

PROFESSIONAL SUMMARY

- Masters in Data Analytics with major in Applied Data Science and Bachelors in Computer Science.
- Seasoned data practitioner with over 6 years experience in Data Management, Statistical Modeling, A/B Testing, Hypothesis testing, Machine Learning and Natural Language Processing.
- Highly motivated team player and proven working experience with cross-functional teams; Demonstrated Communication, Organizational, Analytical and Presentation Skills with business teams.
- Facilitated development of Machine learning and statistical models over complied datasets improving business outcomes and ensured accuracy of models.
- Extensive experience in conducting Market Research, Feasibility studies, Data Analysis, Data Mapping, Data Profiling and Gap analysis for Ad hoc analysis requests.
- Good understanding of Spark Architecture with Databricks; Hands on experience in developing Spark applications for Machine learning models and Big-Data models using PySpark and Spark SQL.
- Experienced in using stored procedures, Triggers, Views, User defined Functions, common table expressions (CTE) for SQL Query Optimization; Developed joins and Sub-Queries to simplify complex queries involving multiples tables.
- Experienced in designing, developing, scheduling reports/dashboards using Tableau and Power BI; Certified Power BI Data Analyst Associate in writing DAX expressions for complex Data Analysis.
- Created and maintained streamlined data flow into Power BI dashboards using Gateway connections to SQL Server, R and Python.
- Hands-on experience in troubleshooting, Performance tuning of Power BI dashboards and resolving issues within Power BI Report Server and dashboards.
- Efficient programming experience in Python and R, with ETL development experience integrating Data from Digital platforms like Google and Facebook Analytics, Shopify Omnichannel to SQL Server.
- Designed and developed A/B testing versions for commercial website and created a Data mining model over customer web footprint activity to support designing sale strategies, Validated observations over Cohort analysis statistics from digital analytics platforms.
- Experienced in Digital Marketing platforms like Google Analytics, Facebook Ads and Shopify.
- Experienced in Business requirements gathering, Data Extraction, Data Transformation and Data Loading (ETL Data pipelines), Business process analysis and Data Modeling, Data Mining and Data warehousing in SQL Server and Oracle environment.

TECHNICAL SKILLS

Programming Languages:	Python, R, SQL
ML Packages:	Pandas, NumPy, NLTK, SpaCy, Scikit-learn, TensorFlow, Keras, Gensim
Data Visualization Tools:	Power BI, Tableau, MS-Excel, Matplotlib, ggplot2, Seaborn
Data Engineering Tools:	Hadoop, Spark, Hive, Docker, Git
Databases & Cloud Services:	IBM Db2, SQL server, Oracle DB, MySQL, Azure Databricks, SSAS(OLAP Data cubes), Amazon S3, EC2 and SageMaker
Certifications:	Microsoft Certified Data Analyst Associate – Power BI

PROFESSIONAL EXPERIENCE

Process and Data Analyst

Sep 2021 – Mar 2023

The Kroger Co., Cincinnati, OH

- Maintained weekly reports in R and Excel and provided necessary code improvements periodically based on the feedback from stakeholders and business owners.

- Created and automated Power BI dashboards, reporting process completion rates and stores' opportunity analysis for over 2400 stores, with streamlining data connection through Gateway from SQL Server and R.
- Designed and developed Drill down and Drill through reports in Power BI and SAP Business Objects.
- Collaborated with business owner to design the reporting framework, and developed SQL and R scripts to integrate data from relational databases to perform Data quality analysis, and Transformation steps.
- Analyzed historical process implementation data to uncover insights on performance metrics and recommended optimistic process completion thresholds benefiting YoY savings by 7% (700 basis points).
- Worked on providing statistical support to identify Revenue and EBITDA benefits in development of business case for Fresh Production process.
- Analyzed the SQL and R scripts to create PySpark notebooks in Azured Databricks for faster performance and automated notebooks using jobs.
- Applied Machine learning framework in Databricks to segment similar stores based on stores' attributes.
- Collaborated with Merchandizing department to work on optimize code to boost reliability and achieved an 85% increase in report usage rate for space-to-sales reporting.
- Imported and managed multiple corporate applications into GitHub code management repository using Git.

Data Analyst Intern

Jan 2021 – Apr 2021

RBB communications, Miami, FL

- Reported Performance Indicators (KPIs) and ROI trends of 5 active digital campaigns in Google Analytics, Facebook Ads, Bing Ads and Shopify Omnichannel to senior leadership.
- Analyzed campaign's reach over budget to recommended scope opportunities and researched target demographics for each specific campaign.
- Created Excel Dashboard using VBA and macros to present weekly, quarterly, and yearly trends on Marketing KPIs.
- Integrated and compiled text data from Facebook post comments and Google reviews using APIs and Developed a Sentiment analyzer model to classify sentiment and compare over cost per acquisition metrics segmented audiences over campaigns.
- Trained the model over pre-existing sentiment labeled datasets and preprocessed the comments to remove noise and standardize text using SpaCy and NLTK.
- Extracted features from preprocessed data using TF-IDF and developed classification model using LSTM technique in TensorFlow resulting 0.9 F1-score compared to 0.6 F1-score of Random Forest model.

Data Analyst Intern

Sep 2020 – Dec 2020

FLKR LYTR, LLC, Boston, MA

- Created and automated an ETL pipeline to integrate digital campaigns' data from web-analytics platforms to MSSQL Server using APIs.
- Developed stored procedures in SQL Server to standardize DML transactions triggered over insert operations from automated scheduled refresh on daily basis to maintain the database up-to date.
- Supported Front-end team in developing and maintaining governance standards on A/B test versions for commercial website; Provided insights on tracking metrics measuring results on A/B test results.
- Applied association rule mining techniques in Python over customer web footprint and hover time for each product to predict customer behavior and develop buyer personas and uncover purchase patterns.
- Collaborated with Marketing team and presented my views from cohort analysis to research and strategize a marketing campaign resulting 50% growth in Quarter over Quarter revenue and observed 10% improvement in customer retention.

Analyst Programmer
RAMCO, Chennai, India

May 2018 – Aug 2019

- Collaborated with product managers to perform cohort analysis that identified an opportunity to reduce pricing for a section of users by 21% and enhanced personalized pricing based on frequency of ordering.
- Designed a model in pilot to increase incentives for drivers during peak hours, increasing driver availability by 15%.
- Cached ML-Platform (AWS SageMaker) Prediction API requests by using Redis and Amazon ElastiCache which cuts down the time by 80% for repeated requests.
- Triggered Auto Notifications to Gitlab and Outlook Mail from the ML-Platform (AWS SageMaker) after ML-Model data loading, training completion for tracking the pipelines with 95% notification sent accuracy.
- Experienced in building data ingestion pipelines using Amazon SageMaker to ingest structured and unstructured data like product performance metrics, reviews, and comments from Amazon S3 and EC2 instances.
- Developed 10+ interactive Tableau and Power BI dashboards and reports with advanced functions, LOD calculations, DAX expressions and dynamic visualizations.
- Creating Multiple Dashboards using SQL as a source and reducing the Performance Issues in the Tableau dashboards with Live connections with the Data Source.
- Responsible for translating Hive/SQL queries and converting Hadoop Big Data objects into Spark transformations using Spark RDD'S and PySpark.

Python Developer Intern
SenecaGlobal, Inc, Hyderabad, India

Jun 2017 – Apr 2018

- Developed and Automated the Employee Pay set Mapping for HCM by similarity-based approach with the historical data and 50% prediction rate, which helps to map the pay set of the employee in python.
- Created ETL Pipeline to connect Google Analytics and Facebook Ads to SQL Server Database using APIs; Designed SQL queries for ad hoc requests and optimized built-in codes using stored procedures, triggers.
- Analyzed daily transactions of food-delivery app to deliver insights to different levels of management with Drill down and Drill through reports in Power BI.
- Developed restaurant sales forecasting model specific to restaurants resulting improvement in relationship with clients and increased client acquisition by 5% by tracking mobile data usage.
- Supported in designing clustering model in identifying similar restaurants based on demand frequency and segmentation of order types.

ACADEMIC PROJECTS

Used Cars Price Prediction:

- Pre-processed vehicle data from Kaggle by implementing multicollinearity analysis using heatmap.
- Performed analytics operations to uncover descriptive and inferential insights over car attributes.
- Analyzed the feature importance related to used cars using Logistic Regression and SHAP; Developed a CatBoostRegressor model resulting 50% improvement in RMSE from the baseline model.

Used Cars Complaints Analysis:

- Implemented Word2vec over tokenized 5 million customer complaints registered in NHTSA, removed stop words, non-alphabetic characters using Gensim, Scikit-learn.
- Classified user complaints to pinpoint underlying issues; Applied Random Forest model, enhancing average F1-score from 0.6 to 0.8 after hyperparameter tuning compared with the F1-score of SVM model.
- Developed a Tableau Dashboard indicating Used cars' most generic complaints based on car attributes.

EDUCATION

M.S. in Data Analytics (Northeastern University, Boston, MA)

Jun 2021

B. Tech in Computer Science (IIIT Tiruchirappalli, India)

May 2018