Sebastian Van Hemert

svanhemert00@gmail.com | 424-400-1223 | LinkedIn | GitHub

EXPERIENCE

Northrop Grumman | Los Angeles, CA
Research Analyst | March 2023-Present
- Wrote the Python logic for satellite cost-predicting tool used to validate estimates for government proposals.
- Conceputalized and built MS Excel tool using VBA and Macros that maps raw accounting data from SAP to industry Work Breakdown Structure (WBS). Output is stored in SQL database and delivered to government for audits.
- Enhanced outdated SQL Server cost history database using SQL code and MS Power Query. Normalized tables for efficiency and introduced inflation-adjusted views. Currently writing ETL script in PySpark for data ingestion.
- Produced Tableau and MS Excel dashboards with dynamic parameters for flexible exploratory data analysis.
- Top Secret (TS) security clearance. Currently awaiting Sensitive Compartmented Information (SCI) clearance.

Loyola Marymount University | Los Angeles, CA
Rains Research Graduate Assistant | June 2023-Present
- Assisted graduate professor and Koginvo AI Founder and CEO, Arin Brahma, Ph.D., in groundbreaking research on "Predicting CVD without any Blood Test using Machine Learning and CVD Progression Pattern Discovery using CT Markov Chain Model," with a particular emphasis on feature selection and feature engineering. Edited Python and R code, wrote documentation, and summarized results in MS Excel. Currently building Streamlit web app front end.
- Currently assisting Assistant Professor Yasaman Ghasemi, Ph.D., in Children's Hospital Los Angeles (CHLA) No-show Prediction research project. Produced descriptive dashboard in Tableau and involved in model selection/tuning.

RTX Corporation | Los Angeles, CA
Data Analyst | August 2022-January 2023
- Supported OneTrust data catalog effort to reconcile data assets spread across different business units after merger.
- Assessed and assigned risk level to third-party companies based on their use of personal data beyond access and auth.
- Leveraged MS Excel and MS Power Query to detect matches on unassociated data based on word placement. Deliverable prevented leadership from completing this task manually.
- Set up spreadsheet and dashboard that streamlined manual data input for data stewards from multiple companies (Pratt & Whitney, Collins Aerospace, etc.), in compliance with OneTrust bulk upload standards.

Retina AI | Los Angeles, CA
Data Analytics Intern | May 2022-August 2022
- Developed customer anomaly-detection tool in PySpark, pandas, and scikit-learn on Databricks. Tool detected anomalies contextually, statistically, and through machine learning algorithms, and was briefed to DS Team.
- Automated notebooks using Databricks widgets, adhering to Retina AI's data pipeline architecture template.
- Monitored website traffic and performance using Google Analytics 4, generating reports to track performance.

PROJECTS

Spotify Project | Spring, 2024: https://spotifind.streamlit.app/
- Created Streamlit web app that allows emerging artists to compare their music to clusters of Spotify editorial playlists.
- ETL pipeline involves Spotify Web API, JupyterHub, AWS RDS MySQL Data Warehouse, Streamlit, and Tableau.

LMU Datathon – First Place Award | Spring 2023
- Won first-place prize in LMU's first-ever Datathon by developing a Python predictive model that supported nonprofit Good Jobs First's mission of exposing wasteful subsidies and corporate misconduct. The multiclass model predicted a range that a subsidy will fall in, and output was validated with a continuous output that fell within these bounds.

EDUCATION

Loyola Marymount University | Los Angeles, CA
Master of Science in Business Analytics (MSBA) | Graduation Date: May 2024

Loyola Marymount University | Los Angeles, CA
Bachelor of Science in Information Systems and Business Analytics (ISBA) | Graduation Date: May 2022

TECHNOLOGIES

- Certified MS Office Excel Expert and Specialist
- Proficient in SQL (MySQL, MS SQL Server), Python for Data Science (pandas, NumPy, Streamlit, scikit-learn, etc.), Tableau, MS Power BI, Tableau Prep, Talend, Azure ML Studio (classic), MS Power Query, MS Access, Lucidchart
- Some PySpark, Databricks, Google Analytics, AWS, VBA, R, ArcGIS Pro, Alteryx, Snowflake, FastAPI
- Understanding of ML, ETL, NLP, Data modeling, Agile Methodology, Cloud Computing, Git
- Native Spanish Speaker