Soha Tariq

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

Summary

Data Analyst/Scientist with 2+ years of experience in data analysis, data modeling and data visualization with education data, economic data, sales data, healthcare data, earthquake and weather data, cryptocurrencies and credit risk data, and stocks data. Looking for a **Data Analyst/Data Scientist** position to solve business problems with Python, R, SQL, Tableau, and Machine Learning, in order to generate actionable insights.

Skills: Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, SciPy, Tensorflow libraries), SQL (Structured Query Language), R programming language, Tableau, Microsoft Excel, Visual Basic Application (VBA), NoSQL, PostgreSQL, SQLAlchemy (ORM), MongoDB, HTML5, CSS, JavaScript ES6+, RESTful APIs, Flask, BeautifulSoup, Amazon Web Services (AWS), Google Cloud Platform (GCP) - BigQuery and Vertex AI, Hadoop, version control GitHub, Jupyter notebook, Google Colab notebook. Exploratory Data Analysis (EDA), ETL pipeline, web scraping, data modeling, data cleansing, data warehousing, data analysis, data visualization, database management, Agile, Machine Learning.

Professional Certifications

Data Analytics	Columbia Engineering Bootcamp	June 2022

Professional Experience

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

Data Analyst
- Developed Machine Learning and Logistic Regression models using Python for a variety of datasets (cryptocurrency data, credit risk data, restaurant data, healthcare, education data).
- Experience working in AWS and carrying out data analysis on datasets therein.
- Created data processing pipelines that take data from multiple data sources and conducted data cleansing, transformation, summarization, and loading of the data into databases, using SQL, Python, MongoDB, and PostgreSQL.
- Employed data analysis and statistical analysis to model, predict, and forecast trends using tools in Excel, Python, SQL, PySpark, and R.
- Constructed web applications and interactive maps (RESTful APIs, JSON formatting), generated in-depth data visualizations (graphs, charts, tables), and conducted web scraping using Python, R, SQL, and JavaScript.

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

- Conducted regular data collection, data cleaning, data analysis, data querying, and visualization on scholars' data, in Microsoft Excel and SQL, to identify the need for academic intervention and specialized services from the Committee on Special Education (CSE).
- Improved the tracking and monitoring process design by introducing advanced data analytics which made early detection of at-risk students more efficient by about 25%.

- Conducted effective meetings and built stakeholder relationships with NYC Department of Education (DOE) 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 relationships with the local communities to assess issues faced when engaging in the politics of infrastructure and in intergenerational occupational mobility.
- Led a team of 9 in data collection from 8,000 households in 15 low-income neighborhoods.
- Conducted mapping of the neighborhoods - made data collection faster by about 50%.
- Developed and maintained a database, conducted data cleansing, data mining, preliminary data and statistical analysis, and data visualization, in R , Microsoft Excel, and STATA.

Innovative Development Strategies (IDS)   Lahore, Pakistan
Data Analytics Intern   June 2014 - August 2014
- Conducted data collection (questionnaires, interviews, and focus groups) in 7 localities, followed by data cleansing, analysis, and data visualization of the aggregate data in Excel.
- Prepared a report and presented the results to government stakeholders on the issues faced by the Internally Displaced People (IDPs), with recommendations to alleviate those issues.

Projects

Amazon Product Reviews Analysis| GitHub: https://github.com/SohaT7/Amazon_Vine_Analysis
Big data analysis to determine if there is a positivity bias in product reviews written by Amazon Vine members (paid versus non-paid members).
- Tools: Amazon Web Services (AWS), SQL, PySpark, Python (Pandas library), pgAdmin

Movies ETL Database | GitHub: https://github.com/SohaT7/Movies-ETL
Created an automated ETL (Extract, Transform, Load) data processing pipeline that extracts data on movies and their ratings from multiple data sources, to then transform (data cleansing, formatting, applying functions) and load the data into a movies database.
- Tools: Python (Pandas and Numpy library), PostgreSQL, and SQL

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

NYC Restaurants | https://github.com/SohaT7/NYC_Restaurant_Analysis#machine-learning-model
Analyzes the NYC DOH Restaurant Inspections data to find correlation between income level of the area and health inspection grade for a local restaurant therein.
- Role: Machine Learning   **|** Tools: Python (Pandas, Scikit-learn, Imbalanced-learn), PostgreSQL, SQLAlchemy, Tableau, JavaScript (Plotly.js library), HTML, CSS, and Bootstrap

Additional Experience

University of Virginia . 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 Math . LUMS	September 2011 - July 2017