Damola Adediran

Machine Learning Project

By: Adedamola Adediran

Customer Loyalty Program as a Platform Service

https://adedirandamola.com/
https://github.com/adexdams

Overview

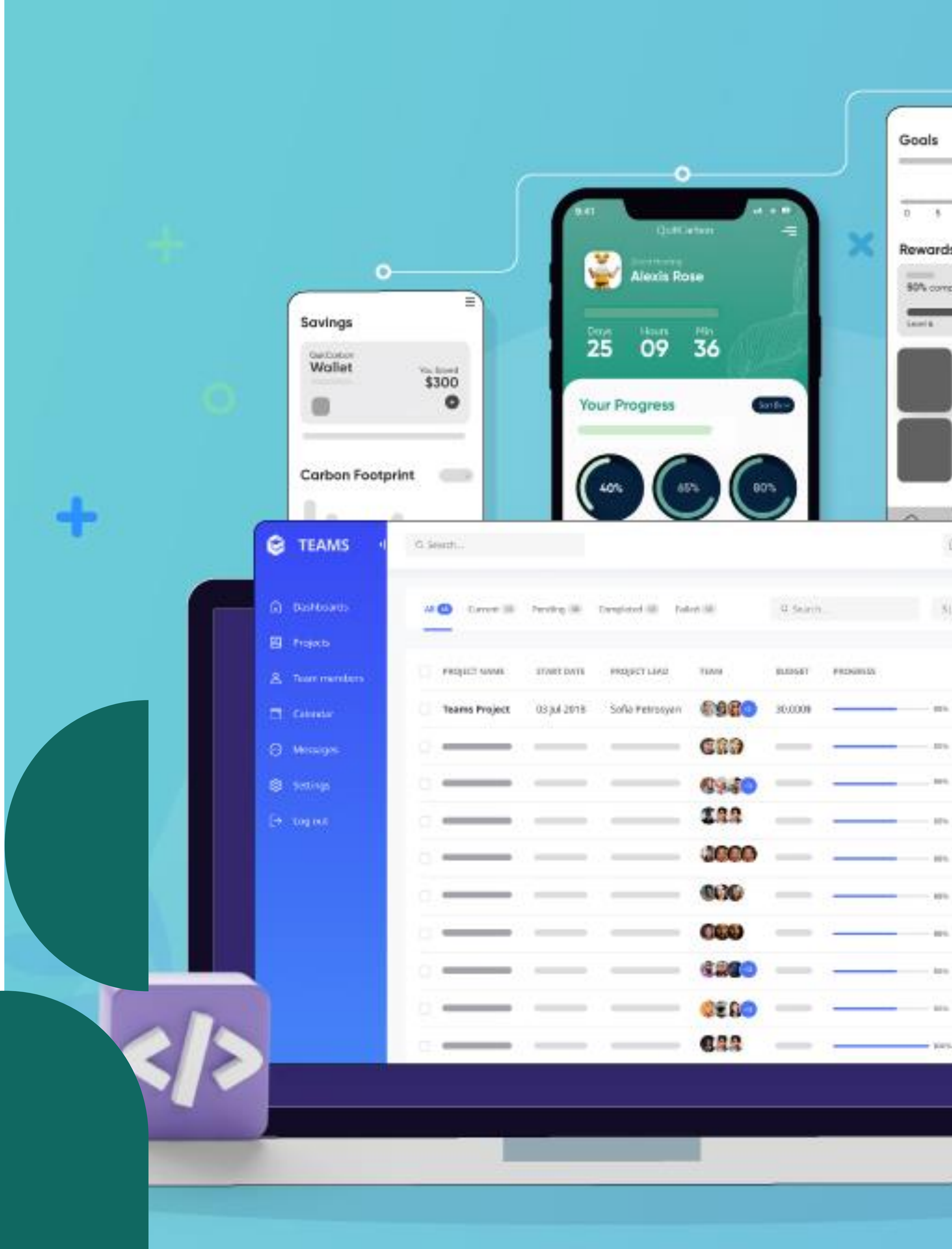
1. Introducing the product
2. Develop the problem
3. Framing the ML hypothesis
4. The Product Goal
5. Framing the ML task
6. Data source and feature requirements
7. Simplified ML problem into models
8. Data Sourcing for MVP

Introducing the product

The product is an AI-driven loyalty program optimization platform that automates the selection of weekly promotional items and personalizes loyalty benefits for both customers and retailers, thereby maximizing customer engagement, satisfaction, and sales.

Background

- Retailers face challenges in efficiently selecting the most effective products for weekly loyalty promotions, leading to suboptimal sales and customer engagement.
- This issue is frequent, given the dynamic nature of retail inventory, changing customer preferences, seasonality changes, and loyalty program schedules.
- My motivation is to enhance retail productivity and profitability while improving customer satisfaction.



Developing the problem

The dynamic nature of retail, changing customer preferences, and the need for personalized loyalty promotions present significant challenges for retailers in selecting the most impactful products for their weekly loyalty programs, as supported by these statistics.

The problem

61% Retailers cite customer retention as their biggest challenge.[1]

77% Consumers now retract their loyalty more quickly than they did three years ago.[2]

The opportunity

80% Retail customers say they are more likely to do business with a firm if it offers personalized experiences.[3]

1. https://rosetta.ai/blog/statistics-on-customer-retention-and-loyalty
2. https://queue-it.com/blog/loyalty-program-statistics/
3. https://queue-it.com/blog/loyalty-program-statistics/

The Target Audience

Retailers and their marketing teams are the primary users, while the end beneficiaries are customers participating in their loyalty programs. Retailers gain improved sales and operational efficiency, while the customers receive more relevant and enticing loyalty benefits.

Framing the ML Hypothesis

Potential ML use case

1. What presale benefits to offer weekly
2. What post-sales rewards to offer customers?
3. What products to select and recommend to customers weekly?
4. What is the appropriate timeline per benefit?
5. How do we distribute our budget against customer tendency to buy?
6. How do we personalize by selecting the groups of customers to benefit from the rewards offered?

The selection criteria:

- Is the problem complex?
- Does it change over time?
- Does it need a solution that scales?
- Do users need an explainable approach?
- Does it need to be 100% accurate?
- Is it ethical to use ML?
- Can it be solved with ML?

Our Hypothesis

We believe retailers have a problem deciding from a plethora of inventory which products out of possibly 10,000 to select for their weekly loyalty promotions. If we propose a tech-based solution to automate this process, we can expect improved productivity and an optimized revenue increase.

Framing the ML Hypothesis

Identifying and Validating the assumptions

We believe retailers have a problem deciding from a plethora of inventory which products out of possibly 10,000 to select for their weekly loyalty promotions. If we propose a tech-based solution to automate this process, we can expect improved productivity and an optimized revenue increase.

Assumption #1 Retailers struggle to decide which products to select for weekly loyalty promotions.		**How to test**
Why?	This assumes retailers do not already have effective systems or methods in place for selecting promotional items.	**Conduct surveys or interviews** with a diverse group of retailers to understand their current processes for selecting promotional items and the challenges they face. This can validate the need for a new solution.
Why?	If retailers already have efficient systems, introducing a new solution might not present enough value to warrant a switch.	
Why?	Adoption of a new technology depends significantly on the perceived improvement over existing processes.	
Assumption #2 An automated tech-based solution can effectively determine the best products for promotions.		**How to test**
Why?	Assumes that technology can analyze and interpret data accurately enough to make promotional decisions better than or equivalent to humans.	**Develop a minimum viable product (MVP)** that uses historical sales, inventory, and customer preference data to recommend promotional items. Measure the accuracy of its recommendations against human selections and its impact on sales in a pilot program.
Why?	The effectiveness of the technology relies on the quality of data, the relevance of the factors considered, and the ability to predict customer behavior.	
Why?	Incorrect predictions or recommendations can lead to reduced sales and dissatisfaction among both retailers and customers.	

Framing the ML Hypothesis

Identifying and Validating the assumptions

We believe retailers have a problem deciding from a plethora of inventory which products out of possibly 10,000 to select for their weekly loyalty promotions. If we propose a tech-based solution to automate this process, we can expect improved productivity and an optimized revenue increase.

Assumption #3 Improved productivity and optimized revenue increase will follow the implementation of this solution.		**How to test**
Why?	Assumes that selecting optimal items for promotions directly correlates with significant productivity improvements and revenue increases.	**Implement the solution with a control group and a test group among participating retailers**. Compare changes in productivity metrics (e.g., time spent on promotional planning) and revenue growth between the groups over a significant period to assess the impact of the solution.
Why?	Productivity gains and revenue increases depend not only on the choice of promotional items but also on other factors like market trends, customer preferences, and competitive actions.	
Why?	The solution might optimize the selection process but does not guarantee overall business performance improvement if other external factors are not addressed.	
Assumption #4 Retailers and customers will adopt and engage with the new loyalty program platform.		**How to test**
Why?	Assumes that both retailers and customers see enough value in the platform to change their behavior or existing systems.	**Create a pilot program with interested retailers** to introduce the platform to a segment of their customers. Track adoption rates, usage patterns, and feedback from both retailers and customers to assess engagement levels and identify areas for improvement.
Why?	Retailers might be resistant to adopting new technology due to costs, complexity, or disruption to existing processes. Customers might not engage if the benefits are unclear or if it requires effort to participate.	
Why?	Adoption rates directly impact the platform's effectiveness and its ability to generate data necessary for optimizing promotions.	

The Product Goal

Our goal is to develop a machine learning model that can analyze historical transaction data, customer behavior, and inventory levels to predict which products will maximize engagement in the weekly loyalty promotions yet tailored to specific customer groups, leading to at least 5% increased sales month-on-month and at least 85% net customer satisfaction from participants. This system will automate the selection process for promotional items, making it more efficient and data-driven.

Framing the ML task

Our problem is best framed as multidimensional regression (combines recommendation systems and predictive analytics) and multi-class classification or clustering (groups customers into segments by their likelihood to purchase).

The model recommends a personalized list of products for promotions and predicts revenue increase for retailers from the likelihood of customers to purchase when fed with historical sales data, inventory levels, customer behaviors, and other relevant factors like seasonality.

Our Assessment metrics

Our success metrics are

1. Increase in participation rate in the loyalty program promotions week-over-week.
2. Improvement in sales revenue from promoted products compared to similar periods without promotions.
3. Enhancement in customer satisfaction scores related to the loyalty program.
4. Reduction in the time retailers spend selecting promotional items.

Our model is deemed a failure if...

1. Decrease or no change in participation rate in the loyalty program promotions.
2. No significant change or decrease in sales revenue from promoted products during promotion periods.
3. Negative feedback or no improvement in customer satisfaction scores concerning loyalty program promotions.
4. Retailers find the process of using the system more cumbersome than manual selection.

The ML Model Output

Expected Model Output

The model should output a ranked list of products predicted to be the most effective for inclusion in the next weekly loyalty promotion.
Each product recommendation should come with a confidence score or expected impact metric to aid retailers in decision-making.

The ranked output would be selected based on customer groups, to ensure recommendations are personalized to individuals.

Output Availability (or timing)

The output from the ML model should be available well in advance of the upcoming promotional period, ideally at least a week before, to allow retailers sufficient time for planning and preparation.

This timing ensures that inventory and marketing strategies can be aligned with the promotion.

Expected Output Usage

Retailers will use the model's output to select which products to include in their weekly loyalty promotions in-store. The ranked list will guide them in prioritizing items that are predicted to yield the highest engagement and sales uplift.

Additionally, the system will update the loyalty recommendation system in the digital commerce channels (mobile apps & websites) once the retailers permit it.

Simplifying the ML Problem into operatable Models

S/n	Model Name & Type	Objective	Input Features	Output/ Target	Model Evaluation Metrics
1	**Product Popularity (or Demand) Prediction** Unidimensional Regression	Predict the future sales volume or popularity of each product based on historical sales data, seasonality, and current market trends.	Historical sales data, time of year, price, promotional history, and product category.	A continuous value representing the expected sales volume or popularity score for each product.	RMSE (Root Mean Square Error) or MAE (Mean Absolute Error)
2	**Customer Purchase Intent** Binary Classification	Predict whether a specific customer segment is likely to purchase a given product if included in the weekly promotion.	Customer demographic information, past purchase history, engagement with previous promotions, and product information.	A binary value (Yes/No) indicates the likelihood of purchase.	F1 Score both the accuracy of the positive predictions and the coverage of actual positive cases are equally important
3	**Product Affinity** Multi-class Single-label Classification or Clustering	Identify groups of products that are often purchased together or have a high affinity.	Purchase history, product categories, and customer feedback.	Class labels or clusters indicating product groupings with high affinity.	F1 Score To measure the model's accuracy across multiple product categories.

Simplifying the ML Problem into operatable Models

S/n	Model Name & Type	Objective	Input Features	Output/ Target	Model Evaluation Metrics
4	**Promotion Effectiveness** Multidimensional Regression	Predict the expected uplift in sales and customer engagement for selected products when included in promotions.	Product information, historical promotion performance, competitor activity, and market trends.	Multiple continuous values represent the expected uplift in sales, customer engagement, and other relevant metrics.	RMSE (Root Mean Square Error) or MAE (Mean Absolute Error)
5	**Inventory Optimization** Unidimensional Regression or Multidimensional Regression	Predict optimal inventory levels for promoted products to meet expected demand without overstocking.	Historical sales and promotion data, lead times, supplier constraints, and storage costs.	A continuous value (or values) indicates the optimal inventory level for each product.	RMSE (Root Mean Square Error) or MAE (Mean Absolute Error)
6	**Loyalty Price Optimization** Multidimensional Regression	Identify the price point that balances increased loyalty participation with sustained profit margins.	[Leverage other itemized models including historical engagement data]	Optimal discount rate/ amount and predicted sale price post-discount	RMSE (Root Mean Square Error) or MAE (Mean Absolute Error)

Data Sourcing for MVP

To get development data source, we can start with data set from the following channels:

Historical Sales Data, Customer Purchase History, and Demographics dataset

1. **Retail Data Analytics dataset on Kaggle** - Contains historical sales data for 45 stores across different departments [1]
2. **E-Commerce Sales Data** - Comprehensive dataset with sales data across channels and financial information [2]
3. **Retail Data Databases & Datasets from Datarade** - Curated selection of top retail data sources, including market research reports and e-commerce trends [3]

Citations

[1] https://www.iguazio.com/blog/13-best-free-retail-datasets-for-machine-learning/

[2] https://www.kaggle.com/datasets/manjeetsingh/retaildataset

[3] https://datarade.ai/data-categories/retail-data

Damola Adediran

Thank you!

For reading my presentation

Adedamola Adediran

https://adedirandamola.com/
https://github.com/adexdams