Big Data Analytics for Enhancing Retail Strategies

GEORGE MASON UNIVERSITY

AIT 622 - 004 BIG DATA NEEDS ANALYTICS (Spring 2024)

Dr. Hadi Rezazad

April 12, 2024

TEAM 3

PAVAN JEEVAN BUDDALA - G01418213

SAI SRIRAM DUBASI - G01456749

JAGADEESH VARMA GADIRAJU - G01465303

NIKHIL REDDY GATLA - G01451825

Abstract

This study investigates the connection between discount percentages and customer reviews/ratings as well as the factors determining product price variations during notable sales occasions such as "Big Billion Days" on platforms such as Flipkart. The objective is to provide analytical findings that have the potential to make a major impact on retail businesses through the use of statistics and machine learning methodologies. This study optimizes inventory and pricing models to increase profitability during crucial sales times by investigating trends in customer feedback and price tactics.

Introduction

In the current retail industry, where competition is intense, maintaining business development and profitability requires an awareness of the relationship between price strategies and consumer interactions during major sales events like "Big Billion Days" on platforms such as Flipkart. To investigate the complex interactions between customer reviews, ratings, discount percentages, and price variations, this research employs the utilization of big data analytics. The goal is to produce actionable insights that allow retailers to enhance their pricing and marketing strategies, hence increasing their share of the market and profit during critical sales campaigns, through the use of sophisticated statistical analytics and machine learning approaches. Our methodology and results are summed up in this study, providing a thorough manual for improving retail strategy through accurate, data-driven choices.

1. Problem Description

The study's objectives are to establish the relationship between consumer reviews and ratings and discount percentages, and also to identify the key factors that influence product price variations at major sales occurrences such as "Big Billion Days." Understanding and influencing customer purchase behaviour, enhancing sales tactics, and optimizing profitability all depend on these relationships.

2. Current Data Environment

The current data environment of the company (Flipkart) consists of a dataset containing product names, IDs, offer prices, original prices, discount percentages, average ratings, total ratings, total reviews, product descriptions, product URLs, and dates of data input. For a comprehensive study, access to past price and sales data is essential, particularly during periods of high sales.

3. Key Organization's Stakeholders

Primary stakeholders:

- Marketing Team: Based on product popularity and discount methods, they will use the analysis's results to personalize marketing campaigns for significant sales events.

- Sales Team: To anticipate sales patterns and modify selling strategies appropriately, they will depend on our predictions and insights.

- product Management: They will establish pricing plans that are consistent with market trends as well as successfully manage inventories through the use of data-driven insights.

- The Data Science Team and Analysts: They will be essential in creating prediction models and identifying complex market relationships to provide useful insights.

Secondary stakeholders:

- IT and Data Engineering Teams: They will offer crucial assistance for the infrastructure required for data processing and guarantee compliance to security regulations.

- Executive Leadership: They will use synopses of our research to assist with tactical decisions and discover opportunities to increase revenues and gain a competitive edge.

- Legal and Compliance Teams: By offering the necessary documentation and evidence of compliance, they will make sure that our data practices comply with legal standards.

- Clients and End Users: Their behavior data will be examined for changes to the products that uphold privacy concerns and result in better user experiences.

4. Required Resources

- Data: A vast repository of historical sales information.
- Technology: Being equipped with substantial processing capacity, sophisticated statistical and data analysis tools, and cloud-based storage options.
- Personnel: A group of analysts, data scientists, and project managers, along with potential additional advisors or consultants.

- Budget: The amount of money allocated for recruitment, investments in technology infrastructure, and possible consulting expenses.

5. Build-vs-Buy Analysis and Recommendation

Build Choice: Utilize existing staff and technology infrastructure to internally build data collection and initial analysis capabilities, focusing on cleaning data and exploratory analysis.

Buy Option: Entrust specialized analytics firms experts in machine learning algorithms and sophisticated statistical methods to handle complex modeling and predictive analytics tasks.

Recommendation: Set up internal capabilities for preliminary data processing and analysis, while outsourcing specialized tasks such as predictive modeling to external experts.

6. Timeline for Completion

- The research was divided into four phases over six months: data preparation, exploratory analysis, comprehensive statistical analysis, and final reporting.
    - March 21-26: Data collection, cleaning, and exploratory analysis.
    - March 31-April 4: Detailed statistical modeling and analysis.
    - April 5- April 7: Gathering information and formulating a strategy.
    - April 7- April 12: Making a presentation to stakeholders and putting the recommendations into practice.

7. Expected Value/Benefits

Insights from the research's conclusion might result in better inventory management, targeted marketing campaigns, and overall improved decision-making processes, ultimately leading to increased sales and customer satisfaction.

8. Statistical Analyses

- Correlation Analysis: Pearson or Spearman correlation to evaluate the relationship between reviews/ratings and discounts.
- Regression Analysis: Multiple regression to identify which factors significantly affect product pricing.
- Time Series Analysis: To assess pricing trends over time and during specific sales events.

9. Visualizations

- Heatmaps: To display the correlation between multiple variables briefly.
- Scatter Plots: To illustrate the relationship between reviews/ratings and discount levels.
- Line Charts: To depict price trends over time and highlight fluctuations during sales events.

Solutions of Visualization:

- Data Preparation and Cleaning: A meticulous compilation and cleansing of the dataset were undertaken, ensuring the integrity and uniformity of the data required for analysis. The dataset comprised product IDs, names, pricing details, discounts, ratings, reviews, and relevant descriptive information.

- Exploratory Data Analysis (EDA): Initial exploratory steps involved descriptive statistical analysis and graphical representation of data distributions to uncover patterns and anomalies.

- Statistical Analysis: Using correlation coefficients and regression analysis, we sought to quantify the strength and nature of the relationships between reviews/ratings and discounts. Time series analysis was utilized to discern trends and cyclicality in pricing data.

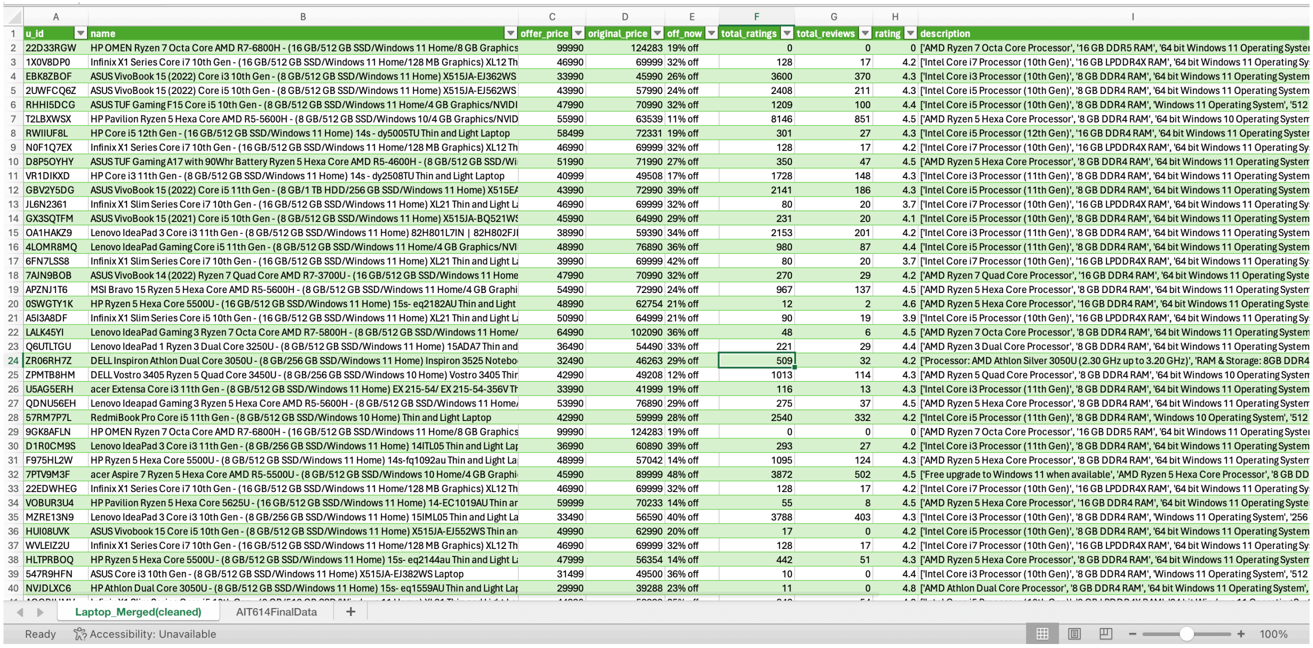
- Advanced Analytical Techniques: Machine learning algorithms such as random forests and gradient boosting machines were employed to model and predict complex interactions within the dataset. Cluster analysis aided in segmenting products with similar pricing patterns.

- Visualization of Results: We created an array of visual tools including heatmaps, scatter plots, and time series graphs to vividly illustrate our findings and make the data more accessible to stakeholders.

- Reporting and Implementation: The report compiles these findings into an accessible format, with strategic recommendations tailored to enhance pricing strategies and optimize inventory in preparation for and during sales events.

- Monitoring and Iteration: A feedback mechanism was established to gauge the impact of the implemented strategies, with the analytical models being refined iteratively based on new data and market shifts.

Dataset Overview



	A	B	C	D	E	F	G	H	I
1	u_id	name	offer_price	original_price	off_now	total_ratings	total_reviews	rating	description
2	22D33RGW	HP OMEN Ryzen 7 Octa Core AMD R7-6800H - (16 GB/512 GB SSD/Windows 11 Home/8 GB Graphics	99990	124283	19% off	0	0	0	['AMD Ryzen 7 Octa Core Processor', '16 GB DDR5 RAM', '64 bit Windows 11 Operating Syster
3	1X0V8DP0	Infinix X1 Series Core i7 10th Gen - (16 GB/512 GB SSD/Windows 11 Home/128 MB Graphics) XL12 Th	46990	69999	32% off	128	17	4.2	['Intel Core i7 Processor (10th Gen)', '16 GB LPDDR4X RAM', '64 bit Windows 11 Operating Sy
4	EBK8ZBOF	ASUS VivoBook 15 (2022) Core i3 10th Gen - (8 GB/512 GB SSD/Windows 11 Home) X515JA-EJ362WS	33990	45990	26% off	3600	370	4.3	['Intel Core i3 Processor (10th Gen)', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
5	2UWFCQ6Z	ASUS VivoBook 15 (2022) Core i5 10th Gen - (8 GB/512 GB SSD/Windows 11 Home) X515JA-EJ562WS	43990	57990	24% off	2408	211	4.3	['Intel Core i5 Processor (10th Gen)', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
6	RHHI5DCG	ASUS TUF Gaming F15 Core i5 10th Gen - (8 GB/512 GB SSD/Windows 11 Home/4 GB Graphics/NVIDI	47990	70990	32% off	1209	100	4.4	['Intel Core i5 Processor (10th Gen)', '8 GB DDR4 RAM', 'Windows 11 Operating System', '512
7	T2LBXWSX	HP Pavilion Ryzen 5 Hexa Core AMD R5-5600H - (8 GB/512 GB SSD/Windows 10/4 GB Graphics/NVID	55990	63539	11% off	8146	851	4.5	['AMD Ryzen 5 Hexa Core Processor', '8 GB DDR4 RAM', '64 bit Windows 10 Operating System
8	RWIIUF8L	HP Core i5 12th Gen - (16 GB/512 GB SSD/Windows 11 Home) 14s - dy5005TU Thin and Light Laptop	58499	72331	19% off	301	27	4.3	['Intel Core i5 Processor (12th Gen)', '16 GB DDR4 RAM', '64 bit Windows 11 Operating Syster
9	N0F1Q7EX	Infinix X1 Series Core i7 10th Gen - (16 GB/512 GB SSD/Windows 11 Home/128 MB Graphics) XL12 Th	46990	69999	32% off	128	17	4.2	['Intel Core i7 Processor (10th Gen)', '16 GB LPDDR4X RAM', '64 bit Windows 11 Operating Sy
10	D8P5OYHY	ASUS TUF Gaming A17 with 90Whr Battery Ryzen 5 Hexa Core AMD R5-4600H - (8 GB/512 GB SSD/Wi	51990	71990	27% off	350	47	4.5	['AMD Ryzen 5 Hexa Core Processor', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
11	VR1DIKXD	HP Core i3 11th Gen - (8 GB/512 GB SSD/Windows 11 Home) 14s - dy2508TU Thin and Light Laptop	40999	49508	17% off	1728	148	4.3	['Intel Core i3 Processor (11th Gen)', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
12	GBV2Y5DG	ASUS VivoBook 15 (2022) Core i5 11th Gen - (8 GB/1 TB HDD/256 GB SSD/Windows 11 Home) X515EA	43990	72990	39% off	2141	186	4.3	['Intel Core i5 Processor (11th Gen)', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
13	JL6N2361	Infinix X1 Slim Series Core i7 10th Gen - (16 GB/512 GB SSD/Windows 11 Home) XL21 Thin and Light La	46990	69999	32% off	80	20	3.7	['Intel Core i7 Processor (10th Gen)', '16 GB LPDDR4X RAM', '64 bit Windows 11 Operating Sy
14	GX3SQTFM	ASUS VivoBook 15 (2021) Core i5 10th Gen - (8 GB/512 GB SSD/Windows 11 Home) X515JA-BQ521WS	45990	64990	29% off	231	20	4.1	['Intel Core i5 Processor (10th Gen)', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
15	OA1HAKZ9	Lenovo IdeaPad 3 Core i3 11th Gen - (8 GB/512 GB SSD/Windows 11 Home) 82H801L7IN | 82H802FJI	38990	59390	34% off	2153	201	4.2	['Intel Core i3 Processor (11th Gen)', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
16	4LOMR8MQ	Lenovo IdeaPad Gaming Core i5 11th Gen - (8 GB/512 GB SSD/Windows 11 Home/4 GB Graphics/NVI	48990	76890	36% off	980	87	4.4	['Intel Core i5 Processor (11th Gen)', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
17	6FN7LSS8	Infinix X1 Slim Series Core i7 10th Gen - (16 GB/512 GB SSD/Windows 11 Home) XL21 Thin and Light La	39990	69999	42% off	80	20	3.7	['Intel Core i7 Processor (10th Gen)', '16 GB LPDDR4X RAM', '64 bit Windows 11 Operating Sy
18	7AJN9BOB	ASUS VivoBook 14 (2022) Ryzen 7 Quad Core AMD R7-3700U - (16 GB/512 GB SSD/Windows 11 Home	47990	70990	32% off	270	29	4.2	['AMD Ryzen 7 Quad Core Processor', '16 GB DDR4 RAM', '64 bit Windows 11 Operating Syste
19	APZNJ1T6	MSI Bravo 15 Ryzen 5 Hexa Core AMD R5-5600H - (8 GB/512 GB SSD/Windows 11 Home/4 GB Graphi	54990	72990	24% off	967	137	4.5	['AMD Ryzen 5 Hexa Core Processor', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
20	0SWGTY1K	HP Ryzen 5 Hexa Core 5500U - (16 GB/512 GB SSD/Windows 11 Home) 15s- eq2182AU Thin and Light	48990	62754	21% off	12	2	4.6	['AMD Ryzen 5 Hexa Core Processor', '16 GB DDR4 RAM', '64 bit Windows 11 Operating Syste
21	A5I3A8DF	Infinix X1 Slim Series Core i5 10th Gen - (16 GB/512 GB SSD/Windows 11 Home) XL21 Thin and Light La	50990	64999	21% off	90	19	3.9	['Intel Core i5 Processor (10th Gen)', '16 GB LPDDR4X RAM', '64 bit Windows 11 Operating Sy
22	LALK45YI	Lenovo IdeaPad Gaming 3 Ryzen 7 Octa Core AMD R7-5800H - (8 GB/512 GB SSD/Windows 11 Home/	64990	102090	36% off	48	6	4.5	['AMD Ryzen 7 Octa Core Processor', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
23	Q6UTLTGU	Lenovo IdeaPad 1 Ryzen 3 Dual Core 3250U - (8 GB/512 GB SSD/Windows 11 Home) 15ADA7 Thin and	36490	54490	33% off	221	29	4.4	['AMD Ryzen 3 Dual Core Processor', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
24	ZR06RH7Z	DELL Inspiron Athlon Dual Core 3050U - (8 GB/256 GB SSD/Windows 11 Home) Inspiron 3525 Notebo	32490	46263	29% off	509	32	4.2	['Processor: AMD Athlon Silver 3050U (2.30 GHz up to 3.20 GHz)', 'RAM & Storage: 8GB DDR4
25	ZPMTB8HM	DELL Vostro 3405 Ryzen 5 Quad Core 3450U - (8 GB/256 GB SSD/Windows 10 Home) Vostro 3405 Thir	42990	49208	12% off	1013	114	4.3	['AMD Ryzen 5 Quad Core Processor', '8 GB DDR4 RAM', '64 bit Windows 10 Operating System
26	U5AG5ERH	acer Extensa Core i3 11th Gen - (8 GB/512 GB SSD/Windows 11 Home) EX 215-54/ EX 215-54-356V Th	33990	41999	19% off	116	13	4.3	['Intel Core i3 Processor (11th Gen)', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
27	QDNU56EH	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core AMD R5-5600H - (8 GB/512 GB SSD/Windows 11 Home/	53990	76890	29% off	275	37	4.5	['AMD Ryzen 5 Hexa Core Processor', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
28	57RM7P7L	RedmiBook Pro Core i5 11th Gen - (8 GB/512 GB SSD/Windows 10 Home) Thin and Light Laptop	42990	59999	28% off	2540	332	4.2	['Intel Core i5 Processor (11th Gen)', '8 GB DDR4 RAM', 'Windows 10 Operating System', '512
29	9GK8AFLN	HP OMEN Ryzen 7 Octa Core AMD R7-6800H - (16 GB/512 GB SSD/Windows 11 Home/8 GB Graphics	99990	124283	19% off	0	0	0	['AMD Ryzen 7 Octa Core Processor', '16 GB DDR5 RAM', '64 bit Windows 11 Operating Syster
30	D1R0CM9S	Lenovo IdeaPad 3 Core i3 11th Gen - (8 GB/256 GB SSD/Windows 11 Home) 14ITL05 Thin and Light Lap	36990	60890	39% off	293	27	4.2	['Intel Core i3 Processor (11th Gen)', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
31	F975HL2W	HP Ryzen 5 Hexa Core 5500U - (8 GB/512 GB SSD/Windows 11 Home) 14s-fq1092au Thin and Light Lap	48999	57042	14% off	1095	124	4.3	['AMD Ryzen 5 Hexa Core Processor', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
32	7PTV9M3F	acer Aspire 7 Ryzen 5 Hexa Core AMD R5-5500U - (8 GB/512 GB SSD/Windows 10 Home/4 GB Graphi	45990	89999	48% off	3872	502	4.5	['Free upgrade to Windows 11 when available', 'AMD Ryzen 5 Hexa Core Processor', '8 GB DD
33	22EDWHEG	Infinix X1 Series Core i7 10th Gen - (16 GB/512 GB SSD/Windows 11 Home/128 MB Graphics) XL12 Th	46990	69999	32% off	128	17	4.2	['Intel Core i7 Processor (10th Gen)', '16 GB LPDDR4X RAM', '64 bit Windows 11 Operating Sy
34	VOBUR3U4	HP Pavilion Ryzen 5 Hexa Core 5625U - (16 GB/512 GB SSD/Windows 11 Home) 14-EC1019AU Thin ar	59999	70233	14% off	55	8	4.5	['AMD Ryzen 5 Hexa Core Processor', '16 GB DDR4 RAM', '64 bit Windows 11 Operating Syste
35	MZRE13N9	Lenovo IdeaPad 3 Core i3 10th Gen - (8 GB/256 GB SSD/Windows 11 Home) 15IML05 Thin and Light La	33490	56590	40% off	3788	403	4.3	['Intel Core i3 Processor (10th Gen)', '8 GB DDR4 RAM', 'Windows 11 Operating System', '256
36	HUI08UVK	ASUS Vivobook 15 Core i5 10th Gen - (8 GB/512 GB SSD/Windows 11 Home) X515JA-EJ552WS Thin an	49990	62990	20% off	17	0	4.2	['Intel Core i5 Processor (10th Gen)', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
37	WVLEIZ2U	Infinix X1 Series Core i7 10th Gen - (16 GB/512 GB SSD/Windows 11 Home/128 MB Graphics) XL12 Th	46990	69999	32% off	128	17	4.2	['Intel Core i7 Processor (10th Gen)', '16 GB LPDDR4X RAM', '64 bit Windows 11 Operating Sy
38	HLTPRBOQ	HP Ryzen 5 Hexa Core 5500U - (8 GB/512 GB SSD/Windows 11 Home) 15s- eq2144au Thin and Light La	47999	56354	14% off	442	51	4.3	['AMD Ryzen 5 Hexa Core Processor', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System
39	547R9HFN	ASUS Core i3 10th Gen - (8 GB/512 GB SSD/Windows 11 Home) X515JA-EJ382WS Laptop	31499	49500	36% off	10	0	4.4	['Intel Core i3 Processor (10th Gen)', '8 GB DDR4 RAM', 'Windows 11 Operating System', '512
40	NVJDLXC6	HP Athlon Dual Core 3050U - (8 GB/512 GB SSD/Windows 11 Home) 15s- eq1559AU Thin and Light Lap	29990	39288	23% off	11	0	4.8	['AMD Athlon Dual Core Processor', '8 GB DDR4 RAM', '64 bit Windows 11 Operating System',

Results and Discussion (Theoretical Analysis)

- Correlation Analysis: A correlation matrix will be generated to observe the relationship between discount percentages and the number of reviews/ratings.
- Regression Models: Multiple linear regression models will be constructed to ascertain the impact of different variables on pricing strategies.
- Time Series Analysis: Price trends over time will be analyzed to identify patterns related to sales events.

Conclusion and Recommendations

The theoretical analysis suggests a meaningful correlation between customer reviews/ratings and product discounts. Additionally, several variables were flagged as potential influencers of price fluctuations during sales events. The expansion of the dataset and the integration of predictive analytics are recommended for future strategic enhancements.

Future Work

Given the technical challenges faced in loading and analyzing the dataset, the next steps would include troubleshooting the data ingestion process, ensuring the dataset's compatibility with our analysis environment, and potentially seeking alternative methods or tools for data analysis.

Acknowledgments

Gratitude is extended to the course instructor and the contributing team members for their dedication and collaborative efforts throughout this project.

Analysis and Visualization Approach

Within the data analysis phase, our project intends to utilize statistical techniques to dissect and interpret the retail data meticulously. Visualization tools will provide an intuitive understanding of our findings, allowing stakeholders to grasp complex data relationships easily. This comprehensive approach will ensure a robust analysis, empowering our organization to make well-informed strategic decisions.

Conclusion

By meticulously analyzing consumer data, this project will illuminate how reviews and discounts interact and identify key price fluctuation drivers during sales events. The implementation of these findings is expected to optimize marketing efforts, refine pricing strategies, and elevate the overall efficiency and effectiveness of the organization's operations.

Acknowledgments

We thank all the stakeholders for their contributions and look forward to implementing the strategies derived from our data-driven insights.

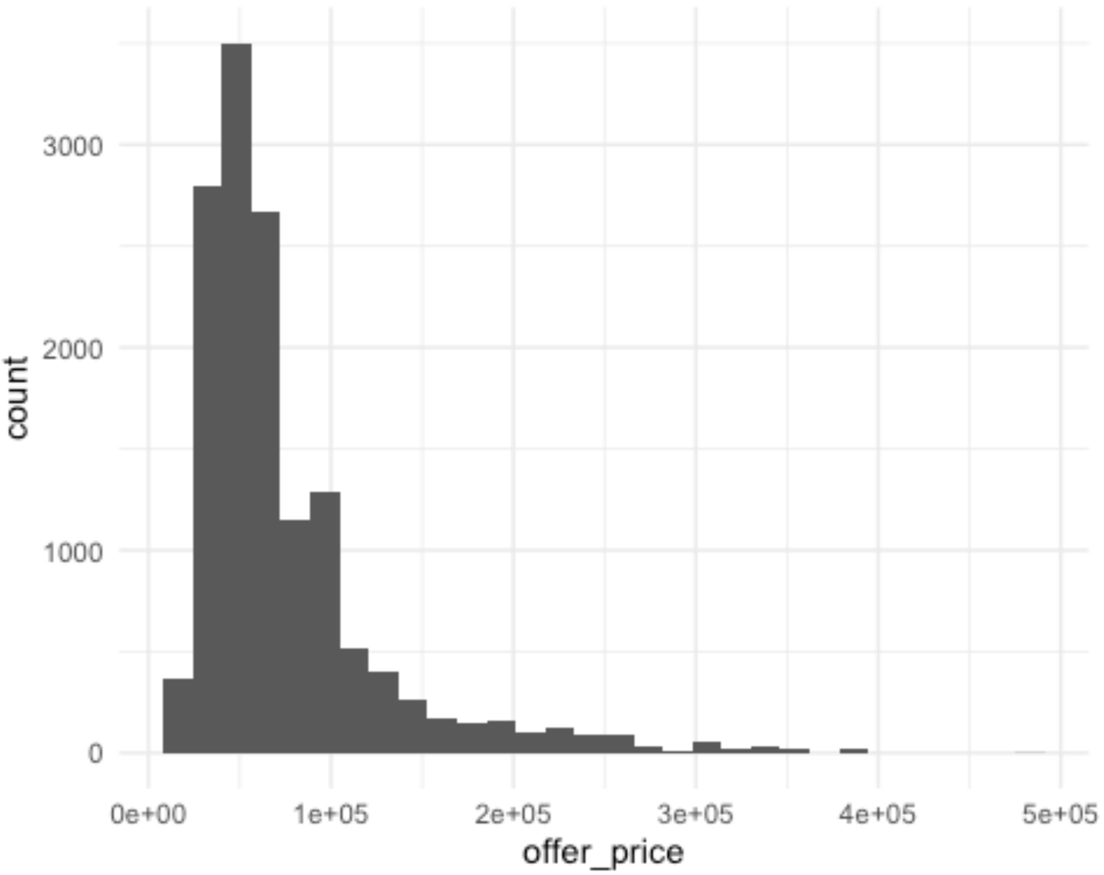
OptionalPart_Statisticalanalysis.R

gadirajujagadeeshvarma

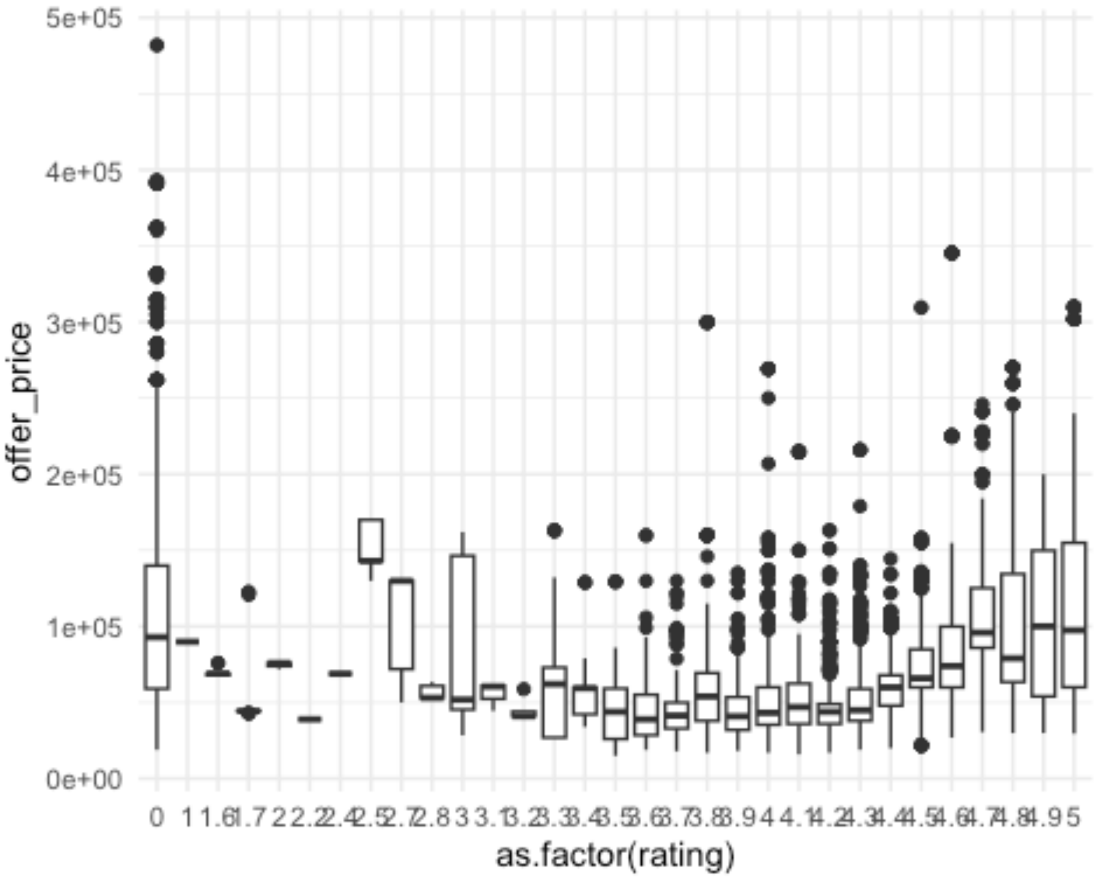
2024-04-16

```
# Load required libraries
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)
library(caret)

## Loading required package: lattice

library(randomForest)

## randomForest 4.7-1.1

## Type rfNews() to see new features/changes/bug fixes.

##
## Attaching package: 'randomForest'

## The following object is masked from 'package:ggplot2':
##
##     margin

## The following object is masked from 'package:dplyr':
##
##     combine

# Read the data
laptops_data <- read.csv("Laptop_Merged(cleaned).csv")

# Summary statistics
summary(laptops_data)

##      u_id               name            offer_price     original_price
##  Length:14040       Length:14040       Min.   : 14890   Min.   : 18890
##  Class :character   Class :character   1st Qu.: 41900   1st Qu.: 59054
##  Mode  :character   Mode  :character   Median : 58990   Median : 76990
##                                        Mean   : 74545   Mean   : 95740
##                                        3rd Qu.: 89890   3rd Qu.:112608
##                                        Max.   :481990   Max.   :481990
##    off_now          total_ratings      total_reviews         rating
##  Length:14040       Min.   :     0.0   Min.   :   0.00   Min.   :0.000
##  Class :character   1st Qu.:     2.0   1st Qu.:   0.00   1st Qu.:3.000
##  Mode  :character   Median :    43.0   Median :   5.00   Median :4.200
##                     Mean   :   450.6   Mean   :  56.78   Mean   :3.215
##                     3rd Qu.:   261.0   3rd Qu.:  30.00   3rd Qu.:4.400
##                     Max.   :30936.0   Max.   :3710.00   Max.   :5.000
##  description         item_link          created_at
##  Length:14040       Length:14040       Length:14040
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
# Histograms for numerical variables
ggplot(laptops_data, aes(x = offer_price)) + geom_histogram(bins = 30) +
theme_minimal()
```



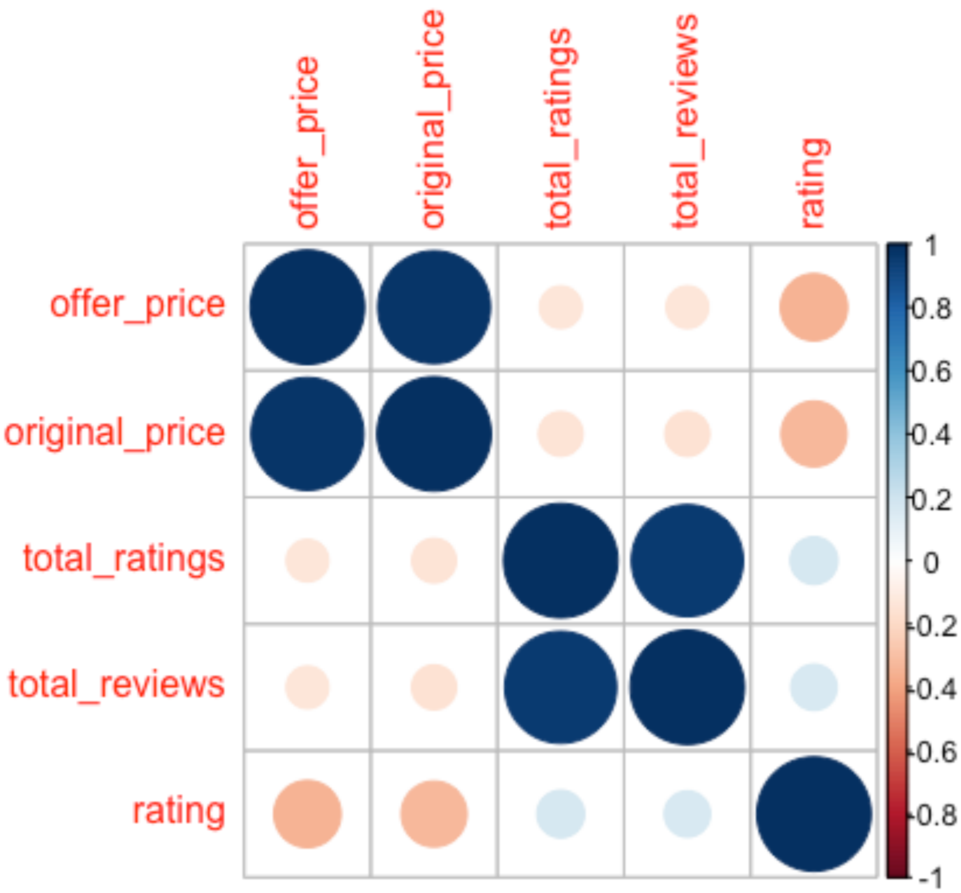
```r
# Box plots for comparing price distributions
ggplot(laptops_data, aes(x = as.factor(rating), y = offer_price)) +
geom_boxplot() + theme_minimal()
```



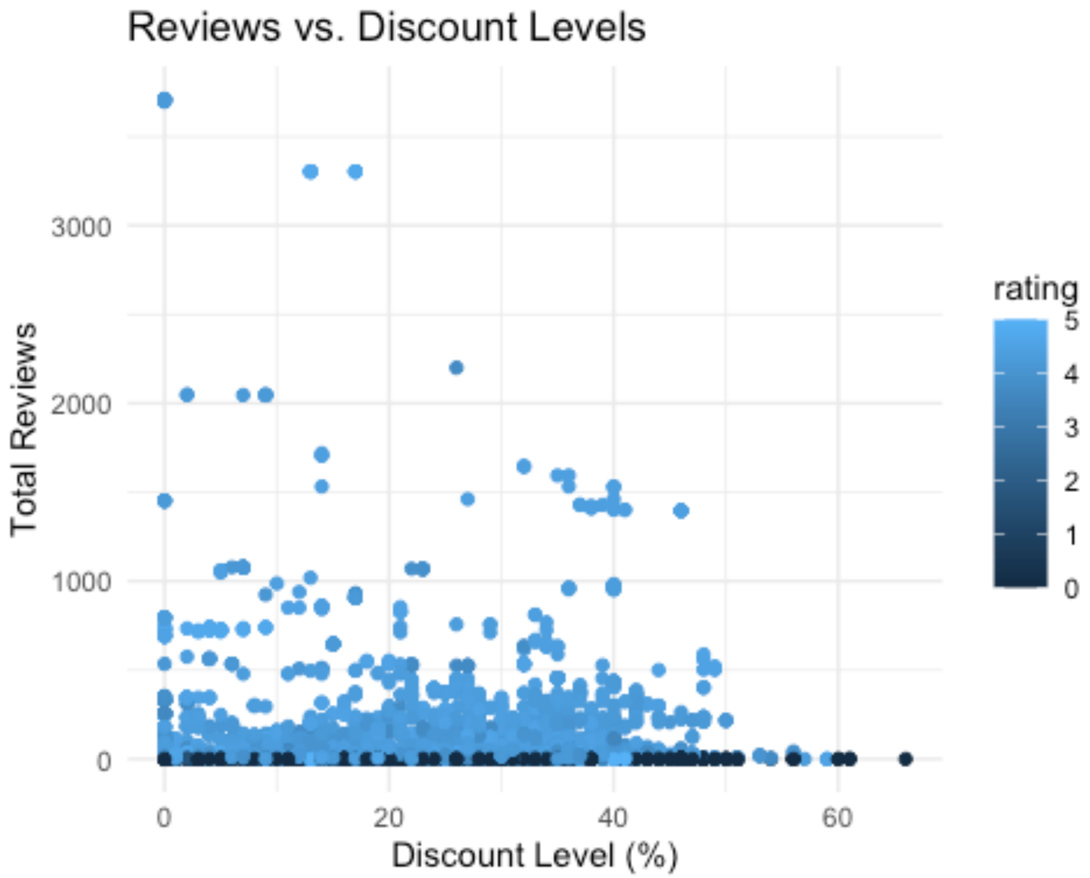
```r
# Correlation matrix
correlations <- cor(laptops_data %>% select(offer_price, original_price,
total_ratings, total_reviews, rating), use = "complete.obs")

# Heatmap
library(corrplot)
```

```
## corrplot 0.92 loaded
```

```r
corrplot(correlations, method = "circle")
```

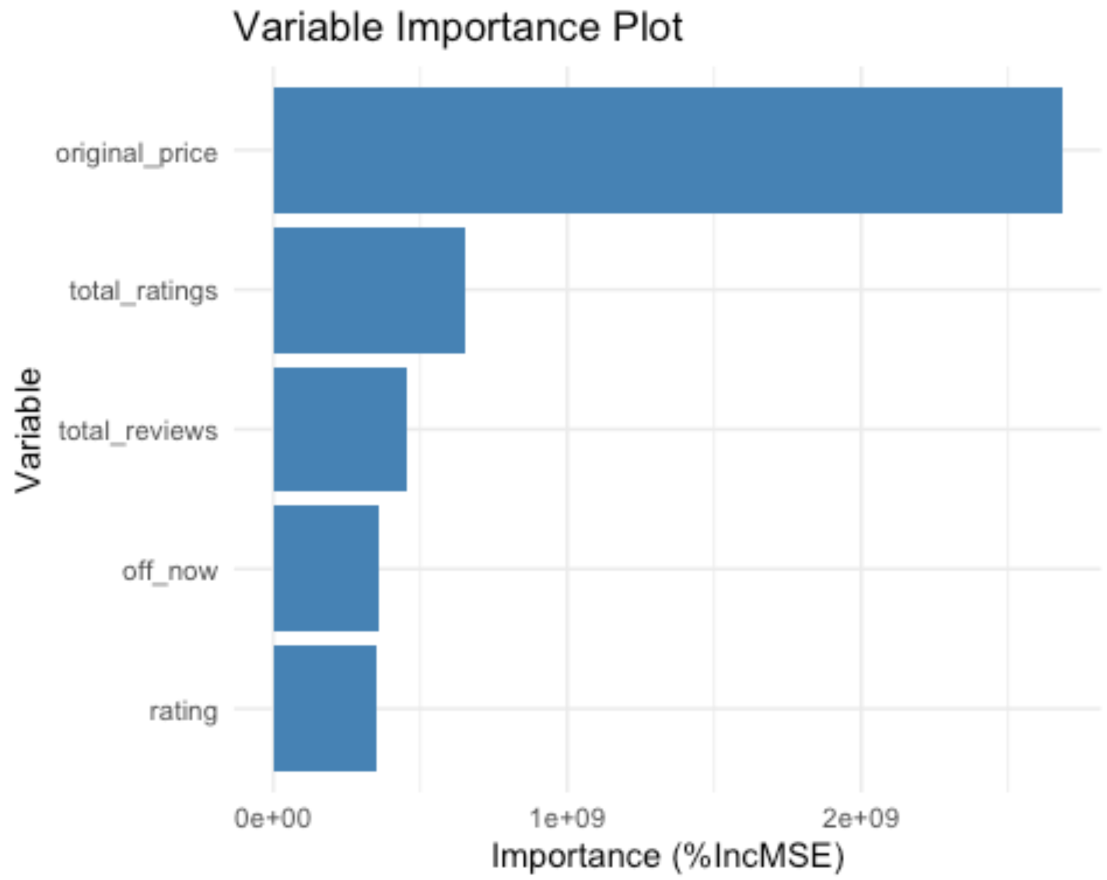


```
# Convert discount levels from percentage to numerical
laptops_data$discount_numeric <- as.numeric(sub("% off", "",
laptops_data$off_now))


# Create scatter plot for Number of Reviews vs. Discount with color based on
Rating
ggplot(laptops_data, aes(x = discount_numeric, y = total_reviews, color =
rating)) +
  geom_point() +
  labs(x = "Discount Level (%)", y = "Total Reviews", title = "Reviews vs.
Discount Levels") +
  theme_minimal()
```



```
# Split the data into training and testing sets
set.seed(123)
train_index <- createDataPartition(laptops_data$offer_price, p = 0.8, list =
FALSE)
train_data <- laptops_data[train_index, ]
test_data <- laptops_data[-train_index, ]

# Build the random forest model
rf_model <- randomForest(offer_price ~ original_price + off_now +
total_ratings + total_reviews + rating,
                         data = train_data, importance = TRUE)

# Get the column names of the importance data frame
importance_cols <- names(rf_model$importance)

# Determine the column name for importance values
importance_col_name <- ifelse("IncNodePurity" %in% importance_cols,
"IncNodePurity", "%IncMSE")

# Create a data frame with variable importance
importance_df <- data.frame(Variable = rownames(rf_model$importance),
                            Importance = rf_model$importance[,
importance_col_name])
```

```
# Plot variable importance››
ggplot(importance_df, aes(x = Importance, y = reorder(Variable, Importance)))
+
  geom_col(fill = "steelblue") +
  labs(x = paste0("Importance (", importance_col_name, ")"), y = "Variable")
+
  ggtitle("Variable Importance Plot") +
  theme_minimal()
```



References

C. Refereneces:

[1] Flipkart - Electronic items prices. (2022, November 10). Kaggle. https://www.kaggle.com/datasets/kiranbudati/mobile-prices-flipkart/data

[2] Anderson, G. (2022, August 29). The 6 most important B2B eCommerce Stakeholders. Corevist. https://www.corevist.com/6-important-b2b-ecommerce-stakeholders/

[3] Agarwal, N. (2024, April 9). The build vs. buy guide for the modern data stack. Monte Carlo Data. https://www.montecarlodata.com/blog-the-build-vs-buy-guide-for-your-modern-data-stack/

[4] R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/