# Take-Home Exam: Final Project

Top of Form

Bottom of Form

This exam is a take-home project. Write your answers as a report (maximum 15 pages, including figures and tables; not the more the better).

**Walmart Store Sales Forecasting.**

In this recruiting competition, job-seekers (for Walmart) are provided with historical sales data for 45 Walmart stores located in different regions. Each store contains many departments, and participants must project the sales for each department in each store. The goal is to use historical data (training set) to predict store sales (test set).   
  
You are asked to analyze this data to answer the questions below. The data, the evaluation metric, and other relevant information can be found at [Kaggle](https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting" \t "_blank).

**What you need to include in your report?**

Section 1: Introduction   
Provide a brief introduction of the goal of this final project. What's it about? Where you get the data? What's the background of the data?   
  
Section 2: Data   
Provide a brief explanation/summary of variables you plan to include in your analysis. Include a couple of graphical displays of the data (don't plot all the data). Also comment on any patterns/characteristics of the data which you find interesting or anything relevant to your later analysis.   
  
Section 3: Method  
You are required to build three prediction models. For each model, include of a description of the methodology, and a description of the implementation if the implementatin is not trival.

* Section 3.1: Start with a simple model. For example, predict the weekly\_sales based on sales from the prior year. In my R code, I use the median of 3 adjacent weeks in the prior year. But you need some special treatment for holiday weeks, and also some stores/depts are in the test set, but not in the training set, i.e., we have no sale information for those stores/depts.
* Section 3.2: Build a linear regression model (or models) to predict the weekly\_sales.   
    
  For example, for each store, you can fit a linear regression model like the following:   
    
  weekly\_sales ~ as.factor(dept) + as.factor(holiday) + as.factor(year) + as.factor(month) + as.factor(wk) + temp + fuel + cpi + uemp ...  
    
  Or you can fit a linear regression model like the following for each department:   
    
  weekly\_sales ~ as.factor(store) + as.factor(holiday) + as.factor(year) + as.factor(month) + as.factor(wk) + temp + fuel + cpi + uemp ...
* Section 3.3: Use randomforest to predict the weekly\_sales.
* For each of the three methods, you should have a score after submitting your prediction on Kaggle. Report your scores in your report.
* Section 3.4 (Optional): You can also try some other methods.

Section 4: Discussions

**What you need to submit on compasst?**

Submit your

* report (in pdf),
* R code (in .R or .txt), and
* submission files in Kaggle required format (at least three)

on Compass (Assignment Dropbox) by midnight, December 19, 2014.

**Rules**

You are NOT allowed to discuss the exam with anyone else. If you have questions, please email the instructor or post your question on the discussion board.  
  
It is fine for you to use online resources. A good starting place will be the Forum section on Kaggle (start with the recent posts; old posts are not much useful since the participants were not allowed to have open discussions at that moment as this is a recruiting competition).   
  
I did some analysis of this data, and [here](https://compass2g.illinois.edu/bbcswebdav/pid-1355694-dt-content-rid-13531654_1/courses/stat_425_120148_120701/R_Project.html" \t "_blank) is my code. You may find the following three sets of R code useful too:

* [https://bitbucket.org/dthal/kaggle\_walmart/src](https://bitbucket.org/dthal/kaggle_walmart/src" \t "_blank)
* [https://github.com/mikeskim/Walmart](https://github.com/mikeskim/Walmart" \t "_blank)
* [https://github.com/ChenglongChen/Kaggle\_Walmart-Recruiting-Store-Sales-Forecasting](https://github.com/ChenglongChen/Kaggle_Walmart-Recruiting-Store-Sales-Forecasting" \t "_blank)

If you find some useful resources online, and want to share with your fellow classmates, although I don't know why you want to do that :), please post it on the discussion board.   
  
How to use online reources?

* You are NOT allowed to directly copy any sentences from others' work (paper, blog, or his/her post on the Forum). You have to either paraphrase or cite the source. Check some online websites on "how to avoid plagiarism".
* Can I use the R code I find on the web to finish this project? If you really cannot do the coding by yourself, you can "borrow" some online code. But then you have to have a section called "Acknowledgment" where you acknowledge the author (or authors) of the code. And of course, you'll lose some points.