

# HIMANSHU PRAJAPATI

Tampa, Florida-33612 ▪ himanshuprajapati@usf.edu ▪ (656) 200-7826 ▪ [LinkedIn](#) ▪ [GitHub](#)

## SUMMARY

- Experienced data engineer with a strong background in Python, AWS, and SQL, dedicated to optimizing data infrastructure for operational efficiency.
- AWS Certified Associate Solution Architect skilled in designing and implementing scalable cloud solutions.
- Committed to leveraging technology to streamline processes and drive business growth in the role of a data engineer.

## EDUCATION

University of South Florida

**Master of Science (MS) in Business Analytics & Information Systems**

Tampa, Florida

AUG 2023 - MAY 2025

Gitam University

Vizag, India

B.Tech. in Electronics & Communication Engineer

JUNE 2017 - AUG 2021

## SKILLS

- Programming language: **Python, SQL, R, C#, XML, XSLT, C, MATLAB**
- Visualization & Analytical tools: **Tableau, Power BI, Qlik, Matplotlib, SSRS**
- Data Analysis and Automation Skills: **VBA**, Microsoft Excel for automation and custom functions
- Cloud technologies & tools: **AWS Lambda, S3, Redshift, Auto Scaling, EC2**
- ETL Tools: **Apache Airflow, Bonobo**, etc.
- IDE: Visual Studio, PyCharm, Idea IntelliJ
- Database: Oracle DB, My SQL, PostgreSQL, pgAdmin, **SSMS**
- Others: **SAS, SciKit Learn, SDLC, PowerPoint, Word, Git, HTML**

## PROJECTS

AUG 2023 – NOV 2023

### R Project - Time Series Forecasting (Business Analytics)

- Conducted linear and multiple regression analysis in **R** to identify key factors influencing sales trends and demand fluctuations.
- Generated forecasts and visualizations using R packages like forecast and **ggplot2**, providing actionable insights for inventory management and resource planning, resulting in optimized operations and cost savings.

### Database - Cricket (Advanced Database Management)

- Developed a comprehensive cricket database with an Entity Relationship Diagram (**ERD**), optimizing its structure.
- Improved server performance by optimizing **SQL** queries and utilized **Microsoft Excel** for detailed analysis of match statistics and player performance.

### Churn Customers (Data Mining)

- Facilitated **requirements-gathering** sessions to understand stakeholder needs and define project objectives effectively.
- Led data preprocessing in **Python**, ensuring quality through importation, **cleaning**, null value checks, and conducted (exploratory data analysis) **EDA**.
- Identified **XG Boost** as the best churn predictor and applied hyperparameter tuning with GridSearchCV, used **Power BI** to analyze churn trends, and concluded that less than \$40K in education-level graduates have more utilization ratio.
- Utilized **Qlik** for interactive data visualization, creating dynamic dashboards to communicate insights and trends effectively.

### ETL – Apache Airflow

- Built **ETL pipelines** in Apache Airflow using operators, connecting **Oracle** databases securely. Configured DAGs for streamlined data retrieval and logging in Apache Airflow, enhancing monitoring and workflow efficiency.

## WORK EXPERIENCE

Bangalore, India

### THOMSON REUTERS

JUNE 2021 - JULY 2023

#### Associate Software Engineer

- Headed team to enhance **application** functionalities, resulting in a 5% system performance boost. Simplified and optimized 15 complex queries, generating 50 SQL queries from an ERD for database development.
- Built serverless application **APIs** using **AWS Lambda Service** and Step Functions in Python.
- Produced **Python** coverage reports for unit, integration, and XSLT tests on GitHub, reducing bug reports by 20%.
- Utilized Selenium, BeautifulSoup, and Chrome driver for Python-based **web scraping**, extracting over 10,000 data records.
- Collaboratively contributed to **IT business** strategy knowledge acquisition, reducing support response times by 25% and costs by 10%.

- Led the implementation of Software Development Life Cycle (**SDLC**) methodologies, ensuring adherence to best practices throughout the software development process.
  - Introduced **SAS** for data analysis and reporting tasks, leveraging SAS procedures to analyze large datasets and extract actionable insights. Implemented **PROC SQL** for data manipulation and performed **statistical analysis** to support decision-making processes.
- 

#### **INTERNSHIP (Machine Learning)**

JUNE 2020 - AUG 2020

- Collaborated in an 8-week internship, training/testing 15+ **machine learning models**, and conducting exploratory data analysis with a diverse team. Cleansed and analyzed 10,000+ records, enhancing data quality by 15%.
- Exhibited leadership in **algorithm** proficiency, comparing the accuracy of machine learning models, and selecting optimal algorithms for tasks, demonstrating expertise in data analysis and model selection.
- Utilized **SciKit Learn** extensively for implementing various machine learning algorithms, including classification, regression, and clustering tasks, enhancing predictive modeling capabilities and facilitating data-driven decision-making processes.