

# 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

Apr. 2020 – Present

- Did my M.Sc. thesis: "*Estimating the Effectiveness of Depression Care Pathways: A Process Mining and Survival Analysis Approach*".
- Was in charge of implementing, A/B testing and monitoring decision logic for customer steering rules.
- Created and maintained a monthly customer traffic report using *SQL*, *Qlikview* & *Piwik*. Around 200 people used the report to find inefficiencies and targets to improve on.
- Performed K-means clustering and time series analyses to highlight the 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

June 2019 – Sept. 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

Sept. 2020 – Dec. 2022

- 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

Sept. 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, Tidyverse, 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.

### Hotel demand prediction [↗](#) | Python, XGBoost, Seaborn

- Predicting demand given booking curves. XGBoost was fitted to reservations on hand for each day until the arrival date.
- This prediction method is more efficient than having multiple time series models for each booking curve length.
- Reasonable accuracy for daily predictions with a weekly horizon (MAPE of 8%).

### Finnish house prices scraper [↗](#), EDA and prediction [↗](#) | Python, Scikit-Learn, XGBoost, Selenium, SQLite

- 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 an accommodation, given its several features like size and location.
- The predictions were done with common regression models and XGBoost. The models performed quite well, given the noisy data.