**Bhavuk Agarwal (Assoc. Project Manager)**

**PMP CERTIFIED**

**DATA SCIENTIST | DATA ANALYST Contact: +1 (409) 213 9747**

**Mail ID:** bhavuka92@gmail.com

**LinkedIn:** [www.linkedin.com/in/BhavukDS](http://www.linkedin.com/in/BhavukDS)

**……………………………………………………………………………………………………………………………………………………………………………………………………………** A young, creative and dedicated engineering professional with proven ability to work individually and in a team to meet the deadlines.

**PROFESSIONAL SUMMARY**

* Overall 9+ years of IT experience and 4+ years of experience in managing Data Science teams, Machine Learning, Data mining projects with large data Sets (Structured and Unstructured Data), Data Validation and Data Visualization.
* Hands on experience in the entire Data Science Life Cycle (SDLC), including **Data Cleaning, Data Wrangling, Exploratory Data Analysis, Data Mining, Validating and Visualizing using Matplotlib and Seaborn libraries.**
* Experience and highly skilled in **Food and Beverages,** **Merchandise, Supply chain and in Retail Marketing Domains.**
* Worked with various datasets, analyzed data **using SQL and Python.**
* Worked with executives and other business leaders to identify opportunities for improvement.
* Collaborated with other team members to collect and analyze the data.
* Used graphs and other methods to visualize data using different python available libraries like Matplotlib, Seaborn.
* Experience in writing **Queries, Sub Queries on MySQL databases.**
* Experience writing queries on large data sets and strong aptitude on learning new technologies and problem solving.
* In-depth knowledge in programming languages **Python and SQL.**
* Knowledgeable in **HDFS, PySpark, Hive SQL tech stack.**
* Extensively used **Pandas (Python) for** **Data Munging**, **NumPy and SciPy** for numerical computations.
* Experience working with various **Database- MySQL, NoSQL- HDFS.**
* Strong Knowledge and experience processing **structured, semi-structured and unstructured data** and handled different file formats like delimited CSV, XML, JSON etc.

**EDUCATION**

**Bachelor of Technology -** Major-Information Technology | UPTU University | 2010-2014 | GPA-3.2

**TECHNICAL SKILLS**

**Analytical Tools:** Matplotlib**,** Seaborn**,** Tableau 10.5 through 8.00, (Desktop, Public, Server, Reader, Online, Prep)

**Programming:** Python, SQL, Machine Learning

**RDBMS:** MYSQL, MS SQL Server, HIVE SQL

**Python/SQL Tools:** SQL Server Management Studio, DBeaver, DB visualizer, Visual Studio Code, Jupyter Notebook,

PYSpark

**Statistical Methods:** Factor Analysis, Regression Models, Logistic Models, Naïve Bayes, Clustering, Conditional Prob.

Confidence Intervals, Sample Size Calculation, Stepwise Regression Model, Hypothesis Test, Principal Component Analysis, MAPE, Precision, Recall, F1 Score, Confusion Matrix

**Scripting Languages:** Python, SQL

**Cluster:** Hadoop Eco-cluster on Azure Cloud, Databricks

WORK AUTHORISATION U.S Lawful Permanent Resident (G.C)

CERTIFICATION ITIL V3 FOUNDATION

**WORK EXPERIENCE**

**Client - PepsiCo, Plano Tx**

**Data Scientist (Assoc. Tech. Project Manager) Sep 2020 to Present**

* Handled and lead a team of 15 people working under different domains in a project.
* Requirement gathering from customers, stakeholders and articulate them for the team to understand the business scenarios.
* Documenting end to end conversations to keep a track of assumptions, discussions and future actions for the team.
* Classical ML/NLP models: Regression & classification models tree & ensemble methods (Random Forests, Gradient Boosting) and kernel methods (SVM)
* Followed agile project methodology to keep a track of project by making epics, features and user stories for the sprint-based architecture to follow
* Implement ML/NLP solutions for clients including models for process optimization, image detection, anomaly detection, automation/pre-emption, recommender systems, fraud detection, text/sentiment analysis, dialog systems
* Create analytics/machine learning driven outcomes for Oil and Gas industry
* Use data science techniques to identify potential efficiency/improvement opportunities in the areas of Design, Construction, Recovery and Production of Oil and Gas
* Design, embed/train NLP models for text, speech, natural language using NLTK, Word2Vec, SpaCy, Gensim for CBOW, skip-grams, tf-idtf, LDA, HMM or RNN LSTMs
* Developed visualization reports using **DOMO** as an alternative to Tableau.
* Developed simple to complex visualization reports using **Tableau** as well as **DOMO**.
* Involved in Tableau Administration to schedule reports and assign rights to the users to view the reports.
* Involved in conversion of existing reports in Tableau and WebFocus to DOMO and DataStream,
* Self learnt new tools like **DOMO, DataStream, GreenField** by myself and implemented various reports and projects using these new tools.
* Used **Teradata SQL Assistance** to implement SQL logic later used in DataStream to perform ETL jobs or to create simple Excel reports.
* Used **DOMO** reporting tool to develope various reports and converted existing Tableau reports into DOMO based on the requirements using the available features in DOMO.
* **GreenField** is also a proprietary tool like Tableau and DOMO developed within Confidential to develop Visualization reports.
* Improved existing WebFocus reports as per the new Requirements and converted few WebFocus reports to either DataStream reports or DOMO reports.
* Support and Administer Db2 Queue replication between DB2 Z/OS and distributed environments
* Testing zparms features and optimizing suggestions to DBMS group for implementation
* Exposure to SMP/E for migrating from DB2 V7 to V8 and installation.
* Leverage Image, text and numeric data within Confidential using AI tools to deliver results
* Worked for on boarding the new resources, provided necessary KT sessions within the team to remove the gaps.
* Raising requests for the new and old team members for specific software’s and AD group accesses.
* Organize and control the activities of the area, assign personnel to the various projects.
* Managed high complexity, cross-functional system implementation, maintenance and integration of projects.
* Monitor progress to ensure project objectives are delivered on time and within budget & business results are realized.
* Pre-Migration testing and validation of all data elements been converted.
* Development of new COBOL/CICS-DB2 Application programs
* Enhancement of existing application(DB2-COBOL/CICS) programs
* Determine the impact of project changes on the business case and re-forecast value creation.
* Helped in conducting interviews for new hires, worked with company HR to follow the guidelines.
* Develop, plan and schedule the development, introduction, communication and maintenance of Project.
* Created Source to target Data mapping document of input /output attributes with the proper transformations which would help for the easy development of SSIS bridge.
* Expertise in Gathering and Analyzing Information Requirements, Data Analysis, Data Architecture, Business **E-R** modeling, Dimensional Modeling, ETL Design
* Used SQL, SSMS and SSIS, Sybase, Excel VBA, Access VBA and Microsoft Applications Excel, Access and Word to design, develop and deploy software applications, including the Comprehensive Credit Analysis and Review CCAR application. used to produce reports submitted to the Fed on a quarterly basis
* Used SSIS to automate ETL processes for manipulation of large volumes of data between SQL and Sybase
* Developed reports and apps using T-SQL, Sybase and VBA to add new and monitor existing Treasury and Corporate Deposit and Loan business and transactions
* Modified/created Balance Sheet Report summary and detailed analysis , Intraday Report, Overnight Report and other financial reports using T-SQL scripts, VBA, SQL/Sybase and Excel
* Interfaced with data layers using SyBase and various systems for use with designed applications used by Traders
* Program/SW application development and database design using VB, VBA, MS ACCESS, EXCEL, SQL, Sybase, MS Office Suite
* Utilized advanced methods of **big data analytics**, **machine learning**, **artificial intelligence**, **wave equation modeling**, and **statistical analysis**. Provided exclusive summary on oil/gas seismic data and well profiles, conduct predictive analyses and data mining to support interpretation and operations.
* Created change request in SNOW and work on approvals.
* Co-ordinated all scrum ceremonies including sprint planning, daily standups, sprint retrospectives, sprint demos.
* Designed and re-architecture Machine learning solution from HD-Insight cluster to databricks.
* Enhanced end to end Salesforce Duration Prediction machine learning Model and deployed in production.
* Enhanced Machine Learning Random Forest Classification preexisting model using Hyper Parameter Tuning and improved accuracy by ~10%. using tree depth, min leaf samples and iterations.
* Developed script in PLSQL to perform end to end testing of the migration for each table being migrated.
* Worked with different models like K-Means Clustering, Linear Regression and Ensemble models to classify the future records to build collaborative model as per business needs and process.
* Enhanced model, helped in substantial reduction of False-Positives which led to an increase in overall accuracy and customer satisfaction.
* Performed pre and post Data Analysis on model results for business understanding which results in retraining of model with updated data.
* Migrated pure Python model from pre-owned clustered environment to Data Bricks environment and used ML Spark.
* Built end-to-end enhanced ML pipeline and improved the efficiency/run time from hours to minutes.
* Migrated code from Test environment to development then to Prod environment.
* Attended daily DSR meetings (Scrum based).
* Built pipeline to extract data from multiple sources like .csv, .xlsx, SQL Database, HDFS.
* Performed POC on Data Bricks environment for code enhancements.
* Used Azure Data Factory to create pipelines and workflows to transfer and orchestrate data from blob to blob and adls to adls.
* Developed **Star** and **Snowflake** schemas based dimensional model to develop the data warehouse.
* Document all data mapping and transformation processes in the Functional Design documents based on the business requirements.
* Used Data Lake as a storage point for structured and unstructured data.
* Google Cloud Platforms ML suite:, Vision API, Video Intelligence, Text-to-Speech, NLP API, Translation, AutoML, Dataflow, Dataproc
* Created presentation/decks based on recommendations and findings for business stakeholders to understand the solution.
* Used Statistical knowledge to identify the patterns and outliers in data, also tried to visualize the patterns using different python libraries like matplotlib and seaborn or ggplot.
* Worked with team of analysts and other associates to process the information.
* Performed Exploratory data analysis on large data sets to find useful information or insights using pandas.
* Evaluated the models using RMSE, Confusion Matrix, ROC and AUC in both dev and production environments.
* Created complex SQL queries to fetch the data from databases.
* Used HIVE to analyze the partitioned and bucketed data and compute various metrics for reporting.
* Knowledgeable on working with presto to query database.
* Worked on Hadoop Eco-cluster for distributed environment and used hive SQL to query database.

WORK AUTHORISATION U.S Lawful Permanent Resident (G.C)

CERTIFICATION ITIL V3 FOUNDATION

**Client - Core logic, Irving Tx**

**Data Engineer/Analyst Aug 2019 to Sep 2020**

* Created Dashboard and reports using **python** to communicate the results to the Business and Management Administrations.
* Lead the effort in converting the logical design to physical design, Data Mart Design by enforcing naming standards, referential integrity across multiple subject areas.
* Based on the technical specification documents, designed logical/ physical data models in Oracle Data Modeler tool to support the warehouse and the data marts.
* Responsible for data mapping and data mediation between the source data table and WMO data tables using MS Access and MS Excel.
* Data Migration testing using SQL and checking source and target data integrity.
* Performed Exploratory data analysis on large data sets to find useful information for business and to decide features to use in ML models.
* Design and develop ETL Informatica mappings using Informatica to support the reporting data mart.
* Worked on identifying facts, **dimensions** and various other concepts of dimensional modeling which are used for data warehousing projects.
* Modified Excel VBA code to connect Excel directly to SQL/Sybase database
* Managed and created extremely large data-sets using SQL/Sybase Queries, Stored Procedures, Triggers, etc.
* Used VBA code to automate data import/export to and from Excel and SQL/Sybase
* Automation of manual processes, programming of SW, application development and database design using VB, VBA, MS ACCESS, EXCEL, SQL, Sybase, MS Office Suite, and other technologies
* Extensively involved in Data Extraction, Transformation and Loading (ETL process) from XML to the staging area and from Staging to ODS using Informatica Power Center.
* **Created SQL queries to fetch the data from database.**
* Created and Designed logical database schema layout structure to prepare for ETL Informatica processes, data mining, extraction, Analysis and Reporting System.
* Design and coding of Conversion of VSAM to DB2 for various programs
* Creation of test databases/ Alteration of DB2 Objects
* Granting access to DB2 Objects
* Provided **Trainings**on **Tableau, DOMO Software** and **Agile Scrum** Implementation.
* Expert in Business Intelligence tools like **Brio, OBIEE, Power BI, Excel, SAP BO**, and **DOMO**Software.
* Developed simple to complex visualization reports using **Tableau** as well as **DOMO**.
* Involved in Tableau Administration to schedule reports and assign rights to the users to view the reports.
* Involved in conversion of existing reports in Tableau and WebFocus to DOMO and DataStream,
* Self learnt new tools like **DOMO, DataStream, GreenField** by myself and implemented various reports and projects using these new tools.
* Worked and enhanced the performance of **Fraud Detection Model** using **Ensemble Model** by 5% accuracy.
* Worked with different models **K-Means Clustering, SVM** for **Anomaly detection** to detect fraud transactions.
* Evaluated the models using **RMSE, Confusion Matrix, ROC and AUC** in both dev and production environments.

**NTT Data (India)**

**Senior Analyst Oct 2017 to Jul 2019**

* Organized data by removing Anomalies, Outliers, Null-values etc.
* Worked with large data sets to get the desired inferences using statistical techniques.
* Attended daily DSR meetings.
* Proficient in applying different tests required for hypothetical testing (**Both numerical and categorical data)**.
* Worked with Data Science teams in building POC’s.
* Used NumPy, Panda’s library for mathematical solutions and pandas to work with data-frames.

**HCL, Noida (India)**

**Analyst (Infrastructure Admin) July 2014 to Sep 2017**

* Management and co-ordination with network, storage, server, security, Wintel, VM Team.
* Developed **scripts** to validate the batch files which are generated on a daily, weekly, fortnightly and monthly basis.
* Worked closely with risk teams to find the default users and generated reports grouped by account transactions using Microsoft Excel.
* Handled 400k+ total accounts of merchants and analyzed most profitable accounts to promote good will to the merchants.
* Worked on development of **SQL** for normalization and renormalization in MYSQL.
* Built SQL queries for performing various operations like create, update, read and delete.
* Involved in testing migrated data field to field mapping.

# WORK AUTHORISATION U.S Lawful Permanent Resident (G.C) CERTIFICATION ITIL V3 FOUNDATION