SETHU S

**Data Scientist | Statistician | Data Analyst**

[**sethus4791@gmail.com**](mailto:sethus4791@gmail.com)| [**github.com/itsmesethus**](https://www.github.com/itsmesethus) | [**Portfolio: sethus4791portfolio.streamlit.app/**](https://sethus4791portfolio.streamlit.app/)

[**kaggle.com/sethu123123**](https://www.kaggle.com/sethu123123)| **Villupuram, Tamil Nadu-605103** | **+91-8122815260 |** [**linkedin.com/in/sethus4791**](https://www.linkedin.com/in/sethus4791)

# Summary

Resolute Data Scientist and Analyst with a strong record of extracting actionable insights from complex data. Proficient in Statistical Analysis, Hypothesis Testing, and Data Modelling, utilizing Python, SQL and R. Demonstrated leadership through project delivery and

a commitment to innovation. Pursuing further expertise in data-driven solutions to contribute to team success.

# Experience

**EVOSCIEN Pvt, Statistician – Bengaluru, India May 2023 - Present**

* Enhanced decision accuracy by 25% through data acquisition, consolidation, and advanced statistical methods and analysis.
* Developed and implemented statistical models, algorithms, and techniques, yielding insights that led to 15% improvements in decision Making processes.
* Executed appropriate statistical tests (t-tests, ANOVA, chi-square tests, and non-parametric tests) to analyze data and uncovered meaningful patterns, trends, and relationships with a 95% confidence interval.
* Collaborated with the Entomology & Engineering departments using a data-driven approach, improving operational efficiency by 15%.
* Produced 20+ detailed reports and interactive data visualizations, clearly communicating analytical findings to both technical and non-technical audiences.

# Education

* **8.61/10 M. Sc in Statistics,** Bharathiar University **|** Coimbatore, Tamil Nadu **Sep 2021 – May 2023**
* **9.39/10 B. Sc in Statistics,** Arignar Anna Govt Arts College **|** Villupuram, Tamil Nadu  **Jun 2018 – May 2021 Courses:** Descriptive Statistics | Sampling Theory | Probability Theory | Statistical Estimation Theory | Statistical Quality Control | Multivariate Statistical Analysis| Distribution Theory | Econometrics | Statistical Inference | Programming in R | Design of Experiments | Stochastic Processes.

# Skills

**Programming**: Python, R, SQL.

**Tools**: Microsoft Power BI, IBM SPSS, MINITAB, STATISTICA, Microsoft Excel, MYSQL, Jupyter Notebook, Git, Github.

**Libraries**: Pandas, Numpy, Matplotlib, Seaborn, Plotly, Scipy, Scikit-Learn, TensorFlow, Keras, Statsmodels, Pingouin, Streamlit.

**Data Preprocessing & Others**: Data Cleaning, Data Analytics, Data Analysis, Exploratory Data Analysis (EDA), Data Mining, Feature Engineering, Feature Selection Techniques, Data Visualization, Outlier Detection, Correlation Analysis, A/B Testing, Ad Hoc Analysis, Model Evaluation, Model Deployment, Advanced Analytics, Hyperparameter Tuninig.

**ML Expertise**: Regression, Classification, Clustering, Predictive Modelling, Quantitative Analysis, Statistical Modelling, Deep Learning, CNN, RNN, LSTM, GRU, Bidirectional Models, Time Series Analysis, Forecasting.

# Projects

**Chrono-Power Forecast: Electricity Demand Forecasting for UK Year-2024** - [link](https://github.com/itsmesethus/CHRONO-POWER-FORECAST-ELECTRICITY-DEMAND-FORECASTING-FOR-UNITED-KINGDOM-FOR-THE-YEAR-2024-) **Dec 2023 ‑ Jan 2024**

* Implemented diverse forecasting models including Exponential Smoothing, ARIMA, SARIMA, Fb-Prophet models in Phase 1.
* Investigated deep learning architectures like RNN, LSTM, GRU, and hybrid Convolutional-RNN models for with 3 output forecasting for Phase 2.
* Enhanced the precision of deep learning models, achieving 20% better results for ND, TSD, EWD.

**Cognizant AI Virtual Internship - Machine Learning for Gala Grocery Retail Pricing** - [link](https://github.com/itsmesethus/COGNIZANT-ARTIFICIAL-INTELLIGENCE-VIRTUAL-INTERNSHIP-Machine-Learning-for-Gala-Grocery-Retail-) **Sep 2023 ‑ Oct 2023**

* Performed in-depth EDA, establishing a sturdy foundation for modelling and optimization across 10+ product categories.
* Applied Ensemble Regressors (Adaboost, Bagging) and ANN-based Multiple Linear Regression, achieving over 94% R² for predictions.
* Validated forecast reliability using 3+ evaluation metrics, refining ML models with RandomizedSearchCV.

**BCG Virtual Internship - Churn Predictive Modelling to Customer Retention** - [link](https://github.com/itsmesethus/BCG-DATA-SCIENCE-VIRTUAL-INTERNSHIP---CHURN-BUSTER-PREDICTIVE-MODELING-TO-CUSTOMER-RETENTION-) **Aug 2023 ‑ Sep 2023**

* Employed various predictive models including Ensemble classifiers and KNeighbors Classifiers, achieving an F1 score exceeding 95%.
* Integrated advanced techniques such as SMOTE and Yeo-Johnson transformation, leading to a 20% reduction in prediction errors.
* Interpreted machine learning insights for customer retention and developed an impactful Power BI dashboard report with 3 visualization sections, with key performance indicators (KPIs).

**Multiclass Disease Classification of Medical Image Data Using CNN** - [link](https://github.com/itsmesethus/MULTICLASS-DISEASE-CLASSIFICATION-OF-MEDICAL-IMAGE-DATA-USING-CONVOLUTIONAL-NEURAL-NETWORK-) **Jan 2023 ‑ Apr 2023**

* Assembled CNNs using TensorFlow and Keras, achieving 94.88% accuracy in Pneumonia X-ray classification.
* Exhibited expertise in multi-class classification with accuracies of 95.67%, 95.18%, and 90.87% for various diseases.
* Adopted VGG19 architecture with strategic layers for superior eight class classification to Gastrointestinal diseases performance.

# Certifications

* IBM Data Science Professional Certificate – (IBM-Coursera, Jul 2023).
* Machine Learning Specialization (Supervised Machine Learning- Regression & Classification, Advanced Learning Algorithms, Unsupervised Machine Learning & Recommender Systems) – (DeepLearning.AI - Coursera, Oct 2023).
* Ensemble Methods in Python – (DataCamp, Sep 2023)
* Sequences, Time Series and Prediction – (DeepLearning.AI- Coursera, Nov 2023).
* Feature Engineering for Machine Learning in Python – (DataCamp, Sep 2023)
* Data Analytics with Python – (NPTEL, Swayam, Apr 2022).
* Databases And SQL For Data Science with Python - (IBM - Coursera, Nov 2022)