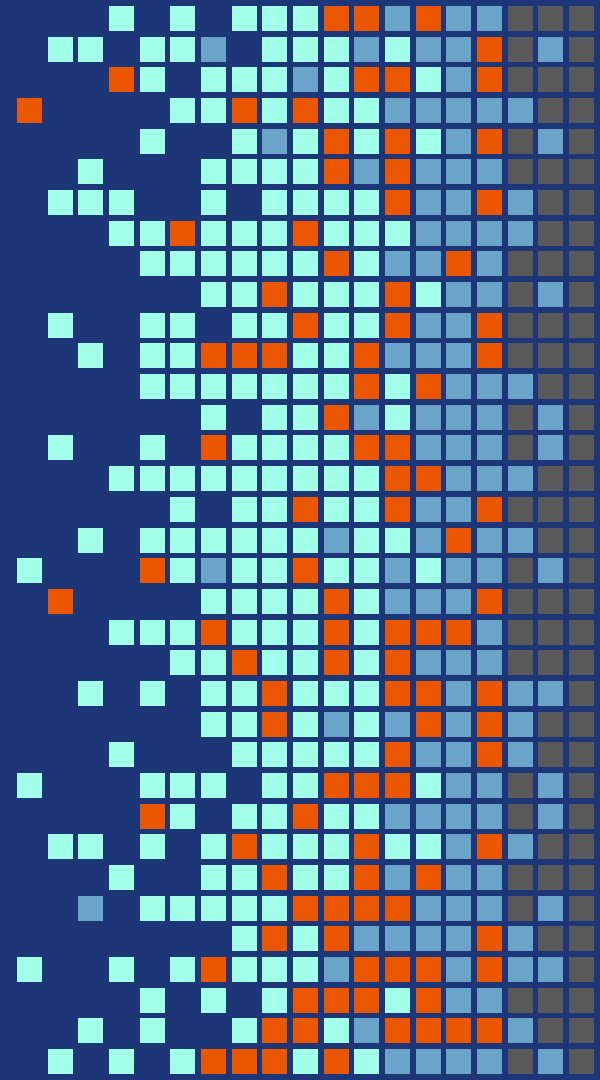


# HomeDescription Demo Day Deliverable

datares

June 2nd



# Meet the Team



Gauresh Kapoor



Joanna Wang



Darren Furtado



Ananya Garg



Jane Zou

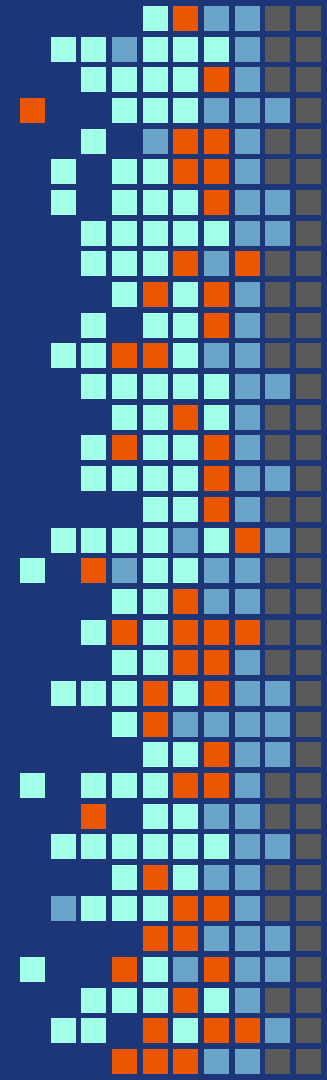


Haoran Jia



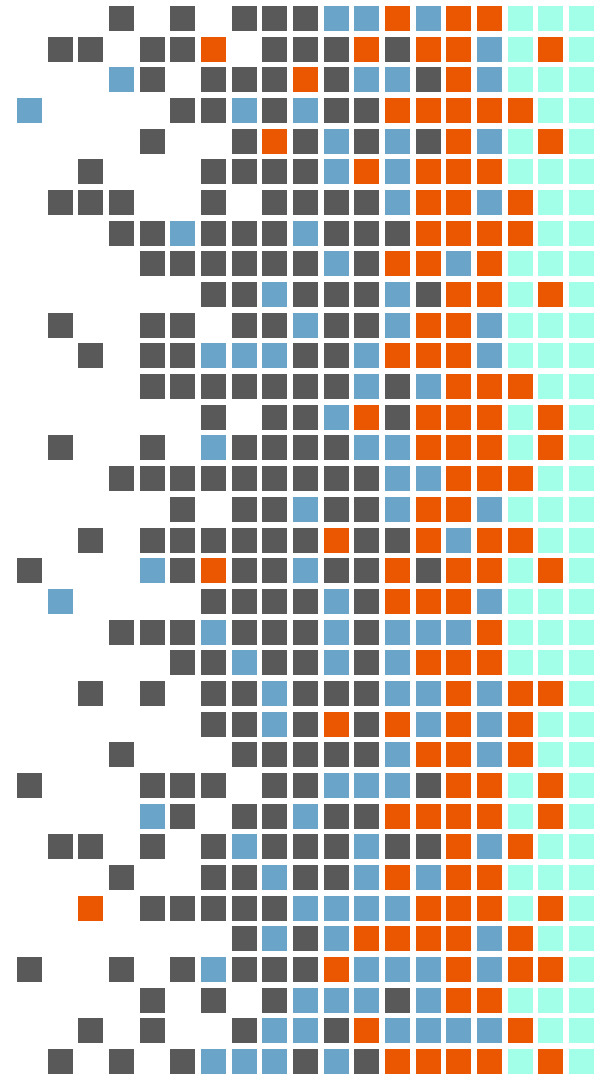
Siddarth Srikanth

# Introduction

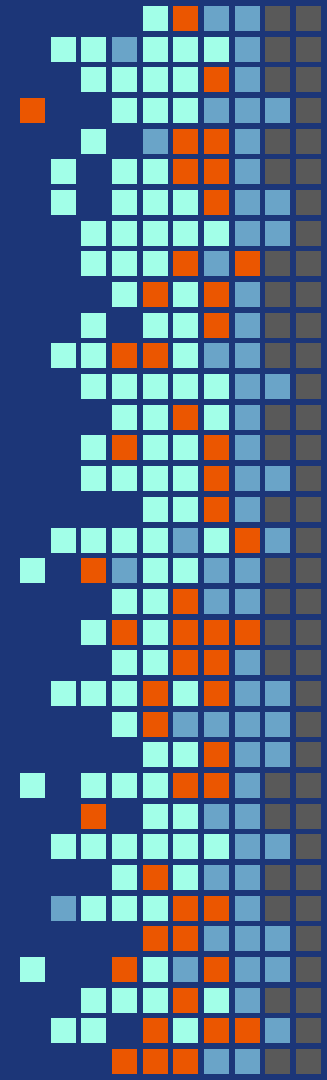


# Project Scope and Regression Analysis

- ❖ Focused on creating a "sellability" index for real estate descriptions on zillow.
- ❖ Quantified the impact that each variable (such as cost, # of bedrooms, Year Built, etc.) has on the price and time to sell on market.
- ❖ Carried out simple and multi-linear regressions between variables to identify trends.
- ❖ Engaged strongly with *text data analysis*, coming up with models for sentiment analysis, agent gender classification, and regressions.

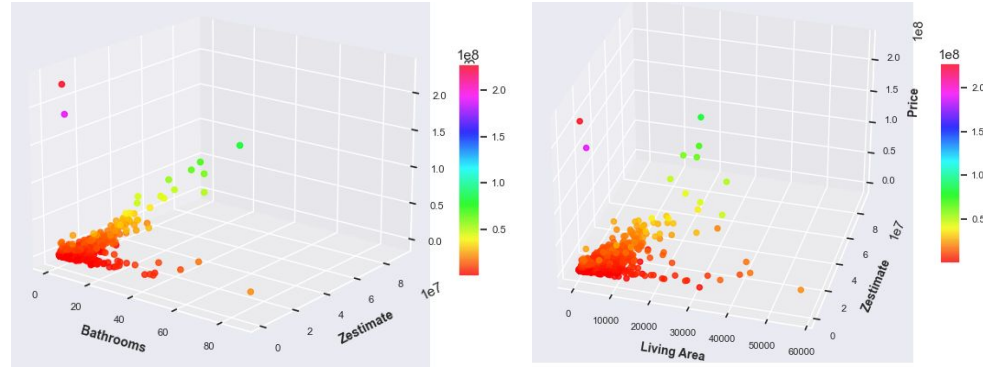
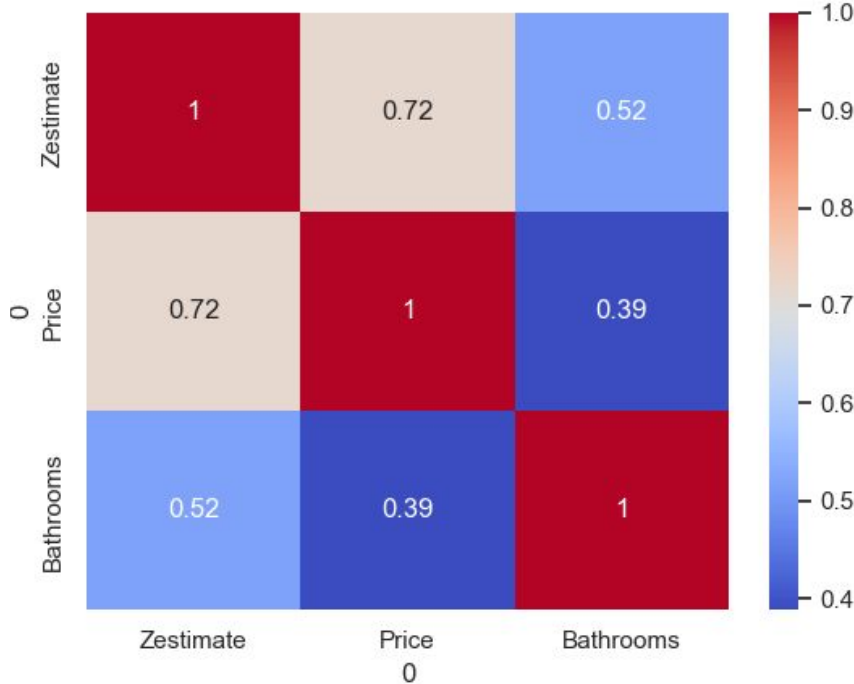


# Non Text Regression



# Exploratory Data Analysis

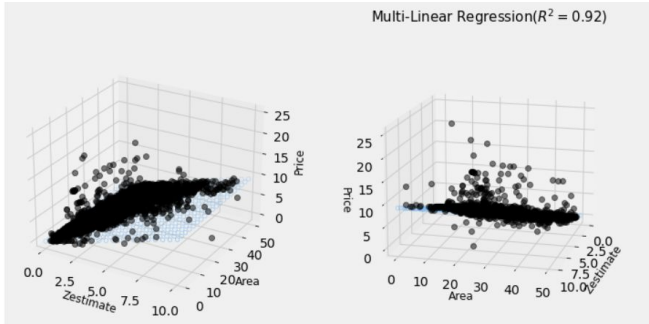
Effective Variables: Zestimate, Bathrooms, Living Area



- ❖ **Zestimate:** directly related to price, highest positive linear correlation with price (0.72 r-value)
- ❖ **Bathrooms:** strongest variable out of available listing specifications (0.52 r-value)
- ❖ **Living area:** positive linear correlation, also strong indicator of price and sellability when combined with Zestimate (0.51 r-value)

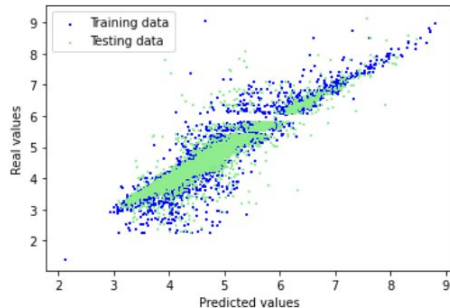
# Regression Testing

- ❖ Gradient boosting model over all variables.



## ❖ Size and Zestimate

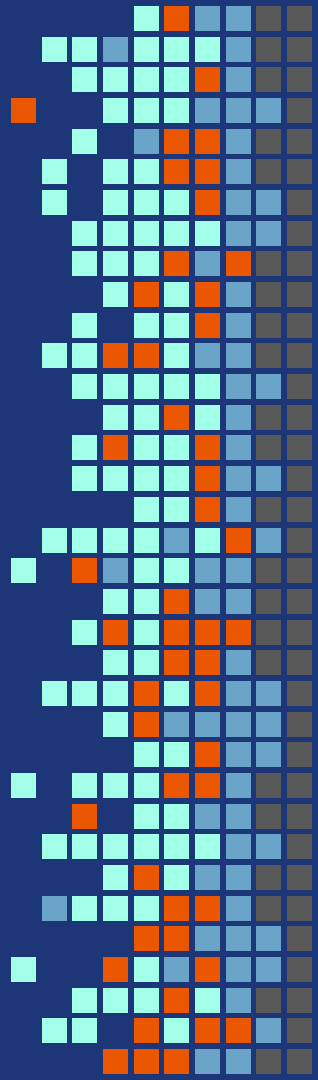
RMSE of training set: 0.1833  
RMSE of testing set: 0.2321  
R square of training set: 0.9272  
R square of testing set: 0.8865



# Data Cleaning Process

- ❖ Data type:
  - E.g. "[Gas]", "\$400 monthly"...
- ❖ Outliers:
  - E.g.  $size > 250,000 \text{ feet}^2$ ,  $price > \$0.1B$
- ❖ Minor groups:
  - E.g. cities with  $< 50$  houses on sale.
- ❖ Null values and Dummy variables

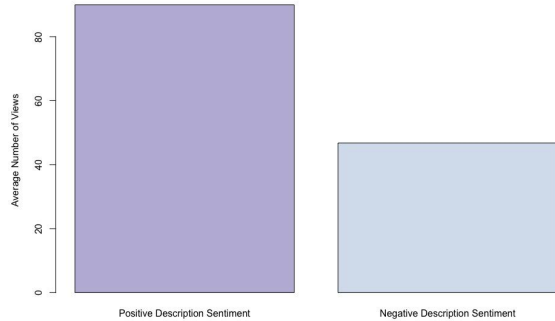
# Text Description Analysis



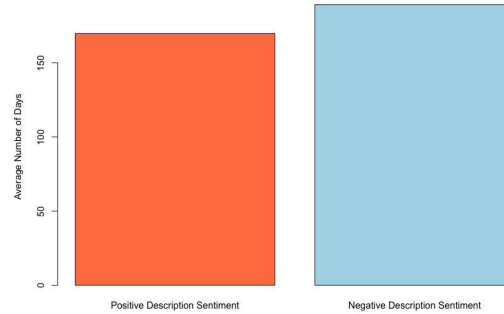


# Exploratory Data Analysis

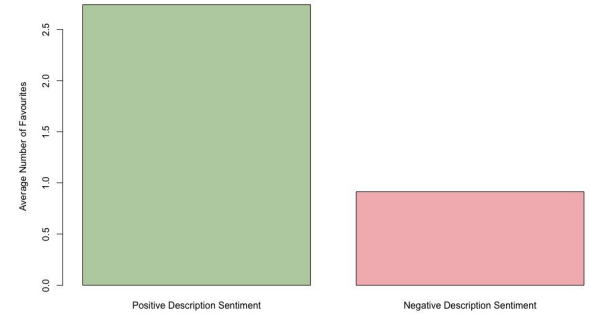
Average Listing View Count for Agencies with Average Positive/Negative Description Sentiment



Average Days on Zillow for Agencies with Average Positive/Negative Description Sentiment

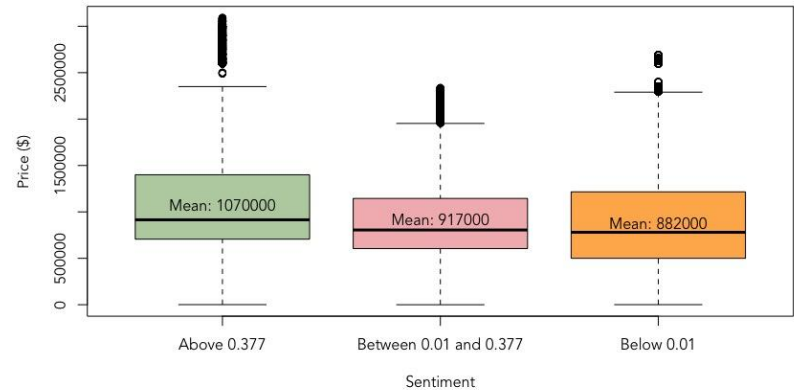


Average Listing Favourite Count for Agencies with Average Positive/Negative Description Sentiment



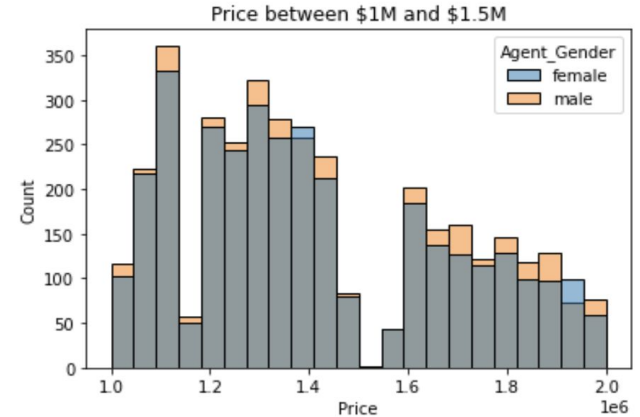
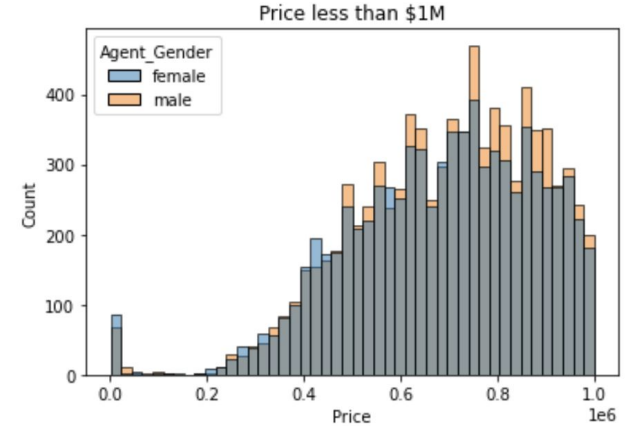
- ❖ Exploring the relationship between sentiment of listing description and listing view count, favourite count, and days on Zillow
- ❖ Estimating the regression relationship between sentiment and price through recursive partitioning

Range of Prices Depending on Description Sentiments



# Sentiment Analysis

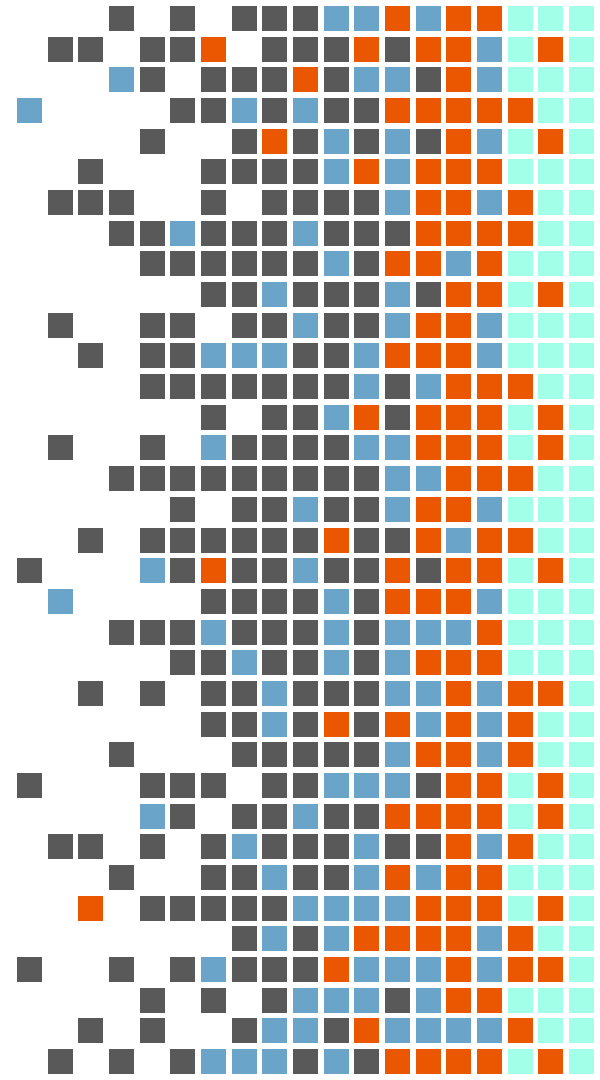
- ❖ Most common words:
  - Similar 'positive' qualities (e.g. *radiant*, *perfection*, *unmatched*, etc.)
  - *Starter*
  - *Swim*
  - *Venice*
- ❖ Price was \$24,620 higher for male agents
- ❖ Zestimate was \$9,873 higher for male agents
- ❖ Houses were listed on Zillow for 4.1 fewer days for male agents
- ❖ About 14,000 male agents and 13,000 female agents



# Conclusion

## Summary of Findings:

- ❖ Prices have positive correlations to zestimates, bathrooms, and living area
- ❖ On average, listing prices lower as description sentiments get more negative
  - “Positive” descriptions: “positive” qualities, *Starter*, *Swim*, *Venice*
- ❖ Low correlations between agent gender vs price and agent gender vs sentiment



# THANKS FOR LISTENING!

Any questions?

