

Forecasting Zestimate Error for Zillow:

A Data Analysis
IFN701 Project

Student: Linni QIN (n9632981)

Supervisor: Dr Guido Zuccon

***Zillow:** A leading real estate agent in U.S.*

***Zestimate:** Zillow's home value estimation model*

***logerror** = $\log(\text{Zestimate}) - \log(\text{SalePrice})$*



**Scopes
Deliverables**

**Motivation
Significance**

Methodologies

**Constraints
& Risks
Management**

Project Scopes & Deliverables



- Zillow Scopes:

- Improve Zestimate model
- Predict Zestimate errors for 6 timepoints:
Oct. Nov. and Dec. of 2016, Oct. Nov. and Dec. of 2017
- Practice on the training data through 2016
3millions of properties in 3 counties in California, USA (LA, OR, VE)
57 variable physical data/ property (eg. room number, location)
Test data of 2017 will be available on October2, 2017

Zillow Deliverables: CSV file, R markdown file

- QUT Scopes & Deliverables

- IFN701: Project Proposal, Project Presentations and Final project report

Project Motivation & Anticipated Significance

- Motivation


- Accurate prediction of future home value get largest ROI
- Helpful for decision making during real estate transaction

- Anticipated Significance

- 2006 – 2017 Zestimate
Median Error Rate 14% → 5%
- Median Error means
50% home value estimation \approx 5%
another 50% prediction > 5%

$$\text{\$1,175,000} * 5\% = \text{\$58,750}$$


- **A trusted platform to monitor home asset**







**111 Archer Ave,
New York, NY 10031**
4 beds • 3 baths • 3,410 sqft

Built in 2009, perfectly blending elegance with functional living space. Excellent floor plan with 3 beds up and 1 on main. Open living, kitchen & dining w/ huge fireplace & Sound views. Spacious kitchen w/ slab granite surfaces & center island. Huge master suite with Jacuzzi tub & separate shower. Features; hwd floors, all

FOR SALE
\$1,175,000
Zestimate®: \$1,275,448

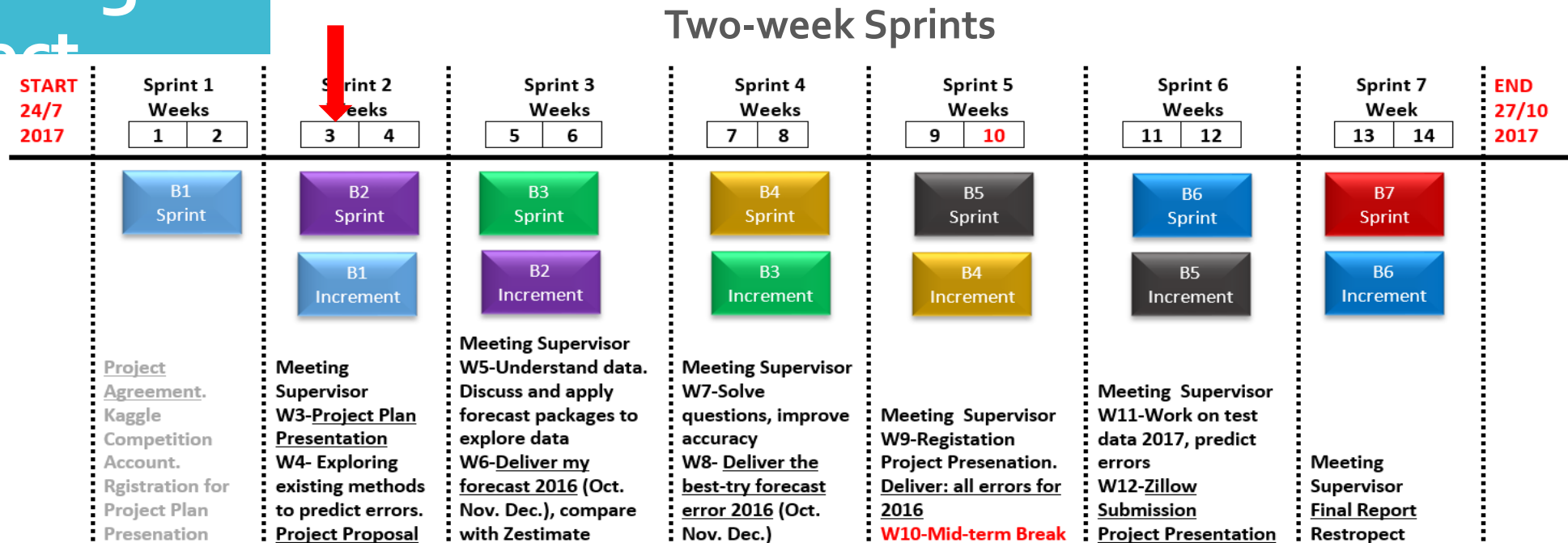
EST. MORTGAGE
\$4,461/mo 
[Get pre-qualified](#)

CONTACT
 Your n
 Phone
 Email
I am inter
NY 10031.

☐ I want f

Methodologies for Project

- Project Categories: A Development Project
- Project Methodologies:
Programming Tools: R and Rstudio
 Steps: 1. Define type of forecast analysis (Time Series)
 2. Find methods to analyse time series data
 3. Identify model to predict the future values
 4. Use the right packages and functions to manipulate the data
- Project Management Approach: Scrum

Complex products and complex environment



Constraints & Risk Management

Constraints (1)

Mitigation

Two final submission for Zillow only

Think and check carefully before submission

Risks (3)

Mitigation

Unavailable Supervisor

Alternative communication channels (Slack, Other group chats)

Loss of the work

Stable backup (GitHub)

Limited computing power of private Laptop or University computer

Discuss solutions with supervisor to find more powerful resources

THANK YOU!