



HITESH GUPTA

PROFESSIONAL SUMMARY

- Highly skilled **Business Analyst/Data Analyst/Data Scientist** with around **4 years** of experience, proficient in leveraging **Python** and **SQL** for comprehensive data analysis, visualization, and reporting to drive strategic business decisions.
- Employed advanced techniques in **Python**, **SQL**, and **R** to analyze and interpret complex datasets, including predictive modeling, machine learning, and deep learning algorithms, to uncover deep insights and trends.
- Proficient in leveraging **Snowflake**, **SQL Server**, and **Hadoop** for comprehensive data analysis, manipulation, and visualization to extract actionable insights from large and complex datasets.
- Designed and implemented complex **ETL** processes using **Data Bricks** to ensure optimal **data flow** and **integrity**.
- Developed interactive dashboards and reports using **Power BI**, **Tableau**, and **MS Excel**, featuring dynamic visualizations, drill-down capabilities, and real-time data integration for stakeholder engagement and decision-making.
- Mastered data wrangling and preprocessing techniques to ensure clean datasets for accurate analysis.
- Applied **advanced statistical methods** and techniques, such as **multivariate analysis**, **time series analysis**, and **Bayesian inference**, to perform comprehensive data profiling and analysis, guiding strategic decision-making processes with actionable insights and predictive models.
- Led cross-functional teams in **Agile (Scrum)** environments, driving collaborative efforts to deliver innovative, data-driven solutions within tight deadlines, leveraging advanced project management tools and methodologies to ensure successful project execution and stakeholder satisfaction.
- Implemented advanced version control strategies using **Git** and **GitHub** to maintain code integrity and facilitate collaboration.
- Skilled in **ML modeling**, **time series forecasting**, **fraud detection**, **pricing**, **A/B Testing**, **risk identification**, **deep learning**, dashboarding, building **Data flows**, **warehousing**, **ETL Pipelines**.

SKILLS

Functional Skills: Program Management, Management Reporting, Business Intelligence, Data Visualization, Data Analytics, Machine Learning, Predictive Modelling, Data Mining, Statistics, Big Data, Product Development, SDLC, SSIS, Agile (Scrum), Waterfall, ETL.

Programming/Technologies: SAS, SQL, Power BI, Tableau, Python, R, Teradata, Hadoop, Snowflake, Hive, Azure, AWS, GIT, Alteryx, PostgreSQL, Selenium Automation, Microsoft Suite (Excel, Word, PowerPoint, Visio)

PROFESSIONAL EXPERIENCE

Business Data Analyst Intern, NiSource, USA

May 2024 – Oct 2024

Streamlined operations and financial reporting for a utility company with ML, forecasting and improved insights.

- Implemented a **SARIMA-based time series forecasting** model to predict field operations overtime costs, achieving **\$1 million** in projected annual savings through improved workforce planning and operational efficiency.
- Coordinated with cross-function teams across **6 US states** to gather requirements to migrate **Financial Capital Close** reporting in Azure cloud, reducing analysis time by **90%** and creating a **Power BI dashboard**. Utilized **SQL** and **Excel** for the Customer Engagement team to identify customer issues across states and calculate KPIs like resolution time, field issues, and repeated callers, enhancing customer support insights.
- Improved budgeting efficiency by 30% and reduced monthly budget variance by training a machine learning model using **Data Bricks** (python) and **Azure ML Ops** to predict pipeline **repair/replacement** costs, achieving a prediction accuracy of **74%** using **ensemble methods**. Wrote scripts for testing model by implementing performance metrics (e.g., MAE, RMSE, R^2) and integrating into **CI/CD** pipelines.
- Leveraged **Azure services**, including **ADF**, **Synapse**, and **Azure Gen Lake**, to design and implement a robust data analytics platform, enhancing processing efficiency for service delivery team, storage scalability, and real-time data availability through automated pipelines for downstream analytics.
- Demonstrated expertise in data quality assurance and testing methodologies to ensure the accuracy and reliability of analytical results.

Data Science Analyst, Code Byte, Australia

Dec 2022 – Aug 2023

Developed pricing strategies and risk models using machine learning, driving optimized solutions for an insurance client.

- Developed a **pricing model** for the underwriting team of insurance clients, reducing pricing errors by **30%**, increasing underwriting efficiency by **25%**, and boosting policy profitability by **15%** through more precise risk-based pricing using ensemble ML models.
- Enhanced analysis and visualization by developing data architecture in **Snowflake** to combine data from multiple tables and integrating it with **Tableau** dashboards, resulting in improved data accessibility for stakeholders. Delivered actionable insights by building 10 self-serve Power BI dashboards, visualizations, and interfaces with Snowflake, empowering the portfolio management team to make data-driven decisions with increased efficiency.
- Conducted **A/B testing** on different insurance packages for clients, optimizing pricing and feature offerings based on data-driven

insights to increase customer conversion rates by 7%.

- Increased business efficiency by administering the evaluation of business requirements and formulating **30+ user stories** on Azure DevOps, streamlining the tracking of work progress and project management.

Senior Data Analyst, Hero Motocorp Ltd, India

Jun 2019 – Dec 2022

Improved warranty analytics by leveraging predictive models, fraud detection, and dashboards.

- Leveraged Python to integrate **SAP ERP** and **API** data sources, exporting and analyzing data to identify key fraud patterns. This resulted in a 25% reduction in fraudulent claims, saving approximately **\$1.2 million annually**.
- Architected robust data storage solutions for warranty data (big data) leveraging **Azure Data Lake Storage** and **Blob Storage** and build **Data Flows** and **ADF Pipelines**, facilitating secure and scalable data management to support big data analytics and machine learning workloads.
- Enhanced business decision-making and efficiency by **23%** through preliminary data exploration on large datasets using **Python** (NumPy and Pandas) and **SQL** for data extraction and manipulation, generating actionable insights to drive strategic initiatives.
- Developed a **time series model** for predicting warranty part failures, enabling proactive inventory management and reducing downtime by **10%** through accurate demand forecasting.
- Analyzed dealers' data using **Power BI Dashboard** (using complex DAX and M Language), providing actionable recommendations to senior management finally reducing warranty cost by 22% in specific models.
- Adhered to Agile sprint planning sessions and effectively communicated ongoing findings and recommendations, which contributed to more timely project completions, enhancing team efficiency and reducing cycle times for project deliverables.
- Demonstrated expertise in **statistical analysis techniques**, including **correlation analysis**, **hypothesis testing**, and **regression modeling**, to derive actionable insights from data.

Business Intelligence Intern, Hero Motocorp Ltd, India

Jan 2019 – Jun 2019

Built dashboards and ETL pipelines to deliver actionable insights and optimize data workflows.

- Worked with **data engineers** to build **ADF pipelines** and **data flows** migrate data from **Azure Blob storage** to dedicated SQL Pool in synapse by using ETL process both in ADF and Databricks (using **PySpark** and **SQL**) for complex data transformations. Designed and implemented a **Power BI** dashboard by integrating with SAP for the Tool Engineering team to monitor key business KPIs (CPC, rejection rates), delivering actionable insights to senior management that led to a 12% cost reduction

EDUCATION

Stevens Institute of Technology, School of Business – Hoboken, NJ

Dec 2024

Master's in Business Analytics & Data Science

(GPA: 4.0/ 4.0)

Relevant Coursework: Corporate Finance, Data Models and Decisions, Machine Learning, Statistics, Linear Programming, Database Management, Applied AI, Natural Language Processing.

PEC University of Technology – Chandigarh, India

May 2020

Bachelor of Technology, Engineering

Rank 9 (GPA: 9.0/ 10.0)

ACADEMIC PROJECTS

- **Customer Segmentation | Market Basket Analysis | Python** – Built an RFM model and market basket analysis in Python for a retail brand, using percentile segmentation and the Apriori algorithm to enhance customer retention strategies.
- **Selenium Data Mining | Automation | Python** – Build a Data Mining Bot as a Research Assistant for Professor helping in web scrapping articles title. Abstracts and keywords from various journals.
- **Customer Churn Analysis| Business Intelligence | Power BI, Power Query, Dax** - Conducted an in-depth Customer Churn Analysis for DoorDash using Power BI, Power Query, and DAX, leveraging advanced data processing techniques and creating dynamic visualizations. Analyzed customer churn patterns by leveraging data on customer demographics, behavior, and interactions, with the aim of identifying key factors influencing churn and implementing strategies to enhance customer retention.
- **Natural Language Processing (NLP) | Python** – Built a website classifier using Natural Language Processing (NLP), comparing text representation methods (Bag of Words, TF-IDF, Word Embeddings) to evaluate their efficiency and feature space effectiveness.
- **NYC Taxi Data Analysis | Python** – Conducted exploratory data analysis on NYC taxi trip data using Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn, identifying patterns and anomalies that optimized taxi fleet management.
- **Fraud Detection in Medi-Claims using Ensemble ML Models | Azure, Data Bricks, Python** – Developed an end-to-end fraud detection pipeline with ADF, ADLS, Databricks, and ML (KNN, GB, RF), using feature engineering and factor.

CERTIFICATIONS

- Certified Project Management Professional (PMI), Microsoft Power BI Analyst (PL-300), Spreadsheet Modelling (HBS), Azure Cloud Fundamentals (AZ-900), Quantitative Methods (HBS), Corporate Finance (HBS), Google Business Intelligence Certificate, AWS Certified Cloud Practitioner, Tableau Essential Training, Snowflake Hands On Essentials - Data Warehouse, Introduction to Data Warehousing, Getting Started with Amazon Redshift