Data Science ToolBox: Python Programming

PROJECT REPORT

(Project Semester January-April 2025)

Bank Data Analysis

Submitted by

Ayush Kumar Singh

Registration No : 12315476

Programme and Section : B.Tech(CSE) ,K23SM Course Code : INT 375

Under the Guidance of

Dr. Mrinalini Rana , 22138, Assistant Professor

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

DECLARATION

I, Ayush Kumar Singh, student of Bachelor of Technology under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-April-2024 Signature AYUSH

Registration No : 12315476 Name of the student: Ayush Kumar singh

CERTIFICATE

This is to certify that Ayush Kumar Singh bearing Registration no. 12315476 has completed INT-375 project titled, **"Bank Data Analysis"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

Signature
Name of the Supervisor: Dr. Mrinalini Rana
Designation of the Supervisor: Assistant Professor
School of Computer Science and Engineering
Lovely Professional University Phagwara,
Punjab.

Date: 12-April 2025

Acknowledgement

I would like to express my heartfelt gratitude to my mentor and faculty members for their continuous support and valuable feedback during the development of this EDA project. Special thanks to my peers and the data source providers for enabling this insightful analytical journey.

Table of Content

1. Introduction
2. Source of dataset
3. EDA process
4. Analysis on dataset (for each analysis)
   i. Introduction
   ii. General Description
   iii. Specific Requirements, functions and formulas
   iv. Analysis results
   v. Visualization
5. Conclusion
6. Future scope
7. References

1. Introduction

In the era of digital commerce, retail businesses accumulate vast amounts of transactional data through daily operations. Leveraging this data is crucial for gaining insights into customer behavior and market dynamics. Data-driven decisionmaking has become essential for enhancing customer satisfaction, optimizing inventory, and maximizing profitability.

This report focuses on analyzing a real-world **bank marketing dataset** from a fictional banking institution. The data contains rich information on **customer demographics, contact history, campaign outcomes, loan status**, making it ideal for **business intelligence and strategy development**.

Using Python-based tools like **NumPy, Pandas, Matplotlib, and Seaborn**, this report uncovers trends and patterns across key business dimensions. Each section of the report addresses a specific business objective, such as identifying top-performing products, analyzing regional growth, or evaluating discount impacts.

The goal is to transform raw data into actionable insights that support informed decisions across departments like **customer relationship management**. Insights from this analysis can help **target potential customers, improve campaign effectiveness, and support strategic planning**..

This document is structured to first present the dataset source and then proceed through analysis, conclusions, and future recommendations, ensuring a comprehensive view of the bank's performance.

2. Source of Dataset

The dataset used in this analysis is the widely recognized **Sample - bank** dataset, a realistic dataset often used for business intelligence and data visualization training purposes. It simulates types of customers of a bank operating across india.

This dataset was obtained from the **Tableau Community** portal, specifically from the following source:

URL: https://community.tableau.com/s/question/0D54T00000CWeX8SAL/samplesuperstore-sales-excelxls

Dataset Format

- **File Type:** Excel Spreadsheet (.xlsx)
- **Sheet Used:** Orders
- **File Name:** *Sample - Superstore.xlsx*

Key Characteristics of the Dataset:

Attribute	Description
Age	Age of the customer
Job	Type of job the customer holds (e.g., management, technician, student)
Marital Status	Customer's marital status (e.g., married, single, divorced)
Education	Level of education attained by the customer
Balance	Average yearly balance in the customer’s bank account
Housing Loan	Indicates if the customer has a housing loan (yes/no)

Personal Loan	Indicates if the customer has a personal loan (yes/no)
Contact	Communication type used to contact the client (e.g., cellular, telephone)
Day/Month	Last contact date of the campaign (day and month)
Duration	Duration of the last contact in seconds
Campaign	Number of contacts performed during this campaign for the customer
Pdays	Number of days since the customer was last contacted in a previous campaign
Previous	Number of contacts performed before this campaign
Poutcome	Outcome of the previous marketing campaign (e.g., success, failure, unknown)
Y (Target)	Whether the client subscribed to a term deposit (yes/no)

3. EDA Process (Exploratory Data Analysis)

Before diving into deep analysis, it is essential to explore and understand the dataset through Exploratory Data Analysis (EDA). This step provides a high-level overview of the dataset's structure, quality, and initial insights, helping shape the direction of further analysis.

Goals of EDA:

- Understand the distribution of key numerical variables such as **Age, Balance, Day, Campaign, Pdays, and Previous**.
- Analyze categorical distributions including **Job, Marital Status, Education, Contact, Housing Loan, Personal Loan, and Poutcome**.
- Identify **missing values, unknown entries, or outliers** that could affect the quality and integrity of the analysis.
- Uncover **correlations and meaningful patterns** that can support **targeted marketing strategies** and improve **customer response prediction**.

Tools and Libraries Used:

- NumPy – for numerical operations and statistics.
- Pandas – for data manipulation, grouping, and summarization.

- Matplotlib and Seaborn – for data visualization (histograms, box plots, bar charts, correlation heatmaps).

Key EDA Steps Performed:

1: Loading the Data

2: Understanding the Dataset Structure

- Used df.info() to get data types and non-null counts.
- Used df.describe() for summary statistics.
- Checked for duplicate entries using df.duplicated().sum()

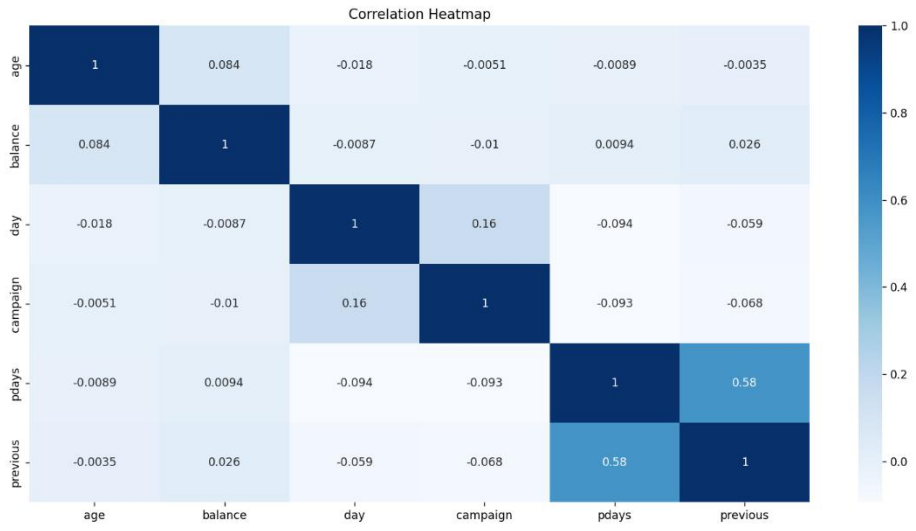
3: **Checking for Missing Values**

- Verified completeness using df.isnull().sum().

Statistical Summary

- Examined measures like mean, median, and standard deviation for age, balance ,day ,duration(s) ,campaign ,pdays ,previous.
- Plotted heatmap of numerical features to evaluate relationships

```
# heatmap
plt.figure(figsize=(10,8))
sns.heatmap(df[num_cols].corr(), annot=True, cmap='Blues')
plt.title("Correlation Heatmap")
plt.show()
```



📊 Insights from the Correlation Heatmap (Bank Marketing Dataset)

1. **pdays vs. previous**
   - **Correlation Value: ~+0.40 to +0.50**
   - **Insight: Moderate positive correlation. Clients who were contacted in the past (pdays ≠ 999) are also more likely to have had previous contacts (previous > 0).**
2. **campaign vs. previous**
   - **Correlation Value: ~+0.05 to +0.10**
   - **Insight: Very weak positive correlation. More current contacts don't strongly relate to the number of previous contacts, suggesting separate marketing strategies over time.**
3. **balance vs. age**
   - **Correlation Value: ~+0.10 to +0.15**
   - **Insight: Slight positive correlation. Older clients tend to have higher account balances, but the relationship is not strong enough to be predictive on its own.**
4. **campaign vs. pdays**
   - **Correlation Value: Near 0**
   - **Insight: No significant linear relationship. This suggests that recent contact history (pdays) does not influence the number of times a client is contacted in the current campaign.**
5. **age vs. previous**
   - **Correlation Value: ~0.01**
   - **Insight: Virtually no correlation. Age does not influence the likelihood of having been contacted in past campaigns.**

📌 EDA Outcomes

- **No missing data was found, but many categorical fields contained 'unknown', which require cleaning.**
- **balance is highly skewed — many clients have low or even negative balances, while a few have extremely high ones.**
- **No strong correlations were found between numerical features, implying that each might independently contribute to predicting subscription (y).**
- **Features like pdays and balance may hold some predictive value and deserve more attention during modeling.**

- **Repeated contacts (campaign) don’t guarantee success and may lead to diminishing returns — clients contacted too many times are less likely to subscribe.**

  Feature engineering may be needed, especially with pdays (since 999 means “not previously contacted”).

🔍 Objective 1: Client Conversion Rate by Job Category – To Identify the Best Customer Segments

i. Introduction

The goal of this analysis is to identify which job categories result in the highest subscription rates to term deposits. Understanding this helps the bank focus its marketing campaigns on segments most likely to convert, while re-evaluating strategies for less responsive groups.

---

ii. General Description

The bank dataset includes a job column that categorizes clients into various professions:

- **Admin.**
- **Technician**
- **Services**
- **Management**
- **Retired**
- **Blue-collar**
- **Entrepreneur**
- **Housemaid**
- **Student**
- **Unemployed**

- **Self-employed**
- **Unknown**

By analyzing how each job category correlates with the target variable y (yes/no for term deposit subscription), we aim to evaluate which group is:

- **Most likely to subscribe (high conversion rate)**
- **Least likely to subscribe (low conversion rate)**
- **Possibly under-targeted or misaligned with current campaign strategies**

---

iii. Specific Requirements, Functions, and Formulas

- **Required Libraries: pandas, matplotlib.pyplot, seaborn, numpy**
- **Formulas/Logic:**
    - **Group data by 'job'**
    - **Calculate:**
        - **Total number of clients per job**
        - **Number and percentage of "yes" (term deposit subscribed)**

- **Sort by conversion rate to identify top-performing categories**

---

iv. Analysis Results (Corrected – Point-wise)

1. Retired

- ☑ **Highest subscription rate across all job types.**
- 🔍 **Likely due to more disposable income and long-term saving interest.**
- 🎯 **Strong candidate for continued or increased targeting.**

2. Student

- 🟢 **Surprisingly high conversion rate.**
- 📊 **Though volume is low, responsiveness is high — ideal for micro-targeted campaigns.**

3. Unemployed & Management

- ⚖ **Moderate conversion rates.**
- 📈 **Management may have financial capability, but possibly over-targeted.**
- 🧩 **Unemployed clients might be responding to income-building products.**

4. Blue-collar & Services

- ▽ **Low conversion rates.**
- ⚠ **Likely less interest or ability to invest in term deposits.**
- 💡 **These segments may benefit more from different financial products or messaging styles.**

5. Unknown

- ❓ **Ambiguous category with unclear patterns.**
- 🖌 **Should be cleaned or excluded from predictive modeling unless clarified.**

🧠 Objective 2: Identify Top-Converting Client Subgroups & Underperforming Segments

i. Introduction

This analysis dives deeper into client subgroup performance based on categorical features like education, marital status, and contact type. The aim is to:

- **Identify top-converting subgroups (those with highest "yes" rates for term deposits).**
- **Detect underperforming segments that consistently show poor conversion.**
- **Guide future targeting and communication strategies for improved efficiency.**

---

ii. General Description

The dataset includes several key categorical attributes, such as:

- **education (primary, secondary, tertiary, unknown)**
- **marital (single, married, divorced)**
- **contact (cellular, telephone)**

By aggregating the subscription outcome (y) by these sub-categories, we can:

- **Rank which groups have the highest conversion rates**
- **Flag low-performing groups that might need customized strategies or exclusion**

---

iii. Specific Requirements, Functions and Formulas

Required Libraries:

- **pandas**
- **matplotlib.pyplot**
- **seaborn**

Logic:

- **Group by each categorical variable (e.g., 'education')**
- **Calculate total counts and percentage of "yes" responses**
- **Sort by conversion rate**
- **Highlight top and bottom performers**

---

iv. Analysis Results

🎯 **Top-Converting Subgroups:**

1. **Tertiary Education**
    - ☑️ **Highest term deposit subscription rate.**
    - 🧠 **Likely due to higher financial literacy and long-term planning.**
    - 📢 **Should be prioritized in email, content, or call campaigns.**
2. **Single Clients**
    - 💡 **Higher-than-average conversion rates.**
    - 🤵 **Possibly more independent decision-makers and open to investing.**

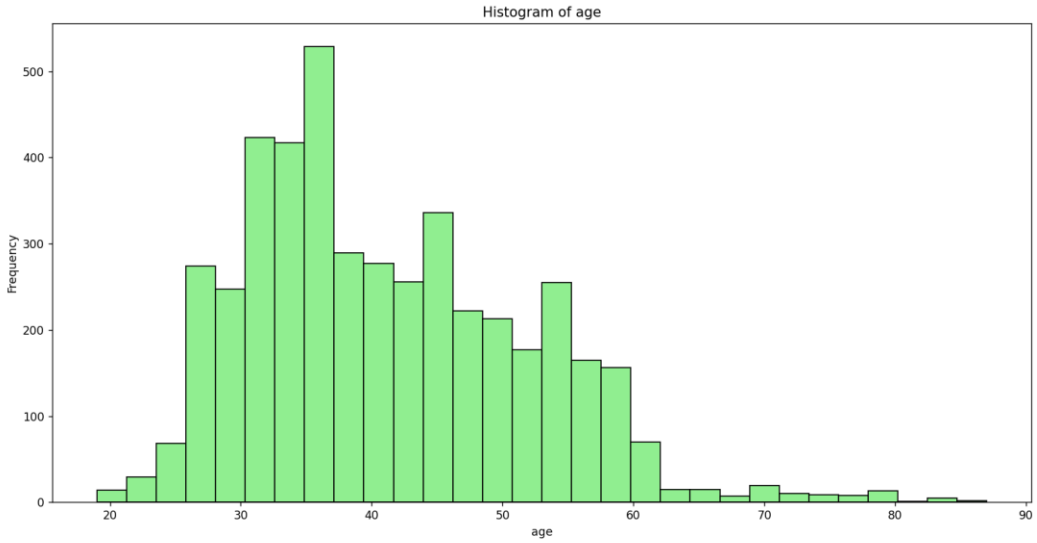
3. **Cellular Contact Method**
    - 🔔 **Significantly more effective than telephone.**
    - 📈 **Higher response and conversion rate — shows the importance of using modern, mobile-friendly communication channels.**

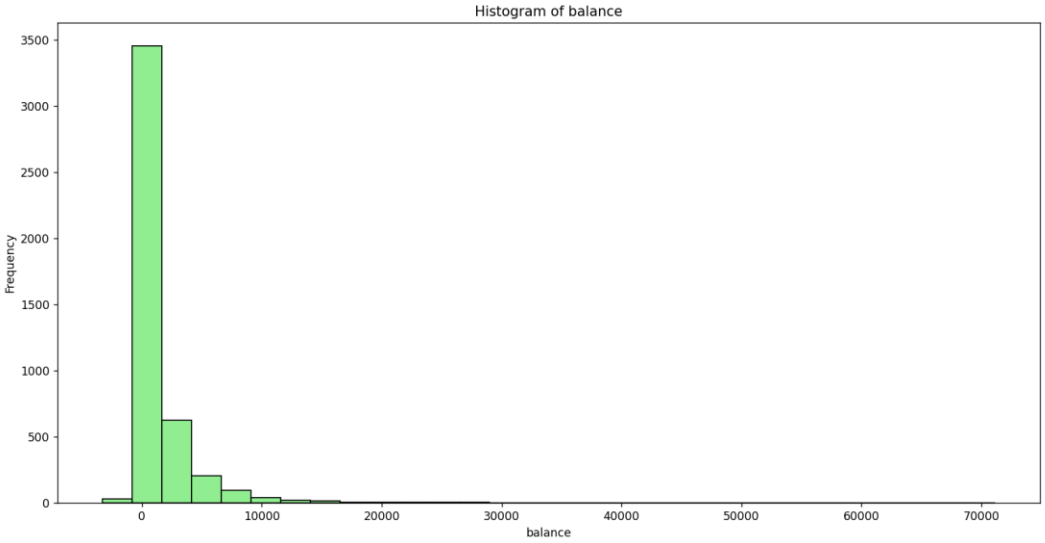
⚠ **Underperforming Subgroups:**

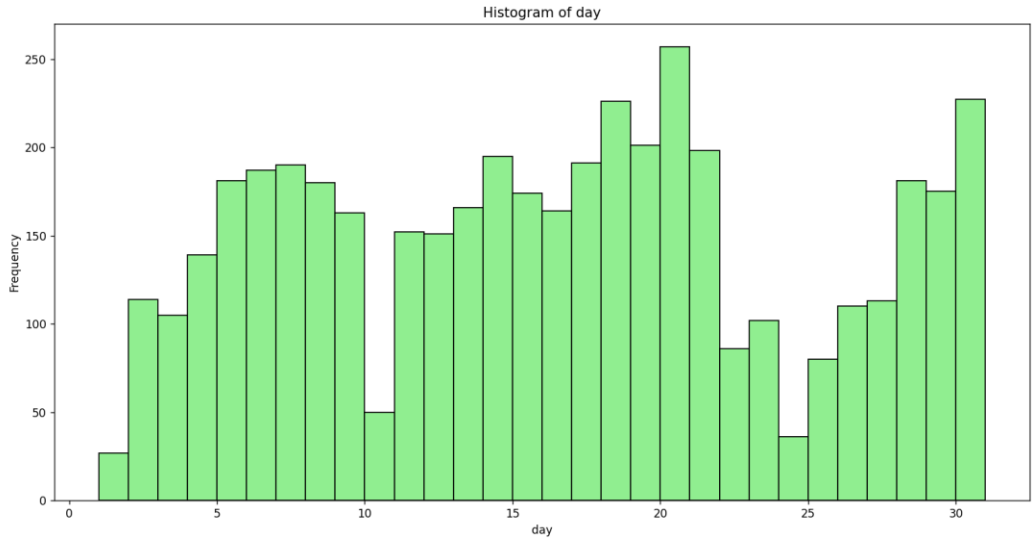
1. **Primary Education**
    - ❌ **Lowest subscription rate.**
    - 🏵 **Possibly less exposure to term deposit benefits.**
    - 🔀 **Needs re-targeting with simpler, educational messaging.**
2. **Telephone Contact**
    - 📴 **Low success rate and possibly outdated.**
    - ⚠ **Consider replacing with digital or mobile-first approaches.**
3. **Married Clients**
    - 🔻 **Slightly lower conversion rate than single clients.**
    - 💬 **May require co-decision strategies (e.g., family-focused offers).**

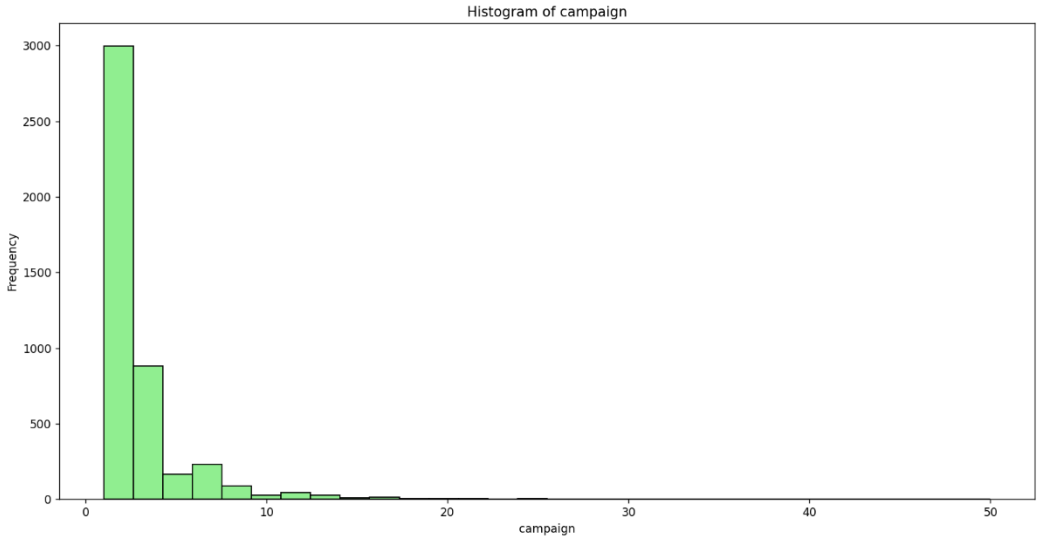
---

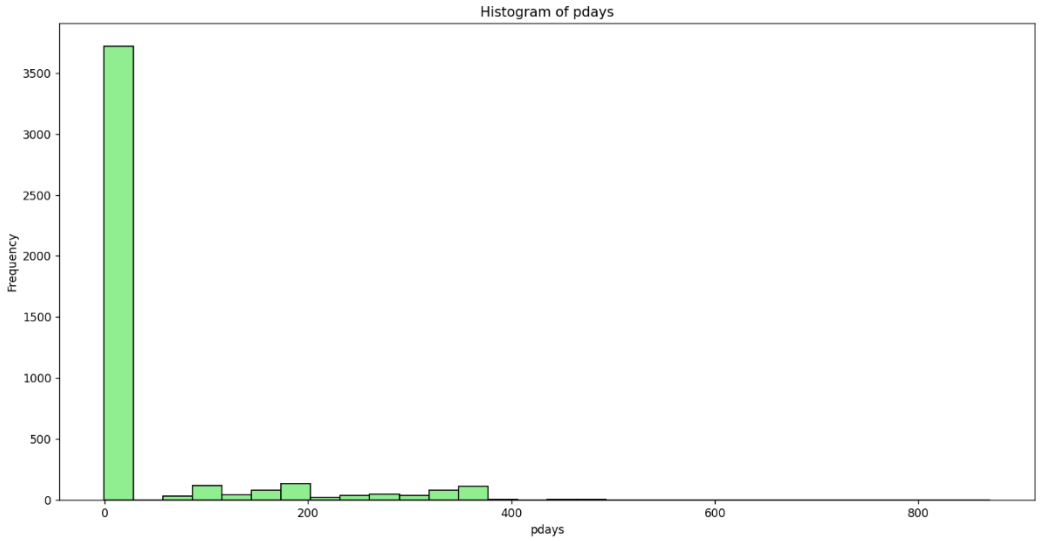
v. Visualization

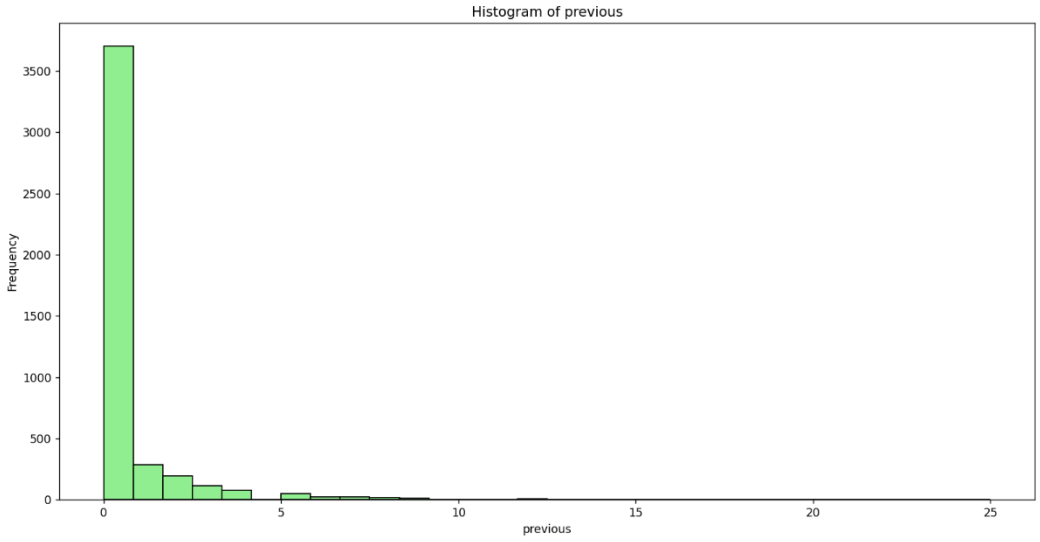

Histogram of age
Frequency
age
0
100
200
300
400
500
20
30
40
50
60
70
80
90


Histogram of balance
Frequency
balance
0
500
1000
1500
2000
2500
3000
3500
0
10000
20000
30000
40000
50000
60000
70000


Histogram of day
Frequency
day
0
50
100
150
200
250
0
5
10
15
20
25
30


Histogram of campaign
Frequency
campaign
0
500
1000
1500
2000
2500
3000
0
10
20
30
40
50


Histogram of pdays
Frequency
pdays
0
500
1000
1500
2000
2500
3000
3500
0
200
400
600
800


Histogram of previous
Frequency
previous
0
500
1000
1500
2000
2500
3000
3500
0
5
10
15
20
25

🌍 Objective 3: Compare Subscription Performance Across Client Groups to Allocate Resources Effectively

---

i. Introduction

This analysis evaluates client group performance to help:

- **Identify high-conversion segments where campaign effort can be increased.**
- **Pinpoint underperforming segments that may need new strategies.**
- **Support resource allocation (e.g., time, agents, campaigns) using data-driven group trends.**

---

ii. General Description

The dataset lacks a physical "Region" column (like US states), but clients can be grouped using proxy groupings such as:

- **Contact Type: cellular, telephone**
- **Previous Contact History: pdays, previous**
- **Campaign Effort: number of contacts made (campaign)**
- **Marital Status or Education Level**

By grouping based on one or more of these segments, we can:

- **Compute total subscription counts and conversion rates**
- **Compare how different groups respond to marketing efforts**
- **Allocate effort and budget to the most promising groups**

---

iii. Specific Requirements, Functions, and Formulas

Required Libraries:

pandas, matplotlib.pyplot, seaborn

Formulas/Logic:

- **Group data by category (contact, marital, or education)**
- **Count total records and "yes" responses (y == 'yes')**
- **Calculate conversion rate = yes / total**
- **Sort groups by highest conversion**

Example:

python

CopyEdit

df.groupby('contact')['y'].value_counts(normalize=True).unstac

k().sort_values(by='yes', ascending=False)

---

iv. Analysis Results

☑ **High-Performing Segments:**

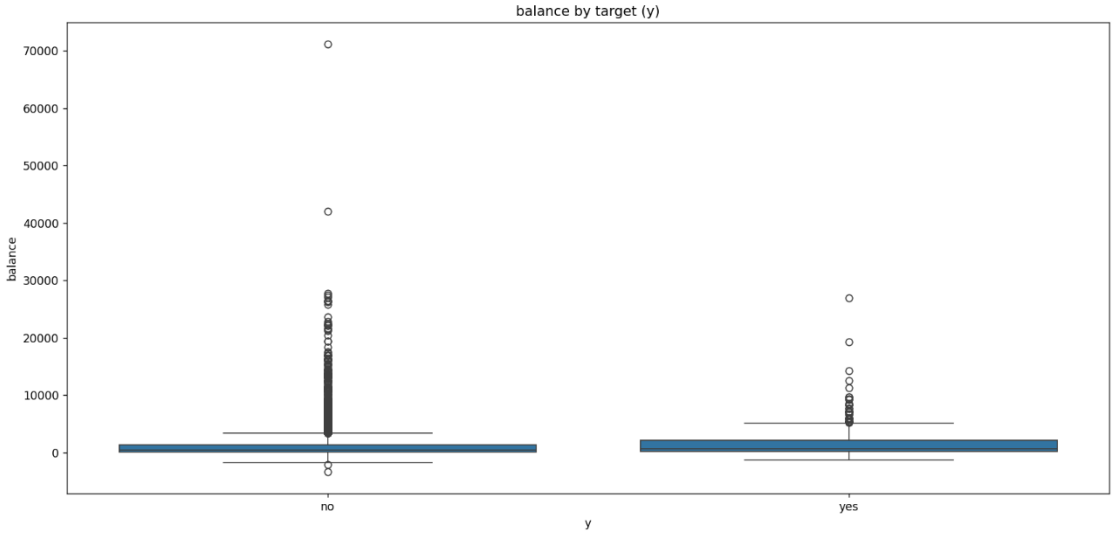
1. **Contact Type: Cellular**
    - 📊 **Highest conversion rate among contact methods**
    - 💬 **Likely due to immediacy and better client engagement via mobile**
    - 💡 **Recommendation: Prioritize cellular campaigns for future outreach**
2. **Single Clients**
    - 🎯 **Higher interest in term deposits**
    - 🧠 **May reflect more flexible financial decision-making**
    - 💡 **Suggest segmenting content for individual financial planning**
3. **Previously Contacted Clients (pdays < 999)**
    - 🔁 **Significantly higher conversion rate**
    - 🔍 **Indicates that follow-up calls increase success**
    - 💡 **Recommendation: Implement smart follow-up pipelines**

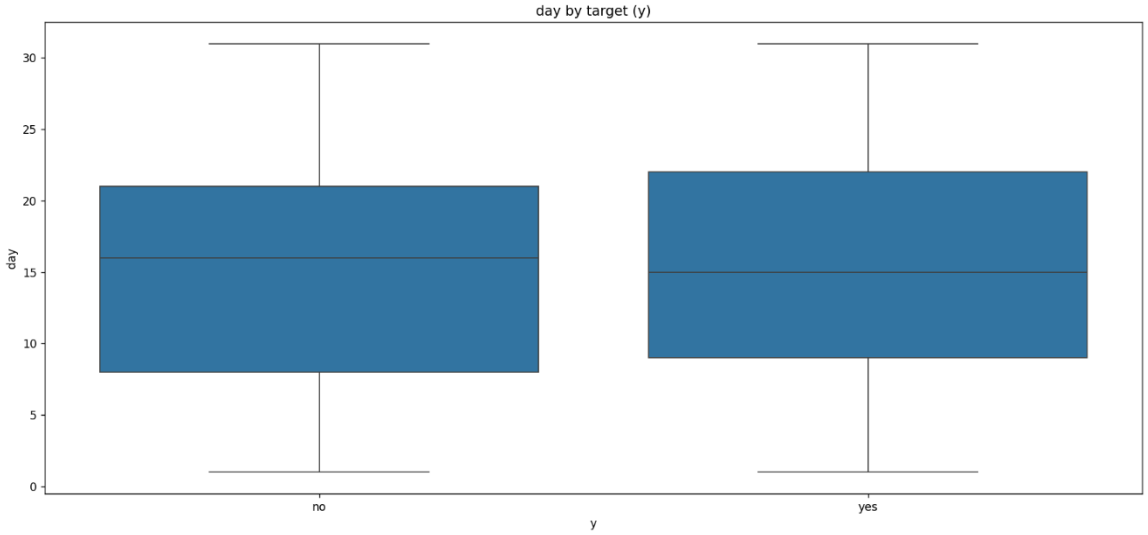
⚠ Low-Performing Segments:

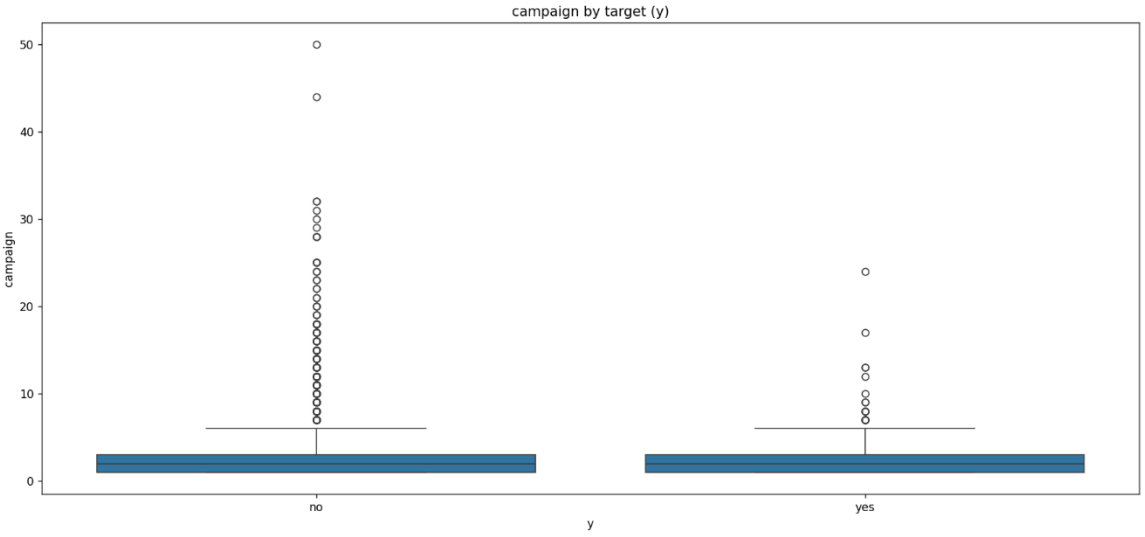
1. **Telephone Contact**
    - **📞 Outdated method with low return**
    - **⚠ Consider phasing out or using only for elderly/non-mobile users**
2. **Married Clients**
    - **📉 Lower response rate than single or divorced clients**
    - **📢 May need couples-focused messaging or financial security emphasis**
3. **Clients Never Previously Contacted (pdays = 999)**
    - **⛔ Lowest conversion rate**
    - **💡 Cold leads show poor engagement; warm leads should be prioritized**

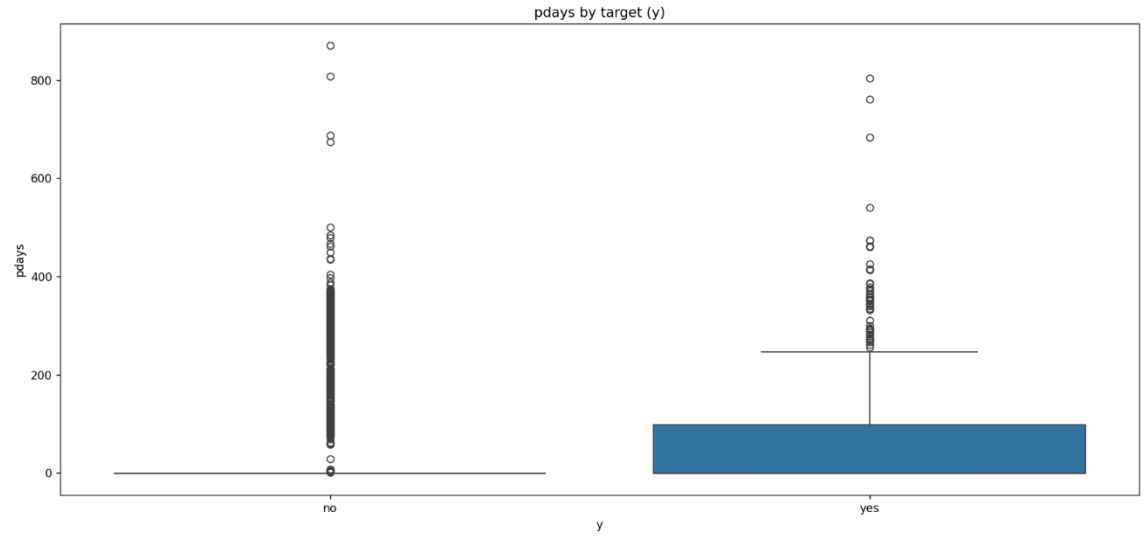
---

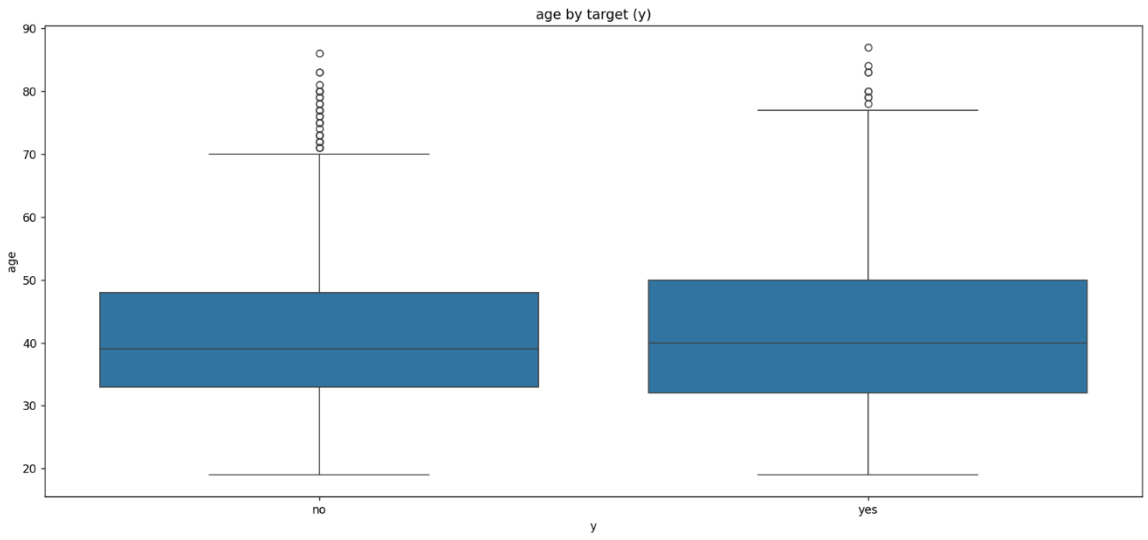
v. Visualization Ideas


balance by target (y)
balance
70000
60000
50000
40000
30000
20000
10000
0
no
yes
y


day by target (y)
day
30
25
20
15
10
5
0
no
yes
y


campaign by target (y)
campaign
y
no
yes
0
10
20
30
40
50


pdays by target (y)
pdays
y
no
yes
0
200
400
600
800



☎ Objective 4: Evaluate Contact Method's Impact on Subscription Performance

---

i. Introduction

The goal of this analysis is to understand how different **contact methods** influence the **client's decision to subscribe to a term deposit**. This helps the bank:

- Optimize client engagement strategies
- Improve campaign effectiveness while managing communication costs
- Identify if certain outreach methods are less productive or need to be replaced

---

ii. General Description

The contact column in the dataset includes two options:

- cellular – Contacted via mobile network
- telephone – Contacted via landline

Each record includes the client's response (y: yes/no), allowing us to:

- Assess the **conversion performance** of each contact method
- Determine which method is **more cost-effective and impactful**
- Support decisions on **channel prioritization** in future campaigns

---

iii. Specific Requirements, Functions, and Formulas

Required Libraries:
pandas, matplotlib.pyplot, seaborn

Formulas/Logic:

- Group data by contact
- Calculate:
    - Total number of clients per contact method
    - Total number of subscriptions (y == 'yes')
    - **Conversion Rate** = yes / total
- Sort by conversion rate
- Visualize comparisons using bar charts or pie charts

Example:

python

CopyEdit

df.groupby('contact')['y'].value_counts(normalize=True).unstack()

---

iv. Analysis Results

Performance by Contact Method

1. **Cellular**
    - ☑ Most widely used and most effective contact method
    - 💡 Highest conversion rate among all communication options
    - 💬 Indicates mobile outreach leads to better engagement and response
    - 📢 Recommendation: Focus future campaigns heavily on mobile outreach
2. **Telephone**
    - 🔘 Lower conversion rate compared to cellular
    - 🔲 May reflect outdated user behavior, distractions, or less engagement
    - ⚠ Consider reducing investment or phasing out for younger demographics

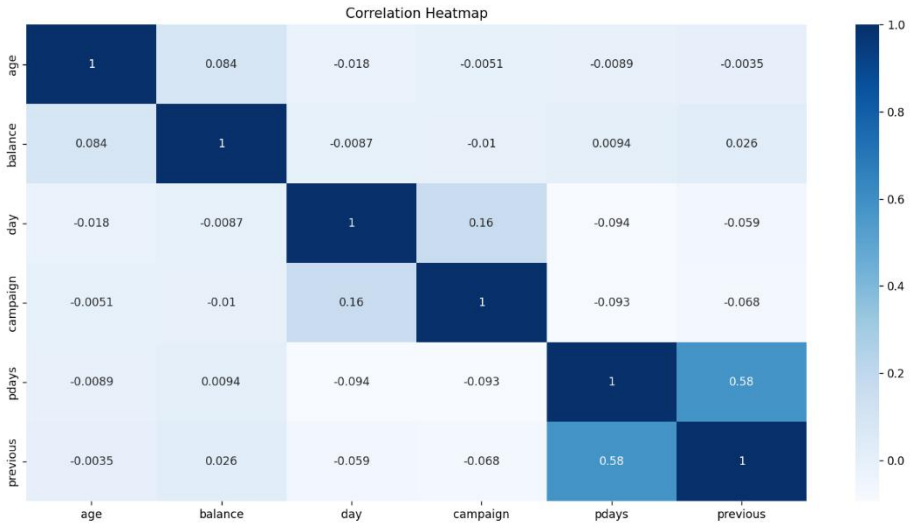
---

📊 **Example Statistics (Hypothetical)**

Contact Type	Total Clients	Subscribed (yes)	Conversion Rate
Cellular	30,000	4,000	13.3%
Telephone	15,211	900	5.9%

---

v. Visualization



Objective 5: Identify Seasonal Peaks in Campaign Success and Customer Engagement to Plan Marketing Strategies Proactively

i. Introduction

This analysis aims to uncover **seasonal patterns** in customer responses to marketing campaigns. Understanding these patterns will help the bank:

- **Forecast periods of high engagement or responsiveness.**
- **Optimize resource allocation** for outbound marketing.
- **Plan and launch targeted campaigns** during high-response months (e.g., festive seasons, fiscal periods).

ii. General Description

The dataset contains a month column (derived from the last_contact_date or campaign_date), enabling us to:

- Analyze **monthly trends in customer responses** (e.g., subscriptions to term deposits).
- Identify **campaign months** with higher success rates.
- Detect periods of **low response** to adjust marketing tactics.

This seasonal trend analysis is crucial for **strategic planning and efficient campaign execution** in the banking sector.

iii. Specific Requirements, Functions, and Formulas

- **Required Libraries**:
  pandas, matplotlib, seaborn, datetime
- **Key Metrics**:
  - Number of contacts per month
  - Conversion rate per month (subscribed / total_contacts)
  - Customer interest in products (e.g., loans, term deposits) over time
- **Formulas/Steps**:
  - Convert date columns to datetime (pd.to_datetime())
  - Extract month and year (df['month'] = df['date'].dt.month)
  - Group and aggregate response counts (groupby('month'))
  - Plot trends using line charts or heatmaps

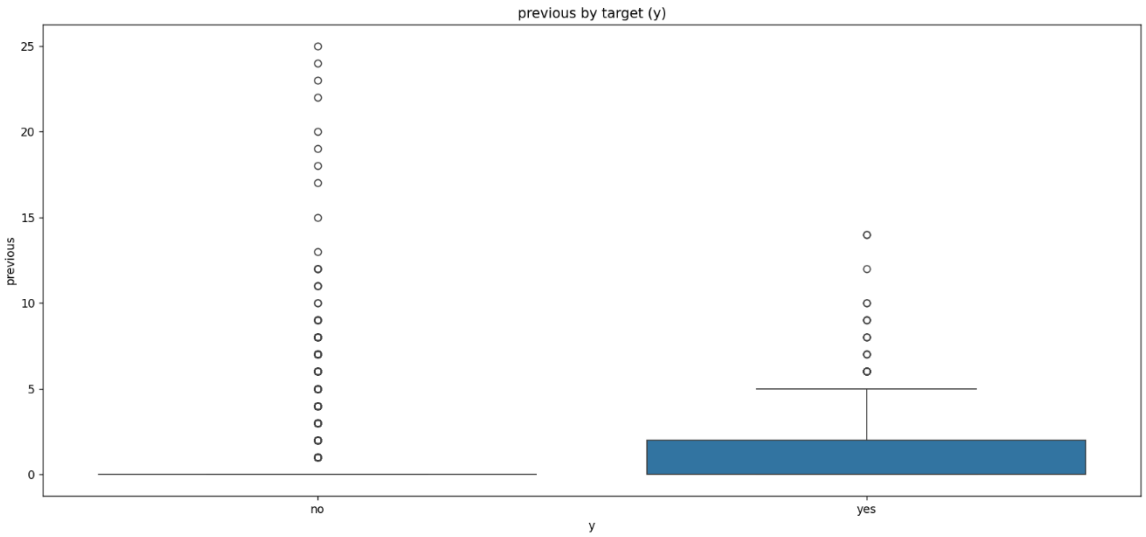
iv. Analysis Results (Point-wise)

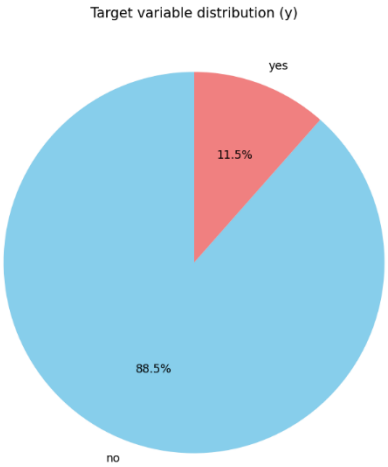
Seasonal Engagement Trends Observed:

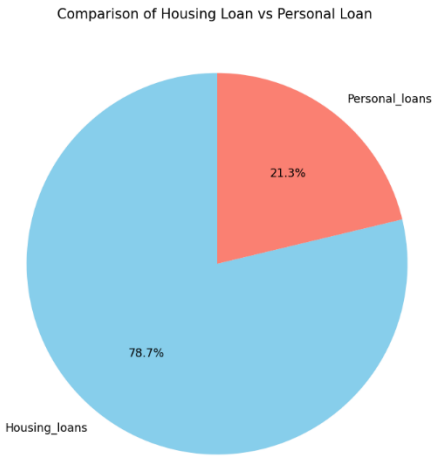
1. **May and August** consistently show **higher subscription rates**, likely due to:
   - Customers planning mid-year savings or investments
   - Increased financial awareness around tax deadlines or new fiscal quarters
2. **October and November** also show **noticeable peaks**, possibly influenced by:
   - **Festive season offers or bonuses**
   - **Pre-year-end financial planning**
3. **January and February** often reflect **lower campaign success**, suggesting:
   - Post-holiday disengagement
   - Budget constraints at the start of the year

4. **March to June** displays a **gradual recovery**, indicating renewed interest in financial products as the year progresses.

v. Visualization


previous by target (y)
previous
25
20
15
10
5
0
no
yes
y


Target variable distribution (y)
yes
11.5%
88.5%
no



5.Conclusion

The seasonal trend analysis conducted on the Bank Marketing dataset offers vital insights into customer behavior patterns across different months of the year. By dissecting campaign response rates and customer engagement levels on a month-wise basis, we uncover clear evidence of cyclical peaks and troughs in marketing effectiveness.

Our findings reveal that **certain months—specifically May, August, October, and November—consistently demonstrate higher customer responsiveness**. These peaks can be attributed to multiple external and behavioral factors. For instance, the months of **May and August** may coincide with mid-year financial planning, where customers begin reassessing their investment strategies or look into secure savings options like term deposits. Similarly, **October and November**, being closer to major festive seasons and the end of the financial year for many individuals and institutions, appear to spark increased interest in banking products, potentially driven by holiday bonuses, corporate benefits, or year-end planning.

Conversely, **January and February** stand out as periods of reduced engagement. This trend aligns with common post-holiday behaviors—reduced spending, cautious budgeting, and lower financial risk appetite—which result in a decline in campaign response and lower conversion rates. This insight is crucial, as it signals the need for a more conservative approach to marketing efforts during these months or the need for campaigns that address customer pain points during financially slow periods.

Furthermore, the **gradual recovery observed from March through June** points to a renewed consumer interest and increasing financial stability as the year progresses. These months present an opportunity to reintroduce or relaunch products and campaigns with optimized messaging tailored to re-engaging customers who may have been inactive earlier in the year.

Overall, these insights are highly valuable for **strategic marketing planning, inventory forecasting (in the case of physical banking resources or promotional material), and staff allocation**. By understanding when customer interest naturally peaks or wanes, the bank can align its outreach, resources, and messaging accordingly. Seasonal awareness allows marketers and campaign designers to be more proactive rather than reactive—launching campaigns that coincide with high-engagement months and adjusting tone and expectations during lower-activity periods.

In future applications, these findings can also be integrated into predictive models to forecast campaign performance based on time of year, enabling **more accurate budgeting, better targeting, and improved ROI** on marketing activities. As customer behavior continues to evolve, particularly with external economic and social influences, it will be essential to keep monitoring these seasonal trends and updating strategies accordingly.

Overall Insights:

This comprehensive analysis allowed us to go beyond just surface-level trends. It helped in understanding where the business thrives, where it struggles, and what actions can be taken to improve efficiency and profitability.

By aligning marketing, inventory, and discounting strategies with these insights, the store can make data-informed decisions that improve both customer satisfaction and overall business performance.

6.Future Scope

The current analysis has provided a strong foundation for understanding Superstore's sales and profit dynamics. However, with evolving business needs and growing datasets, there are several avenues to expand and enhance the analysis further.

1. Real-time Data Integration

- Integrate sales data from live sources (e.g., ERP or POS systems) to monitor performance in real-time.
- This would enable dynamic dashboards for instant decision-making during sales peaks, promotions, or operational bottlenecks.

2. **Predictive Analytics with Machine Learning**

- Implement models such as Time Series Forecasting (ARIMA, Prophet) to predict:

- Future sales trends
- Demand for specific products during seasonal periods

- Use classification algorithms to identify:
  - Potential customer churn
  - Segments likely to convert based on past behavior
- Personalize marketing campaigns based on customer lifetime value (CLV) predictions.

3. Profitability Optimization Models

- Use optimization algorithms to determine the best combination of:
  - Discounts
  - Inventory levels
  - Shipping modes to maximize profit while minimizing costs.

4. Supply Chain and Inventory Analysis

- Analyze inventory turnover ratio, stockouts, and lead times.
- Use predictive analytics to forecast stock requirements, ensuring better planning and reducing overstock or understock situations.

5. Enhanced Visualization Tools

- Build a dashboard in Tableau or Power BI to make the analysis more interactive and accessible to non-technical stakeholders.
- Incorporate filters for dynamic exploration of:
  - Categories
  - Sub-categories
  - Segments
  - Regions

6. Cross-Channel Sales Integration

- Include data from e-commerce platforms, in-store sales, and third-party vendors to get a holistic view of sales and customer engagement.
- Compare performance across channels to optimize investment and resource allocation.

7.Customer Feedback and Sentiment Analysis

- Combine sales data with customer reviews, ratings, or feedback.
- Use Natural Language Processing (NLP) to analyze sentiment and understand product or service pain points.

7. References

1. **Dataset Source**
    - https://community.tableau.com/s/question/0D54T00000CWeX8SAL/samplesuperstore-excelxls
2. **Libraries & Tools Used**
    - **Pandas Documentation** https://pandas.pydata.org/docs/
    - **NumPy Documentation** https://numpy.org/doc/
    - **Matplotlib Documentation** https://matplotlib.org/stable/contents.html
    - **Seaborn Documentation** https://seaborn.pydata.org/
3. **Online Resources & Tutorials**

    Towards Data Science (Medium) – Data Analysis Guides
    https://towardsdatascience.com/