

**TEJAS GALANDE**

Python-Data Scientist – Capgemini, Pune (https://www.capgemini.com/)

Phone: +91- 9890772287

Email: tejasgalande001@gmail.com

**Technology Specialization:**

Currently working with client **Marks & Spencer, United Kingdom** as Data Scientist for Master Enhanced Intelligent Reports.

**Objective**

To employ myself in a progressive organization that provides scope to update my practical knowledge and skills in accordance with the latest trends and be part of the team that dynamically works towards growth of the organization.

Looking for a software Python-Data Scientist job where I can utilize my programming skills to develop Data Science/ MachineLearning. To achieve a challenging position as a python programmer in a result-oriented company, where acquired skills and education will be utilized towards continuous growth and advancement.

**Technical Qualification Headlines:**

Currently working as a **’Python-Data Scientist’** for Master Enhanced Intelligent Reports with client Marks & Spencer, United Kingdom. Strong work experience with Technology ETL / Big Data / Business Intelligence / Analytical Language using PySpark.

**Experience with Core Data Science/ ML Libraries/ Packages:** PySpark | Pandas | Numpy | Seaborn | Matplotlib | Scikit Learn | Keras **|** TensorFlow | SciPy | Plotly

**Major Tools, Associated Technologies & Environment**

* Unstructured Database Used – Mongo DB
* Web Scraping Library – Beautiful Soup 4
* Web Framework involvement : Django
* BI Tool used - Tableau

**Experience with Machine Learning Algorithm:** Linear Regression | Logistic Regression | KNN | Decision Tree | Random Forest | K-Means | Naïve Bayes | Support Vector Machine | Principal Component Analysis | XG Boost

**Process worked with:** DevOps + Agile

**Software Programming Language Experience:** Python |SQL |Java |

**Business Implementation Programming Experience:** Machine learning – Python | ETL – Python | Business Intelligence

**Professional Work Experience**

**Data Science /Machine learning /ETL/BI Task**

* Highly efficient Data Scientist/Data Analyst with overall 2.9 years of experience in Data Analysis, Machine Learning, Data mining with large data sets of Structured and Unstructured data, Data Acquisition, Data Validation, Predictive modeling, Data Visualization.
* Overall 2.9 years of Experience in design, development, testing and implementation of various stand - alone and client-server architecture based enterprise application software in Python on different domains
* Experience in statistical programming languages with Python including Big Data technologies like Hadoop, Hive and Pyspark.
* Proficient in managing entire data science project life cycle and actively involved in all the phases of project life cycle including Data Acquisition, Data Cleaning, Data Engineering, Features Scaling, Features Engineering, Statistical Modeling (Decision Trees, Regression Models, Clustering), Dimensionality Reduction using Principal Component Analysis (PCA) and Factor Analysis, Testing and Validation using ROC plot, K - fold Cross-Validation and Data Visualization.
* Experience and deep understanding of Statistical modeling, Multivariate Analysis, Model Testing, Problem Analysis, Model Comparison and Validation.
* Expertise in transforming business requirements into Analytical Models, Building Models, Developing Data Mining and Reporting solutions that scale across a massive volume of structured and unstructured data.
* Skilled in performing Data Parsing, Data Manipulation and Data Preparation with methods including describe data contents, compute descriptive statistics of data, regex, split and combine, remap, merge, subset, reindex, melt and reshape.
* Experience in using various Libraries and Packages in python-like NLP, Pandas, NumPy, Seaborn, SciPy, Matplotlib, Sci-kit-learn, Beautiful Soup, Json, CSV, Highly skilled in using visualization tools like Tableau for creating dashboards.
* Extensive experience in Text Analytics, generating data visualizations using Python and creating dashboards using tools like Tableau.
* Hands on experience with big data tools like Hadoop, Spark, Hive, Pig, PySpark, Spark SQL. Hands on experience in implementing LDA, Naive Bayes and skilled in Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, neural networks, Principle Component Analysis.
* Experience in designing stunning visualizations using Tableau software and publishing and presenting dashboards, Storyline on web and desktop platforms.
* Experience and Technical proficiency in Designing, Data Modeling Online Applications, Solution Lead for Architecting Data Warehouse/Business Intelligence Applications.
* Experience with Data Analytics, Data Reporting, Ad-hoc Reporting, Graphs, Scales, PivotTables and OLAP reporting.
* Highly skilled in using Hadoop (pig and Hive) for basic analysis and extraction of data in the infrastructure to provide data summarization.
* Worked and extracted data from various database sources like Oracle, SQL Server, DB2, regularly accessing JIRA tool and other internal issue trackers for the Project development.

**Key Skills**

* Programming Languages: Python, HTML, SQL.
* Version Control: Git, GitHub.
* Cloud Computing: AWS EC2
* Analytic Tools: PySpark.
* Databases: Oracle 10g, My SQL, MongoDB.
* IDEs/ Development Tools: PyCharm, Anaconda – Jupyter Notebook, Microsoft Visual Studio.
* Frameworks: Django
* Operating Systems: Windows, Linux, Unix .

**Working Zone Organization**

Currently working as **Python-Data Scientist** with Capgemini, Pune (https://www.capgemini.com/) since May 2019 to till date.

**Achievements**

* Received “*Star Performer Award”* from Capgemini for good performance.
* Received appreciation for E2E Delivery of Master Enhanced Intelligent Reports from client Marks & Spencer, United Kingdom

**Projects**

**Project Sequence 1**

**Project Name** : Master Enhanced Intelligent Reports

**Vertical** : E- commerce, Payments, CRM

**Client** : Marks & Spencer, United Kingdom

**Technology & Tool:** Python, Pyspark, SQL, Pandas

**Roles** : Python - Data Scientist

**Detail Project Overview and Workflow:**

The overview provides a view of user overall performance. This includes the standard metrics like conversion, average buyer rate, revenue, and more. Report Analytics solution provides a visualization of consumer behavior on user website. It shows customer a horizontal funnel of what consumers are doing on user site from adding products to the shopping cart to completing a purchase.

The funnel can show different dimensions such as device category, browser, and country so user get a better understanding of how optimized customer funnel is at each stage of the process. The Checkout Behavior Report is a visualization report that zooms in on the checkout process.

Like the Shopping Behavior report, user can view how many users made it through each step of the process, and where the most drop-offs occur. Product Performance report provides a look into how user products are performing at a granular level. Metrics in this report include product revenue, average price, and unique revenue. There are some metrics, however, that are only available in the enhanced reports.

In the enhanced reports, the cart-to-detail rate shows the percentage of users who have added items to the cart against those viewed it on the site. The buy-to-detail rate is the percentage of users who purchased a product against those that viewed it. You can also filter the report by product category, product SKU, and the name of the brand.

Product List Performance report: This report presents a logical grouping of user products based on tags or categories. These groups can represent catalog pages, up-sell blocks, and search result pages. User can also optimize Customer product lists to determine which groups attract the most clicks, or generate the most revenue.

**Task Handled:**

* Implemented Data Exploration to analyze patterns and to select features using Python and SciPy.
* Built Factor Analysis and Cluster Analysis Models using Python and SciPy to classify customers into different Target Groups.
* Supported MapReduce Programs running on the cluster.
* Evaluated business requirements and prepared detailed specifications that follow project guidelines required to develop written programs.
* Participated in Data Acquisition with Data Engineer team to extract historical and real-time data by using Hadoop-MapReduce and HDFS.
* Performed Data Enrichment jobs to deal missing value, to normalize data, and to select features.
* Developed multiple MapReduce jobs in Java for Data Cleaning and Pre-processing, analyzed the partitioned and bucketed data and compute various metrics for reporting.
* Developed Hive queries for analysis, and exported the result set from Hive to MySQL using Sqoop after processing the data.
* Created HBase tables to store various data formats of data coming from different portfolios. Worked on improving performance of existing Pig and Hive Queries.
* Created reports and dashboards, by using D3.js and Tableau 9.x, to explain and communicate data insights, significant features, models scores and performance of new recommendation system to both technical and business teams.Utilize SQL, Excel and several Marketing/Web Analytics tools (Google Analytics, Bing Ads, AdWords, AdSense, Criteo, Smartly, SurveyMonkey, and Mailchimp) in order to complete business & marketing analysis and assessment.
* Used Git 2.x for version control with Data Engineer team and Data Scientists colleague. Used Agile methodology and SCRUM process for project developing.
* KT with the client to understand their various Data Management Systems and understanding the data, creating meta-data and data dictionary for the future data use/ data refresh of the same client.
* Structuring the Data Marts to store and organize the customer's data.
* Running SQL scripts, creating indexes, stored procedures for data analysis, Data Lineage methodology for Data Mapping and maintaining data quality.
* Prepared Scripts in Python and Shell for Automation of administration tasks.
* Maintained PL/SQL objects like packages, triggers, procedures etc. Mapping flow of trade cycle data from source to target and documenting the same.
* Performing QA on the data extracted, transformed and exported to excel.
* Participated in all phases of Data Mining; Data Collection, Data Cleaning, Developing Models, Validation, Visualization and performed Gap Analysis.
* Extracted data from HDFS and prepared data for Exploratory Analysis using data Munging.
* Built models using Statistical techniques like Bayesian HMM and Machine Learning Classification Models like XG Boost, SVM, Random Forest. etc.
* A highly immersive Data Science program involving Data Manipulation & Visualization, Web Scraping, Machine Learning, Python programming, SQL, GIT, Unix Commands, NoSQL, MongoDB, Hadoop.
* Used Pandas, Numpy, Seaborn, Scipy, Matplotlib, Scikit-learn in Python for developing various Machine Learning Algorithms.
* Worked on different data formats such as JSON, XML and performed machine learning algorithms in Python.

**Project Sequence 2**

Project Name : Digital Receipt Management

Vertical : Payments

Client : Braintree, Australia

Technology & Tool: Python, SQL, Pandas

Roles : Python - Data Scientist

**Detail Project Overview and Workflow:**

This Payment solution Accelerate Decision-Making, Assess, analyze and translate complex payment data sets into meaningful business insights through powerful visualizations. Payment Analytics refers to integrating and processing payments data from various sources like cards, mobile wallets, and bank transfers. If used efficiently, it can benefit businesses by providing insights into their revenues, payment trends, and customer shopping behavior. Improve user sales conversion rate and marketing ROI by analyzing and identifying patterns through customer payments data, such as customers’ preferred payment methods, high-performing stores or the payments value and volumes during normal and peak seasons. Analytics-driven strategies can lead to improved profitability by both cutting cost and optimize revenue in various contexts.Various Modules are Forecast transaction volume, Fraud and risk management, Merchant analytics

Merchant analytics simply means using data analysis to improve commercial operations and increase customer brand loyalty. Payment providers can monetize the data by helping merchants understand their market and create more personalized loyalty campaigns. Merchants can regularly evaluate whether their rewards programs are effective in increasing the number of visits and purchases. There’s a lot of insights that can be gained from a digital invoice and merchants can optimize their profitability by understanding where to look.

**Task Handled:**

* Implemented Data Exploration to analyze patterns and to select features using Python and SciPy.
* Built Factor Analysis and Cluster Analysis models using Python SciPy to classify customers into different target groups.
* Implemented Data Exploration to analyze patterns and to select features using Python and SciPy.
* Built Factor Analysis and Cluster Analysis models using Python and SciPy to classify customers into different target groups.
* Created HBase Tables to store various data formats of data coming from different portfolios.
* Provided summary statistics of key performance metrics and other measures deemed significant business units.
* Coordinated and managed Data Analytics activities with stakeholders
* Performed Multivariate Analysis, Predictive Modelling, Cluster, Market Basket Analysis using sophisticated statistical techniques.
* Interpreted & translated analytic output into insights.
* Determined the source of the data, coordinate extraction and acquisition.
* Assess data quality, identify gaps in the data and eliminate irrelevant data.

**Personal Details: -**

**C/O Name** : - Tejas Galande

**Current Address** : - Dahiwadi, Satara, 415508

**Date of Birth** : - 01/06/1997

**Gender** : - Male

**Marital Status** : - Single/Unmarried

**Nationality** : - India

**Languages** : - English, Hindi and Marathi.

(Tejas Galande)