

SALVIN

Data Analyst cum Data Engineer

PROFESSIONAL SUMMARY

- Over 5 years of experience as a **Data Analyst and Data Engineer** with hands-on expertise in designing scalable **data pipelines**, conducting **product analytics**, and real-time analytics.
- Strong proficiency in **SQL, Python, and R**, with experience in both data analysis, behavioral analysis and engineering projects across various industries.
- Skilled in building and maintaining **data warehouses**, automating workflows using **Apache Airflow**, and developing ETL pipelines.
- Proficient in visualization tools like **Tableau, Power BI, Metabase, and Looker Studio** to deliver actionable business insights.
- Proven track record in identifying and tracking **core product metrics (retention, engagement, conversion)** to guide business decisions.
- Experience working with cloud platforms like **Google Cloud Platform (GCP)** and **AWS** for data infrastructure and automation.
- Experienced in **A/B testing**, experimentation frameworks, and applying **causal inference** to assess impact and improve product strategy.
- Built observability dashboards using Grafana (demo setup) integrated with Prometheus to monitor Airflow DAGs and ETL failures.
- Implemented basic instrumentation using OpenTelemetry to trace data pipeline execution and latency across services.
- Applied anomaly detection algorithms on ETL log data to identify failed jobs and unusual delays, reducing false alert fatigue.
- Familiar with Kafka for real-time ingestion of event data; working on a proof-of-concept for event-driven ETL orchestration.
- Built dashboards and data products that directly impacted decision-making for **Sales, Marketing, and Product teams**.
- Worked with **Salesforce, Google Analytics**, and various CRM tools for **user funnel analysis** and performance tracking.
- Exposure to **LLMs and GenAI**, with active participation in Analytics Vidhya's DataHack Summit 2024.
- Experienced in deploying solutions using **Docker, Kubernetes**, and version control with Git.
- Adept at collaborating across cross-functional teams and documenting solutions for transparency and scalability.

SKILLS

Databases: Snowflake, PostgreSQL, MySQL

Programming Languages: Python, R language

Visualization Tools: Tableau, PowerBI, Metabase, Google Looker Studio

Cloud Tools: Google Cloud VM, AWS

Applications: VS Code, Jupyter Notebooks, Apache Airflow, Postman, R-Studio, DBT Libraries & Tools:

Pandas, GSpread(Google Sheets), PySpark

Deployment Tools: Git, Kubernetes – Docker, Rancher

Business Apps: Salesforce & Google Analytics

Recent Activity - Aanded Analytics Vidhya DataHack Summit 2024, focusing on LLMs & GenAI.

EXPERIENCE

Insightogram Dec 2023 - Present Data Engineer

- Developing and managing data pipelines to process, clean, and transform large datasets for business insights.
- Utilizing Snowflake and PostgreSQL for efficient data storage and retrieval.
- Automating data workflows using Apache Airflow and Python scripting.
- Collaborated with PMs to define KPIs, track product usage metrics, and build dashboards using Tableau/Power BI.
- Enhancing data infrastructure for improved performance and scalability.
- Improved feature adoption tracking and supported A/B test result monitoring.
- Built observability dashboards using Grafana (demo setup) integrated with Prometheus to monitor Airflow DAGs and ETL failures.

WeWork India Nov 2022 - Nov 2023 Lead

- Currently building LLM models for business key Stakeholders
- Oversee & manage Data warehouses integrated with Salesforce and other Org Databases.
- Drove End-to-End Automation of Business Reports utilizing Python, DBT, Apache Airflow & GCP, reducing 48 hours of manual human hours into 3 minutes, in addition to avoiding human errors.
- Collaborated with Marketing and Product teams to develop behavioral segmentation, feature usage tracking, and conversion funnel analysis.
- Implemented A/B tests to evaluate engagement across communication channels.
- Created dashboards tracking core product KPIs like session frequency, user paths, and retention curves.
- Lead partner in planning and developing data infrastructure to fetch, process, and provide reporting for multiple teams.

Senior Data Analyst

- Performed cohort and behavioral analysis to uncover purchasing patterns across B2B pharmacy clients.
- Enhanced conversions and cut acquisition costs by collaborate data analytics with marketing teams.
- Developed precise customer segmentation models (SQL, Python, data tools) for high-engagement campaigns, including A/B testing & Conversion Channels.
- Utilized SQL, Python, and Google App Scripts for robust data extraction, transformation, and analysis.
- Build a data warehouse for analytics by positioning across CRM platforms, Salesforce, and Salesforce (In-house App).
- Built a simple lead management system using Google APIs, significantly improving lead reachouts.
- Identified inefficiencies in the process, recommended enhancements, and safeguarded product revenue.
- Strong believer in documenting processes and findings using tools like Notion, Confluence, and ensuring transparency and traceability.

Biddano May 2022 - Oct 2022 Junior Data Analyst

- **Context** A company focused on B2B Distribution of Pharmacy Products.
- Optimized 5 hours of delivery time into 2 hours using operational data, increasing throughput.
- Supporting Developers with optimized SQL queries for client dashboards.
- Providing Cohort Analysis to Clients for Purchase behavior and providing key Insights on the areas that need business attention
- One-stop person for data related to Finance, Operations, & Business Stakeholders
- Developed Metabase & PowerBI reports, showcasing reports to Vertical Heads

- Developed Python Scripts to read 5000+ human-made Excel ledgers
- Google Data Studio that tracks work logs & Jira tickets for optimization of workloads.
- Motivated and taught my peers to write efficient SQL queries.

UNICOURT Oct 2020 - April 2022 Data Analyst

- Conducted root cause analysis of bad, streamlined data and built a Postgres Data Warehouse that queried and processed approximately 4 million records from 3000+ databases using Kubernetes within a few seconds.
- Generated weekly KPIs, Net Promoter Score, Customer behavior analytics, Ad-hoc Analytics, churn rate, Revenue, and core organizational reports for strategic planning of the business using Google Spreadsheets, Metabase & Python.
- Pre-processed large raw datasets from data entry users to identify key ones for a business strategy using Data Mining Techniques.
- Implemented innovative automation in weekly/monthly reports, saving 80% of the time in generating these reports.
- Proactively collaborated with various teams like SEO & DevOps to manage and gather intelligent data. Communicated & presented key insights to the senior leadership of the organization.

ACADEMIC AND PERSONAL PROJECTS

1. Project - Sales Forecasting Model

Technologies Used: Python, Machine Learning (Scikit-learn), Pandas, NumPy, OpenTelemetry, Grafana, kafka

Duration: Oct 2024 - Mar 2025

Description:

Built a regression-based model to predict future sales trends using historical data.

Roles & Responsibilities:

- Collected and preprocessed historical sales data for training and testing.
- Applied and tuned multiple regression models for better accuracy.
- Evaluated model performance using metrics like RMSE and R^2 score.
- Interpreted results to identify key business drivers.
- Developed a predictive model to forecast sales trends based on historical data.
- Utilized regression analysis to optimize business sales strategies.

2. Project - ETL Pipeline Automation

Technologies Used: Apache Airflow, Python, PostgreSQL, OpenTelemetry, Grafana, kafka

Duration: Apr 2024 - Sep 2024

Description:

Developed a scalable ETL pipeline to automate data movement across sources.

Roles & Responsibilities:

- Designed DAGs in Airflow for daily data extraction and transformation.
- Wrote custom Python scripts to handle data validation and cleaning.
- Integrated pipeline with PostgreSQL for structured storage.
- Monitored and debugged pipeline failures using Airflow logs.
- Designed and implemented an automated ETL pipeline to extract, transform, and load data.
- Improved data processing efficiency by automating data ingestion from multiple sources.

3. Project - Real-Time Analytics Dashboard

Technologies Used: Tableau, Snowflake, SQL, OpenTelemetry, Grafana, kafka

Duraon: Dec 2023 - Mar. 2024

Descripon:

Created a real-me dashboard to visualize and monitor key business metrics.

Roles & Responsibilities:

- Wrote opmized SQL queries to fetch and join large datasets from Snowflake.
- Built calculated fields and KPIs in Tableau for visual insights.
- Worked with stakeholders to define data requirements.
- Implemented data refresh logic and dashboard scheduling.
- Built an interacve real-me analycs dashboard for business reporng.
- Opmized database queries to improve dashboard performance.

4. Project - Customer Analycs and Segmentaon of Online Marketplace

Technologies Used: Python, SQL, OpenTelemetry, Grafana, kafka, RFM Analysis, NLP, Churn Analysis

Duraon: Nov 2022 - Nov 2023

Descripon:

Analyzed customer behavior and segmented users to target retenon strategies.

Roles & Responsibilities:

- Applied RFM analysis to categorize customers by purchasing behavior.
- Conducted churn predicon using classificaon models.
- Performed senment analysis on reviews using NLP.
- Suggested product-specific offers based on segment behavior.
- A project focusing on improving Customer value and Customer Equity (CE). Segmenng customers based on Recency, Frequency, and Monetary values (RFM Analysis). Customers were grouped based on their loyalty to launch new products.
- Evaluated customers using Churn Analysis to increase profits and assess loss using Natural Language Processing (NLP) on reviews and targeted segmented groups.

5. Project - Autonomous Visual Assistance (AVA)

Technologies Used: Python, Deep Learning (TensorFlow, Dlib), OpenTelemetry, Grafana, kafka,

OpenCV, NumPy, IMUTILS, PIL, Jupyter Notebook **Duraon:** Nov 2020 - Nov 2021

Descripon:

R&D iniave to assist visually impaired individuals through AI-powered object and face recognition.

Roles & Responsibilities:

- An R&D project focused on helping visually impaired people using Deep Learning Techniques.
- Trained Dlib face recognition model with data augmentaon technique, achieving **99.38%** accuracy.
- Achieved Transferred Learning in TensorFlow's pre-trained SSD mobile net v2 for Object Detecon and Vehicle Detecon.
- Tools and Libraries: Python NumPy, TensorFlow, IMUTILS, PIL, Jupyter Notebook, Tkinter.

6. BCA Alumni

Technologies Used: MySQL Server, HTML, PHP, CSS, Bootstrap,

- Developed a scrum project by building a community forum website for Alumni of St. Aloysius College.

EDUCATION

St. Aloysius Instute of Management & Informaon Technology September 2020 *Master of Science in Big Data Analycs.*

St Aloysius College

Bachelor of Computer Applicaon **August 2018**