# APPLICATION DEVELOPMENT REPORT


**Department of Computer Science & Engineering**

**MALLA REDDY UNIVERSITY, HYDERABAD**


**2023-2024**

# Forecasting Coupon Redemption

**Designed and Developed by**


1. R.Nagesh 2111CS010303
2. D.Padma Srija 2111CS010347
3. K.Nithin Varma 2111CS010333


**Guided by**

**Dr. Raviteja Kocherla**


## Department of Computer Science & Engineering

**MALLA REDDY UNIVERSITY, HYDERABAD**


**2023-2024**

# MALLA REDDY UNIVERSITY

(Telangana State Private Universities Act No.13 of 2020 and G.O.Ms.No.14, Higher Education (UE) Department)

## CERTIFICATE

This is to certify that this is the Application development lab record entitled "**Forecasting Coupon Redemption**", submitted by **R.NAGESH – 2111CS010303, D.PADMA SRIJA-2111CS010347, K.NITHIN VARMA-2111CS010333,**B. Tech **III** year **II** semester, Department of CSE during the year 2023-24. The results embodied in this report have not been submitted to any other university or institute for the award of any degree or diploma.

| | |
|---|---|
| **Internal Guide** | **HOD-CSE** |
| **Dr. Raviteja Kocherla** | **Dr. Shaik Meeravali** |

**External Examiner**

MALLA REDDY UNIVERSITY
(Telangana State Private Universities Act No.13 of 2020 and G.O.Ms.No.14, Higher Education (UE) Department)

# DECLARATION

I declare that this project report titled **Forecasting Coupon Redemption** submitted in partial fulfillment of the degree of B. Tech in CSE is a record of original work carried out by me under the supervision of **Dr. RAVITEJA KOCHERLA,** and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgements have been made wherever the findings of others have been cited.


1. R.Nagesh 2111CS010303
2. D.Padma Srija 2111CS010347
3. K.Nithin Varma 2111CS010333

ACKNOWLEDGEMENT

We extend our sincere gratitude to all those who have contributed to the completion of this project report. Firstly, we would like to extend our gratitude to Dr**. V. S. K Reddy, Vice-Chancellor**, for his visionary leadership and unwavering commitment to academic excellence.

We would also like to express my deepest appreciation to our project guide **Dr. Raviteja Kocherla,** who's invaluable guidance, insightful feedback, and unwavering support have been instrumental throughout the course of this project for successful outcomes.

We are also grateful to **Dr. Shaik Meeravali, Head of the Department of CSE,** for providing us with the necessary resources and facilities to carry out this project.

We would like to thank **Dr. Kasa Ravindra**, **Dean, School of Engineering**, for his encouragement and support throughout my academic pursuit.

My heartfelt thanks also go to **Dr. Harikrishna Kamatham, Associate Dean School of Engineering** for his guidance and encouragement.

We are deeply indebted to all of them for their support, encouragement, and guidance, without which this project would not have been possible.

# ABSTRACT

In today's competitive market, understanding customer behavior and optimizing marketing strategies are paramount for businesses. One crucial aspect of marketing is the distribution of coupons, which can significantly influence consumer purchasing decisions. Predicting coupon redemption rates accurately can empower businesses to tailor their marketing efforts effectively and maximize their return on investment. In this study, we propose a machine learning approach for forecasting coupon redemption. Leveraging historical coupon redemption data and various machine learning algorithms, including but not limited to decision trees, random forests, and neural networks, we aim to develop models capable of predicting the likelihood of coupon redemption by customers. By analyzing past redemption patterns and relevant customer attributes, our models seek to provide actionable insights for businesses to optimize coupon distribution strategies and enhance marketing effectiveness. Through empirical evaluation on real-world coupon redemption datasets, we demonstrate the efficacy of our approach in accurately forecasting coupon redemption rates, thereby assisting businesses in making informed decisions to drive sales and customer engagement. The client’s promotions are shared across various channels including email, notifications, etc. A number of these campaigns include coupon discounts that are offered for a specific product/range of products. The retailer would like the ability to predict whether customers redeem the coupons received across channels, which will enable the retailer’s marketing team to accurately design coupon construct, and develop a more precise and targeted marketing strategies.

# TABLE OF CONTENT

# LIST OF FIGURES

# CHAPTER – 1

# INTRODUCTION

# Forecasting Coupon Redemption

## 1.1 Introduction to Coupon Redemption

Coupons offer discounts or incentives to consumers, serving as catalysts for purchase decisions and brand engagement.Predicting which customers are likely to redeem coupons and under what circumstances can enable businesses to optimize their marketing strategies and allocate resources effectively.the advent of machine learning offers a promising avenue for enhancing the precision and efficiency of coupon distribution. By leveraging historical data and advanced algorithms, machine learning empowers businesses to uncover intricate patterns and insights hidden within vast datasets.we seek to develop predictive models capable of accurately estimating the likelihood of coupon redemption. Our approach entails the utilization of various machine learning techniques, including decision trees, random forests, and neural networks, to extract actionable insights from coupon redemption data.Through empirical evaluation on real-world coupon redemption datasets, we aim to demonstrate the effectiveness of our proposed approach in forecasting coupon redemption rates. By providing businesses with reliable predictions and actionable insights, our study aims to equip marketers with the tools necessary to optimize coupon distribution strategies, enhance customer engagement, and drive sales growth.

Discount marketing and coupon usage are very widely used promotional techniques to attract new customers and to retain & reinforce loyalty of existing customers. The measurement of a consumer's propensity towards coupon usage and the prediction of the redemption behaviour are crucial parameters in assessing the effectiveness of a marketing campaign.ABC promotions are shared across various channels including email, notifications, etc. A number of these campaigns include coupon discounts that are offered for a specific product/range of products. The retailer would like the ability to predict whether customers redeem the coupons received across channels, which will enable the retailer's marketing team to accurately design coupon construct, and develop more precise and targeted marketing strategies.XYZ Credit Card Company collaborates with ABC, an established Brick & Mortar retailer, to enhance their discount marketing process using machine learning.

ABC frequently conducts marketing campaigns, including coupon discounts, to attract new customers and retain existing ones. The effectiveness of these campaigns hinges on predicting customer coupon redemption behavior.

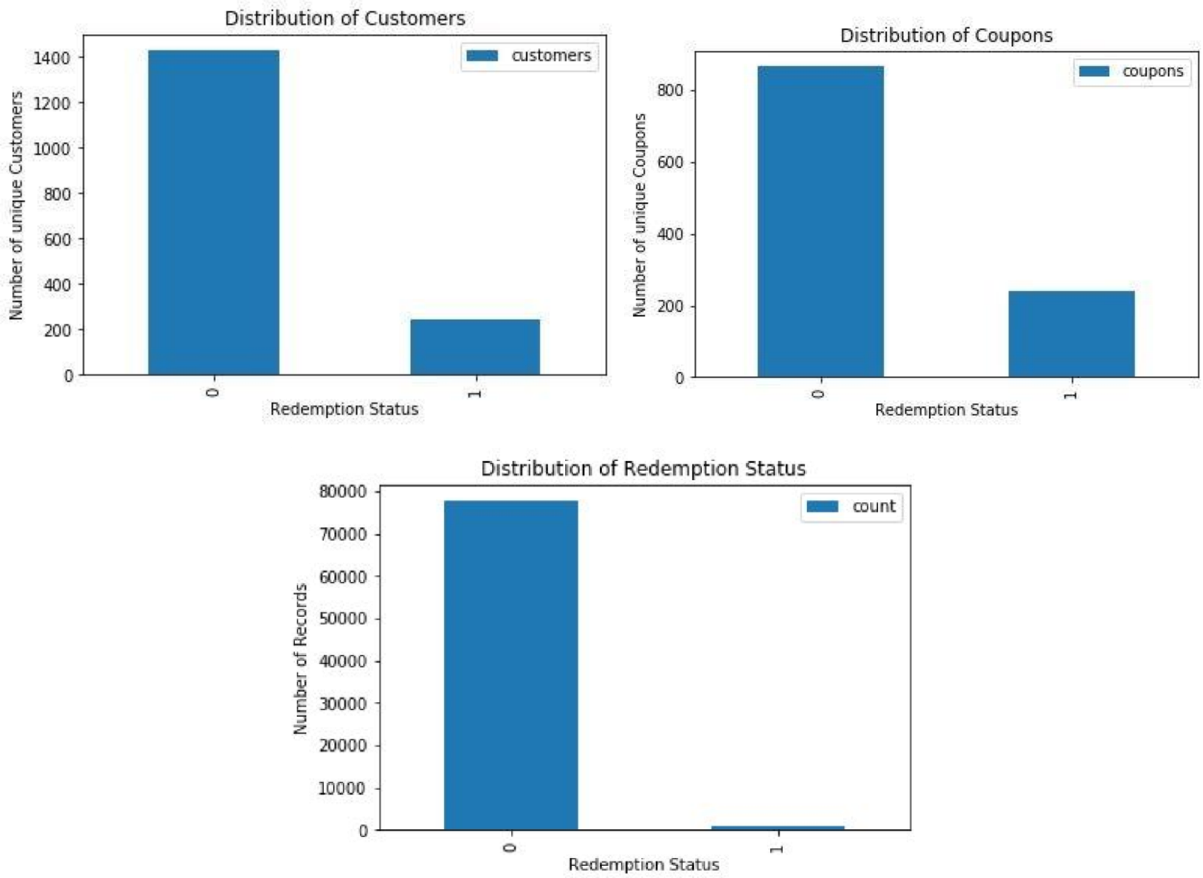



**Fig: 1.1: REDEMPTION Analysis**

## 1.2 Problem Statement

Discount marketing and coupon usage are very widely used promotional techniques to attract new customers and to retain & reinforce loyalty of existing customers. The measurement of a consumer’s propensity towards coupon usage and the prediction of the redemption behaviour are crucial parameters in assessing the effectiveness of a marketing campaign. The client’s promotions are shared across various channels including email, notifications, etc. A number of these campaigns include coupon discounts that are offered for a specific product/range of products. The retailer would like the ability to predict whether customers redeem the coupons received across channels, which will enable the retailer’s marketing team to accurately design coupon construct, and develop a more precise and targeted marketing strategies. The strategic distribution of coupons is a widely employed marketing tactic aimed at incentivizing consumer purchases and fostering brand loyalty. However, the efficacy of coupon campaigns hinges on the ability to accurately predict coupon redemption rates. Without precise forecasts, businesses risk misallocating resources, distributing coupons to customers unlikely to redeem them, and missing opportunities to maximize return on investment. Traditional approaches to forecasting coupon redemption often rely on simplistic heuristics or manual analysis, which may overlook subtle patterns and fail to capture the complex interplay of factors influencing consumer behaviour. Moreover, the sheer volume and complexity of coupon redemption data present significant challenges for conventional analytical techniques.

## 1.3 Objective of the project

The primary objective of the coupon redemption project using data analytics is to predict whether customers will redeem coupons. By analyzing historical data, we aim to understand customer behavior and create accurate models that empower marketing teams to design targeted strategies. The project's success lies in optimizing coupon constructs and improving overall campaign effectiveness

1. **Reinventing Coupons: Strategies for a Successful Coupon Campaign**:

2. In 2012, marketers distributed approximately 310 billion valued coupons, totaling $484 billion[1].

3. Coupons are a powerful sales tool, influencing buying patterns and driving sales.
    a. Key considerations for successful coupon campaigns:
        i. **Attracting New Customers**: Coupons can acquire new customers and encourage existing ones to spend more.
        ii. **Balancing Margins**: Poorly planned coupon promotions may hurt margins due to revenue cannibalization.
        iii. **Coupon Design**: Critical for maximizing profits and minimizing cannibalization.
        iv. **Clarity on Marketing Goals**: Coupons should align with marketing objectives[1].

4. **Predictive Analytics for E-Coupon Redemption**:
    a. Analyzing data from an insurance e-commerce platform in East China, researchers examined factors influencing e-coupon redemption.
    b. Behavioral predictive analytics tools were used to understand customer behavior after receiving coupons[2].
5. **Coupon Acquisition and Redemption Metrics**:
    a. Businesses can optimize coupon strategies by analyzing redemption rates.
    b. Factors affecting redemption include discount value, expiration date, and ease of redemption

## 1.4 Goal of the project

The primary goal of a coupon redemption project using data analytics is to predict customer coupon redemption behavior. By analyzing historical data, we aim to understand patterns, optimize coupon constructs, and improve overall campaign effectiveness. This empowers marketing teams to design targeted strategies and enhance the success of discount marketing campaigns. Traditional approaches to forecasting coupon redemption often rely on simplistic heuristics or manual analysis, which may overlook subtle patterns and fail to capture the complex interplay of factors influencing consumer behavior. Moreover, the sheer volume and complexity of coupon redemption data present significant challenges for conventional analytical techniques. By analyzing historical data, we aim to understand customer behavior and create accurate models that empower marketing teams to design targeted strategies. The project's success lies in optimizing coupon constructs and improving overall campaign effectiveness. Optimize model performance by adjusting hyperparameters (e.g., learning rate, max depth). Conduct controlled experiments to compare different coupon strategies. Oversampling: Creating synthetic instances of the minority class.Undersampling: Reducing instances of the majority class.SMOTE (Synthetic Minority Over-sampling Technique): A combination of oversampling and interpolation.

# CHAPTER – 2
# PROBLEM IDENTIFICATION

## 2.1Existing System

The existing system for this project consists of the following components:

- Create personalized coupon-based campaigns.
- Upload coupons from third-party systems via CSV or API.
- Autogenerate custom coupons with various attributes (shape, expiration date, discount percentage).
- Set custom coupon redemption rules.
- Distribute digital coupons across channels (email, SMS, mobile apps, etc.) via API or webhooks.
- Prevent fraud by defining coupon limits.
- Extend existing system coupon capabilities.

## 2.2 PROPOSED SYSTEM

**Enterprise Blockchain for Digital Coupons:**

- Objective: Improve coupon creation, distribution, and redemption processes.
- Challenges Addressed:
- Complexity: Traditional coupon management involves multiple parties and steps.
- Fraud: Digital fraud limits coupon adoption.
- Key Features:
- Secure Distribution: Blockchain ensures tamper-proof coupon issuance and distribution.
- Deals Marketplace: Customers can search for deals based on various criteria.
- Loyalty Store: Exchange loyalty points for digital coupons.
- Referral Program: Reward users for referring others.
- Technology: Enterprise or permissioned blockchains (e.g., Ethereum).
- **Secure Distribution**:Coupons are securely issued and distributed via blockchain
- Immutable records prevent fraud.

**Deals Marketplace**:

- Customers search for deals on coupon sites.
- Blockchain ensures transparency and trust.

**Loyalty Store**:

- Loyalty points exchanged for digital coupons.
- Smart contracts handle redemption.

## 2.3 OVERALL DESCRIPTION

**Data Collection:**

- Gather historical data on customer interactions with coupons. This dataset should include information such as:
- Customer demographics (age, gender, location, etc.)
- Coupon details (discount percentage, expiration date, type of coupon, etc.)
- Channel through which the coupon was delivered (email, app notification, etc.)
- Customer behavior (purchases, visits, etc.) before and after coupon receipt
- Redemption status (whether the customer used the coupon or not)

**Data Preprocessing**:

- Clean the data by handling missing values, outliers, and inconsistencies.
- Feature engineering:
- Create relevant features such as:
- Time since last purchase/visit
- Total spending by the customer
- Frequency of coupon usage
- Interaction history with previous coupons

**Exploratory Data Analysis (EDA)**:

1. Visualize the data to gain insights:

    - Distribution of coupon redemption (target variable)
    - Correlations between features and redemption
    - Seasonal trends in coupon usage
    - Identify patterns and anomalies.

2. Model Building:
    - Split the data into training and validation sets.
    - Choose appropriate machine learning algorithms (e.g., logistic regression, random forests, gradient boosting).
    - Train the models using the training data.
    - Evaluate model performance using metrics like accuracy, precision, recall, and F1-score.
3. Feature Importance:
    - Determine which features contribute most to coupon redemption.
    - Understand the impact of different variables on the outcome.
4. Model Interpretability:
    - Use techniques like SHAP (SHapley Additive explanations) to explain model predictions.
    - Understand why certain customers are more likely to redeem coupons.

# CHAPTER- 3

# REQUIREMENTS

## 3.1 Software Requirements

### Functional Requirements:

- **Coupon Management**:
  - Create new coupons.
  - Allocate coupons to users or specific demographics.
  - Manage coupon activation and deactivation.
  - Define rules (e.g., age restrictions, minimum cart value).
  - Set expiration dates.
- **User Management**:
  - Authentication, registration, and profile management.
  - User roles and permissions (e.g., regular users vs. administrators).
- **Voucher Management**:
  - Allocate vouchers.
  - Track voucher redemption status.
- **Administrative Capabilities**:
  - Create, activate, and deactivate coupons.
- **Security Mechanisms**:
  - Robust authentication to protect user data.
- **Performance and Scalability**:
  - Handle concurrent users and transactions efficiently.

### Non-Functional Requirements:

- **Performance**:
  - Handle large user loads without significant degradation.
- **Scalability**:
  - Accommodate growth in user base and transaction volume.

- **Reliability**:
  - Ensure high availability and data integrity.
- **Security**:
  - Prevent unauthorized access.

## Technical Requirements:

**1.Software:** The system should be built using a programming language such as Python or R.

**2.Database:** The system should be able to store and retrieve large datasets of financial data.

**3.Statistical Analysis Software:** The system should be able to use statistical analysis software such as Excel or Python for data analysis.

**4.Data Visualization Tools:** The system should be able to use data visualization tools such as Tableau or Power BI to present complex data in a clear and concise manner.

**5.Operating System:** The system should be compatible with a variety of operating systems, including Windows, macOS, and Linux.

## Infrastructure Requirements:

**1.Hardware:** The system should be able to run on a high-performance computing infrastructure with multiple CPU cores and ample memory.

**2.Storage:** The system should have access to a large storage capacity to store and retrieve large datasets of financial data.

**3.Network:** The system should have a reliable and fast network connection to ensure seamless communication with external databases and data sources.

# CHAPTER- 4

# Design and Implementation

## 4.1 Design

Data Merging:-

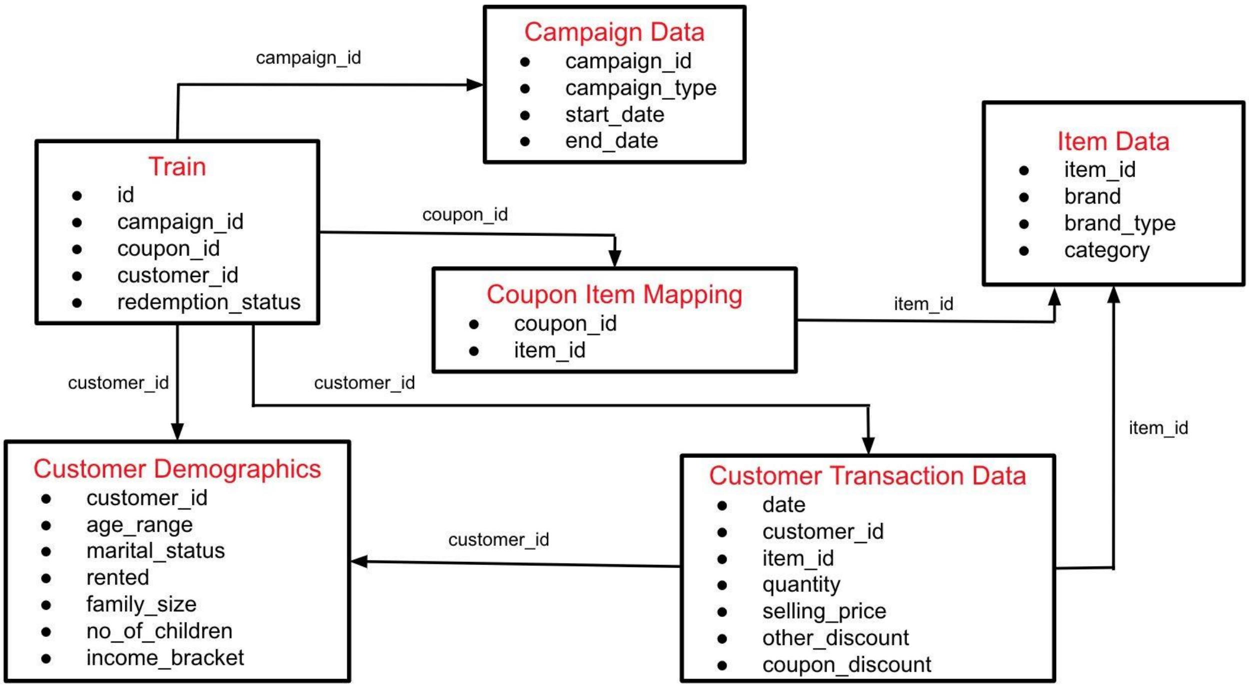



**Fig. 4.1: Design**

Discount marketing and coupon usage are very widely used promotional techniques to attract new customers and to retain & reinforce loyalty of existing customers. The measurement of a consumer's propensity towards coupon usage and the prediction of the redemption behaviour are crucial parameters in assessing the effectiveness of a marketing campaign.ABC promotions are shared across various channels including email, notifications, etc. A number of these campaigns include coupon discounts that are offered for a specific product/range of products. The retailer would like the ability to predict whether customers redeem the coupons received across channels, which will enable the retailer's marketing team to accurately design coupon construct, and develop more precise and targeted marketing strategies.

XYZ Credit Card Company collaborates with ABC, an established Brick & Mortar retailer, to enhance their discount marketing process using machine learning. ABC frequently conducts marketing campaigns, including coupon discounts, to attract new customers.

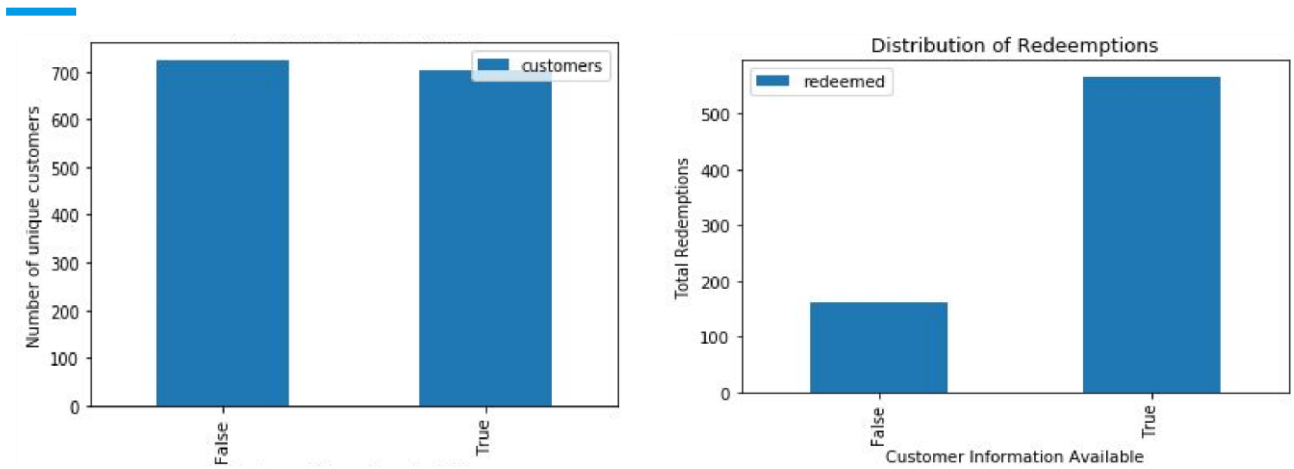



**Fig. 4.2: Return Values**

**Data Cleaning**:-

**After analysing the dataset, below issues were found.**

**• Missing Data**

- no_of_children: Assuming it to be Zero
- marital_status: If family_size - no_of_children > 1
- then Married else Single
- Customers without Information: Handle them with algorithm

**• Outliers**

- Many outliers in the customer's transactions
- Prediction algorithm has to be trained with outliers

**Data Merging:-**

**Coupon Information**

- Extracting summary variables from coupon information

**Customer Behaviour**

- Extracting summary variables from customer's transactions

# CHAPTER- 5

# CODE

## 5.1 Source Code

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
```

# Exploratory Data Analysis

Initial Questions

1. Distribution of redemption status with respect to coupon, customer and campaign?
2. What is the trend in the campaign duration?
3. How many customers whose information is not present?
4. What is the trend in the Customer transactions?

### Redemptions

```
In [2]: trainset = pd.read_csv("../data/train/train.csv", index_col='id')
        trainset.head()
```

Out[2]:

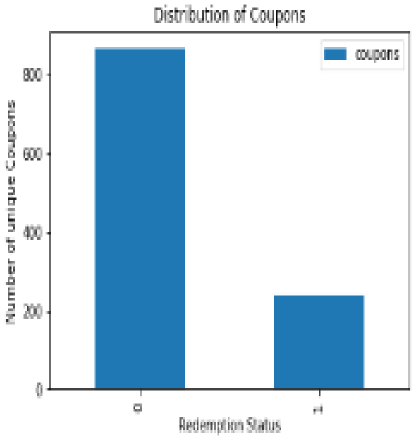
| id | campaign_id | coupon_id | customer_id | redemption_status |
|---|---|---|---|---|
| 1 | 13 | 27 | 1053 | 0 |
| 2 | 13 | 116 | 48 | 0 |
| 6 | 9 | 635 | 205 | 0 |
| 7 | 13 | 644 | 1050 | 0 |
| 9 | 8 | 1017 | 1489 | 0 |

```
In [3]: group_by_redemption = trainset.groupby('redemption_status').agg({'coupon_id': 'nunique', 'customer_id': 'nunique', 'redemption_st
        group_by_redemption.columns = ['coupons', 'customers', 'count']
        group_by_redemption = group_by_redemption.reset_index()
        group_by_redemption
```

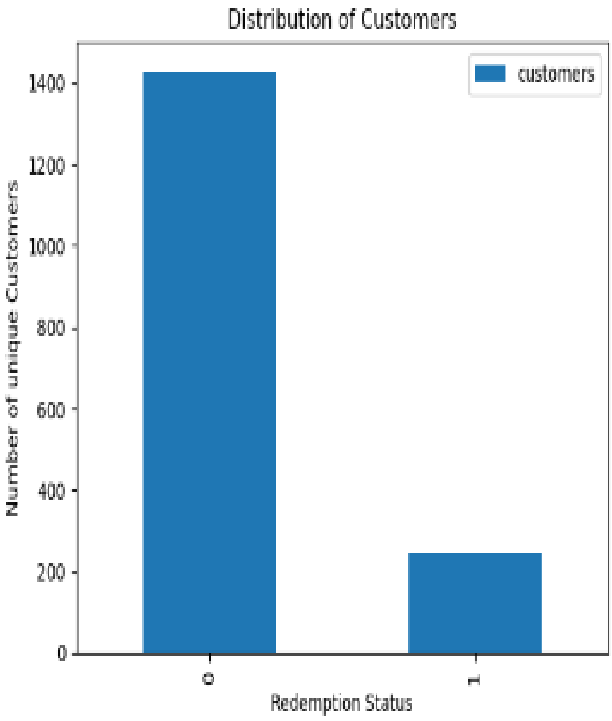
Out[3]:

| | redemption_status | coupons | customers | count |
|---|---|---|---|---|
| 0 | 0 | 866 | 1428 | 77640 |
| 1 | 1 | 239 | 247 | 729 |

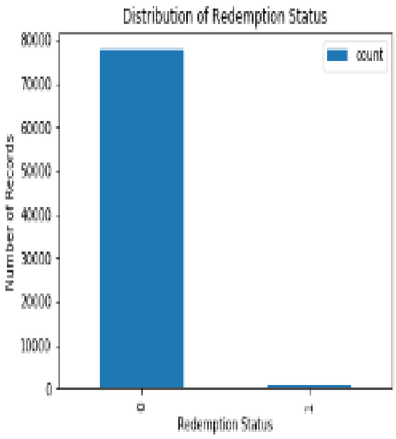
In [4]:
```
group_by_redemption.plot('redemption_status', 'coupons', kind='bar')
_ = plt.xlabel("Redemption Status")
_ = plt.ylabel("Number of unique Coupons")
_ = plt.title("Distribution of Coupons")
plt.savefig('figure/coupon_distribution.png', bbox_inches="tight")
plt.show()
```




In [4]:
```
group_by_redemption.plot('redemption_status', 'customers', kind='bar')
_ = plt.xlabel("Redemption Status")
_ = plt.ylabel("Number of unique Customers")
_ = plt.title("Distribution of Customers")
plt.savefig('figure/customer_distribution.png', bbox_inches="tight")
plt.show()
```



```
In [6]: group_by_redemption.plot('redemption_status', 'count', kind='bar')
_ = plt.xlabel("Redemption Status")
_ = plt.ylabel("Number of Records")
_ = plt.title("Distribution of Redemption Status")
plt.savefig('figure/redemption_distribution.png', bbox_inches="tight")
plt.show()
```




Below are the findings from the above distributions

1. There are 627 coupons which are not redeemed by any customers
2. There are 1181 customers who has not redeemed any coupons
3. The data is highly Imbalanced

## Campaigns

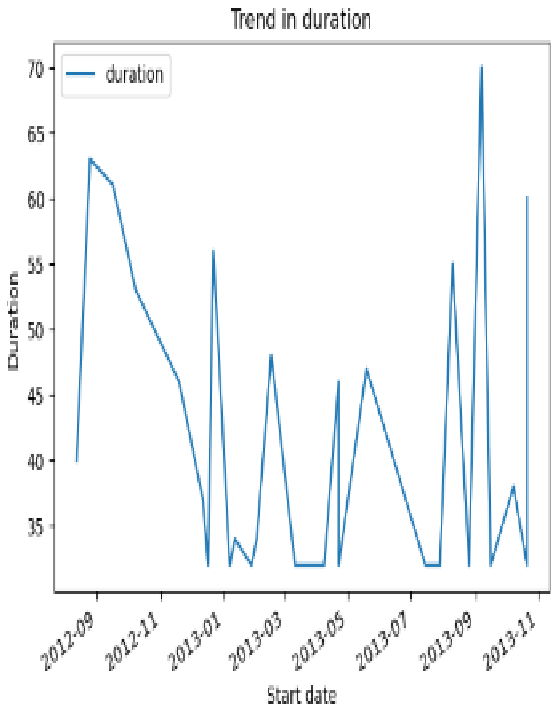
```
In [5]: campaign_data = pd.read_csv('../data/train/campaign_data.csv', parse_dates=['start_date', 'end_date'], dayfirst=True)
campaign_data['duration'] = (campaign_data['end_date'] - campaign_data['start_date']).dt.days
campaign_data = campaign_data.sort_values('start_date')
campaign_data.head()
```

```
C:\Users\RAGOJI NAGESH\AppData\Local\Temp\ipykernel_6720\3631248235.py:1: UserWarning: Could not infer format, so each element
will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a form
at.
  campaign_data = pd.read_csv('../data/train/campaign_data.csv', parse_dates=['start_date', 'end_date'], dayfirst=True)
C:\Users\RAGOJI NAGESH\AppData\Local\Temp\ipykernel_6720\3631248235.py:1: UserWarning: Could not infer format, so each element
will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a form
at.
  campaign_data = pd.read_csv('../data/train/campaign_data.csv', parse_dates=['start_date', 'end_date'], dayfirst=True)
```

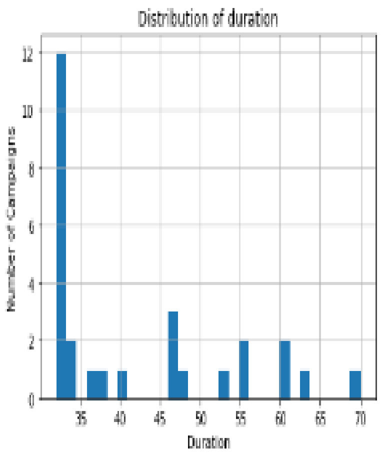
Out[5]:

| | campaign_id | campaign_type | start_date | end_date | duration |
|---|---|---|---|---|---|
| **27** | 26 | X | 2012-08-12 | 2012-09-21 | 40 |
| **26** | 27 | Y | 2012-08-25 | 2012-10-27 | 63 |
| **25** | 28 | Y | 2012-09-16 | 2012-11-16 | 61 |
| **24** | 29 | Y | 2012-10-08 | 2012-11-30 | 53 |
| **23** | 30 | X | 2012-11-19 | 2013-01-04 | 46 |

```
In [6]: campaign_data.plot('start_date', 'duration')
_ = plt.xlabel("Start date")
_ = plt.ylabel("Duration")
_ = plt.title("Trend in duration")
plt.savefig('figure/duration_trend.png', bbox_inches="tight")
plt.show()
```




```
In [9]: campaign_data.duration.hist(bins=30)
_ = plt.xlabel("Duration")
_ = plt.ylabel("Number of Campaigns")
_ = plt.title("Distribution of duration")
plt.savefig('figure/duration_distribution.png', bbox_inches="tight")
plt.show()
```



## Customer's Information

In [10]:
```
customer_data = pd.read_csv('../data/train/customer_demographics.csv')
customer_data.head()
```

Out[10]:

| | customer_id | age_range | marital_status | rented | family_size | no_of_children | income_bracket |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 70+ | Married | 0 | 2 | NaN | 4 |
| 1 | 6 | 46-55 | Married | 0 | 2 | NaN | 5 |
| 2 | 7 | 26-35 | NaN | 0 | 3 | 1 | 3 |
| 3 | 8 | 26-35 | NaN | 0 | 4 | 2 | 6 |
| 4 | 10 | 46-55 | Single | 0 | 1 | NaN | 5 |

In [11]:
```
customer_info = trainset.merge(customer_data, how='left', on='customer_id')
customer_info['info_available'] = customer_info.family_size.notna()
customer_info.head()
```

Out[11]:

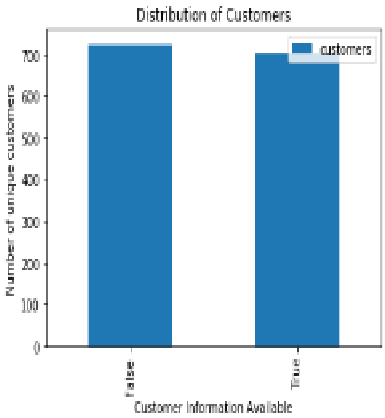
| | campaign_id | coupon_id | customer_id | redemption_status | age_range | marital_status | rented | family_size | no_of_children | income_bracket | info_available |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 13 | 27 | 1053 | 0 | 46-55 | NaN | 0.0 | 1 | NaN | 5.0 | True |
| 1 | 13 | 116 | 48 | 0 | 36-45 | Married | 0.0 | 2 | NaN | 3.0 | True |
| 2 | 9 | 635 | 205 | 0 | 46-55 | Married | 0.0 | 2 | NaN | 7.0 | True |
| 3 | 13 | 644 | 1050 | 0 | NaN | NaN | NaN | NaN | NaN | NaN | False |
| 4 | 8 | 1017 | 1489 | 0 | 46-55 | Married | 0.0 | 2 | NaN | 3.0 | True |

In [12]:
```
group_by_available = customer_info.groupby('info_available').agg({'customer_id': 'nunique', 'redemption_status': 'sum'})
group_by_available.columns = ['customers', 'redeemed']
group_by_available = group_by_available.reset_index()
group_by_available
```

Out[12]:

| | info_available | customers | redeemed |
|---|---|---|---|
| 0 | False | 725 | 161 |
| 1 | True | 703 | 568 |

In [13]:
```
group_by_available.plot('info_available', 'customers', kind='bar')
_ = plt.xlabel('Customer Information Available')
_ = plt.ylabel('Number of unique customers')
_ = plt.title('Distribution of Customers')
plt.savefig('figure/information_customers_distribution.png', bbox_inches="tight")
plt.show()
```



## Customer transactions

In [15]:
```
transaction_data = pd.read_csv('../data/train/customer_transaction_data.csv', parse_dates=['date'])
transaction_data['month'] = transaction_data.date.dt.month + transaction_data.date.dt.year.replace({2012: 0, 2013: 12})
transaction_data['coupon_discount'] = transaction_data['coupon_discount'] < 0
transaction_data.head()
```

Out[15]:

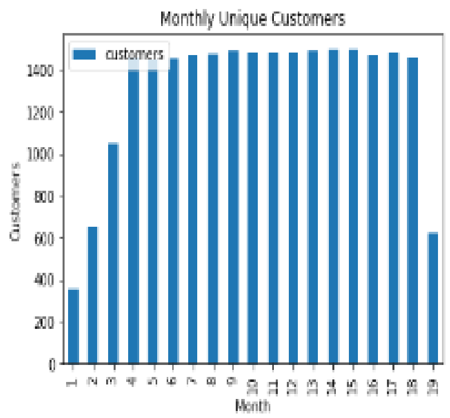
| | date | customer_id | item_id | quantity | selling_price | other_discount | coupon_discount | month |
|---|---|---|---|---|---|---|---|---|
| 0 | 2012-01-02 | 1501 | 26830 | 1 | 35.26 | -10.69 | False | 1 |
| 1 | 2012-01-02 | 1501 | 54253 | 1 | 53.43 | -13.89 | False | 1 |
| 2 | 2012-01-02 | 1501 | 31962 | 1 | 106.50 | -14.25 | False | 1 |
| 3 | 2012-01-02 | 1501 | 33647 | 1 | 67.32 | 0.00 | False | 1 |
| 4 | 2012-01-02 | 1501 | 48199 | 1 | 71.24 | -28.14 | False | 1 |

In [16]:
```
group_by_date = transaction_data.groupby('month').agg({'customer_id': 'nunique', 'item_id': 'nunique', 'quantity': 'count', 'cou
group_by_date.columns = ['customers', 'items', 'count', 'coupon_discount']
group_by_date = group_by_date.reset_index()
group_by_date.head()
```

Out[16]:

| | month | customers | items | count | coupon_discount |
|---|---|---|---|---|---|
| 0 | 1 | 351 | 7432 | 12382 | 187.0 |
| 1 | 2 | 654 | 11482 | 25073 | 287.0 |
| 2 | 3 | 1050 | 16307 | 46730 | 476.0 |
| 3 | 4 | 1463 | 19801 | 68340 | 868.0 |
| 4 | 5 | 1444 | 20334 | 79093 | 923.0 |

In [17]:
```
group_by_date.plot('month', 'customers', kind='bar')
_ = plt.xlabel('Month')
_ = plt.ylabel('Customers')
_ = plt.title('Monthly Unique Customers')
plt.savefig('figure/transactions_customers.png', bbox_inches="tight")
plt.show()
```



# Data Cleaning

### Item Information

In [3]:
```
item_data = pd.read_csv('data/train/item_data.csv')
item_data['brand_type'] = item_data['brand_type'].replace({'Established': 1, 'Local': 0})
item_data.head()
```

Out[3]:

| | item_id | brand | brand_type | category |
|---|---|---|---|---|
| 0 | 1 | 1 | 1 | Grocery |
| 1 | 2 | 1 | 1 | Miscellaneous |
| 2 | 3 | 56 | 0 | Bakery |
| 3 | 4 | 56 | 0 | Grocery |
| 4 | 5 | 56 | 0 | Grocery |

In [4]:
```
item_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 74066 entries, 0 to 74065
Data columns (total 4 columns):
item_id       74066 non-null int64
brand         74066 non-null int64
brand_type    74066 non-null int64
category      74066 non-null object
dtypes: int64(3), object(1)
memory usage: 2.3+ MB
```

In [5]:
```
total_items = item_data['item_id'].nunique()
total_brands= item_data['brand'].nunique()
total_brand_types = item_data['brand_type'].nunique()
total_categories = item_data['category'].nunique()

print("total_items: {}".format(total_items))
print("total_brands: {}".format(total_brands))
print("total_brand_types: {}".format(total_brand_types))
print("total_categories: {}".format(total_categories))
```

```
total_items: 74066
total_brands: 5528
total_brand_types: 2
total_categories: 19
```

## Data Preprocessing

Combine train and test data

```
In [23]: columns = train_data.columns[train_data.columns != 'redemption_status']
         total_data = train_data[columns].append(test_data, sort=True)
         total_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 128595 entries, 0 to 50225
Data columns (total 4 columns):
campaign_id    128595 non-null int64
coupon_id      128595 non-null int64
customer_id    128595 non-null int64
id             128595 non-null int64
dtypes: int64(4)
memory usage: 4.9 MB
```

### Data Transform

Transforming Coupon-Item mapping into coupon specific variables

```
In [24]: coupon_data = coupon_item.groupby('coupon_id').agg({
             'item_id': ['nunique'],
             'brand': ['nunique', most_frequent, least_frequent, most_frequent_count, least_frequent_count],
             'brand_type': ['nunique', most_frequent, least_frequent, most_frequent_count, least_frequent_count],
             'category': ['nunique', most_frequent, least_frequent, most_frequent_count, least_frequent_count]
         })
         coupon_data.columns = ['c_unique_items', 'c_unique_brand', 'c_freq_brand', 'c_rare_brand',
                                'c_items_freq_brand', 'c_items_rare_brand', 'c_unique_brandt', 'c_freq_brandt',
                                'c_rare_brandt', 'c_items_freq_brandt', 'c_items_rare_brandt',
                                'c_unique_category', 'c_freq_category', 'c_rare_category', 'c_items_freq_category',
                                'c_items_rare_category']
         coupon_data['c_coverage_item'] = coupon_data['c_unique_items'] / total_items
         coupon_data['c_coverage_brand'] = coupon_data['c_unique_brand'] / total_brands
         coupon_data['c_coverage_brandt'] = coupon_data['c_unique_brandt'] / total_brand_types
         coupon_data['c_coverage_category'] = coupon_data['c_unique_category'] / total_categories

         coupon_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1116 entries, 1 to 1116
Data columns (total 20 columns):
c_unique_items         1116 non-null int64
c_unique_brand         1116 non-null int64
c_freq_brand           1116 non-null int64
c_rare_brand           1116 non-null int64
c_items_freq_brand     1116 non-null int64
```

# CHAPTER- 6

# RESULTS AND CONCLUSION

## 6.1 Results

Coupons offer discounts or incentives to consumers, serving as catalysts for purchase decisions and brand engagement.Predicting which customers are likely to redeem coupons and under what circumstances can enable businesses to optimize their marketing strategies and allocate resources effectively.the advent of machine learning offers a promising avenue for enhancing the precision and efficiency of coupon distribution. By leveraging historical data and advanced algorithms, machine learning empowers businesses to uncover intricate patterns and insights hidden within vast datasets.we seek to develop predictive models capable of accurately estimating the likelihood of coupon redemption. Our approach entails the utilization of various machine learning techniques, including decision trees, random forests, and neural networks, to extract actionable insights from coupon redemption data.Through empirical evaluation on real-world coupon redemption datasets, we aim to demonstrate the effectiveness of our proposed approach in forecasting coupon redemption rates. By providing businesses with reliable predictions and actionable insights, our study aims to equip marketers with the tools necessary to optimize coupon distribution strategies, enhance customer engagement, and drive sales growth.

Discount marketing and coupon usage are very widely used promotional techniques to attract new customers and to retain & reinforce loyalty of existing customers. The measurement of a consumer's propensity towards coupon usage and the prediction of the redemption behaviour are crucial parameters in assessing the effectiveness of a marketing campaign.ABC promotions are shared across various channels including email, notifications, etc. A number of these campaigns include coupon discounts that are offered for a specific product/range of products. The retailer would like the ability to predict whether customers redeem the coupons received across channels, which will enable the retailer's marketing team to accurately design coupon construct, and develop more precise and targeted marketing strategies.

XYZ Credit Card Company collaborates with ABC, an established Brick & Mortar retailer, to enhance their discount marketing process using machine learning. ABC frequently conducts marketing campaigns, including coupon discounts, to attract new customers and retain existing ones.

## 6.2 Conclusion

After performing different experiments with features, missing values, parameters, etc. The ensemble of tuned lightGBM with num_leaves 15, 20 and 25 performs better than others. The evaluation metric of ***area under the ROC curve*** final model is getting the score of **0.9132**. The same model got private score of **0.8979**.

Further Improvements can be done to improve the model. Below are some ideas that can be explored.

- The model uses features, where each feature has a small contribution towards the model performance. Merging the features will result in better speed of model and hence, more experiments can be performed.
- Different way of handling many-to-many relations between coupons and customer's transactions.
- Hyper-parameter tuning might lead to local minima, further tuning can be tried for better results.
- Confusion matrix was not analysed. It can be analysed and fix the models for the issues.

# REFERENCES


1) Blattberg, R. C., & Neslin, S. A., "Coupon Redemption: Theory and Evidence," Journal of Marketing, Vol. 49, 1985.
2) Bawa, K., & Shoemaker, R. W., "Analyzing Incremental Effects of Coupons on Brand Choice Behavior," Journal of Marketing Research, Vol. 24, 1987.
3) Chandon, P., Wansink, B., & Laurent, G., "A Benefit Congruency Framework of Sales Promotion Effectiveness," Journal of Marketing, Vol. 64, 2000.
4) Raghubir, P., "Coupon Value: A Signal for Price?" Journal of Marketing Research, Vol. 35, 1998.
5) Leeflang, P. S. H., & Wittink, D. R., "The Effects of Promotional Restrictions on Actual Coupon Redemption," Journal of Marketing Research, Vol. 39, 2002.
6) Gupta, S., "Impact of Sales Promotions on When, What, and How Much to Buy," Journal of Marketing Research, Vol. 25, 1988.
7) Gedenk, K., Neslin, S. A., & Ailawadi, K. L., "Sales Promotion," in Handbook of Marketing Decision Models, Springer, 2010.
8) Pauwels, K., Hanssens, D. M., & Siddarth, S., "The Long-Term Effects of Price Promotions on Category Incidence, Brand Choice, and Purchase Quantity," Journal of Marketing Research, Vol. 39, 2002.
9) Srinivasan, S., Pauwels, K., Hanssens, D. M., & Dekimpe, M. G., "Do Promotions Benefit Manufacturers, Retailers, or Both?" Management Science, Vol. 50, 2004.
10) DelVecchio, D., Henard, D. H., & Freling, T. H., "The Effect of Sales Promotion on Post-Promotion Brand Preference: A Meta-Analysis," Journal of Retailing, Vol. 82, 2006.