Read the scenario below about a data scientist’s self-proclaimed failed project (based on <https://www.youtube.com/watch?v=FW2dqQksRwo&feature=youtu.be>, “PyData Ann Arbor: Jacqueline Nolis | When Data Science Projects Fail”).

**#1: The Adler Alert (start at 13:07 in video)**

The data scientist was working for an eCommerce company. Everything customers bought from the company was bought through their website. One day, the marketing department noticed that there were significantly lower sales than expected. They started to suspect there might be a bug in the website that was interfering with sales, so they called the software engineering department to see if there was a problem. There was, and the engineers fixed it. By the time the bug was discovered and patched, the bug costed the company millions of dollars.

In response to this fiasco, a marketing executive asked the data scientist to create a tool that would show sales from the same day of the week from the last 4 weeks, so it would be easier to see if there was another website problem in the future.

Rather than just report the raw sales numbers from the past 4 weeks, the data scientist suggested that she create a more complete time series model of what sales should be on a given day, taking into account factors like growth over time, seasonality, holidays, etc. She would then use this more complete model to create a tool that flagged potential issues the marketing department should be aware of. She did a really good job inspiring the executives with this idea, so they supported her project, and even gave her a team of people to help her implement it. They decided that they would extend their efforts to not only flag issues with sales, but also issues with 20 other key performance indicators (KPIs), including metrics like revenue, average order value, and conversion rates. They also decided they would try to tailor their predictions for different regions of the company. They reported “flags” through an excel dashboard like the one displayed in her slide below (each row represents a day, and each column represents a KPI).



Despite the fact that the dashboard was completed on time, ultimately, the project was scrapped. Overall, the data scientist considers the project a failure.

Imagine that the data scientist now wants your advice about what she could have done better to help the project succeed and to help its outputs be put into production. As a team, please prepare answers to these questions to share with the class:

* What was the original problem the company had? What problem did the data science team end up solving?
* Were there other stakeholders that might have been able to help solve the company’s problem better?
* Who was the intended audience for the tool the data scientist created, and do you think that was the correct audience to target?
* Why didn’t the dashboard they used get adopted?
* What communication strategies could the data scientist have tried or improved at different stages of the project to make it more likely the project would ultimately be considered a “success”?

Read the scenario below about a data scientist’s self-proclaimed failed project (based on <https://www.youtube.com/watch?v=FW2dqQksRwo&feature=youtu.be>, “PyData Ann Arbor: Jacqueline Nolis | When Data Science Projects Fail”).

**#2: Full Customer Analysis for a Retail Company (starts at 17:50 in the video)**

The data scientist was consulting for a retail company that had many years of “Big Data” they had never mined before. The data they had included things like purchasing records, in-store surveys, and marketing metrics (ex: email click-through rates), but a lot of the data was stored in disparate places, preventing it from being analyzed holistically or all at the same time, especially because they didn’t have any data scientists on staff. When the company hired the data scientist as a consultant, they tasked her with using data science tools to find something useful in the data that would improve the business. The data scientist enthusiastically accepted, thinking: “Yes, that’s what I do! I find relationships in data!”

The data scientist did a great job combining the disparate data sets, and implemented a lot of impressive data wrangling skills to get the data ready for effective analysis. During the analysis phase, she found many interesting relationships and patterns that she presented in a compelling presentation to the executives of the company. Overall, the data scientist was very proud of her work. Nonetheless, the executives were not happy at the end of her final presentation. Despite the fact that the results the data scientist presented required a lot of technical skills the company had never used before, the executives already knew the “story” of her results from other types of analyses the business did on a regular basis. The data scientist didn’t present any patterns the executives didn’t know about already. In addition, when the executives asked “Ok, we already knew about that…what should we do about it?”, the data scientist didn’t have any recommendations prepared. Overall, the executives concluded the project was a failure.

Imagine that the data scientist wants your advice about what she could have done better to help the project succeed and to help the executives feel the project had a positive impact on their company. As a team, please prepare answers to these questions to share with the class:

* What resources in our course videos or readings seem most relevant to this data scientist’s story? I have a specific name for these types of projects…what do I call them?
* What was the original problem the company had? What problem did the data scientist end up solving?
* According to the YouTube video, the lesson the data scientist took away from the experience was that the problem with the project wasn’t science, the report, or how she went about it, but rather, the fact that the company didn’t have a clear objective. Do you agree?
* What communication strategies could the data scientist have tried or improved at different parts of the project to make it more likely the project would be considered a “success”?

Read the scenario below about a data scientist’s self-proclaimed failed project (based on <https://www.youtube.com/watch?v=FW2dqQksRwo&feature=youtu.be>, “PyData Ann Arbor: Jacqueline Nolis | When Data Science Projects Fail”).

**#3 Automating QA testing (starts at 20:36 in video)**

The data scientist was consulting for a very large company that released numerous software apps. The company employed a huge team of people to test the apps to ensure good quality control. These quality assurance (QA) specialists would run hundreds of tests on each app before they were released. The data scientist was hired jointly by the company’s quality assurance team (who ran the tests) and the company’s data science team to help with a project that was already in progress. The goal of the project was to use digital logs from all the ways people interacted with the app and the way the app interacted with hardware to automate testing so that the company would no longer need so many humans to do the app quality assurance testing.

Just a couple of days after starting the project, the data scientist started to worry that the project was doomed. The data logs from the apps recorded things like when buttons were pressed and when an app successfully connected to a server, but she couldn’t see a reasonable way those data could be used to replace the tests, which assessed more abstract things like whether or not the menus used appropriate grammar. As she anticipated, the models the project team created to predict test performance simply were not accurate enough to replace humans. The data scientist felt that it was clear the project was not likely to be successful because there wasn’t enough relevant signal in the logged app data for the predictive models to ever be accurate enough, but neither the data science team nor the quality assurance team wanted to end the project early because they didn’t want to appear pessimistic or like they weren’t team players. The project went on for a full year before it was finally abandoned. The data scientist concluded that the lesson she should learn from this experience is that “A project can fail because of poor leadership”.

Imagine that the data scientist wants your advice about what she could have done better to help the project succeed, or at least to have helped end the project sooner. As a team, please prepare answers to these questions to share with the class:

* What was the original problem the company had? Can you imagine ways to modify the goal of the project that would still help address the problem, but that would be more tolerant of imperfect predictive models?
* Were there other stakeholders the teams could have involved that might have been able to help address the problem?
* What communication strategies could the data science team have tried or improved at different phases of the project to make it more likely the project would be considered a “success”?
* Do you agree with the data scientist’s conclusion that the main lesson she should learn from this experience is that “A project can fail because of poor leadership”? Why or why not? Are there other lessons to be learned, or alternative conclusions you might want to draw? If you agree with her conclusion, what should the leaders of the teams have done differently?

Read the scenario below about a data scientist’s self-proclaimed failed project (based on <https://www.youtube.com/watch?v=FW2dqQksRwo&feature=youtu.be>, “PyData Ann Arbor: Jacqueline Nolis | When Data Science Projects Fail”).

**#4: Kickstarter Prediction (starts at 25:08 in the video)**

A data scientist decided to do an exciting project on her own, because she thought it sounded fun. She found a website that analyzed data from kickstarter (a company that allows people to raise money for an idea or project through crowdsourcing) to predict whether a kickstarter campaign was likely to meet its fundraising target by its stated deadline. She determined that the website’s prediction was based on a simple algorithm that assumed the rate of revenue generation would be constant over time. She hypothesized that this assumption might not be correct, and that instead, there might be a big “bump” in fundraising at the very beginning and the very end of a campaign. She suspected the website’s algorithm might be dramatically more successful if these hypothesized phenomena were taken into account, so she emailed the owner of the website and asked if the owner would share his/her data with her so she could see if she could improve the algorithm. To her surprise, the owner agreed!

After weeks of effort, the data scientist determined that she couldn’t improve the model’s performance in any meaningful way. Her hypotheses were not supported by the data, so she made the model more and more complicated in order to try to incorporate modified, more nuanced hypotheses about possible “bumps” in fundraising. Nothing she tried made much of a difference in the model’s accuracy. Further, she developed the model in R, but didn’t know how to create an API in R, so the model she implemented was not easy to integrate into the owner’s website. When she finally decided to end the project, she sent her model and code to the website’s owner, but never heard back. She deemed the project a failure and concluded, “Be prepared for models not to be accurate. You never know when there isn’t a signal in there.”

Imagine that the data scientist wants your advice about what she could have done better to help the project succeed. As a team, please prepare answers to these questions to share with the class:

* What problem was the data scientist trying to solve? If the original model turned out to be reasonably accurate, why do you think the original approach had similar results to the data scientist’s more complicated approach?
* Although we haven’t discussed common logical fallacies or psychological pitfalls that influence data stories yet, how could the data scientist have improved the reasoning she used to guide her project?
* Even if she couldn’t improve on the website’s model, are there any other features she could have considered adding that would still provide value?
* What communication strategies could she have tried or improved at different phases of the project to make it more likely the project would be considered a “success” in either her eyes or the website owner’s eyes?
* Would you consider this project a failure? Why or why not?

Read the scenario below about a data scientist’s self-proclaimed failed project (based on <https://www.youtube.com/watch?v=FW2dqQksRwo&feature=youtu.be>, “PyData Ann Arbor: Jacqueline Nolis | When Data Science Projects Fail”).

**#5 Retail Company Coupon Simulation (starts at 28:52 in the video)**

A data scientist was working for a large retail company. Like most retail companies, the company found it challenging to understand and quantify the effects of its marketing techniques, like offering coupons. The reason why the effects were difficult to quantify is that it is hard to measure and account for all the factors that influence a technique’s success. For example, the data you receive is often from a biased sample (for instance, in the coupon scenario, the people who use coupons tend to be customers who would buy from the store anyway). As another example, it is hard to track what marketing techniques competitors are using at the same time, so you don’t know what environmental variables to take into account.

Interestingly, it is the financial team in a company rather than the marketing team that is usually responsible for decided whether a proposed marketing strategy is financially viable, and therefore whether it can move forward. The financial team usually makes their determination by creating a model in Excel using certain assumptions they have researched. Upon learning this, the data scientist thought: “Surely we can do better than these Excel models!” To improve a company’s marketing modeling efforts, she decided to use machine learning to simulate a customer’s purchasing behavior and predict what each customer would order next, and when. Then she estimated the effects of a given marketing technique, and aggregated the effects of the marketing technique across all of the simulated customers. She turned this approach into a product that companies could buy in order to predict the effects of a marketing campaign over the course of many months. To her delight, many companies did buy the product. However, she was stunned to experience that after using the product, all of the customers would eventually ask her to translate the results of the model into a simple excel formula they could share with the finance department. She felt this request defeated the purpose of her efforts, because her entire goal was to avoid overly-simplistic models in Excel! The data scientist concluded “The clients didn’t want a true prediction of their return on investment. They wanted the smallest amount of work necessary to prove to the finance department that their proposed marketing strategy was a financially viable idea. In the end, all they wanted me to do was prove the company wouldn’t go bankrupt as a result of their marketing idea.”

Imagine that the data scientist wants your advice about what she could have done better to help the project succeed. As a team, please prepare answers to these questions to share with the class:

* What was the original problem the company had? What problem did the data scientist end up solving?
* Why did the data scientist end up giving up on the tool, even though companies were paying for it?
* What measures of success did the stakeholders use for the project? What measures of success did the data scientist use for the project?
* Were there other stakeholders she could have talked to that might have been able to help her solve the company’s problem better?
* What communication strategies could the data scientist have tried or improved at different parts of the project to make it more likely the project would ultimately be considered a “success”?