KEERTHI KRISHNA AIYAPPAN

📞 315-374-9628 ✉ keerthikrishna.official@gmail.com 🔗 linkedin.com/in/keerthi-krishnaa

EXPERIENCE

GSK (ViiV Healthcare) - via Atrium   Jun 2024 – Present
Data Strategy and Operations Associate   *Durham, NC, USA*
- Deployed a Hypothesis Testing algorithm in R, measuring the impact of a parallel data strategy on patient data stewardship on 3+ metrics in an Azure Databricks environment, validating a 5% reduction in data duplicates
- Automated the generation of 150+ slides of HIV claims data using Plotly, 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
- Collaborated as a team of two to build an anomaly detection and data monitoring dashboard using Power BI, implementing logics to identify irregular data changes and improve data quality

NA Sportz Interactive   Nov 2021 – Jun 2023
Data Science and Analytics Associate   *Mumbai, MH, India*
- Engineered an Expected Goals predictive model using xGBoost and Gradient Boost with 8 features, transforming the company's analytics capabilities by yielding an accuracy of 93% in predicting the likelihood of a shot ending in a goal
- Developed a multivariate A/B Testing model, using SciPy and Statsmodels, to optimize the MatchCentre icon placement on the home page of a client website, resulting in a 9% increase in page clicks
- 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
- 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 performance evaluation capabilities

BilledRight Healthcare Solutions   Jun 2020 – May 2021
Data Scientist   *Chennai, TN, India*
- Crafted over 10 dashboards in Tableau identifying 10+ KPIs, enhancing the visualization of weekly reporting processes'
- Operated as the technical lead to build a health claims risk stratification model using Random Forest and Gradient Boost, assessing the risk of policyholders with 85% accuracy, enabling clients to set personalized insurance premiums
- Handled 100+ ad-hoc queries, using Window Functions and CTEs in SQL, improving client data understanding
- Implemented a clustering solution to bucket 70+ key doctors based on their product preferences and communicative tendencies using K-Means Clustering and DBSCAN, enhancing sales representatives' tailoring of product pitches

PROJECTS

Bayesian Marketing Mix Modeling | *Scikit-Learn, SHAP Methods, SciPy Optimizers*   Feb 2024 – May 2024
- Designed a Marketing Mix Modeling solution for a shampoo brand leveraging Shapley Methods and Scipy Optimizers using Sagemaker and S3 buckets within AWS, optimizing spending across marketing channels to increase revenue by 6%
- Used Naïve Bayes and Multiple Linear Regression to build the final solution with an accuracy of 87% after hyperparameter tuning using GridSearch, taking Adstock and Carryover effects into consideration

Price Elasticity Modeling and Optimization | *PySpark, SciPy Optimizers*   Oct 2024 – Dec 2024
- Designed a Price Elasticity Modeling solution for a shoe manufacturer using PySpark for modeling and L-BFGS-B and SLSQP methods in SciPy for optimization to determine optimal shoe pricing, increasing revenue by 3.6%

Data Warehousing Solution for NY Metro Authority | *SQL, Snowflake, dbt, Power BI*   Sep 2024 – Dec 2024
- Implemented a data warehousing solution covering 5 business processes for the NY Metro Authority, adopting Enterprise Bus Architecture techniques, using Snowflake for data storage, dbt for ETL processes, and Power BI for reporting

EDUCATION

Syracuse University   Aug 2023 – May 2025
Master of Science in Applied Data Science   *Syracuse, NY, USA*

Bharathiar University   Aug 2016 – May 2019
Bachelor of Science in Visual Communication and Electronic Media   *Coimbatore, TN, India*

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

- Served as the Lead Data Consultant for the 2022-23 season of the Indian Super League, acting as the data science liaison between broadcasters, organizers and my employers