



Klaus Paranko

Data Engineer & Analyst

I recently graduated with a master's from Aalto University, majoring in Finance and minoring in Machine Learning, Data Science, and Artificial Intelligence. As my studies progressed and we had our first programming course, I realized I really enjoy programming and computer science. From that point onward I directed my studies towards data science, and started self-learning everything about computers and programming.

I enjoy being able to bring order into chaos by taking messy and complex raw data and turning it into understandable clean data. I also enjoy discovering the patterns underlying the data, which often open interesting and useful windows into real world phenomena. In programming I enjoy taking a process, figuring out the best way to do it once, and from there on letting the computer do it for me through automation. I am also fascinated by the magic of how everything we use on a computer, no matter how complex, eventually boils down to simply zeroes and ones. I use linux as my daily driver and enjoy using the command line. I use neovim as my editor of choice, although I am happy working in any environment (eg. the Snowflake/Databricks UIs)

As a person I am open-minded and tend to get along with everyone. I enjoy different cultures, and have lived in Vietnam and Australia. I am always learning about one subject or another, whether it is through youtube videos, books, or online courses. I am also an avid gym-goer, and have been doing so since I was 15 years old.

Contact

Phone
+358 4036681

Email
kparanko@gmail.com

LinkedIn
<https://www.linkedin.com/in/klaus-paranko/>

Github
<https://github.com/KlausParanko>

Expertise

- SQL
- Python
 - PySpark
 - Pandas
 - Numpy
 - Matplotlib (Seaborn, Plotly)
 - PyTorch
 - HuggingFace
- Databricks
- Snowflake
- Power BI
- Git/Github
- Linux

Language

Finnish
English

Education

MSc Business and Administration, Aalto University

Major: Finance

Minor: Machine Learning, Data Science, and Artificial Intelligence

Grade Average: 4.53/5 (Degree Completed with Honors)

2019-2023

For my thesis I gathered a novel dataset of company 8K reports using Python and applied state-of-the-art deep learning algorithms to find relationships between the content of the reports and the reporting companies' stock price movements. Gathering the dataset included automating making requests for the 8K reports, cleaning the data, and parsing the reports (which were originally intended for human consumption) into a structured form that could be statistically analyzed. Afterwards, I gave the dataset to a phd student of my advisor, who asked if he could use it for his own research. The grade I got for the work was 5/5.

BSc Business and Administration, Aalto University

Finance

Exchange in Vietnam

Grade Average: 4.33/5

2016-2019

For my bachelor's thesis I studied whether attention companies receive as measured by the amount of Google searches correlates with their price movements. I used Python to gather the data from the Google Trends API and matched this to price data of companies from CRSP. The grade I received was 4/5.

Projects

Finnish Tech Market Analysis

Stack: Python, Databricks (Spark, Delta Lake), and Power BI

The purpose for this project is to get know Finnish tech landscape from my perspective as a job seeker. I do this by analysing both tech company financials and job postings for data engineering roles. The questions I answer from the financials dataset is which companies are financially profitable, which ones are least risky regarding their debt amount, and which ones are hiring (have a increasing employee count). From the job postings I analyze which technologies (eg. Databricks, Power BI) have the most demand from data engineering recruiters. I also combine the datasets to rank the job postings by the financial success of the recruiting company.

I get the financials from the Orbis database and the job postings by scraping LinkedIn jobs by creating a Python web scraping script. I parse both datasets into a structured form using python. I then upload these to a database in Databricks. To answer my questions I analyze the datasets using PySpark and create data visualizations using Power BI.

My Financial and Fitness Data

Stack: Python, Snowflake, and Power BI

I look at two things in this project: my financials (money coming in/going out) and my fitness (calories coming in/going out, amount of sleep). The purpose is to analyze the data and see if there are things I am spending a lot of money on that provide me little value, or if I am eating things that are not that enjoyable but contribute a lot of calories. I also look at the interplay of my health and financials, like do I spend more money on days I haven't slept well. I also want to automate this as much as possible, so that it is easy to look at on a monthly basis.

I get my financials from downloading my transaction log from Nordea and parsing the details with Python. I get my calorie consumption and sleep amount from my Fitbit wristband from Fitbit's website as a csv. I get my calorie consumption from my calorie diary app FatSecret as a csv. I parse both financials and health data with Python. I store the data in a Snowflake database. I analyze the data with SQL queries and visualize the results with Power BI.

Jobs

Research Assistant, Tampere University of Technology

Summers 2017 and 2018

Performance Review Grade: Excellent

Reference: Teemu Laine, Professor of Industrial Engineering & Management