Aakash Parwani
Data Scientist
Email- aakashparwani@gmail.com
Phone – (551)-225-9572

PROFESSIONAL SUMMARY

A former **Database Developer** with newly acquired skills in **data analytics and visualization tools**, an insatiable intellectual curiosity, and the ability to mine hidden gems located within large sets of structured, semi-structured and unstructured data. Leverages wide range of **statistical and machine learning methodologie**s to obtain project requirements and implement solutions that drive bottom line. Recognized for providing **outstanding support in data life cycle development**. Quick problem solver while dealing with new concepts, systems, and procedures. Recognized for accuracy, quality of work and a **strong work ethic**.

- Experience in using **SSIS, CTE, temp tables and effective DDL/DML Triggers** to facilitate efficient data manipulation.
- Experience with **Google Analytics**, **Tableau** and **SQL Server Reporting Services** (SSRS).
- Strong experience in **Software Development Life Cycle (SDLC)** including Requirements, Specifications Analysis/Design and Testing as per the Software Development Life Cycle.
- Extensive experience in writing **functional specifications**, translating business requirements to technical specifications, **created/maintained/modified database design** document with detailed description of logical entities and physical tables.
- Expertise in writing **T-SQL Queries, Dynamic-queries, sub-queries and complex joins**.
- Excellent knowledge in creating **Databases**, **Tables**, **Stored Procedure**, **DDL/DML Triggers**, **Views**, User defined data types, effective functions, Cursors and Indexes.
- Able to **communicate** effectively with **multifunctional teams**, programmers and technical staff at all levels. Strong customer service orientation.
- Participated in requirements analysis reviews and Design sessions to understand the requirements and designing **Reporting Solutions**.
- Experience in transforming raw data into actionable **strategic knowledge** in order to gain insight into business processes, and thereby guide and help businesses in their decision-making and run efficiently.
- Skilled in providing analytic support including **data importing**, **data wrangling** and **data visualization**.
- Real-Time experience using PYTHON & R packages like **dplyr**, **tidyr**, **ggmap**, **caret**, **forecast, SnowballC**, **wordcloud**, **corrplot**, **rpart**, **reshape2**, **ggplot2, tm, Scikit-Learn and MatPlotLib**.
- Experience in designing stunning visualizations using **Tableau** software and publishing and presenting dashboards on web and desktop platforms.
- Familiar with fundamentals of **ETL, SAS**, **Hadoop Architecture**, **HDFS Framework** and components of its ecosystem like **Map Reduce and HIVE**.
- Committed **Team player** with excellent communication skills and capable of working independently.

TECHNICAL SKILLS

Databases	**Expert:** Relational Database - SQL Server 2008, MySQL. **Novice:** DB2.
Programming	**Expert:** C++, Visual Basic, Python, SQL, HTML, XML. **Intermediate:** MATLAB, JavaScript, R, SAS. **Novice:** Django, Hive.
Operating System	**Expert:** Linux, Windows XP/7/8/10, MAC OS
Business Intelligence Tools	**Intermediate:** Tableau, Plotly, Google Refine, MS Excel -

	Analytical Solver, SSRS.
Statistics	**Intermediate:** Mean, Median, Mode, Data Distributions, Standard Deviation and Variance, Hypothesis Testing (p-values) and Test for significance (z-test, t-test and ANOVA).
Frameworks	**Intermediate:** Apache Hadoop, Hive, Zeppelin. **Novice:** Talend.
Machine Learning	**Intermediate:** Feature Selection, Supervised and Unsupervised Learning, Correlation Analysis, Linear Regression, Multiple Linear Regression, Logistics Regression, Clustering, Classification, Decision Tree, Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbors (KNN), Clustering, K-Means Clustering, Random Forest, Principal Component Analysis, Factor Analysis, Correlation Matrix, Time-Series Analysis.
R Packages	**Intermediate:** dplyr,caret, datatable, reshape, ggplot2, NLP, sqldf, ggmap, ggvis, dplyr, PerformanceAnalytics, randomForest, rpart, lm, glm,nnet, xgboost, ksvm, lda, qda, adabag, adaboost , lasso, tm.
Python Packages	**Expert:** scikit-learn, matplotLib, numpy, scipy, pandas, iPython, NLTK, Seaborn, Pattern, Statsmodels, ggplot, plotly, sqlite3.

PROFESSIONAL EXPERIENCE

Core Card Software, INC. **Bhopal, INDIA**
Role: Database Developer **July'11 – June'15**

Key Projects Handled

CNB (www.centralnational.com)
Developed application to allow college students perform transactions in prepaid domain.
FirstView (www.firstviewfinancial.com)
Designed business prepaid card solution to specifically assist client in streamlining employee expense management.
PexCard (www.pexcard.com)
Implemented procedures to detect poor database performance and fine tune production database.

Responsibilities

- Develop application to allow college students perform transactions in prepaid domain.
- Provide program management and processing services supporting a diverse array of prepaid card solutions.
- Understand business architecture and design business prepaid card solution to specifically assist companies in streamlining employee expense management.
- Interacting with the client for system study, requirements gathering and analysis.
- Understand application architecture then undertaking requirement and impact analysis.
- Determined the missing data, outlier and invalid data and applied appropriate data management techniques.
- Arranging test cases and doing Unit and String test. Going through review and maintaining the project document as per the company and client standard.
- Wrote simple and advanced **SQL queries** and **scripts** to create standard reports for senior managers.
- Handled database archiving, database replication, database partitioning and resolved deadlock issues in production environment.
- Handled **performance tuning** for production environment after examining the most expensive queries in the environment.

Environment: Python, Micorsoft SQL Server 2008, C++, Visual Basic, XML, DBBIDE, Asp.Net, WCF, Microsoft Excel, Microsoft Word, Microsoft PowerPoint.

ACADEMIC EXPERIENCE

Aarhus City Weather Data Analysis **Jersey City, US**
Role: Data Scientist **Jan'17 – May'17**

This project was a part of final semester **capstone** subject at saint peter's university. Arhus is Denmark's second-greatest city and the cash related concentration of the Central Denmark Region. The city has a catchment zone of 1.2 million people inside a one-hour travel go and is especially connected with Copenhagen and Hamburg. The purpose of this project was to apply statistics & modeling techniques on big data and perform review of how atmosphere data examination and representation attempt can help in building framework for sharp urban ranges to stimulate the introduction of smart city applications for atmosphere envisioning.

Responsibilities

- Used **Zeppelin** environment to perform analysis of weather data. Understood the importance of features like: **dewpoint, humidity, air pressure** for prediction of **air temperature.**
- Developed strategy to use ***Short Weather Forecasting*** model and outperform ***Long Range Forecasting***.
- **Analyzed**, **retrieved** and **aggregated data** from multiple client to perform **data mapping**, in order to precisely append either incorrectly mapped or missing data.
- Loaded the data file in **R** and utilized R packages (**Dplyr** & **Tidyr**) to perform **exploratory data analysis (EDA)** to gain an overall picture of the data.
- Performed **Kstest** to check whether environmental variables are normally distributed or not. If not then normalize the variables first.
- Performed **data manipulation** and prepared the training and testing sets for modeling.
- Used **Pearson correlation** to get an idea of relationship between dependent and independent variable and it also informs about weighting coefficient of independent variables.
- Created the visual summaries to understand the shape and distribution of the data using R package **ggplot2** and **plotly**.
- Performed time-series analysis (**AR, MA, ARMA, ARIMA**) and linear regression model (**OLS Regression**) to predict **air temperature** using machine learning packages in R.
- Ensured best practices are applied and **integrity of data** is maintained through security, documentation, and change management.

Environment: Zeppelin, R, R Studio, Data Warehousing, SQL, Microsoft Office, Tableau.

Project URL: Weather Data Analysis

House Prices: Advanced Regression Techniques **Jersey City, US**
Role: Data Scientist/Data Analyst **Jan'17 – May'17**

This project was a part of final semester **marketing analytics: kaggle competition** at saint peter's university. The objective of this project was to practice advanced regression techniques on house prices dataset. This data set was constructed for the purpose of an end of semester project for an undergraduate regression course. The original data (obtained directly from the Ames Assessor's Office) is used for tax assessment purposes but lends itself directly to the prediction of home selling prices. Data set is describing the sale of individual property in **Ames, Iowa** for year **2006 to 2010**.

Responsibilities

- Performed **data manipulation** and prepared the training and testing sets for modeling.
- Used **matplotlib** to explore the dataset and to check **skewness** in the data.
- Removed the **skewness** and performed **normalization** of data.
- Performed **PCA (principal component analysis)** to figure out independent variables having strong relation with target variable.
- Wrote simple and advanced **SQL queries** for extracting data and created **dashboard** and stories.
- Utilized libraries such as **xgboost & scikit-learn** to build statistical model such as **Ridge Regression, Lasso Regression, Xgboost Regression & Random Forest.**
- Trained the model on training dataset and evaluated **root mean square error & accuracy scores** to select the best statistical model.
- Performed prediction of **house prices** on testing dataset.

Environment: Python, DATABRICKS, Jupyter, Data Warehousing, SQL, Microsoft Office, Tableau, Plotly.

Project URL: House Price Prediction

Credit Card Fraud Detection Model **Jersey City, US**
Role: Data Engineer/Data Analyst **Jan'16 – May'16**

This project was a part of second semester **big data analytics** subject at saint peter's university. The purpose of this project was to practice data analysis & visualization techniques on transaction database and build a fraud miner application that can accurately predict an incoming credit card transaction as a fraud or a legitimate transaction.

Responsibilities

- Conducted team meetings & ensured project documentation is completed.
- Responsible for **data aggregation**, data pre-processing, missing value imputation and descriptive and inferential analysis.
- Ensured all data has a complete and accurate definition.
- Performed **data manipulation** and prepared the training and testing sets for modeling.
- Used **Plotly and Tableau** to perform data visualization and understand the trends in the dataset.
- Performed **frequent activity mining** to train the model on existing transactions and understand transaction pattern of customers.
- Used **scikit-learn** library to build classification statistical method and more specifically **KNN (K- Nearest Neighbors)** for prediction of incoming credit card transaction.

Environment: Python, Jupyter, Data Warehousing, SQL, Microsoft Office, Tableau, Plotly.

Project URL: Credit Card Fraud Detection Model

Fraud Insurance Claims Tracking Model **Jersey City, US**
Role: Data Analyst/Analytics **Aug15 – Apr16**

This project was a part of first semester analytics competition organized by **verisk analytics** at saint peter's university. The goal of this project was to **detect the insurance fraud and identify the false claims** by customers. During this project, I analyzed the fraudulent insurance claims and developed a text mining mechanism to identify suspects with

high accuracy as possible. My responsibilities were to perform text mining in python and parse text documents using Stanford NER and python NLTK packages to identify the fraudulent activity, location and suspect information.

Responsibilities

- Determining business requirements through research, **facilitation** and/or **analysis**.
- Gathering **business requirements**, defining **business processes** and coordinating testing activities with other team members.
- Implemented **Stanford NER**, **NLTK** and **geocoder packages in python** to perform text and parts of speech analysis and identified the fraudulent information from the data.
- Worked on **text processing**. Implemented various techniques like **tokenization**, **parsing** and **lemmatization**.
- Worked on **sentiment analysis** in python to analyze the severity of crime/fraudulent activity from the articles.
- Worked on regular expressions to effectively identify the patterns in the documents, which improved the efficiency of identifying fraudulent activity.
- Performed syntactic (POS tagging) and semantic analysis (word vector analysis, IR techniques).
- Conducted team meetings & ensured project documentation is complete.

Environment: Python (Stanford NER, NLTK, geocoder)

Project URL: Fraud Insurance Claims

EDUCATION:

- **Bachelor of Engineering, Computer Science & Engineering**
  Radharaman Institute of Technology & Science, Bhopal, India

- **Master of Science in Data Science with Concentration in Business Analytics**
  Saint Peter's University, Jersey City, NJ, USA

CERTIFICATES:

- Introduction to R, **DataCamp**
- Intro to Python for Data Science, **DataCamp**
- Big Data Foundations, **Big Data University**
- Hadoop Foundations, **Big Data University**
- SQL Fundamentals, **SOLO LEARN**

HONORS & AWARDS:

- Received **Star Performer Award** by Core Card Software team for February month 2015 for developing optimized and error free applications.

REFERENCES:

- Professor Sylvain Jaume, Director Data Science Department, Saint Peter's University, sjaume@saintpeters.edu.
- Om Prakash Sharma, Team Lead Prepaid Domain, CoreCard Software, opsharma@corecard.com.