

Forecasting Zestimate Error for Zillow:

A Data Analysis
IFN701 Project

Student: Linni QN (ng632981)

Supervisor: Dr Guido Zuccon



Project Overview



- Scopes:

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

- Zillow: Improve Zestimate model
- Zillow: Predict errors for 6 timepoints:
Oct. Nov. and Dec. of 2016, Oct. Nov. and Dec. of 2017
- Zillow: Practice on the training data through 2016 (90k properties)
3 millions of properties in 3 counties in California, USA (LA, OR, VE)
Test data of 2017 will be available on **October 2, 2017**
- Zillow: Round 1 deadline of submission: **October 16, 2017**

•

- IFN701: Project Proposal: **August 20, 2017**
- IFN701: Project Presentations: **Week 3, Week 12**
- IFN701: Project Report Submission: **October 27, 2017**

- Deliverables:

CSV file, R markdown file

Project Proposal, Project Presentations, Final project report

Project Motivation & Anticipated Significance

- Motivation

- Application of my educational skills
- Pursuit of financial data analysis
- Dr Guido Zuccon and his teaching team
- Kaggle prize US\$1,200,000 (round 1: \$50,000; round2: \$1,150,000)

- Anticipated Significance

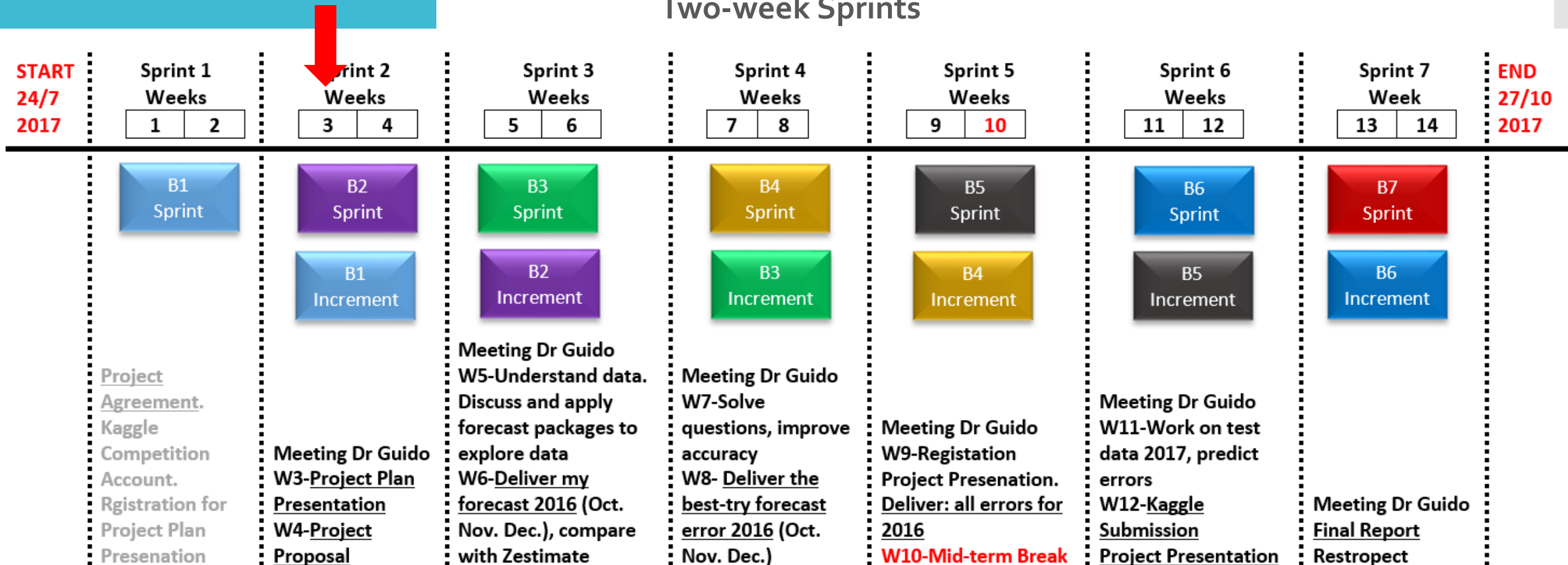
- 2006 – 2017 Zestimate Error Rate 14% → 5.4%

$$\begin{aligned} \$722,000 * 5.4\% &= \$38,988 \\ &(\$39,000) \end{aligned}$$

- Project Categories: A Development Project
- Project Management Approach: Scrum

Complex products and complex environment

Two-week Sprints



THANK
YOU!

Constraints & Risk Management

Constraints (4)	Mitigation
Submission will be evaluated on Mean Absolute Error only	Find the right forecast package with returning the relevant data for prediction accuracy reference
The log errors are accurate the 4 th decimal places	Adjust decimal formats
Accept CSV file for final submission only	Find the right way to generate the CSV file from R without any lost data
5 entries to the Kaggle forum each day and 2 final submission only	Think before doing
Risks (4)	Mitigation
Unavailable Supervisor	Contact the Project Coordinator immediately to check if any other backup resources
Lost of the work	Sufficient Backup (H Drive, Google Drive, GitHub etc.)
Sick or other personal urgent issues	Lost the chance to win the prize Apply for extension to the faculty
Poor/Congestive Network at last minute	Facilitate tasks as early as possible Multiple trials of file submission