My Project

- My project involves research and statistical analysis of the data produced by a number of internal tools.
- The final goal is to identify patterns and relationships between various metrics and gain insight into projects’ life cycles.
- As part of my internship, I will be presenting my research at a summit in Mountain View, CA to approx. 500 Google engineers.
- By the end of my internship at Google, I am expected to write and publish a research paper on my findings.

What I learned (so far)

- Big Data is noisy and difficult to analyse. Smoothing and filtering is almost always required to get meaningful results.
- Processing gigabytes of data at a time is a non-trivial task. Must parallelise for computation to finish in finite time.
- When carrying out research hypotheses must be well defined from the very beginning to avoid confirmation bias.
- With statistical analysis, graphing the results helps to verify the correctness of your data.
- It is worthwhile to spend some time on code design before starting. This way you will write cleaner code.

Technologies Used

- Golang
- MapReduce
- Time Series Analysis
- Python
- Bigtable
- FlumeJava

What does Google do?

- Widely known for its search engine, Google is multinational tech company with a broad range of products such as Gmail, Youtube, Google Chrome and Android.
- Google’s mission is “to organize the world’s information and make it universally accessible and useful.”
- With offices all around the world I got the chance to work in Zürich, Switzerland on a team focused on Engineering Productivity.
- Engineering Productivity is about building tools and infrastructure to make Google’s 2 billion lines of code cleaner and more robust.