Soham Prasad Ambekar

Dallas, TX; 75252; (469) 265-9450; sohamambekar21@gmail.com; linkedin.com/in/sohamambekar/

SUMMARY

Data professional with hands-on experience in SQL, Python, & statistical methods for data mining, building dashboards, data pipelines, & automated reporting systems. Skilled in translating complex datasets into actionable insights & continuously learning new techniques to support decision-making across healthcare & B2B supply chain environments.

EDUCATION

MS Information Science – *University of Texas at Dallas, Richardson, TX* **August 2022 - May 2024**

Relevant Coursework: Cloud Computing (AWS S3, EC2, Redshift), Business Data Warehousing, Business Analytics with SAS

BE in Computer Engineering – *University of Mumbai, Mumbai, India* **July 2015 - May 2019**

Relevant Coursework: Database Management Systems, Distributed Databases, Big Data

TECHNICAL SKILLS

Languages & Scripting: SQL (T-SQL, PL/SQL), Python (Pandas, NumPy, Matplotlib)
Data Visualization & BI: Power BI, Tableau
Cloud & Data Platforms: Hive, Spark, Hadoop, AWS (S3, EC2, Redshift)
Tools & Techniques: Excel, Data Validation, KPI Automation, Data Governance
Statistical & ML Methods: Logistic Regression, Clustering, Decision Trees, Hypothesis Testing, Bayesian Analysis
Certifications: Google Data Analytics; Oracle Cloud Infrastructure; Data Management; Databricks Lakehouse

PROFESSIONAL EXPERIENCE

Ecco Select (Site – NKC Health) – *Business Analyst* **September 2024 – Present**

- Identified inefficiencies in reporting architecture, proposed **HIPAA**-compliant enhancements to bolster data security.
- Optimized data workflows by implementing **incremental loading**, reducing processing overhead & cutting costs by 7%.
- Mapped data flows & **catalogued** 9+ data sources from clinical & operational systems to Health Catalyst (analytics platform), documenting end-to-end lineage to strengthen compliance.
- Developed SQL-based **ETL** pipelines & dimensional models to integrate Lawson ERP data into Health Catalyst, enabling unified analytics across clinical & supply chain teams.
- Delivered insights via **Power BI dashboards** & exec-level reports boosting adoption & data literacy in 3 departments.

Kroger Co. – *Technology Intern* **May 2023 – August 2023**

- Mitigated **Snyk**-reported vulnerabilities by implementing recommended fixes & enhancing application security by 30%.
- Documented test coverage improvements in **Confluence** & **JIRA**, increasing code quality & data transparency by 45%.
- Participated in sprint reviews & retrospectives, reinforcing test coverage goals & **backlog prioritization** for features.

Ingram Micro – *Business Systems Analyst* **June 2019 – April 2022**

- Led design & delivery of 10+ BI dashboards by translating business data needs into **scalable** models & enabling faster, self-service insights across teams, cutting turnaround time by 25%.
- Identified "dead carts" & proposed a 30-day auto-delete mechanism to **reduce inventory holding** on large B2B orders.
- Optimized 10+ **SQL** views powering dashboards for finance, logistics, & ops, improving query speed by 40%.
- Streamlined data exports to vendors via MFT, achieving 99% accuracy & timely distribution of client-bound files.
- Owned data integration efforts for B2B **data transmission** via Sterling Integrator, ensuring 100% on-time delivery.
- Fulfilled 20+ ad-hoc BI requests using **Excel (VLOOKUP, PivotTables)**, translating raw data into insights for stakeholders.
- **Queried** order data from Oracle SQL databases to extract SIM order numbers for tracking issues & business visibility.
- Automated sorting of SIM order numbers in Excel for **status message tracking**, reducing manual processing time by 60%.

PROJECTS

Instacart Market Basket Analysis (Tableau)

- Focused on identifying the top 10 products, peak traffic hours, best-selling product category, customer buying cycles, & most frequently repurchased items to uncover the bigger picture in consumer purchasing behavior & build a data story.

Credit Risk Analysis (Pandas, Scikit-learn, Plotly)

- Conducted credit risk analysis on German banking data using Python, applying logistic regression, random forest, & XGBoost to classify imbalanced data, improving recall by 30% over baseline & enabling better creditworthiness decisions.

Amazon Delivery Dataset analysis (AWS)

- Built an end-to-end logistics BI pipeline to analyze 40K+ deliveries & generate speed, distance, & delay risk metrics.
- Reduced reporting latency by 60% through metric engineering & pipeline optimization.