Mirza Ahmer Gull

LinkedIn: linkedin.com/in/mirzaahmergull
Github: github.com/mirzaahmergull

Email: mirzaahmergull@outlook.com
Mobile: +92-301-4385785
Website: codewithmag.dev

EXPERIENCE

- **CityScape Technologies** — On site
  Data Engineer (Full-time) — *December 2022 - Present*
  - **Sourcing Data pipelines**: Developed and implemented various sourcing pipelines using Python-based ETL/ELT processes to extract property listings data from multiple MLSs. Successfully deployed these pipelines on Jenkins and Airflow, ensuring continuous delivery of updated data.
  - **PIPO**: Designed and deployed a Point-in-Polygon operation (PIPO) using Python, Redshift, RDS, and GeoSpatial Pandas to provide Cityscape's teams with geographical hierarchy information for listings based on their latitude and longitude, using Canada's geospatial data.
  - **Latest Listings Sourcer**: Developed the Latest Listing Sourcer using Python, RDS, AzureSQL, and Redshift, deployed and scheduled on Jenkins to ensure the delivery of up-to-date, accurate, and clean data for each property in Canada, regardless of the MLS source. This data was used by the Analytics Team, the Real Estate Generative Assistant Team, and on the CityScape property portal (cityscapeone.com).
  - **Cleaning Pipelines on Airflow**: Deployed and managed Airflow and created cleaning pipelines to generate Silver and Gold Standard Data that complies with Canadian property business rules. This high-quality data was used by various teams. The pipelines were developed using Python, Advance pandas cleaning processes and MODIN, seamlessly deployed on Jenkins and Airflow for frequent batch cleaning of sourced data.
  - **Looker Studio Dashboards**: Designed interactive Looker Studio dashboards to provide actionable insights on data and pipeline monitoring, significantly enhancing business efficiency. These dashboards proved valuable to the Data Engineering Team, increasing their productivity and data integrity.
  - **Data Flow Architecture**: Designed the architecture for data processing pipelines, creating a streamlined flow for the execution of various pipelines. This enabled the successful and conflict-free achievement of their goals.

- **MLSense.ai** — On site
  Data Engineer & Full Stack Developer (Full-time) — *June 2021 - December 2022*
  - **Sedex**: Built automated Python-based data ETL/ELT pipelines to load raw supply chain data from various sources to Snowflake. Cleaned and standardized the data using Materialized Views and Python in SnowPark. Applied Row Level Access and Dynamic Data masking policies to mask PII information data. The data was further used by other teams and was used on ThoughSpot for AI-driven data decisions.
  - **Text Analysis using AWS Textract**: Built pipelines to extract Forms/Tables data from image-based PDFs of UK company's bills. Cleaned and stored the extracted forms and tables in AWS Glue. Used AWS Athena and AWS QuickSight to show analysis on the data. The whole pipeline was event-driven using S3 Buckets. This project enabled the company to move from legacy paperwork to a more modernized digital solution, producing useful insights that helped them make better decisions.
  - **ETL Scapper**: Created a scheduled Airflow pipeline to scrape data from different sites (confidential), clean it, and load it into BigQuery. Almost 27 million records were scraped initially, followed by more than 100K daily. Python, BigQuery, and Airflow were used in the process..
  - **IPA for Kering (Full Stack Web application)**: In this project for Kering (parent company of Gucci, Balenciaga, etc) I Led a team of four developers to create an End-to-End BI Web Application in Django and React, showing marketing data insights in beautifully presented graphs and tables created using ChartJS, ApexCharts, and more. Created ETL pipelines using Cloud Functions to populate raw data in CloudSQL from the Meta(Facebook) API, Trueclicks API, and Google Display and Video API. Further cleaned and standardized the data according to business rules. Created an admin dashboard for managing this WebApp. The application itself was deployed on Google App Engine.

- ○ **DMS (Full Stack Web application)**: Worked as Secondary Team lead of a team of 6 Engineers to develop a document management system for the UK Government. This web platform allowed users to store, OCR, share documents securely (with optional expiry and masking of PII information), and search within documents. My role was to develop REST API using Django DRF and develop the front end using React and Material UI. Also, used ElasticSearch to search within documents, for which I developed an event-driven pipeline using CloudStorage. On trigger, a Python Cloud Function was used to extract/OCR text and push it into ElasticSearch. Developed GitHub Actions CI/CD pipeline to auto-deploy the web application on Google App Engine.
- ○ **Full Stack Application Development (Dictionaries)**: Developed a Full Stack Application Web Application using Django DRF and React on the frontend; ElasticSearch was used as a search engine; this app could take any language and give its meanings, similar words/sentences from a total of 27+ Dictionaries and APIs and translate that word along with its meaning/sentence to any specified language within milliseconds. The application was deployed on the GCP App engine using GitHub actions.

- **Professional Freelancer Fiverr** — Freelance
  Freelancing (Part-time) — *January 2020 - July 2021*
  - ○ **Automation Scripts**: Created automation scripts which ran on daily basis different clients using Shell Scripting and Python.
  - ○ **GUI Automation Using PyautoGui**: Created GUI automation scripts using Python and AutoHotKey for clients, this enabled them to do repetitive tasks easily.
  - ○ **Web Scraping**: Created scrapping scripts for many clients, These scripts ran on Sites with many web pages or dynamic web pages using multi-threading and site exploitation techniques. Scraped and colledted data into csv/database/xml etc

EDUCATION

- **Punjab University of Information Technology** — Lahore, Pakistan
  Bachelor of Computer Sciences — *October 2018 - November 2022*

  ### PROJECTS

  - ○ **Credit Card Fraud Detection**: Created a basic fraud detection system for online credit card transactions in Python using SciKit Learn, NumPy, and pandas.

SKILLS SUMMARY

- ○ **Languages**: Advanced Python, C++, JavaScript, SQL, NoSQL, SnowSQL
- ○ **Frameworks**: Django, DRF, JWT, Flask, HTML, CSS, React, Elastic Search
- ○ **Tools**: Linux, VS Code, Docker, Git, GitHub, PostgreSQL, MySQL, SQLite, MongoDB, Elastic Search Platform, Snowflake, Postman, Tableau, Google Chart Studio
- ○ **Platforms**: Linux, Windows, AWS, GCP, Snowflake Data Warehouse, Kibana
- ○ **Technologies & Services**: App Engine, Cloud Storage, Cloud Run, Cloud Functions, Airflow, GitHub Actions, Cloud Pub/Sub, BigQuery, Cloud SQL, Composer, AWS, Azure, AWS RedShift, AWS RDS, Jenkins, PowerBI
- ○ **Soft Skills**: Top Quality Leadership, Quick Learning, Efficient Problem Solving, Team Work, Writing, Public Speaking, Time Management

HONORS AND AWARDS

- Selected two times as Teaching Assistant for Data Structures and Algorithms Course - May 2021
- Second Runner's in Preliminary ICPC Competition in Lahore, Pakistan
- Bronze Medal in ICPC Asia Topi Regional On-site Programming Contest, 2021
- 10th rank in The ICPC Asia Topi Regional On-site Programming Contest, 2019