

Karan Jadhav

Data Analyst

📍 Cork
☎ +353894876066
✉ karanjdvh44@gmail.com
in [linkedin.com/in/karan-jadhav-27b78b100](https://www.linkedin.com/in/karan-jadhav-27b78b100)
🐙 github.com/KaranLM10/Data_Science

Education

Masters Data Science & Analytics

2019 - 2020

Cork Institute Of Technology

Computational Analytics, Explanatory Data Analysis, Advanced Statistics, Data Acquisition, Relational Database, Distributed Computing, Machine Learning, Time Series Analysis, Factor Analysis, Deep Learning, Python, R, Data Visualization.

Grade: 1

Masters in Computer Applications

2014 - 2017

Maharashtra Institute Of Technology

Data Structure & Algorithms, DBMS, Operating System, Mobile Technologies, OOAD, OOPS, Python, Data Warehousing, Linux, Organizational Behavior

Grade: 1

Bachelor Of Science(Information Technology)

2011 - 2014

K M Agarwal College

HTML, PHP, Unix, C, C++, Information Systems, Computer Organization, Maths & Modelling, Core Java, Advance Java, SQL

Grade: 2.1

Experience Summary

- Overall Experience of 2.6 years working with **PRGX India Pvt. Ltd** (Jan 2017-July 2019) as **Data Analyst**.
- Extensive **Talend Hadoop and Talend SQL Development** experience more than 2 years with SQL, Hadoop, and Oracle in the Retail industry with a strong understanding of data & analytics.
- Developed ETL program to Extract, Transform, and Load into HDFS, Oracle and SQL databases, mappings development, performance tuning, debugging, migration, deployment, and scheduling job.
- Sound knowledge of **Data Warehousing Concepts, Dimensional modeling, HDFS**.
- Strong data visualization with the implementation of data modeling techniques using **Power BI, Tableau**.
- Expert in data analysis, solving issues with files such as delimiter issues in ASCII files using **Python, R, Java**.
- Experience in handling of source files such as delimited files, ASCII fixed files, Multi-schema files, Excel Files, and Oracle Dump files.
- Good exposure to analyzing the problem, team handling, decision-making, and assisting team lead. Excellent Verbal & Written Communication skills and strong in Documentation and Presentation.

Relevant Work experience

Data Analyst

Jan 2017 - July 2019

[PRGX India Pvt Ltd](#)

PROJECT: Period: Jan - 2017 to July - 2019

Domain: BIG DATA AND ANALYTICS

Project Name: HiPer Development

Clients: Sainsbury, Boots, Tesco, Dixons and Waitrose

Team Size: 8

Project Description :

Talend Job Development(SQL & Hadoop)

- Data Ingestion: Load various types of client data in the HDFS environment using Talend. (ASCII- delimited files, CSV files, excel files, EBCDIC files, report files (Monarch files), ASCII- Fixed files, Oracle dump files, etc.

- Pre - data push process. (All files receive are checked, Communication with counterparts, Create Mappings, Create Talend Jobs, File Fix Code development in Python/Java and Talend Code fix - for existing jobs).
- Post - data push process. (Testing and Quality Check of loaded data in HDFS & SQL, Communication with counterparts). Implement Quality Checks.
- Maintain MDR (Master Data Repository)
- Creating dashboards, reports about the client-related data using Tableau.

Data Migration

- Design and create data marts in Hadoop for the purpose of recovery Audit by applying data transformation principles based on Audit director's requirements. (POS, AP, Vendor Master, DSD, etc. -- Procure to pay Cycle).
- Convert challenging requirements that contain a lot of Joins, Complex calculations, Complex audit conditions to hive code on the Development cluster. Coding of requirements in the hive from scratch.
- Created UDFs for data transformation.
- Convert existing MS SQL Server environment build requirement to HIVE.
- Migration of code to the production environment for the execution of yearly/Quarterly/monthly builds.
- Deliver data to the SQL server using SQOOP.
- Automated Oracle import data pipeline using Automation Anywhere & Python. Helped in reducing manual efforts. Won PRGX's Hack award for this automation.
- Developed SQL scripts and stored procedures for sales and line item data.
- Created data models for Accounts Payable & Accounts Receivable SAP data to derive profit & loss incurred using Power Query.
- Created visual reports, dashboards to provide insights on reasons responsible for losses using PowerBI, Tableau.

Academic Projects

Built tuned model for regression & classification

- Dealt with data tidying (tidyverse , deplyr, mice), dummy creation, log transformation, NA removal, data imputation using R, Python.
- Performed exploratory data analysis for outliers & normality detection, the correlation between variables, Multi-co-linearity issues.
- Built different regression & classification models using hyperparameter tuning & cross-validation methodology to avoid miss-classification issues.
- Predicted on the test data using the efficient tuned model which was created using various machine learning algorithms such as Naive Bayes, decision tree, random forest, SVM, Bagging and Boosting, Adaboost, XgBoost, Linear regression, Logistic regression.

Understanding multi co-linearity issues with predictor variables in the regression model

- Analyzed how co-related variables affect the performance of the model.
- Investigated VIF & AIC approach reliability in solving the problem.

Code for data analyses and reporting

- Developed code focusing on structured, unstructured & image data analysis in Python.
- Structured data analysis focused on creating visual reports using different graphs to give insights into the data.
- Unstructured data analysis involved tokenization, lemmatization, stemming, count vectorizer, TF-IDF, pickle, cross-validation & interpreting classification report.
- Image data analysis involved loading, resizing, normalization, flattening, dimensionality reduction.

Aspect based Sentiment Analysis to differentiate customer feedback

- Unstructured text analysis to identify feedback.
- Results helped in understanding customer feedback quickly & efficiently.

Time Series Analysis

- Performed trend & season analysis to forecast the data using appropriate models.

Technical Skills

ETL, Hadoop, HDFS, Python, R, PowerBI, Tableau, SQL, HiveQL, Oracle, Data Analysis, Data Integration, Data Reporting, Java, C++, C#, Statistics, Machine Learning, Unix, Linux, Azure, Kibana, Kubernetes, Minio, Airflow, JIRA, Agile, Scrum, MS Office, SAS, Git

ACHIEVEMENTS & CERTIFICATES

- Won PRGX India Pvt Ltd's Hackathon(2019) for implementing Robotic Process Automation(RPA).
- Won Quarterly Award at PRGX India Pvt Ltd for projecting Quality & Excellence.
- Tableau Data Analyst Certification - June 2020