

# KEERTHI KRISHNA AIYAPPAN

☎ 315-374-9628 ✉ [keerthikrishna.official@gmail.com](mailto:keerthikrishna.official@gmail.com) 🔗 [linkedin.com/in/keerthi-krishnaa](https://www.linkedin.com/in/keerthi-krishnaa) 🌐 [keerthikrishna.vercel.app](https://keerthikrishna.vercel.app)

## SUMMARY

Data Science professional with **3+** years of experience delivering high-impact data solutions across sports, healthcare, insurance, and technology domains. Proven track record in **Predictive Modeling**, **Data Pipelines**, **Automated Reporting** and **Optimization** techniques, driving measurable business outcomes. Expertise in developing full-stack analytics solutions from **ETL** to **visualization** using **Python**, **R**, **SQL**, **Apache Spark**, and cloud technologies (**AWS**, **Azure**).

## EXPERIENCE

### GSK (ViiV Healthcare)

Jun 2024 – Present

*Data Strategy and Operations Associate*

*Durham, NC, USA*

- **Hypothesis Testing:** Deployed a **Hypothesis Testing** algorithm using **R** and **Apache Spark** within an **Azure Databricks** environment to evaluate a parallel data strategy, validating the reduction of duplicate patient records from 4.7% to 0.05% across two datasets over a 6-month period, influencing the team's decision to implement the new data stewardship strategy
- **Automated Claims Reporting Pipeline:** Automated the generation and reporting of 150+ slides of HIV claims data using **Plotly** and **Databricks UI** by creating a scalable data pipeline in **Databricks**, reducing manual slide creation time by 80%
- Analyzed weekly sales data for two products, using **DAX** in **Power BI**, **Window Functions**, **Views** and **CTEs** in **SQL** in an **Azure Databricks** environment, delivering trend and forecast reports to 5+ internal business partners
- **Anomaly Detection:** Collaborated as a team of two to build an **anomaly detection** and data monitoring dashboard using **Power BI** and **SQL**, implementing logics to identify irregular data changes and improve incoming data quality

### NA Sportz Interactive

Nov 2021 – Jun 2023

*Data Science and Analytics Associate*

*Mumbai, MH, India*

- **Marketing Mix Modeling:** Collaborated as a team of three to build a **Marketing Mix Modeling** solution for a sports broadcasting client leveraging **Shapley Methods**, **ARIMA**, **Multiple Linear Regression** and **Bayesian Optimization**, optimizing ad spending across digital and traditional channels to increase viewership by 8%
- **XG Predictive Model:** Designed an Expected Goals predictive model using **XGBoost** and **Gradient Boost** with 8 features, achieving an F1 Score of 93% when evaluated on 350,000 shots, enabling precise prediction of scoring probabilities
- **Marketing Optimization:** Developed an **A/B Testing** model using **Python** to optimize MatchCentre icon placement, increasing click-through rate from 4% to 13% with over a 3-week testing period, enhancing user engagement metrics analysis
- **KPI Reporting:** Identified and automated 25+ KPIs to track soccer player and team performance metrics, leveraging **tidyverse** and **ggplot** in **R**, and **CTEs**, **Window Functions**, and **Joins** in **SQL**, delivering daily reports to clients and senior management
- **Player Clustering Model:** Built a clustering model using **K-Means Clustering**, **Principal Component Analysis** and **Scikit-learn** to group 60+ Indian soccer players based on their roles using 30+ features, enhancing player evaluation capabilities
- **Data Pipeline Architecture:** Engineered and maintained 3 scalable **ETL pipelines** leveraging **Apache Spark**, **Snowflake**, and **dbt**, transforming real-time scoring data into tables for dashboarding and reporting within the data warehouse ecosystem
- **Optimization:** Streamlined the creation of sport highlight packages using **SciPy Optimizers**, cutting video editing time by 80%
- **Forecasting Match Results:** Implemented a forecasting model using **ARIMA** to predict match outcome probabilities before ISL games with 76% accuracy, enabling commentators to deliver data-driven narratives about potential game outcomes

### BilledRight Healthcare Solutions

Jun 2020 – May 2021

*Data Scientist*

*Chennai, TN, India*

- **Scalable ETL Pipelines:** Managed the **ETL** of 10+ medical billing related business processes using **Glue** and **Athena** within **AWS**, transforming business processes into data tables for reporting and dashboards
- **Insurance Risk Detection:** Led development of a health claims risk stratification model using **Random Forest** and **Gradient Boost**, achieving 82% ROC-AUC, incorporating 20+ features on 50,000+ records to enable personalized premium structuring
- **Ad-hoc Queries:** Handled 100+ ad-hoc medical billing queries, using **Window Functions** and **CTEs** in **SQL** to analyze claim patterns and reimbursement trends, improving client understanding of denial rates and revenue cycle performance
- **Provider Segmentation:** Performed data cleaning, transformation, and feature engineering for a provider segmentation initiative leveraging **Python** and **PySpark**, enabling identification of high-value providers generating significant volume and revenue, contributing to improved collection and allocation strategies

## EDUCATION

### Syracuse University

Aug 2023 – May 2025

*Master of Science in Applied Data Science*

*Syracuse, NY, USA*

## SKILLS

**Programming Languages and Data Visualization :** Python, R, SQL, Spark, Tableau, MS Excel

**Cloud and ETL :** Snowflake, Docker, dbt, Hadoop, Airflow, Kubernetes, BigQuery, Kafka, MongoDB, Minio, Cassandra, Hive

**Libraries :** Scikit-Learn, Tensorflow, Keras, PyTorch, Plotly, Streamlit, Statsmodels, Spacy, SciPy, Shiny, Caret, SparkMLlib

**Modeling :** Regression, Classification, Clustering, Anomaly Detection, NLP, Deep Learning, Forecasting, Generative AI

**AWS Services :** Lambda, Athena, S3, Sagemaker, Redshift, QuickSight, CloudWatch, Glue, RDS, EC2

**Azure Services :** Databricks, Data Factory, Data Lake, Synapse, Power BI

## LEADERSHIP

- **Lead Data Consultant, Indian Super League (2022–23):** Acted as liaison between broadcasters, organizers, and data teams, demonstrating cross-functional collaboration skills, while at NA Sportz Interactive
- **Technical Lead, Automated Claims Reporting at ViiV Healthcare:** Spearheaded end-to-end automation of HIV claims reporting, resulting in 80% reduction in manual processing time, streamlining the generation of 150+ data slides