Movie Revenue Prediction

Team Movie Badgers

Group Member:

Anna Huang, Chen Ren, Jingyun Chen

Junmeng Zhu, Weichen Xu

Agenda

Background

- **Motivation**
  - Factors affecting box-office
- **Objective**
  - Basic analysis of movie information
  - Prediction of movie's revenue
- **Features**
  - Data visualization
  - Machine learning
  - Graphic User Interface

Data Source

- **Two Databases**
  - **TMDB:** https://www.themoviedb.org/
  - "Revenue"," Budget" , "IMDB_key"
  - **OMDB:** http://www.omdbapi.com/
  - e.g. "Released Date", "Production", "Runtime", "IMDB_key" etc

- **API request**
  - Packages used: Requests
  - Order by: page(20), year(2006-2016), popularity

- **Limitation**
  - Time, resource, combination

Use cases

- **Movie fans/ Theater Owner**
  - Input the features of a specific movie they want to get its future revenue
  - Select a regression model
  - System presents its estimated revenue
- **Researchers interested in movie revenue prediction**
  - Import the submodules for independent analysis
  - Collect their own dataset based on self defined parameters
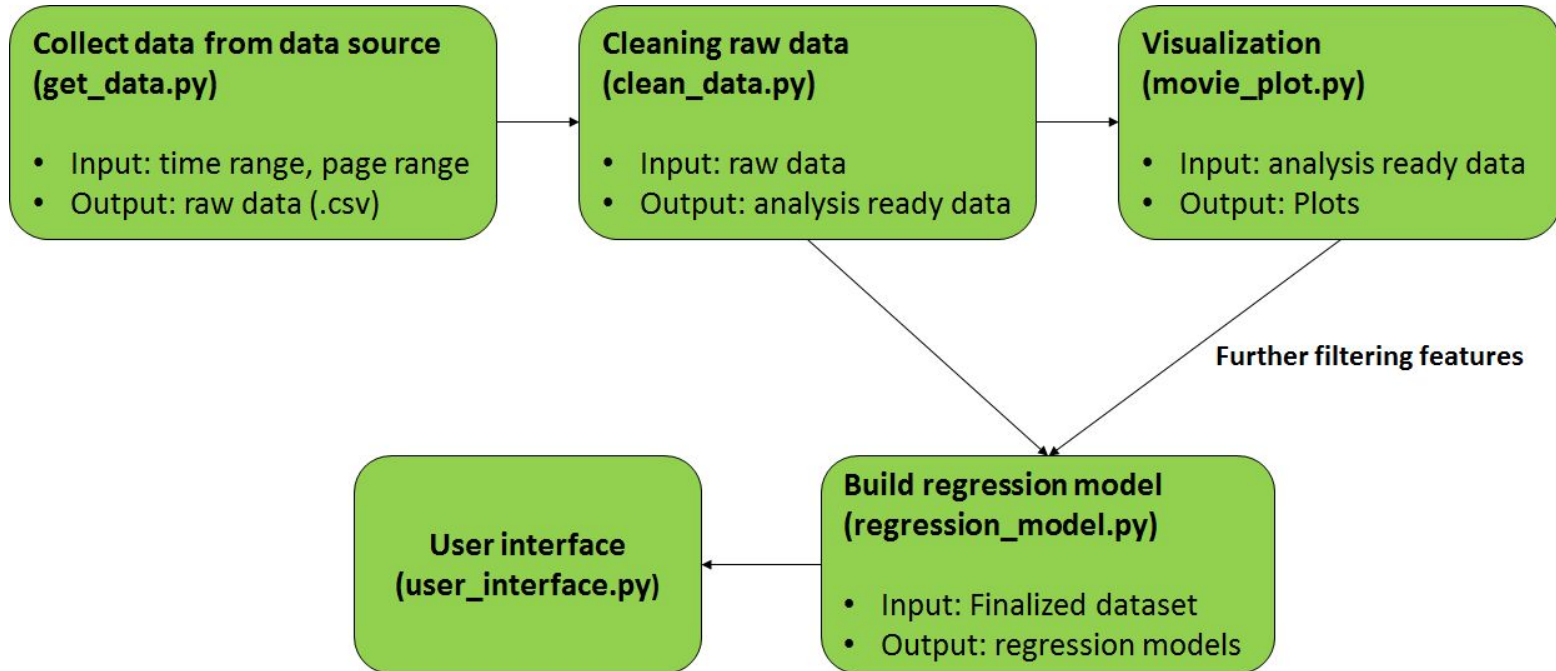 - Select features to build model using visualization tools

Design

- **Getting (get_data.py)**
  - Get movie information and profit data
  - Merge them into one dataframe
- **Cleaning (clean_data.py)**
  - Get actor and director average popularity
  - Processing data that ready for regression and analysis
- **Modelling (regression_model.py)**
  - Build and evaluate several models using cross validation
  - Save model as pickle file to local machine
- **Plotting (movie_plot.py)**
  - Scatter plot and box plot
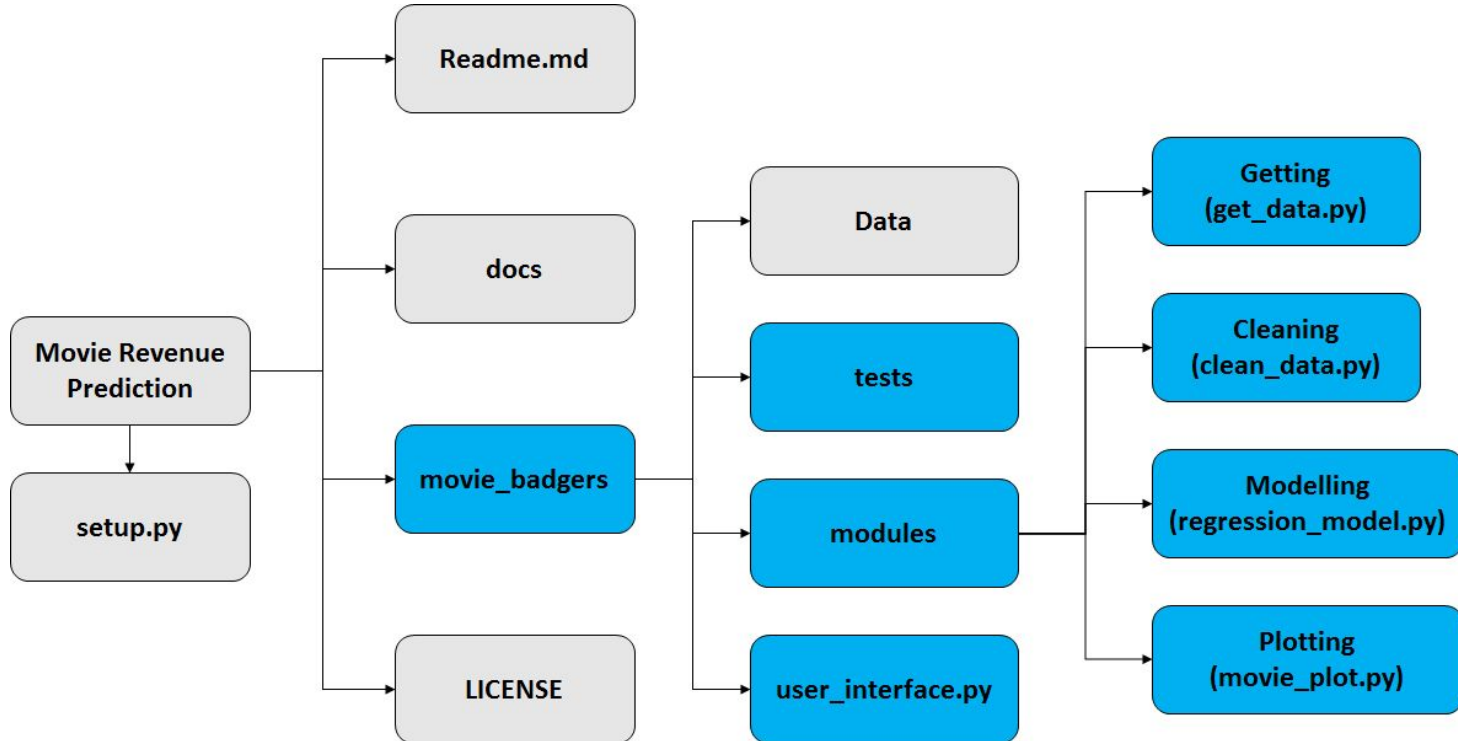- **UI (user_interface.py)**

Design component diagram

Collect data from data source (get_data.py)

- Input: time range, page range
- Output: raw data (.csv)

Cleaning raw data (clean_data.py)

- Input: raw data
- Output: analysis ready data

Visualization (movie_plot.py)

- Input: analysis ready data
- Output: Plots

Further filtering features

Build regression model (regression_model.py)

- Input: Finalized dataset
- Output: regression models

User interface (user_interface.py)

DEMO

Repository Structure

Lessons Learned & Next Steps

Lessons learned

- Python programming
- Git
- Version control
- Software design
- Programming styles
- Unit test

Next Steps

- Import user defined models into UI
- Different models for movie not released and movie already released (with opening weekend boxoffice)
- Better and more user friendly user interface

Thank you!