SAUMYA SINHA

Contact: +1 (425) 370-9038
Email: saumyasinha0510@gmail.com

Github: https://github.com/saumya-sinha01
LinkedIn: https://www.linkedin.com/in/saumya-sinha-09b318b5/
Tableau: https://public.tableau.com/app/profile/saumya.sinha01

Experienced Data Analyst skilled in Python, SQL and Tableau, seeking to generate actionable insights and optimize business process.

WORK EXPERIENCE

Data Research Assistant, San Jose State University   Feb 2024 - Present
- Processed form-based image datasets by removing duplicates, resizing, and normalizing images, and fine-tuned LLMs (LayoutLMv3, YOLOv8) to improve baseline performance by 10%.
- Utilized docTR and PyTesseract for text detection and extraction, and developed a Python-based web application providing text search, addition, and removal functionalities.

Data Analyst, Optum Inc. (UnitedHealth Group)   May 2019 - April 2021
- Analyzed healthcare claims data using SQL and Python to identify cost-saving opportunities and trends in healthcare utilization, leading to actionable insights that optimized average claims processing time by 15%.
- Created interactive Tableau dashboards to visualize cost patterns, track key metrics such as claim approval rates, processing time, and denial reasons, enhancing visibility for stakeholders and improving decision-making abilities.
- Led a project to improve the quality of data collected and used within the All Savers portal, identifying data quality issues, developing solutions, and implementing best practices for data management, while training team members.

Member Technical, Automatic Data Processing (ADP)   Sep 2016 - May 2019
- Conducted exploratory data analysis (EDA) using SQL and Python to examine employees' punch-in/punch-out data, absence, and overtime patterns, providing a robust foundation for further insights.
- Developed interactive Tableau dashboards to visualize the analyzed data, resulting in optimized shift schedules and improved tracking of productivity and engagement metrics.
- Enhanced stakeholders' understanding of data patterns through clear and actionable visualizations, leading to a 20% improvement in decision-making efficiency and a 15% increase in employee engagement.

ACADEMIC PROJECTS

Analysis of Vehicle Accidents and Injury Prediction Using Machine Learning, Data Analytics, SJSU   Aug 2023 - Dec 2023
- Analyzed crashes, road conditions, and weather in San Jose to identify factors linked to accidents and predict accident severity.
- Cleaned data, performed EDA & tuned ML models (Decision Trees, XGBoost), achieving 0.48 F1 macro, 0.88 AUC, and 0.70 accuracy.
- Utilized Tableau to show geospatial visuals of crash severity, analyzing injury levels, key features like age & sex for fatal injuries.

Diabetes Prediction, Data Mining, SJSU   Jan 2023 - May 2023
- Developed a diabetes prediction system using NHANES 2013-14 data, focusing on data preprocessing, performing EDA to uncover patterns and anomalies, and utilizing Tableau for visualization and feature selection with ANOVA, Chi-squared, and RFE.
- Trained SVM, Logistic Regression, KNN, and Random Forest classifiers, achieving 89% AUC and 85% accuracy.

Get Your Perfect Fit - H & M, Database Systems for Analytics, SJSU   Jan 2022 - May 2022
- Utilized H&M dataset to offer personalized outfit recommendations based on body shape, skin condition, and fabric preferences.
- Performed ETL using AWS Redshift, S3, and Glue to efficiently load, process, and prepare data for analysis.
- Utilized Redshift queries & Tableau visualizations to analyze customer demographics, spending data, inventory & sales trends.

EDUCATION

- Master of Science in Data Analytics, San Jose State University   Jan 2022 - Dec 2023
- Bachelor of Technology, Computer Science, Biju Patnaik University of Technology   Aug 2012 - May 2016

TECHNICAL SKILLS

Programming: Python, Java, R , JavaScript   **Cloud**: AWS (S3, Redshift, Glue, Sagemaker)

Databases: MySQL, SQL Server, NoSQL (DynamoDB), BigQuery   **Machine Learning**: SKlearn, HuggingFace, Keras, TensorFlow

Visualization Tools: Tableau, PowerBI, Matplotlib, Seaborn, Plotly, MS Excel

Analytical Skills: Data Visualization, Exploratory Analysis, Data Mining, Statistical Modeling

Software/Other Skills: Jira, Rally, Critical Thinking, Presentation Skills, Problem-solving, Decision-Making, Storytelling, Google Colab

CERTIFICATIONS

- Google Data Analytics Professional   • Tableau   • MS Excel   • Amazon Redshift   • Amazon SageMaker