Dear Hiring Manager,

As a detail-oriented and analytical professional with a strong background in data analysis and software development, I am confident in my ability to excel as a Data Analyst. Below are key skills and experiences that align with the job requirements:

• Data Analysis and Visualization:
  • Extracted, analyzed, and visualized reports from Workday and PeopleSoft in Tableau to enhance data-driven decision-making (Business and HR Analyst, Arizona State University)
  • Conducted comprehensive exploratory data analysis on large-scale logistics datasets, uncovering critical KPIs that were strategically integrated into Tableau dashboards (Data Analyst Intern, Cube Logistics)

• Data Engineering and Pipeline Development:
  • Developed a data pipeline leveraging Apache Spark to process cricket match data, handling large-scale data ingestion from Amazon S3, transforming and validating schemas, and optimizing performance for fast querying and analytics (Sports Data Processing Pipeline)
  • Orchestrated and optimized data pipelines using scikit-learn, incorporating GridSearchCV for hyperparameter tuning, achieving price prediction accuracy of 91% (Machine Learning Intern, Tequed Labs)

• SQL and Database Management:
  • Enhanced SQL query performance by optimizing SARGability, implementing composite indexing, and restructuring queries, resulting in a 50% reduction in execution times and improved efficiency in a Snowflake-based data warehouse (Data Analyst Intern, Cube Logistics)
  • Leveraged statistical methods like ANOVA for data preprocessing and classification, while developing and implementing outlier detection models such as MAD and Boxplot, achieving high F-scores (up to 0.99) in detection across diverse datasets (Outlier Detection and Statistical Modeling for Sensitivity-Driven Data Analysis)

• Cloud Computing and AWS:
  • Engineered scalable cloud-based image recognition services using AWS, implementing an advanced auto-scaling algorithm (Elastic Image Recognition Service on AWS)
  • Developed a real-time data processing pipeline using Apache Kafka for streaming stock market data, AWS S3 for storage, and AWS Glue to transform raw stock market data into a queryable format and automate the loading process (Real-Time Stock Market Data Pipeline)

Sincerely,
Harsh Victor Challa