# ANUSHA N

anushaswamy.n@gmail.com | +91 9980854247 | Bengaluru, IN | GitHub | LinkedIn

## DATA SCIENTIST

Motivated data scientist with 3+ years of experience as a freelance data scientist. Passionate about building models that fix problems. Relevant skills include machine learning, problem solving, programming, and creative thinking. Armed with a passion to solve real-world business challenges using data analytics. Coming with proficiency in Microsoft Office applications.

## TECHNICAL SKILLS

**Languages**: Python 3.7, MySQL, Shell Script, PostgreSQL, Snowflake
**Packages**: NumPy, Pandas, Matplotlib, Seaborn, Plotly, Scikit-Learn, Stats Models
**Statistics/ML**: Linear & Logistic Regression, K-Means, Hierarchical Clustering, PCA, Decision Tree, Random Forest, XGBoost.
**Tools**: Tableau, Excel, PowerPoint, SharePoint

## KEY SKILLS

- Machine Learning
- Data Pre-processing
- Data Visualization & Sanitization
- Exploratory Data Analysis
- Data Mining & Data Wrangling
- Database Management
- Statistics
- Natural Language Processing
- Team Player & Leadership

## EDUCATION

- Masters on Data Science |Liverpool John Moores University | Sept ’20 – Jul’21
- PG Diploma in Data Science | IIIT Bangalore & UpGrad | Sept ’19- Sept ‘20
- BE- Information Science | Bangalore University | 80.07% - First Class with Distinction | Jun ’14 - Jun ‘18

## KEY DATA SCIENCE & MACHINE LEARNING PROJECTS

**Masters: Customer Segmentation – Online Retail Data | Dec’20 – May’21 | Tech Stack: Python, Jupyter Notebook**
• Objective: Identifying the customers into their respective segments for larger reach over customers
• Solution: Data Preparation, Calculate RFM value, analysed a variety of models and selected K Means which best worked on the data. After which items were suggested based on FP Growth Data Mining technique.

**Payment Card Industry – Credit Industry | Apr '20 | Tech Stack: Python, Jupyter Notebook**
• Objective: A Credit card company wants to detect Fraud transactions using the data that is highly sparse on fraud cases.
• Solution: Handled Class imbalance, analysed a variety of models and selected XGBoost which best worked on the data.

**Domain: Telecom | Nov '19 | Tech Stack: Python and Jupyter Notebook**
• Objective: A Telecom company wanted to build a predictive model to identify customers at high risk of churn and identify the main indicators of churn.
• Solution: Designed a logistic classification model to predict the high risk of churn customers and identified main indicators

**Domain: Taxi Industry | Jan '20 | Tech Stack: Hive, HDFS**
• Objective: A taxi Corporation wants to explore their data and get meaningful insights out of it to get a better understanding of the taxi system in the city.
• Solution: Analysed the dataset using Hive queries and got useful insights about the trip behaviour

**Domain: Online Education | Oct '19 | Tech Stack: Python, Jupyter Notebook**

- Objective: An education company wanted to increase its lead conversion rate by correctly identifying potential leads
- Solution: Designed a Logistic regression model which could predict the potential leads and increase the conversion rate.

**Domain: Automotive | Aug '18 | Tech Stack: Python, Jupyter Notebook**

- Objective: An Auto company interested to enter into manufacturing in the US market wanted to know factors in the American Market on which the price of an automobile would depend.
- Solution: Came up with a linear regression model which specified the indicators along with their importance in deciding the price of the automobile.

## PROFESSIONAL EXPERIENCE

**Company Name – Tata iQ**

**Client – Tata Power Text Classification**

- The objective of this project is to classify the similar items together.
- Data Pre-processing, Explore the data, Data Cleaning, Understand the similarity using Cosine Similarities.
- Train the model using Spacy, Predict the item name, dimension and category using trained model.

**Client – Tata Projects Forecasting**

- The objective of this project is to build the Metal Price Forecasting Model (Steel/Copper/Aluminium).
- Data Preparation, Data Cleaning, Missing value treatment, EDA for the data
- Identify important variables using Grid Search & Granger Causality technique
- Perform VAR Modelling to forecast the indices price for long term & short term period.

**Client – Tata Power Forecasting**

- The objective of this project is to build the Coal Price Forecasting Model for various type of coal (API2/API4/NewC).
- Data Preparation, Data Cleaning, Missing value treatment, EDA for the data
- Perform VAR Modelling to forecast the indices price to next months.

**Client - TSSS Customer Base Segmentation**

- The objective of this project is to segment the customer based on the various data points available for segregation of customer. Ex: nature of the campaign response, leads and program response (MGM, Spin and Smile).
- Data Preparation, Data Cleaning, EDA for the various forms of data.
- Clustering should be done in order to divide of a set of consumers into persons with similar needs and wants.

**Client – Croma NPS Calculation**

- Extract the data from various sources (ECOM, BCOM & Carry-In)
- Data Preparation, Data Cleaning, Data Integration, Data Enrichment by using Python.
- Based on NPS Promoter tag their experience into Detractor, Passive and Promoter.

**Client – TSSS FNF Advocacy**

- TSSS FNF Advocacy project aims in rule based segmentation of customers based on their behaviour throughout.
- Initially their nature of referring, campaign responses, page visits against each individual brands are created
- Data Preparation, Data Integration, EDA and Data Profiling from various form of data