

Rahul Prajapati

[+91 7039542072](tel:+917039542072) | rahulprajapati2072@gmail.com | [LinkedIn](#) | [Github](#)

Objective

M.Sc. Statistics Graduate with strong skills in SQL, Advanced Excel, Python, and statistical analysis, including regression, hypothesis testing, and time series analysis. Experienced in cleaning, validating, and analyzing large datasets to generate reports and clear, actionable insights. Interested in analytics and data quality, with a strong focus on accuracy and data-driven decision-making.

Skills

Programming & Tools: python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn), SQL (SSMS, MYSQL), R, Advanced Excel (Pivot Tables, Power Query, VLOOKUP/XLOOKUP, INDEX-MATCH, Charts), Power BI

Statistical Technique: Hypothesis Testing, Regression Analysis, Time Series Analysis, Logistic Regression, Chi-square Test, t-test

Machine Learning: Classification, Clustering, Decision Trees, KNN, DBSCAN, Random Forest, XGBoost, Ensemble Learning

Other: Data cleaning, Feature Engineering, Analytical skill, Model Evaluation, EDA

Work Experience

Data Science Intern - Encryptix

Mar 2025 - Apr 2025

- Cleaned and preprocessed datasets to ensure data accuracy and consistency.
- Performed exploratory data analysis to identify patterns, trends, and key insights.
- Built and evaluated predictive models, delivering data-driven insights for decision-making

Projects

Retail Sales Pattern Analysis using SQL [[Link](#)]

- Designed and managed a Retail Sales database in SQL Server by creating relational tables and maintaining structured sales data for analysis. .
- Wrote SQL queries using GROUP BY, CASE, RANK(), and DATEPART to analyze sales by category, customer demographics, and time of day
- Identified the top 5 highest-spending customers and found that Clothing led all categories in total sales and order volume
- Found that November was the best-performing month and afternoon hours (12 PM–5 PM) had the highest order volume, highlighting peak sales periods

Consumer Preferences for Coca-Cola [[Link](#)]

- Conducted a research study using a Likert scale survey to analyze factors influencing consumer perception of Coca-Cola
- Applied Factor Analysis to identify key factors: Brand Promotion, Affordability, Packaging, Advertisement & Lifestyle, Product Availability, and Taste Preference.
- Developed a Multinomial Logistic Regression model with 60.18% accuracy to predict purchase likelihood.
- Performed statistical interpretation of model outputs to generate actionable business recommendations.
- Interpreted results to highlight risk drivers such as pricing sensitivity, availability, and branding

Financial Transaction Fraud Detection using Machine Learning [[Link](#)]

- Built an end-to-end fraud detection pipeline on a highly imbalanced financial transaction dataset, performing data cleaning, exploratory analysis, and feature engineering to identify suspicious transaction patterns.
- Engineered features like org_diff, dest_diff, and balance_ratio to capture abnormal balance transitions and improve fraud detection
- Handled class imbalance using SMOTE and trained an XGBoost classifier, selecting the final model based on fraud precision rather than accuracy alone
- Optimized the model by comparing feature combinations and imbalance-handling techniques, achieving 93.3% fraud precision while minimizing false positive alerts

Education

Masters of Science, Statistics - Ramniranjan Jhunjhunwala College

Aug 2024 – Jun 2026

Bachelor of Science, Statistics - Ramniranjan Jhunjhunwala College

Aug 2021 – Apr 2024

Certifications & Extracurricular Activities

- Data Analysis with Python (IBM)
- Data Analysis Essential (Cisco)

Student Coordinator - Statistika Fest 2024

- Led the planning and execution of Statistika Fest 2024, driving data-oriented initiatives that built enthusiasm for Statistics and collaborative learning