

JAYASHREE RAMAN

DATA | ANALYTICS | PRODUCT

Sammamish, WA – 98029

Email: jramanu@gmail.com

LinkedIn: www.linkedin.com/in/jayashree-raman

Mob: (206) 972-9037

Github: <https://github.com/jayashreeraman>

Tableau: <https://public.tableau.com/profile/jayashree.raman#!/>

PRIMARY SKILLS & AREAS OF FOCUS

Data Engineering, Data Platform Services & Analytics Lifecycle

- 5+ years' experience in Exploratory Data Analysis, Data Warehousing & Automated Deployment of data pipelines
- Expertise in building highly available data services, including automation, infrastructure-as-code, monitoring, and scaling
- Excellent communication skills to collaborate with cross functional teams and independently drive projects & decisions
- Strong skills in Data Modeling, Reporting and Dashboard Design for communicating insights to stakeholders at all levels
- Strong understanding and experience in applied statistics and machine learning and natural language processing
- Software engineering mindset and ability to write elegant, maintainable code, and follow engineering best practices
- Adept at data scraping, data wrangling and data cleaning with large-scale, complex data sets (using SQL, R & Python)
- 3+ years' experience in implementing data storage, query optimization and processing in Cloud Based Environments

SKILLS & TECHNICAL EXPERTISE

Programming Languages

- Python(pandas, numpy, scipy, sklearn, matplotlib, seaborn)
- R (dplyr, tidyr)
- SQL
- Java

Data Analysis & Visualization Tools

- D3/Tableau
- SQL Server/ SSIS
- PostgreSQL/MySQL
- Jupyter/Anaconda
- RStudio

Statistical Analysis & Machine Learning

- Linear Regression
- Gradient Descent
- SVM/PCA/kNN
- Natural Language Processing (NLP)

Cloud Computing

- Hadoop, Spark, Hive
- AWS: Amazon Redshift, S3, Kinesis, Lambda, AWS CDK
- Google Analytics

WORK EXPERIENCE

Data Engineer II

Sep 2020 – Present

Amazon LLC - Seattle, WA, USA

- Engineered cloud based data pipelines using AWS for a People Analytics Data Platform serving multiple teams, using a variety of AWS Services viz. Redshift, Kinesis, Lambda and AWS CDK
- Worked on re-engineering data pipelines to provide near real time stream processing instead of ad-hoc data refreshes
- Optimized Redshift data warehouse to minimize operational issues and reduce query processing times for team-owned datasets
- Collaborated and translated data requirements to build a scalable data engineering solution for tracking P0 metrics for products across the business organization, resulting in savings of 20+ hours of time savings per month, and advanced product usage insights
- Built 30+ data-intensive metrics and KPIs to monitor and give management visibility into Talent Movement and Talent Acquisition
- Onboarded structured datasets into Graph Data format to facilitate data science and machine learning initiatives
- Collaborated with engineers, product managers and data scientists to understand data needs, and build data pipelines to satisfy the varied use cases related to employee demographic, performance, talent acquisition and talent management
- Built data expertise and leveraged data controls and archived system tables to ensure privacy, security, compliance, data quality and operations for all data pipelines owned by the team

Sr. Data Engineer

Apr 2020 – Aug 2020

SAP Concur - Bellevue, WA, USA

- Create data pipelines and specifications for **Fact and Dimension tables for bringing in new datasets using Apache Spark**
- Developed success metrics and KPIS to help clients achieve specific outcomes w.r.t Spend Management
- Catalyst Fellowship: Guided Customer Help to improve Customer Search Experience – Performed statistical analysis & evaluated metrics for automated chatbot service developed for all SAP products, using **Python ML packages – scipy, numpy and pandas, and Apache Spark and Zeppelin**
- Catalyst Fellowship: Data Lake as a Service – helped build out the data lake service for Customer Search applications, focus on Product Parts Management Service (PPMS) application data – **using SQL, JSON schema, Python and SAP Datahub**
- Performed NLP on a free-text field in Salesforce to help structure the “Notes” field in Success Plan for better tracking
- Performed **data discovery and structuring** to bring in S & P Market Intelligence data from CapIQ for **benchmarking T&E spend**

Business Analyst

Apr 2019 – Mar 2020

SAP Concur - Bellevue, WA, USA

- Developed success metrics, benchmarks & KPIS to help clients measure & monitor spend and make recommendations using **SQL**
- Worked extensively with collating multiple data sources, performing Exploratory Data Analysis and defining a focused Solutions Strategy, **improving time efficiency for reporting by ~80% by automating the production of PPTs from PDFs**
- Performed proof-of-concept implementation for several machine learning initiatives targeted at identifying Invisible (Fraudulent) Spend and Improvement of Reporting Infrastructure
- Leveraged strong understanding of **operational data (O-data)** for aggregating and developing data relationships, thus combined large data sets and sources to generate new insights, **using SQL and Python**

Data Analytics Intern

June 2018 – Nov 2018

GE Ventures, Industrial Data Exchange - Seattle, WA, USA

- Arbitrated product requirements with Sales/Marketing teams to build a customized solution, through launch and two iterations
- Worked extensively with **Amazon Redshift and S3** for large scale data analysis to improve execution time by over 35%
- Wrote **R Scripts** to clean data & perform Exploratory Data Analysis for medical equipment in secondary marketplace
- Used **Tableau** to visualize distribution of medical equipment in secondary markets as well as their price distribution /prediction
- Performed statistical analysis/time series analysis and built a multiple linear regression model for price prediction

Senior Software Engineer

BNP PARIBAS ISPL, Mumbai, India

Nov 2015 - Aug 2017

Worked on the Primary and Secondary Markets applications for BNP Paribas

- Built a hybrid framework (keyword and data driven) to automate the testing for web application testing
- Independently developed and implemented Continuous Integration using Jenkins
- Worked extensively with SQL and Python to compare and validate data on front-end and back-end
- Expanded Test Coverage of application from 45% to 87%; showcased the effort to clients and higher management
- Analyzed trends of previous testing processes and reduced the project turnaround time by 40%
- Won the Catalyst Award for outstanding performance in Q1- 2017

Systems Engineer

TATA CONSULTANCY SERVICES, Mumbai, India

Dec 2012 – Oct 2015

- Created a hybrid framework for automated testing of Web/Mainframe applications for Citi Bank
- Analyzed external logs and created visualizations, effected savings of 60 hours/week in Operations Department by reusing automation scripts designed for testing
- Created parsers for XML files using utility functions developed in C#, thus saving almost 100 hours/month
- Won the Star Performer Award for the month of July 2015

EDUCATION

UNIVERSITY OF WASHINGTON, INFORMATION SCHOOL, Seattle, WA

June 2019, GPA: 3.89

Master of Science in Information Management - Specialization in Data Science & Business Intelligence

Relevant Coursework: Advanced Analytics Methods, Business Intelligence Systems, Population Health Informatics, Applied Data Science, Machine Learning, Big Data and Cloud Systems

UNIVERSITY OF MUMBAI, Mumbai, India

July 2012, GPA: 3.5

Bachelor of Engineering in Electronics & Telecommunications

Relevant Coursework: Data Structures and Algorithms, Wireless Networks, Software Programming
Consistent high performer, with a focus on practical implementation of technologies

COURSE & RESEARCH PROJECTS

- Worked with Professor Abhishek Borah (Foster School of Business) to analyze customer complaints/company response data for customer satisfaction; using Python and R to perform statistical analysis of effective response patterns
- Built a sentiment analysis project using NLP to determine trend of company stocks from their earnings calls held every quarter as part of the Capstone project
- Data Analysis of Population Health Data with R (wrangling with **dplyr**, visualization with **ggplot2**)