Soha Tariq

st.sohatariq@gmail.com | (240) 899-8167 | Washington DC
LinkedIn: https://www.linkedin.com/in/soha-tariq-5226143b/ | GitHub: https://github.com/SohaT7
Portfolio: https://sohat7.github.io/Portfolio/

Summary

Data Analyst/Data Scientist with three years of experience in data analysis, modeling, visualization, and machine learning, working with a variety of structured and unstructured data, such as education, economic and public, labor, ecommerce (sales transactions), weather, and financial data. Looking for a **Data Analyst/Data Scientist** position, to solve problems with Python, R, SQL, Tableau, Machine Learning, and Cloud, in order to generate actionable insights. Willing to relocate.

Skills: Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, SciPy, Tensorflow libraries), SQL, R, STATA, Tableau, Microsoft Excel, VBA, NoSQL, MySQL, PostgreSQL, JavaScript/HTML/CSS, RESTful APIs, PySpark, Amazon Web Services (AWS), Google Cloud Platform (GCP) - BigQuery and Vertex AI. Exploratory Data Analysis (EDA), ETL pipeline, web scraping, API Data Retrieval, data modeling, data warehousing, data visualization, database management, Agile, Machine Learning, Computer Vision.

Professional Certifications

Data Analytics & Visualization	Columbia University	June 2022
Agile Foundations	NASBA Certified	August 2022
Introduction to SQL for BigQuery and Cloud SQL	Google Cloud Training	October 2022
Computer Vision Fundamentals	Google Cloud Training	November 2022

Relevant Experience

Columbia University	New York, NY
Data Analytics & Visualization Bootcamp	December 2021 - June 2022

- Developed Machine Learning and Deep Learning predictive models (regression, classification, neural networks (NN), and deep neural networks (DNN)), using Python (TensorFlow, Scikit-learn, Keras) for a variety of structured and unstructured datasets.
- Experience working with Amazon Web Services (AWS) RDS and AWS S3 buckets, moving data across multiple platforms, and carrying out ETL and data analysis on the data.
- Created data processing pipelines that take data from multiple sources, and cleaned, transformed, summarized, and loaded data into databases, using SQL, Python (Pandas, Numpy, Matplotlib, Seaborn), MongoDB, and PostgreSQL.
- Analyzed data and ran statistical analysis to model, predict, and forecast trends with Excel, VBA, Python, SQL, PySpark, and R.
- Constructed web applications and interactive maps (RESTful APIs, JSON formatting), generated data visualizations and dashboards, and conducted web scraping using Python, R, Tableau, and JavaScript/HTML/CSS.

Success Academy	New York, NY
Associate, Data Analyst (Special Education)	June 2021 - November 2021

- Owned the data collecting, loading, and standardizing processes, and analyzed data in Excel and SQL to identify achievement gaps and recommended academic interventions to the Committee on Special Education.

- Implemented data analytical solutions which improved the workflows for early detection of at-risk students in the school by about 25% - improved database querying, dashboards and visualizations to identify trends and outliers.
- Built effective stakeholder relationships with NYC Department of Education to finalize academic plans for students in need.

Lahore University of Management Sciences (LUMS)   Lahore, Pakistan
Data Research Assistant   June 2017 - March 2018; September 2013 - December 2013
- Built outreach to local communities to understand their perspective on infrastructure and intergenerational occupational mobility.
- Led a team of 9 in data collection from 8,000 households in 15 low-income neighborhoods.
- Conducted mapping of neighborhoods unavailable on Google Maps, which made the entire data collection process faster by about 50%.
- Developed and maintained a database, conducted data mining, preliminary data and statistical analysis (regression, model building, tests), and data visualization, in R, STATA, and Excel.

Innovative Development Strategies (IDS)   Lahore, Pakistan
Data Analyst Intern   June 2014 - August 2014
- Conducted data collection (survey research, interviews, and focus groups) in 7 localities, followed by data cleaning, analysis, and data visualization of the aggregate data in Excel.
- Prepared reports and presentations for the government stakeholders on the issues faced by 5000+ Internally Displaced People (IDPs), with recommendations to alleviate those issues.

Projects

Image Classification | GitHub: https://github.com/SohaT7/Image_Classification
Computer Vision analysis (image classification) in GCP. Custom Machine Learning Model building and evaluation - Deep Neural Network (DNN), Convolutional Neural Network (CNN), and Deep Convolutional Neural Network (DCNN) - using 60,000 labeled images in the MNIST dataset.
- Tools: Google Cloud Platform (GCP), Vertex AI, Python (TensorFlow, Keras, NumPy, Matplotlib, Seaborn), APIs (Vertex AI API, Notebooks API)

Purchase_Analysis | GitHub: https://github.com/SohaT7/Purchase_Analysis
Uses BigQuery to explore the Google Analytics dataset via SQL queries and build a Machine Learning Model (in SQL) to predict whether a visit to the website will result in a sales transaction or not.
- Tools: Google Cloud Platform (GCP), BigQuery, Vertex AI API, SQL, Python, BigQuery Client

Credit Risk Analysis | GitHub: https://github.com/SohaT7/Credit_Risk_Analysis
Supervised machine learning models built and evaluated to predict credit loan risk. Evaluation of the resampling algorithms and ensemble techniques applied to the logistic regression classifier models.
- Tools: Python (Scikit-learn, Imbalanced-learn, Pandas, and NumPy libraries)

NYC Restaurant Analysis |GitHub: https://github.com/SohaT7/NYC_Restaurant_Analysis
Data visualization of the NYC restaurant data, and a Machine Learning Model to gauge whether a restaurant located in a high-income area receives a higher health inspection grade or not.
- Tools: Python (Pandas, Scikit-learn, Imbalanced-learn), PostgreSQL, SQLAlchemy, Tableau, JavaScript (Plotly.js library), HTML, CSS, and Bootstrap

Additional Experience

University of Virginia	Graduate Teaching Assistant	August 2019 - December 2020
United We Reach	Curriculum Writer	November 2018 - June 2019
AIESEC in Lahore & Kenya	Project Manager, Marketing Advisor	September 2011 - August 2014

Education

Masters of Arts, Anthropology	University of Virginia	August 2019 - December 2020
**Bachelors of Science, Economics & Politics	Minor in Mathematics**	LUMS	Sept 2011 - July 2017