Mina Sonbol

+201222392979 | minasonbol@gmail.com | linkedin.com/in/msonbol | github.com/el-grudge | datascienceportfol.io/el_grudge

Personal Statement

Experienced Data professional with over 15 years of experience in data management, advanced analytics, machine learning and AI. My expertise include designing, implementing, and maintaining Data Warehouses for Telecom operators and E-Commerce platforms, creating and orchestrating Change Data Capture streams and ETL data pipelines, delivering actionable insights through reports and dashboards for Telecom operators, Digital Marketing platforms and Customer Experience AI providers.

Skilled in Python, SQL, and R; experienced with open source ETL frameworks such as Mage and Airbyte, as well as proprietary tools such as IBM DataStage; managed the integration, implementation, and management of data warehouses using cloud solutions such as GCP BigQuery and on-premise solutions such as Teradata.

In addition, I enjoy building innovative tools using RAG pipelines, LLMs and NLP techniques, such as Deep-pod, a streamlit app I built that allows users to interact with their favorite podcasts through Chat, Summarization, and Topic detection.

Work Experience

Silverkey, Cairo, EG — *Sr Data Engineer*
11/2024 - present

- Designed, implemented, and maintained a data warehouse for Dabdoob, a multinational e-commerce platform operating in the GCC region with over 500K users and 1500 daily orders using GCP BigQuery as an OLAP data warehouse and Airbyte for data integration, creating a single source of truth and providing real time insights by improving data ingestion speeds by x3.
- Incorporated data from Dabdoob's production database (including sales, inventory, payment transactions), as well as data from Mixpanel that tracks user journeys on the app, from Adjust to track Ad spend, and from the Promo Engine to track and target users for promotions to increase interaction. The BigQuery warehouse currently has 4 Billion records consuming 1 TB of data.
- Designed and implemented an automated merchandising report system using Python and SQL, integrating data from multiple sources (POS and inventory management systems) to provide real-time restocking alerts, resulting in a 15% reduction in stock-outs and a 10% increase in sales for flagged items.
- Improved the efficiency of the Power BI data models of a government agency working on renewable energy improving the availability of its 15 dashboards and over 300 charts.
- Led the planning of a government analytics platform by creating a structured roadmap with task-level deliverables and timelines to enable descriptive and predictive reporting in coordination with the Ministry of Planning.

Afiniti, Washington, DC — *Sr Data Analyst, Client Data Services*
04/2023 - 09/2024

- Created the monthly reconciliation ETL pipeline using PL/SQL that streamlined the monthly revenue reconciliation process, reducing time by 95%.
- Analyzed customer tenure data for the commercial strategy team showing a strong positive correlation between tenure and stickiness, an insight that helped inform our pricing models and offerings.
- Developed and maintained the monthly lift persistence Power BI dashboard, tracking lift persistence and decay over 30, 60, and 90 days, facilitating timely strategic interventions.
- Used advanced metrics to track baseline performance on a daily basis, monitoring revenue per call and call counts that help with early detection of issues that can impact our operations.

Persado, New York, NY — *Sr Data Scientist*
08/2021 - 01/2023

- Led over 600 A/B tests across 15 accounts, enhancing client interactions by 21.74%, increasing conversion rate by 9.59%, and contributing to a 10.33% increase in revenue.
- Developed and maintained 35 dashboards using SQL, Sisense Periscope, Sigma, and Snowflake, meticulously tracking KPIs across the funnel.
- Led the expertise on statistical analysis and Persado technology, delivering actionable insights to client services teams; enhanced overall product understanding.

Orange, Cairo, EG — *Sr DWH Supervisor*
05/2010 - 12/2019

- Adopted Teradata Communication Data Model to fit France Telecom's business requirements.
- Executed and refined the Churn model using Teradata Warehouse Miner, successfully retaining 77% of high-risk churners in the top revenue segment, resulting in nearly $3M in revenue preservation.
- Managed ETL operations using IBM DataStage, ensuring smooth data processing workflows.
- Generated daily performance reports using SQL and BusinessObjects InfoView, detailing key performance indicators (KPIs) such as aggregated consumption trends.
- Conducted routine platform integrity checks to ensure data consistency across various environments, preventing significant revenue loss for the company.
- Developed and maintained the GeoSegmentation model using Spark SQL, identifying subscribers' home and work locations to enable the commercial team to craft personalized experiences, enhancing customer satisfaction and engagement.

Education

University of Central Florida
M.Sc. in Data Analytics, 01/2020 - 05/2021
Exceptional Graduate Student Award in Data Analytics

Arab Academy for Science and Technology
B.Eng. in Computer Engineering, 09/2002 – 02/2008

Portfolio Projects

Deep-pod: Podcast Interaction System
September 2024

Designed and implemented an AI system that enables users to interact with podcast through chat, summarization, and text extraction.

- Implemented search functionalities using indexes and vector databases, including Elasticsearch and ChromaDB.
- Developed prompts and evaluated their performance with Hit-Rate and MRR.
- Used LLM-As-A-Judge to evaluate the RAG pipeline to compare the performance FLAN-5 with GPT-4o
- Implemented semantic search and experimented with various embedding models (TF-IDF, T5, OpenAI).

Repo available here: https://github.com/el-grudge/simple-deep-pod
Tool available here: https://simple-deep-pod.streamlit.app/

Money Diaries
May 2024

An end-to-end data engineering project that provides a bird's eye view of the Money Diaries blog, tracking blog entries since 2017 and providing insights on product prices, salary distribution, wealth-debt overview, and inflation tracking.

- Implemented an ETL pipeline with Mage that scrapped the data from the blogs and loaded it in a BigQuery
- Transformed the data with dbt and created a data mart that feeds the dashboard
- Used partitioning and clustering for optimized query performance
- Deployed cloud resources with Terraform
- Created a front-end dashboard with Streamlit

Repo available here: https://github.com/el-grudge/money-diaries
Dashboard available here: https://moneydiaries.streamlit.app