Syed Azfar Ali

Karachi, Pakistan | +92-340-2007760 | azfarali021.aa@gmail.com

linkedin.com/in/azfarali-saa0021 | github.com/azfarali16 | azfarali16.github.io

SUMMARY

Data Scientist with experience in building ML models, crafting data-driven insights, and deploying data solutions. Skilled in Python, SQL, ETL, and machine learning with a strong foundation in data engineering.

EDUCATION

NED University of Engineering and Technology Karachi, Pakistan
Bachelor of Science in Computer Science *Nov. 2021 – June 2025*

CERTIFICATIONS

Data Analyst with Python – DataCamp
SQL Associate – DataCamp

EXPERIENCE

Firstplant Remote
Graduate Data Engineer (Part-Time) *May 2025 – Present*
- Contributed to automation scripts and data ingestion pipelines using Apache Airflow to streamline workflows, enabling automated collection and processing of sustainability-related product data from diverse web sources and technical documents.
- Designed data extraction workflows to parse and consolidate product and environmental information from diverse documents.

Firstplant Remote
Data Engineer Intern *Nov 2024 – May 2025*
- Built custom web scrapers to extract sustainability-related product and environmental data from targeted websites.
- Cleaned, structured, and standardized raw scraped data for integration into downstream data pipelines.

StaunchTie Innovations Remote
Software Developer *Oct 2024 – Dec 2024*
- Developed analytical dashboards using Dash and Plotly to visualize key data insights.
- Collaborated on frontend development and interface improvements to enhance user experience.
- Implemented data visualization features supporting business decision-making and performance tracking.

PROJECTS

PhishTracer: Phishing Detection System
- Enabled real-time phishing detection by deploying an ML-powered system as a Flask microservice, integrated with a Node.js frontend. Processed and cleaned over 50,000 email samples, applying optimized feature engineering to improve detection accuracy. Achieved an F1 score of 85.71, supporting reliable and automated threat classification.

E-Store ETL Data Pipeline
- Streamlined sales reporting by building a ETL pnipeline that extracted data from Amazon RDS, transformed it into a star schema, and loaded it into MySQL. The project significantly improved query performance and enabled data-driven decisions through interactive Power BI dashboards for sales and inventory trends.

Daraz Product Review Sentiment Analysis
- Supported product evaluation efforts by developing a tool that scraped unstructured reviews from Daraz and applied sentiment analysis using VADER. This provided actionable insights into customer opinions, helping quantify product feedback at scale.

SKILLS

Languages: Python, SQL
Frameworks: Flask, TensorFlow, PyTorch, Dash
Tools: Git, REST APIs, Power BI
Data Skills: Pandas, Scikit-learn, Machine Learning, EDA, Data Warehousing Concepts, ETL Pipelines, Data Cleaning, Data Validation, Automation, Web Scraping
Core Skills: Data Analytics, Critical Thinking, Attention to Detail