# Aradhya Agrawal

agrawalaradhya121@gmail.com | **724-647-8390** | www.linkedin.com/in/agrawalaradhya0204/ | **Houston, TX (Open to Relocate)**

## Skills

- **Programming & Databases:** Python, SQL, R, C++, SAS, MySQL, PostgreSQL, Oracle, Redis
- **Big Data & Data Engineering:** Apache Spark (PySpark), Hadoop, BigQuery, MapReduce, Kafka, Apache Airflow
- **Cloud & Platforms:** AWS, Google Cloud, Azure, Docker, GitHub, Streamlit
- **Visualization & BI Tools:** Tableau, Power BI, Looker, Matplotlib, Plotly
- **Python Libraries & ML/AI:** Pandas, NumPy, SciPy, Scikit-learn, TensorFlow, Keras, GeoPandas, Geopy
- **Other Skills & Analytics:** Machine Learning, Generative AI, Data Modeling, Data Architecture, Data Quality, Root Cause Analysis, Ad-hoc Reporting, Business Reporting, Customer Analytics, Data Strategy, Data Wrangling, System Design, n8n, Antigravity

## Work Experiences

**Data Analyst - Academy Sports + Outdoors** **Feb 2024 – Present**

*Python, SQL, BigQuery, KPIs, T-Test, Excel, Snowflake, Treasure Data, Presto, Looker* **Houston, TX**

- Led customer data analysis, partnering with stakeholders to address business questions, driving informed decisions across departments and optimizing resource allocation, resulting in 40-50% improvement in decision-making speed.
- Composed Looker Studio dashboards, providing self-service analytics capabilities and slashing report generation time by over 80%; automated weekly reporting for the marketing department, saving 15 hours per week.
- Led data collection, feature preparation, and model development support for a machine-learning-driven Media Mix Model contributing to a 20% reduction in inefficient media spend.
- Pioneered Customer Lifetime Value (CLV) models using the Starbucks methodology to isolate high-value customers, triggering personalized marketing, and boosted overall retention 14-18% in high-risk segments.
- Streamlined a logistic regression model, predicting high-value "S1" customers with over 90% accuracy, enabling teams to target valuable customers earlier and improve customer engagement.
- Created a customer propensity model, using machine learning algorithms (K-Nearest Neighbors) to generate likelihood scores for A/B testing, achieving a lift in top decile > 3.2x vs random targeting and accelerating marketing campaign effectiveness.
- Revolutionized year-over-year KPI generation in BigQuery across 70 billion rows by refining SQL queries, improved query performance by 80% and providing faster insights for decision-making.
- Revamped the method of defining geography for e-commerce customers, incorporating spatial libraries in Python to improve geography definition for over 12 million customers and enhance accuracy and personalization in marketing efforts.
- Led statistical hypothesis testing (t-tests, A/B experiments) in Python and SQL across 100+ campaigns, validating performance improvements with 95% confidence, increasing conversion rates by 15% and reducing ineffective promo spend by 20%.
- Standardized Excel-based reporting to visualize key performance indicators, delivering granular insights into merchant and department sales trends, and was adopted by 90% of the analytics team.
- Identified and classified 1+ million previously untracked customers using a novel clustering algorithm, refining KPI calculations and slashing inaccuracies in year-over-year sales reports by approximately 21%.
- Spearheaded the creation of visualizations that enhanced year-over-year analysis of over 30 million customers, resulting in management leveraging insights to make key business decisions.
- Led the STEM and Analytics group at ASO, leading monthly/ quarterly meetings and discussions for a cross-functional team of over 50 members, fostering collaboration and executing group objectives.

**Research Assistant (Analytics Engineer) - San Diego State University** **Jun 2023 – Feb 2024**

*Python, Machine Learning, ETL, AWS* **San Diego, CA**

- Automated an ETL pipeline integrating data from transactional systems, digital analytics, and marketing platforms, processing 20–40 GB daily and storing it on AWS S3; developed a data governance system centralizing metric definitions and automating data quality checks, reducing discrepancies by 60–70%.
- Engineered a handbag defect detection model using YOLOv3 with PyTorch and Darknet, cutting manual inspection time by 60% and improving quality control processes.
- Trained and deployed a Random Forest Regression model on AWS to estimate handbag sales based on popularity and price, boosting a 9% sales increase through targeted incentives; orchestrated data processing and ML workflows using Pandas, NumPy, and Apache Airflow, consolidating 6 disparate sources into a structured daily dataset.

**Data Engineer - Motionworks International** **Sep 2022 – May 2023**

*Python, SQL, Airflow, GCP services (Bigquery and GCS), Git, Postman, JIRA, Docker* **San Diego, CA**

- Revolutionized spatial data accessibility by designing a state-level data model for India, mitigating query times by 50% and allowing analysts to process 500K+ spatial data points daily; streamlined ETL pipelines using Apache Airflow and PySpark to ingest 1M+ rows into BigQuery, increasing daily data volume to 20GB while cutting runtime to under 1 hour.
- Developed and optimized a trip classification model using Uber H3 edge data, achieving 82% accuracy with Decision Tree and enhancing performance by 15% through Random Forest and XGBoost feature reduction to 35 features.
- Conducted large-scale comparative analysis of routing APIs with coordinates from 80M devices, identifying optimal solutions that reduced geocoding errors by 3.2% and enhanced location data quality.

**Data Scientist - Amazon Incorporation** **May 2022 – Aug 2022**

*Python, SQL, Statistical Modeling, and Data Visualization tools*. **Seattle, WA**

- Improved sales predictions by 12% by predicting lost sales during product unavailability and vendor lead time shortages, embedding the model with the inventory management system.
- Simplified a mathematical model for lost sales prediction in a Product Line, contributing to a projected $5M revenue increase through better stock availability.
- Conducted A/B testing, comparing the new lost sales prediction model against existing baseline forecasting methods, which reduced stockout errors by 15-20%.

**Research Assistant - Metabolism of Cities Living Lab** **Jan 2022 – May 2022**

*Python, SQL, Machine Learning and Cloud Technologies (AWS).* **San Diego, CA**

- Extracted and processed 500+ satellite images daily using web scraping and an ELT pipeline, storing data in AWS S3; established a YOLOv3-based debris detection system to automatically identify plastic waste, decreasing manual inspection costs by 90%+ and accelerating environmental response initiatives.

## Project Experiences

**AI-Powered Workflow Automation & Chatbot (n8n + OpenAI)**

- Automated a no-code automation system in **n8n** using the *Trigger → Nodes → Action* architecture, automating multi-step workflows and lowering manual effort by **50%+**; scaled reusable patterns to internal use cases, strengthing operational efficiency by **40%**.
- Architected an AI-powered ecommerce chatbot ingesting **1,000+ CSV product records**, integrating OpenAI with dynamic prompt engineering and webhook triggers to achieve **90%+ accuracy**; implemented data transformation and fallback/escalation logic, optimizing unresolved queries by **30%**.

**Research Data Pipeline and Analytics Dashboard (OpenAlex API)**

- Engineered an end-to-end pipeline ingesting **5K+ OpenAlex records** into cloud-hosted **PostgreSQL (Neon)**, enabling sub-second queries; automated ETL and validation processes to achieve **99%+ accuracy** and cut manual preprocessing by **80%**.
- Built an optimized **Streamlit dashboard** with indexed SQL queries, improving analysis time by **60%** and load time by **40%**; accelerated development by **30%** using AI-assisted tools and managed code via GitHub best practices.

## Education

Master of Science, **Big Data Analytics**, San Diego State University **Aug 2021 - May 2023**

Bachelor of Technology, **Computer Science**, Medi-Caps University **Aug 2017 - May 2021**

## Recognitions

**MVP Award -** Academy Sports and Outdoors **Jun 2025**