Sharazad Ali

Berkeley, CA 94705 | sia3@berkeley.edu | (571) 471-7081| www.linkedin.com/in/sharazad-ali/

EDUCATION

University of California, Berkeley, Berkeley, California,
Masters in Statistics May 2026

Rice University, Houston, Texas May 2025
Bachelor of Science in Statistics, Bachelor of Arts in Psychology, Minor in Data Science **GPA: 3.87 / 4.00**

Relevant Coursework: Machine Learning, Linear Regression, Data Science Tools and Models, Advanced Probability, Computational Thinking, Research Methods, Data Visualization, Market Models, Applied Time Series and Forecasting

Awards: Cum Laude (2025), 4x President Honor Roll, Rice Trustee Distinguished Scholarship (2021), UC Berkeley Opportunity Award (2025)

PROFESSIONAL EXPERIENCE

HP Inc., Remote | ***Data Engineering Co-op*** August 2024 –Present
- Designed and implemented a volume monitoring system using Databricks and Airflow to detect anomalies in data products across teams, ensuring real-time data integrity and reducing errors in cross-functional pipelines.
- Collaborated with team to successfully migrate all data materials from Amazon Redshift to Databricks Lakehouse.
- Optimized ETL workflows in Databricks by improving Spark configurations, reducing data processing time by 25%, and enhancing efficiency for enterprise analytics and reporting.

HP Inc., Vancouver, Washington | ***Data Engineering Intern*** May 2024 – August 2024
- Analyzed end-to-end ETL pipeline latency from printer to PSU using Python, AWS, SQL, and Apache Spark.
- Improved telemetry data processing efficiency by 30% by building an alert system to detect latency shifts from seasonal fluctuations and job scheduling in Databricks.
- Developed interactive Databricks and Power BI dashboards for real-time monitoring of big data pipelines for the DataOS team.

Procter & Gamble (P&G), Cincinnati, Ohio | ***IT Data Analytics & Insight Intern*** May 2023 – August 2023
- Engineered an advanced forecasting model resulting in a 20% enhancement in the accuracy of market estimate projections.
- Improved sales forecast by ~$1 billion, facilitating superior sufficiency planning and product consumption estimates.
- Developed novel Python script to analyze fair share of distribution opportunities in multicultural neighborhoods and activated a plan to go after a $22 million dollar sales opportunity.
- Utilized skills in data mining and analysis to conduct a comprehensive exploration of new data tables containing 1 million+ rows of demographic and sales data, providing actionable insights for optimizing sales among multicultural consumers.

PROJECTS

Rice Data to Knowledge (D2K) Lab: Predicting Epilepsy Patient Travel Trends January 2025 – May 2025
- Utilized Pandas, NumPy, and Spark for data aggregation of 2 million medical claims and conducted geospatial analytics with Geopandas and Shapely to uncover travel patterns for Livanova's VNS treatment product.
- Applied supervised machine learning models, such as random forest regression and support vector machines, using scikit-learn for statistical modeling to understand patient journeys and treatment accessibility.

Machine Learning Clustering Methods in Biomedicine August 2023 – December 2023
- Demonstrated clustering methods' efficacy in revealing patterns in unlabeled biomedical data through theoretical and real-world analysis of a Leukemia gene expression dataset, achieving an average accuracy improvement of 15% compared to baseline methods.
- Utilized R and Python to highlight the versatility of different clustering algorithms, including K-means, DBSCAN, and hierarchical clustering, with K-means exhibiting the highest precision (85%) in identifying distinct gene clusters.

LEADERSHIP

University of California, Berkeley, D-Lab | *Data Science Consultant & Fellow* August 2025 – Present
- Advise faculty, graduate students, and researchers on research design, data acquisition, and statistical/machine learning workflows in Python and R, streamlining data science methods for social science and humanities research.

University of California, Berkeley, Statistics Department | *Outreach Peer Ambassador* July 2025 – Present
- Represented the Statistics Master's Program at 7+ outreach events and conferences, collaborating with a team of 8 ambassadors.
- Mentored prospective students from underrepresented backgrounds, strengthening community engagement and inclusivity at Berkeley

Rice University, National Society of Black Engineers | *Vice President* April 2022 – August 2024
- Collaborated with corporate sponsors to secure funding and create agendas for club general body meetings consisting of 60+ members.
- Executed club participation of 60+ members in all NSBE related conferences, such as Nationals and the Fall Regional Conference

SKILLS

Technical Languages: Python, Java, SQL, R, SPSS, HTML, Spark, Scala, React, Node.js

Platforms: AWS, Databricks, Splunk, GitHub, Microsoft Azure, Power BI, Jira, Airflow