

Ali Hamza Detho

dethoaliamza@gmail.com | github.com/average-ali
03073132641 | www.linkedin.com/in/ali-hamzab24424236

Education

The City School Jinnah Campus | O-levels(8 A*/As), A-levels(3 A*/As) **Aug 2015 - Aug 2019**
NED-UET, B.E Computer Systems Engineering | Class of 2023 | CGPA: 3.17/ 4.00 **OCT 2019 - OCT 2023**

Achievements

- Google Cloud Certified Professional Data Engineer
- Google Cloud Certified Professional Cloud Database Engineer
- DataCamp Certified Associate Data Analyst
- 50+ badges, courses and projects completed on DataCamp, Google Cloud Skills Boost, Microsoft Learn

Relevant Work Experience

Data Engineer, Nebbiu Limited

Jan 2023 - Oct-2024

- Created an efficient data model and migrated client's HVAC product data from different Excel workbooks to BigQuery, eliminating data silos, enhancing data management, and enabling advanced analytics.
- Achieved a 25% improvement in data quality by performing exploratory analysis and implementing standardised validation checks on the client's data warehouse.
- Wrote ad-hoc SQL queries for a French online polling platform to analyze user engagement by extracting insights from anonymized poll data, including summary statistics and voter behaviour patterns.
- Optimised the ETL process for updating HVAC product data by replacing fragmented Python files with templated SQL pipelines in Dataform, enhancing speed, access control, and data validation.
- Achieved an 80% speedup over VLOOKUP-based methods for generating price-book spreadsheets for HVAC-R wholesalers by developing automated ETL pipelines on Google Cloud.
- Developed a customer purchase Likelihood prediction model using BigQuery ML for a European ferry reservation website. Documented the process and conducted training sessions for knowledge transfer.
- Achieved 60x reduction in model serving latency (from 1 minute to 1 second) by migrating client's ML workload to Vertex AI and Cloud Functions.

Data Science Intern, NCCS-NED

Aug 2022 - Dec 2022

- Worked in a team of four to develop a user entity behaviour analysis system for our FYP.
- Obtained a rich feature set by applying statistical and NLP techniques on the CMU CERT r4.2 insider threat dataset. This feature set was used to train an autoencoder model for anomaly detection.
- Created a visual report containing time-series charts to analyse the behaviour patterns of malicious actors.
- Reported progress to the project supervisor and documented the workflow for inclusion in the final report.

Projects

DataCamp Associate Data Analyst Exam Case Study

The practical component of this two-part certification exam tested my data management and exploratory skills. I performed EDA, validation, cleaning and aggregation tasks before creating insightful visuals to answer key questions about claims customers made against a restaurant chain in Brazil.

Link to the Github repo: https://github.com/average-ali/DataCamp_associate_DA_exam/tree/main

Real-time dashboard for NYC TLC trips dataset

Employed Dataflow to build a data pipeline that streamed data from a Pub/Sub source into Looker Studio via BigQuery. Visualised relationships between key metrics such as total revenue, total rides, and operational time to monitor the performance of the taxi fleet.

Skills

Languages: Python(pandas, seaborn, matplotlib, scikit-learn, keras, PySpark), SQL, HTML, DAX

Tools: Microsoft Power BI, PostgreSQL, Excel, Airflow, Jira

GCP: BigQuery, Dataform, Cloud Storage, Looker Studio, Cloud Functions, Vertex AI, Cloud Shell