Shahan Shahid Nawaz (shahan.shahid@berkeley.edu)

Portfolio: shahanshahidnawaz.github.io

SUMMARY

I am a Master of Information Management from UC Berkeley for which I studied Data Science, Data Engineering, Machine Learning, Deep Learning, Computer Vision, and Data Visualization. I have extensive research, algorithmic problem-solving experience, and applied science experience. I am looking for my first data-adjacent full-time role in the United States, and hold a F-1 STEM OPT valid till May 2026.

SELECTED EDUCATION

University of California, Berkeley – School of Information (Berkeley, CA)   08/21-05/23
Master of Information Management and Systems – Applied Data Science Certificate

Lahore University of Management Systems (LUMS) (Lahore, Pakistan)   08/12-06/16
Bachelor of Science (Honours) – Major: Economics, Minor: Sociology & Anthropology

SELECTED WORK EXPERIENCE

Data Scientist & Researcher – Othering & Belonging Institute

Berkeley, CA   08/21-Present

- Analyzed, visualized, and presented (as maps and reports) the extent of single-family zoning across California using spatial (ArcGIS) and data analysis (R); restrictive zoning contributes to the state's housing shortfall of 7 million units
- Led the refresh of the "Inclusiveness Index," a global snapshot of inclusivity by building data pipelines, and analyzing data from 136 countries and 30 indicators to provide valuable insights to policymakers, and highlight global leaders (New Zealand, Sweden, Norway)

Data and Communications Strategy Summer Fellow – Breakwater Strategy

New York, NY   06/22-08/22

- Brought data visualizations from various consumer insights datasets (including Juul, Snap, Coinbase) using the Microsoft Office suite—including Excel and PowerPoint—in line with visualization best practices as in the latest research enhancing the decision-making potential
- Steered a culture of data visualization by leading multiple company-wide data visualization discussions, and authoring a data-viz focused organizational plan to differentiate Breakwater Strategy's in the industry
- Recognized for the quality of analytical input, my energy, enthusiasm, curiosity upon the Fellowship's conclusion

Research Data Analyst – Centre for Asian Philanthropy and Society

Hong Kong, China   09/18-09/20

- Led end-to-end data analysis as the sole data scientist performing data collection, analysis, and presentation in the form of visualizations and reports
- Championed data science best-practices such as reproducibility (transitioning from Excel to Stata), and team upskilling (leading trainings on statistics fundamentals and Stata) resulting in more reliable and sharper data-driven insights
- Led the refresh of the "Doing Good Index," that ranks 18 Asian economies on support for philanthropy by updating the methodology and data sources, presenting a more accurate understanding of the Asian philanthropic landscape
- Led the design of digital products including an interactive data dashboard, and a scrollytelly web platform for the Doing Good Index, a first for the Centre
- Conducted survey data analysis and in-depth interviews for "Business for Good," a study of social entrepreneurship in 6 Asian economies to chart the future of sustainable capitalism in the region
- Served as an effective collaborator to fit into a small but high-performing team demonstrating my collaborative muscle

Research Analyst > Senior Data Analyst – Centre for Economic Research in Pakistan

Lahore, Pakistan   05/15-08/18

- Spearheaded end-to-end data analysis for a pioneering academic Randomized Controlled Trial (RCT)—the global impact evaluation gold standard—in the development/behavioral economics domain leading preliminary research, survey design, data cleaning, analysis, reporting, and presentation

- Successfully executed the project with a sample of 63,000 homes in two Pakistani metropoles delivering services demanded by citizens and measuring change in citizen sentiment towards taxation
- Created a GIS database that facilitated the identification of treatment and control regions resulting in streamlined information management, clean attribution, and no treatment spillovers
- Conducted extensive algorithmic problem solving, namely: Stata and Python scripts to automate various processes, including webscraping, payroll analysis, and the generation of 63,000 citizen-facing information sheets
- Worked under eminent Development Economists, demonstrated expertise in data analysis, project management, and programming, and contributed to the success of the RCT

SELECTED DATA SCIENCE PROJECTS

<u>Visualizing the Housing Crisis</u> (Master's Final Project)   6/23
- Analyzed the housing crisis examining historic, social, economic, and policy factors using Richmond, CA as a case study
- Conducted extensive data analysis in Python (pandas) on the disparate impacts of unaffordable housing along race, and visualized this data in a bespoke, isotype-based data visualization created in Figma

<u>Analyzing the Extent of Single-Family Zoning in Monterey, CA</u>   7/23
- Led spatial and data analysis in GIS software and Python (geopandas) of single-family zoning in Monterey highlighting its extent, and noting cities most urgently in need of reform
- Used key statistics and linked data from the census to situate the housing supply constraining effects of single-family zoning in the state's context

<u>Predicting the Trajectory of COVID-19 in the United States using Machine Learning</u>   1/22
- Successfully predicted with Artificial Intelligence (AI), in Python (sci-kit), case counts and vaccinations with political affiliation, celebrity deaths, and time (1.76% Mean Absolute Percentage Error)
- Successfully implemented and evaluated techniques such as Logistic Regression, Principal Component Analysis (PCA), K-Fold Cross Validation, and Ridge Regression

<u>Adobe Analytics Challenge 2022</u> (Data Science Business Case Competition)   10/21
- Offered a holistic, three-pronged revenue maximization solution from an analysis (e.g., tracking customer journeys through funnels, evaluating search performance, and noting price inelasticity for products) of shopdisney.com user data
- Led the design of the presentation creating several bespoke visualizations with a deliberate eye towards color and style themed along Disney IP
- Learned and successfully conducted effective data analysis in Adobe Analytics, in the span of 24 hours

<u>How Does the *New York Times* Front Page Work?</u>   12/21
- Developed from end-to-end a Natural Language Processing (NLP) and statistical analysis project in Python (NLTK)
- Conducted original analysis that confirmed the lexical uniqueness of headlines, as described in existing literature

<u>Ensuring Player Selection Fairness for a Renowned International Cricket League worth US$300m</u>   01/21
- Developed end-to-end a statistical model in Python (pandas) to ensure teams contesting the Pakistan Super League have a fair chance to build their best squad; the system has been in use successfully since 2018