The lecture slides for the first lecture: https://docs.google.com/presentation/d/1-jkoseBpWyCS-Kepw_uJi3tV_EGZj0QEZ_mFRSqnFbw/edit?usp=sharing

The course is structured into 6 main lessons. Other than this lesson, the premise of each lesson is considering the ideal data science experience and then discussing what makes it ideal and more realistic settings.

(lesson 0) Introduction to the course

1. The data pull is clean
2. The experiment is carefully designed, concepts
3. The experiment is carefully designed, things to do
4. Results of analyses are clear
5. The decisions are obvious
6. The analysis product is awesome

Let's drill down into each lesson to discuss the content.

Introduction to the course. In this lesson, we give an overview of the course, the premise and an outline.

1. The data pull is clean. In this lesson, we discuss the ideal setting where creating the analytic data set goes perfectly. Since this never happens, we give some strategies for managing the creation of analytic data sets. The first is creating summary tables with means, medians, other quantiles and standard deviations. The second is looking at diagnostics from regression output. The third is a strange little fact called Benford's law. The final suggestion is to perform sampling based data quality queries.
2. The experiment is carefully designed, principles. In this lesson, we cover experimental design and what happens when we have observational data. We also discuss causality and randomization's role in uncovering it. Then we move on to discuss confounding, a central problem of observational studies and adjustment, the first line of defense for confounding.
3. The experiment is carefully designed, things to do. OK now that we know some of the concepts we can talk about things to do. First, we discuss A/B testing, the adjustment for confounding. We also discuss methods for combating sampling bias.
4. The results of analyses are clear. In this lesson, we cover issues when the results of analyses aren't easy to understand, especially focusing on hypothesis tests. First, we broadly discuss the need to consider more than hypothesis tests. Then we consider issues and strategies with testing and statistical results. Next, we discuss multiple comparisons and their role in hypothesis testing. Then we suggest comparing results of unknown phenomena with that of known. Finally, we discuss negative control analyses.
5. The decision is obvious. There are many reasons why decisions are not obvious. First, the data can be very equivocal about the hypotheses. We discuss power as it relates to this issues. Secondly, even in the presence of significant results, one may not have measured the right quantities.
6. The analysis product is awesome. In this module we discuss the final report. Here we make two recommendations. The first is enforcing reproducible of analysis documents. We suggest tools for doing this. Secondly, we suggest version controlling the software and give recommended solutions.

Throughout, we focus on managerial topics. That is, high level conceptual issues, or practical recommendations that can be made to data scientists.