# COUPON PREDICTION

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

**BACHELOR OF TECHNOLOGY**

in

**COMPUTER SCIENCE & ENGINEERING**

by

**T.VISHAL MARY (2103A51495)**

**G.ASHMITHA (2103A51512)**

Under the guidance of

**Mr. S. Naresh Kumar**

Assistant Professor, Department of CSE.



## Department of Computer Science and Artificial Intelligence



**Department of Computer Science and Artificial Intelligence**

**CERTIFICATE**

This is to certify that this project entitled **“COUPON PURCHASE PREDICTION”** is the bonafied work carried out by **T.Vishal Mary,G.Ashmitha** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE & ENGINEERING** during the academic year 2022-2023 under our guidance and Supervision.

**Mr. S Naresh Kumar Dr. M. Sheshikala**

Asst. Prof, Department of CSE. Assoc . Prof .&HOD(CSE),

S R University, S R University,

Ananthasagar, Warangal. Ananthasagar, Warangal.

# ACKNOWLEDGEMENT

We express our thanks to Course co-coordinator **Mr.** **S. Naresh Kumar, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude toHead of the department CS&AI, **Dr. M. Sheshikala , Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

# ABSTRACT

Our project is based on user data analysis whichis used to predict what coupon a user will get based on his previous purchases and transactions from the site. Issuing discount shopping coupons and barter purchase is a popular way to promote sales in E-commerce environments. In order to improve the coupon usage ratio, it is important to predict the probability in which a user will get a coupon. To treat the coupon redemption probability prediction problem as a binary classification problem, we use machine learning methods to analyze users' coupon usage behavior and coupon purchase prediction.

Coupon prediction refers to the process of using data analytics and machine learning techniques to forecast the likelihood of a customer getting a particular coupon or offer. This practice is essential for businesses in the retail and e-commerce industries, as it allows them to optimize their marketing strategies by targeting customers who are most likely to use coupons and offers. By analyzing customer data such as past purchase and transactions, and other relevant factors, businesses can predict which customers are more likely to get coupon promotions and tailor their marketing efforts accordingly. Predictive models can also help businesses optimize the timing and distribution of coupons and offers to maximize their effectiveness. The benefits of coupon prediction include improved targeting, increased redemption rates, and a better return on investment for marketing efforts. Overall, coupon prediction is a valuable tool for businesses looking to optimize their marketing strategies and improve customer engagement.

## Table of Contents

**Chapter No. Title Page No.**

1. Introduction

.1. Overview 1

.2. Problem Statement 1

.3. Existing system 1

.4. Proposed system 2

.5. Objectives 2

.6. Overall architecture 3

1. Literature survey 4-5
2. Data pre-processing 6-7
   1. Describe dataset 8
   2. Data cleaning 9-10
   3. Data augmentation 11
   4. Data Visualization 11-13
3. Methodology
   1. Procedure to solve the given problem 14
   2. Model architecture 15
   3. Software description 16
4. Results and discussion 17-18
5. Conclusion and future scope 19
6. References 20

**1.INTRODUCTION**

## 1.1 OVERVIEW

A coupon prediction project involves applying data analytics and machine learning techniques to develop a predictive model for coupon prediction to the customers. The project typically involves collecting and analyzing customer data, such as past purchase, previous transcations and other relevant factors that may influence coupon usage. The data is then used to train a machine learning model that can predict which customers are most likely given a coupon and which coupons are likely to be redeemed. The benefits of a coupon prediction project include increased targeting, higher redemption rates, and improved return on investment for marketing efforts. A successful coupon prediction project can help businesses to optimize their marketing strategies and improve customer engagement, leading to increased sales and customer loyalty.

**1.2 PROBLEM STATEMENT**

## Many businesses in the retail and e-commerce industries struggle to optimize their coupon and offer distribution strategies due to a lack of data-driven insights. They often rely on a trial-and-error approach to distribute coupons and offers, which can result in low redemption rates and wasted marketing resources. To address this challenge, we aim to develop a predictive model that can identify which customers are most likely to respond to coupon promotions and which coupons are likely to be redeemed. By using customer data and machine learning algorithms to predict coupon usage, we aim to improve the targeting and effectiveness of our coupon distribution strategy and increase customer engagement and sales.

## 1.3 EXISTING SYSTEM

Customer segmentation is one of the existing model it is the process of dividing customers into groups based on common characteristics such as demographics, behaviour, preferences, and needs. A customer segmentation model is a statistical tool used to identify meaningful segments of customers that have similar characteristics, behaviours, or needs. It uses statistical methods to cluster customers into distinct groups based on the segmentation criteria.

**1.4 PROPOSED SYSTEM**

Here we want to predict the coupon to the customer based his/her transaction history and depending upon their item purchases the customer gets the coupon discount . This is a machine learning algorithm to predict which coupons will be most effective for different types of customers based on their purchase history and other factors. This involves looking at past coupon usage data to identify patterns and trends that can be used to make predictions about future purchases.

### 1.5 DEFINE OBJECTIVES

The main objectives of coupon prediction are to:

* Increase coupon redemption rates: By predicting which customers are most likely to get a coupon, we can target their promotions more effectively and increase the redemption rates of their coupons.
* Improve customer satisfaction: By offering customers coupons that are more relevant to their needs and preferences, we can improve customer satisfaction and increase customer loyalty.
* Increase sales: Coupon prediction can help us to optimize promotions to maximize sales and revenue.
* Reduce marketing costs: By targeting coupons more effectively, retailers can reduce marketing costs by avoiding the distribution of coupons to customers who are unlikely to use them.
* Optimize inventory management: By predicting which products are most likely to be purchased in order to predict coupon, retailers can optimize their inventory management and ensure that they have enough stock on hand to meet demand.

Overall, coupon prediction can help retailers improve the effectiveness of their coupon campaigns, reduce costs, and gain a better understanding of their customers' behaviour and preferences.

## 1.6 OVERALL ARCHITECTURE

DATA COLLECTION

DATA

CLEANING

DATA

VISUALISATION

EXPLORATORY

DATA ANALYSIS

TRAINING

TESTING

RESULTS

# 2.0 LITERATURE SURVEY

# Coupon prediction refers to the process of using data analytics and machine learning algorithms to forecast which customers are likely to use coupons and when. This can help businesses optimize their couponing strategies, target the right customers with the right coupons, and increase their ROI. Here is a literature survey on coupon prediction.

Coupon prediction is a popular research topic in the field of marketing and e-commerce. The following literature survey covers some of the recent studies that have focused on coupon prediction:

"Predicting Coupon Redemption Rates: A Deep Learning Approach" by Zeng et al. (2021) - This paper proposes a deep learning model to predict coupon redemption rates based on the customer's historical purchase data and demographic information. The model achieved higher accuracy compared to traditional machine learning methods.

"Coupon Recommendation Using Deep Learning Based Collaborative Filtering" by Chen et al. (2020) - This study proposes a collaborative filtering-based approach to predict which coupons customers are likely to redeem. The proposed model outperformed traditional collaborative filtering methods.

"Customer Coupon Usage Prediction Based on Time Series Analysis and Random Forest Algorithm" by Liu et al. (2020) - This paper presents a time-series-based prediction model to forecast customer coupon usage. The model integrates random forest algorithm with time series analysis to improve the accuracy of coupon usage prediction.

"Predicting Coupon Redemption in E-commerce: A Multi-task Learning Approach" by Chen et al. (2019) - This study presents a multi-task learning approach to predict coupon redemption in e-commerce. The proposed model leverages customer purchase history, product features, and customer demographic information to predict coupon redemption rates.

"A Coupon Recommendation System for E-commerce" by Yang et al. (2018) - This paper proposes a coupon recommendation system for e-commerce based on customer purchase history and behavior data. The proposed system uses collaborative filtering and association rule mining techniques to recommend coupons to customers.

Overall, these studies demonstrate the effectiveness of using machine learning and deep learning models to predict coupon redemption rates and recommend coupons to customers.

# Overall, these above information demonstrate the potential of machine learning and data analytics for coupon prediction, and provide insights into the best practices and techniques for building accurate and effective couponing strategies.

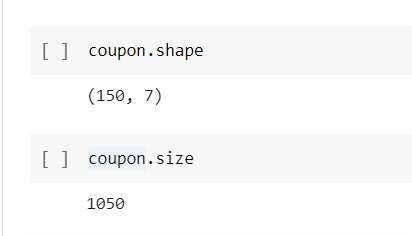
Li et al. (2018) proposed a data mining approach to predict coupon redemption in e-commerce. The study used a dataset containing coupon usage and user behavior data from a Chinese e-commerce platform. The dataset was preprocessed, and feature selection was performed to select the most relevant features for the prediction task.

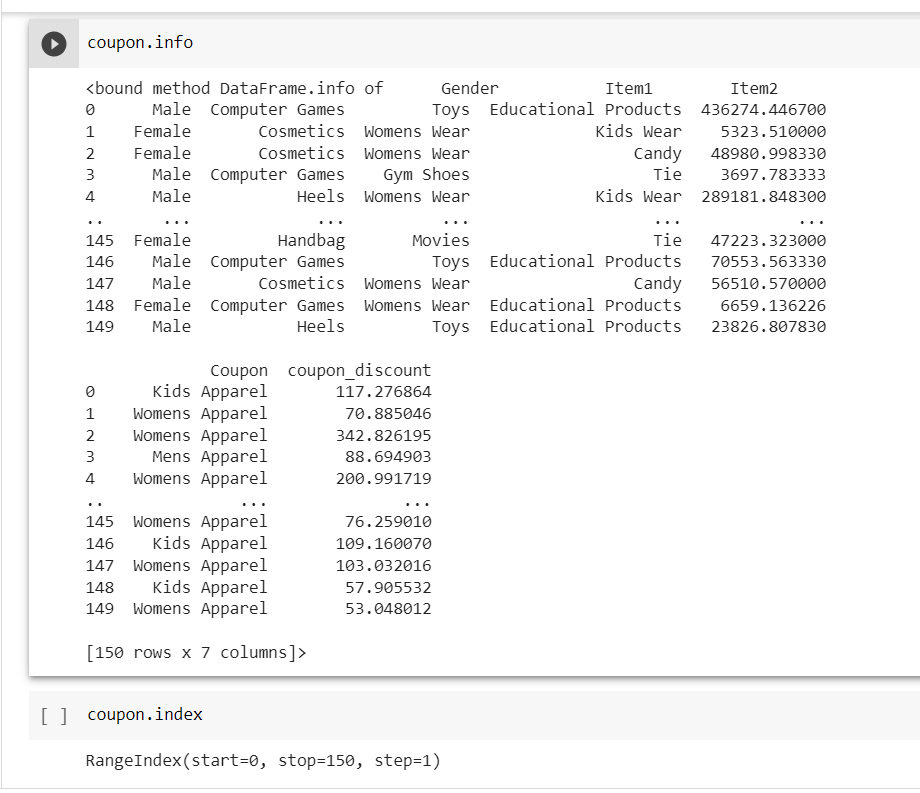
The authors used four classification algorithms, namely logistic regression, decision tree, random forest, and support vector machine (SVM), to predict coupon redemption. The performance of each algorithm was evaluated using metrics such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC).

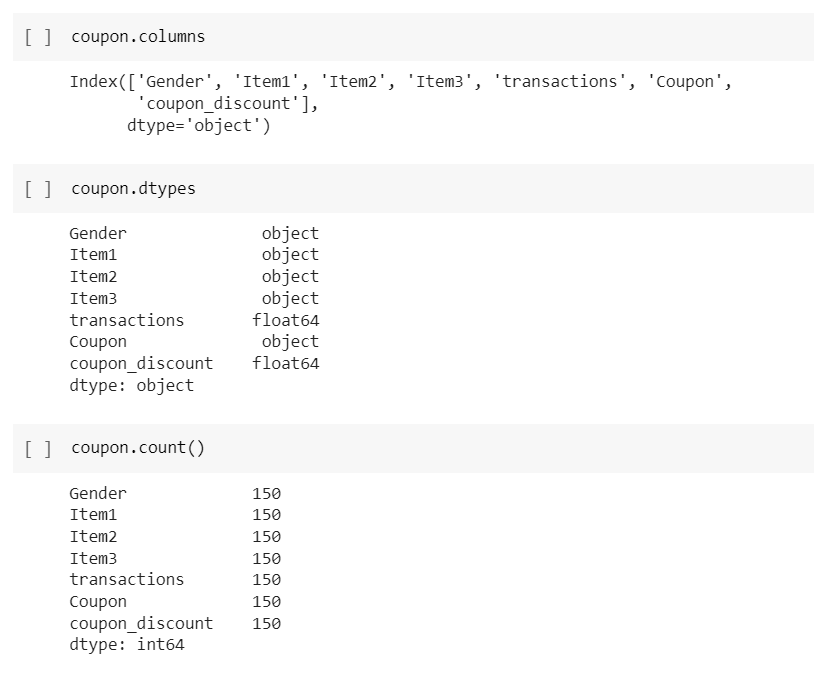
The results showed that SVM outperformed the other three algorithms, achieving an accuracy of 76.18% and an AUC-ROC of 0.792. The authors also performed a sensitivity analysis to evaluate the impact of different feature subsets on the prediction performance. The results showed that the inclusion of user behavior data significantly improved the prediction performance.

The study contributes to the field of coupon prediction by proposing a data mining approach that can effectively predict coupon redemption in e-commerce. The approach is based on a combination of feature selection and classification algorithms, and the results show that it can achieve high prediction accuracy. The study also highlights the importance of including user behavior data in the prediction task.

# 3.0 DATA PREPROCESIING







## 3.1 DATASET DESCRIPTION

## This dataset contains information related to Retail/E-commerce sales and the type of rewards selected by the customers. The dataset consist of only 150 records.

## The 7 main attributes of our dataset are GENDER , ITEM1, ITEM2, ITEM3, TRANSACTIONS, COUPON, COUPON DISCOUNT.

## .GENDER: The Gender of the customer to predict the coupon regarding their uses with help of Gender.

## Categorical variable. It denotes the customer's gender.

## 2.ITEMS: It is the group of the items.

## Computer Games

## Toys

## Educational Products

## Cosmetics

## Womens Wear

## Kids Wear

## Handbag

## Movies

## Board Games

## Mens Wear

## Candy

## Gym Shoes

## Tie

## Heels

## 3.COUPONS: These are the coupons assigned or predicted to the customers and the following are they

## Kids Apparel

## Womens Apparel

## Mens Apparel

## 4.PREVIOUS TRANSACTIONS:

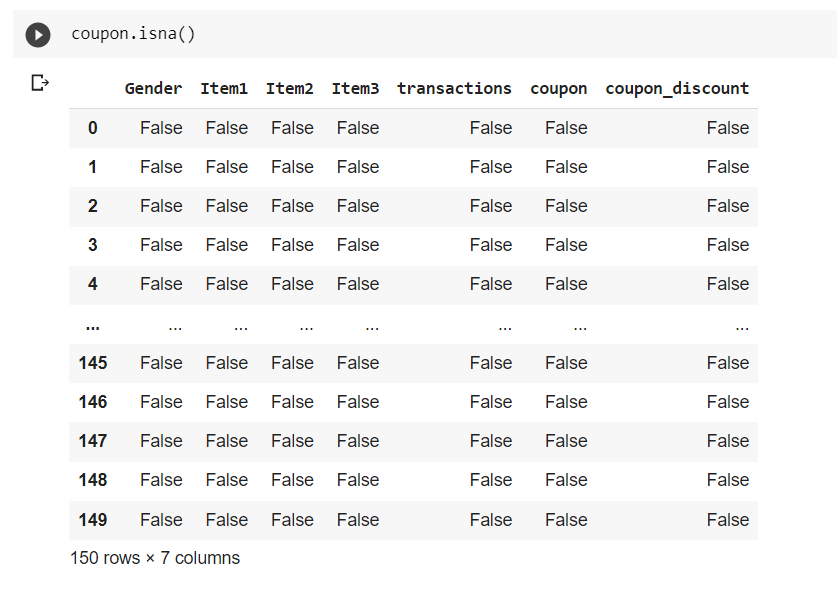
## 5.COUPON DISCOUNT 8

**3.2 DATA CLEANING**

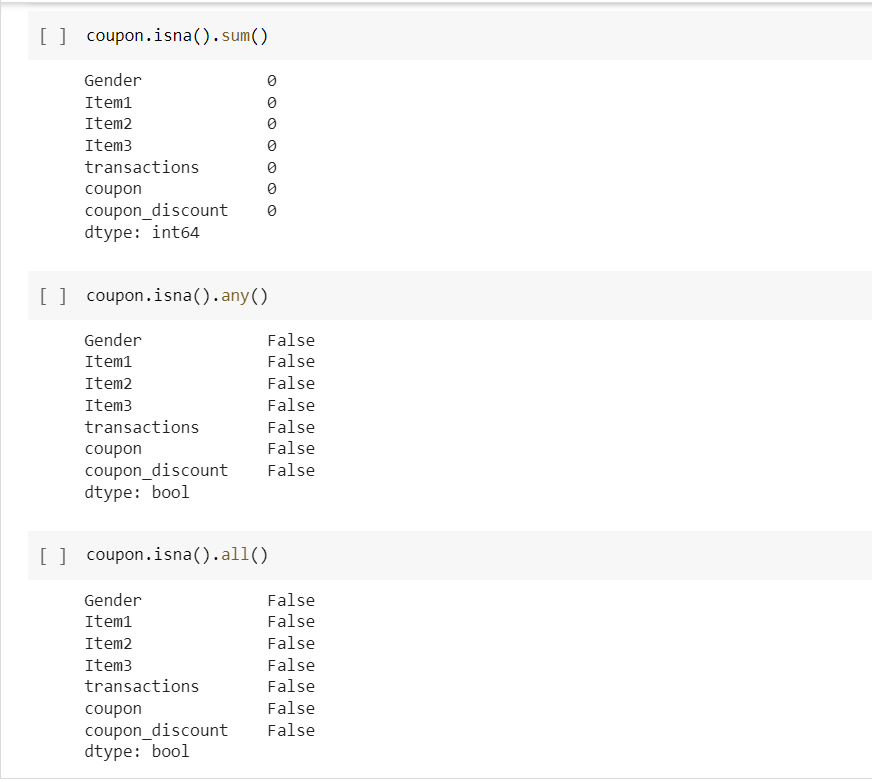
Data Cleaning is a process of removing null values and inaccurate values.

In this particular dataset we have checked for null and irrevalant values by importing the dataset. If the null values are found it can be removed by isnull().

This can be implemented as follows:



9



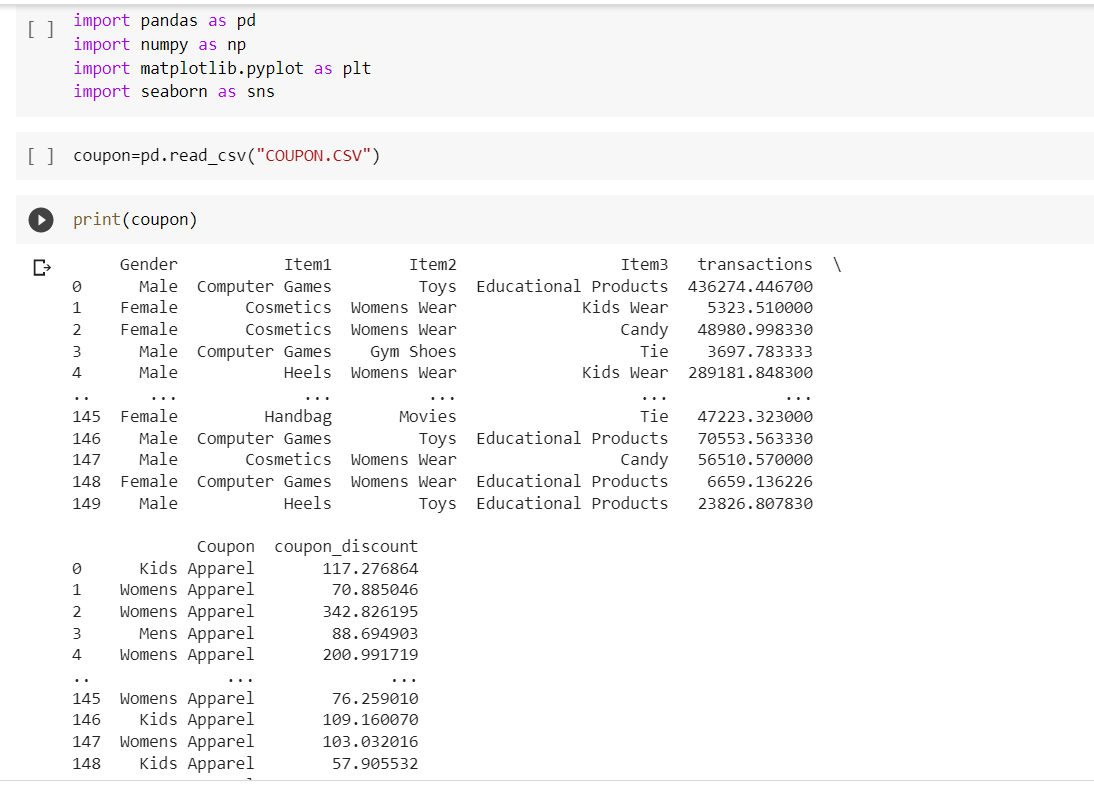
10

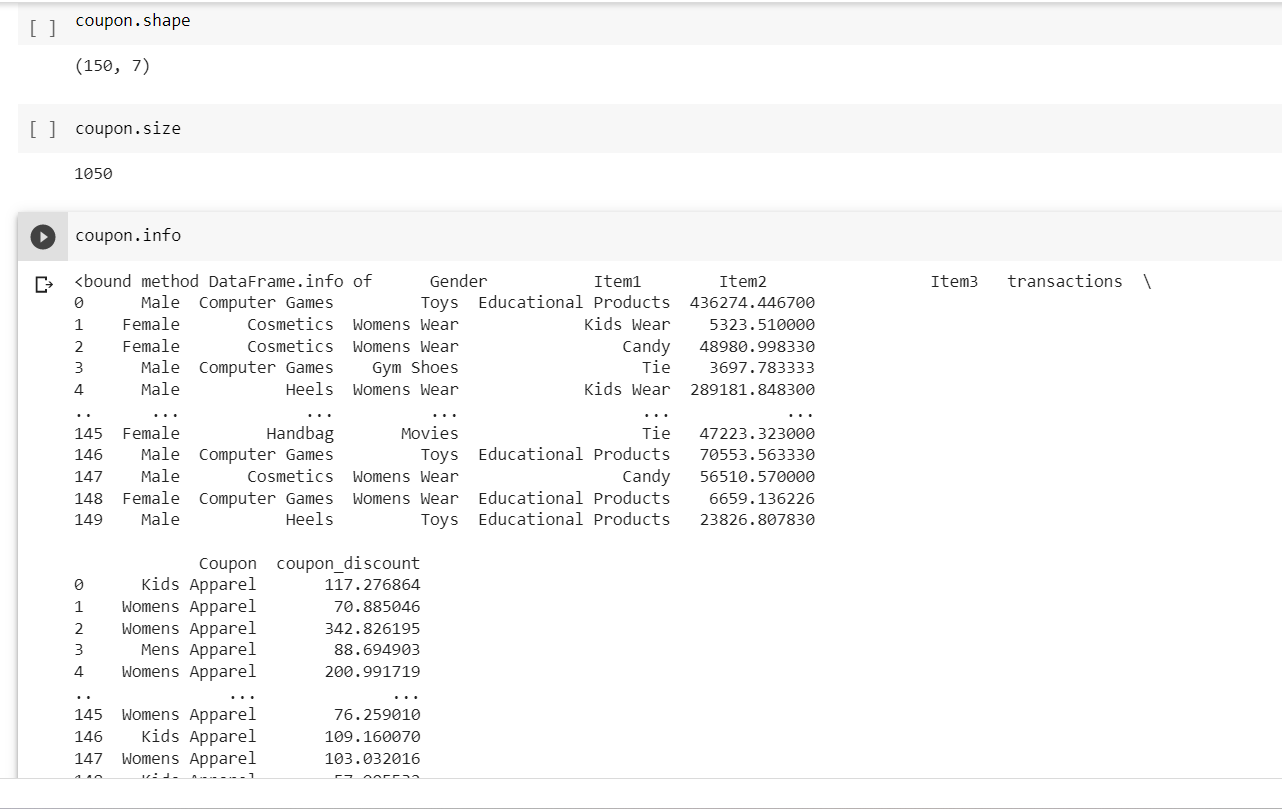
## 3.3 DATA AUGMENTATION

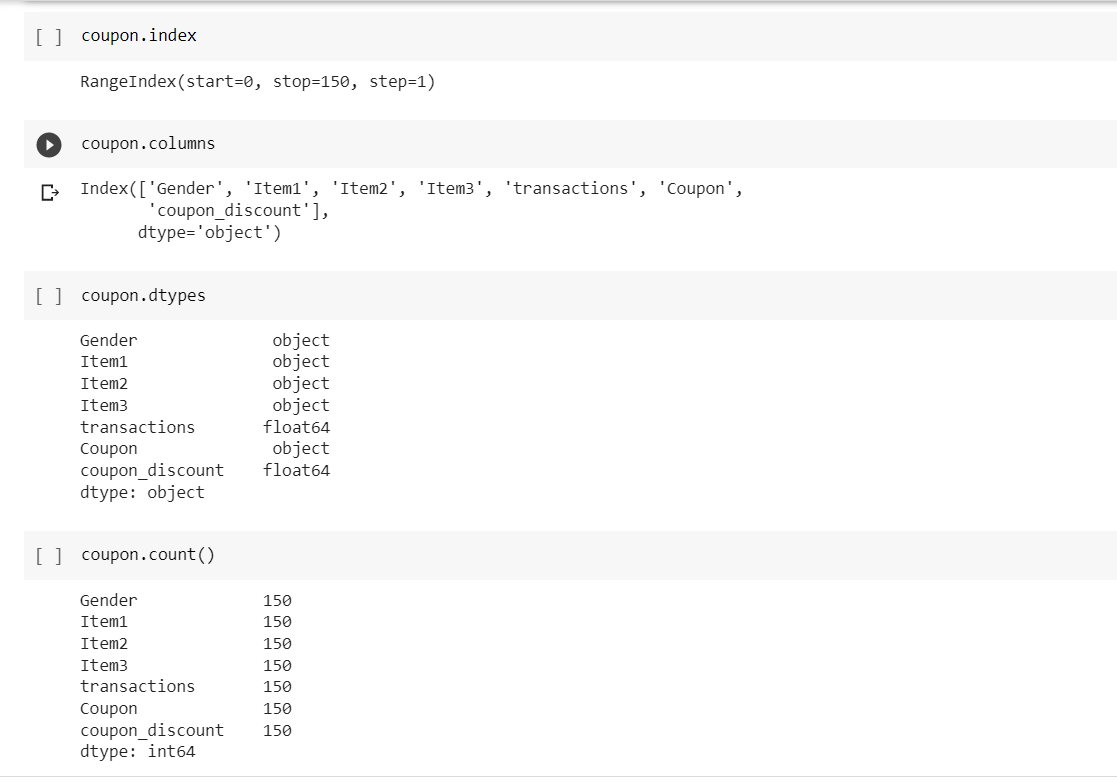
Data augmentation is a technique used in machine learning to increase the size and diversity of a training dataset by creating new training examples from existing ones. This can be particularly useful in coupon prediction tasks, where having a larger and more diverse training set can lead to better performance of the predictive model. Overall, data augmentation can be a powerful tool in coupon prediction tasks, allowing the model to learn from a larger and more diverse set of training examples and improving its ability to generalize to new and unseen data.

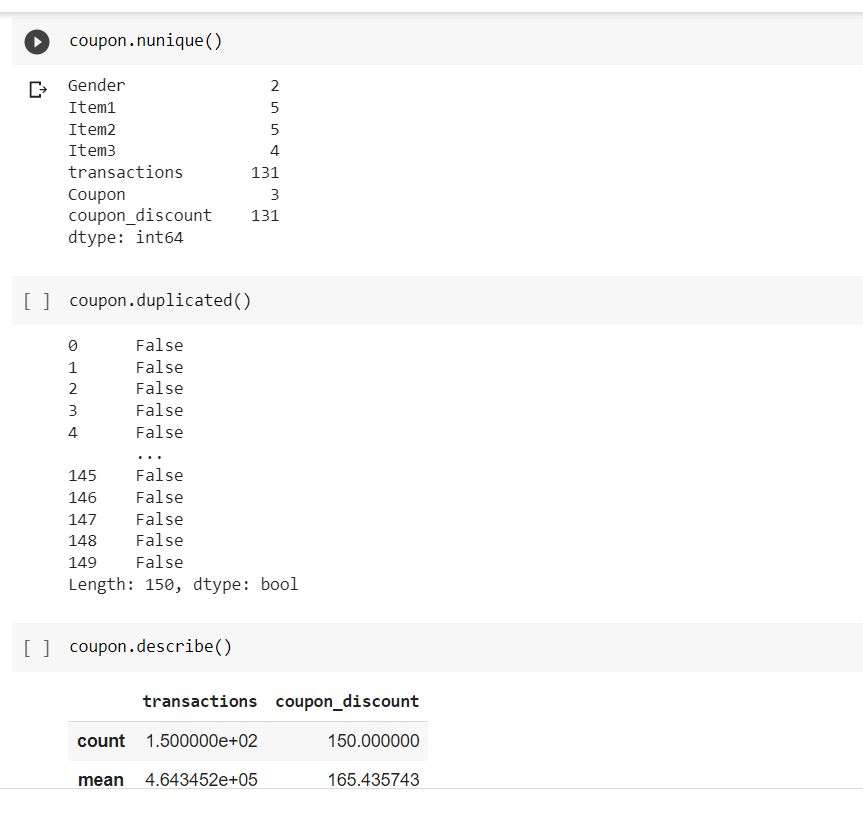
**3.4 DATA VISUALISATION**

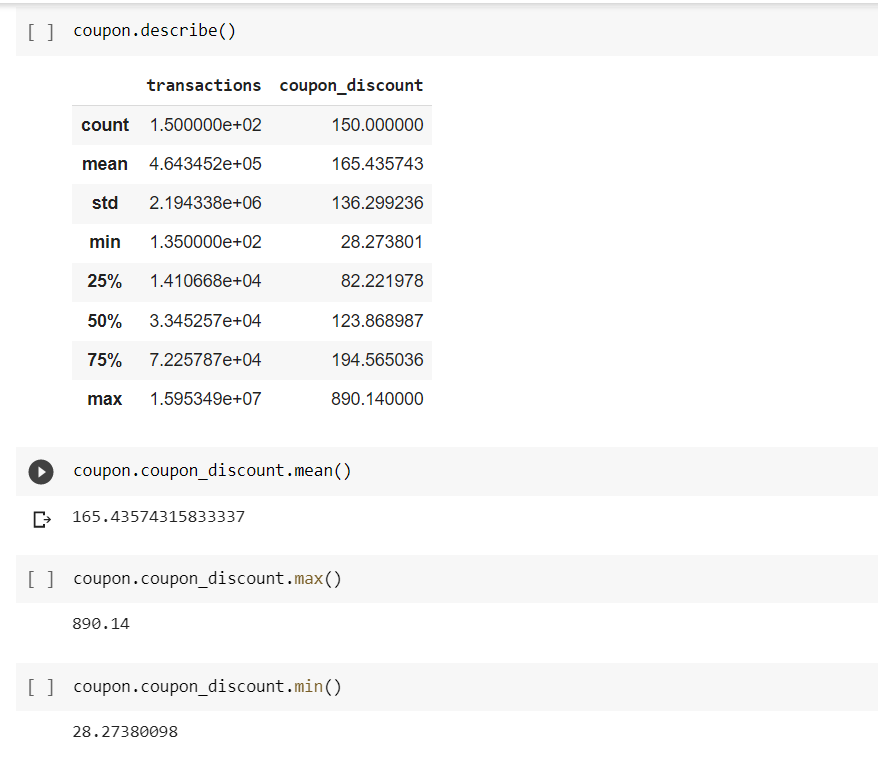
Data visualization is the portrayal of data in charts, graphs, maps and other visual formats to help you identify trends and relationship in your data.

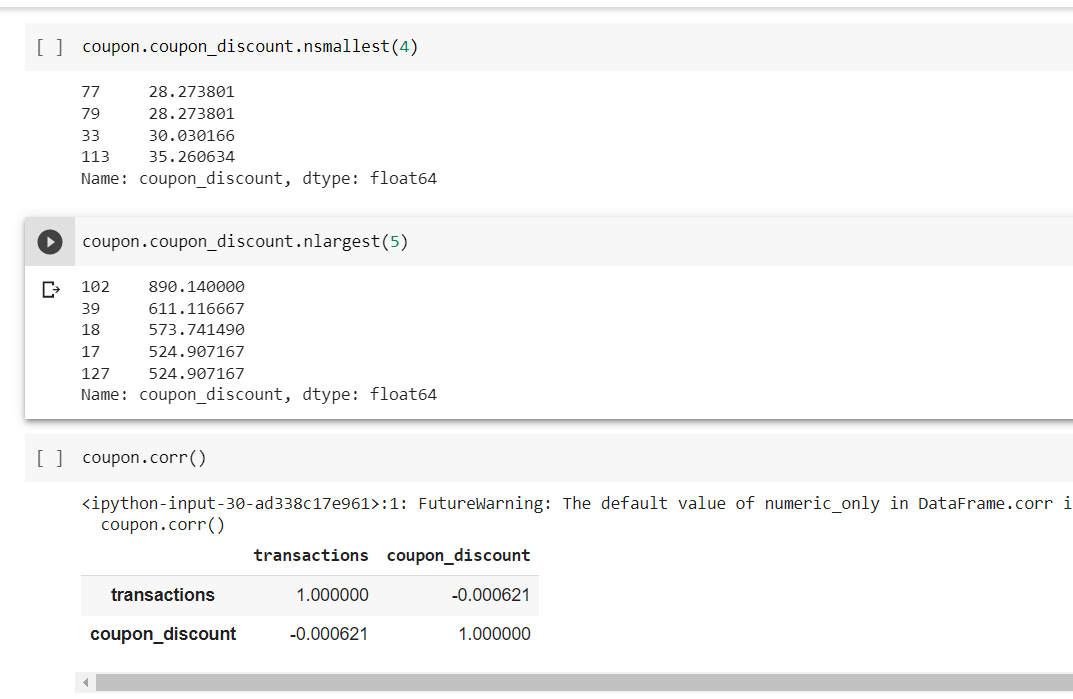


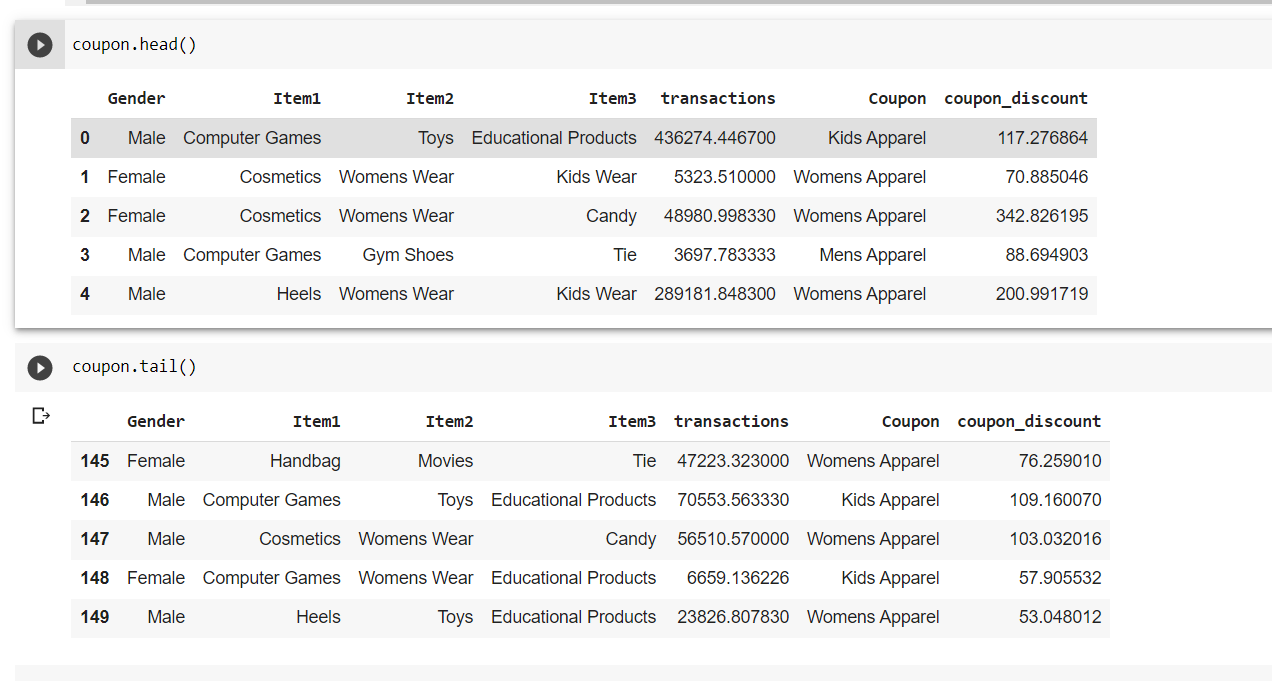


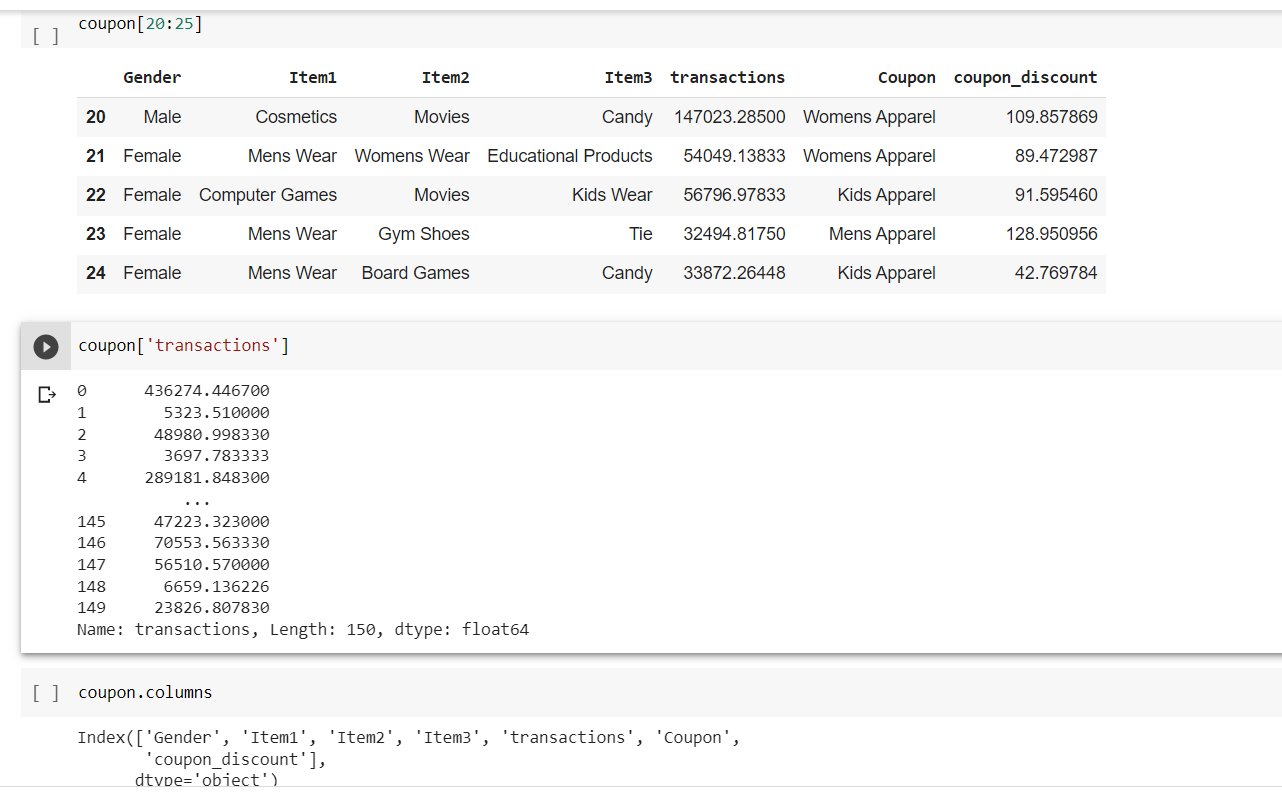


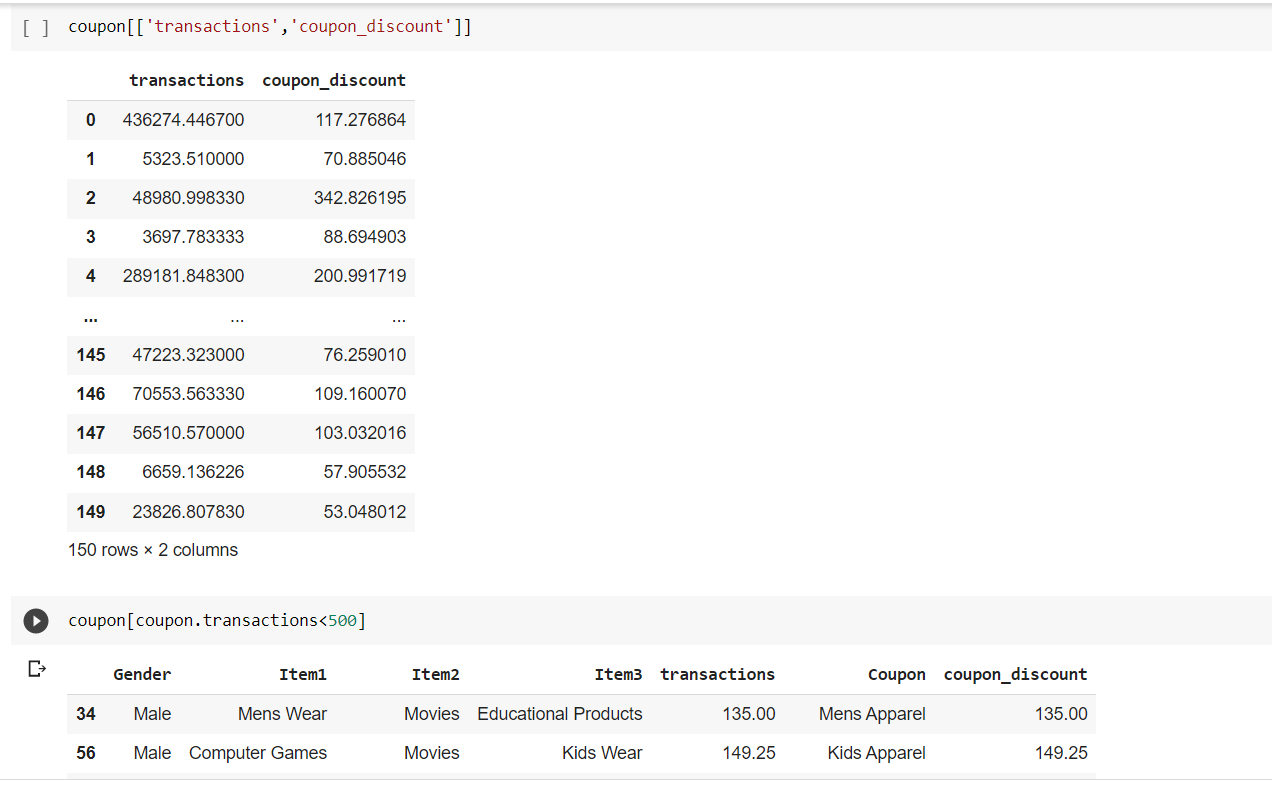
****

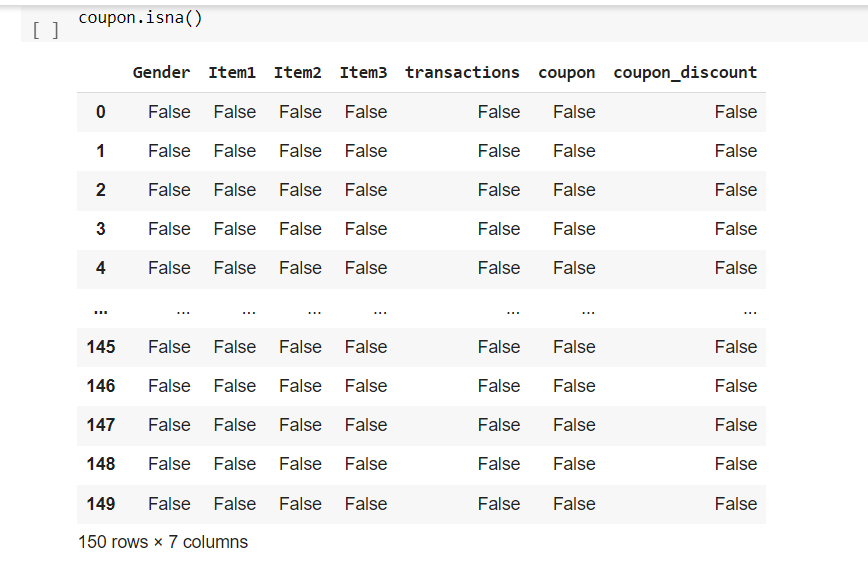
****

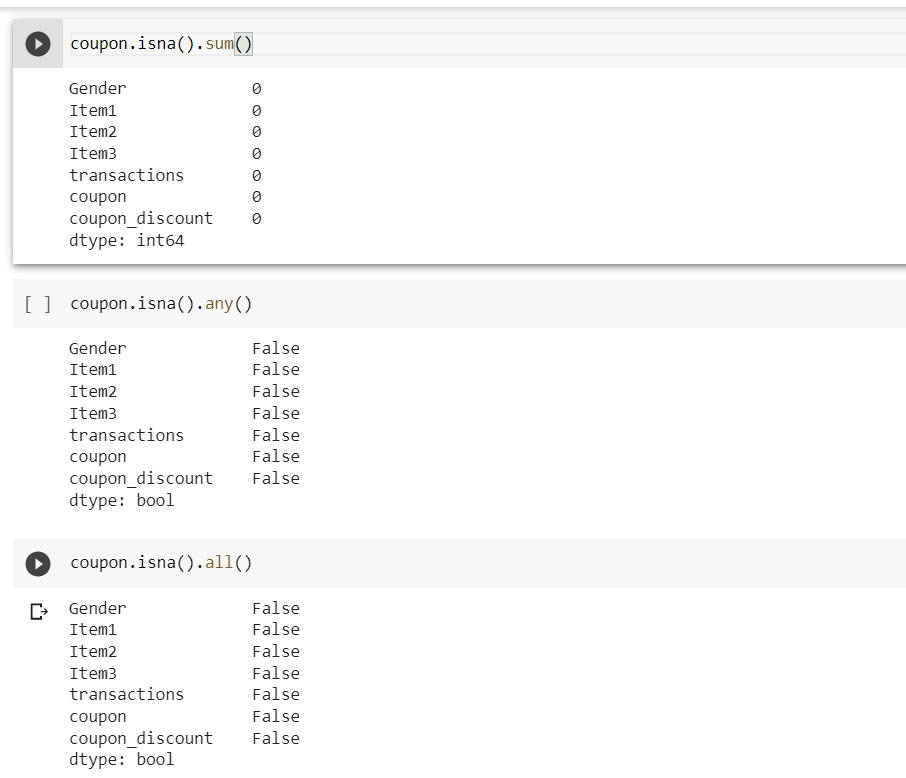
****

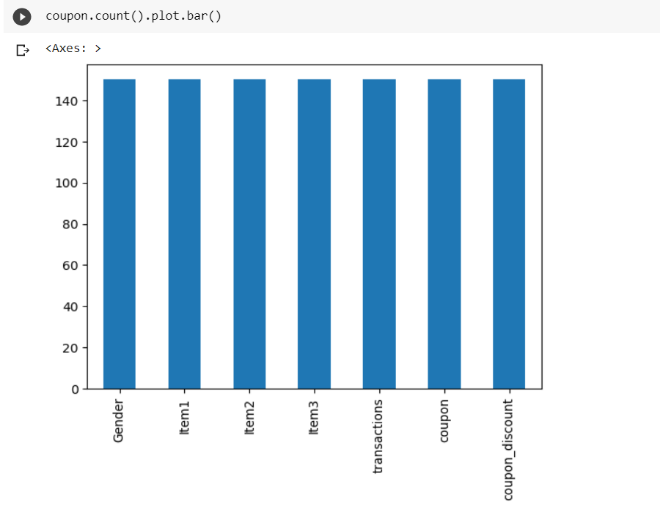
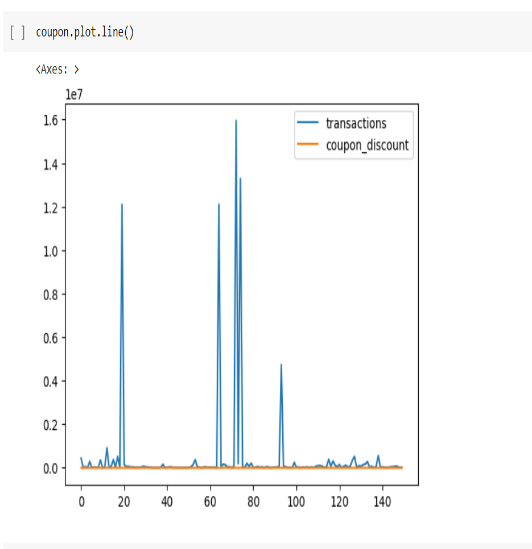
****

****

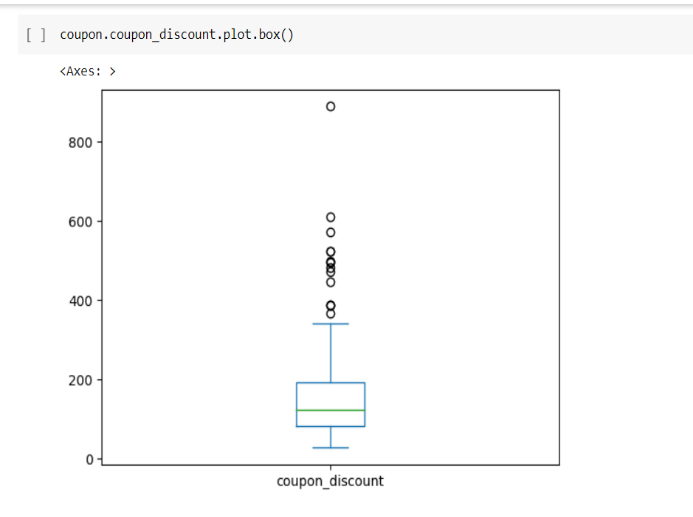
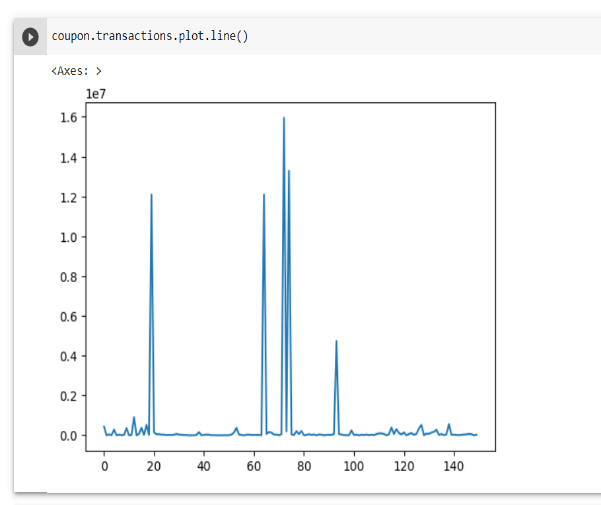
****

****

****

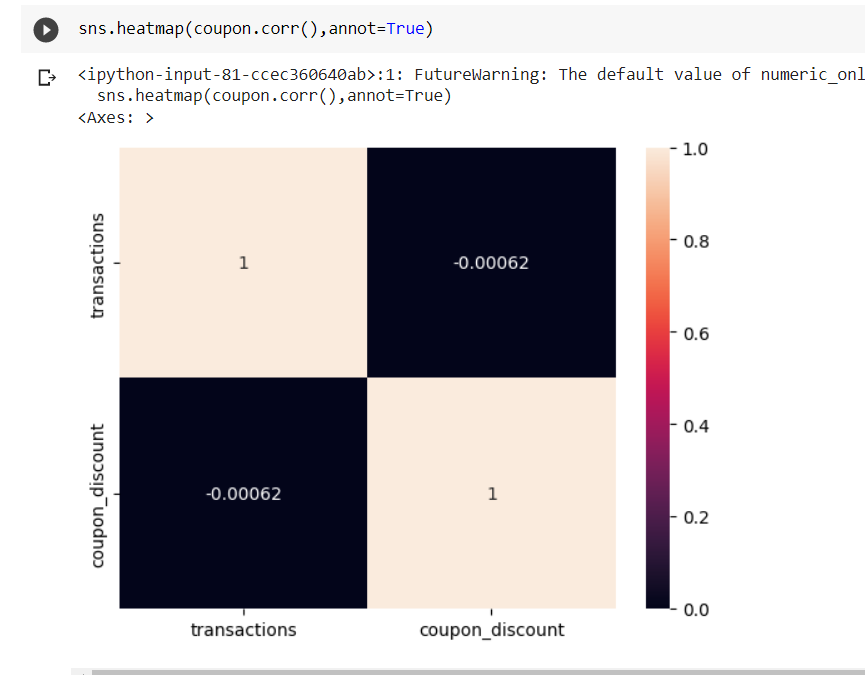
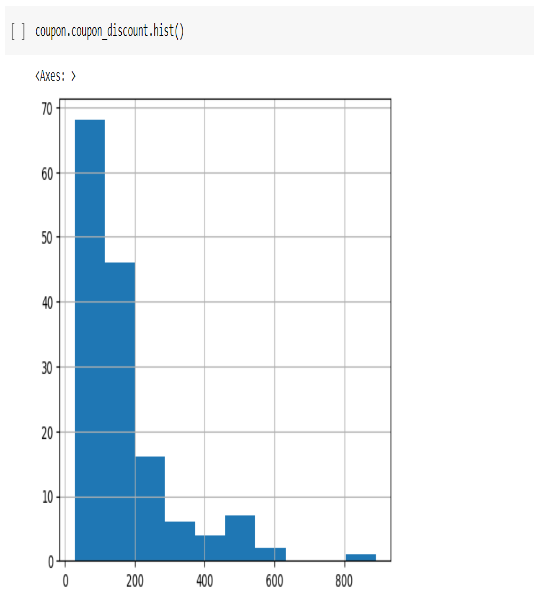
** ** Visual representation of data in form of barplot plot line it is a tool used to visually represent the se

events in a story.

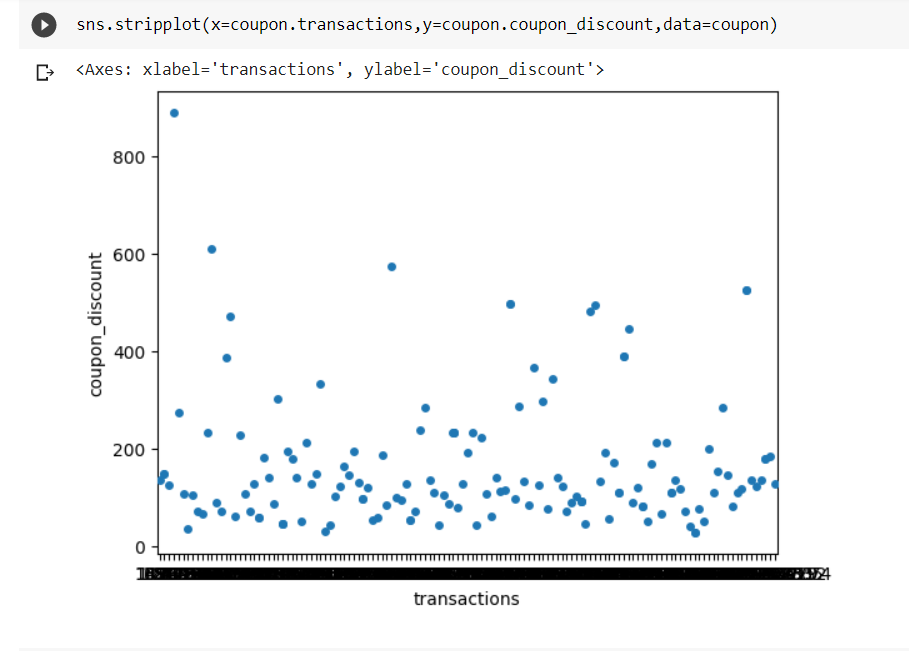
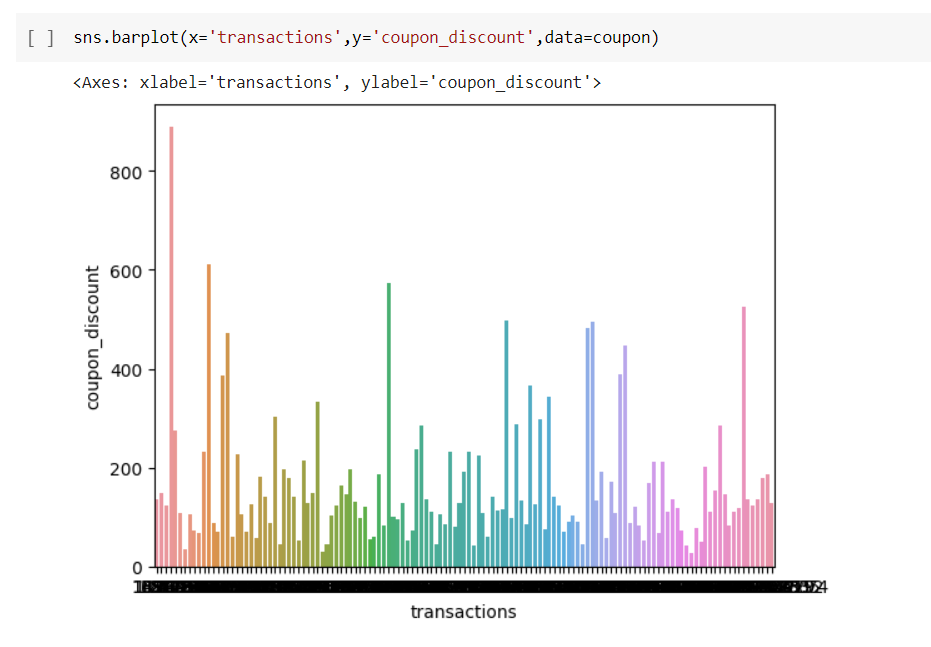
**** ****

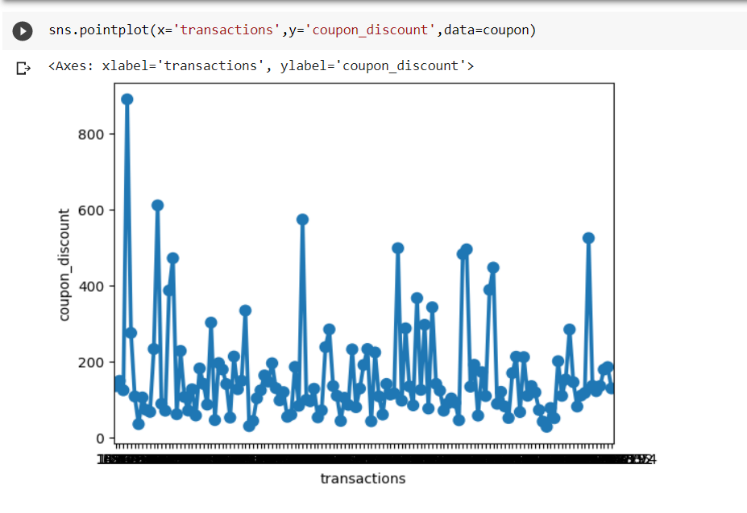
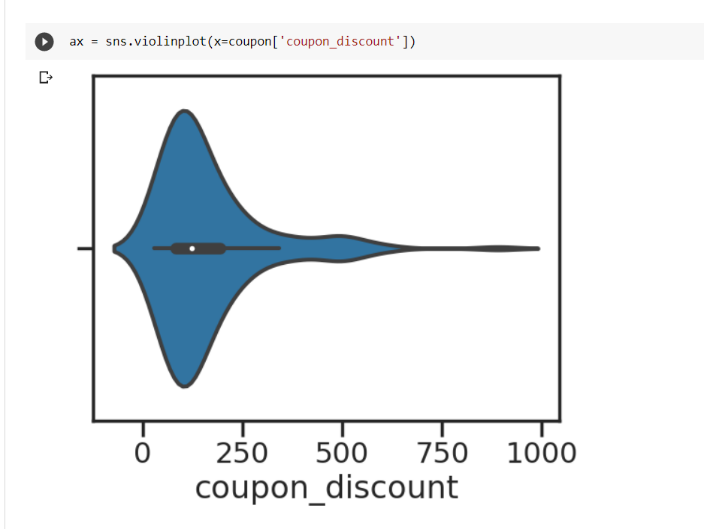
It is used to show the distribution of the data

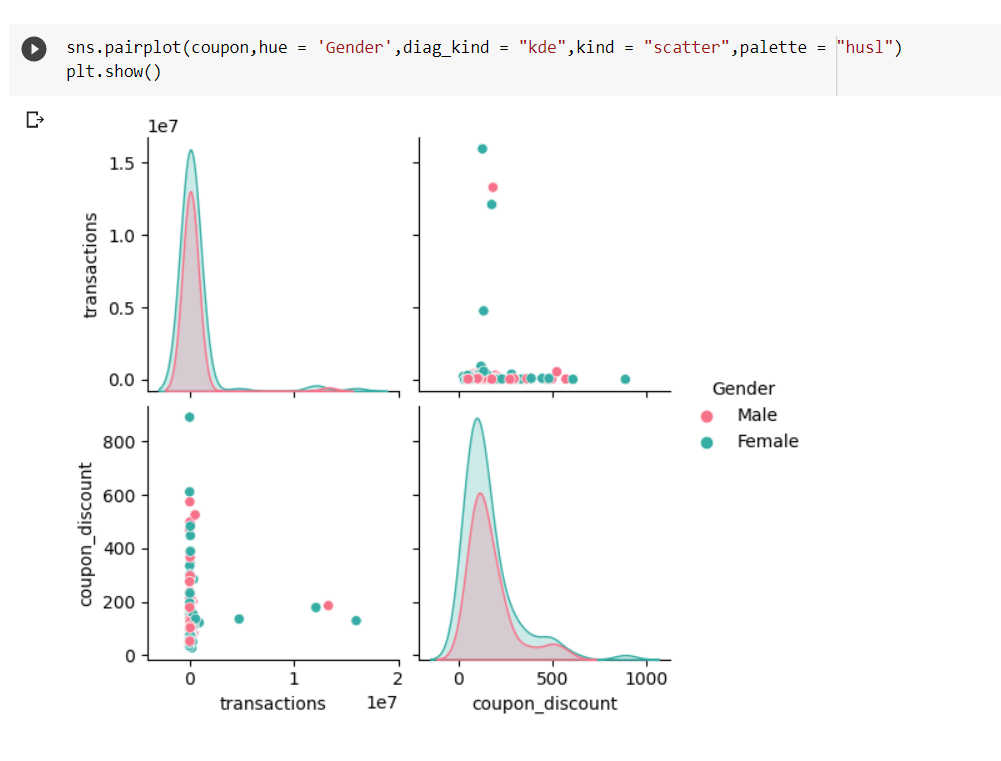
and to identify outliers

****

It is a way to show the frequency of different values or ranges of values in a dataset

****

****

****

A pair plot is a type of data visualization that displays pairwise relationships between multiple variables in a dataset. It is commonly used in exploratory data analysis to identify correlations or patterns in the data

# 4.0 METHODOLOGY

## 4.1 PROCEDURE

To predict the coupon for a particular customer , we have collected the data about the transactions ,purchases and items that undertaken by the customers .We have considered the independent variables as x(Transactions) and dependent variables as y(coupon discount).From this data we want to predict the coupons in future for a customer . We have used linear regression algorithm in order to solve this problem .It is defined as the relationship between one independent variable and one independent variable.

y = ax + b

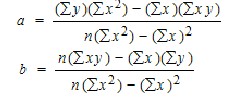
where y = Targe

x= features

a = slope

b = y-intercept constant

The slope and intercept can be calculated by:



15

### 4.2 Model architecture

Input

(

Gender

,

Item1,item2,item3,

Coupons,

Coupon discounts)

Regressor

Train

e

d

Mo

del

Unknown

In

put

Predicted

output

Here the input include Transactions and which are considered as independent variables ‘x’. From this we will predict the coupon discount which is considered as a dependent variable ‘y’. Based on the transactions occurrence in the past , we will predict what coupon will occur in future , if it occurs then we will predict the discounts.

16

### 4.3 SOFTWARE DESCRIPTION

**CONDA COLAB**: Conda is a package manager that allows users to easily manage and install software packages and dependencies. Colab , on the other hand, is a free online platform that allows users to write and execute Python code in a Jupyter notebook-like environment. Conda-Colab is a combination of these two tools that allows users to run Conda commands in a Colab notebook. Conda-Colab enables users to install and manage software packages in a virtual environment within Colab, which helps to ensure compatibility and avoids conflicts with other installed packages. It also provides a seamless integration with Colab's GPU and TPU capabilities, making it easy to set up and run machine learning experiments.Overall, Conda-Colab is a useful tool for data scientists and machine learning practitioners who want to quickly set up and manage their environment within Colab.

**Libraries used in project are as follows:**

NUMPY- Numpy is a very popular python library for large multi-dimensional array and matrix processing, with the help of a large collection of high-level mathematical functions. It is very useful for fundamental scientific computations in Machine Learning.

PANDAS-Pandas is a Software Libarary in Computer Programming and it is written for the Python Programming Language its work to do data analysis and manipulation.

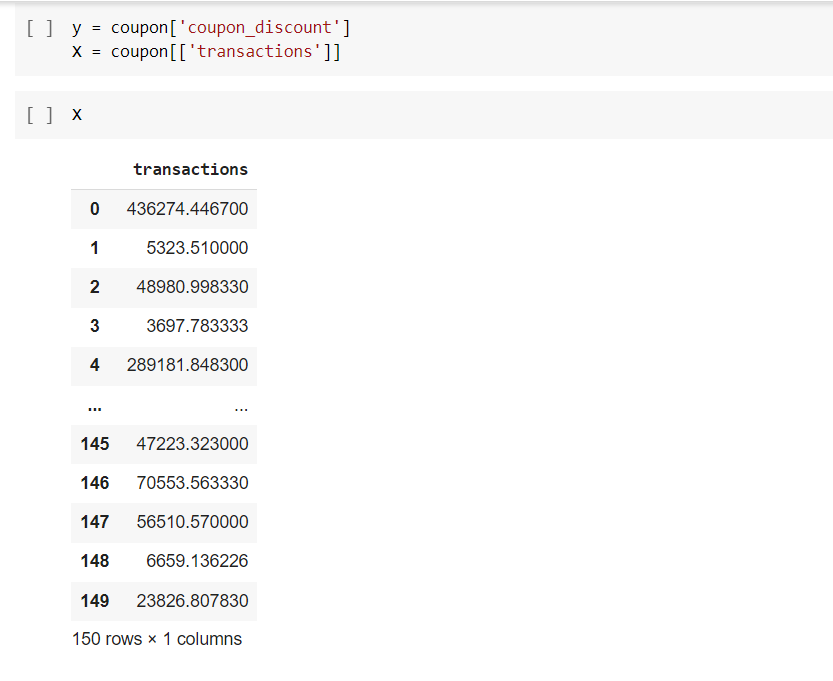
SEABORN- Seaborn is a Python data visualization library based on the Matplotlib library. It provides a high-level interface for drawing attractive and informative statistical graphs. MATPLOTLIB- Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open source alternative to MATLAB. Developers can also use matplotlib's APIs to embed plots in GUI applications.

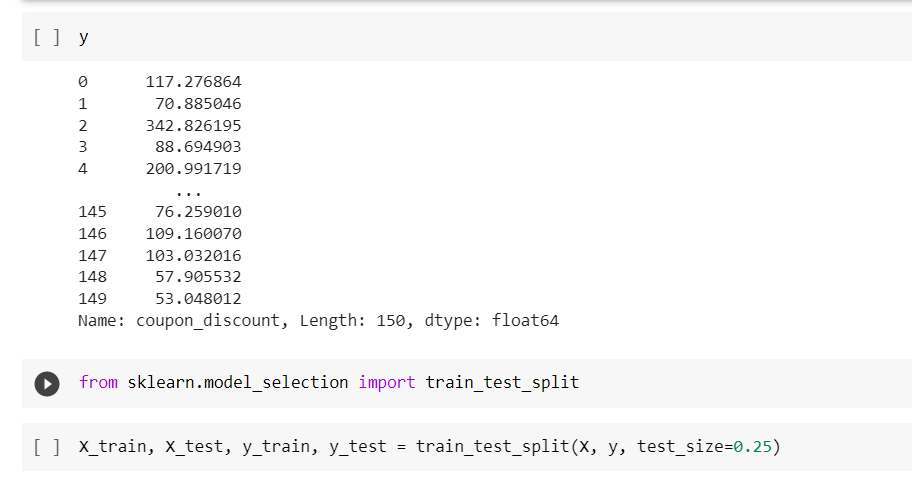
SKLEARN- Simple and efficient tools for predictive data analysis . Accessible to everybody, and reusable in various context.

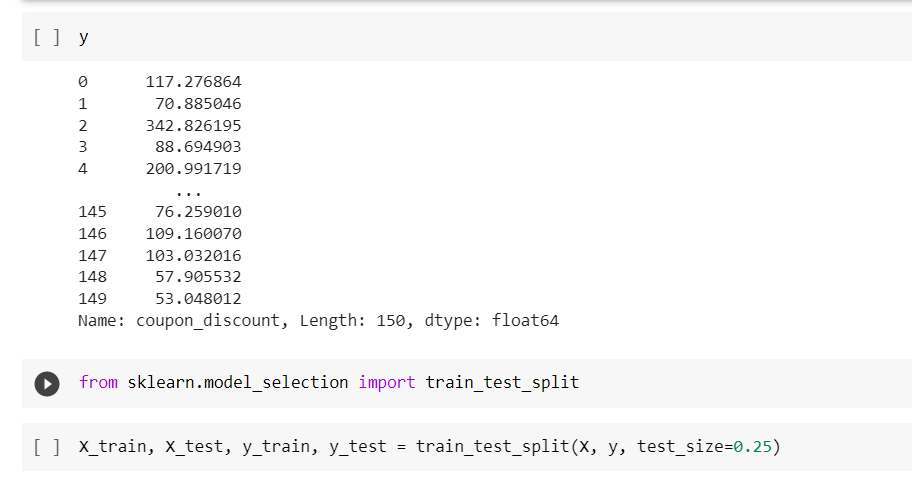
METRICS-Metrics are used to monitor and measure the performance of a model (during training and testing), and don't need to be differentiable.

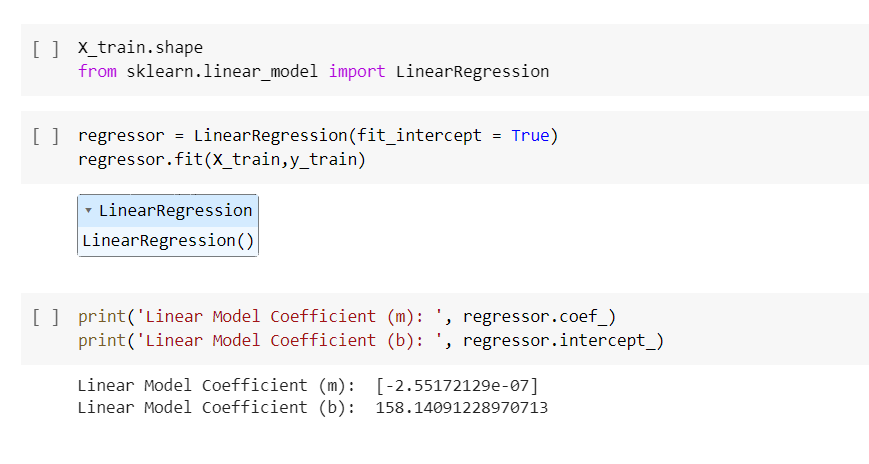
17

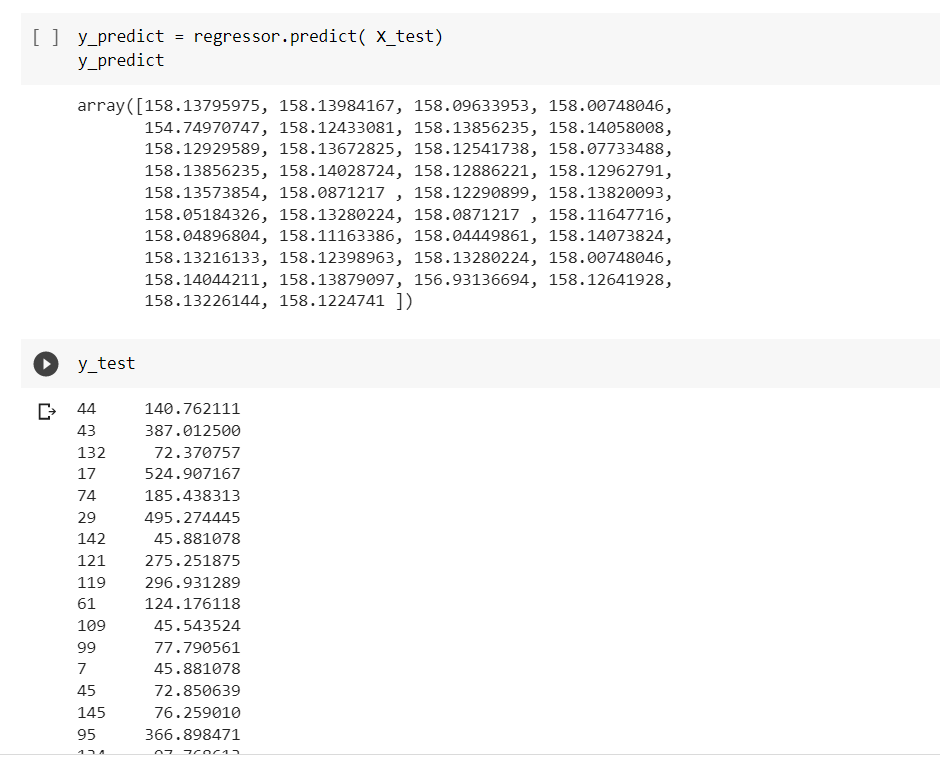
## 5. RESULTS AND DISCUSSION

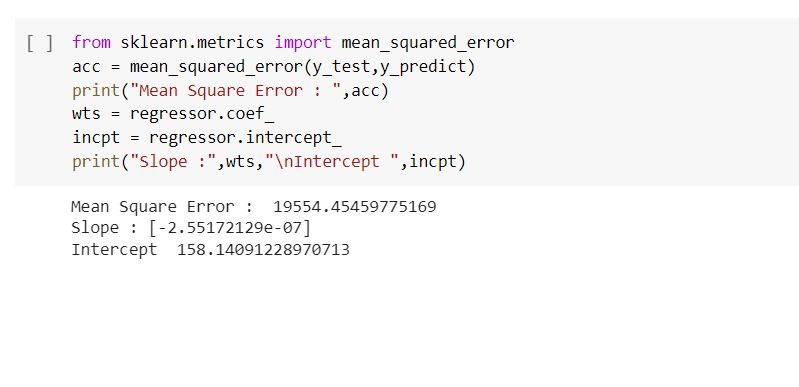












### 6.CONCLUSION AND FUTURE SCOPE

**Conclusion:** In conclusion, coupon prediction is an important problem for businesses that offer coupons as a marketing tool. The goal of coupon prediction is to forecast the probability of coupon based on various predictors .

Linear regression is a commonly used method for coupon prediction, as it can model the linear relationship between coupon and the predictors. However, other regression models such as logistic regression, decision trees, or random forests may also be used depending on the nature of the problem and the available data.

Overall, accurate coupon prediction can help businesses optimize their coupon campaigns, improve customer engagement, and increase sales. By leveraging the power of data analytics and machine learning, businesses can gain valuable insights into customer.

**Future Scope:** The future scope for coupon prediction is promising, as advancements in data analytics and machine learning are enabling businesses to gain deeper insights into customer behaviour and preferences.

Personalization,Real-time prediction,Integration with loyalty programs,Multi-channel prediction,Integration with supply chain are some potential areas of development for coupon prediction

### 7.REFERENCES

<https://www.kaggle.com/>

<https://www.kaggle.com/datasets/vysakhvms/coupons>

<https://colab.research.google.com/drive/1vxw_MLqJBvBqqO7KQtMJrxM0kxaTN5bl#scrollTo=vado7J-blSQh>

<https://colab.research.google.com/drive/1WY1JhR-VE_UldJ6Jc7SW73Aa57wxkDrS#scrollTo=NwQrm8_IVDhg>

https://colab.research.google.com/drive/1WY1JhR-VE\_UldJ6Jc7SW73Aa57wxkDrS#scrollTo=NwQrm8\_IVDhg