

Joseph J. Bokan III

📍 New York ✉ jjbokan3@gmail.com ☎ (518) 698-8886 📺 in/josephbokan 🌐 josephbokan.io

SUMMARY

Detail-oriented Data Professional with over 4 years experience in data science and engineering. Specializes in processing large-scale datasets, building data pipelines, leveraging cloud technologies, and applying analytics to drive business insights.

EXPERIENCE

Beauty Co-Lab (Omnicom Media)

Senior Analyst, Data Science

New York, NY
November 2024 - Present

- Manage a 10TB project on BigQuery and optimize scalable data pipelines for efficient ingestion and processing of raw data.
- Conduct rigorous QA by mapping raw data to expected schemas based on a comprehensive data dictionary, ensuring high data quality.
- Design and implement staging and production schemas with normalized table structures to streamline data transformation.
- Leverage SQL and Python to integrate 1p, 2p, and 3p data through ID matching, delivering actionable insights for competitive market analysis.

Michael Kors

Customer Data Analyst

New York, NY
October 2022 - November 2024

- Automate reporting and data cleansing processes with Python, enhancing productivity and data insight quality.
- Optimize and implement automated ETL pipelines using Python and SQL to ingest, clean, and transform social media leads into the CRM database, reducing manual data entry time by 80% and ensuring 99% data consistency and accuracy.
- Utilize SQL and Snowflake to efficiently query datasets over 1.5TB, directly supporting decision-making for marketing and customer segmentation strategies.
- Actively collaborate with cross-functional teams to communicate data-driven insights, promoting transparency, and fostering a data-centric culture across departments for improved decision-making.
- Enhanced data quality by implementing rigorous validation processes, reducing data errors by 20%, and achieving 98% accuracy in customer records within six months using SQL and Python-based scripts.

Radiate The World, Inc.

Data Science Intern

New York, NY
March 2022 - July 2022

- Leverage Metabase and SQL to query, analyze, and extract actionable insights, resulting in a 30% improvement in reporting efficiency for business decisions.
- Harness Python and advanced data analytics libraries to forecast user growth, enabling a 15% increase in projected user retention through strategic adjustments.
- Collaborate with the product team to drive data-driven recommendations for A/B testing, UX/UI optimization, and feature enhancements.
- Facilitate cross-functional collaboration in a fast-paced, agile startup environment to deliver impactful solutions.

PROJECTS

MTA Subway Tracker

November 2024 - Present

- Built a real-time NYC subway map and timetable by extracting, validating, and cleaning transit data using Python, Pydantic, and Folium for geospatial visualization.
- Developed and deployed a FastAPI-based RESTful API hosted in Docker on AWS EC2, providing users with directions to nearest subway stations and accurate train arrival predictions.
- Established a structured database archiving historical arrival and timing data to support future transit analytics and performance optimization.

EDUCATION

Master of Science in Business Analytics

Fordham University, Gabelli School of Business • 3.9/4.0

New York, NY
August 2023

Bachelor of Science in Information Systems

Fordham University, Gabelli School of Business • 3.7/4.0

Bronx, NY
May 2022

TECHNICAL EXPERTISE

Programming & Frameworks: Python (Polars, PySpark, Pydantic, FastAPI), SQL, Scala

Data Engineering & Orchestration: Dagster, DBT, DLT, BigQuery, Snowflake, Postgres, DuckDB

Cloud & Tools: GCP (BigQuery), AWS, ETL/ELT Pipelines, Data Modeling