

# Parv Jain

San Jose, CA | 559-728-2755 | [parvjain2503@gmail.com](mailto:parvjain2503@gmail.com) | [linkedin.com/in/parv-2503](https://linkedin.com/in/parv-2503) | [github.com/Parvj26](https://github.com/Parvj26)

## EDUCATION

California State University, Fresno, CA

B.S Computer Science

**Aug 2016 – May 2021**

Relevant Coursework: **Big Data Analytics**, Web Programming, **Algorithm & Data Structure**, Database Systems, Operating Systems

## TECHNICAL SKILLS

- Programming Languages: **Python**, C/C++
- Tools: Git, Docker, JIRA, **Databricks**, **Tableau**
- Databases: **SQL**, DynamoDB, RDBMS
- Cloud Technologies: Amazon Web Services **AWS - S3, EC2, VPC**, Route53, **IAM**, ASG, CloudFront
- Web Technologies: HTML5, CSS3, JavaScript, React
- Frameworks: Amazon API Gateway, **Apache Spark**
- Other Technologies: Agile, Scrum

## PROFESSIONAL EXPERIENCE

**Web Developer** | California State University, Fresno, CA | [Website](#) | [Blog](#)

**March 2019 - Current**

- **Technologies:** **HTML5, CSS3, JavaScript, SQL**
- Developed a Web application to check out university vehicles by working collaboratively with senior members.
- Requirement gathering from end-users and interpreted the client specifications to the developers.
- Resolved bugs by testing multiple versions of the application.
- Enhanced Search Engine Optimization (SEO) for 20+ web pages to improve organic web traffic.
- Consistently updated existing content for 13+ sub-department web pages for more than 8,000 students.

**Software Developer Intern** | Studycity

**March 2019 - Current**

- Developed a DNS based productivity and web blocking/logging application in Python.
- Exposed to all aspects of software development life cycle (SDLC) such as Analysis, Planning, Developing, Testing, Implementing, and Post-production analysis of the project.
- Worked in a fast-paced environment, followed Agile software methodology for project development using JIRA.

## PROJECTS

**Data Analyst** | Airbnb Data Analysis: [GIT](#)

- **Technologies:** **Python, Databricks, Apache Spark, SQL, S3, EC2**
- Created clusters and workspace on Databricks, pulled dataset from AWS S3 bucket, cleaned the dataset, and created tables.
- Performed data transformation, data aggregation, and exploratory data analysis on the dataset.
- Sentiment analysis on the data using keyword extraction to check the properties with the best reviews.

**Data Science** | Telecom Churn Machine Learning

- **Technologies:** **Python, Databricks Apache Spark, SQL, Spark ML**
- Performed descriptive and exploratory data analysis on the dataset to find the relation between churn and the factors leading to it.
- Created train and test datasets to feed the machine learning pipeline. Used PySpark to create the ML pipeline.
- Performed logistic regression on the dataset and

**Data Analyst** | Covid-19 Data Analysis [GIT](#)

- **Technologies:** **Python**, Jupyter Notebook, Apache Spark, Hadoop
- ETL process of the data using SQL and Apache Spark.
- Performed prescriptive, descriptive, and predictive analysis on Covid-19 data using python.
- K-means clustering to find the relation between temperature and affected cases of Covid-19.

**AWS Devops** | Fault-Tolerant WordPress Website using AWS

- **Technologies:** AWS EC2, S3, CloudFront, ASG, RDS, VPC, Route53,
- Deployed a VPC to host a website with load balancing, Autoscaling using CloudFront, Route53, ELB, and ASG.
- Hosted the website on two EC2 Linux instances in different Availability zones using Apache and WordPress. Used Amazon RDS to deploy multi-AZ for disaster recovery.
- Used Security groups and NACLs to ensure security.

## CERTIFICATIONS

- AWS Solutions Architect - Udemy
- Data Analysis using PySpark Science Bootcamp - Udemy
- Analyzing Big Data with SQL – Coursera
- Introduction to Machine Learning - Kaggle