**DEEPTHI Annem**

+1 (314)-329-1090| [deepthi.a1147@gmail.com](mailto:deepthi.a1147@gmail.com)

[LinkedIn](http://www.linkedin.com/in/deepthi-a-249a04318) | [Portfolio](https://deepthiannem.github.io/) | [GitHub](https://github.com/deepthiannem)

**PROFESSIONAL Summary**

* Data Analyst with 7 years of experience worked with complex datasets in retail, telecom, and aviation sectors. Skilled at extracting valuable insights and translating data into actionable recommendations to support revenue growth and operational efficiency.
* Having strong hands-on experience in extraction of data from various source systems ranging from **Mainframes** like DB2, Flat Files, VSAM files, and IMS, to **RDBMS** like Oracle, SQL Server, Teradata, MySQL, and PostgreSQL.
* Proficient in **SQL** for complex data manipulation, including writing advanced queries, stored procedures, and user-defined functions to extract, transform, and analyze large datasets from relational databases.
* Strong in **Python** for data analysis, skilled in **EDA**, data cleaning, and visualization using pandas, NumPy, and matplotlib to process large datasets and generate insightful reports.
* Proficient in using **ETL** tools **Informatica PowerCenter 9.x** and **SQL Server Integration Services (SSIS)** for extracting, transforming, and loading data from multiple sources.
* Skilled in **Tableau**, **IBM Cognos**, and **Looker** for creating dynamic dashboards and visualizations, effectively handling and presenting large datasets to drive data-informed decision-making across the organization.
* Proficient in developing multidimensional data models with **SSAS** and creating detailed, interactive reports using **SSRS**, optimizing complex data structures for efficient querying and analysis.
* Experienced in using **HDFS** for distributed storage of large datasets and **MapReduce** for parallel processing of complex data, enabling efficient analysis and extraction of insights from massive volumes of structured and unstructured data.
* Specialized in statistical analysis, **A/B Testing**, and **Hypothesis Testing** to extract actionable insights from complex datasets. Experienced in designing experiments and applying analytical techniques to inform business decisions and optimize key performance metrics.
* Proficient in applying machine learning algorithms including linear and logistic regression**, K-Nearest Neighbors** (KNN), **Support Vector Machines** (SVM), and **Naive Bayes** (NB) for predictive modeling and classification tasks.
* Strong experience in implementing data warehouse solutions using **AWS Redshift**, **RDS**, and **S3**, including migrating data from on-premises databases, and extensive expertise in cloud databases and data warehouses across both **AWS** (Redshift/RDS) and **Azure** (SQL Database).
* Skilled in developing and optimizing **SQL Server Reporting Services** (SSRS) solutions, creating parameterized reports, implementing drill-through functionality, and designing custom datasets for efficient **Ad-Hoc analysis** and automated report generation.
* Experienced in implementing projects using **Agile** and **Scrum** methodologies, with proficiency in **Git** for version control and **JIRA** for project tracking and collaboration.
* Experienced in managing multiple concurrent projects, consistently meeting tight deadlines while collaborating effectively with **Cross-Functional Teams** to deliver high-quality data solutions that align with diverse stakeholder requirements.

**Tools and technologies**

* **Programming Languages:** SQL, Python, R, C, C++
* **Databases:** MySQL, PostgreSQL, Oracle, Microsoft SQL Server, MongoDB
* **Data warehousing:** Snowflake, Amazon Redshift
* **Big Data Technologies:** Azure Synapse Analytics, Databricks, PySpark
* **Amazon Web Services (AWS):** Redshift, S3, EC2, Lambda, Glue, SageMaker
* **Cloud Platforms:** AWS, GCP**,** Microsoft Azure
* **Visualization Tools:** Power BI, Tableau, Looker, Excel, Google Charts, IBM Cognos, QlikView
* **ETL Tools:** SSIS, SSRS, Informatica PowerCenter, Hadoop,
* **Version Control:** Git

**WORK Experience**

**Southwest Airlines Data Analyst**

**April 2022 – Present  
Project Description:**

To Develop and implement a data-driven strategy to optimize flight pricing and improve route network efficiency, aiming to increase revenue and enhance operational performance.

**Responsibilities:**

* Designed and implemented a hybrid data architecture using **SQL Server** for transactional data and **Azure Synapse Analytics** for data warehousing.
* Conducted Exploratory Data Analysis (EDA) using Python and **Databricks** to uncover patterns in flight data, customer behavior, and market trends. Developed and optimized complex **SQL Queries** and stored procedures to improve data retrieval efficiency.
* Created and maintained **ETL pipelines** using **Azure Data Factory**, ensuring seamless data flow between SQL Server and Azure Synapse Analytics.
* Developed a machine learning model using scikit-learn and **Azure Machine Learning** service to forecast demand for each route, increasing prediction accuracy by compared to previous methods.
* Implemented time series analysis techniques **ARIMA** and **Prophet** for short and long-term demand forecasting, enabling more accurate capacity planning.
* Collaborated with the Network Planning team to optimize flight schedules using linear programming, resulting in significant improvement in aircraft utilization.
* Developed a dynamic pricing engine using **Python** and **Azure Functions**, which adjusts ticket prices in real-time based on demand forecasts and competitor pricing.
* Created interactive **Power BI** dashboards to visualize **Key Performance Indicators** (KPIs), enhancing data-driven decision making. Implemented an **A/B testing** framework using Azure AI to evaluate different pricing strategies.
* Conducted customer segmentation analysis using clustering algorithms in **Azure Databricks**, providing actionable insights to the marketing team. Optimized data retrieval and storage processes in **Azure Data Lake Storage** for large historical datasets.
* Implemented **Data Quality Checks** and monitoring systems using SQL Server Agent jobs to maintain data integrity.
* Collaborated with **Cross-Functional Teams** to integrate the new pricing engine with existing systems, ensuring seamless data flow and minimizing disruptions.
* Presented monthly performance reports to senior management, effectively communicating complex data insights and contributing to strategic decision-making.

**Tools:** SQL, SQL Server, Azure Synapse Analytics, Python, Azure Data Factory, Azure Functions, Azure Machine Learning, Power BI, Databricks

**Charters Communications Data Analyst February 2020 – March 2022  
Project Description:**

Developed and implemented a comprehensive customer churn prediction model and data-driven retention strategy to reduce customer attrition and improve service quality.

**Responsibilities:**

* + Designed and implemented an **ETL pipeline** using **AWS Glue** to process daily customer data from **CRM** **Systems** and network logs which significantly improved data processing efficiency.
  + Conducted in-depth statistical analysis using **Python** (scipy, statsmodels) to identify key churn indicators, revealing service outages and billing issues as primary drivers of customer dissatisfaction.
  + Developed a feature engineering framework in **PySpark** and created behavioral and temporal features to enhance model accuracy.
  + Build an ensemble machine learning model combining **XGBoost** and neural networks using **Scikit-Learn** and **TensorFlow**, achieving high churn prediction accuracy.
  + Implemented a real-time scoring system using **AWS SageMaker**, enabling near real-time churn risk assessment for all customers.
  + Designed interactive Looker dashboards with custom **LookML** models to visualize churn metrics and model performance across different product lines.
  + Developed a Python library for causal inference analysis to assess the impact of retention initiatives, leading to improved campaign effectiveness.
  + Created a customer segmentation pipeline using **K-means Clustering**, identifying distinct customer profiles for targeted retention strategies.
  + Optimized **SQL** queries in **Amazon Redshift**, substantially improved data retrieval speed for critical churn analysis workflows.
  + Collaborated with the marketing team to develop personalized retention strategies based on churn predictions, resulting in a significant reduction in customer attrition.
  + Conducted regular knowledge sharing sessions with the analytics team on advanced **Machine Learning** techniques and best practices in churn prediction.
  + Presented monthly churn analysis reports to senior management, providing actionable insights for strategic decision-making.

**Tools:** Python, PySpark, AWS Glue, AWS SageMaker, Amazon Redshift, scikit-learn, TensorFlow, XGBoost, Looker, SQL

**O’Reilly Auto Parts Data Analyst January 2019 – January 2022  
Project Description:**

Developed and implemented a data-driven inventory management system to optimize stock levels across stores and distribution centers, resulting in improved supply chain efficiency and reduced costs.

**Responsibilities:**

* + Designed and implemented ETL workflows using **Informatica PowerCenter 9.x**, integrating data from multiple source systems including point-of-sale and supplier databases.
  + Utilized **Snowflake's** cloud data warehouse to create a scalable analytics environment, significantly improving query performance for complex inventory reports.
  + Conducted **Exploratory Data Analysis (EDA)** on historical sales data using Python, identifying key patterns and seasonality trends influencing product demand across different store locations.
  + Created interactive **Tableau** dashboards to visualize inventory levels, demand forecasts, and supply chain **KPIs,** enhancing stakeholder decision-making processes.
  + Implemented time series forecasting models using Python's libraries to predict product demand across 500+ store locations, resulting in a reduction in stockouts.
  + Developed a custom SQL-based scoring system to identify slow-moving inventory items, improving inventory turnover for targeted product categories.
  + Collaborated with the IT team to integrate the new inventory management system with the existing **ERP**, ensuring seamless data flow and improving overall **Data Consistency**.
  + Utilized **JIRA** to manage project tasks and track progress, facilitating collaboration between analytics, IT, and business teams.
  + Implemented data quality checks and validation rules in **Informatica PowerCenter**, enhancing the reliability of inventory reports.
  + Optimized SQL queries in **Snowflake**, improving data retrieval speed for critical inventory analysis workflows.
  + Generated monthly performance reports, providing insights that led to reduced carrying costs and improved product availability across the network.

**Tools:** Informatica PowerCenter, Snowflake, Tableau, Python, SQL, SAS, JIRA, Excel

**Ahex Technologies Private Limited Data Analyst June 2017– December 2018  
Project Description:**

To develop a comprehensive sales performance analytics dashboard for DMart, a major retail chain. The goal is to provide real-time insights into sales data across multiple store locations, enabling data-driven decision-making for inventory management, promotional strategies, and store performance optimization.

**Responsibilities:**

* + Designed and implemented an ETL pipeline using **Python** to extract daily sales data from Oracle-based point-of-sale systems and load it into a **Microsoft SQL Server** data warehouse.
  + Developed complex SQL **Stored Procedures** and **views** to aggregate and analyze sales data, calculating key metrics such as daily/monthly revenue and product category performance.
  + Created an interactive **Power BI** dashboard featuring real-time sales **KPIs,** trend visualizations, and drill-down capabilities, resulting in a reduction in decision-making time for management.
  + Utilized **Pandas** and **NumPy** for advanced data manipulation, including time series analysis of sales trends and seasonal patterns.
  + Implemented a machine learning model using **Scikit-Learn** to predict stock-outs, improving inventory management efficiency.
  + Developed automated daily and weekly sales performance reports using Python and **Plotly**, significantly reducing manual reporting tasks.
  + Collaborated with stakeholders to define and prioritize **Key Performance Indicators** (KPIs) for sales analysis, ensuring alignment with business objectives.
  + Performed **Exploratory Data Analysis (EDA)** using various statistical techniques, uncovering insights that led to improved promotional campaign effectiveness.
  + Created a custom Python library for standardizing common data processing tasks, enhancing team efficiency.
  + Conducted regular knowledge sharing sessions on data analysis techniques and tools, contributing to overall team skill improvement.
  + Optimized database queries and dashboard performance, reducing average report generation time.

**Tools:** Python, SQL, Microsoft SQL Server, Power BI, Pandas, NumPy, Plotly, scikit-learn

**Education**

* **Saint Louis University St. Louis, MO** 
  + - **Master of Science in Computer Science**