

# SAI ASHISH BEHARA

Kent, OH | +1 (234) 296-3301 | [ashishflash21@gmail.com](mailto:ashishflash21@gmail.com) | [LinkedIn](#) | [Portfolio](#)

## SUMMARY

Data Analyst with 2 years of professional experience specializing in business intelligence, predictive analytics, and cloud-based data solutions. Proficient in Python, SQL, and visualization platforms with hands-on expertise in AWS, BigQuery, and real-time streaming technologies. Demonstrated success in optimizing data workflows, reducing operational costs, and delivering actionable insights that drive strategic decision-making across Fortune 500 enterprises.

## EDUCATION

**Master of Science in Data Science**, Kent State University

Expected December 2025

GPA: 3.95/4.0

Kent, OH

Relevant Coursework: Data Analytics, Statistical Forecasting, Business Intelligence, Project Timeline Optimization

## CERTIFICATIONS

**Microsoft Certified: Azure Data Scientist Associate**

Microsoft

## TECHNICAL SKILLS

### Data Analysis & Visualization

Power BI, Tableau, Looker Studio, Excel (Pivot Tables, Power Query, Macros), KPI Dashboards, Data Storytelling

### Programming & Databases

Python (Pandas, NumPy, Matplotlib, Seaborn), R, SQL (MySQL, PostgreSQL, Snowflake), MS Access, SAS

### Data Science & Machine Learning

Scikit-learn, TensorFlow, PyTorch, Regression, Classification, Clustering, ANN, CNN, RNN, Feature Engineering, Model Evaluation, A/B Testing, Forecasting, Time Series

### Generative AI & LLM Tools

OpenAI API, LangChain, Hugging Face Transformers, Ollama, Pinecone (Vector Databases), RAG, Prompt Engineering, LlamaIndex, Streamlit, Flask (Deployment)

### Data Engineering & ETL

Apache Spark, PySpark, Airflow, Kafka, Data Warehousing (Snowflake, BigQuery, Redshift), ETL Pipelines, Data Quality, Governance

### Statistical Techniques

Hypothesis Testing, Correlation Analysis, Predictive Modeling, Sampling Methods, Experimental Design, Probability Distributions

### Business Intelligence

Power BI Service, DAX, SSRS, SAP Analytics Cloud, Power Automate, Report Optimization

### Cloud & Big Data Platforms

AWS (S3, Redshift, SageMaker, QuickSight), Azure (Data Factory, Synapse, Cognitive Services), GCP (BigQuery, Vertex AI), Databricks, Delta Lake

### Version Control & Collaboration

Git, GitHub, Jira, Confluence, Agile/Scrum, CI/CD

## EXPERIENCE

### Product Analytics Intern

Geekl

Feb 2021 - Sep 2021

Remote

- Analyzed user engagement metrics and product usage patterns across 50K+ active users, tracking CTR (Click-Through Rate), conversion rates, and session behavior to optimize user acquisition funnels.
- Built predictive CLV (Customer Lifetime Value) models using Python and Scikit-learn to forecast revenue potential and identify high-value customer segments, achieving 82% prediction accuracy.
- Supported the launch of 3 new product features by establishing success metrics, designing measurement frameworks, and monitoring post-launch performance through real-time dashboards.
- Conducted A/B testing experiments for feature variations, analyzing statistical significance across 10K+ users and providing recommendations that improved feature adoption by 18%.

- Performed root cause analysis (RCA) on sudden drops in user engagement and conversion rates, identifying technical bugs and UX issues that led to resolution and 15% recovery in metrics.
- Created interactive Tableau dashboards visualizing product KPIs including Daily Active Users/Monthly Active Users (DAU/MAU) ratios, feature utilization rates, CTR by channel, and customer retention cohorts.
- Executed customer segmentation analysis using Recency, Frequency, Monetary (RFM) methodology and k-means clustering to enable targeted marketing campaigns and personalized user experiences.
- Collaborated with product managers and engineering teams to define event tracking requirements for new features, ensuring comprehensive data instrumentation and accurate measurement.
- Delivered weekly product performance reports to stakeholders, translating complex analytical findings into actionable insights that informed product roadmap prioritization.

**Data Analyst - Full Time Employee**

Oct 2022 - Nov 2023

Cognizant Technology Solutions

- Designed end-to-end analytics solutions for Fortune 500 manufacturing clients, translating business requirements into technical specifications and delivering insights that shaped quarterly strategic planning.
- Managed data operations for multiple client accounts simultaneously, overseeing data quality, pipeline monitoring, and troubleshooting to ensure 99.5% uptime across all analytical systems.
- Conducted in-depth exploratory data analysis on manufacturing operational datasets, uncovering inefficiency patterns that led to process improvements saving clients an estimated \$200K annually.
- Mentored 3 junior analysts on best practices for SQL optimization, Python scripting, and dashboard design, fostering technical skill development within the team.
- Participated in client workshops and requirements gathering sessions, serving as the technical liaison between business stakeholders and development teams to ensure alignment on deliverables.
- Enhanced existing data models by incorporating additional data sources and refining transformation logic, expanding analytical capabilities by 35% without infrastructure changes.
- Performed root cause analysis on data quality issues, implementing corrective measures that reduced incident tickets by 60% over a 6-month period.
- Contributed to proposal development for new client engagements, providing technical feasibility assessments and effort estimations for data analytics projects.
- Standardized documentation practices for data pipelines and analytical processes, creating comprehensive knowledge bases that reduced onboarding time for new team members by 3 weeks.
- Monitored and reported on SLA compliance metrics, maintaining transparent communication with clients regarding system performance and project milestones.

**Data Analyst Intern**

Feb 2022 - Sep 2022

Cognizant Technology Solutions

- Assisted senior analysts in data extraction, cleaning, and preparation tasks for client deliverables, handling datasets ranging from 100K to 5M records with high attention to detail.
- Conducted comparative analysis of client operational metrics against industry benchmarks, identifying performance gaps and opportunities for competitive advantage.
- Participated in weekly client status meetings, preparing progress reports and presenting preliminary findings to project stakeholders.
- Performed ad-hoc analytical requests with 24-48 hour turnaround times, supporting urgent business decisions with quick yet accurate data insights.
- Researched and evaluated emerging analytics tools and technologies, presenting recommendations to leadership for potential adoption in client projects.
- Created comprehensive test cases for data validation scenarios, ensuring analytical outputs met quality standards before client delivery.

- Supported data governance initiatives by documenting data lineage, business definitions, and transformation rules for key datasets used across multiple projects.
- Analyzed user feedback on existing dashboards and reports, implementing iterative improvements that enhanced usability and adoption rates by 40%.
- Collaborated with offshore development teams to troubleshoot technical issues, facilitating communication and ensuring timely resolution of production incidents.
- Built prototypes and proof-of-concepts for potential analytical use cases, demonstrating feasibility and value proposition to secure client buy-in for full implementations.

## **ADDITIONAL QUALIFICATIONS**

---

- Architected cloud data infrastructure on AWS leveraging S3, EMR, and Redshift, optimizing storage costs by 25% while simultaneously enhancing query performance for high-volume analytics workloads.
- Developed Python-based automation frameworks for continuous data quality monitoring and validation, reducing manual verification time by 60% and ensuring consistent data integrity.
- Performed advanced database optimization on PostgreSQL and MongoDB systems, improving application response times by 40% for high-traffic educational platforms.
- Implemented event-driven streaming pipelines using Kafka and PySpark, processing 500,000+ daily events to power real-time dashboards with minimal latency.
- Applied sophisticated statistical techniques including Central Limit Theorem, hypothesis testing, and confidence intervals to derive meaningful insights from large-scale transactional datasets exceeding 550,000 records.
- Led data governance and cataloging initiatives to strengthen data lineage documentation, improve asset discoverability, and ensure regulatory compliance.