# SHRIYANSH SINGH

+1 930 333 5141 — shriyansh.singh24@gmail.com — LinkedIn — GitHub

## EDUCATION

**Indiana University Bloomington** Aug 2023 – May 2025
Master of Science in Data Science Indiana
Relevant Courses: Information Visualization, Data Mining, Applied Machine Learning, Statistics, Big Data Applications, Cloud Computing, Graph Analytics, Applied Database Technologies, Intelligent Systems

## PROFESSIONAL EXPERIENCE

**Data Analyst Intern** April 2024 - Present
Hyphenova AI Los Angeles, California

- Analyzed large datasets using Python (Pandas, NumPy) and SQL to extract actionable insights on brand-creator matches, employing regression analysis and data visualization techniques in Tableau, resulting in a 30% increase in successful matches and a 15% boost in user satisfaction.
- Improved data quality and reliability by implementing real-time data validation scripts in Python (using Pandas for data manipulation and PyOD for outlier detection), which targeted missing values and outliers, increasing data accuracy by 30% and reducing analytics errors by 20%, directly enhancing reporting precision.
- Developed and optimized data pipelines using Apache Spark and AWS (Glue, S3), incorporating partitioning strategies and efficient data storage formats, reducing data processing times by 40%, which enabled real-time reporting and facilitated quicker decision-making for stakeholders in campaign management.
- Conducted feature engineering and data preprocessing, including normalization, handling missing data, and feature scaling, to refine datasets for analysis, leading to a 25% improvement in campaign performance and more accurate data-driven recommendations.
- Collaborated with cross-functional teams to design interactive dashboards in Tableau, visualizing key performance indicators (KPIs) such as engagement rates and conversion metrics, which enhanced data-driven strategy formulation for client campaigns.

**Data Analyst Intern** May 2022 - Oct 2022
Enterprise Business Technologies Pvt Ltd Mumbai, India

- Analyzed project data using Python (Pandas, NumPy) and Excel to implement the OKR framework, using data-driven insights to align project strategies with business objectives, resulting in a 25% increase in project completion rates and a 20% improvement in client satisfaction.
- Revamped Power BI dashboards by refining data models and automating validation processes using DAX and M language, which improved report reliability by 15% and reduced report generation time by 40%.
- Conducted comprehensive market analysis using Python (Pandas, NumPy, Statsmodels) for data wrangling, and applied linear regression and time series forecasting techniques, enhancing forecast accuracy by 18% and contributing to a 10% increase in quarterly revenue.
- Developed and presented data visualizations in Power BI to communicate market trends and performance metrics to stakeholders, enabling data-driven decisions that supported business growth initiatives, including optimizing marketing strategies based on forecasted trends.

## PROJECTS

**Retail Inventory Optimization**

- Analyzed historical sales and inventory data using Python and Pandas to identify demand trends and optimize inventory levels across 50 retail locations, leading to a 15% reduction in holding costs and a 10% decrease in stockouts.
- Developed and deployed Exponential Smoothing models for demand forecasting, achieving a 90% accuracy rate, which informed inventory replenishment decisions.
- Conducted ABC classification of inventory, prioritizing high-value items and improving turnover rates by 12% through targeted stock management.
- Implemented Monte Carlo simulations to evaluate inventory scenarios, optimizing reorder points and reducing excess stock by 8%.

**Sentiment Analysis on Social Media Data**

- Extracted and preprocessed over 100,000 social media posts using Python, Beautiful Soup, and NLP techniques to prepare data for sentiment analysis, addressing challenges in noise reduction and text normalization.
- Built a sentiment classification model with NLTK and Scikit-learn, fine-tuning hyperparameters to achieve an F1-score of 0.85, effectively differentiating positive, negative, and neutral sentiments.
- Created dynamic Tableau dashboards to visualize sentiment trends and correlate spikes in negative feedback with product delivery issues, providing actionable insights.
- Delivered insights to the marketing team that informed targeted campaigns, resulting in a 25% increase in positive sentiment and improved customer perception within three months.

**Sales Data Analysis and Forecasting**

- Collected, cleaned, and integrated sales data from multiple sources, including CRM and ERP systems, using Python and Pandas, standardizing over 10,000 records for analysis.
- Built time series forecasting models using ARIMA and Prophet, incorporating external variables like seasonality, holidays, and promotions, leading to a 20% improvement in forecast accuracy.
- Performed feature engineering to identify key drivers of sales trends, enhancing predictive power and enabling more accurate inventory planning.
- Created interactive Tableau dashboards to visualize forecasts and trend analyses, supporting stakeholders in making data-driven decisions for inventory management and marketing strategies.

**Customer Segmentation Using Clustering**

- Extracted and cleaned customer data from various sources using SQL and Python, handling missing values and normalizing over 100,000 records to ensure data quality for clustering analysis.
- Employed K-means clustering with PCA for dimensionality reduction, successfully segmenting customers into 5 distinct groups based on purchasing behavior and demographic attributes, enhancing model interpretability.
- Assessed clustering quality through silhouette scores and iteratively refined parameters to optimize results, achieving clear and actionable customer segments.
- Developed comprehensive segmentation dashboards in Tableau, enabling targeted marketing efforts that increased campaign effectiveness by 30% through personalized customer interactions.

**Employee Performance Analysis**

- Collected and analyzed employee performance data across multiple departments using Python and SQL, processing over 5,000 records to identify key drivers of high performance, such as training frequency and peer collaboration.
- Developed multiple regression models to quantify the impact of individual and team-level factors on performance, validating models with cross-validation techniques to ensure robustness.
- Utilized feature selection methods to refine the models, enhancing interpretability and predictive accuracy by focusing on the most impactful variables.
- Designed interactive Power BI dashboards to visualize performance insights, enabling HR teams to craft data-driven retention strategies that boosted employee retention by 15%.

**Finance: Credit Risk Analysis**

- Extracted and cleaned loan applicant data using Python and SQL, including over 50,000 records with credit histories, financial statuses, and demographic details, standardizing data for predictive modeling.
- Built a logistic regression model to predict credit risk, incorporating domain-specific features like credit utilization and payment history, which significantly improved model accuracy.
- Validated model performance using ROC and AUC metrics, achieving a 25% improvement in prediction accuracy compared to baseline models, directly supporting risk management strategies.
- Developed interactive Tableau dashboards to present risk scores and model insights, enabling financial analysts to make data-driven decisions that reduced loan default rates by 10%.

**Healthcare: Patient Readmission Prediction**

- Processed and analyzed hospital and patient records using Python and SQL, focusing on 20,000+ patient cases to identify factors contributing to 30-day readmissions, including prior visits and medication adherence.
- Developed decision tree models, applying SMOTE to balance the dataset, which effectively reduced false negatives and lowered the readmission rate by 18%.

- Conducted a feature importance analysis to prioritize critical predictors, providing actionable insights for healthcare providers to focus on high-risk factors like chronic conditions.
- Implemented a real-time monitoring dashboard in Power BI, allowing healthcare teams to proactively intervene with at-risk patients, resulting in $2M annual cost savings through reduced readmission rates.

**E-commerce: Product Recommendation System**

- Aggregated and processed over 1 million transactions and user behavior records from a major e-commerce platform using Python and AWS data lakes, ensuring data integrity for recommendation modeling.
- Developed a hybrid recommendation system combining collaborative and content-based filtering techniques, employing matrix factorization with Scikit-learn to personalize user experiences.
- Enhanced model performance through extensive hyperparameter tuning and A/B testing, resulting in a 15% increase in average order value and improved recommendation relevance.
- Deployed the recommendation system on AWS SageMaker with autoscaling capabilities, efficiently managing peak demand loads and reducing cart abandonment rates by 10%, enhancing overall customer satisfaction.

**Predictive Maintenance for Manufacturing Equipment**

- Processed over 1 million IoT sensor data points from manufacturing equipment using Python and SQL, performing data cleaning and transformation to ensure high-quality inputs for predictive modeling.
- Developed a predictive maintenance model using logistic regression and random forests, fine-tuning algorithms to achieve 92% accuracy in forecasting equipment failures.
- Integrated anomaly detection techniques, such as Z-score analysis, to flag irregularities in equipment performance, reducing false positive rates by 10%.
- Designed and deployed a real-time monitoring dashboard in Power BI, enabling maintenance teams to track equipment health proactively, leading to a 20% reduction in unplanned downtime and a 15% cut in maintenance costs.

**Customer Lifetime Value (CLV) Prediction**

- Conducted RFM analysis on transaction data from 100,000+ customers using Python to segment customers by recency, frequency, and monetary value, providing a foundation for CLV prediction.
- Developed and validated a regression model to predict Customer Lifetime Value, utilizing behavioral and demographic features with recursive feature elimination, achieving an 85% accuracy rate.
- Applied K-means clustering to segment customers into distinct groups, focusing on high-value clusters for targeted marketing, leading to a 20% increase in ROI.
- Built interactive CLV dashboards in Tableau, enabling real-time tracking of customer value segments and supporting retention strategies that improved customer loyalty by 15%.

## SKILLS

- **Languages & Tools:** Python, R, SQL, NoSQL (MongoDB, Neo4j), Excel, Java, C/C++, Bash
- **Data Analysis & Visualization:** Data cleaning, data wrangling, exploratory data analysis (EDA), data visualization (Tableau, Power BI, Excel, Matplotlib, Seaborn)
- **Statistical Analysis & Modeling:** Descriptive statistics, inferential statistics, regression analysis (Linear, Logistic), hypothesis testing, time series analysis, A/B testing, clustering (K-means, DBSCAN)
- **Database Management:** Database querying and manipulation, ETL processes, data modeling (PostgreSQL, MySQL)
- **Data Wrangling & Automation:** Data manipulation, automation of data processes, scripting for data tasks (Pandas, NumPy)
- **Data Engineering & Big Data:** Data pipeline development, data extraction, transformation, and loading (ETL), big data processing (Apache Spark, Hadoop, AWS Redshift, Google BigQuery)
- **Machine Learning (Relevant to Data Analysis):** Predictive modeling, classification, clustering, model evaluation (Scikit-learn, TensorFlow, XGBoost)
- **Business Intelligence & Reporting:** Dashboard creation, KPI tracking, data storytelling, and report generation (Tableau, Power BI, Google Data Studio)
- **Version Control & Collaboration:** Git, GitHub, GitLab for version control and team collaboration
- **Cloud & Deployment Platforms:** AWS (S3, Redshift, Glue), GCP, Azure for data storage, processing, and model deployment
- **Soft Skills:** Analytical thinking, problem-solving, communication skills, stakeholder engagement, presentation of insights, teamwork