

Patrick Antonacci

Toronto, ON · 647-637-4609 · patrick.antonacci@gmail.com · </in/patrick-antonacci/> · github.com/pantonacci

SUMMARY

Results-driven Data Engineer with over 7 years of experience designing and optimizing scalable data architectures, pipelines, and models on cloud platforms, including GCP. Proven expertise in BigQuery, Airflow, dbt, Vertex AI, and Python to deliver efficient ETL/ELT workflows that support advanced analytics, product, and machine learning initiatives. Adept at collaborating with cross-functional teams to translate business requirements into technical solutions learned from my time as a consultant. GCP Professional Data Engineer certified and skilled in performance optimization, data governance, and automation, with a focus on cost-effective processing of large datasets.

EXPERIENCE

- | | | |
|--------------|---|-------------|
| 2022-Present | SAGARD HOLDINGS
<i>Senior Data Scientist</i> | Toronto, ON |
| | <ul style="list-style-type: none">• Implemented data ingestion pipelines using web scraping, batch file ingestion, and API ingestion into data pipeline for deal sourcing on GCP using Cloud Functions, Cloud Scheduler, and Cloud Composer, processing 3.5M + weekly records.• Created an ELT infrastructure using Airflow, dbt, and BigQuery to filter company leads for our application, and live dashboarding using Looker Studio for our performance metrics.• Trained and implemented a logistic regression machine learning model in production to predict likelihood of company leads for investors using Google Vertex AI on our application.• Incorporated generative AI using Gemini LLM models on GCP into our data pipeline to categorize companies, extract data from documents, and extract data from text based online news articles.• Created data architecture of a product for investment signals through various data sources for fintech startups, working on a team with a product manager, developer, designer, and the CTO.• Managed data engineering, science, and analytics agile sprint work for full stack data pipelines on Google Cloud from multiple data ingestion sources, data processing and warehousing, machine learning predictions, to our application. | |
| 2021-2022 | MONTREAL ANALYTICS
<i>Senior Data Consultant</i> | Toronto, ON |
| | <ul style="list-style-type: none">• Consulted on modern cloud based data architecture in GCP for a variety of businesses using data warehouses, dbt, and Looker, creating custom solutions and outlining detailed roadmaps for client engagements during the sales process.• Built, audited and refactored business intelligence and KPI models and dashboards in dbt and Looker for businesses by developing scalable models for clients with big data.• Trained clients' developers on best practices and defined coding standards and development processes, including CI/CD and blue-green deployment.• Managed data and analytics engineering agile sprint work for clients' data models and pipelines in GCP and BigQuery / Snowflake / Redshift resulting in long term partner engagements. | |
| 2016-2021 | ALTUS ASSESSMENTS INC.
<i>Senior Data Scientist</i> | Toronto, ON |
| | <ul style="list-style-type: none">• Provided thought leadership for company wide data architecture by building ETLs in our data warehouse using AWS Glue and PySpark to implement and maintain Looker for data visualization, reporting, KPI tracking.• Used R, Python, and Spark for various ETLs from company transactional MongoDB into a structured SQL format to create dashboards and reports for data-driven company wide executive decisions and performance tracking.• Created automated response rating prediction algorithm in Python using NLP and various modeling methods, able to achieve a quadratic weighted kappa of 0.6, a higher level of accuracy than our human raters.• Used AWS architecture to create a data pipeline which transcribed video files in S3 to implement NLP tools for sentiment analysis, topic summarization, and topic modeling on videos for production.• Developed automated fraud detection for plagiarism on large unstructured text data sets with natural language processing using cosine similarity resulting in an automated plagiarism reporting tool for our product. | |

EDUCATION

- | | |
|-----------|--|
| 2013-2015 | UNIVERSITY OF TORONTO , Toronto, ON
<i>Master's in Mechanical Engineering</i> |
| 2008-2012 | UNIVERSITY OF TORONTO , Toronto, ON
<i>Bachelor's in Physics and Mathematics</i> |

CERTIFICATIONS

- | | |
|------|---|
| 2022 | GOOGLE CLOUD - PROFESSIONAL DATA ENGINEER |
| 2022 | DBT LABS - DBT CERTIFIED DEVELOPER |
| 2021 | LOOKER - LOOKML DEVELOPER CERTIFICATION |
| 2017 | UNIVERSITY OF TORONTO - DATA SCIENCE CERTIFICATE |

SKILLS

PROGRAMMING: Python, SQL, R, Agile Project Management, Git

DATA SKILLS: Cloud Based Data Architecture, Data Warehouse Design, Data Pipelines, Data Modeling, ELT/ETLs, Serverless Data Ingestion, Exploratory Data Analysis and Data Visualization

MACHINE LEARNING / GEN AI: Vertex AI, GenAI (Gemini), LLM prompting, BigQuery ML, Regression modeling, AutoML, A/B Testing, Natural Language Processing

CLOUD ARCHITECTURE: BigQuery, Snowflake, Redshift,
PostgreSQL, GCP, AWS, dbt, Airflow, Looker, Spark, MongoDB