

SUMIT GUNDAWAR

+44 7810 081335 | sumitgundawar3@gmail.com | London, England, United Kingdom |

linkedin.com/in/sumit-gundawar-759470129/ | github.com/sumitgundawar | www.sumitgundawar.com/

PROFESSIONAL SUMMARY

Results-driven **Data Engineer** with a strong foundation in **ETL pipeline** development, **data modeling**, and **cloud-based** data solutions. Skilled in building **scalable data architectures** that optimize **analytics**, enhance **decision-making**, and **drive business** impact. Adept at collaborating with **cross-functional** teams to implement **data-driven** strategies and **automation** processes, ensuring efficient **data integration** and **reporting**.

SKILLS

Programming & Querying: SQL, Python, R, Java, NoSQL, MySQL, PostgreSQL, SAS

Data Engineering & ETL: Apache Spark, Databricks, Redshift, Oracle, NoSQL, ETL Pipeline Development, RESTful APIs, JSON, XML

BI & Analytics: Tableau, Power BI, Amazon Quicksight, Statistical Analysis, Customer Segmentation, Customer Behavior Analysis, Marketing Analytics

Big Data & Cloud: AWS (S3, Redshift, DynamoDB, EC2), Microsoft Azure

Machine Learning & Data Science: Forecasting, Predictive Modeling, Statistical Analysis, R, SAS, Matlab

Data Governance & Automation: Data Integrity, CI/CD, Data Warehousing, A/B Testing, Experimentation Design

PROFESSIONAL EXPERIENCE

LatentView Analytics

Chennai, TN, India

Data Analyst / Data Engineer

June 2021 - June 2023

- Engineered a multi-layered ETL pipeline that integrated data from multiple sources (Azure Data Lake, Teradata, Presto, SQL Server) to support demand forecasting across 25M+ customer-product combinations.
- Optimized data pipeline performance, reducing processing time from 13 to 8 hours by refining Spark jobs, minimizing redundancies, and improving query execution.
- Developed complex SQL-based data models to track procurement performance, supplier compliance, and inventory trends, ensuring accurate and actionable insights.
- Implemented advanced error handling and data validation mechanisms to maintain high data integrity, preventing inconsistencies in demand forecasts.
- Collaborated with PepsiCo's IT team to secure and allocate optimal compute resources, improving processing speed and reducing cloud costs.
- Designed and deployed interactive Power BI dashboards, enabling real-time monitoring of demand trends, supply chain risks, and procurement efficiency, contributing to a \$60M annual revenue impact

PROJECTS

Demand Planning Disaggregation

Chennai, TN, India

PepsiCo

September 2021 - June 2023

- Engineered a multi-layered ETL pipeline that integrated data from multiple sources (Azure Data Lake, Teradata, Presto, SQL Server) to support demand forecasting across 25M+ customer-product combinations.
- Optimized data pipeline performance, reducing processing time from 13 hours to 8 hours by refining Spark jobs, minimizing redundancies, and improving query execution.
- Developed complex SQL-based data models to track procurement performance, supplier compliance, and inventory trends, ensuring accurate and actionable insights.
- Implemented advanced error handling and data validation mechanisms to maintain high data integrity, preventing inconsistencies in demand forecasts.
- Collaborated with PepsiCo's IT team to secure and allocate optimal compute resources, improving processing speed and reducing cloud costs.
- Integrated geo-location data from PepsiCo's database into ETL pipelines to enhance regional demand forecasting across stores in multiple countries, optimizing supply chain decisions.
- Designed and deployed interactive Power BI dashboards, enabling real-time monitoring of demand trends, supply chain risks, and procurement efficiency, contributing to a \$60M annual revenue impact and visualizing geographic sales trends and regional performance insights.

Customer Behavior Analytics

Hilton

Chennai, TN, India

June 2022 - November 2022

- Engineered a comprehensive customer analytics pipeline analyzing 100+ behavioral features including booking patterns, travel preferences, demographics, and stay duration, enabling sophisticated customer segmentation and targeted marketing
- Developed and implemented advanced machine learning models to analyze complex patterns across multiple customer dimensions including booking lead time, room preferences, income levels, trip purpose (business/leisure), and length of stay
- Created automated data processing workflows to integrate diverse data sources and generate actionable insights for Hilton's marketing and revenue management teams
- Built interactive dashboards to visualize customer segmentation patterns and booking behavior trends, enabling data-driven decision making for marketing campaigns and personalized offerings
- Achieved significant business impact with 26% improvement in customer retention rate and approximately \$10M in annual revenue through targeted marketing initiatives and improved customer experience
- Implemented robust feature engineering pipelines to process and analyze large-scale customer behavior data, ensuring high-quality inputs for machine learning models

Understanding Socio-Demographic Factors Influencing COVID-19 Mortality

University of East London

London, England, United Kingdom

October 2023 - January 2024

- Conducted comprehensive statistical analysis using Python and R to investigate relationships between socio-demographic factors and COVID-19 mortality rates across England and Wales
- Implemented multiple linear regression models and factor analysis techniques, achieving 45% variance explanation in mortality predictions
- Engineered data preprocessing pipelines handling complex demographic datasets, performing correlation analysis and feature importance ranking
- Created insightful visualizations using Python libraries to represent mortality patterns across different demographic segments
- Applied advanced statistical tests (KMO, Shapiro-Wilk) and feature selection methods (Backward Elimination) to optimize model performance
- Developed a robust methodology combining multiple analytical approaches including correlation matrices, principal component analysis, and statistical hypothesis testing

EDUCATION

University of East London

Master's, Data Science

September 2023 - September 2024

GPA: 8.5

Vellore Institute of Technology

Master's, Computer Science

July 2019 - June 2021

GPA: 9.26

Savitribai Phule Pune University

Bachelor's, Computer Science

June 2016 - June 2019

GPA: 9.1

CERTIFICATIONS

Applied AI with Deep Learning (Coursera) – January 2024
Advanced Machine Learning and Signal Processing (Coursera) – September 2023
Analytics Bootcamp Program (Great Learning) – July 2021
Google Cloud Platform Fundamentals (Coursera) – May 2020
AI From the Data Center to the Edge (Intel) – March 2020

ACHIEVEMENTS

- Programming Competition Winner (Apr 2018) - Savitribai Phule Pune University
- Query Solving Competition Winner (Apr 2019) - Savitribai Phule Pune University
- Merit Scholarship (Sep 2020) - Vellore Institute of Technology
- Outstanding Performance Award (Sept 2021, Dec 2021) - LatentView Analytics
- Access and Participation Student Board Member (Jun 2024) - University of East London

WORK AUTHORIZATION

Authorized to work in the UK on a Graduate visa with full-time work rights. Open to relocation for international opportunities