

# Basit Faisal

<https://www.linkedin.com/in/basitfaisal/> | basitfaisal03@gmail.com | [Portf](#)

## EDUCATION

---

### Ghulam Ishaq Khan Institute of Engineering Sciences & Technology

Pakistan

Bachelor of Science in Data Science

2025

- CGPA: 3.64 (Dean's Honor List)
- Relevant Courses: (Data Structures, Artificial Intelligence, Data Engineering, Data Mining, Inferential Statistics, Data Warehousing & BI, Databases, Engineering Economics)

## WORK EXPERIENCE

---

### Motive

June 2025 - Current

Data Quality Specialist

- Scraped, cleaned & uploaded **2500+** prospect accounts creating a pipeline value of **35** million USD.
- Developed an automation tool to match thousands of accounts with millions of contacts, reducing processing time from **4 hours to 0.5 hours** via parallel processing and big data tools like PySpark.
- Orchestrated Airflow pipeline to scrape **50k unions** documents from department of labor, parse it via open-sourced OCR models and feed it to LLM to build the perfect unions data pipeline and storage.

### DAO PropTech

January 2025 – May 2025

Junior Data Scientist

- Pioneered the development of a Retrieval-Augmented Generation (RAG) chatbot, leveraging **Milvus** as a vector database and OpenAI's LLM to create an agentic solution (Lang-chain framework), increasing customer engagement by **25%** within the first quarter of deployment.
- Engineered Bayesian statistical models to analyze user activity across company platforms, resulting in a **15%** increase in user-to-client conversion rates by designing targeted campaigns based on behavioral insights.
- Building Data pipelines using Airflow, click-house DB to improve dashboard loading time and increase data access.

### S&P GLOBAL

June 2024 - September 2024

Data Governance Intern (Market Intelligence)

- Enhanced data integrity by **30%**, deploying Python scripts on GCP to validate **50+** datasets and flag discrepancies.
- Increased content generation efficiency by **55%**, fine-tuning GPT-4 to generate **1000+** new and existing S&P index descriptions.
- Improved stakeholder decision-making by designing a dashboard for real-time tracking of **200+** datasets.
- Reduced search time by **40%**, standardizing index retrieval in S&P databases by integrating Bloomberg naming conventions.
- Achieved **1st** place in a global intern innovation challenge, enhancing data accessibility by **35%** through creating a Retrieval Augmented Generation (RAG) feature on **CapIQ** using GENAI.

### CITI Bank

June 2023 - August 2023

Summer Data Analyst (Securities & Services Division)

- Amplified portfolio returns by an average of **20%**, developing 3 investment models.
- Elevated investment precision by processing and refining over **10** datasets.
- Reduced trade settlement discrepancies by **30%**, implementing accurate tracking using Excel Dashboards.
- Decreased time wastage by **30%**, creating tax reports in Excel using built-in functions.

## LEADERSHIP EXPERIENCE & Co-Curriculars

---

ICPC (International Collegiate Programming Contest)

August 2021 – November 2021

Participant

- Solved algorithmic challenges under time constraints in a team of three, developing skills in data structures, dynamic programming, and graph theory. Developed efficient algorithms in Python to solve problems.

## SKILLS & INTERESTS

---

**Skills:** Quantitative modeling, Python, SQL, R, Machine Learning, Data visualization (PowerBI, Tableau), Cloud computing (AWS, GCP), Statistical analysis (SPSS, SAS), MS Excel (Data Modeling, Pivot-Tables), Big Data technologies (Hadoop, Databricks), Docker, Git/Github, Salesforce, Snowflake, Airflow.

**Interests:** Financial markets, Quantitative research, Data-driven decision making, Algorithmic trading.