Data Analyst / Statistical Modeling

ML / Python / R / SQL / Big Data / Tableau

Or Ben-Haim | Ramat-Gan | 054-2018844 | [orbenhaim20@icloud.com](mailto:orbenhaim20@icloud.com) | [LinkedIn profile](http://www.linkedin.com/in/or-ben-haimmm)

**Summary**

* **3 years** of practical experience in **data analysis, statistical modeling,** and **machine learning** using **Python, R, SQL,** and **big data** tools.
* **B.Sc. in Statistics, Data Science & Business Management**, The Hebrew University
* Skilled in extracting and analyzing **complex data**, experienced in building **advanced visualizations**
* Experienced in **Tableau, Excel**, and **SQL,** and environments like **Apache Spark** and **Colab**.
* Proficient in **statistical learning** and applying **supervised** and **unsupervised ML models**, including **SVM, neural networks,** **random forest**, and **k-means.**
* Strong command of **data structures, algorithms,** and simulations for modeling stochastic processes and using Tabular data modeling such as **F-tests & ANOVA**.
* Experience **applying NLP** and time series modeling (**linear regression, SARIMAX**) to explore relationships between textual sentiment and financial time series.

**Education & Courses**

2020 - 2024 **Dual B.Sc./ B.A. in Statistics, Data Science, & Business Management**, The Hebrew University

2015 - 2019  **Studies** **in Physics and Economics**, The Hebrew University of Jerusalem

2023 - 2024 **Tableau Certified Data Analyst**, Udemy

2023 - 2024 **SQL Bootcamp(PostgreSQL and pgAdmin)**, Udemy

**Sentiment Signal NASDAQ Forecast Project**

* **Project Goal -** Explore whether the tone of financial news articles is linked to movements in the NASDAQ index.
* **Core Idea -** Understand if shifts in media sentiment can reflect or even anticipate changes in the stock market.
* **Data Collection -** Gathered financial headlines from CNBC, Reuters, and The Guardian. Used a language model to generate short summaries where descriptions were missing, creating a more complete textual dataset.
* **Sentiment Analysis -** Each headline + description pair was scored for sentiment using **pre-trained NLP models** from **Hugging Face**. Sentiment scores were aggregated on a monthly level.
* **Data Alignment -** Monthly sentiment scores were aligned with the monthly average closing values of the NASDAQ index (retrieved via yfinance). Resulted in two synchronized time series: one for public sentiment and one for market performance.
* **Modeling Approach -** Compared two models: A simple linear regression using sentiment scores to predict the NASDAQ. **A SARIMAX time series model** incorporates both past NASDAQ behavior and sentiment as an external signal.

**Experience**

2020 - 2024 **Practical Academic Experience**, The Hebrew University

* Conducted **data extraction** and **analysis** using **SQL, Python,** and **R** to support decision-making processes across various projects.
* Built **advanced Excel models** (including PivotTables, VLOOKUP, XMATCH, and What-If Analysis) to forecast trends and evaluate financial scenarios such as interest rate impacts on amortization schedules.
* Designed and presented **data visualizations** using Excel and R, including heatmaps and world maps, to effectively communicate **complex datasets.**
* **Implemented simulations** in R to analyze stochastic processes, including elevator usage patterns and traffic load distributions.
* Worked in **big data environments** (**BigQuery, Apache Spark, Colab**) to read and process large datasets using Python and **Unix** commands, followed by generating meaningful insights through automated reports and graphs.
* **Applied statistical** and **machine learning techniques** such as F-tests, ANOVA, logistic and **linear regression**, **SVM**, **neural networks**, and **clustering models** (K-means, CPA) to identify patterns and predict performance outcomes.
* Developed and optimized **custom algorithms** and **applied data structures** to ensure efficient data processing and accurate analysis.

**Tools and Technologies**

Proficient with MS Office, including **Excel;** Statistical analysis: **R**

Programming languages & Databases: **Python, C++**, **SQL,** and **Visual Basic**

Additional functions: **Count, Counta, Vlookup, Index, Match**

Financial formulas: IRR, XIRR, NPV, XNPV, PV, PMT, PPMT, IPMT

**Languages & Military Service**

**Hebrew** - Native | **English** - Full proficiency

Combat medic and squad commander, Magal Corps