Marketing Data Analysis

Dr. Lema LS

2025-09-10

Instructions

```
# Do not modify this chunk
library(tidyverse)
library(readxl) # package that will help you to load MS Excel data in R.
library(flextable)
library(ggplot2) # will help with advanced plotting if necessary.
```

In this assignment, we perform basic data exploration and visualization on marketing data. This is to get insights on how customers behave to what `FreshDirect` company offers. FreshDirect is the leader in online grocery delivery. Their marketing data captures customer information such as demographics, transaction behavior, and ordering patterns to enable loyalty analysis, segmentation, and predictive modeling.

Your task is to load the data using the package `readxl` allowing you to load excel files. Identify the function from the package with specific chosen parameters from it to get rid of some big issues that may come with data importing. Then you also need to perform some data cleaning for some computations.

Below is the data description.

Column Name	**Description**
LOYALTY_SEGMENT	Shopper classification based on purchase frequency (e.g., Weekly, Bi-Weekly, Monthly)
AGE	Customer's age
INCOME	Household income (may be grouped into ranges)
GENDER	Gender of the primary shopper
ZIP_CODE	Residential ZIP code of customer
DMA	Designated Market Area (media/advertising region)
GEOGRAPHY	Broader geographic grouping
ACQUIRED_DATE	Date the customer first registered or became active
12 Mo. DELIVERY_FEE_PAID	Total delivery fees paid in the last 12 months
12 Mo. DELIVERYPASS_USED	Number of times DeliveryPass subscription was used in 12 months
12 Mo. DISCOUNT_AMOUNT	Total discounts applied in 12 months
12 Mo. Orders	Total number of orders in the past 12 months
12 Mo. ORDERS_W_PROMO	Number of orders that included a promo in 12 months
12 Mo. Sales	Total sales generated by customer in 12 months

Column Name	Description
24 Mo. DELIVERY_FEE_PAID	Total delivery fees paid in the last 24 months
24 Mo. DELIVERYPASS_USED	Number of times DeliveryPass was used in 24 months
24 Mo. DISCOUNT_AMOUNT	Total discounts applied in 24 months
24 Mo. Orders	Total number of orders in the past 24 months
24 Mo. Orders w. Promo	Number of orders with a promo in 24 months
24 Mo. Sales	Total sales generated by customer in 24 months
SUNDAY ORDERS 12 MO.	Number of orders placed on Sundays (12 months)
MONDAY ORDERS 12 MO.	Number of orders placed on Mondays (12 months)
TUESDAY ORDERS 12 MO.	Number of orders placed on Tuesdays (12 months)
WEDNESDAY ORDERS 12 MO.	Number of orders placed on Wednesdays (12 months)
THURSDAY ORDERS 12 MO.	Number of orders placed on Thursdays (12 months)
FRIDAY ORDERS 12 MO.	Number of orders placed on Fridays (12 months)
SATURDAY ORDERS 12 MO.	Number of orders placed on Saturdays (12 months)

Use R to attempt each of the following questions. We recommend you to write your interpretation in your English.

Part 1.

1. After importing data in R. Check which column has the highest count of missing information.

```
# Import data in a variable named df

# Check the dimension of df

# Display the data structure of df
```

Point out any issue with the data (asterix serve as bullet points in markdown. Add * as much as possible.)

-
-

2. How many unique customers are in the dataset?

Interpretation:

3. What is the age distribution of customers?

Interpretation:

4. What is the income distribution (mean, median, spread or standard deviation)?

Interpretation:

5. After decoding the variable GENDER (replace F and M by Female and Male respectively), what are the count of male and female customers in the dataset?

Interpretation:

6. Which part of America has the highest customer counts?

Interpretation:

7. How many customers fall into each LOYALTY_SEGMENT category? You can draw a pie/bar chart and interpret it.

Interpretation:

8. Compute the average number of orders per customer in 12 months.

Interpretation:

9. Compute the average sales per customer in 12 months?

Interpretation:

10. How many customers used DeliveryPass at least once?

Interpretation:

Part 2.

11. Do higher-income customers place more orders?

Interpretation:

12. Based on gender, is there any difference between average sales?

Interpretation:

13. How do `LOYALTY_SEGMENTS` differ in terms of twelve month sales?

Interpretation:

14. Do younger customers, for customer aged less than 30, use promos more than older ones?

Interpretation:

15. How does discount amount vary across income brackets?

Interpretation:

16. Is DeliveryPass usage associated with higher total sales?

Interpretation:

17. Do frequent shoppers (Weekly, Bi-Weekly) pay less in delivery fees?

Interpretation:

18. What percentage of total sales comes from each LOYALTY_SEGMENT?

Interpretation:

19. Are customers with earlier acquisition dates (older customers) more loyal in terms of orders?

Interpretation:

20. Which ZIP codes have the highest per-customer spending?

Interpretation:

Part 3.

What is the day of the week with the highest average orders?

Interpretation:

21. Is weekend ordering (Sat+Sun) higher than weekday ordering?

Interpretation:

22. Do different LOYALTY_SEGMENTS prefer different days of the week?

Interpretation:

23. Do promo orders cluster on specific days (e.g., Fridays)?

Interpretation:

24. Are sales more evenly distributed across days or skewed to a few?

Interpretation:

24. (Bonus) Draw at least 2 graphics and carefully interpret results.

Interpretation:

24. (Bonus) Perform at least one statistical test and interpret results.

Interpretation:

Submission

Submit the `.Rmd` and the knitted PDF files using the correct naming.