Big Data HW1

Data mining
in real estate sale and media data

Instruction

- According to the case study, **formulate** a forecasting model.
- **Explain** and **Justify** your modeling procedures step by step
  - General statement of the problem
    - What is your definition of best allocation?
  - Collection of data relevant to the model
    - What is the dependent variable in your model ?
    - What are the independent variables ?
    - Do you include other sources of open data? And why?
  - Selection of the model
    - Formulate your forecasting model, be it linear regression model or other non-linear forms.
  - Statistical evaluation of the model
    - Propose the diagnostics used in your forecasting model.

Submission

- Form groups of 1 - 4 members for this homework and fill in the group members in hw1_group_sheet
- Note the group need not be the same as the group in final project
- The requirement is the same for each group

Submission

- Upload the file as follows to ceiba
  - ❏ <group_name>.zip (e.g. example.zip)
    - ❏ report.pdf
    - ❏ readme (optional)
- Upload by one group member is enough

Submission

- The evaluation will be based on model feasibility and depth of detail in the report
- Encourage literature survey and creative but reasonable model construction
- Due at 04/13/2015 22:00
- Late Penalty: 10% per day