

ANSHUMAN SHARMA

Irvine, California | (949) 279-7105 | anshums3@uci.edu | [linkedin.com/in/anshumansharma23/](https://www.linkedin.com/in/anshumansharma23/) | [anshumansharma23.github.io](https://github.com/anshumansharma23)

OBJECTIVE

Data professional with an MS in Data Science and approximately 3 years of industry experience specializing in Data Engineering, Data Science and Business Intelligence. Proficient in Machine Learning, Statistics, and adept at building and integrating data solutions, including real-time analytics and reporting. Lean Six Sigma Green Belt certified, with a track record of driving efficiency and quality improvement initiatives.

EDUCATION

University of California Irvine, Donald Bren School of Information and Computer Science

Master of Science in Data Science

Honor: Program Ambassador MDS

Irvine, CA

GPA: 3.61/4.0, 2023

Amity University, Amity School of Engineering and Technology

Bachelor of Technology in Computer Science and Engineering

Leadership: Convener AmityWE (An Entrepreneurship Club), Hostel & Student Committee Head

Noida, India

GPA: 3.88/4.0, 2020

SKILLS

Programming: Python, R, SQL, Shell Scripting

Frameworks/ Packages: Apache Airflow, Spark, Data Warehousing (Redshift, Snowflake, SAP BW), Pandas, Numpy, Docker, Hadoop, Git, NoSQL, Apache Kafka, Cassandra, Tableau, Power BI, Looker, Look ML, AWS, Azure

Statistical Modeling: Linear, Logistic & Multivariate Regression, Hypothesis Testing, Chi-Squared tests, Bayesian Methods, Time Series

Other: Project Management (Agile), Stakeholder Communication, JIRA for Task Management, A/B Testing

EXPERIENCE

Dell Technologies | Round Rock, Texas

Jun 2023 - Aug 2023

Data Analyst Intern

- Conducted in-depth interviews exceeding 50 sessions to meticulously gather requirements, pinpointing gaps, and strategizing solutions to fortify data infrastructure.
- Identified gaps in data integration processes and implemented a robust solution, enabling more than 6 different data-providers to integrate into one centralized warehouse. This initiative streamlined data workflows, enhancing efficiency and reliability across the organization.
- Orchestrated the consolidation of insights from 9+ reports, resulting in significant annual time savings of 5,000+ hours. This transformative initiative contributed to a remarkable 12% boost in overall productivity.
- Promoted dashboard adoption, seamlessly integrating with the PAN ISG application to target 10,000+ hours in annual time savings for enhanced organizational efficiency.

Accenture | Gurugram, India

Oct 2020 - Jul 2022

Business Intelligence Analyst

- Led projects with Sales Executives & Data Engineers to define sales KPIs, identify trends, and present performance results using Tableau.
- Designed robust, scalable data pipelines in collaboration with the Lead Data Engineers, processing millions of records daily.
- Optimized data pipelines, resulting in a 20% reduction in data transfer delay for the North American region.
- Led a high-performing team of 5 to surpass SLA benchmarks, achieving an impressive 98% on-time delivery rate for critical BI reports.
- Utilized SAP BW techniques, such as data partitioning and selective deletion, to optimize system performance and enhance analytics workflows.

Samsung Research Institute (SRI-D) | Noida, India

Feb 2020 - Apr 2022

Data Engineer Intern

- Implemented Apache Kafka to capture real-time logs of tests executed by Robot Framework in Python, establishing an alert system to promptly identify failures. This effort resulted in a 80% reduction in system monitoring, streamlining QA processes.

PROJECTS

Internal ChatGPT | City of Irvine (Llama2, Python, Streamlit, Natural Language Processing, Machine Learning)

- Engineered an efficient ChatGPT system for the City of Irvine, optimizing query processing and user interactions with a vast dataset of over 1 million documents. Utilized Llama2, Python, and Streamlit for seamless functionality.

Interchange.com | UCI - Database Development for an E-Commerce website (PostgreSQL, ER - Modelling)

- Developed and optimized an ER-Model database for an E-Commerce website, enhancing performance, scalability, and security measures against unauthorized breach.

Airflow-Powered Data Pipeline Development (Apache Airflow, PostgreSQL)

- Deployed a comprehensive data pipeline using Apache Airflow, managing tasks from PostgreSQL table creation to advanced operations such as API availability checks, user data extraction, processing, and secure storage in the password database.

CERTIFICATIONS

• Lean Six Sigma Green Belt Certified.

• Microsoft Azure Data Fundamentals.

• Looker and Data Specialization.