Vishal Karur

Vxk210096@utdallas.edu | 1-469-493-6566 | LinkedIn | Portfolio | GitHub | Dallas , TX 75080

EDUCATION

The University of Texas at Dallas, Texas Aug 2022 - May 2024
Master of Science - Business Analytics (Data Science and Data Engineering Specialization)
Courses: Big Data, Applied Machine Learning

Visvesvaraya Technological University, Bangalore, India. June 2019
Bachelors in Electronics and Communication Engineering **GPA-3.6**

TECHNICAL SKILLS

Programming Languages: Python, R, Structured Query Language (SQL)

Tools/Software:	Azure Data Studio, DataBricks, Pyspark, Azure, Tableau, Microsoft Excel, Power BI, Oracle SQL, MySQL, NoSQL, PostgreSQL, Amazon Redshift & S3, AWS, ETL , Hadoop
Libraries:	Pandas, NumPy, Seaborn, Matplotlib, ggplot2, dplyr, Plotly.
Version Control Tools:	Git, GitHub
Domain Knowledge:	Data Analytics, Business Intelligence, Data Modelling & Warehousing, Quantitative Analysis, problem-solving, Computer Science, AB Testing, hypothesis testing, Statistical methods, ad-hoc analysis,
Certifications:	Tableau, Python, Advanced SQL, DataBricks

WORK EXPERIENCE

Navy Federal Credit Union, Vienna, VA May 2023 – Aug 2023
Data Engineer Intern (Azure DataBricks, Pyspark)

- Leveraged Azure Data Studio for efficient data exploration and analysis of customer member transcripts and messages, extracting valuable insights to support strategic decision-making by the leadership team.
- Significantly improved data pipeline efficiency by implementing code optimizations in Databricks and PySpark, resulting in a savings of 10 hours per week in data processing and transformation, allowing the Modeling team to access high-quality datasets faster.
- Single-handedly led the analysis of the Wave 0 rollout, analyzing call transcripts and messages. This effort enabled us to identify the top 3 issues that customers are concerned with regarding the new platform.
- Collaborated in an Agile environment, working in sprints, and utilizing Azure DevOps for seamless project management, version control, and continuous integration/continuous delivery (CI/CD), enabling streamlined teamwork and efficient project execution.

Accenture, Bangalore, India Sep 2019 - May 2022
Data Analyst (SQL, ETL, Python)

- Built a high-performance data pipeline using ETL transformations on multiple source data so that the BI team can analyze.
- Structured data was presented to various Business Units/Data Warehouse teams by integrating over 10 million rows of raw data from diverse sources such as Excel, JSON, API, and Database which helped the team for further analysis.

Key Contributions

- Improved Business KPI Metric for RFP (Percentage of lines cleared without manual intervention) from 88.6% to 94.21%.
- Built the data and reporting architecture from the scratch using SQL and ETL(Informatica) to implement Business KPIs.
- Designed & developed a 95 % accurate reporting system to correctly identify the returned products that boosted the monthly collection of about $100k in chargeback payments.
- Developed and deployed a Python script to automate the ETL pipelines by downloading CSV from a website, modifying, and loading data into a database using business logic via a daily interface thus saving 4 hours of human effort per week.

IPrimed, Bangalore, India June 2018 - Jan 2019
Data Analyst Intern (Excel)

- Cleaned and Validated the Bike Sales dataset using VLOOKUP and other Excel functions to remove duplicates and noise in the dataset.
- Performed Exploratory Data Analysis (EDA) on Bike Sales dataset using Advanced SQL to the correlation between different parameters in the dataset.
- Developed a dynamic analytical chart to determine the top 5 regions by profit margin to further invest and increase market spending.

ACADEMIC PROJECTS

Web-Scraping Amazon e-commerce (Python & Tableau) Jan 2023

- Created a web scraping model to identify and download (CSV) the top 10 products from Amazon across several categories.
- Improved the model by adding the smtplib module, which sends automated e-mails when the price of a specific product reduces by 10%.

New York Cab Analysis (NumPy, Pandas, Power BI) Aug 2022

- Cleaned and analyzed the New York cab dataset using NumPy and Pandas to uncover crucial insights and correlations between metrics such as peak hours, average tip per ride, average distance travelled, and average speed.
- Visualized the Data in Power BI to determine the most profitable time range for both cab drivers and clients.

Covid Data Analysis (Advanced SQL) Dec 2021

- Conducted a detailed Exploratory Data Analysis (EDA) on covid data set using SQL to determine the ongoing patterns.
- Identified crucial insights such as the country with the highest number of current cases, the highest number of recoveries, and the vaccination rate per country, among others.