

KARAN PRATAP SINGH

+1(469) 332-2869 • kxs220023@utdallas.edu • [LinkedIn](#) • [Github](#) • Dallas, Texas

EDUCATION

Master of Science (M.S.) - Business Analytics

Aug 2022 - May 2024

The University of Texas at Dallas, USA

Bachelor of Technology (B.Tech.) - Computer Science

Aug 2014 - Jun 2018

Dr. A. P. J. Abdul Kalam Technical University, India

SKILLS

Certifications:

AWS Solutions Architect Associate, Azure Data Engineer Associate, Pendo Product Analytics, Google Data Analytics.

Programming Languages:

Python, Java, R, SAS, STATA, Bash, SQL Server, Oracle, Mysql, Nosql.

Libraries:

Numpy, Pandas, Dplyr, Scikit-learn.

Cloud Platforms:

Azure Databricks, AWS, GCP, Snowflake.

Big Data Technologies:

Apache Spark, Hive, Airflow, Databricks, Hadoop, HDFS, Flume, Kafka, HBase.

Data Visualization /Analysis:

Power BI, Tableau, Alteryx MS Excel.

Version Control/ CI-CD:

Git, Jenkins, Docker.

Statistics & AML:

Hypothesis Testing, Logistic Regression, ANOVA, K-Means, KNN, Random Forest.

Data Mining:

Data Cleaning, Data Wrangling, Data Modelling, Data Exploration, Data Analysis, Data Visualization, Data Interpretation, Data Presentation.

EXPERIENCE

Software Engineer Intern

Mar 2024 - May 2024

Round The Clock Technologies

Dallas

- Streamlined and managed end-to-end CI/CD pipelines, ensuring seamless software delivery and enhancing code quality for the global retail industry leader, leading to a **40% reduction in deployment errors** and a **25% increase in deployment frequency**.
- Devised the architecture and database for the RRO project that **improved query response time by 50%**.
- Boosted Outbound and Inbound optimization engines by **5% using NoSQL (Cosmos) databases** and built RESTful APIs to address the intricate supply chain issues, **resulting in 2% improvement**, in close collaboration with data science and product teams.

Senior Software Engineer

Apr 2021 - Jul 2022

Qualitest

Noida

- Established a batch processing pipeline using **PySpark in Azure Databricks** to identify the replenishment data based on the purchase order transactions across all the channels and carried out root cause analysis, **resulting in a 30% improvement in real-time insights**.
- Designed a self-serve platform using Python to perform data ingestion of 8 TB into the Snowflake database. Utilized Snow SQL to visualize analytical data across all platforms, improving data transparency, and application design and **reducing troubleshooting time by 8%**.
- Designed a comprehensive solution architecture utilizing Azure Databricks' delta tables and matched the ingested data with the redesigned ML models, resulting in a 25% increase in data processing efficiency.
- Established and tested ETL data pipeline for data retrieval to extract sales data from Hive and send to target vendor for UPC forecast sales analysis, **achieving a 20% increase in speed of data analysis**.

Software Engineer

Jun 2018 - Apr 2021

QA Infotech

Noida

- Employed PySpark and SparkSQL within Apache Spark on Amazon EMR to execute data transformations based on STMs, orchestrated via Apache Airflow to automate workflow sequencing, adding business value and reducing data processing time by 30%.
- Optimized a data pipeline solution design by migrating from Hive to Apache Spark and accomplished debugging of migration algorithms **resulting in 25% faster processing** and improving data architecture.

- Implemented ETL pipeline leveraging AWS Lambda, AWS Glue, and Amazon Athena to analyze partnered credit sales trends and customer segmentation, leading to a 30% reduction in query response times for decision-making.
- Implemented and automated data pipelines and data stores in Snowflake on service-oriented use cases, created test plans, and performed unit testing and code review, increasing performance by 15%, and **reducing ETL maintenance efforts by 20%**.
- Delivered ad-hoc financial reports and dashboards in Tableau with actionable insights, and improved cost-effectiveness by 40%.

Data Analyst Intern

Jun 2017 - Nov 2017

Nirwani Technologies Pvt. Ltd., India

Delhi

- Conducted comprehensive analysis of customer experience key metrics trends, providing senior leadership and product managers with actionable insights to drive product development, process enhancements, infrastructure requirements, root cause analysis, and integrated marketing and monetization strategies
- Developed automated notification pipelines to ensure the smooth operation of data warehouses.
- Performed data wrangling and ensured data quality controls by validating data, leading to a 15% reduction in errors and improving accuracy by 10%.

PROJECTS

Conagra Hackathon (FMCG Data)

Feb 2023 - May 2023

- Led analysis of present FMCG market trends, providing growth recommendations for Conagra Brands. Identified opportunities in this segment using Python and PowerBI and revealed a previously **untapped market of \$67 million**, enhancing the product portfolio, and recommending effective cost-effective merchandising strategies.
- The analysis prompted the introduction of an additional product for substantial weekly sales enhancement.

Analyzing Powerlifting Dataset

Sep 2022 - Dec 2022

- Used data analysis techniques and assessed the impact of different lifts on powerlifting competition outcomes. The project included **machine learning models** committed to synthesizing data.
- Used Data analytics methods including EDA, decision trees, neural networks, PCA, and **regression analysis** in the project.

Truck Driver Hazard Identification

Aug 2023 - Dec 2023

- Conducted comprehensive analysis of truck fleet data using **Hive**, resulting in precise transportation forecasts and extraction of actionable insights to optimize operations.
- Employed **Tableau** for advanced data visualization, enabling identification of patterns and application of statistical innovation methods to identify high-risk truck drivers, significantly improving road safety measures.

ACHIEVEMENTS AND EXTRACURRICULAR ACTIVITIES

- Spearheaded strategic initiatives as **President** of [The Product League at UTD](#), propelling the organization to achieve a remarkable 3X growth in membership. Directed the coordination and execution of the university's **largest panel discussion** event centered on Product Management, securing high-profile industry speakers and ensuring flawless day-of execution.
- Served as a **Mentor** of **Data Science Club At UTD**, helping students develop essential skills in **exploratory data analysis** and solving queries, thus enabling them to gain a deeper understanding of datasets and extract valuable insights.
- Nominated for the **Global Leadership Award** for Exceptional Leadership in an E-commerce Project by **Qualitest**, Recognized **Amongst 500+ Employees** for leading the software engineering team, fostering collaboration, and surpassing project expectations through effective communication and **project management**.