

# KARAN PILLAY

## DATA ANALYST



### SUMMARY

- Around 3 years of experience in Data Mining with large datasets of Structured and Unstructured Data, Data Acquisition, Data Validation, Predictive Modelling, and Data Visualization.
- Proficiency in all phases of requirements gathering, analysis, designing, coding, testing, integration, and documentation, and extensive experience in Agile and Waterfall Methodologies.
- Knowledge of Python and R programming with various packages including NumPy, Pandas, SciPy, and Matplotlib.
- Knowledge of AWS command line client and management console to interact with AWS resources and APIs.
- Experience with common Machine Learning Algorithms like Regression, Classification, NLP and Random Forest.
- Proficient in developing and manage SQL, Python and R code bases for data cleansing and data analysis using Git version control.



### EXPERIENCE

#### Data Analyst

Goldman Sachs Group, CA | Jan 2021 – Mar 2021

- Followed Agile Software Development Methodology to build the application iteratively and incrementally.
- Programmed a utility in Python that used multiple packages like SciPy, NumPy, Matplotlib and Seaborn.
- Developed predictive models using Python and R to predict customer churn and classification of customers.
- Created new EC2 instance in AWS, allocate volumes and giving Provisionals using IAM.
- Created action filters, parameters, and calculated sets for preparing dashboards and worksheets in Tableau.
- Used R Studio for data visualization and generated scatter plot and high low graph to identify relation between different variables.
- Design and build a text classification application using different text classification models.
- Used Python, Jupyter Notebook, PyCharm, and PostgreSQL for statistics and mathematics modeling.
- Developed Tableau data visualization using scatter plots, geographic maps, pie charts and bar charts, and density chart.

#### Data Analyst

Zensar Technologies, India | Jan 2017 – July 2019

- Created the Technical Specification document for the reports developed as part of the project life cycle and followed the SDLC using Waterfall.
- Generated various capacity planning reports (graphical) using Python packages like NumPy, Matplotlib, SciPy.
- Utilized Power BI to design multiple scorecards and dashboards to display information required by different departments and upper-level management.
- Assisted Supply chain analysts with automating reporting functionality using Power BI tools, Power BI reporting, Dashboards & Scorecards (KPI).
- Extensively used Python libraries like NumPy, Pandas, SciPy for data wrangling and analysis, while visualization libraries of Python like Seaborn, Matplotlib for graphs plotting.
- Involved in project cycle plan for the data warehouse, source data analysis, data extraction process, transformation and loading strategy designing.
- Used Git for version control with the Data Engineer team and Data Scientists.

### CONTACT

**Phone:** (901) 206-4539

**Email:** [karan.p@cvemail.com](mailto:karan.p@cvemail.com)

**Location:** CA

<http://www.karanpillay.com>

### EDUCATION

**Master's in data science,**

University of the Pacific, Stockton, CA

### SKILLS

**Methodologies:** Agile, Waterfall

**Programming Language:** Python, R, SQL

**IDEs:** Jupyter Notebook, PyCharm, R Studio

**Packages:** ggplot2, NumPy, Pandas, Matplotlib, SciPy, Seaborn

**ML:** Regression, Classification, NLP, Random Forest

**Visualization Tools:** Tableau, Power BI

**Database:** MySQL, PostgreSQL

**Cloud:** AWS

**Other Tools:** Git, MS Office, MS access

**Operating System:** Windows, Linux

### CERTIFICATES

**Google Data Analytics**

**IBM DataScience**