

**UNIVERSITY of WASHINGTON**  
DATASCI 450: Deriving Knowledge from Data at Scale (4694)

Aleksey Kramer

[Home](#) ► [DATASCI 450 4694 Au15](#) ► [Course Overview](#) ► [Project Part1](#)**Administration**[Course administration](#)[My profile settings](#)**My Course Help**

- [My Course Help](#)
- [Contact Our Support Staff](#)

**Project Part1**

Class Project:

This class project worth 40% of your course grade. Select one of the designated Kaggle competition projects, or a non-Kaggle project with instructor's approval. The emphasis of this course is practicing data science, not theoretical understanding of methodologies. The project should have enough depth and breadth to illustrate your analytic skills and your understanding of statistical/machine learning methodologies. You could form a team or work alone on this project. I highly recommend you consider the Kaggle projects because they have been vetted and have some reasonable structure to problem definition.

The project will be submitted in 3 parts.

Due on Nov 16

Part 1: Define the objective and scope of the project. Gather and organize the data for the project.

1. a) Conduct exploratory data analysis such as visualizing the data through graphs, tables, summary statistics, and other means to understand the data.

1. b) Identify any issues associated with data gap, data size, data type, data manipulation, data storage and data retrieval for analysis.

**Navigation****Home**

Current course

[DATASCI 450 4694 Au15](#)[Participants](#)[Course Overview](#) [Latest News and Announcements](#) [General Discussion Forum](#) [Lecture 1](#) [Lecture 2](#) [Quiz 0 - Lecture 1](#) [Lecture 3](#) [Homework 1 - Lecture 3](#) [Quiz 1 - Lecture 3](#) [Lecture 4](#) [Homework 2 - Lecture 4](#) [Quiz 2 - Lecture 4](#) [Lecture 5](#) [Homework 3 - Lecture 5](#) [Lecture 6](#) [Homework 4 - Lecture 6](#) [Project Part1](#) [Quiz 3](#) [Lecture 7](#) [Homework 5 - Lecture 7](#) [Lecture 8](#) [Lecture 9](#) [Lecture 10](#) [Final Course Project](#)**My courses**

Structured or unstructured data?

1. c) Describe the high level analytic problem needs to be resolved: supervised learning, unsupervised learning.

Due on Nov 30

Part 2: Model construction and evaluation

1. a) Construct analytic model(s) to address the project objective
1. b) Evaluate the model outcomes
1. c) Iterate and improve the model when necessary
1. d) Justify the final model and its output

Due on Dec 7

Part 3: Document the findings

1. a) Write a report summarizing the previous two parts. Clearly summarize your steps and your analytic process. Don't just provide screen shots of algorithm outputs.
1. b) Compare your results with the Kaggle score board if you select a Kaggle project

You will be graded on your contribution to this project. Please indicate your work clearly in your team report so that credit will be given to the individual. At times, it is hard to separate a tight collaboration. Hence, team members would share the credit. Please indicate share credit in your submission.

**Submission status**

Submission status	No attempt
Grading status	Not graded
Due date	Monday, November 16, 2015, 11:55 PM
Time remaining	5 days 3 hours
Last modified	Saturday, November 7, 2015, 8:39 AM
Submission comments	<a href="#">▶ Comments (0)</a>

[Add submission](#)

Make changes to your submission

©2009-2015 UNIVERSITY OF WASHINGTON. ALL RIGHTS RESERVED.

MY COURSE IS AN ONLINE LEARNING SYSTEM FOR UW COURSES OFFERED THROUGH OR IN PARTNERSHIP WITH UW EDUCATIONAL OUTREACH.

NO PART OF THIS PUBLICATION MAY BE REPRODUCED IN ANY FORM OR BY ANY MEANS WITHOUT PERMISSION IN WRITING FROM THE PUBLISHER.