Venkata Sai Bharadwaj Velamakanni

Plainsboro, NJ | Willing to relocate anywhere in the US
vvsbharadwaj24@gmail.com | +1 714-860-2987

linkedin.com/in/venkatavelamakanni
github.com/venvelamakanni
venvelamakanni.github.io

PROFESSIONAL SUMMARY

Data Engineer & Analyst with 3 years of designing scalable ETL pipelines, building ML models, and automating dashboards in SQL/Python. Proven track record optimizing performance at scale—seeking to apply cloud-native data engineering expertise to drive enterprise-wide strategy and insights.

EXPERIENCE

Senior Data Engineer   June 2024 – Present
BulkMagic   *Jersey City, NJ*

- Engineered a real-time pricing engine pipeline ingesting 100K+ SKUs/day via Spark Streaming into PostgreSQL—boosted average order margin from 20% to 30% and cut price-lookup latency by 35%.
- Implemented a Redis caching layer (Valkey) for competitor pricing lookups in Python services—reduced external API calls by 40% and lowered query latency from 150 ms to 90 ms.
- Developed fallback logic in PySpark for pricing data ingestion—improved pipeline uptime from 70% to 99.5% over Q4 2024.
- Orchestrated weekly competitive pricing DAGs (Directed Acyclic Graph) in Prefect—automated job scheduling and error handling, raising pipeline reliability from 65% to 95%.
- Designed and deployed executive Tableau dashboards powered by FP-Growth (NetworkX, Apriori) to reveal $100K+ cross-sell opportunities, enabling targeted sales strategies.
- Built an image-transformation microservice using FastAPI, containerized with Docker and deployed on Kubernetes (AWS EKS); integrated Prometheus/Grafana monitoring to reduce manual processing effort by 80% and maintain 99.9% availability.
- Conducted daily standups and led sprint planning for a team of 5 data engineers, reviewed pull requests, and onboarded 2 new teammates—improving team velocity by 20%.
- Implemented PyTest and unittest frameworks to achieve 95% unit test coverage on PySpark modules—caught regressions before deployment and increased code reliability.

Data Analyst   Feb 2022 – Dec 2022
Vistex, Inc   *Hyderabad, India*

- Streamlined SAP SD rebate process for 2 enterprise clients by configuring automated rebate schedules and developing Python scripts—cut manual reconciliation time by 30% and ensured 100% contract compliance.
- Created dynamic Excel models with pivot tables, VLOOKUPs, and macros to audit rebate liabilities and forecast payouts—reduced spreadsheet effort by 40% and improved accuracy by 25%.
- Built real-time SSRS dashboards and Power BI reports by querying SAP HANA and SQL Server—embedded drill-downs and alerts for KPIs, reducing review cycle from weekly to daily and cutting payroll errors by 20%.
- Optimized GTMS analytics pipelines in Databricks by refactoring PySpark jobs, adding Python validation checks, and running HiveQL on Hadoop for data staging—improved data accuracy by 35% and preserved $4.2M in recurring revenue.

Data Analyst   Jan 2021 – Sept 2021
Suven Consultants   *Hyderabad, India*

- Built a Python/SQL segmentation pipeline using K-means and RFM (Recency, Frequency, Monetary) on 100K+ transactions—clustered high-value segments and flagged risky accounts, driving a 20% lift in campaign ROI.
- Developed advanced Excel models to reconcile campaign spend vs. performance metrics—saved 5 hrs/wk of manual reporting.
- Partnered with stakeholders to design and analyze A/B experiments for payment flow changes—wrote SQL queries to pull A/B cohorts, analyzed results in Python (pandas), and recommended strategy shifts that reduced abandonment from 40% to 30%.

PROJECTS

End-to-End Machine Learning Pipeline for Fraud Detection   Project Link

- Built an end-to-end Python pipeline ingesting 1M+ payment transactions into PySpark on Databricks, engineered 50+ risk features (velocity, geolocation, device), trained a logistic-regression model (scikit-learn) to predict fraud with 92% accuracy, deployed as a FastAPI microservice in Docker orchestrated via Airflow DAGs, and integrated real-time scoring with Snowflake.

SKILLS

Analytics & Data Engineering: SQL (Teradata, PostgreSQL, BigQuery), Python (pandas, scikit-learn), R, predictive modeling (logistic regression, clustering), PySpark, Databricks, Alteryx Hadoop/Hive, Airflow, AWS Glue, Snowflake

Reporting & DevOps: A/B test, KPI dashboards (Tableau, Power BI), Excel (pivot tables, VLOOKUPs, macros), PyTest, unittest, Git/GitHub Actions, Docker, CI/CD (Jenkins, Azure DevOps), Agile/SCRUM (daily standups, sprint planning), Jira

EDUCATION

California State University, Fullerton   Fullerton, CA
Masters in Information Systems, Concentration in Business Analytics–GPA-3.79   *Jan 2023 – Dec 2024*

Vardhaman College of Engineering   Telangana, India
B.Tech, Mechanical Engineering (GPA: 8.5/10)   *Jul 2018 – May 2022*

Certifications

AWS Data Engineer - Associate Certification, Google Project Management Professional Certificate