LIFECYCLE OF A
DATA SCIENCE PROJECT

STAGE #1

Identify The Problem

Ever heard the phrase "Here's some data, can you find some insights?" Right? Too often stakeholders approach Data Scientists with vague or even undefined goals. Understanding the end goal is very important and sets up the rest of the project for success. This phase takes up about 10% of the time in the project's lifecycle.

STAGE #2

Prepare the Data

By far, everybody's least favourite stage, but perhaps the most important one. Data can come from many sources, be in the wrong format, have anomalies and a myriad of other problems. A single mistake in this stage can render the rest of the analysis useless. That's why typically, up to 70% of the time is spent here.
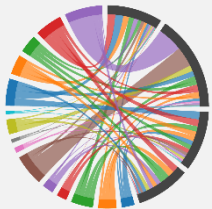
STAGE #3

Analyse the Data

Creating models, performing data mining, running text analytics, setting up simulations – the list goes on! This is the most fun and exciting part and if the previous stages have been done correctly, analyzing the data and deriving insights will feel like a breeze. Time consumption: 10%

STAGE #4

Visualize Insights

Visualizing comes hand-in-hand with analyzing. This is a very powerful technique as seeing the data in various forms and shapes can help uncover insights that are otherwise not evident. Also some projects such as BI dashboards don't require much analysis but rely heavily on visualization instead. Time consumption: 10%

STAGE #5

Present Findings

We've reached 100% the project is over! Actually, not yet. Presenting findings is a whole separate "Bonus" stage. You need to not only convey the insights in your audience's language but also get buy-in from them to take action based on those insights. This is an art in its own right. Time consumption: extra 80%

Join The SuperDataScience Facebook Group And Improve Your Career