**Sonali Singh**

[datascientist628@gmail.com](mailto:datascientist628@gmail.com)

**678-799-7914**

# PROFESSIONAL SUMMARY:

* Professional qualified Data Scientist with around 5 years of experience in Data Science and Predictive Analytics including Machine Learning, Data Mining and Statistical Analysis.
* Extensive experience of Software Development Life Cycle (SDLC) methodologies like Waterfall, Agile and RUP.
* Created and executed Test Plan, Test Scripts and Test Cases based on Design document and User Requirement document for testing purposes.
* Involved in the entire data science project life cycle and actively involved in all the phases including data extraction, data cleaning, statistical modeling and data visualization with large data sets of structured data, created ER diagrams and schema.
* Skilled in Advanced Regression Modeling, Correlation, Multivariate Analysis, Model Building, Business Intelligence tools and application of Statistical Concepts.
* Experienced with machine learning algorithm such as logistic regression, random forest, XGboost, Bagging and Boosting, KNN, SVM, neural network, linear regression, lasso regression and k-means.
* Strong skills in statistical methodologies such as A/B test, experiment design, hypothesis test, ANOVA.
* Extensively worked on Python 3/2.7 (Numpy, Pandas, Matplotlib, NLTK and Scikit-learn).
* Experience in implementing data analysis with various analytic tools, such as Anaconda 4.0 Jupiter Notebook 4.X, R 3.0 (ggplot2, Caret, dplyr) and Excel.
* Proficient in Predictive Modeling, Data Mining Methods, Factor Analysis, ANOVA, Hypothetical testing, normal distribution and other advanced statistical and econometric techniques.
* Proficient in Tableau and R-Shiny data visualization tools to analyze and obtain insights into large datasets, create visually powerful and actionable interactive reports and dashboards.
* Experienced in Data Integration Validation and Data Quality controls for ETL process and Data warehousing using MS Visual Studio SSIS, SSAS, and SSRS.
* Extensive experience in MS SQL Server 2008/2012 / 2014, Business Intelligence in MS SQL Server Integration Services, MS SQL Server Reporting Services and MS SQL Server Analysis Services.
* Experienced in cooperating with BA to create business requirement document, functional and non-functional requirement document and use case.
* Excellent communication and interpersonal skills, good analytical reasoning and ability to learn quickly, high adaptability to new technologies and tools.

# TECHNICAL SKILLS:

* **Concepts** - Machine Learning, Time Series Forecasting, Text Mining, Recommender System, Data Visualization, Business Intelligence, Fraud Analytics, Churn analysis, Customer segmentation, Sentiment analysis, Agile Methodology
* **Analytical Tools** – SQL, Python 2.7, R, SAS, SAS JMP, Tableau 9, Power BI, Weka 3.8, MS Excel
* **Languages** – Python 2.7,JAVA, JavaScript, HTML
* **Database**- MS SQL, PostgreSQL, MySQL, Oracle 11g
* **Others**- Skilled in Microsoft office 2016 – Word, Excel, PowerPoint, Git and Bitbucket

# PROFESSIONAL EXPERIENCE:

**Inter-American Development Bank, Washington, DC July 2017 –Present**

**Role: Data Scientist**

**Description**: Working as Data Scientist in Investment Operations Department (INO) of IDB Invest, involved in classifying all the bad loans and doing demographic analysis on the customers. Also, involved in gathering and documenting the requirements. The business objective of the project is to minimize the risk for the bank. And second objective is to support INO efforts related to the analysis of business activities across all Divisions as well as provide and coordinate business information and qualitative reporting to key stakeholders.

**Responsibilities:**

* Research and transform information from raw data into an easily understood analysis that identifies trends and insights for the organization.
* Pinpoint a set of variables to evaluate and work with when deciding on the range of analysis and scope of information sought.
* Use a variety of sources inside and outside of the company to collect, aggregate and analyze data
* Discussed results of database analysis with various members of management in an organization, and led staff members to realize the significance of the data.
* Discovered industry trends based on data collection methods and analysis strategies, and used the information to help the company make production and product adjustments to increase efficiency by 12%.
* Migrated data from different sources (text based files, Excel spreadsheets, and Access) to SQL Server databases using SQL Server Integration Services (SSIS).
* Managed data operations team and collaborated with data warehouse developers to meet business user needs, promote data security, and maintain data integrity.
* Consulted with and trained analysts to effectively access data and optimize Tableau dashboards.
* Maintained SQL scripts to create and populate tables in data warehouse for daily reporting across departments.
* Created customized reports and processes in SAS and Tableau Desktop
* Generated Sub-Reports, Drill down reports, Drill through reports and Parameterized reports using SSRS.
* Designed, Developed and Deployed reports in MS SQL Server environment using SSRS 2014.
* Extensively used Joins and Sub-Queries to simplify complex queries involving multiple tables.
* Worked with database objects like Tables, views and indexes.
* Performed Report Testing and Verification as per Internal Reporting system requirements.
* Administrated interface to organize reports and data sources, schedule report execution and delivery, subscription of reports and track reporting history.
* Generated Daily, Weekly and Monthly Trigger based reports regarding assigned projects to designated Managers and Clients.
* Planned a complete backup on the database and restored database from disaster recovery.
* Performed daily tasks including backup and restore by using SQL Server 2014 tools like SQL Server Management Studio, SQL Server Profiler, SQL Server Agent, and Database Engine Tuning Advisor

**Environment: SAS, Tableau Desktop, Teradata, SQL Server 2008, SQL BI Suite (SSIS, SSRS), Enterprise manager, MS PowerPoint, MS Project, MS Access & Windows Server 2008**

**Infosys, New york, NYC Jan 2017 –May 2017**

**Role**: **Data Scientist**

**Description:** Worked as Data Scientist, was involved in capturing the fraudulent claims on daily basis. Also,

involved in gathering and documenting the requirements. The business objective of the project was to capture all

the behavioral and application kinds of fraud activities happening on credit card transactions. And second

objective was to prevent the bank from false chargebacks from their customers.

* Performed descriptive analysis using SAS and SQL and outlier detection techniques to identify novel fraud cases on credit card related transactions for the North America region
* Apply SMOTE method to handle unbalanced minor class, which increased precision by above 10%
* Built effective classification models like Logistic regression using SAS to anticipate potentially fraudulent transactions
* Succeeded to lower false alarm rate to 3.5% in a quarter by optimizing predictive models through feature engineering(PCA), parameter tuning, and cross-validation techniques
* Created weekly Tableau dashboards for company executives to monitor application fraud for different geographies, achieved 8-10% cost reduction over a one-year period
* Involved in Planning, Defining and Designing database using Erwin on business requirement and provided documentation.
* Created SSIS packages to extract data from OLTP to OLAP systems and scheduled Jobs to call the packages and Stored Procedures.
* Designed SSIS Packages to transfer data from various sources like Text Files, XML Files, SQL Server, Excel and Access to SQL Server using Business Intelligence Development Studio.
* Applied various data transformations like Slowly Changing Dimension, Aggregate, Sort, Multicasting, Conditional Split, Derived column.
* Wrote T-SQL queries, Stored Procedures and used them to build packages.
* Created ad-hoc reports using SQL server 2014 Reporting Services (SSRS).
* Played a major role in production support of SSIS jobs.
* Worked closely with the claims processing team to obtain patterns in filing of fraudulent claims.
* Involved in creating ad hoc analysis and reporting requests in a timely manner for BML Risk related to fraud Approval's, Declines daily, monthly and yearly
* Extracted data from Oracle and migrated into SQL Server 2014 using SSIS.
* Designed various E-R diagrams for different modules using Erwin and Visio.
* Created and presented executive dashboards and scorecards to show the trends in the data using Excel and VBA-Macros.

**Environment: Windows 7, Windows Server 2012, SAS, MS SQL Server 2014, MS SQL Server Reporting Services 2014, MS SQL Server Integration Services, Excel, Acrobat Reader.**

**Bank corp Wilmington DE Sep 2016– Dec 2017**

**ROLE: Jr. Data scientist**

**Description:** Responsible for implementing data mining and statistical machine learning solutions to various business problems such as Churn Analysis and targeted marketing. Work closely with various teams across the company to identify and solve business challenges utilizing large structured, semi-structured, and unstructured data in a distributed processing environment.

**Responsibilities:**

* Filtered and cleaned data, and review reports, summaries, and performance indicators to locate and correct code problems
* Employed customer segmentation based on demographic tendencies for improving marketing strategies using Logistic Regression. Reduced marketing budget expenses by 10%
* Interacted with clients to identify critical questions that need to be answered for the Analytics team to provide effective Key performance indicators (KPI’s), actionable insights, and impactful reports that drive critical decision making
* Reduced attrition rate of customer by 12% in each quarter by ensuring promotional plan through incisive analysis
* Devised a model to convert Life pro plus customers to Life pro-survival plan resulting in a $2M profit in a year
* Designed, developed, and debugged stored procedures in MS SQL Server 2012 to load the data into consolidated database.
* Developed T-SQL queries for table creation, wrote stored procedures, views, triggers, and tested tables in company’s BI star schema structure.
* Maintenance of large data sets, combining data from various sources by Excel, SAS Grid, Enterprise, Access and SQL queries.
* Created Dimensions with the cube wizard and individually and Deployed Cubes to the production server.
* Rebuilding / Monitoring the indexes at regular intervals for better performance.
* Responsible for monitoring and making recommendations for performance improvement in hosted databases which involved index creation, index removal, index modification, file group modifications, and adding scheduled jobs to re-index and [Update](http://www.indeed.com/r/Dany-Divi/fb85b74fcbad4c87#84325321) statistics in databases
* Involved in setting up hierarchies and defined scope of analysis for providing drill down/drill through options in the reporting environment.
* Scheduled Cube Processing from Staging Database Tables using SQL Server Agent.
* Wrote MDX queries and expressions for use in Front End BI Applications.
* Developed reports out of the consolidated data warehouse for the business users using MS SQL Server Business Intelligence Development Studio (BIDS) 2012 (SSRS).
* Designed and created data views for ad-hoc reporting models.
* Developed batch jobs to load the consolidated data into database at scheduled times.
* Implemented security for the reports, user tabs, report models and folders in the report manager at the application, database and the report manager level.
* Created variables and defined formula to capture user responses to a prompt and to display document information for reports using office performance point 2010.

**Environment: MS SQL Server 2012, T-SQL, MS SQL Server Reporting Services (SSRS), Report Builder, Visual Studio 2012, Window Server 2008, SAS, MS SQL Server Integration Services (SSIS), XML.**

**Tata Consultancy Services - Mumbai , India Jan 2013 – Dec 2015**

**ROLE: Data Scientist**

**Description:**  Worked as a Data scientist, to create a sales and service data mart required for Business Intelligence and reporting purposes. The Existing system which allows the agents to enter, modify new sales/service data is used as one of major source systems.

**Responsibilities:**

* Gathered and analyzed Business and System requirements with the customers.
* Performed data analysis and data profiling using complex SQL on various sources systems.
* Written SQL scripts to test the mappings and Developed Traceability Matrix of Business Requirements mapped to Test Scripts to ensure any Change Control in requirements leads to test case update.
* Created Dashboards on Performance Point for managers and directors.
* Participated in designing and developing data marts and cubes.
* Wrote stored procedures and triggers based on requirements.
* Extensively used SQL for accessing and manipulating database systems.
* Hands-on architecture experience with data base schema design, query design and optimization.
* Extracted data from various sources and imported into data warehouse.
* Performed quantitative analysis for financial data and financial statements.
* Involved in writing several SQL queries and procedures to provide multiple database services.
* Performed data cleansing, data mapping, and data profiling using SSIS.
* Assisted in developing ETL standards and best practices.
* Developed reports on SSRS and created multiple types of reports such as parameterized, conditional, table and matrix.
* Developed SSIS packages for ETL to migrate data from different sources including insurance groups etc. to the data warehouse.
* Designed and developed OLAP using MS SQL Server Analysis Services (SSAS).
* Designed and developed MS SQL Server Reporting Services (SSRS) under SQL Server 2008.
* Generated packages using different Transformations like Lookups, Derived Column, Merge Join, Fuzzy Lookup, For Loop, For Each Loop, Conditional Split, Union all, Script component and etc.
* Generated periodic reports based on the statistical analysis of the data using SQL Server Reporting Services (SSRS).
* Created Drill-down, Parameterized and Sub-Reports using RDL and promoted RDLs to Reporting8 Service Server (SSRS).
* Developed Report Model for generating reports at run time.
* Developed Management Information Reports in SSRS on monthly, quarterly, and annual bases.
* Designed SSIS Packages to transfer data from various sources like Oracle, DB2, SQL Server, Excel and Access to SQL Server using Business Intelligence Development Studio.
* Designed various packages for Parallel ETL mechanism using SSIS Control flow and Data Flow components.
* Developed Technical Specification documents and Data Discovery Documents.
* Involved in bug fixing and resolving issues for ETL Packages.
* Optimized queries and ETL packages for better and faster performance.
* As a Report developer, wrote SQL/T-SQL queries, stored procedures for reports and testing.
* Used Stata to perform regression analysis for managers to make prompt decisions with.

**Environment: SQL Server 2008, SQL BI Suite (SSIS, SSRS), Enterprise manager, XML, MS PowerPoint, MS Project, MS Access & Windows Server 2008, Power pivot, Power view**