Stat 6160

Course Project Info

Xiwei Tang
Department of Statistics, University of Virginia

General Info

- Each project can be completed by **a team with up to three students**
- The goal of class project is to apply the skills we learned from this class to analyze a real data problem.

- General requirement
  - **Final Written Report**: 10 pages limit (not include graphs and tables). **<u>Due on Monday, May, 9th</u>**
  - **Presentation slides** (for non-individual teamwork only, no presentation!): 10 slides limit. **<u>Due on Monday, May, 9th</u>**

Submission

- Please indicate whether you are working individually or in a team on proposal
- Every team member shall submit a copy to Collab
- Both final project report and presentation slides (optional) will be submitted online (Collab) electronically.
- **Due at 11:59 pm, Monday, May 9th.**

Datasets

- **Self-collected Data!!!**

- **Kaggle**: https://www.kaggle.com/competitions

- **Webscope**: https://webscope.sandbox.yahoo.com/

- **UCI ML Respiratory:** http://archive.ics.uci.edu/ml/index.php

- **Google datasets**: https://datasetsearch.research.google.com/

Outline of Project

- PART I: Problem Description:
- PART II: Experiment Analysis
- PART III: Statistical Analysis
- PART IIII: Conclusion of your results

Some Details

- Experiment Analysis:
  - 1) Whether the current experiment is well designed? Any potential issues?
  - 2) How to deal with each factor? Quantitative or Categorical
  - 3) Any data splitting or integration? Transformation? ….
  - 4) ….

- Statistical Analysis (not limited to models covered in this course):
  - 1) Fast screening on factors, e.g., 2^k factorial analysis
  - 2) Identify significant factors, e.g., ANOVA or linear regression model
  - 3) Model diagnosis and assessment
  - 4) Predictive Power
  - 5) ….