hiteshguptaus@gmail.com | https://www.linkedin.com/in/hiteshg1318/
https://github.com/hiteshg1318 | (201)-469-1754

HITESH GUPTA

PROFESSIONAL SUMMARY

- Highly skilled **Data Analyst/Business Analyst** 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 Pipelines** processes using **Data Bricks** to ensure optimal **data flow, data warehousing** and **integrity**.
- Applied **advanced statistical methods** and techniques, such as **multivariate analysis**, **time series analysis**, **A/B Testing,** and **Bayesian inference**, to perform comprehensive data profiling and analysis, guiding strategic decision-making processes with actionable insights and predictive models.

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.
- Integrated datasets from **SQL Server** and **AWS Redshift**, reducing report generation time by 30%, ensuring reports are available within 10 minutes of data extraction
- 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.

***Senior Data Analyst*, KPMG,** India **Aug 2020 – July 2023**
Improved analytics by leveraging predictive models, fraud detection, and dashboards.

- Participated in a special financial analysis project, leveraging **Python** (**NumPy** and **Pandas**) to extract, clean, and analyze large financial datasets, ensuring data quality and preparing it for in-depth analysis, writing efficient code weekly.
- Conducted statistical analysis and hypothesis testing using **Python** and **R** libraries, providing data-driven recommendations that significantly contributed to customer retention and sales growth.
- Conducted comprehensive auditing of financial data using **SQL**, identifying discrepancies and ensuring data integrity across multiple financial databases, which improved compliance and accuracy in financial reporting.
- Utilized **MySQL** database to query, manipulate, and manage data, ensuring efficient retrieval and storage of critical business information, executing complex **SQL** queries, and maintaining databases.
- Created interactive and insightful dashboards in **Power BI** to visualize key financial metrics and present data-driven insights to stakeholders, developing and maintaining dashboards and reducing financial report generation time.
- Employed **Git** for version control to manage code changes, collaborate efficiently with team members, and maintain a comprehensive record of data analysis projects, handling code commits and pull requests weekly.
- Utilized **Azure data lake** for scalable storage solutions ensuring data durability and high availability, and Synapse for fast querying and

analysis of large datasets to support data-driven decision-making.

- Conducted rigorous **A/B testing** on models and strategies annually, providing empirical data that influenced the optimization of business processes and marketing campaigns
- Implemented the **Waterfall methodology** for structured and sequential project execution, ensuring thorough documentation and clear milestones at each phase of data analysis projects, resulting in enhanced project clarity and adherence to timelines.

***Technical Data Analyst*, Vedanta,** India **July 2019 – Aug 2019**

Streamlined data processes and automation to boost efficiency, visibility, and decision-making.

- Engineered optimized **SQL queries** for real-time data retrieval, reducing reporting delays by 40%, resulting in faster decision-making across operational teams.
- Created and deployed automated **Power BI** dashboards, reducing troubleshooting time by 30% and improving visibility into production KPIs by 20%.
- Designed a **centralized MySQL database solution**, cutting query processing time by 15 minutes per query and improving access to critical production data.
- Streamlined **ETL processes** using **SSIS** and **Talend**, automating data transformation and saving over 40 hours per month in manual processing.
- Applied **data mining techniques** to analyze production data, uncovering inefficiencies that led to a 20% improvement in production line efficiency.

***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.
- **E-Commerce Customer Segmentation | KMeans Clustering, Python, Tableau –** Developed a customer segmentation model to enhance targeted marketing strategies and improve user experience by identifying distinct customer groups through data collection, cleaning, feature engineering, normalization, and applying KMeans clustering, with visualization using PCA.
- **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