**Data Analyst**

**SWATHI Email: swathism1413@gmail.com**

**Data Analyst Mobile: 732-723-8906**

**PROFESSIONAL SUMMARY:**

* Data Analyst with **7+** years of experience in **Data Analysis, Business Analysis, Data Mining,** **Data Cleansing, Data Visualization, Statistical modeling** and **Machine Learning**.
* Highly skilled in Advanced **MS Excel** - proficiency with creating **formulas, pivot tables, report charts, V-lookups**.
* Experience in working with distinct phases of Software Development Life Cycle (**SDLC**) in **Agile** methodology.
* Excellent knowledge in Relation Database Management System (**RDBMS**) such as SQL Server, Oracle, MySQL, Teradata and **NoSQL** database like MongoDB
* Expert in **SQL** for writing SQL Queries, Dynamic-queries, sub-queries and complex joins for generating User-defined Functions and Views
* Tremendously involved in Setting up **Tableau** Data Extracts for Bulk Data Reports. Combining Data Extracts and Scheduling Data Extracts Refreshes (Full and Incremental) to circumvent Live Connection when needed.
* Hands on experience with **Tableau** Filter Hierarchy especially using Data Source Filters, Context Filters to fine tune dashboards.
* Good Experience at creating User Groups, Assigning User Permissions and Sharing Dashboards/reports over **Tableau** Online/Server to target User Groups.
* End to end experience in designing and deploying data visualizations using Tableau
* Experience in **Predictive** and **Statistical** **modelling** using **machine** **learning** in **R**
* Experience in **R** using **CARET**, **GGPLOT2**, **DPLYR** and **TIDYR** packages for data preprocessing, data wrangling and data manipulation.
* Experience in Extract-Transform-Load (**ETL**) packages using **SSIS** and **Informatica** to insert datasets to SQL server and generate project reports using **SSRS**, **Tableau** ,**MicroStrategy** and SQL.
* Experience in working on various **python** packages such as **NumPy, Matplotlib, Pandas.**
* Experience in translating legacy scripts code to **python** scripts.
* Experience in Scheduling jobs in **Python** to run scripts to produce weekly reports.
* Experience in using cloud services Amazon Web Services (AWS) including  **S3, EMR**.
* Proficient in **Machine learning algorithms** like Linear Regression, K-Nearest Neighbors, Logistic Regression, Support Vector Machine, Linear Regression, Polynomial Regression, Decision Tree, Random Forests , Apriori ,Time series using LM.
* Performed multiple **Data Mining** techniques like Classification, Clustering, Outlier detection and derived new insights from the data during exploratory analysis
* Extensively followed development methodologies such as **Agile (SCRUM)** with high quality deliverables delivered on-time.
* Experience in Datawarehouse Techniques, Dimensional Modelling, **Star** and **Snowflake** schema
* Experience in building Data Integration and Workflow Solutions for data warehousing using SQL Server Integration Service (**SSIS**).
* Experience in various programming languages (SQL, XML) and operating hardware and software applications including TOAD, Query Analyzer and SQL Navigator.
* Strong understanding of DW concepts and implementations.
* Good command on full life cycle application development.
* Experience in Unit testing using SQL scripts.
* Experience in analyzing and handling large database using Oracle Relational Data Architecture.
* Experience with Build Automation Tool Maven for writing the SCM Tools like Git, GitHub.
* Knowledge in Hadoop eco system such as Pig, and **Spark**
* **Fully Authorized to work in United States for any employer without sponsorship.**

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Operating Systems** | Unix, Windows |
| **Programming Languages** | SQL, T-SQL, NoSQL, Python, R |
| **Databases** | Oracle 10/11g, MySQL, SQL Server, PostgreSQL, MongoDB |
| **BI Tools** | Tableau, MicroStrategy, MS Excel, Rapid Miner, Informatica, SSIS and SQL Server reporting Services (SSRS), TOAD, Teradata SQL Assistant, |
| **Data Modeler** | Erwin |
| **Version Control** | Git ,GitHub |
| Quality Tools | HPQC, JIRA |
| **Cloud/Big Data** | EMR ,S3 buckets, EC2 instance, RDS,Hadoop, Pig, Apache Spark, AWS Redshift |

**EDUCATION:**

**Certification in Data Analytics, UT Austin(2018-2019)**

**Master of Science in Software Engineering from VIT, INDIA (2011)**

**PROFESSIONAL EXPERIENCE*:***

**Mass Mutual, Dallas, Texas Feb 2018 - Present**

**Data Analyst**

**Description:**

Mass Mutual is a Life Insurance company with multiple financial products such as life insurance, disability income insurance, long term insurance and retirement plans and services. The project involved in developing innovative computational analysis strategies to appraise, collect, analyze data across a various structure to find the key metrics and vulnerabilities to empower data-driven decisions.

**Roles & Responsibilities:**

* Developed **R scripts** for data preprocessing and data manipulation
* Used **R** programming language to graphically analyses the data and perform data mining.
* Performed Data Analysis using **T-SQL** to profile data to identify data quality issues analyze for accuracy, to extract specific information, Data Testing and Data flows.
* Extensively used R packages like **GGPLOT2, CARET, DPLYR** on huge data sets.
* Responsible for managing and maintaining the integrity of the SQL server database by normalizing the data and using appropriate constraints in **SQL Server** Management Studio hence reducing the error margin by up to 75%.
* Built reports and created dashboards in **Tableau** to report the data to make Business sense and presented the same to the board of clients.
* Analyzed **market trends** for different age groups and its effect on the market share of the company.
* Applying the **PCA** algorithm to reduce higher dimensionalities to the dataset.
* Involved in implementing **Machine learning algorithms** like Linear Regression, K-Nearest Neighbors, Logistic Regression, Support Vector Machine, Linear Regression, Polynomial Regression, Decision Tree, Random Forests, Apriori, Time series using LM.
* Analyzing feature importance and model parameter tuning to promote the best fit algorithms.
* Developed interactive Dashboards using Visualizations like Dual Axis Bar Graphs, Bar charts and Scatter Plots using **Tableau**.
* Developed scripts to extract data from various file formats like CSV and applied data transformations to business requirements and loaded the data in database.
* Used **RMSE/MSE, R squared, Accuracy and Kappa** to evaluate different models' performance.
* Performed feature selection and feature scaling for feature normalization and data preprocessing.
* Proficient in importing/exporting large amounts of data from files to SQL Server and vice versa.
* Designed stunning visualization dashboards using **tableau** desktop and publishing dashboards on **tableau** server and desktop reader.
* Built **Tableau** visualizations by connecting to various Data Sources Relational Databases (Oracle, SQL Server) hosted on **AWS** RDS environment.
* Designed and developed several tableau reports using data extract with incremental refresh to reduce data traffic out of data sources hosted on AWS.
* Developed **Tableau** visualizations to support ad-hoc analyses connecting to data source hosted in AWS EC2 system and interacted with Business users in understanding their requirements.
* Worked on **AWS Redshift** and RDS for implementing models and data on RDS and Redshift
* Designed and deployed reports with Drill Down, Drill Through and Drop-down menu option and Parameterized and Linked reports using **Tableau**.
* Implement point of view security to **Tableau** dashboards to facilitate visibility across various levels of the Organization.
* Created advanced chart visualizations in Tableau using Dual Axis, Box Plots, Bullet Graphs, Tree maps, Bubble Charts, Pie Chart, Gantt chart, Histograms.

**Environment:** R, T-SQL, Tableau, SQL Server, AWS EC2,EMR and S3

**Delta Airlines, Atlanta, GA Feb 2017 - Jan 2018**

**Data Analyst**

**Description:** This position involved supporting the Technical Operations Division (aircraft maintenance/reliability facility) with identifying, analyzing and normalizing data and to generate daily automated database reporting, ad-hoc data pulls, and analytical projects aimed at continually improving airline performance.

**Roles & Responsibilities:**

* Developed **Python** scripts for manipulating the data reading from various Teradata and convert them as one CSV Files.
* Developed **python** scripts to manipulate arrays using libraries like NumPy and Python.
* Worked on numerous **ad-hoc** data pulls for business analysis and monitoring.
* Designed and developed various monthly and quarterly business monitoring excel reports by writing **Teradata SQL** and using in MS Excel pivot tables.
* Interacted with Business Users for gathering, analyzing, and documenting business requirements and data specifications.
* Performed in depth analysis on data and prepared ad-hoc reports in **MS Excel** and **SQL** scripts.
* Performed verification and validation for accuracy of data in the monthly/quarterly reports.
* Developed Teradata **SQL** scripts using various characters, numeric and date functions.
* Created multi-set tables and volatile tables using existing tables and collected statistics on table to improve the performance.
* Developed and reviewed SQL queries with use of joins clauses (inner, left, right) in **Tableau** Desktop to validate static and dynamic data for data validation.
* Developed Teradata SQL scripts using OLAP functions like rank and rank() Over to improve the query performance while pulling the data from large tables
* Utilized ODBC connectivity to connect Teradata from MS Excel to automate the data pull and refresh graphs for weekly and monthly reports.
* Experience in performing Dual Data Validation on various Business critical reports working with another Analyst.
* Developed **UNIX** shell scripts to run batch jobs and communicate logs to the users.
* Automated Teradata SQL scripts in UNIX by using shell scripting.
* Designed stunning visualization dashboards using tableau desktop and publishing dashboards on tableau server and desktop reader.
* Designed and deployed reports with Drill Down, Drill Through and Drop down menu option and Parameterized and Linked reports using **Tableau**.
* Implement point of view security to **Tableau** dashboards to facilitate visibility across various levels of the Organization.
* Performed SQL Testing on **AWS redshift** databases.

**Environment**: SQL, Python, Teradata, AWS redshift, Tableau, MS Excel

**TracFone, Medley, FL Dec 2015 - Jan 2017**

**Data Analyst**

**Description:**  TrancFone is a prepaid mobile virtual network operator in the North and South America.The project involved analyzing data and provide metrics to Internal quality auditing team for analysis of complaints data and to identify, develop, execute and analyze the key metrics and process improvements.

**Roles & Responsibilities:**

* Working with business users in gathering, analyzing, documenting requirements and translate them into KPI's and metrics for reporting to the business users.
* Worked on creating the Entity resolution tables with ODM (on-Device Management) data for business teams using AWS (EMR and S3) and Python
* Developed **MapReduce/Spark Python** modules for machine learning & predictive analytics in **Hadoop** Implemented a Python-based distributed random forest via Python streaming.
* Wrote Python scripts to parse **XML** documents and load the data in database.
* Used Python to extract weekly information from XML files.
* Developed Python scripts to clean the raw data.
* Implementing Python for database connectivity in **MongoDB**
* Developed Ad-hoc reports using Tableau Desktop.
* Performed Data Analysis using SQL and Excel skills, to profile data to identify data quality issues analyze for accuracy, to extract specific information, Data Testing and Data flows and creating appropriate Dashboards.
* Built **Tableau** visualizations by connecting to various Data Sources Relational Databases (Oracle, SQL Server) hosted on AWS environment.
* Developed Interactive Trends Dashboards regarding Job Trends for Senior Management for creating visualization using stacked bars, histograms, side by side bars.

**Environment**: Python, SQL, AWS (EMR and S3), Spark, Hadoop, Tableau, MongoDB, XML

**Myntra, India Feb 2013 - Apr2015**

**Data Analyst**

**Description:** Myntra is a fashion ecommerce company . There are a 100 plus range of categories and a vast range of subcategories available.The project focused upgrading and normalizing the data and providing day to day deals to its customers by identifying trends and interpreting customer’s needs based on sex and age to promote sales.

**Responsibilities:**

* Experienced in participating meetings with business users to gather the requirements.
* Used Python libraries like **Pandas, NumPy** for data extraction, analysis, data manipulation and **SSRS** for data visualization.
* Built, tested, and deployed **SSIS** packages to deliver provider and member information accurately.
* Created T-**SQL** Joins, sub-queries and other SQL code to implement business rules.
* Performance tuning, and optimization of procedures and queries.
* Designed and developed SSIS Packages (.dtsx files) to import and export data from MS Excel, SQL Server 2008 and Flatfiles.
* SSIS packages to update tables with monthly data and automated it through scheduled jobs for the beginning of every month.
* Designed packages which utilized tasks and transformations such as Execute SQL Task, Data Flow Task, Sequence Container, For Each Loop Container, Lookup, Aggregate and Derived Column.
* Developed packages in SSIS with error handling.
* Extensively used T**-**SQL in constructing Stored Procedures, Table and Scalar Variables, Triggers, Tables, User Defined Functions, Views and Indexes
* Deploying and scheduling Reports using **SSRS** to generate all daily, weekly, monthly and quarterly Reports including current status.
* Generated Drill down, Drill through, Matrix Reports, and Parameterized reports using expressions to calculate, group, filter, and parameterize report using **SSRS**.
* Involved in designing, developing and deploying reports in MS SQL Server environment using SSRS in Business Intelligence Development Studio

**Environment:** Python ,SSIS ,SSRS ,T-SQL ,SQL Server

**ILabz technology,India May 2011-Jan 2013**

**Data Analyst**

**Description:** This project was created for a healthcare client to consolidate pharmacy benefit operations and to outline the requirements to deliver Claims information with user defined fields for Paid/Reject claims.

**Roles & Responsibilities:**

* Documented logical, physical, relational and dimensional data models. Designed the DataMarts in dimensional data modeling using star and snowflake schemas.
* Prepared documentation for all entities, attributes, data relationships, primary and foreign key structures, allowed values, codes, business rules, and glossary evolve and change during the project
* Coordinated with DBA on database build and table normalizations and de-normalizations
* Created, documented and maintained logical & physical database models.
* Identified the entities and relationship between the entities to develop Conceptual Model using ERWIN.
* Creating Data mappings, Tech Design, loading strategies for **ETL** to load newly created or existing tables.
* Perform indepth data analysis and load customer details from datawarehousing to analyze, generate comprehensive reports to decision makers and other affected by the results.
* Extensively worked on documentation of Data Model, Mapping, Transformations and Scheduling batch jobs.
* Developed dimensional model for Data Warehouse/OLAP applications by identifying required facts and dimensions.
* Generated the DDL using forward engineering Worked in merging and complete compare of Physical models.
* Designed STAR schema for the detailed data marts and plan data marts consisting of confirmed dimensions.
* Performed data analysis and data profiling using complex SQL on various sources systems including Oracle.
* Hands on experience performing, Data mining and hardcopy or electronic document study to improve and expand the databases in the application.
* Worked on SQL queries in a dimensional data warehouse as well as a relational datawarehouse.
* 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.
* Recommend additional methods to analyze collect and manage data to improve data quality and efficiency of the data systems across the application.
* Perform administrative tasks, including creation of database objects such as database, tables, and views, using SQL DCL, DDL, and DML requests.
* Involved in extensive data validation by writing several complexes SQL queries and involved in back-end testing and worked with data quality issues.
* Designed and developed database models for the operational data store, data warehouse, and federated databases to support client enterprise Information Management Strategy.

**Environment:** Informatica, ERWIN, SQL, Oracle

SoftSol India Limited**,India Dec 2010 - Apr 2011**

**BI developer-Intern**

**Description:** The primary focus of the project was to upgrade the existing reports by adding new functionalities and consolidation of various documents with interlinking them and making them interactive by adding variety of prompts to allow the user to decide what data they choose to examine.

**Roles & Responsibilities:**

* Responsible in Coordinating with Counterparts for Project Transition, setting up core team responsibilities.
* Created **MSTR** architect component: tables, attributes, facts from scratch.
* Created Intelligent Cubes and reports out of them.
* Created Drill Maps to drill up, down and across unrelated hierarchies in reports.
* Developed Dashboards using bubble chart, interactive bubble chart, heat map, gauge, microcharts and simple graphs
* Prepared project plan by gathering/analyzing client input, updated business users regarding.
* Responsible for extracting ,data curation and cleaning from a wide variety of sources using Rapid miner
* project status, analyzed requirements and designed project workflow
* Identified **KPI’s** to improve the performance of claims processing.
* Regularly involved in weekly scrum meetings with the scrum master and BA
* Responsible for Interlinking all the 5 dashboards.
* Enhanced and finetuned the existing dashboards/reports.
* Performance tuned the reports that were executing for long time.
* Re-designing of the old reports to new format as per the business standards.
* Enhancing, updating and maintaining the Views, Triggers, Cursors and Indexes.
* Conducted unit testing for the developed reports.

**Environment:** MicroStrategy, RapidMiner, SQL, TOAD, MySQL