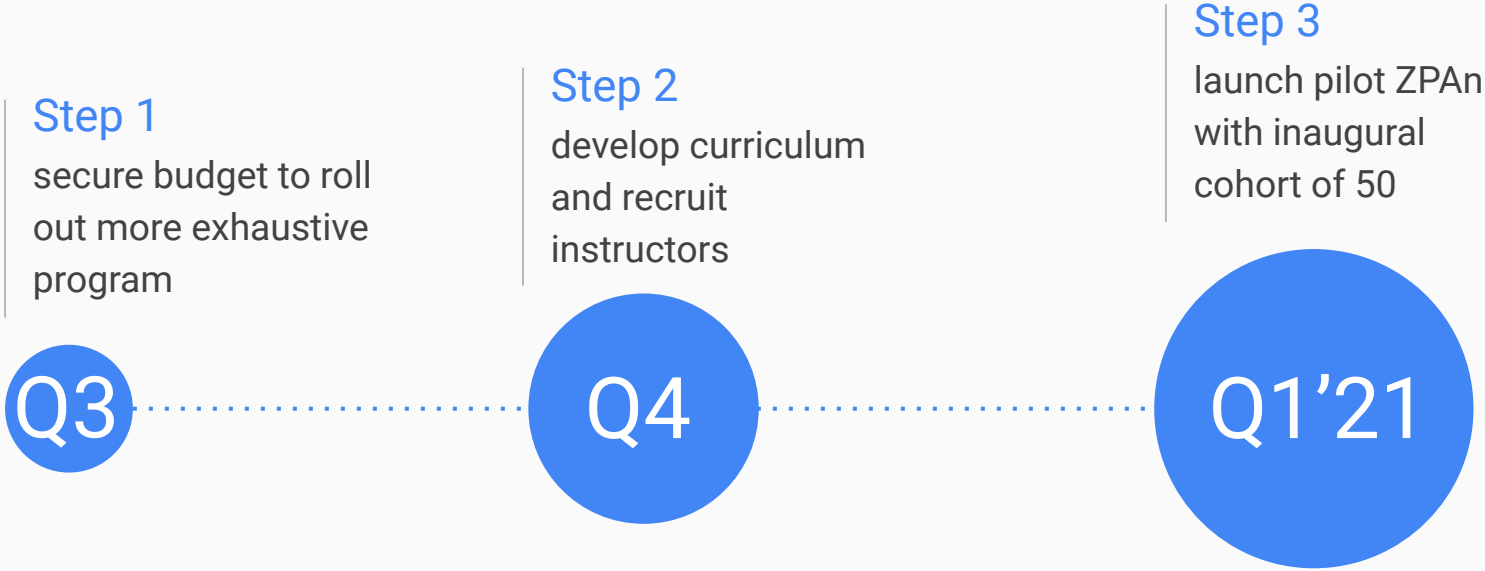
**R2 explains variation in model. An R2 of .89 says: 89% of the variation in sale price can be explained by model

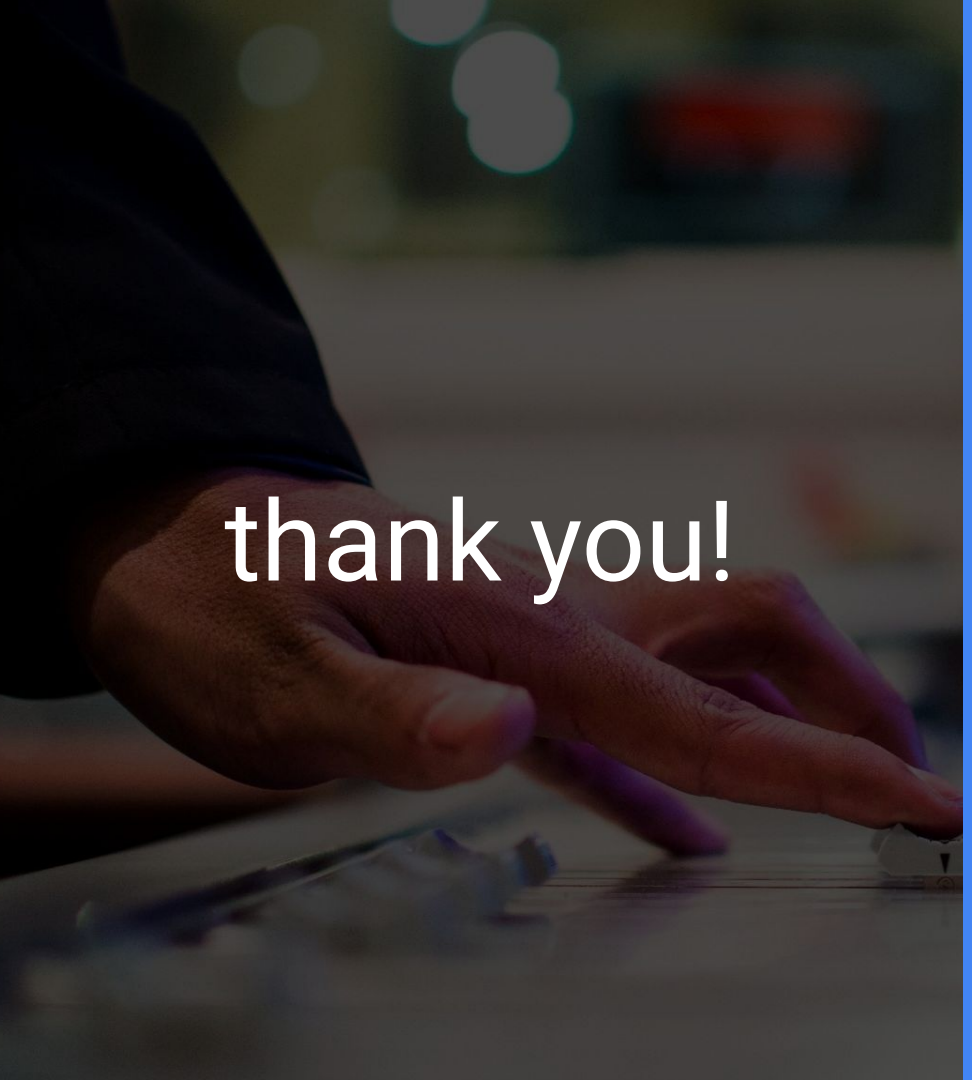
key takeaways

- results from proof of concept are promising
- as a tuition-based program, ZPAn provides additional non-recurring revenue stream...
- ...as well as indirect recurring revenue once Analysts are partnered with Agents

what’s next?

Next Steps


Step 1
secure budget to roll out more exhaustive program
Q3
Step 2
develop curriculum and recruit instructors
Q4
Step 3
launch pilot ZPAn with inaugural cohort of 50
Q1’21



thank you!

questions?