

Skills

- Tableau, Numpy, Pandas, Seaborn, Matplotlib , Python, SQL, Excel, Probability and Statistics, Customer Segmentation, Root Cause Analysis, Machine Learning, Project Management, Product Analysis, Predictive Analytics, Data Cleaning and Preprocessing, Time Series Analysis, Supervised learning, Unsupervised learning , Recommendation systems, Git and Github, NLP

Experience

Data Scientist

Mar 2024 - Present

169PI

Mumbai

- Developed cutting-edge conversational multimodal RAG search engine AI chatbot, Pi, for real-time information delivery.
- Implemented AI-driven assistants using Mistral LLM and OpenAI, enhancing information discovery and chat functionality.
- Integrated features for contextual response generation, efficient information retrieval, chat summarization, and text extraction.
- Successfully built an end-to-end RAG system for parsing long pdfs and interacting with the document or summarizing it.
- Developed a custom function for a Retrieval-Augmented Generation (RAG) model to integrate inline citations within the bot's responses, enhancing the accuracy and traceability of generated content.
- Technology Stack: Python, Data Visualisation, StreamLit, SQL, Docker, MLFlow, Llama3, ChromaDB, Ollama, postman, LangChain.

Data Analyst

Apr 2023 - Nov 2023

Omnify, Inc

Bangalore

- Analyzed top 5 profitable campaigns and performance of ad groups on Tableau for Omnify's Google ad campaign, aiding 14% MoM growth.
- Analyzed and measured key performance indicators (KPIs) for the organization, including a 4% month-over-month revenue growth rate, average revenue per client, gross margin, net profit margin, and client retention rate.
- Developed and presented insights through a comprehensive Tableau dashboard to support datadriven decision-making.
- Created tableau reports to enhance data interpretation for business initiatives, leading 25% boost in influencing strategic decisions and stakeholder communication.
- Defined key metrics such as revenue and trends, keeping the team informed and proactive in adapting to changing consumer patterns.
- Validated data model is in line with projected data growth, non- functional requirements and changing consumer patterns.
- Partnered with the development team to embed reports into client web applications, resulting in a 20% increase in report accessibility and usability.
- Tech Stack:Python Libraries - Pandas , numpy, matplotlib, plotly, Seaborn, Data Analysis, Excel, Tableau, Streamlit.

Electrical Engineer Supervisor

Oct 2021 - Jan 2023

S.R Engineering Works

Dhanbad, Jharkhand

- Led and guided electrical teams to achieve project goals through effective communication and collaboration.
- Overseen the installation and maintenance of electrical systems, ensuring compliance with safety standards and regulations.
- Identified and resolved electrical issues promptly, conducting root cause analysis to prevent future problems.

- Effectively communicated with clients, understanding their requirements, and providing solutions to meet their expectations.

Projects

YULU- EDA AND HYPOTHESIS TESTING.

https://github.com/DebershiMitra/SCALER-DATASCIENCE/blob/main/YULU_HYPOTHESIS_TESTING_.ipynb

- Conducted EDA and Hypothesis testing to Analyze a dataset, Exploring demand determinants for shared electric cycles in Yulu's Indian market amidst revenue challenges.
- Executed a Univariate and Bivariate Analysis to Extract actionable Insights for Yulu aiding their Exploration of shared electric cycle demand in the Indian market.
- Utilized Hypothesis testing to identify the significant Features for predicting shared Electric Cycle demand in the Indian market.
- Found the lowest Average count of Rental Bikes Between 1 AM to 4 AM.
- Discovered that despite 68.6% of cycles being rented on 'Working-Day' compared to 31.4% on 'Non-Working Day', the T-Test analysis doesn't provide enough evidence to conclude a significant effect of the 'Working Day' feature on the number of cycles rented.

Business Case - Scaler - Clustering Job Profiles for Scaler's Learners.

<https://github.com/DebershiMitra/Business-Case-Scaler---Clustering>

Scaler, an innovative online "tech-versity" by InterviewBit, delivers intensive, industry-aligned computer science and data science programs. Through live classes taught by top tech leaders and subject matter experts, Scaler equips software professionals with modern, cutting-edge skills to excel in the competitive tech landscape. Scaler's meticulously structured curriculum covers the latest industry trends and technologies, making it one of the most sought-after platforms for upskilling. As a data scientist in Scaler's analytics team, you are focused on creating data-driven insights into the job market landscape for Scaler's learners. Specifically, you have been tasked with profiling the top companies and job positions using the extensive Scaler database and clustering learners based on various attributes, such as job profiles, companies, and additional features. These clusters will enable Scaler to understand patterns and groupings within its learner community, enhancing personalization and career guidance services.

Project Scope The project consists of the following key steps:

- 1. Data Exploration and Preprocessing Data Review: Begin with a thorough exploration of the learner dataset to understand the features related to job profiles, companies, and additional attributes that may aid clustering. Data Cleaning: Handle missing values, outliers, and inconsistencies to ensure data quality. Feature Engineering: Create new features or modify existing ones to better capture nuances in job roles, company attributes, and learner backgrounds.
- 2. Clustering Model Development Defining Similarity Measures: Establish metrics to measure similarity across learner attributes, ensuring that clusters reflect meaningful groupings. Clustering Algorithm Selection: Choose a clustering algorithm such as K-Means, Hierarchical Clustering, or DBSCAN to achieve ideal segmentation based on experimentation and evaluation. Hyperparameter Tuning: Optimize the number of clusters and other model parameters to maximize intra-cluster similarity and inter-cluster diversity.
- 3. Cluster Evaluation and Validation Silhouette Score and Davies-Bouldin Index: Use these metrics to assess the cohesiveness and separation of clusters. Cluster Interpretability: Analyze the characteristics of each cluster to ensure they make intuitive sense and align with Scaler's objectives. Expected Outcomes The final clusters will group learners with similar job profiles and company associations, offering Scaler actionable insights into: Top Companies and Roles: Identify which companies and positions attract similar types of professionals, aiding in targeted career guidance. Learner Personalization: Tailor Scaler's services and recommendations to better fit distinct learner segments. Market Positioning: Enable Scaler to align its curriculum with in-demand roles and companies, enhancing its value proposition for prospective learners.
- Conclusion By clustering learners based on their professional profiles and job attributes, this project aims to provide Scaler with a data-driven understanding of career trends, ultimately helping learners make informed career decisions and positioning Scaler as an essential resource for advancing tech careers.

OLA - Ensemble Learning- Driver Attrition Prediction

<https://www.linkedin.com/feed/update/urn:li:activity:7170898429970239488/>

- Predicted driver attrition to reduce turnover costs and improve retention strategies.

Key Responsibilities & Achievements: Data Exploration & Analysis

- Imported and conducted exploratory data analysis on driver datasets for 2019 and 2020.
- Checked data structure and characteristics; identified missing values.
- Performed univariate and bivariate analysis to understand variable distributions and relationships.

Data Preprocessing:

- Imputed missing values using KNN imputation.
- Engineered features, including creating a target variable indicating driver attrition and a variable for monthly income increase.
- Conducted correlation analysis among independent variables.
- Applied one-hot encoding to categorical variables.
- Treated class imbalance using appropriate techniques.
- Standardized training data for model readiness.

Model Building & Evaluation:

- Developed ensemble learning models using Bagging and Boosting methods with hyper-parameter tuning.
- Evaluated model performance using Classification Report and ROC AUC curve.

Insights & Recommendations:

- Provided actionable insights based on exploratory data analysis, including comments on data shape, attribute ranges, outliers, and variable relationships.
- Highlighted critical factors influencing driver attrition and recommended strategies for improving driver retention.

Technologies & Tools Used:

- Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn)
- Machine Learning Algorithms (Ensemble methods: Bagging, Boosting)
- Data Imputation Techniques (KNN)
- Statistical Analysis and Visualization

Target Retail - SQL Project

<https://github.com/DebershiMitra/SCALER-DATASCIENCE/blob/main/SQL%20BUSINESS%20CASE-%20DEBERSHI%20MITRA%20DSML%20nov.pdf>

- Established a notable 136.6% increase in the cost of the orders from 2017 to 2018.
- Executed advanced queries to provide insights into the month-on-month number of orders placed using different payment types, helping identify trends and preferences in consumer behaviour and payment methods over time.
- Applied SQL in Google BigQuery to Analyze and Extract valuable insights and recommendations by examining order status, pricing attributes, and customer reviews.

Achievements

- Secured victory in the U-19 District Badminton Tournament and earned the distinction of serving as both the School's Sports and House Captain. Demonstrated leadership and organizational skills by actively participating in numerous cultural programs throughout primary to higher secondary school, serving as both a participant and coordinator.

Education

Scaler

2024

Specialized in Data Science & Machine Learning

<https://www.scaler.com/academy/profile/2021a547f409/>

- Currently pursuing data science and machine learning with proven abilities to analyze intricate data sets, extracting actionable insights to facilitate informed decision-making.

Oriental Institute of Science and Technology, Bhopal

2021

BE/B.Tech/BS in Electrical