Online Retail Customer Retention Analysis

Alex

2025-04-09

Introduction

This report analyses customer purchasing patterns to improve retention for an online retail company

Data Loading and Cleaning

```
# Load the tidyverse package
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readr)

# Load the dataset
retail_data <- read_csv("OnlineRetail.csv")
```

```
## Rows: 541909 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): InvoiceNo, StockCode, Description, InvoiceDate, Country
## dbl (3): Quantity, UnitPrice, CustomerID
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Remove rows with missing CustomerID and negative quantities
clean_data <- retail_data %>%
  filter(!is.na(CustomerID), Quantity > 0)

# Convert InvoiceDate to a date format
clean_data <- clean_data %>%
  mutate(InvoiceDate = as.Date(InvoiceDate, format = "%m/%d/%Y %H:%M"))

# Look at the first few rows
head(clean_data)
```

```
## # A tibble: 6 x 8
##   InvoiceNo StockCode Description      Quantity InvoiceDate UnitPrice CustomerID
##   <chr>     <chr>     <chr>               <dbl> <date>          <dbl>      <dbl>
## 1 536365    85123A    WHITE HANGING H~        6 2010-12-01       2.55      17850
## 2 536365    71053     WHITE METAL LAN~        6 2010-12-01       3.39      17850
## 3 536365    84406B    CREAM CUPID HEA~        8 2010-12-01       2.75      17850
## 4 536365    84029G    KNITTED UNION F~        6 2010-12-01       3.39      17850
## 5 536365    84029E    RED WOOLLY HOTT~        6 2010-12-01       3.39      17850
## 6 536365    22752     SET 7 BABUSHKA ~        2 2010-12-01       7.65      17850
## # i 1 more variable: Country <chr>
```

Analysis

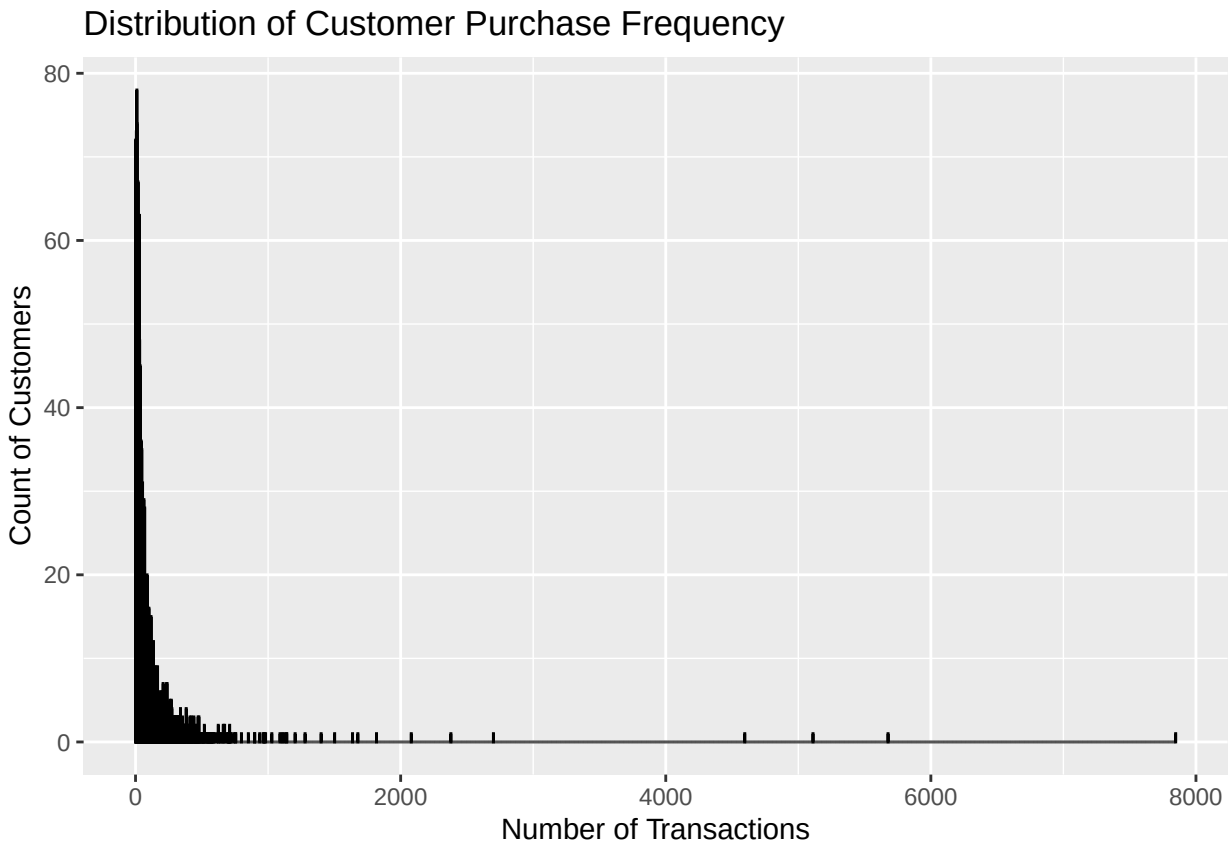
Let's explore customer purchasing patterns.

```r
# Count transactions per customer
customer_frequency <- clean_data %>%
  group_by(CustomerID) %>%
  summarise(Transactions = n()) %>%
  arrange(desc(Transactions))

# Show top 5 customers
head(customer_frequency, 5)
```

```
## # A tibble: 5 x 2
##   CustomerID Transactions
##        <dbl>        <int>
## 1      17841         7847
## 2      14911         5677
## 3      14096         5111
## 4      12748         4596
## 5      14606         2700
```

```r
# Plot a histogram of transaction counts
ggplot(customer_frequency, aes(x = Transactions)) +
  geom_histogram(binwidth = 1, fill = "blue", color = "black") +
  labs(title = "Distribution of Customer Purchase Frequency",
       x = "Number of Transactions", y = "Count of Customers")
```




```
# Calculate total spend per transaction and average per customer
clean_data <- clean_data %>%
  mutate(TotalSpend = Quantity * UnitPrice)

avg_spend <- clean_data %>%
  group_by(CustomerID) %>%
  summarise(AvgSpend = mean(TotalSpend)) %>%
  arrange(desc(AvgSpend))

head(avg_spend, 5)
```

```
## # A tibble: 5 x 2
##   CustomerID AvgSpend
##        <dbl>    <dbl>
## 1      12346   77184.
## 2      16446   56158.
## 3      15098   13306.
## 4      15749    4453.
## 5      15195    3861
```

Findings

Most customers buy only once, but a small group of repeat buyers drives sales. The top spenders average higher transaction values.

Recommendations

- Offer discounts to one-time buyers to encourage repeat purchases.
- Reward top spenders with loyalty perks.