

# JOHANNES MÄKINEN

[joh.makinen@gmail.com](mailto:joh.makinen@gmail.com) | [linkedin.com/in/johmakinen](https://linkedin.com/in/johmakinen) | [johmakinen.github.io](https://johmakinen.github.io) | Helsinki, Finland

## WORK EXPERIENCE

### Data Analyst • Terveystalo

April 2020 – Present

- Did my M.Sc. thesis: "*Estimating the Effectiveness of Depression Care Pathways: A Process Mining and Survival Analysis Approach*". Insights from this are predicted to produce a 1.5M€+ yearly profit increase.
- Was in charge of implementing and testing decision logic for customer steering rules, which led to a yearly profit increase of 500K€.
- Created and maintained weekly and monthly customer traffic reports using *SQL*, *Qlikview* & *Piwik*. Around 200 people used the reports to find inefficiencies and targets to improve on.
- Performed K-means clustering and time series analyses for finding most profitable marketing campaigns to run.
- Helped increase self-service levels across all customer segments, which led to increased profit margins.

### Research Assistant • Aalto University, Systems Analysis laboratory

Jun. 2019 – Sep. 2019

- Did my B.Sc. thesis on "*Estimating the protection provided by islands against anti-ship missiles*" [🔗](#)
- The thesis was involved in a larger study on enhancing combat simulation models with adversarial risk analysis.

## EDUCATION

### M.Sc. (Tech.) Mathematics and Operations Research • GPA 5/5 • Aalto University

Sep. 2020 – Dec. 2023

- Minor in Machine Learning & Data Science
- Emphasis on optimization, statistical inference and mathematical modelling

### B.Sc. (Tech.) Mathematics and Systems Sciences • GPA 4.8/5 • Aalto University

Sep. 2017 – May 2020

- Minor in Computer Science

## SKILLS

- |                |                         |                         |                     |
|----------------|-------------------------|-------------------------|---------------------|
| • Optimization | • Hypothesis testing    | • Times series analysis | • Survival analysis |
| • Clustering   | • Regression techniques | • Bayesian inference    | • Classification    |

## TECHNICAL SUMMARY

**Languages:** Python, R, SQL, C

**Technologies:** Git, Jupyter, Docker, MS Office

**Data Science:** Pandas, Numpy, Scikit-learn, Lifelines, XGBoost, Matplotlib, Plotly, Selenium, Tidymodels, ggplot2, Flask, Cython

## PROJECTS More projects and info at [johmakinen.github.io](https://johmakinen.github.io)

### Data imputation app [🔗](#) | Python, Streamlit, Scikit-Learn, XGBoost, pytest

- Streamlit app that imputes missing values for a user-given dataset.
- Clear UI and easy to use, works for both numerical and categorical data.
- Two models in use: Sklearn SimpleImputer & XGBoost. More to come in the future.

### Automatic portfolio optimization [🔗](#) | Python, Flask, Docker, HTML, Javascript, GCP

- A tool to fetch given assets' adjusted closing prices from Yahoo Finance.
- Using the historical prices, optimize a portfolio with the assets weighted using the *Markowitz model*.
- Dockerized the tool, served it with Gunicorn and deployed with the Google Cloud Platform (GCP).

### Finnish house prices scraper, EDA and prediction | Python, Scikit-Learn, XGBoost

- Implemented an OOP-style data scraper using Selenium to get data on houses and condos. Then cleaned it and saved it to an SQLite database.
- Exploratory Data Analysis (EDA) on the scraped data. Implemented a model to predict the price of accommodation, given its several features like size and location.
- The prediction was done with the Extreme Gradient Boosting Regression model. The model performed quite well, given the noisy data.