



Contact Details

Mobile: ☎ + 971 58 633 7266
 Email: stat.sebijoseph@gmail.com
 Skype id: sebijoseph123
<https://www.linkedin.com/in/sebijoseph123>

Address for Communication

7Even Building, Block F, G15,
 Dubai Investment Park 1, Dubai,
 UAE, P.O Box 125804

Programming Languages

- Python
- R
- SQL

Data Science Techniques

- CatBoost
- Classifications
- Cluster Analysis
- Deep Learning (ANN, CNN & RNN)
- Dimension Reduction
- Feature Engineering
- NLP and Text Analytics
- Regressions
- Sampling Techniques
- XGBoost

Data Visualization Tools

- MS Excel
- Power BI
- Tableau

Big Data Tools

- Hadoop
- Spark

Cloud Platform

- AWS
- GCP

Professional Summary

An extensive career working in data analytics for more than 16 years, capable of developing advanced insight into any business aspect. Excellent statistics, mathematics, and calculation skills with daily practice. Proficient in Python, R, SQL, ML, and DL. Also, I excel in the areas of data visualization, business intelligence, analytical CRM, market research, report generation, conceptual thinking, risk planning, and 360° customer view analytics.

Employment History

Freelance Data Scientist, Dubai, UAE/Cochin, India.

📅 Sep2021-Present

Use data science methodologies and technologies to analyze data and solve business problems for international clients. Write code in Python, R, or similar as needed. Predicted the probability that a house could be sold using machine learning for a real estate client.

Data Scientist at PBW Software Pvt. Ltd. in Cochin, India.

📅 Apr2021-Sep2021

Identified trends and opportunities for growth through the analysis of complex data sets. Created best-practice reports based on data mining, analysis, and visualization. Used ML and NLP techniques to solve business problems.

Freelance Sr. Statistical Analyst/Sr. Data Analyst, Cochin, India.

📅 Oct2017-Mar2021

Identified valuable problems to solve, gathered the data from various sources, and preprocessed and cleaned the data for analysis. Created interactive dashboards using Tableau, Power BI, and Excel. Performed statistical analyses and communicated the findings effectively. To solve business problems, EDA and ML techniques were used.

Sr. Data Analyst at Jacobsons Direct Marketing Services in Dubai, UAE.

📅 Sep2012-Sep2017

Worked closely with various teams across the company to identify and solve business challenges utilizing structured and unstructured data. Analyze customer data based on rules and methods set by the business. Prepared reports to improve customer relationships and interaction. Monitored and reviewed KPIs such as customer satisfaction and retention. Developed dashboards using BI tools.

Sr. Business Analyst at Tata Consultancy Services in Bangalore, India.

📅 Apr2010-Aug2012

Participated in several statistical projects that involved applying the theory of statistics, such as population updates, sampling, creating weights, and feature engineering in the retail measurement services for the UK retail (CPG or FMCG) market.

Data Analyst at Cross-Tab Marketing Services (Course5i) in Bangalore.

📅 Feb2009-Apr2010

Worked with a specific international project called NPS among the online users and performed statistical analysis such as EDA, segmentation, and linear regression.

Research Asst. / Data Management Consultant at Catalyst in Bangalore.

📅 Jan2006-Feb2009

Designed survey research studies, including questionnaire development, sample selection, checking survey results, and conducting SPSS programming and statistical analyses.

Lecturer in Statistics at Ilahia College of Arts and Science, Muvattupuzha, Kerala, India. In 2005

Academic Qualifications

M.Sc. Statistics with 65% marks from Mahatma Gandhi University, India. 📅 Jun2002-Jun2004

B.Sc. Statistics with 78% marks from Mahatma Gandhi University, India. 📅 Jun1999-Mar2002

Certifications

Basic Image Classification with TensorFlow.	Issued by Coursera in Apr2022
Deploying machine learning models with flask for beginners	Issued by Coursera in Mar 2021
Machine Learning, Data Science and Deep Learning with Python.	Issued by Udemy in Feb2021
AWS Certified Machine Learning Specialty 2020 - Hands On!	Issued by Udemy in Aug2020
Taming Big Data with Apache Spark and Python - Hands On!	Issued by Udemy in Jun2020
The Ultimate Hands-On Hadoop – Tame your Big Data!	Issued by Udemy in Jun2020
Data Scientist with Python	Issued by DataCamp in Jul 2019
Data Scientist with R	Issued by DataCamp in Apr 2019

PROJECT EXPERIENCE

1. To predict the probability that a house can be sold

📅 Apr2022-Sep2022

The main goal of this machine-learning project was to predict the probability that a house can be sold. Predicting house sales involves using statistical models to analyze historical data on housing prices and sales as well as factors that may influence the demand for houses, such as economic, demographic, and local market conditions. The goal was to use this information to make predictions about future housing prices and sales. Exposure to web frameworks such as Django and Flask. URL: <http://sebiml50.pythonanywhere.com/>

2. Diabetes Prediction Using Machine Learning

📅 Oct2021-Mar2022

The objective of the project was to predict whether the patient had diabetes or not. Performed data cleaning, data analysis, data visualization, and model building. Used ML models such as decision trees, random forests, SVM, and the XGBoost classifier. Saved the best model by using pickle to make the prediction from real data. Exposure to python libraries such as NumPy, Pandas and Scikit-learn.

3. Data Science for the COVID-19 Project

📅 Jul2020-Aug2020

The objective of the project was to segment countries and states based on COVID-19 cases and deaths. Also, what are the non-medical determinants of the COVID-19, namely the case and death rates?. Exposure to data science techniques such as web scraping, data wrangling, feature engineering, k-means clustering, correlation, and regression. Developed interactive dashboards using BI tools.

4. Predictive Modeling for Consumer Purchasing Behavior

📅 Oct2016-Mar2017

The objective of the project was to predict whether a customer would purchase a vehicle. Input data contained a variety of important factors that influence a customer's purchasing decision. This project helped to run an effective marketing campaign at a lower cost and in less time. Exposure to ML techniques such as logistic regression, decision trees, and random forests.

5. Predicting the Price of Used Cars using Machine Learning Techniques

📅 Mar2016-May2016

The objective of this project was to predict the price of used vehicles using showroom data. We used the data set with vehicle's price histories and related attributes. We used the features, such as mileage, age, fuel type, HP, engine CC, and weight, to predict the price.

6. Total Cost of Ownership (TCO) for Utility Commercial Vehicles

📅 Jan2015-Jun2015

This research was carried out for FCA in the Middle East market. The purpose of the study was to provide concrete values on the competitor's price offering, which would empower the decision-making process for upcoming models. TCO is an estimate of the total cost to own and operate a vehicle for five years. It includes all the expenses spent on fuel, insurance, maintenance, repairs, service, and interest on loan payments, as well as the losses incurred due to the depreciation of the vehicle at the end of the same period. A high-level analysis of the TCO comparison had been done to find out which company was the best in class.

7. Monthly Customer Promoter Survey and Feedback Analysis for Fiat Chrysler Automobiles (FCA)

📅 Jan2013-Aug2017

The survey covered all the FCA customers in the Gulf region. Customers' vehicle sales and service details are among the information gathered. The customer feedback on brand, vehicle, and dealer helped us to derive some important KPIs. The customer promoter score (CPS) was mathematically derived from the survey data. Build a word cloud based on customer feedback. BI monthly reports and interactive dashboards were created using MS Excel and Tableau.

8. Retail Measurement Science for Nielsen UK CPG

📅 Apr2010-Aug2012

This project worked out well for the Nielsen Company in the UK market. EPOS technology had been used for capturing sales data from census retailers. We used the industry's leading sample-based methodology to capture data from sample-based retailers. The following activities were performed during the period: (i). Define, measure, and update the universe (ii). design, select, and maintain a sample (iii). Feature-engineering loading and storing data (iv). Create weights (v). Validate the data inputs (vi). Query resolution, as well as (vii). Monitor and review KPIs.

9. Net Promoter (End-user satisfaction) Study for Microsoft

📅 Feb2009-Apr2010

It was an online survey among internet users conducted in 11 competitive and 21 noncompetitive markets across the globe. Data collected includes internet users' feedback on client (and competitor) services such as portals, homepages, search, email, IM, toolbars, and social networks. The net promoter score (NPS) was mathematically derived from the recommendation question. The survey data is used to derive key drivers for recommendations. Exposure to data science techniques such as multiple linear regression and segmentation.

Personal Details

Sex : Male
DOB : 10-May-1982
Marital Status : Married
Nationality : Indian
Visa Status : Visit Visa

Place: Dubai, UAE
Date:23-Feb-2023

References Furnished Immediately upon Request