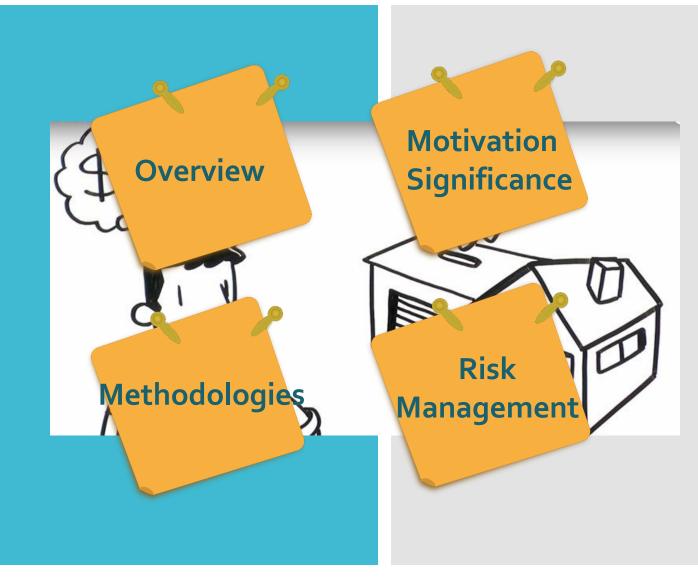
Forecasting Zestimate Error for Zillow:

A Data Analysis IFN701 Project

Student: Linni QN (n9632981)

Supervisor: Dr Guido Zuccon





Project

Overview

Scopes:

logerror = log(Zestimate) - log(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)
 3millions of properties in 3 counties in California, USA (LA, OR, VE)
 Test data of 2017 will be available on October2, 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

 \rightarrow 2006 – 2017 Zestimate Error Rate 14% \rightarrow 5.4%

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

		Complex	products and c	omplex environ	ment		
Two-week Sprints							
Sprint 1 Weeks 1 2	rint 2 Weeks 3 4	Sprint 3 Weeks 5 6	Sprint 4 Weeks 7 8	Sprint 5 Weeks 9 10	Sprint 6 Weeks 11 12	Week	END 27/10 2017
B1 Sprint	B2 Sprint	B3 Sprint	B4 Sprint	B5 Sprint	B6 Sprint	B7 Sprint	
	B1 Increment	B2 Increment	B3 Increment	B4 Increment	B5 Increment	B6 Increment	
		Meeting Dr Guido					
Project		W5-Understand data.	Meeting Dr Guido				
Agreement.		Discuss and apply	W7-Solve		Meeting Dr Guido		
Kaggle	Massing Dr Cuida	forecast packages to explore data	questions, improve	Meeting Dr Guido	W11-Work on test		
Competition Account.	Meeting Dr Guido W3-Project Plan	W6-Deliver my	accuracy W8- Deliver the	W9-Registation Project Presenation.	data 2017, predict errors		
Rgistration for	Presentation	forecast 2016 (Oct.	best-try forecast	Deliver: all errors for	W12-Kaggle	Meeting Dr Guido	
Project Plan	W4-Project	Nov. Dec.), compare	error 2016 (Oct.	2016	Submission	Final Report	
Presenation	: 	with Zestimate		W10-Mid-term Break	Project Presentation	Restropect	



Constrains & Risk Manageme

	Constrains (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 4th decimal places	Adjust decimal formats
5	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