

Data Scientist Data Scientist Data Scientist - L'Oreal USA Berkeley Heights, NJ Data Scientist/Data Analyst around 8+ years of Experience in Data Science and Analytics including Data Mining, Statistical Analysis with domain knowledge in Retail, Healthcare and Banking industries. Involved in Data Science project life cycle, including Data Cleaning, Data extraction, Visualization, with large data sets of structured and unstructured data, created ER diagrams and schema. Experience with Machine Learning algorithms such as logistic regression, KNN, SVM, random forest, neural network, linear regression, lasso regression and k-means. Good experience in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions, to various business problems and generating data visualizations using R, Python and Tableau. Experience in implementing data analysis with various analytic tools, such as Anaconda 4.0 Jupiter Notebook 4.X, R 3.0 (ggplot2, dplyr, Caret) and Excel Experienced the full software lifecycle in SDLC, Agile, DevOps and Scrum methodologies including creating requirements, test plans. Strong skills in statistical methodologies such as A/B test, experiment design, hypothesis test, ANOVA Good Knowledge and experience in deep learning algorithms such as Artificial Neural network (ANN), Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), LSTM and RNN based speech recognition using TensorFlow. Working Experience on Python 3.5/2.7 such as NumPy, SQLAlchemy, Beautiful soup, pickle, Pyside, Pymongo, SciPy, PyTables. Ability to write and optimize diverse SQL queries, working knowledge of RDBMS like SQL Server 2008, NoSQL databases like MongoDB 3.2 Experience in Big Data technologies like Spark 1.6, Spark SQL, PySpark, Hadoop 2.X, HDFS, Hive 1.X. Experience in Data Warehousing including Data Modeling, Data Architecture, Data Integration (ETL/ELT) and Business Intelligence. Good Experience in using various Python libraries (Beautiful Soup, NumPy, Scipy, matplotlib, python-twitter, Pandas, MySQL dB for database connectivity). Having experienced in Big Data technologies including Apache Spark, HDFS, Hive, MongoDB. Used the version control tools like Git2.X and build tools like Apache Maven/Ant. Worked on Machine Learning algorithms like Classification and Regression with KNN Model, Decision Tree Model, Na ve Bayes Model, Logistic Regression, SVM Model and Latent Factor Model. Experience and knowledge in provisioning

virtual clusters under AWS cloud which includes services like EC2, S3, and EMR. Good knowledge on Microsoft Azure. Knowledge and understanding of Devops(Dockers). Experience in writing Sub Queries, Stored Procedures, Triggers, Cursors, and Functions on MySQL and PostgreSQL database. Understanding of Python Best Practices (PEP-8) and package management system (PIP) in Python. Extensive experience in Data visualization tools like, Tableau 9.X, 10.X for creating dashboards. Experience in development and designing of ETL methodology for supporting data transformations and processing in a corporate-wide environment using Teradata, Mainframes, and UNIX Shell Scripting Used SQL Queries and Stored Procedures extensively in retrieving the contents from MySQL. Good in implementing SQL tuning techniques such as Join Indexes (JI), Aggregate Join Indexes (AJI's), Statistics and Table changes including Index. SQL loader for direct and parallel load of data from raw file to database tables. Experience in development of T-SQL, OLAP, PL/SQL, Stored Procedures, Triggers, Functions, Packages, performance tuning and optimization for business logic implementation. Strong SQL Server programming skills, with experience in working with functions, packages and triggers. Good at handling complex processes using SAS/ Base, SAS/ SQL, SAS/ STAT SAS/Graph, Merge, Join and Set statements, SAS/ ODS. Good industry knowledge, analytical & problem solving skills and ability to work well with in a team as well as an individual. Great team player and ability to work collaboratively and independently as required. Experience with Biological data sets (genomic, transcriptomic, microbiome, etc.,) Authorized to work in the US for any employer Work Experience Data Scientist L'Oreal USA - Berkeley Heights, NJ September 2017 to Present Python Description: L'Oreal Group is one of the world's largest cosmetics and Beauty company. Online Product Trading (B2B & B2C) brings dealers, distributors, manufacturers, fleet owners and traders from all around the world together and facilitates trade in an easy, secure, transparent and cost efficient way. Buyers can find sellers, and sellers can find the best possible buyers. This application can be used both by customers and manufacturers. Customer can browse and select their choice of L'Oreal Product or put their specifications if not getting the matching choice, so the manufacturers can make their choice of product and delivered to them by L'Oreal product Trader. Responsibilities:

Communicated and coordinated with other departments to gather business requirements.

Gathering all the data that is required from multiple data sources and creating datasets that will be used in analysis. Participated in the installation of SAS/EBI on Linux platform. worked on Data Modeling tools Erwin Data Modeler to design the data models. Designed tables and implemented the naming conventions for Logical and Physical Data Models in Erwin 7.0 Worked on development of data warehouse, data Lake and ETL systems using relational and non-relational tools like SQL, No SQL. Created SQL tables with referential integrity and developed queries using SQL, SQL*PLUS, and PL/SQL. Design, coding, unit testing of ETL package source marts and subject marts using Informatica ETL processes for Oracle database Developed various QlikView Data Models by extracting and using the data from various sources files, DB2, Excel, Flat Files and Big data Handled importing data from various data sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS Interaction with Business Analyst, SMEs, and other Data Architects to understand Business needs and functionality for various project solutions.

Identifying and executing process improvements, hands-on in various technologies such as Oracle, Informatica, business Objects. Worked on data cleaning and ensured data quality, consistency, integrity using Pandas, NumPy. Participated in feature engineering such as feature intersection generating, feature normalize and label encoding with Scikit-learn pre-processing. Improved fraud prediction performance by using random forest and gradient boosting for feature selection with Python Scikit-learn. Used Python (NumPy, Scipy, Pandas, Scikit-Learn, Seaborn), and Spark 2.0 (PySpark, MLlib) to develop variety of models and algorithms for analytic purposes. Utilized spark, Scala, Hadoop, HBase, Kafka, Spark Streaming, MLlib, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction etc. Implemented, tuned, and tested the model on AWS EC2 to get the best algorithm and parameters. Setup storage and data analysis tools in Amazon Web Services cloud computing infrastructure. Designed and developed machine learning models in Apache - Spark (MLlib). Used NLTK in Python for developing various machine learning algorithms. Implemented deep learning algorithms such as Artificial Neural network (ANN) and Recurrent Neural Network (RNN), tuned hyper-parameter and

improved models with Python packages TensorFlow. Installed and used Caffe Deep Learning Framework. Modified selected machine learning models with real-time data in Spark (PySpark).

Worked with architect to improve cloud Hadoop architecture as needed for Research. Worked on different formats such as JSON, XML and performed machine learning algorithms in Python. Participated in all phases of datamining; data collection, data cleaning, developing models, validation, visualization and performed Gap analysis. Worked very close with Data Architects and DBA team to implement data model changes in the database in all environments. Used Pandas library for statistical Analysis. Communicated the results with operations team for taking best decisions. Collected data needs and requirements by Interacting with the other departments.

Environment: Python 3.2/2.7, hive, oozie, Tableau, Informatica 9.0, HTML5, CSS, XML, MySQL, MS SQL Server 2008/2012, JavaScript, AWS, S3, EC2, Linux, Jupyter Notebook, RNN, ANN, Spark, Hadoop. Data Analyst CVS Health - Boston, MA September 2015 to August 2017 Description: Worked with pharmacy chain client. As the retail pharmacy, it sells prescription drugs and a wide assortment of general merchandise, including over-the-counter drugs, beauty products and cosmetics, film and photo finishing services, our goal of the project was to design, develop, and field data mining solutions that have direct impact to Patients and Janssen. Responsibilities: Investigated market sizing, competitive analysis and positioning for product feasibility. Conducted research on development and designing of sample methodologies, and analyzed data for pricing of client's products. Collaborated with database engineers to implement ETL process, wrote and optimized SQL queries to perform data extraction and merging from SQL server database. Worked on Business forecasting, segmentation analysis and Data mining. Developed Machine Learning algorithm to diagnose blood loss. Generated graphs and reports using ggplot2 package in R-Studio for analytical models. Developed and implemented R and Shiny application which showcases machine learning for business forecasting. Developed predictive models using Decision Tree, Random Forest and Na ve Bayes. Performed time series analysis using Tableau. Developed various workbooks in Tableau from multiple data sources. Created dashboards and visualizations using Tableau desktop. Later used Alteryx to blend the data. Performed analysis

using JMP. Perform validation on machine learning output from R. Written connectors to extract data from databases. Environment: R, Python 2.x, Excel 2010, Machine Learning, Tableau, Quick View, JMP, Segmentation analysis Data Scientist The Cellular Connection July 2014 to August 2015

Description: TCC is the largest Premium Wireless Retailer company. I work with the lead Data scientist to perform statistical analyses and predictive models on datasets to address various business problems in Benefits and Benefit operations through leveraging advanced statistical modeling, operations research, machine learning, or data mining techniques. Responsibilities: Involved in Data Profiling to learn about user behavior and merge data from multiple data sources. Participated in big data processing applications to collect, clean and normalization large volumes of open data using Hadoop ecosystems such as PIG, Hive, and HBase. Designed the prototype of the Data Mart and documented possible outcome from it for end-user Worked as Analyst to generate Data Models using Erwin and developed a relational database system. Designing and developing various machine learning frameworks using Python, R and MATLAB. Processed huge datasets (over billion data points, over 1 TB of datasets) for data association pairing and provided insights into meaningful data association and trends Participated in all phases of data collection, data cleaning, developing models, validation, and visualization and performed Gap analysis. Good knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Secondary Name Node, and MapReduce concepts. Handled importing data from various data sources, performed transformations using Hive, MapReduce, and loaded data into HDFS Collaborate with data engineers to implement ETL process, write and optimized SQL queries to perform data extraction from Cloud and merging from Oracle 12c. Collect unstructured data from MongoDB 3.3 and completed data aggregation. Conducted analysis of assessing customer consuming behaviors and discover the value of customers with RMF analysis; applied customer segmentation with clustering algorithms such as K-Means Clustering and Hierarchical Clustering. Participate in features engineering such as feature intersection generating, feature normalize and Label encoding with Scikit-learn preprocessing. Used pandas, NumPy, Seaborn, Scipy, Matplotlib, SKLearn and NLTK (Natural Language Toolkit), in Python for developing

various machine learning algorithms Utilized machine learning algorithms such as Decision Tree, linear regression, multivariate regression, Naive Bayes, Random Forests, K-means, & KNN. Parsing data, producing concise conclusions from raw data in a clean, well-structured and easily maintainable format. Determine customer satisfaction and help enhance customer experience using NLP. Developed various QlikView Data Models by extracting and using the data from various sources files, DB2, Excel, Flat Files and Big data Perform data integrity checks, data cleaning, exploratory analysis and feature engineer using R 3.4.0 Worked on different data formats such as JSON, XML and performed machine learning algorithms in R Worked on MapReduce/Spark Python modules for machine learning & predictive analytics in Hadoop. Perform data visualizations with Tableau 10 and generated dashboards to present the findings. Work on Text Analytics, Na ve Bayes, Sentiment analysis, creating word clouds, and retrieving data from Twitter and other social networking platforms Use Git2.6 to apply version control. Tracked changes in files and coordinated work on the files among multiple team members. Environment: Python 3.2/2.7, hive, Tableau, R, QlikView, MySQL, MS SQL Server 2008/2012, AWS, S3, EC2, Linux, Jupyter Notebook, RNN, ANN, Spark, Hadoop. Data Analyst- Python PIMCO - Manhattan, NY May 2012 to June 2014 Description: Pacific Investment Management Company is an American investment management firm provides mutual funds and other portfolio management and asset allocation solutions for millions of investors worldwide. I was involving in developing statistical models and algorithms to predict, classify, quantify, and/or forecast business metrics. Partner with business units and Analytics colleagues outside the workgroup to simulate and validate processes to maximize success metrics. Responsibilities: Used python libraries like Beautiful Soap, NumPy Created various types of data visualizations using Python and Tableau. Monitoring and tracking process performance using analytics tools like Tableau dashboard, R Utilized standard Python modules such as csv, robot parser, iterators and pickle for development. Created views in Tableau Desktop that were published to internal team for review and further data analysis and customization using filters and actions. Worked on Python OpenStack APIs and used NumPy for Numerical analysis. Used Python scripts to update content in the database and manipulate files. Used

Python creating graphics, data exchange and business logic implementation. Performed troubleshooting, fixed and deployed many Python bug fixes of the applications and involved in fine tuning of existing processes followed advance patterns and methodologies. Skilled in using collections in Python for manipulating and looping through different user defined objects. Installed numerous python packages using pip and easy install. Environment: Python 2.7, Tableau, R, Windows XP, UNIX, HTML, SQL server 2005 SQL Developer Johnson & Johnson - Bengaluru, Karnataka May 2011 to 2012 Responsibilities: Used DDL and DML for writing triggers, stored procedures, and data manipulation. Interacted with Team and Analysis, Design and Develop database using ER Diagram, involved in Design, Development and testing of the system Developed SQL Server Stored Procedures, Tuned SQL Queries (using Indexes) Created Views to facilitate easy user interface implementation and Triggers on them to facilitate consistent data entry into the database. Implemented Exceptional Handling. Worked on client requirement and wrote Complex SQL Queries to generate Crystal Reports. Created different Data sources and Datasets for the reports. Tuned and Optimized SQL Queries using Execution Plan and Profiler. Rebuilding Indexes and Tables as part of Performance Tuning Exercise. Involved in performing database Backup and Recovery. Documented end user requirements for SSRS Reports and database design. Education Bachelor's Skills Python 3.2/2.7, hive, oozie, Tableau, Informatica 9.0, HTML5, CSS, XML, MySQL, MS SQL Server 2008/2012, JavaScript, AWS, S3, EC2, Linux, Jupyter Notebook, RNN, ANN, Spark, Hadoop (8 years) Additional Information SKILLS MATRIX Languages C, C++, XML, R/R Studio, SAS Enterprise Guide, SAS, R, Python 2.x/3.x , Java, C, SQL, Shell Scripting NO SQL Databases Cassandra, HBase, MongoDB, Maria DB Statistics Hypothetical Testing, ANOVA, Confidence Intervals, Bayes Law, MLE, Fish Information, Principal Component Analysis (PCA), Cross-Validation, correlation. BI Tools Tableau, Tableau server, Tableau Reader, Splunk, SAP Business Objects, OBIEE, SAP Business Intelligence, QlikView, Amazon Redshift, or Azure Data Warehouse Algorithms Logistic regression, random forest, XG Boost, KNN, SVM, neural network rk, linear regression, lasso regression, k-means. Big Data Hadoop, HDFS, HIVE, PuTTY, Spark, Scala, Sqoop Reporting Tools MS Office

(Word/Excel/PowerPoint/ Visio/Outlook), Crystal Reports XI, SSRS, Cognos 7.0/6.0. Database Design Tools and Data Modeling MS Visio, ERWIN 4.5/4.0, Star Schema/Snowflake Schema modeling, Fact & Dimensions tables, physical & logical data modeling, Normalization and De-normalization techniques, Kimball & Inmon Methodologies

Name: Kelly Jones

Email: james88@example.net

Phone: +1-958-484-5994x30566