

JOSE BABY

📍 New Delhi, India

✉ josebgeorge99@gmail.com

☎ +91-9717185662

🌐 [in/jose-baby](https://www.linkedin.com/in/jose-baby)

🔗 [jose-102](https://github.com/jose-102)

EDUCATION

M.Sc. in Statistics, Loyola College (Affiliated to the University of Madras)

2020 - 2022

CGPA: 9.28 / 10.0

Chennai, India

Relevant Coursework: Applied Regression Analysis, Categorical Data Analysis, Multivariate Analysis, Data Warehousing and Data Mining, Testing of statistical Hypothesis, Advanced Distribution theory and Programming using Python & R.

B.Sc. in Statistics, Loyola College (Affiliated to the University of Madras)

2017 - 2020

CGPA: 8.54 / 10.0

Chennai, India

PROFESSIONAL EXPERIENCE

Cognizant India Pvt Ltd - Programmer Analyst (Domain AI/ML)

July 2022 – Present

- **Social Determinants of Health Prediction**
 - **Objective:** Trained multiple models for predicting Social Determinants of Health in member population using claims information and member demographics.
 - **Responsibilities:**
 - Extracted the data from multiple sources (like TeraData, Redshift and MS SQL Server) utilizing the SQLAlchemy and performing the feature engineering using Polars and Pandas.
 - Models pipeline was created using Scikit learn and various ensembled and stacked models were trained.
 - Created dashboards utilizing various metrics to check for the model biasness using Power BI.
 - An accuracy of 64.48% was achieved.
- **Mental Health and SDOH keyword Identification**
 - **Objective:** To identify the various conditions related to Mental Health and SDOH issues the patient is suffering using an elaborate EMR record.
 - **Responsibility:** From a dictionary of words pertaining to Mental health and SDOH, the EMR record had to be compared to extract out various conditions the patient is suffering from. This way the health provider could quickly identify the patient conditions. Have applied the same use case using Open AI's Large Language Model GPT-3.5 Turbo also.
 - **Technology:** Python, Lemmatization, Stop Words removal, Streamlit, NLP Libraries like NLTK and Spacy.
- **Creating a Feature Store**
 - **Objective:** To create Proof Of Concept which utilizes AWS Feature Store service.
 - **Responsibilities:** Create a feature store which would enable features to be registered, discovered, and used both as part of model training pipelines as well as inference pipeline.
 - **Technology:** AWS Sagemaker, Boto3 SDK, Polars, Teradata, Pandas and Scikit Learn and Python.

Medialis Pvt Ltd, UK - Internship

June - July 2021

- **Position:** Biostatistician
- **Projects:** Drafting Statistical inputs to study protocol and Statistical Analysis Plan (SAP) for clinical and epidemiological study designs.

The Sparks Foundation – Internship

May - June 2021

- Mini Project using Machine learning Algorithms.

Bumaas Engineering Solutions and Technology - Internship

Dec - January 2020

- Demand Forecasting analysis of Fasteners for the month of January.

SKILLS

- **Programming and scripting Languages:** R, Python, Tableau, SQL, SPSS, Basics of SAS, MATLAB and HTML
- **Data Science specific Libraries:** Tenson Flow, PySpark, Scikit Learn, Seaborn, Matplotlib, Numpy, Pandas, Polars, Langchain, HuggingFace Transformers, Flask and Streamlit.
- **Technology:** Deep Learning, Machine Learning, Natural Language Processing (NLP), DBMS, Vector Databases, Basics of AWS, Docker Containerization, Packaging ML Models, MLFlow, Terraform basics, Retrieval Augmented Generation (RAG), Power BI and basics of Cloud.
- **Other key skills:** Problem solving, logical and analytical thinking, strong presentation and communication skills.

MAJOR CERTIFICATIONS

- AWS Certified Cloud Practitioner
- Generative AI with Large Language Models - Coursera