

# Ja-Yuan Pendley

(347) 496 -2445 | jay.pendley1218@gmail.com | [LinkedIn](#) | [My Portfolio](#) | Brooklyn, NY

## SUMMARY

An accomplished Business/Data Analyst with a master's degree in data science and over a decade of experience. Proven expertise in enhancing business efficiency, driving revenue growth, and managing complex data-driven projects. Adept in process simplification and management, offering excellent communication skills to influence peers and senior management effectively.

## PROFESSIONAL EXPERIENCE

### **Data Scientist - AI Training (Freelance)| Outlier| Brooklyn, NY | November 2023 – Present**

- Solving complex data problems, like data generation, data cleaning, data visualization, and machine learning modeling
- Wrote robust test cases to confirm code works efficiently and effectively.
- Actively participate in discovering and applying standardization methods in calculating natality rates
- Wrote explanations of how data science can solve problems and evaluate various solution approaches.
- Collaborate with shareholders and stakeholders across departments.
- Developed and implemented advanced machine learning algorithms to enhance AI accuracy, resulting in a 15% increase in prediction precision.
- Collaborated cross-functionally with engineering teams to optimize AI model architecture, reducing training time by 20%.

### **Business Data Analyst Consultant | Maryland Department of Health | Brooklyn, NY | September 2022 - June 2023**

- Executed detailed data analysis and cleaning, leveraging SQL, Python, and ETL processes, leading to optimized business operations and decision-making.
- Automated the ETL data pipeline employing Python, SQL, and Apache Airflow, aligning with business-specific outcomes.
- Conducted data point gap analysis, enhancing systems' functionality and effectiveness, and performed regular data quality checks.
- Create interactive dashboards almost a year.
- Build and maintain relationships with healthcare stakeholders.
- Analyze data for trends and patterns and interpret data.
- Collect Health Records to make sure data is clean readable and easy to work with
- Collaborated with cross-functional teams and utilized advanced data analytics tools to collect, analyze, and interpret market trends, resulting in a 15% increase in sales revenue.
- Utilized CRM systems and worked closely with the IT department to streamline data collection processes, reducing data entry errors by 20%.
- Led training sessions for team members on best practices for collecting and utilizing data effectively, leading to a 30% improvement in data accuracy.

### **Technical Data Analyst| New York Institute of Technology | New York, NY | September 2021 – September 2022**

- Managed data collection, cleaning, transforming, and validation processes to derive meaningful business insights.
- Developed business reports and financial forecasting models using Python, SQL, and PowerBI, contributing to strategic decision making.
- Conduct demographic research analysis services to include custom statistical programming with confidential data.
- Improved process efficiencies by coordinating with stakeholders, ensuring business and technology alignment.
- Maintained data archives for reference, contributing to accurate and efficient business operations.
- Implemented data validation protocols and quality control measures within the team, resulting in a 30% improvement in data accuracy and reliability across all reports.
- Utilized advanced analytical tools to interpret complex datasets gathered from various sources, leading to the identification of key trends and insights that influenced strategic decision-making.
- Collaborated with cross-functional team members and data systems to streamline data collection processes, resulting in a 20% increase in efficiency and a 15% reduction in data errors.

### **Data Analysis | Franklin Delano Roosevelt High School| Brooklyn, NY |February 2017– June 2020**

- Data Collection and Cleaning: Collect, clean, and preprocess large datasets to ensure accuracy and reliability for analysis 3+years.
- Analysis and Interpretation: Conduct exploratory data analysis to identify trends, patterns, and correlations. Interpret findings to provide meaningful insights to support business objectives.
- Report Generation: Develop and generate regular reports for various departments, summarizing key metrics and highlighting areas for improvement.
- Database Management: Assist in maintaining and updating databases, ensuring data integrity and accessibility for analysis.
- Implemented data visualization tools to create interactive dashboards for department heads, improving decision-making processes and increasing efficiency in reporting.
- Developed automated reporting system for all departments, reducing report generation time by 50% and increasing accuracy of key metrics by 30%.
- Analyzed data from various departments to identify bottlenecks and inefficiencies, resulting in a 15% increase in overall

productivity across the organization.

## **TECHNICAL PROJECT EXPERIENCE**

### **Airbnb Data Analysis**

November 2022

- Applied SQL and Python for data extraction, transformation, and loading (ETL) processes to analyze Airbnb data.
- Identified trends and patterns through data analysis, contributing to strategic decision-making.
- Used Tableau for data visualization, facilitating easy comprehension and usage in business discussions.
- App link: [Airbnb Website](#)

### **Song Recommendation System using Spotify's API**

November 2022

- Performed correlation analysis, data visualization and k mean clustering, to uncover patterns and insights within Airbnb's open dataset.
- Visualized the analysis using Folium and Plotly.
- App link: [Spotify Recommendation API](#)

### **NYC Taxi and limo Data Analyst**

November 2022

- Performed correlation analysis, data visualization and k-mean clustering, to uncover patterns and insights within NYC Taxi and limo.
- Visualized the analysis using Seaborn.
- A/B Testing and Linear Regression

App link: [NYC Taxi and Limo Website](#)

### **Web application using Kubernetes**

December 2023

- Deployed a web application on AWS Elastic Kubernetes Service by creating a Docker container.
- Performed cost-benefit analysis of using ECS (Elastic Container Service) vs EKS (Elastic Kubernetes Service).
- Created a repository for the Docker containers in AWS Elastic Container Registry.

### **ETL Pipeline using Azure Data Factory, Databricks and Snowflake**

February 2023

- Created a complex ETL pipeline to extract data from Azure Blob Storage Container.
- Imported a Databricks Notebook in Data Factory to perform the complex transformations and perform Machine.
- Learning Tasks on the data.
- Loaded the data into Snowflake Data Warehouse for further analysis.

## **EDUCATION**

Master of Science, Data Science | New York Institute of Technology

Bachelor of Computer Systems Technology | New York City College of Technology

## **CERIFICATION**

- Google Advanced Data Analytics is authorized by Google.
- AWS Cloud Technical Essentials authorized by Amazon Web Services
- Architecting Solutions on AWS authorized by Amazon Web Services
- Data Wrangling, Analysis and AB Testing with SQL authorized by UC Davis,
- Microsoft Azure Data Engineering authorized Microsoft

## **Technical Skills**

Programming Languages:	Python, R, Java, PySpark, R, JavaScript, C++, SQL, PL/SQL, QlikView, Jira, Excel, SAS, SPSS, HTML
Libraries and Frameworks:	Django, Flask, scikit-learn, Xgboost, Keras, TensorFlow, Pytorch, Pandas, NumPy, SciPy, Kubernetes, Spark, Beautiful Soup, ad-hoc, Hadoop.
AWS Cloud Services:	AWS Lambda, Secrets Manager, IAM, Redshift, AWS Elastic Beanstalk, AWS EKS, AWS EC2, AWS DynamoDB, AWS SNS, AWS Lex, SQS, Event bridge
Azure Cloud Services:	Azure Synapse, Cosmos DB, Azure Databricks, Azure AKS, Azure Blob Storage, Step Functions, Azure Data Factory (ADF), Azure Logic Apps, Azure Function Apps
Databases:	Redshift, Snowflake, PostgreSQL, MySQL, Neo4j, Azure Synapse, SQL Server, MongoDB, DynamoDB, Aurora DB, Cloud Bigtable, Entity Relationship Diagrams
SDLC Methodologies:	Agile, TDD, Waterfall, SCRUM, Portfolio Management
Data Pipeline Orchestration:	Docker, GitLab CI/CD, Airflow, Jenkins, Kubernetes, Quality assurance, Write and review requirements, ESP, Verifying and populating patient clinical data in databases.
Data Visualization:	Matplotlib, Looker, Tableau, PowerBI, Qlik Sense, Visio, Charts, Graphs, Dashboards

Machine Learning Methods:

A/B testing, Natural Language Processing (NLP), Classification, SVM, Regression, Prediction, Statistical Modeling, Recommendation System, Time Series Forecasting, Computer Vision, Transformers, Large Language Models (LLM)

ETL, ML and Big Data Tools:

Machine Learning, Deep Learning, Databricks, ML Flow, Dataiku, Data Informatica, HDFS, MapReduce, Hive, UAT, Data Cleaning, Data Reporting, KPIs, Data Import, User Stories, Audit, identify trends, patterns, and data insights, Digital Data