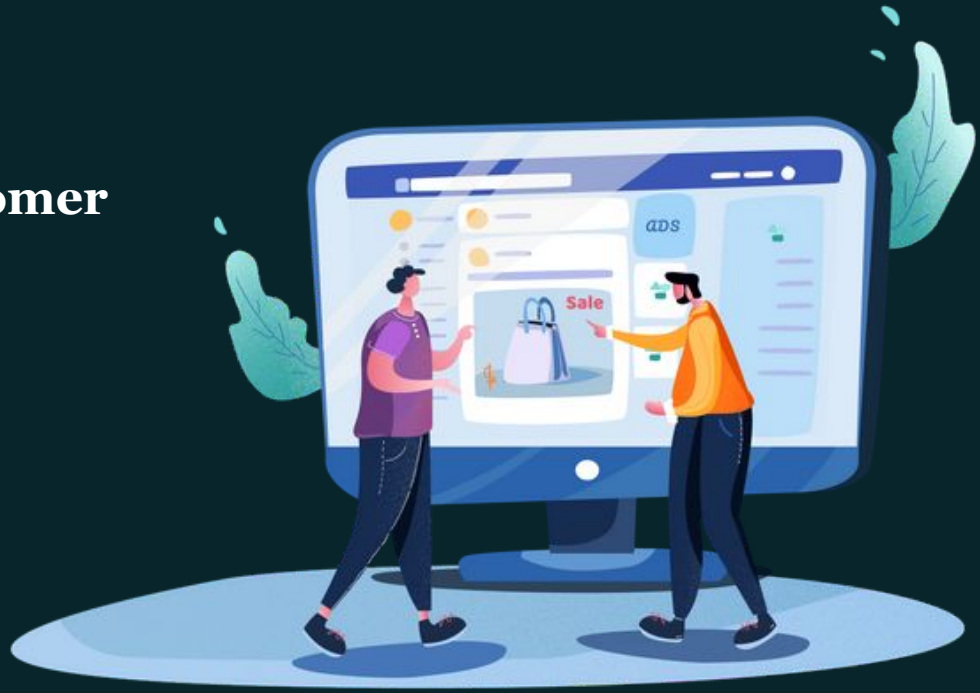
AMDARI®

Data Science Project

Cohort Analysis for assessing customer retention in E-commerce industry

Data Science Project



Specialization: Sales Analytics

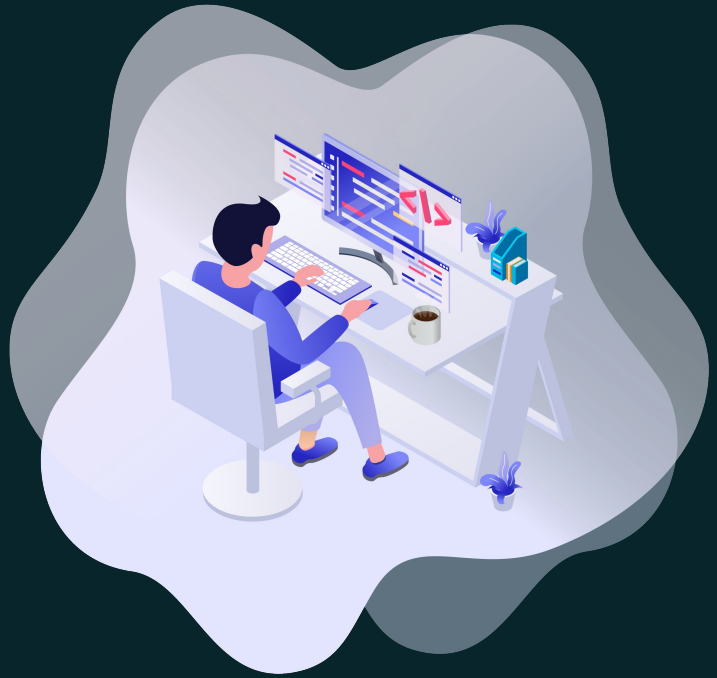
Business Focus: E-commerce

Tool: Python & Git

AMDARI®

Project Learning Opportunity

Sharpen your analytical skills and uncover trends, patterns, insights that can drive strategic decisions. In this project, you will gain invaluable insights into how Cohort Analysis can unlock a deeper understanding of customer behavior, helping businesses pinpoint retention opportunities and optimize marketing efforts. but also a transformative learning experience that can empower you in the ever-evolving world of e-commerce.



Learning Skill

- Exploratory Data Analysis
- Retention rate analysis
- Data visualization
- Clustering and segmentation
- Time-based Cohort analysis with Python
- Version Control using git

AMDARI®

Business Introduction

Company: E-Shop Pro
Industry: E-Commerce

E-Shop Pro is a leading online retailer offering a wide range of products, including electronics, apparel, and home goods. With millions of customers globally, the business prides itself on its fast delivery, customer-centric policies, and personalized shopping experience. Some major achievements highlighted by the company include:

- Growth in user base by 30% annually over the last five years.
- Implementation of a personalized recommendation system, improving customer satisfaction.
- Achieved 80% retention rate among VIP customers in the first two years of loyalty program implementation.

Despite its numerous successes, the company has encountered a formidable challenge that has left its leadership team determined to find a solution: an alarmingly high shopping cart abandonment rate.

AMDARI®

Business Problem

The company is facing a significant challenge in retaining customers over time. While the company experiences high customer acquisition rates, they are seeing a drop-off in returning customers after the first purchase. Specific challenges include:

- **Declining repeat purchase rate**: New customers are not returning after their initial purchase.
- **Customer churn**: A noticeable percentage of customers leave after a brief period of activity, increasing marketing and acquisition costs.
- **Low engagement**: Customers show decreased interaction with the platform after their first few visits, indicating a failure to maintain long-term interest.

Resolving these issues is critical for improving customer lifetime value (CLV) and overall business growth.

AMDARI®

Rationale for cohort analysis

Cohort Analysis is the most effective tool to assess customer retention by dividing customers into groups (cohorts) based on shared characteristics or behaviors over a specific period. By tracking cohorts over time, companies can identify:

1. **Customer retention rates** over different time frames (e.g., weekly, monthly).
2. **Patterns of churn** and identify when customers are most likely to stop engaging.
3. **Effectiveness of retention strategies** such as loyalty programs or promotions.
4. **Customer segmentation** to tailor retention strategies based on behavior.
5. **Optimization opportunities** for the onboarding process and post-purchase engagement.

This project will help the company in identifying trends in customer retention and take proactive steps to improve long-term loyalty.

AMDARI®

Aim of Project

The aim of the project is to conduct a retention rate time-based cohort analysis on the dataset.

Objectives of the Project:

01

Retention Rate Analysis:

To calculate and analyze the retention rates of different customer cohorts over time.

02

Cohort Segmentation:

To segment customers into distinct cohorts based on common characteristics or behaviors, such as acquisition date or first purchase date, enabling a more granular analysis of customer retention.

03

Identify Retention Trends:

To identify trends and patterns in customer retention within each cohort.

04

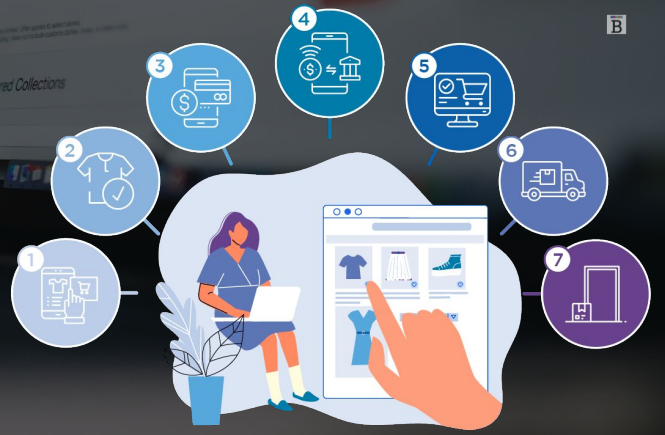
Churn Analysis:

To pinpoint the reasons and timing of customer churn within various cohorts.And build a predictive model that can be used to predict churn behavior

05

Recommendation Strategies:

To develop data-driven strategies and recommendations for improving customer retention based on the insights gained from the cohort analysis.



Data Description

The dataset required for this project will include:

- **InvoiceNo**: A unique identifier for each invoice or transaction, often used for tracking and reference purposes.
- **StockCode**: A code or identifier associated with a specific product or item in the e-commerce store's inventory, used for cataloging and tracking purposes.
- **Description**: A categorical feature that provides a brief textual description of the product or item being sold, offering clarity to customers about what they are purchasing.
- **Quantity**: The quantity or number of units of a product that were included in the transaction, indicating the purchase volume for each item.
- **InvoiceDate**: The date and time when the transaction or invoice was generated, offering insights into when purchases were made and allowing for temporal analysis.
- **UnitPrice**: Indicating the total cost of the items purchased.
- **CustomerID**: A unique identifier associated with each customer or shopper, allowing for customer-specific analysis and tracking of individual purchasing behavior.
- **Country**: The name of the country where the customer is located or where the transaction occurred.

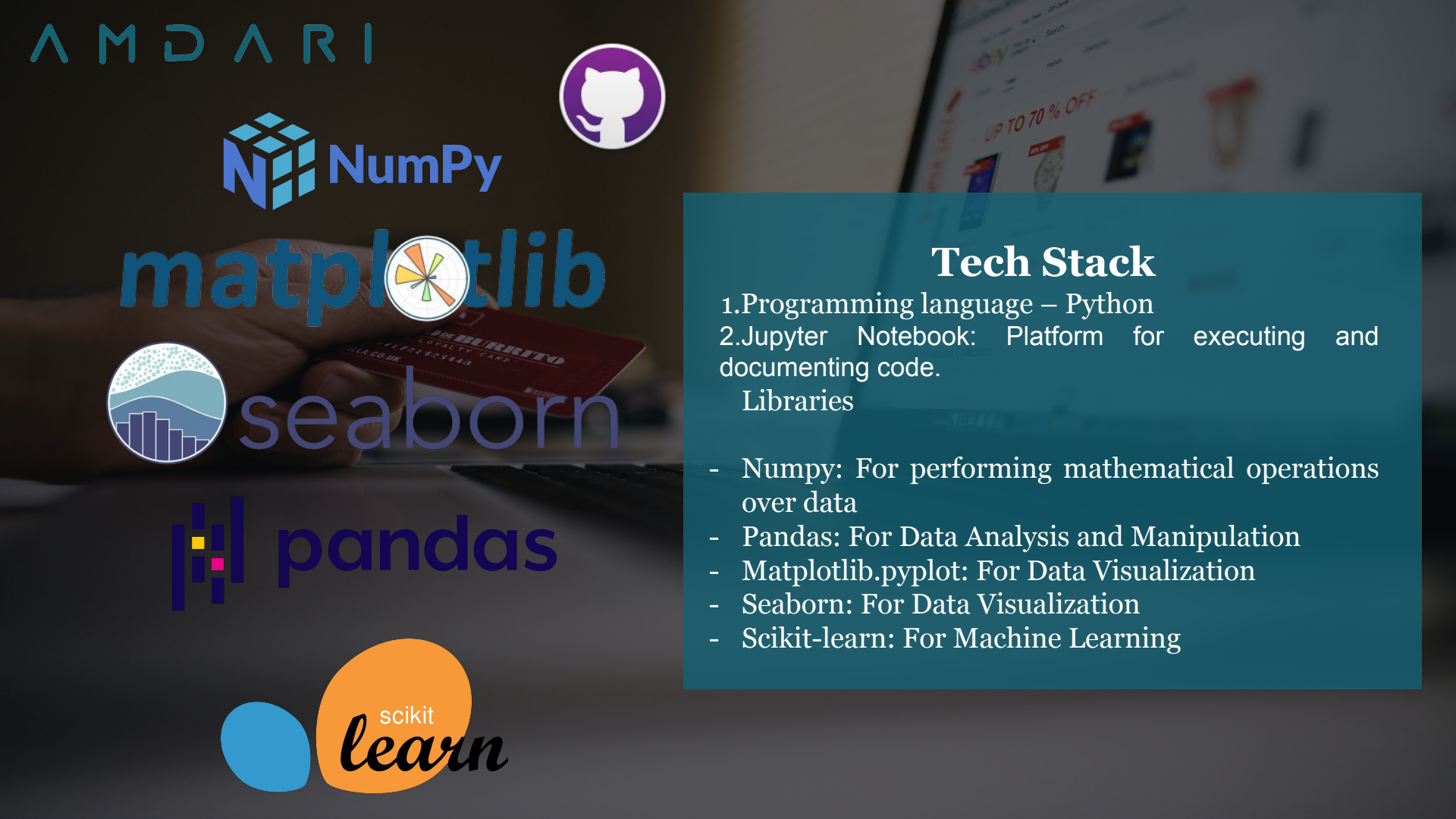
AMDARI®

Project Enhancement

To enhance the project deliverables, you are required to segment the data and implement a Version Control system using git

Customer segmentation: in addition to cohort analysis, we will be utilising a clustering algorithm to to create clusters and perform segmentation analysis, to help us understand the purchasing behavior of the customers

Git : you are required to set a version control system using git that will enable you track version changes and code history from your work on github



Tech Stack

1.Programming language – Python
2.Jupyter Notebook: Platform for executing and documenting code.

Libraries

- Numpy: For performing mathematical operations over data
- Pandas: For Data Analysis and Manipulation
- Matplotlib.pyplot: For Data Visualization
- Seaborn: For Data Visualization
- Scikit-learn: For Machine Learning

AMDARI®

Data Science Project Scope

Ingest Data
Ingest the data and perform data cleaning

Exploratory Data Analysis
the data is thoroughly explored and analysed to gain insights and understand its characteristics. This involves statistical analysis, data visualization, and other exploratory techniques to identify patterns, correlations, anomalies, and potential issues in the data.

Feature Engineering
Feature engineering is the process of creating or selecting relevant features (input variables) from the available data that will be used for modeling. It may involve feature extraction, transformation, scaling, or the creation of new features to improve the predictive power of the models.

Model Development
Various modeling techniques are applied to the prepared data in this stage to build predictive or descriptive models. This can include machine learning algorithms, statistical models, or other analytical approaches. The models are trained, validated, and fine-tuned to optimize their performance.

Model Evaluation and Selection
Once the models are developed, they are evaluated using appropriate evaluation metrics and validation techniques. This involves assessing their performance, generalizability, and robustness. The best-performing model(s) are selected for further deployment or refinement.

Reporting and Strategy Recommendations
Generate a good report and develop strategies and recommendations to curtail churn rate.