**Gajarajan Appandarajan**

United States |331-243-6002| [gajarajan28@gmail.com](mailto:gajarajan28@gmail.com) |LinkedIn - [Gajarajan](https://www.linkedin.com/in/gajarajan-ap/)

# Summary

Experienced Data engineer with a two-and-a-half year and pursuing a master’s in information systems. Proficient in SQL, Python, and Power BI for designing strategic data solutions, extracting insights, driving data-driven decisions in complex business environments, and seeking opportunities to apply expertise in complex business environments.

# Work Experience

**Royal Cyber,** **Naperville, IL (Internship) May 2023- Oct 2023**

* **Assisted in implementing data cleansing techniques**, utilizing tools like **Snowflake** and **Python** to improve the accuracy and reliability of large datasets. This effort led to a 20% reduction in data errors, enhancing the overall data quality.
* **Collaborated with senior analysts** to design and automate data pipelines using **Apache Spark** and **Azure Data Factory**. Contributed to streamlining Big Data processing, reducing manual efforts by 30%, and expediting data delivery for quicker decision-making.
* **Supported the development of predictive models** using Python for customer data analysis. Leveraged **machine learning algorithms** and **Big Data technologies** like **Spark** and **Snowflake** to improve sales and demand forecasting accuracy by 15%.
* **Played a key role in creating interactive Power BI dashboards** that segmented customers based on transaction patterns, locations, and other key performance indicators (KPIs). These dashboards provided valuable insights, contributing to a 20% increase in Q1 revenue.
* **Assisted in generating detailed cost analysis reports**, working closely with the finance team. Utilized **Spark**, **Azure Data Factory**, and **Snowflake** to provide critical insights that informed pricing decisions, resulting in a 5% improvement in profit margins.

**Power Links, Chennai (Data Analyst Associate) May 2020- June 2022**

* Collaborated with team members to understand client requirements and the design of dashboards in **Tableau**, ensuring the layout, look, and feel met client specifications.
* Designed and developed various dashboards and reports using Tableau Visualizations such as bar **graphs, line diagrams, funnel charts, scatter plots, pie charts, heat maps, and tree maps**. Utilized data storytelling to effectively present insights and track **key performance indicators (KPIs**).
* Conducted exploratory data analysis **(EDA)** in Python, resulting in a 25% improvement in data accuracy and quality by addressing null values, irrelevant data, duplicates, and typos. Applied **statistical analysis, data cleaning, and data validation techniques**.
* Leveraged Python for API integration to facilitate seamless data extraction and deployment on the Google Cloud Platform. Built and managed an ETL pipeline using Airflow, which automated data processing and reduced manual effort by 40%. Employed data integration and transformation skills to prepare data for detailed analysis and visualization in Tableau.
* Developed and optimized complex SQL queries, **including joins, sub-queries, and stored procedures**. Gained expertise in query optimization and performance tuning in MySQL, achieving a 40% improvement in query performance and reducing production downtime by 25%.
* Created detailed sales reports to track performance metrics and trends. Data analysis and reporting tools provided actionable insights that informed sales strategies and decision-making.
* Engaged with stakeholders to gather requirements and feedback, ensuring that dashboards and reports met business needs and provided valuable insights. Used data visualization best practices to enhance stakeholder presentations and support data-driven decisions.
* Applied advanced technical skills in **Python, SQL, Tableau, and Google Cloud Platform** to deliver high-quality data solutions. Demonstrated strong problem-solving abilities by addressing data discrepancies and performance issues, contributing to process improvements and operational efficiency.
* Utilized data visualization techniques to translate complex data sets into meaningful visual representations, facilitating better understanding and strategic decision-making for stakeholders.

# Skills

**Data Visualization** Power BI, Tableau, GCP Looker, QlikView

**Programming Skills:** R - tidyr, tidyverse, dplyr, ggplot, and other data science libraries

Python - numpy, Pandas, SciKit, Sklearn, TensorFlow, Keras, matplotlib, seaborn, plotly

Java, JavaScript, SAS, Scale, XML

**Web technology** ReactJS, REST API, GraphQL, HTML, CSS, Postman

**Database:** GCP cloud storage, MS SQL, RDBMS, MongoDB, Oracle, Microsoft SQL Server, Azure SQL

**BI Tools:** Google Analytics, Microsoft Office Suite (Microsoft Word, MS Excel – Pivot, X-lookup, V-lookup,

macros), Kafka, Big data, SAS, Databricks, ETL (Alteryx), Azure Data Factory, spark,

Hadoop, Airflow, Informatica ,SPSS

**Technical Skills:** Clustering Techniques, Probability and Statistics, Machine Learning Concepts, Descriptive

Analysis, Predictive Analysis, Statistical Modeling, GitHub, MS PowerApps

**Soft skills:**  Communication, multitasking, Teamwork, leadership, Analytical, and Critical Thinking

**Cloud:** GCP, Snowflake, AWS, Azure Data Lake, Azure SQL Database, Azure ML, and Azure Data Factory

# Project

**Superstore Sales Analysis**

* Led a comprehensive Superstore Sales SQL project (2014–2017) focusing on customer segmentation and operational enhancements, leading to optimized decision-making.
* Automated **ETL pipeline** to extract, transform, and load data from multiple sources, executed indexing and query restructuring to **optimize database by 80% and swift data retrieval**.
* Created an **interactive Power BI dashboard** to identify trends and segments, showcasing SQL's pivotal role in analysis and strategic decision-making

**Patient Readmission Risk**

* Collected and preprocessed a comprehensive dataset from electronic health records (EHRs), including patient demographics, medical history, lab results, and discharge summaries. Used **Python** with libraries like **Pandas** and **NumPy** for data cleaning, normalization, and handling missing values.
* Created and engineered relevant features such as comorbidity scores, medication adherence metrics, and previous hospitalization history. Applied **feature scaling** and **encoding** techniques to prepare the data for model training.
* Performed EDA using **Matplotlib** and **Seaborn** to visualize data distributions, correlations, and patterns related to patient readmission. Identified key risk factors and trends that informed model development.
* Developed and evaluated several machine learning models including **Logistic Regression**, **Random Forest**, and **Gradient Boosting Machines (GBM)** using **scikit-learn**. Achieved a high performance with an AUC-ROC score of 0.85.
* Implemented a deep learning model using **TensorFlow** and **Keras**. Designed a **Multilayer Perceptron (MLP)** with several hidden layers to capture complex patterns in the data. Fine-tuned the model with **dropout** and **batch normalization**, achieving an AUC-ROC score of 0.88.
* Applied **GridSearchCV** and **RandomizedSearchCV** for hyperparameter tuning to optimize model performance and prevent overfitting.
* Assessed model performance using precision, recall, F1-score, and confusion matrix metrics. The final deep learning model demonstrated superior predictive power and reliability.
* Developed interactive dashboards using **Power BI** to visualize prediction results, risk factors, and patient profiles. Presented findings to healthcare professionals, providing actionable insights to reduce readmission rates and improve patient care.

# Education

|  |  |  |
| --- | --- | --- |
| **Cleveland State University, Cleveland, OH**  Master’s in information systems **GPA:3.63 (MIS)**  **Course-** business analysis, Enterprise database, system thing, modern database | | Aug 2022- May 2024 |
| **Madras Institute of Technology, India**  Bachelor of Electronics and Communications Engineering | Aug 2017 - May 2021 | |

**Certificate**

Cisco Certified Network Associate- CSCO14413144 JULY 2023- JULY 2026

Certified Ethical Hacker (CEH)- ECC3629758104 AUG 2023- AUG 2026

PL-300 - Udemy