JADEN OCA

jadenesoca@gmail.com • 281-467-4955 • Dallas, TX (Open to relocation) • LinkedIn • Portfolio

SUMMARY

2 years of experience in data science and analytics for large-scale SaaS products in financial services. Built ML pipelines that automated 95% of claim classifications, delivered KPI analytics supporting 70% YoY revenue growth, forecasted pricing across 8K tenants using ensemble models, and deployed AI tools adopted by 100-person department. Skilled in Python, SQL, and Tableau.

SKILLS

Programming Languages: Python (Pandas, NumPy, Matplotlib), SQL, R
Software: Snowflake, Git, Claude, Gemini, Power BI, Tableau, Quicksight, Excel, Azure

PROFESSIONAL EXPERIENCE

Capital One – *Associate Business Analyst* **Plano, TX**
Fortune 100 bank with $470B+ in assets across 100M+ customers **August 2025 – Present**

- Support 70% YoY revenue growth for auto lending SaaS platform by monitoring churn, ARR, and customer lifetime value across 19K dealers, informing forecasts and GTM decisions
- Reduce research time by 90% for 100-person department by building custom Gemini assistant grounded in GTM policies, reporting metrics, and internal terminology, accelerating document synthesis, pricing lookups, and trend analysis
- Inform GTM strategy for B2B SaaS products by building and maintaining quarterly report from scratch, using SQL to pull and analyze transaction data from 19K auto dealerships, ensuring reproducibility with Git
- Drive pricing decisions by delivering competitive analyses across SaaS product portfolio, benchmarking pricing and positioning against key competitors

Cotality – *Data Science Intern* **Irvine, CA**
Property intelligence leader with 5.5B U.S. property records serving 98 of top 100 mortgage lenders **June 2025 – August 2025**

- Automated 95% of claims review process through building machine learning pipeline in Python and SQL with TF-IDF vectorization and Logistic Regression model
- Built scalable NLP feature engineering pipeline by designing ETL framework to clean and standardize 40K peril records into a single source of truth in Snowflake

Truist Financial Corporation – *Data Science Intern* **Atlanta, GA**
Financial services corporation managing over $527B in assets with over 15M clients **June 2024 – August 2024**

- Increased recovery for past-due payments by 11% through analyzing 23M transactions to identify trends between when alerts are sent and effectiveness in generating payments within three days
- Enhanced decision-making for a 50-member department by creating a Tableau dashboard analyzing 13 months of client behavior across 6M accounts, leading to improved assessment of payment trends

Link Logistics Real Estate – *Data Analyst Intern* **New York, NY**
Industrial real estate operator managing 535M square feet in assets **May 2023 – March 2024**

- Forecasted leasing prices for 8K tenants in five major markets using Random Forest and XGBoost models, uncovering trends to inform pricing strategies, presenting findings to Chief Data Officer
- Improved model performance by 12% by incorporating geographic and demographic factors using Azure and GeoPandas
- Identified three underutilized investments with projected 8% ROI uplift through regression based financial modeling across 4K properties in 51 geographic markets

PROJECTS

Spotify Streaming Insights – *Language processing and time series analysis project* **November 2024 – December 2024**

- Built a pipeline to transform Spotify stream and lyric data into NLP-ready features, fine-tuned GPT-2 for lyric generation, and visualized insights using BERT-based topic modeling
- Facilitated text analysis by scraping lyrics to user's top 790 tracks across top 10 English artists using Genius.com API
- Increased interpretability and reduced testing loss by 25% by feature engineering songs into 22K lyrical lines using PyTorch
- Detected patterns in listening behavior over six-year period of 400K streams using seasonal decomposition

ADDITIONAL EXPERIENCE

Recovery From Stress Lab – *Research Assistant* **August 2024 – May 2025**

- Identified actionable trends between burnout, workload, and demographic variables across 500 employees, contributing findings to academic papers using linear regression models and correlation matrices in RStudio
- Reduced analysis time by 50% through developing reusable R script to summarize trends between 12 different response variables, standardizing analysis pipeline for team of 5

EDUCATION

Texas A&M University **College Station, TX**
Bachelor of Science in Statistics | Minors in Business and Psychology **May 2025**