

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/>

Data Analyst with a certificate in Data Analytics and Visualization from Columbia Engineering, an education in Mathematics (Statistics), Politics and Economics during Bachelor's, and 2+ years of experience working with data. Past work includes data analysis, data visualization, data management, data mining, data cleaning, data collection, statistical analysis, and building stakeholder relationships. Experienced in handling large datasets in SQL, R, Excel, Python and Tableau, and collaborating across multiple teams to help generate actionable insights from the data. Strengths include analytical problem-solving, empathic teamwork, attention-to-detail, and effective communication.

Technical Skills

Tools: Python (Pandas, Matplotlib, Seaborn, Scikit-learn, Tensorflow libraries), Structured Query Language (SQL), Tableau, Microsoft Excel, Visual Basic Application (VBA), R, 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.

Skills: Exploratory Data Analysis (EDA), ETL pipeline, web scraping, data modeling, data cleansing, data analysis, data visualization, database management (queries), agile methodology, machine learning.

Projects

Amazon Product Reviews Analysis | Data Analyst

May 2022

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 | Data Analyst

March 2022

GitHub: <https://github.com/SohaT7/Movies-ETL>

Creates an automated ETL (Extract, Transform, Load) 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 library), PostgreSQL, and SQL

Bikesharing | Data Analyst

April 2022

GitHub: <https://github.com/SohaT7/Bikesharing>

Creates a dashboard to analyze the NYC Citi Bike sharing data, in order to convince investors to start a similar bike sharing program in another city.

- Tools: Tableau and Python (Pandas library)

Kickstarter Campaigns Analysis | Data Analyst

January 2022

GitHub: <https://github.com/SohaT7/Kickstarter-Analysis>

Analyzes a dataset consisting of 4,000 crowdfunding projects to discover hidden trends (campaign performance based on launch dates and funding goals) through pivot tables and pivot charts.

- Tools: Microsoft Excel

Relevant Experience

Success Academy	Bronx, NY
Data Analyst	June 2021 - November 2021
<ul style="list-style-type: none">• 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
<ul style="list-style-type: none">• 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 Microsoft Excel, R and STATA for the report.	
Innovative Development Strategies (IDS)	Lahore, Pakistan
Data Analytics Intern	June 2014 - August 2014
<ul style="list-style-type: none">• Conducted data collection (questionnaires, interviews, and focus groups) in 7 localities, followed by data cleaning, 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.	

Additional Experience

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

Education

Certificate, Data Analytics and Visualization	
Columbia University, New York	December 2021 - June 2022
Masters of Arts, Anthropology	
University of Virginia, Charlottesville	August 2019 - December 2020
Bachelor's of Science (Honors), Economics & Politics	
Lahore University of Management Sciences, Lahore	
Minor in Mathematics (Applied Statistics)	September 2011 - July 2017