Sreeja Kurapaty

+1 (480) 310-8212 | [sreejakurapatydt@gmail.com](mailto:sreejakurapatydt@gmail.com)

# PROFESSIONAL SUMMARY

Senior Data Analyst with 6+ years of expertise in transforming complex datasets into actionable business intelligence. Demonstrated record of leading cross-functional analytics initiatives, optimizing data pipelines, and delivering insights that drive strategic decision-making. Skilled in advanced statistical analysis, predictive modeling, and data visualization, with a proven ability to translate technical findings into clear recommendations for stakeholders. Known for combining deep technical expertise with strong business acumen to identify growth opportunities and solve complex business challenges through data-driven solutions.

# EDUCATION

Master of Science in **Applied Data Science | Purdue University | USA**

Bachelor of Engineering in **Computer Science and Engineering | Jawaharlal Nehru Technological University | India**

**SKILLS**

# Programming Languages: Python, Java, R Data Analysis Tools: Spreadsheets, Google Analytics, JIRA, GitHub, Altyrex Databases: MySQL, Snowflake, MongoDB, Aqua Data Studio Visualization Tools: Tableau, Power BI, Confluence, SharePoint Machine Learning & Statistical Techniques: Exploratory Data Analysis (EDA), A/B Testing, Data Wrangling, Statistical Modeling, Feature Engineering, Deep Learning, ETL, Business Process Improvement Operating systems: Linux, Unix, Windows Project Management & Methodologies: Scrum, Agile, ServiceNow Middleware technologies: Apache Kafka, TIBCO EMS, TIBCO Rendezvous, IBM MQ, RabbitMQ

# PROFESSIONAL EXPERIENCE

**Goldman Sachs**

**Sr. Data Analyst Mar 2022- present**

* Developed interactive dashboards using Tableau to visualize KPIs for trading and risk management, improving insights and enabling data-driven financial decisions, resulting in a 16.3% increase in decision-making speed across teams.
* Proficient in messaging middleware technologies to facilitate real-time data integration, high-performance messaging, and efficient communication across distributed systems, supporting 100+ applications in a high-frequency trading environment
* Optimized financial data flows by implementing robust messaging architectures, ensuring low latency and fault-tolerant communication for high-frequency trading, risk analysis, and real-time market monitoring.
* Integrated messaging systems (Kafka, RabbitMQ, MQ) to enable seamless communication between disparate financial applications, enhancing data synchronization and trade execution efficiency, reducing operational risk by 25% in volatile market conditions.
* Collaborated with cross-functional teams to deploy messaging solutions that support millions of financial data points per day, ensuring timely reporting, transaction auditing, and compliance with industry regulations, reducing report generation time by 30%.
* Streamlined financial data exchange across systems by configuring and managing messaging queues, enabling real-time updates to risk management dashboards and financial decision-making tools, resulting in faster response times and improved risk mitigation.

**Aetna**

**Data Analyst June 2020 - Mar 2022**

* Led a team of 4 in the development of end-to-end business intelligence solutions, leveraging SQL, SSIS, and Tableau to deliver actionable insights to multiple lines of business and improve decision-making processes.
* Designed and optimized stored procedures and SQL queries to streamline data extraction and reporting, significantly reducing query execution time by 20% for daily operations and analytical dashboards.
* Conducted data analysis and data cleaning to ensure high-quality, accurate datasets for reporting and business intelligence applications, using Python for automation of repetitive data processing tasks.
* Utilized A/B testing to evaluate the effectiveness of different business strategies, working in an Agile environment with iterative sprints and stories tracked in Rally software.
* Developed automated reporting solutions by integrating spreadsheets and Tableau dashboards, allowing business stakeholders to quickly access key performance indicators and metrics in real time.

**Travelers**

## Programmer Analyst Feb 2018 - June 2020

* Optimized SQL queries for auto insurance data analysis, improving query performance by 30% by implementing advanced techniques like CTEs, temporary tables, and complex joins within MySQL, resulting in faster report generation for claims and underwriting teams.
* Developed automated reporting solutions in Power BI for Travelers Insurance, integrating data from multiple sources, including claims and policy databases, enabling real-time performance tracking and insights into premium collections, claims frequency, and customer retention rates.
* Built and maintained robust ETL pipelines to process large volumes of auto insurance data from various sources, including claims and policy systems, reducing data processing time by 30% and ensuring seamless integration with Travelers' Redshift data warehouse.
* Utilized HP ALM to manage test cases and defect tracking for auto insurance applications, collaborating with QA and development teams to ensure high-quality deliverables and alignment with business requirements in the agile development cycle.
* Streamlined financial reporting for auto insurance premiums and claims by leveraging spreadsheets and Power BI to automate month-end reporting processes, cutting report generation time by 25% and improving data accuracy for senior management and regulatory compliance.

# ACADEMIC PROJECTS

## Time Series Analysis and Forecasting of Pharmacy Sales | Tools: SAS, Excel

* Forecasted sales of 8 categories of drugs in SAS EG for a pharmacy store based on historical sales data of the past 10 years.
* Performed trend and seasonality analysis on the dataset. ARIMA and ARIMAX models were used for forecasting and the

**ARIMA model** had the lowest AIC value of **666.6**.

**Breast cancer classification** **| Tools: Python, Tableau,** **Excel**

* Evaluated and compared the performance of Random Forest, Decision Tree, and Logistic Regression models using accuracy, precision, recall, and F1-score across training, testing, and validation datasets, with the Random Forest model achieving an **88.6%** accuracy, outperforming the Decision Tree and Logistic Regression models by **8%** in overall performance.
* Conducted statistical testing to verify that the Random Forest model demonstrated strong generalization, achieving 88.6% accuracy on untested data, which was significantly higher than the Decision Tree and Logistic Regression models, and showed a **20%** improvement in F1-score.