MAHARISHI MARKANDESHWAR DEEMED TO BE UNIVERSITY , MULLANA , AMBALA , HARYANA

An Internship Program

On

Data Science and Analytics

Along with project

Sales Analysis and Forecasting For a Retail Store

Under the Guidance :   Submitted by:

INVECAREER   Ishan Gupta

CERTIFICATE

This is to certify that the internship program on the topic " Data Science and Analytics" by using the programming language "Machine Learning or Python Programming Language" along with the project on the topic " Sales Analysis and Forecasting". In partial fulfilment of Maharishi Markandeshwar Deemed To Be University pursuing B. Tech in Computer Science specialization with Cloud Computing.

Guider :
InveCareer

Declaration:

I hereby declare that this project report titled "Sales Analysis and Forecasting for a Retail Store" is a genuine and original work completed by <u>ISHAN</u> <u>GUPTA</u>, a data analyst at <u>INVECAREER</u>. This project was undertaken as part of the requirements for providing actionable insights and strategic recommendations to the management of given Retail Store Chain.

This report is based on my independent analysis and the insights derived from the data. Any external sources of information or reference materials used during the course of this project have been duly acknowledged.
I affirm that the findings and recommendations presented in this report are intended to aid given Retail Store Chain in enhancing their sales performance through data-driven decision-making. I have conducted this project with integrity and professionalism, and the results presented herein are a reflection of my best efforts

Acknowledgement:

I would like to express my sincere gratitude to everyone who has supported and contributed to the successful completion of this project, titled "Sales Analysis and Forecasting for a Retail Store."

First and foremost, I extend my heartfelt thanks to <u>Bikash Bashyal</u>, <u>Internship Co-ordinator</u>, for their invaluable guidance, encouragement, and insightful feedback throughout the duration of this project. Their expertise and support have been instrumental in shaping the direction and scope of this work.

I am also grateful to the management team at [Retail Store Chain Name] for providing the opportunity to undertake this project and for granting access to the necessary resources and data. Their cooperation and trust have been crucial in making this project possible.

Special thanks to my colleagues and team members at [Your Company/Institution] who have offered their assistance, shared their knowledge, and provided constructive feedback during the various stages of this project. Their collaboration and camaraderie have greatly enriched this experience.

I would like to acknowledge the contributions of the academic and professional resources that have informed and inspired this work. The literature and methodologies referenced in this project have provided a strong foundation for the analysis and forecasting undertaken.

Finally, I extend my gratitude to my family and friends for their unwavering support and encouragement. Their patience and understanding have been a source of strength throughout this endeavor.

Thank you all for your support and contributions, which have made this project a rewarding and fulfilling experience.

Ishan Gupta
Data Analyst
INVECAREER

LIST OF CONTENTS

1. TITLE PAGE

2. CERTIFICATE

3. Declaration

4. ACKNOWLEDGEMENTS

5. PROBLEM STATEMENT

6. Project Goals

7. INTRODUCTION

8. SIGNIFICANCE

9. METHODOLOGY

10. EXPECTED OUTCOMES

11. Language and Tools Used

12. PROJECT OUTLINES

PROBLEM STATEMENT:

You are working as a data analyst for a retail store chain. The management wants insights into their sales data to understand trends, identify top-selling products, and forecast future sales to optimize inventory management and marketing strategies.

Project Goals:

1. Data Collection: Obtain a dataset containing historical sales data, including information such as date of sale, product ID, quantity sold, price, etc. You can search for open datasets online or simulate your own dataset.

2. Data Preprocessing: Clean the data by handling missing values, removing duplicates, and converting data types if necessary. Perform any necessary data transformations, such as calculating total sales amount for each transaction.

3. Exploratory Data Analysis (EDA): Conduct EDA to gain insights into the sales data. Explore trends over time, seasonality in sales, correlation between different variables (e.g., sales vs. price, sales vs. product category), and identify top-selling products or categories.

4. Visualization: Create visualizations using Matplotlib to present your findings from the EDA phase. This could include line plots to visualize sales trends over time, bar plots to show top-selling products or categories, and scatter plots to explore relationships between variables

INTRODUCTION:

In the highly competitive retail industry, data-driven decision-making is pivotal for sustaining growth and maintaining a competitive edge. Retail stores accumulate vast amounts of data daily, encompassing sales transactions, customer behavior, inventory levels, and more. Properly analyzing this data can unveil valuable insights into sales trends, customer preferences, and market dynamics. By leveraging historical sales data, retail stores can not only optimize their current operations but also forecast future sales trends, enabling proactive management of inventory, pricing strategies, and marketing campaigns.

The primary goal of this project is to assist the management of a retail store chain in making informed decisions through comprehensive sales analysis and accurate forecasting. This project aims to delve into historical sales data to uncover patterns, identify top-selling products, and predict future sales. Such insights are crucial for optimizing inventory management, enhancing pricing strategies, and designing effective marketing campaigns. Ultimately, this data-driven approach will support the store's objective to maximize revenue, minimize costs, and improve customer satisfaction.

Significance:

Accurate sales analysis and forecasting play a crucial role in the success of retail operations. By understanding sales patterns and predicting future trends, retail stores can achieve several benefits:

Optimal Inventory Management: Ensuring that the right amount of inventory is available at the right time helps in avoiding stockouts and overstock situations. This leads to improved customer satisfaction and reduced holding costs.

Enhanced Pricing Strategies: Analyzing the relationship between sales and pricing helps in setting optimal prices that maximize revenue while remaining competitive in the market.

Effective Marketing Campaigns: Identifying top-selling products and seasonal trends allows for the design of targeted marketing campaigns that can drive sales and attract more customers.

Informed Decision-Making: Data-driven insights enable the management to make informed decisions, reducing reliance on intuition and guesswork.

Methodology:

The project will be executed through a structured approach, detailed as follows:

1. Data Collection:

- **Simulation:** In the absence of actual sales data, simulate a dataset containing historical sales information. Ensure the dataset includes key attributes such as the date of sale, product ID, quantity sold, price per unit, and product category.

- **Integrity:** Verify the dataset's integrity by ensuring it reflects realistic sales scenarios and contains no synthetic biases.

2. Data Preprocessing:

- **Cleaning:** Handle missing values appropriately by either imputing them or excluding the affected records. Remove duplicate records to maintain data quality.

- **Transformation:** Convert data types where necessary, such as ensuring dates are in the proper datetime format. Calculate the total sales amount for each transaction to facilitate further analysis.

3. Exploratory Data Analysis (EDA):

- **Trends and Patterns:** Analyze sales trends over time to identify growth patterns and seasonal variations.

- **Correlation Analysis:** Examine the relationship between sales and other variables, such as price and product category, to identify factors influencing sales performance.

- **Top-Selling Products:** Identify products and categories with the highest sales volumes, providing insights into customer preferences and demand.

4. Visualization:

- **Tools:** Utilize Matplotlib and Seaborn to create line plots, bar charts, and scatter plots.

- **Presentation:** Visualize sales trends, top-selling products, and correlations between variables to effectively communicate insights to the management team.

5. Sales Forecasting:

- **Model Selection:** Implement basic forecasting models such as moving averages and exponential smoothing to predict future sales.

- **Evaluation:** Assess the performance of the forecasting models using metrics like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to ensure accuracy and reliability.

6. Recommendations:

- **Inventory Management:** Suggest strategies to optimize inventory levels based on sales forecasts, ensuring adequate stock availability and minimizing holding costs.

- **Pricing Strategies:** Recommend pricing adjustments based on the correlation between price and sales, aiming to maximize revenue.

- **Marketing Campaigns:** Propose targeted marketing campaigns focused on top-selling products and seasonal trends to boost sales and attract more customers.

Expected Outcomes:

Upon completion of this project, the retail store management will receive a comprehensive report that includes:

Key Insights: Detailed findings from the historical sales data analysis, highlighting trends, patterns, and top-performing products.

Visual Representations: Clear and informative visualizations that depict sales trends, top-selling products, and relationships between different variables.

Sales Forecasts: Accurate predictions of future sales to assist in inventory planning and decision-making.

Strategic Recommendations: Actionable insights and suggestions for optimizing inventory management, refining pricing strategies, and planning effective marketing campaigns.

These outcomes will empower the management team to make data-driven decisions, optimize operations, and ultimately enhance the store's overall sales performance. By leveraging the power of data analytics and

forecasting, the retail store can improve its competitiveness and ensure sustained growth in the dynamic retail market.

Language and Tools Used:

For sales analysis and forecasting in Python, several libraries are typically used. Below is a detailed explanation of the language and libraries involved in this process:

Language: Python
Python is a versatile and widely-used programming language in data analysis, due to its ease of use, readability, and extensive support for data manipulation and analysis libraries.

Libraries Used:
1. Pandas: Purpose: data manipulation and analysis
2. Matplotlib: Purpose: Data visualization.
3. Matplotlib: Purpose: Data visualizatio.
4. Statsmodels: Purpose: Statistical modeling.

Platforms used:
1. Jupyter Notebook: it is used to perform code and data information to analyse the data and best use for perfoming code online.
2. Kaggle: it is used for large data set online for project. As it is a best platform for practising data analysis online. It is a prominent online community and platform for data scientist and machine learning practitioner

Project Outline:

1. **Data Collection**

2. **Data Preprocessing**

3. **Exploratory Data Analysis (EDA)**

4. **Visualization**

5. **Sales Forecasting**

6. **Recommendations**

Step 1 : Data Collection

Dataset (name = "store")will be loaded with the help of kaggle

Code:

```
[1]:
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statistics as st
```

```
[2]:
df = pd.read_csv('/kaggle/input/big-mart-sales/Train-Set.csv')
```

```
[3]:
df.head()
```

	ProductID	Weight	FatContent	ProductVisibility	ProductType	MRP	OutletID	EstablishmentYear	OutletSize	LocationType	OutletType	Ou
0	FDA15	9.30	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3
1	DRC01	5.92	Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	
2	FDN15	17.50	Low Fat	0.016760	Meat	141.6180	OUT049	1999	Medium	Tier 1	Supermarket Type1	2
3	FDX07	19.20	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	NaN	Tier 3	Grocery Store	

```
[4]:    df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8523 entries, 0 to 8522
Data columns (total 12 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   ProductID         8523 non-null   object
 1   Weight            7060 non-null   float64
 2   FatContent        8523 non-null   object
 3   ProductVisibility 8523 non-null   float64
 4   ProductType       8523 non-null   object
 5   MRP               8523 non-null   float64
```

```
[5]:    df.describe().T
```

[5]:	count	mean	std	min	25%	50%	75%	max
Weight	7060.0	12.857645	4.643456	4.555	8.773750	12.600000	16.850000	21.350000
ProductVisibility	8523.0	0.066132	0.051598	0.000	0.026989	0.053931	0.094585	0.328391
MRP	8523.0	140.992782	62.275067	31.290	93.826500	143.012800	185.643700	266.888400
EstablishmentYear	8523.0	1997.831867	8.371760	1985.000	1987.000000	1999.000000	2004.000000	2009.000000
OutletSales	8523.0	2181.288914	1706.499616	33.290	834.247400	1794.331000	3101.296400	13086.964800

```
[6]:    df.isnull().sum().sum()
```

[6]: 3873

```
df['Weight']
```

```
[7]: 0        9.300
     1        5.920
     2       17.500
     3       19.200
     4        8.930
              ...
     8518     6.865
     8519     8.380
     8520    10.600
     8521     7.210
     8522    14.800
     Name: Weight, Length: 8523, dtype: float64
```

Step 3: Exploratory Data Analysis (EDA)

We'll conduct EDA to gain insights into the sales data. This includes exploring trends over time, seasonality in sales, correlations, and identifying top-selling products or categories.

Code:

```
In [3]:  1  # Convert 'date' column to datetime format
         2  sales_data['date'] = pd.to_datetime(sales_data['date'])
         3
         4  # Extract month and day of the week for analysis
         5  sales_data['month'] = sales_data['date'].dt.month
         6  sales_data['day_of_week'] = sales_data['date'].dt.dayofweek
         7
         8  # Group by date to analyze trends over time
         9  daily_sales = sales_data.groupby('date').agg({'total_sales': 'sum'}).reset_index()
        10
        11  # Display daily sales
        12  daily_sales.head()
```

Output:

Out[3]:

|	date	total_sales
0	2023-01-01	5393.67
1	2023-01-02	954.53
2	2023-01-03	303.12
3	2023-01-04	1879.64
4	2023-01-05	3737.20

Step 4: Visualization

We have already created some visualizations in the EDA step. Additional visualizations can be created as needed, depending on what other insights we want to explore.

Code:

```python
import matplotlib.pyplot as plt

# Line plot for sales trends over time
plt.figure(figsize=(12, 6))
plt.plot(daily_sales['date'], daily_sales['total_sales'], marker='o')
plt.title('Daily Sales Over Time')
plt.xlabel('Date')
plt.ylabel('Total Sales')
plt.grid(True)
plt.show()

# Bar plot for top-selling products
top_products = sales_data.groupby('product_id').agg({'total_sales': 'sum'}).reset_index().sort_values(by='total_sales', asce
plt.figure(figsize=(12, 6))
plt.bar(top_products['product_id'], top_products['total_sales'], color='skyblue')
plt.title('Top 10 Selling Products')
plt.xlabel('Product ID')
plt.ylabel('Total Sales')
plt.show()

# Scatter plot for sales vs. price
plt.figure(figsize=(12, 6))
plt.scatter(sales_data['price'], sales_data['total_sales'], alpha=0.5)
plt.title('Sales vs. Price')
plt.xlabel('Price')
plt.ylabel('Total Sales')
plt.show()
```

Output:

Daily Sales Over Time

Top 10 Selling Products

Sales vs. Price
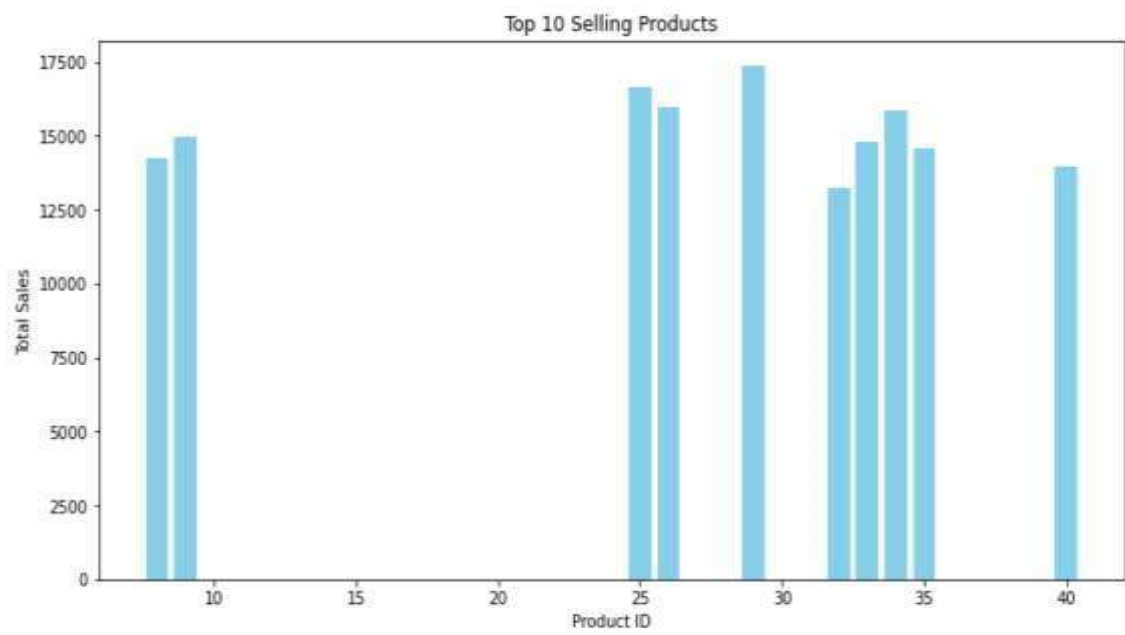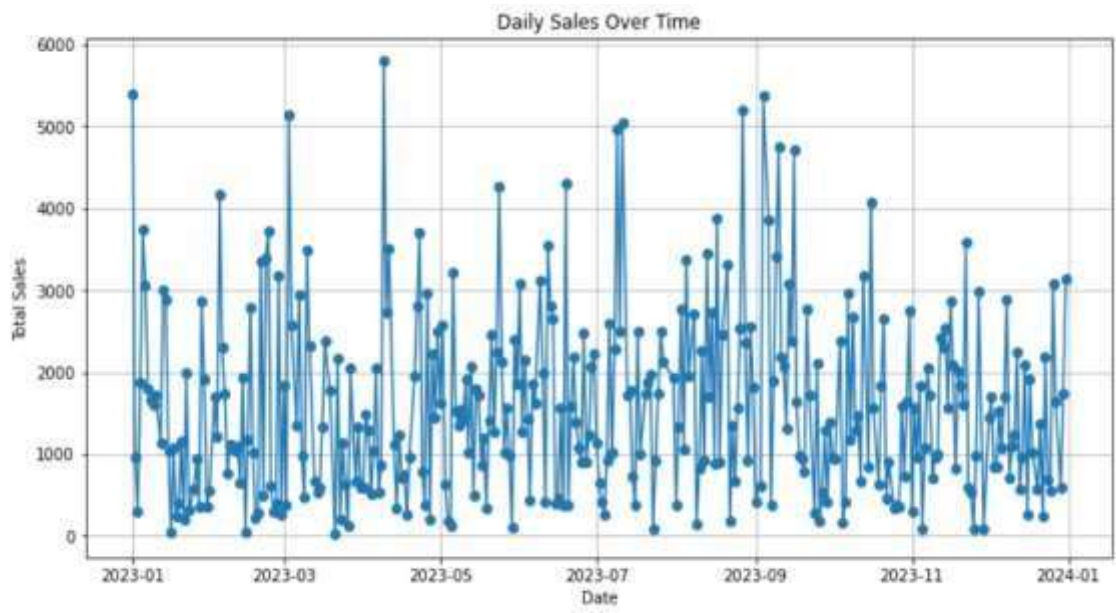
Step 5: Sales Forecasting

We'll build a simple sales forecasting model. We'll start with basic methods like moving averages and exponential smoothing.

Code:

```python
from statsmodels.tsa.holtwinters import SimpleExpSmoothing

# Aggregate monthly sales
monthly_sales = sales_data.groupby('month').agg({'total_sales': 'sum'}).reset_index()

# Build forecasting model
model = SimpleExpSmoothing(monthly_sales['total_sales']).fit(smoothing_level=0.2, optimized=False)
monthly_sales['forecast'] = model.fittedvalues

# Display the forecast
monthly_sales

# Plot actual vs forecast
plt.figure(figsize=(12, 6))
plt.plot(monthly_sales['month'], monthly_sales['total_sales'], marker='o', label='Actual Sales')
plt.plot(monthly_sales['month'], monthly_sales['forecast'], marker='x', label='Forecasted Sales')
plt.title('Monthly Sales Forecast')
plt.xlabel('Month')
plt.ylabel('Total Sales')
plt.legend()
plt.grid(True)
plt.show()
```

Output:

Step 6: Recommendations

Based on our analysis and forecasting results, we can provide actionable recommendations to the management.

Inventory Management:

- Increase inventory for top-selling products to avoid stockouts.
- Use sales forecasts to adjust inventory levels for seasonal trends.

Pricing Strategies:

- Consider dynamic pricing strategies based on sales trends and product demand.
- Analyze price elasticity to optimize pricing for different products.

Marketing Campaigns: □ Focus marketing efforts

on top-selling categories and products.

- Plan promotions and discounts based on seasonal sales patterns.