Pradeep G

Former Data Research Analyst [pradeepreddy0298@gmail.com](mailto:pradeepreddy0298@gmail.com)

LinkedIn: [linkedin.com/in/pradeepg98/](https://www.linkedin.com/in/pradeepg98/) Phone:+1 (703) 296-4110

# PROFILE

* Experienced Data Analyst with expertise in managing and architecting ETL processes, data warehouses, and data delivery systems, improving workflows and reducing processing times by 40%.
* Skilled in automating data extraction and reporting using SQL and Python, reducing retrieval times by 85% and enhancing data-driven decision-making across teams.
* Proven ability to develop interactive dashboards and visual reports using Power BI, Tableau, and Excel, delivering key insights for departments and executive leadership.
* Proficient in applying machine learning (supervised and unsupervised) algorithms to drive predictive analytics, improving marketing campaigns and aligning with market demands.
* Experienced in deploying machine learning models using AWS SageMaker and familiarity with cloud technologies such as Snowflake, S3, Redshift, enabling real-time analytics and optimization.
* Advanced knowledge in A/B testing, statistical analysis, optimizing marketing strategies and improving actionable insights through enhanced data processes.
* Strong collaboration in Agile environments with cross-functional teams, driving project delivery efficiency by 30% and saving $250K annually by streamlining data-driven operations.

# WORK EXPERIENCE

**Data Research Analyst – *George Mason University, VA*** Oct 2022 – May 2024

* Managed and architected ETL processes, data warehouses, and data delivery systems, refining workflows to improve data quality, resulting in a 40% reduction in processing times and enhancing over 100 internal reports.
* Automated data extraction and reporting tasks using SQL and Python, reducing retrieval time by 85% and allowing teams to focus on high-value analytical activities.
* Supported departments by retrieving and organizing necessary data for reporting, facilitating informed decision-making.
* Communicated results of data analysis to executive leadership, contributing to strategic planning and operational improvements.
* Developed interactive dashboards and visual reports in Power BI to present key insights on student engagement and employment outcomes to career counselors, academic departments, and university administrators.
* Utilized Microsoft Excel for detailed reporting and analysis, incorporating advanced pivot tables, data modeling, and forecasting to identify career trends and inform program improvements.
* Collaborated with career services teams to enhance data-driven decision-making, supporting program development and student career readiness through comprehensive data analysis.
* Applied predictive analytics using Python to forecast employment outcomes and job market trends, helping the university align its programs with future demands.

**Data Analyst – *Kantar, India*** May 2021 – Aug 2022

* Led cross-functional teams to optimize data-driven processes, boosting project delivery efficiency by 30%, saving $250k annually through effective resource allocation and workflow automation.
* Deployed machine learning models for predictive analytics (using Regression, Classification, CNNs and RNNs like LSTM) on large datasets (10 million+ data points), enhancing marketing campaign performance by 25% through Multi-Touch Attribution and Market Mix Modeling.
* Implemented A/B testing frameworks and hypothesis testing (P-Value, T-tests) across marketing units, improving actionable insights by 30%, which refined marketing strategies.
* Leveraged AWS SageMaker to train, test, and deploy machine learning models, enabling real-time analytics that optimized campaign performance.
* Visualized and communicated complex data insights through Tableau, and Excel to senior stakeholders, ensuring data-driven decisions across marketing efforts.
* Automated data pipeline processes using Python, reducing data extraction and transformation times by 40%, streamlining reporting across marketing operations.
* Applied advanced machine learning techniques such as Random Forests to drive predictive analytics and improve marketing and sales outcomes.
* Collaborated in an Agile environment with marketing and data science teams to prioritize and deliver data analysis tasks on time, ensuring alignment with business goals.

# SKILLS

**Programming Languages:** SQL, Python (NumPy, Pandas, Seaborn, Scikit Learn, TensorFlow, Keras, OpenCV)

**Big Data Platforms:** Hadoop, Spark, Apache Kafka

**Data Modelling:** Kimball Methodology, Star Schema, Snowflake Schema

**Data Visualization:** Tableau, Power BI, PowerPoint, Microsoft Excel (VLOOKUP, Pivot Tables, Statistical Analysis)

**Inferential Statistics & Hypothesis Testing:** Distributions (Sampling, Binomial, Poisson, Normal)**,** Critical Value Method, P-Value Method, A/B Testing, T-tests, Z-Test, Chi-square, ANOVA

**Machine Learning:** Regression, Classification, Clustering, Time series models (ARMA, ARIMA, SARIMAX),

Decision Trees, Random Forest Models, CNNs, Transformers

**Evaluation Metrics:** Accuracy, Precision, Recall, F1-score, ROC Curve, AUC

**Cloud:** Snowflake, Amazon S3, RDS, Amazon Redshift, AWS SageMaker

**Containerization & Orchestration:** Docker, Apache Airflow, Kubernetes

**CI/CD:** Jenkins

**Version Control:** Git

**Methodologies:** Agile, Waterfall

# PROJECTS

* **Analysis on Fentanyl Production:** Conducted a comprehensive analysis in collaboration with TraCCC at the university, using Python, SQL, MS Excel, and Tableau to uncover significant links between the PRC government and fentanyl manufacturers. This research provided essential data, contributing to the development of regulatory measures and informing policy decisions to combat illicit fentanyl production and distribution.
* **Chest X-Ray Image Classification for Effusion Detection:** Developed a deep learning model using CNN and ResNet to classify chest X-ray images as 'effusion' or 'no finding' addressing class imbalance with weighted cross-entropy, which improved the AUC from 0.57 to 0.72, enhancing diagnostic accuracy for lung condition detection.
* **Automated Vehicle Detection and Classification for Road Construction Planning:** Developed an OpenCV-based system for vehicle detection and classification using video frames, leveraging contour detection to track and count vehicles crossing a virtual line. Implemented a CNN classifier to categorize cropped vehicle images into distinct classes (e.g., cars, trucks), providing critical data to aid government road construction projects.
* **Optimization of Marketing Strategies through Multi-Touch Attribution:** Leveraged advanced ML techniques and implemented Multi-Touch Attribution models using Python and SQL, achieving a 20% boost in marketing ROI by accurately attributing sales to specific marketing channels.
* **Marketing Campaign Effectiveness with A/B Testing:** Developed and implemented an A/B testing project to evaluate two marketing campaigns using Python, analyzing metrics like cost per mille (CPM), click-through rate (CTR), impressions, and purchase rates. Cleaned and merged datasets for comprehensive analysis, visualizing results with Seaborn. The test campaign demonstrated improved user engagement and conversions, despite higher impression costs. Key insights included increased click-through rates and reduced cart abandonment, leading to a higher purchase rate in the test campaign.
* **Sales Forecasting for Big Mart:** Developed sales forecasting models utilizing smoothing techniques and ARIMA methods. After decomposing the time series data into trend, seasonal, and residual components, applied models including Simple Exponential Smoothing, Holt-Winters, ARIMA, and SARIMA. Compared model accuracy using Mean Absolute Percentage Error (MAPE), achieving the best forecast with SARIMA, which had the lowest MAPE of 7.92.
* **Customer Churn Prediction for Telecom Firm:** Developed a predictive model using Random Forest and Python to analyze customer demographics, services availed, and expenses to identify churn patterns. Achieved an accuracy score of 91.28% and a recall score of 78.55% in predicting customer churn, enabling the firm to implement targeted retention strategies and reduce churn rates by 15%.

# EDUCATION

**M.S in Data Analytics Engineering– *George Mason University – Fairfax, VA***

**Thesis Research Paper:** Identifying the Origins and Manufacturers of Synthetic Opioid Precursors

**P.G Diploma in Data Science (**Business Analytics Specialization**) – *IIIT – Bangalore, India***

**Thesis Research Paper:** Optimizing Operations and Marketing Budget using Market Mix Modelling

# CERTIFICATIONS

* **AWS Certified Machine Learning Engineer |** [**Credentials**](https://www.credly.com/badges/30b73520-50e6-4c48-92b8-94b8552220c9/public_url)
* **upGrad & IIIT-B |** [**Credentials**](https://www.credential.net/profile/pradeepreddy846815/wallet)
* **Reporting Analytics |** [**Credentials**](https://verify.skilljar.com/c/nxcw4ufncpvu)