DSC 423 Class Project  
Milestone 5: Project Proposal

The project is well under way! You should note that nearly half the project is simply getting ready to begin your model building. This is not uncommon in real-world projects. Now that you have selected your dataset, processed it and performed your first exploratory analysis, it is now time to formalize a plan for the remainder of the quarter. Your deliverable for this milestone is a detailed strategy for the upcoming weeks. We will go over this strategy together in a 30-60-minute interview (M06: Midflight Interview).

Let’s begin by what \*not\* to do. Do not divide the project into a series of steps. At no time should anyone be relying on someone else to complete their work. These are meant to be individual milestone, an opportunity for you to demonstrate the skills you have learned in the class. You should divide the project into parallel but distinct lines of work.

There are two common strategies I often recommend. First, you can divide the project “by rows”. If you are predicting the sales price of condos in Chicago, one student might predict prices on the South Side, while others predict prices on the West Side, North Side or in the Loop. A fifth student might use all the data.

Another strategy is to divide the project “by columns”. If you are predicting the GRE scores of high school students, one member of the group might focus on their scholastic data. Another might focus on their socioeconomic data and a third might focus on their family history. A fourth member of the group might use all the data to predict GRE scores.

There may be other ways to divide the project, but those are the most common. I am willing to consider others, but I look for these specific qualities…

* Everyone must have a clearly stated and actionable goal. I do not want anyone wondering what they are expected to do.
* Everyone’s goal must be independent of the other group members. I do not want anyone hamstrung, waiting on the results of another group member.
* The individual milestones must be related. This will allow you to share code, discuss obstacles, and identify insights.

One member of the group should start a Google Doc. Give all group members access to the document. Using the Google Doc, the group should \*\*collaboratively\*\* prepare a strategy that includes...

* Paragraph 1: The context (e.g. “Diabetes is a serious problem for millions of Americans. It is…”)
* Paragraph 2: The specific problem (e.g. “One of the effects of diabetes can be blindness as the result of diabetic retinopathy, swelling or other changes in the ….”).
* Paragraph 3: The pitch (e.g. “This project will investigate diabetic retinopathy through regression analysis, specifically we will predict…”).
* Paragraph 4: The data (e.g. “To facilitate this exploration into diabetic retinopathy, we will use a dataset provided by the Centers for Disease Control…).
* Paragraph 5: Goals (e.g. “Using the CDC data we hope to answer these specific questions: 1) What are the…”).
* Paragraph 6-10: Individual Efforts. For each group member, describe the work they are going to perform. (e.g. “We believe diabetic retinopathy may present differently for different ages and sexes. Males over 40 years shower higher rates of diabetic retinopathy and are less likely to seek treatment. Rob Smith will build models focused on this group. In particular, he will...”)
* Paragraph 11: Conclusion. (e.g. “In sum, retinopathy caused by diabetes is a serious health risk impacting millions of Americas. This study will evaluate the predictors of diabetic retinopathy and will look at specific groups defined by both age and sex…”)

The result should be about three pages. Include the group name and members on the top of the page. Submit a **PDF** version of the report to the group’s D2L dropbox. Notice that this is a group submission; only one member need submit the report. It must be a PDF. Email me questions, but please do not wait until the last minute.

Once the plan of attack is prepared, your group should schedule a 60-minute interview with me so we can go over your plan in detail (M06: Midflight Interview). I will post my availability through the news widget. Email me to schedule your interview. In your email…

1. Include the group name in the title of the email.
2. Include the day/time your group would like to meet.
3. Make sure each group member is cc’d on the email.
4. Please note that my schedule is very busy and while I can push things around to create the time to go over your project, I would appreciate it if you could schedule things as early as possible

**REMEMBER that the peer review forms are also due this week. They must be done individually. They are confidential. Please upload them to the D2L’s submission folder.**