

KARAN PRATAP SINGH

+1(469) 332-2869 • karan.corporate25@gmail.com • [LinkedIn](#) • [Github](#) • Dallas, TX

EDUCATION

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

Aug 2022 - May 2024

The University of Texas at Dallas, Texas, TX

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

Aug 2014 - Jun 2018

Dr. A. P. J. Abdul Kalam Technical University, Ghaziabad, U.P

SKILLS

Programming Languages:

SQL, Python, R, SAS, SSIS, Linux, SQL Server, Oracle, Mysql, Nosql.

Libraries:

Numpy, Pandas, Dplyr, Scikit-learn

Cloud Platforms:

AWS S3, EC2, Redshift, QuickSight, Lambda, AWS Athena, ADLS, GCP, Azure Data Factory, Azure Databricks, Snowflake

Testing:

RPA, Pytest, Selenium, Postman, REST Assured, SoapUI

Big Data Technologies:

Hive, Hadoop, HDFS, Splunk, Apache Spark, Apache Flume, Apache Kafka, HBase.

Data Visualization /Analysis:

Alteryx, Power BI, Visio, Cognos, Tableau, MS Excel.

Version Control:

Git, Jenkins, Docker.

Statistics & AML:

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

Frameworks:

Data Warehousing, Data Modeling, REST API, CI/CD, Agile Methodology.

Certifications:

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

EXPERIENCE

Senior Software Testing Engineer

Apr 2021 - Jul 2022

Qualitest

Noida, India

- Validated SQL queries for data integrity, performed CRUD operations, and verified database schema against defined specifications, adding value for WM leadership and reducing data processing time by 30%.
- Led Manual, Ad-Hoc, and Exploratory Tests, expanding testing coverage and identifying Store inventories and used RPA UiPath to automate the manual Global Logistic System Application, improving the speed by 20
- Tested ETL end-to-end data pipeline for data transformation to extract forecast data from S3 buckets and send to target vendor for UPC forecasted products analysis, achieving a 20% increase in speed of data analysis.
- Integrated Jenkins CI for continuous integration, automating QA job executions, and ensuring consistent and reliable test automation processes.

Software Engineer

Jun 2018 - May 2021

QA Infotech

Noida, India

- Established a batch processing pipeline using PySpark to identify the replenishment data based on the purchase order transactions across all the channels and performed root cause analysis for management reporting, resulting in a 30% improvement in real-time insights.
- Coordinated QA estimates, test plans, and regression suites for Canada, Mexico, and Chile's WM Distribution & Fulfillment Centers. Managed production support, and UAT assistance for new feature implementations, ensuring smooth go-live activities and system integrity, reducing troubleshooting time by 8%.
- Conducted quality assurance testing on the ETL pipeline and evaluated the implementation of a new feature, transitioning from Putty (4GL) to GLS, utilized in Warehouse Management Systems (WMS) for enhanced warehouse management.
- Implemented Log4J logging in Python to capture and analyze test execution logs, facilitating effective debugging and troubleshooting during test automation, decreasing manual intervention by 12%.
- Executed API testing with POSTMAN tool and Rest Assured Library. Developed scripts for RESTful API Testing using Rest Assured Framework, facilitating Regression and Smoke testing.

Data Analyst Intern

Jun 2017 - Nov 2017

Nirwani Technologies Pvt. Ltd., India

Delhi, India

- Created SQL queries and provided data solutions to the R&D team from customer usage reports, improving data quality by 15% and enhancing team collaboration.
- Automated failure handling with a scripting language like Python and created scripts for data flow and data collection decreasing manual intervention by 12%. Designed and developed a comprehensive solution architecture utilizing Azure Databricks and matched the ingested data with the Machine learning models, resulting in a 25% increase in data processing efficiency.

- Performed data wrangling and ensured data quality controls by validating data and using advanced problem-solving analytical skills and technical skills, leading to a 15% reduction in errors and improving accuracy by 10%.

PROJECTS

- **Conagra Hackathon (FMCG Data):** Led analysis of present FMCG market trends, providing growth recommendations for Conagra Brands. Identified opportunities in this segment using XGBoost and Random Forest , revealed a previously untapped market of \$67 million, enhanced the product portfolio, and recommended effective cost-effective merchandising strategies. The analysis prompted the introduction of an additional product for substantial weekly sales enhancement
- **Stock Market Analysis** Implemented a full-stack real-time stock market analytics pipeline with stream ingestion with Alpha Vantage API, leveraging Apache Flume, Python, HDFS, and Spark RDDs for 5 stocks. (
- **Analyzing Powerlifting Dataset:** Utilized data analysis to evaluate lifts' impact on powerlifting outcomes, deploying ML models: linear regression, decision trees, random forests, SVMs, and neural networks, for insights. Showcased expertise in data analytics, stats, RDBMS, and R programming.
- **Cab Fares Data Analysis :** Examined the data for Uber and Lyft rides in Boston to understand what factors affect cab prices using statistical modeling, business intelligence, and data quality management.
- **Truck Driver Hazard Identification :** Conducted comprehensive analysis of truck fleet data using Impala, resulting in precise transportation forecasts and the extraction of actionable insights to optimize operations. Employed Tableau for advanced data visualization, enabling the identification of patterns and the application of statistical innovation methods to identify high-risk truck drivers, thus significantly improving road safety measures