

# JASMINE PINTO | Data Engineer

+1(315)-886-0519 | [LinkedIn](#) | [GitHub](#) | [jpinto01@syr.edu](mailto:jpinto01@syr.edu)

## EDUCATION

Syracuse University, Syracuse, NY | Master of Science in Computer Science (GPA: 3.67)

May 2024

## PROFESSIONAL EXPERIENCE

### Tungsten Automation Formerly Kofax, CA, USA | Data Engineer

Jun 2023 - Present

- Developed and deployed Apache Airflow and Python data pipelines, boosting data handling efficiency by 85%, and integrated them with CI/CD to expedite business capabilities through improved data reliability.
- Designed and maintained advanced data solutions in Snowflake, leveraging Python scripting to optimize data storage and retrieval processes. This initiative directly contributed to a 70% improvement in deployment efficiency and supported the implementation of dynamic data visualization tools, enhancing decision-making and business intelligence capabilities.
- Directed the development of scalable data integration pipelines using SQL and Apache Airflow within Agile frameworks, enhancing system reliability by 75%. Provided expert support and troubleshooting, ensuring system integrity and seamless alignment of technical strategies with overarching business objectives.
- Developed sophisticated statistical models and interactive dashboards using SQL and Tableau, integrated with backend systems powered by Snowflake. This work increased forecasting accuracy by 15% and improved stakeholder data accessibility, facilitating the conversion of complex datasets into clear, actionable insights for strategic decision-making.
- Spearheaded major data infrastructure initiatives, focusing on the integration of IT teams and stakeholders using Python, SQL, and Snowflake. This leadership role involved overseeing the comprehensive support and ongoing enhancement of data-driven solutions, aligning data engineering practices with both business and technical strategies to foster innovation.

### Larsen and Toubro Infotech, India | Data Engineer

Oct 2020 - Jun 2022

- Enhanced data processing efficiency by 30% through the implementation of PySpark code optimizations for data transfer and storage on a Hive server, achieving seamless integration and scalability in high-volume data environments.
- Led the optimization of Tableau data visualizations for business intelligence, reducing the decision-making process time by 30% by generating efficient Hive Query reports for datasets exceeding 1TB.
- Directed the development of robust data models and comprehensive data dictionaries to improve data quality and integration, utilizing SQL and Python to establish and maintain best practices in data management across the team.
- Developed and deployed advanced analytics solutions that streamlined data workflows and reduced processing bottlenecks, directly impacting project timelines and resource allocation.
- Collaborated with cross-functional teams to identify and resolve data-related issues, ensuring data consistency and accuracy in reporting, which supported strategic business initiatives and growth.

## TECHNICAL SKILLS

**Languages:** Python, R, SQL

**Web Technologies:** HTML/HTML5, CSS/CSS3, XML, jQuery, JSON, Rootstrap, Angular JS

**Python Libraries/Packages:** NumPy, SuPy, Doto, Pickle, Pyside, Pylables, Data Frames, Pandas. Matplotlib, SQLAlchemy

**Tools:** Tableau, Power BI, Advanced Excel (Pivot Tables, VLOOKUP)

**IDE:** PyCharm. PyScripter. Spyder. PyStudio. PyDev. IDLE. NetBeans. Sublime Text Visual Code

**Machine Learning and Analytical Tools:** Supervised Learning (Linear Regression, Logistic Regression, Decision Tree, Random Forest, SVM, Classification, Unsupervised Learning (Clustering, KNN, Factor Analysis, PCA), Natural Language Processing, Google Analytics Fiddler, Tableau.

**Cloud Computing:** AWS, Azure, Rackspace, OpenStack

**AWS Services:** Amazon EC2, Amazon S3, Amazon Simple DB, Amazon MQ, Amazon ECS, Amazon Lambdas, Amazon Sagemaker, Amazon RDS, Amazon Elastic Load Balancing, Elastic Search, Amazon SQS, AWS Identity and access management

**Databases/Servers:** MySQL, Apache Airflow, Snowflake, PostgreSQL CouchDB, MongoDB, Teradata, Apache Web Server.

**ETL:** Informatica, Data Stage, SSIS.

**Build and CI tools:** Docker, Kubernetes, Maven, Gradle, Jenkins, Hudson, Bamboo

**SDLC/Testing Methodologies:** Agile/ Scrum, Waterfall, TDD

## PROJECTS

### Bank Market Capitalization ETL Pipeline | (Python, ETL, Web Scraping)

Mar 2023

- Developed and optimized a global ETL pipeline for the top 10 banks worldwide, enhancing data analytics and logging protocols to improve operational efficiency by 20% through advanced web scraping and multi-currency normalization.

### Fake Job Posting Prediction | (Python, Data Mining, Machine Learning)

Sep 2023

- Developed a job posting authenticity classifier using Naïve Bayes and Stochastic Gradient Descent, achieving a 97.4% accuracy rate in detecting fraudulent advertisements with advanced text processing via Python NLTK.

### Text Summariser | (Python, Machine Learning, Web, Data Analysis)

Mar 2020

- Developed a text summarization system using NLP, unsupervised learning, and text detection, achieving a 70% reduction in document length and a 50% decrease in reading time, significantly enhancing user productivity and readability.