Final Project Report

on

Supermarket Sales Analysis and Forecasting

MCA

Computer Applications

by

Sparsh Bhardwaj

(22001602062)

Under supervision of

Dr. Manjeet Singh

Department of Computer Applications

JC BOSE UNIVERSITY OF SCIENCE & TECHNOLOGY
YMCA FARIDABAD-121006
November 2023

DECLARATION

We hereby declare that the project work entitled **"Supermarket Sales Analysis and Forecasting"** submitted to **JCBUST, YMCA, Faridabad (Haryana)**, is a record of an original work done by us under the guidance of **Dr. Manjeet Singh**, Professor in Computer Applications, J.C. Bose University of Science and Technology (YMCA), Faridabad. This project is submitted in partial fulfilment of the requirements for the award of the degree of MCA.

Sparsh Bhardwaj
(Name of student)
Date: November 2023

CERTIFICATE

This is to certify that **Sparsh Bhardwaj** of J.C. BOSE UNIVERSITY OF SCIENCE AND TECHNOLOGY (JCBUST), YMCA has successfully completed the project work titled **Supermarket Sales Analysis and Forecasting** in partial fulfilment of the requirement for the completion of the UG course.

This project report is the record of authentic work carried out by them during the period from **OCT 2023** to **DEC 2023**. They had worked under my guidance.

Signature:

Mentor Name: **Dr. Manjeet Singh**

ACKNOWLEDGEMENT

This project would not have taken shape, without the guidance provided by **Dr. Manjeet Singh**, our mentor who helped in the modules of our project and resolved all the technical as well as other problems related to the project and, for always providing me with a helping hand whenever we faced any bottlenecks, in spite of being quite busy with her hectic schedules.

Table of Contents

INTRODUCTION

Project Statement:

The supermarket industry faces the challenge of efficiently managing sales data to optimize operations, inventory, and customer satisfaction. Inaccurate sales forecasts and inefficient analysis can lead to overstocking, stockouts, and revenue loss. The motivation behind this project is to address these challenges by employing advanced data analytics techniques for precise sales analysis and forecasting in supermarkets.

It will contain:

1. Graphs based on provided data.
2. Detailed analysis of sales data.
3. Insights of data.
4. Forecasting which will improve the decision-making process.

Problem Statement and Motivation

The primary challenge faced in supermarket sales analysis and forecasting lies in the accuracy of the collected data. Often, sales records might not be accurately inputted by the original personnel, introducing errors and inconsistencies. For instance, a cashier might mistakenly input the wrong product code or quantity sold, leading to inaccurate sales data. Additionally, manual entry of sales figures can be prone to human error, affecting the reliability of the data for forecasting purposes. Implementing a system that ensures accurate data collection without solely relying on manual input would be crucial to enhance the accuracy of sales analysis and forecasting.

Moreover, the time-consuming nature of processing sales data poses another hurdle. Analyzing a vast number of sales records manually can be highly inefficient. Python offers powerful tools and libraries for data processing and analysis, yet the challenge lies in optimizing these processes to handle large volumes of sales data efficiently. Automating data extraction, cleaning, and analysis using Python scripts and machine learning algorithms can significantly reduce the time required for sales analysis and forecasting, improving overall efficiency.

Furthermore, accessibility to sales information for relevant stakeholders, such as managers and decision-makers, is vital for informed decision-making. The lack of easy access to real-time sales data can hinder timely decision-making and forecasting accuracy. Enhancing the system to provide accessible and user-friendly dashboards or reports generated through Python-based tools can empower stakeholders with instant access to crucial sales insights, facilitating better decision-making and forecasting accuracy.

OBJECTIVE OF PROPOSED PROJECT

The objectives of the project are listed below:

a) The primary goal of the project is to enhance decision-making processes for supermarket management by providing accurate and insightful sales analysis and forecasting.

b) Improve the efficiency of sales data collection and analysis, aiming for a faster turnaround time in generating sales reports compared to the existing system.

c) Implement advanced forecasting algorithms to predict future sales trends, enabling the supermarket to optimize inventory levels, plan promotions effectively, and enhance overall supply chain management.

d) Ensure the accuracy and reliability of sales data by implementing robust data validation processes, reducing the chances of errors and inconsistencies in the analysis.

e) Incorporate machine learning and data mining techniques to identify patterns and correlations within sales data, providing valuable insights for strategic decision-making.

f) Identify patterns, seasonality, and recurring trends to understand the cyclical nature of sales.

TECHNOLOGIES AND ONLINE SERVICES USED

1. Tools Required/Platform Used

Tools Required to make this project:

a) Pentium-pro processor or later.
b) RAM 512MB or more.
c) Windows 7(64-bit) or above.
d) Python 3.0 or above.
e) Jupiter or Google Colab.
f) Strong Internet Connection

This project requires the knowledge of following:

a) Python
b) Numpy
c) Panda
d) Seaborn
e) Arima
f) Matplotlib

FILES INCLUDED

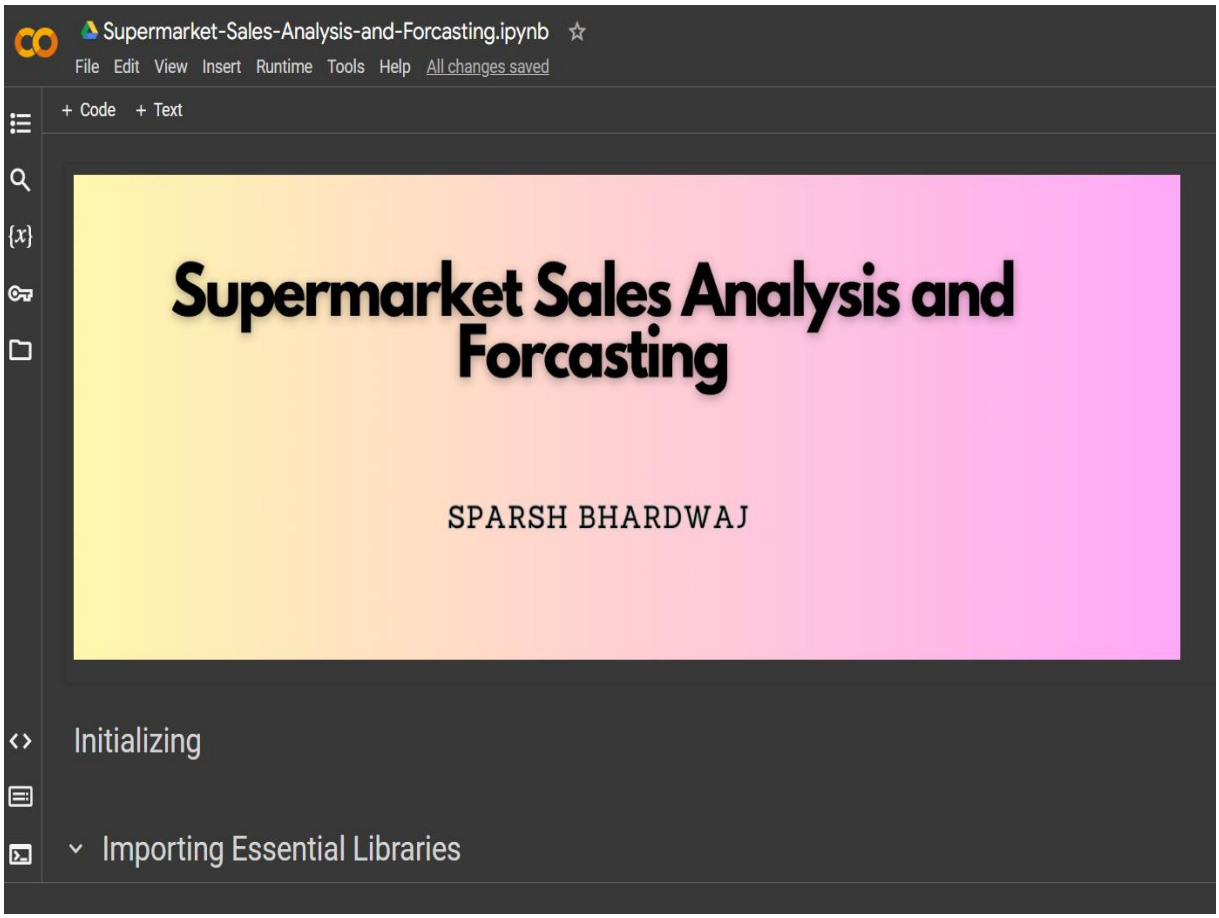
Following files are included in this project which are as:

- Jupyter Notebook (Python Script):
- Dataset File:
- Sales Forecasting Model:
- Visualization of Forecasting Results:
- Google Colab Integration:
- Documentation and Comments:

OVERVIEW

First, we need to import necessary python libraries like pandas, numpy, matplotlib.pyplot, seaborn and arima.

To initiate the program, execute each command individually to gain a comprehensive understanding of how the program operates, progressing from data cleaning and analysis to forecasting. This step-by-step approach will enhance your comprehension of the program's functionality.



1. **Overview of the Dataset:** Reading and Getting the overview of the dataset.

Supermarket-Sales-Analysis-and-Forcasting.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

1. Introduction to the Dataset

Reading the Dataset

```
df = pd.read_csv(dataset_path)
print(df)
```

```
      Invoice ID Branch       City Customer type  Gender  \
0    750-67-8428      A     Yangon        Member  Female
1    226-31-3081      C  Naypyitaw        Normal  Female
2    631-41-3108      A     Yangon        Normal    Male
3    123-19-1176      A     Yangon        Member    Male
4    373-73-7910      A     Yangon        Normal    Male
..           ...    ...        ...           ...     ...
995  233-67-5758      C  Naypyitaw        Normal    Male
996  303-96-2227      B   Mandalay        Normal  Female
997  727-02-1313      A     Yangon        Member    Male
998  347-56-2442      A     Yangon        Normal    Male
999  849-09-3807      A     Yangon        Member  Female

                 Product line  Unit price  Quantity   Tax 5%      Total  \
0         Health and beauty       74.69         7  26.1415   548.9715
1    Electronic accessories       15.28         5   3.8200    80.2200
2        Home and lifestyle       46.33         7  16.2155   340.5255
3         Health and beauty       58.22         8  23.2880   489.0480
4         Sports and travel       86.31         7  30.2085   634.3785
..                      ...         ...       ...      ...        ...
995       Health and beauty       40.35         1   2.0175    42.3675
996      Home and lifestyle       97.38        10  48.6900  1022.4900
```

0s completed at 9:59 PM

2. **Cleaning Data:** Cleaning the Dataset and removing the columns which are not needed for our analysis and forecasting.

Supermarket-Sales-Analysis-and-Forcasting.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Removing Invoice Column as it is not usefull in our Analysis

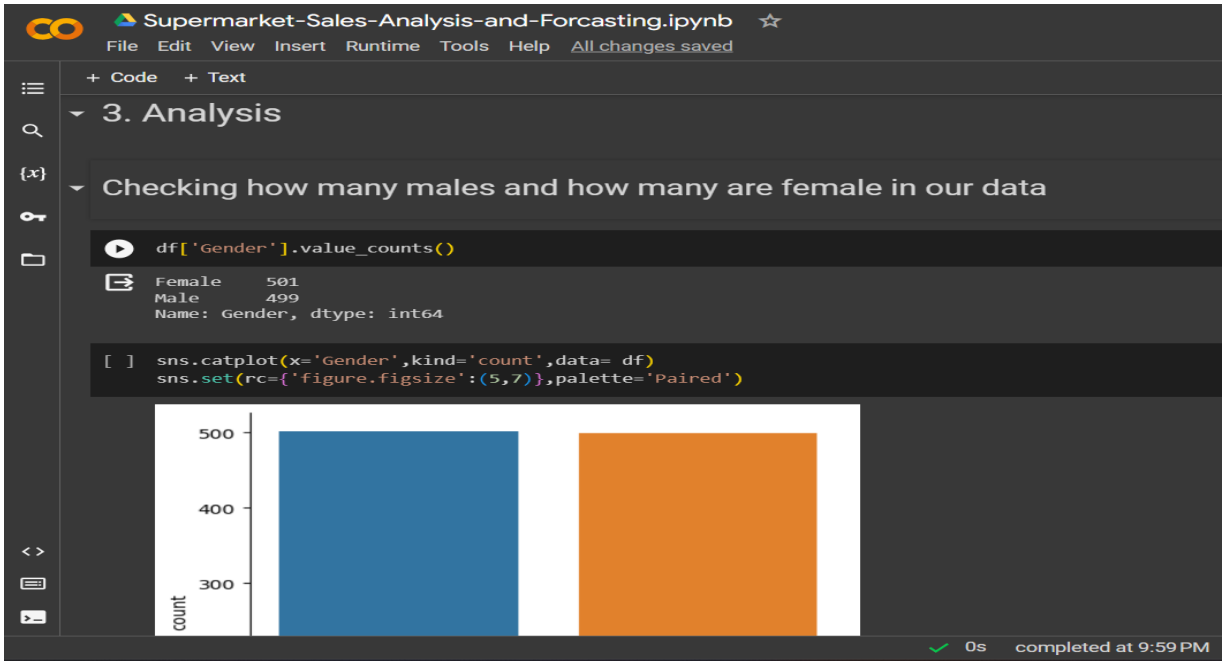
```
df=df.drop(['Invoice ID'],axis=1)
```

```
#checking that the Invoice Column is removed Successfully or not!
df.info()
```

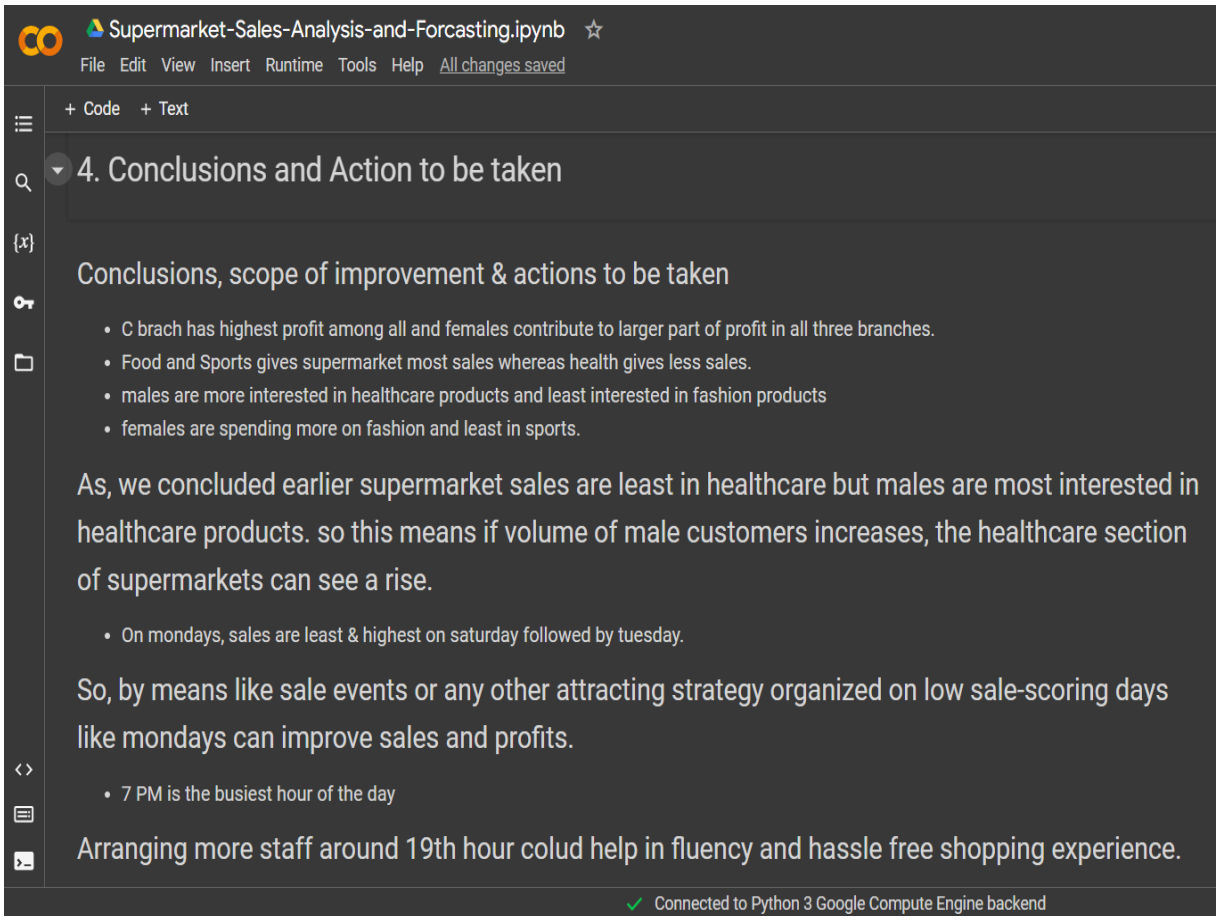
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 16 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Branch                   1000 non-null   object
 1   City                     1000 non-null   object
 2   Customer type            1000 non-null   object
 3   Gender                   1000 non-null   object
 4   Product line             1000 non-null   object
 5   Unit price               1000 non-null   float64
 6   Quantity                 1000 non-null   int64
 7   Tax 5%                   1000 non-null   float64
 8   Total                    1000 non-null   float64
 9   Date                     1000 non-null   object
 10  Time                     1000 non-null   object
 11  Payment                  1000 non-null   object
 12  cogs                     1000 non-null   float64
 13  gross margin percentage  1000 non-null   float64
 14  gross income             1000 non-null   float64
```

0s completed at 9:59 PM

3. **Analysis:** Analyze the dataset by employing various Python functions to extract insights and gain a comprehensive understanding of the data.



4. **Conclusions and Actions to be Taken:**



5. **Prediction and Forecasting:** Predicting the sales for upcoming one month.

Supermarket-Sales-Analysis-and-Forcasting.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

Forcasting

5. Prediction and Forcasting

we make test cases using arima library and sarimax library and furthermore predicting the sales for upcoming one month.

```
[ ] from statsmodels.tsa.stattools import adfuller
    dftest=adfuller(df['cogs'], autolag='AIC')
    dfoutput=pd.Series(dftest[0:4], index=['Test Statistic','p-value','Lags Used','Number of Observations Used'])
    for key,value in dftest[4].items():
        dfoutput['Critical Value (%s)'%key] = value
    print(dfoutput)
```

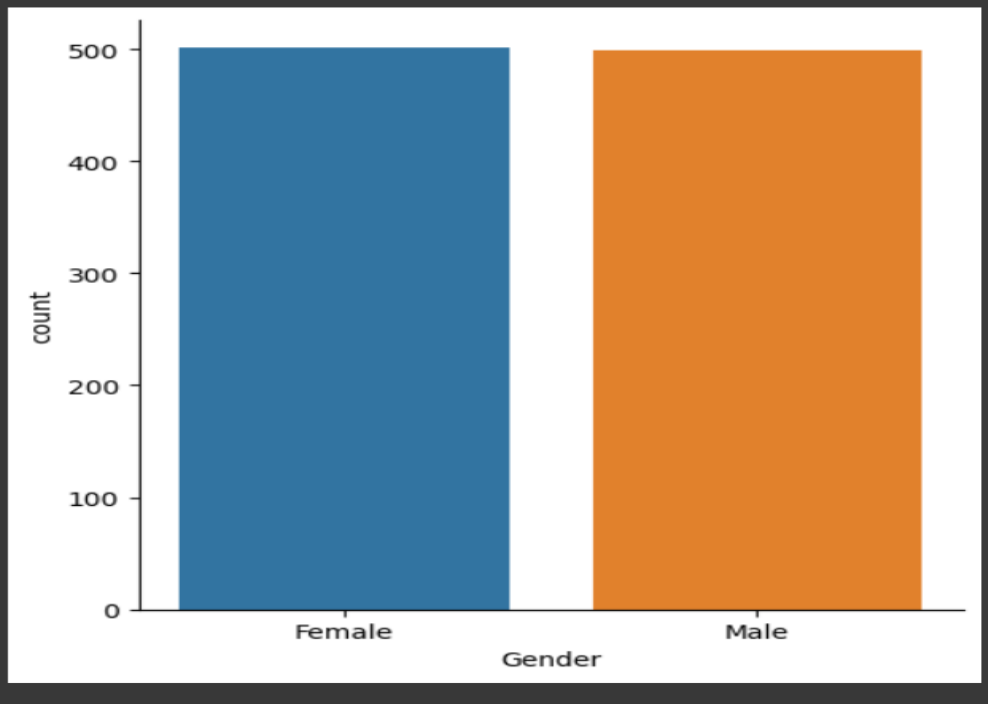
```
Test Statistic                 -30.603524
p-value                          0.000000
Lags Used                        0.000000
Number of Observations Used    999.000000
Critical Value (1%)             -3.436913
Critical Value (5%)             -2.864437
Critical Value (10%)            -2.568313
dtype: float64
```

```
[ ] !pip install pmdarima
```

Connected to Python 3 Google Compute Engine backend

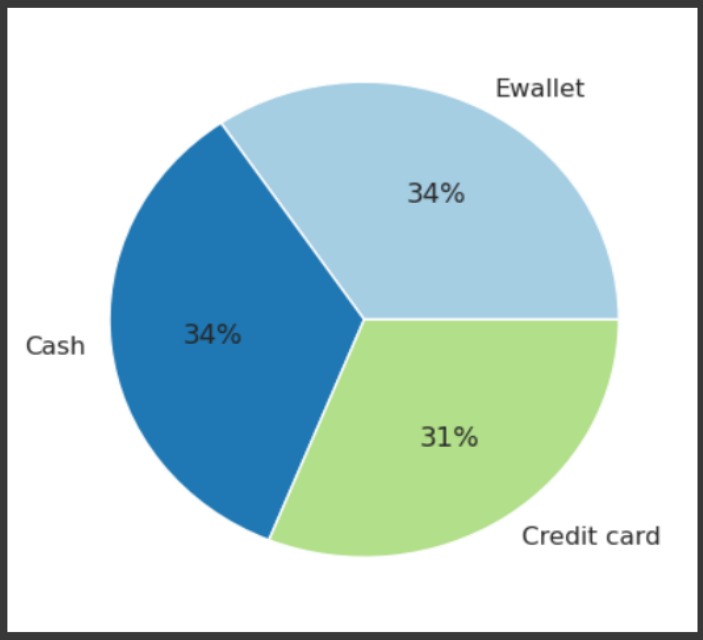
Some Snippets of Code

```
sns.catplot(x='Gender',kind='count',data= df)
sns.set(rc={'figure.figsize':(5,7)},palette='Paired')
```



Lets see what is the most popular payment method used by customers

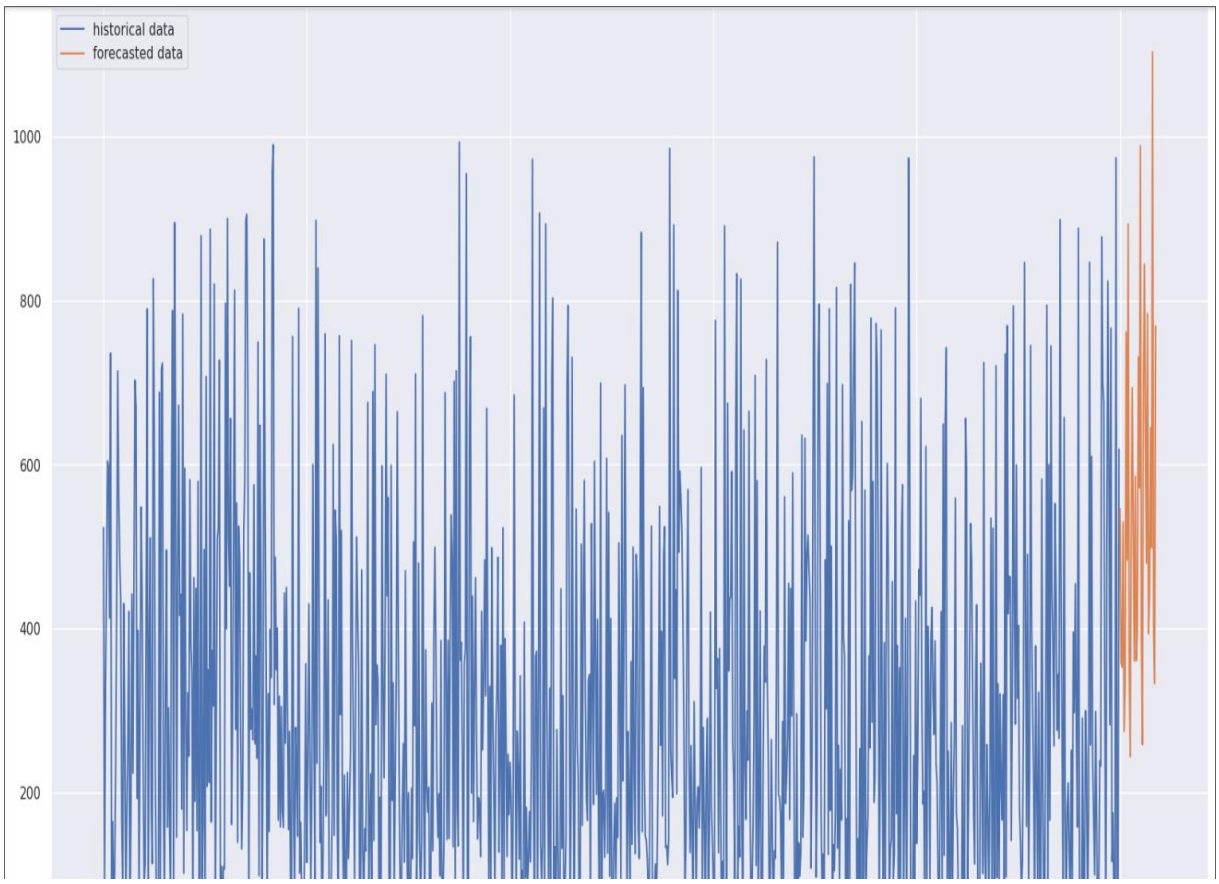
```
plt.pie(df['Payment'].value_counts(),labels=df['Payment'].unique(),autopct='%0.0f%%')
plt.show()
```



Does gross income affect customer rating?

```
[ ] sns.lmplot(x='Rating',y='gross income',data=df,col='Customer type')
```

```
<seaborn.axisgrid.FacetGrid at 0x7dff3c88b7c0>
```



Some of its Uses

In the realm of supermarket sales analysis and prediction and forecasting projects, the traditional approach involved manual tracking of sales transactions and inventory, which was both time-consuming and prone to errors. Subsequently, advancements in technology led to the implementation of data-driven solutions, transforming the way supermarkets operate. Here's a use case highlighting the significance of such a project:

1. **Optimized Inventory Management:**
   - The project analyzes historical sales data to identify peak sales periods, helping supermarkets optimize inventory levels.
   - Prediction models anticipate upcoming demand, reducing instances of overstocking or stockouts.

2. **Dynamic Pricing Strategies:**
   - By analyzing customer purchasing patterns, the system can recommend dynamic pricing strategies.
   - Supermarkets can implement targeted promotions and discounts based on forecasted demand, maximizing revenue.

3. **Improved Customer Experience:**
   - Enhanced inventory management ensures that popular products are consistently available, improving the overall customer experience.
   - Predictive models help supermarkets anticipate customer preferences, offering a tailored shopping experience.

4. **Data-Driven Decision Making:**
   - Store managers and executives have access to insightful visualizations and reports derived from sales analysis.
   - Informed decision-making is facilitated, allowing for adjustments in real-time based on current market trends.

The supermarket sales analysis and prediction and forecasting project revolutionize the retail landscape by offering data-driven insights, improving operational efficiency, and facilitating informed decision-making for supermarkets in an ever-evolving market.

Project Scope and Direction

The primary objective of the supermarket sales analysis and forecasting project is to revolutionize traditional retail management by harnessing the power of data analytics and predictive modeling. The project unfolds in a systematic manner, beginning with the importation of essential Python libraries, including Pandas, NumPy, Seaborn, Matplotlib, and ARIMA. These libraries lay the foundation for a sophisticated analytical approach.

Step 1: Importing Necessary Libraries:
The project initiates by importing vital Python libraries, setting the stage for comprehensive data analysis. Libraries such as Pandas and NumPy enable efficient data manipulation and computation, while Seaborn and Matplotlib facilitate the creation of insightful visualizations. The inclusion of ARIMA (AutoRegressive Integrated Moving Average) demonstrates a commitment to employing advanced time-series forecasting models.

Step 2: Overviewing the Dataset:
A thorough examination of the dataset follows, providing a contextual understanding of the data at hand. This involves studying the structure, format, and variables within the dataset, laying the groundwork for subsequent analytical processes.

Step 3: Cleaning the Dataset:
To ensure the accuracy and reliability of the analysis, data cleaning is imperative. This step involves the removal of unnecessary columns that do not contribute to the forecasting model. Additionally, any null values are addressed, enhancing the dataset's integrity.

Step 4: Analyzing the Dataset Using Graphs and Plots:
The project employs Python's powerful graphing capabilities to visualize and interpret the dataset. Graphs and plots, generated through Seaborn and Matplotlib, unveil trends, patterns, and correlations within the sales data. This visual exploration sets the stage for informed decision-making.

Step 5: Applying ARIMA Model for Sales Prediction:
The crux of the project lies in applying the ARIMA model for sales prediction and forecasting. ARIMA, a robust time-series analysis method, is employed to leverage historical sales data and project future trends. This step involves model training, validation, and fine-tuning to ensure optimal performance.

Impact, Significance and contributions

The supermarket sales analysis and forecasting, executed through Python, signify a paradigm shift in retail management, promising profound contributions to societal and economic dynamics. At its core, the project optimizes the operational landscape of supermarkets, introducing advanced data analytics to streamline processes such as inventory management, staffing, and resource allocation. By leveraging predictive modeling, supermarkets can operate more efficiently, reducing waste and enhancing overall productivity. This not only results in economic benefits for supermarkets but also aligns with sustainability goals by minimizing unnecessary resource consumption and environmental impact.

Moreover, the project places a premium on customer-centricity, tailoring supermarket offerings to individual preferences through a meticulous analysis of historical sales data and customer behaviors. This personalized shopping experience not only satisfies consumer needs but also fosters loyalty and engagement. Importantly, the advent of this data-driven strategy signifies a broader cultural shift in decision-making, steering away from traditional approaches toward a more responsive and adaptable retail landscape.

Beyond its immediate impact on supermarkets, the project contributes to societal development by fostering job creation and skill development. The demand for professionals proficient in data analytics and forecasting grows as supermarkets embrace technology-driven solutions. This not only enhances workforce capabilities but also aligns with broader trends in the evolving intersection of retail and technology. Ultimately, the supermarket sales analysis and forecasting project using Python exemplify a powerful synthesis of technological innovation, economic sustainability, and positive societal impact, offering a blueprint for the future of retail management.

Conclusions

In summary, the Supermarket Sales Analysis and Forecasting Project represents a paradigm shift in retail operations. By harnessing Python libraries and the ARIMA model, the project not only tackles existing limitations but also charts a course toward a data-driven, efficient, and dynamic future for supermarket management.

In essence, the Supermarket Sales Analysis and Forecasting Project stand as a beacon of innovation in retail. Through the strategic use of Python libraries and the application of the ARIMA model, it not only addresses historical inefficiencies but also establishes a foundation for supermarkets to thrive in an era of data-driven decision-making. The project's meticulous approach to dataset overview, cleaning, and analysis, followed by the implementation of advanced forecasting, reflects a commitment to precision and adaptability. As supermarkets navigate the complexities of the market, this project heralds a transformative journey, fostering resilience, efficiency, and strategic foresight.