Baker Zaitoun

📞 +961 76539424

✉ bakerzaitoun62@gmail.com

in https://www.linkedin.com/in/baker-zaitoun-02b448368/

Data Engineering Intern – Claude Ghanem Shipping
1st Jun 2025 – 1st Sep 2025

- Cleaned and standardized shipment and delivery data from internal databases and CSV files.
- Built and automated ETL pipelines using Azure Data Factory for data integration.
- Designed and managed a Snowflake data warehouse to support reporting and analytics.
- Wrote SQL queries to analyze delivery performance, delays, and customer support requests.
- **Technologies**: Python, SQL, PySpark, Azure Data Factory, Snowflake, GitHub

Projects

1. Care Utterance Analysis
Description: Simulated a customer service improvement project by building basic data pipelines and analyzing customer support transcripts to understand common call reasons and suggest enhancements to a sample IVR system.

- Built data extraction scripts to collect and clean sample customer utterance data from mock Oracle tables (CSV format).
- Used basic clustering techniques (K-means) with an open-source tool to categorize utterances and identify common issues.
- Proposed improvements to a hypothetical IVR menu structure based on grouped utterance patterns.
- Demonstrated how such analysis could reduce call transfers and improve customer experience.
  = Technologies: Py Spark, SQL, Azure Data Factory, Snowflake, Azure DevOps(CI/CD)

2. Intraday Retail Analytics
Description: Developed a mini data pipeline project simulating store performance forecasting for a small retail chain using sample datasets, aiming to predict store traffic and product demand.

- Created ETL scripts to ingest and process sample retail datasets (e.g., store visits, staffing logs, sales history).
- Built basic machine learning models (linear regression, random forest) to forecast customer headcount and sales for a set of mock stores.
- Evaluated model performance and simulated how forecasts could help in better staffing decisions and sales planning.
- Practiced code versioning and deployment using Git and GitHub Actions as a basic CI/CD process.
  = Technologies: Python, Py Spark, SQL, PostgreSQL, Databricks, ADLS Gen2, GitHub Actions

Technical Skills
- Languages: Python, SQL
- Big Data Technologies: Spark, PySpark, Spark SQL
- ETL & Data Processing Tools: Databricks, Azure Data Factory
- Cloud Platforms: Microsoft Azure (Synapse Analytics, HDInsight)
- Data Warehousing: Snowflake, SQL Server
- DevOps & Orchestration: Airflow, Azure DevOps (CI/CD), Docker
- Machine Learning: Pytorch, Hugging Face, Transformers
- LLM & AI Frameworks: LangChain

Education

Bachelor's Degree in Information Technology (IT)

Reference for Projects

tejasborkar7020@gmail.com
Linkedin:linkedin.com/in/tejas-borkar20a