Project Report: Black Friday Sales Data Analysis

Project Title: Black Friday Sales Data Analysis
An Exploratory Analysis Using Python and R

Author: Shijin Ramesh

Objective: To uncover customer purchase patterns from a large Black Friday retail dataset and generate actionable insights for marketing and product optimization using **Python** and **R**.

Note: R was used as a supplementary analysis to Python

Dataset Overview: - **Source:** Kaggle
- **Records:** 537,577 rows
- **Fields:** User demographics, product categories, city type, purchase amount

---

Tools Used: - **Python:** pandas, matplotlib, seaborn
- **R:** ggplot2, dplyr, tidyr
- **Platform:** Jupyter Notebook, RStudio, GitHub, RPubs

Key Analyses & Insights:

1. Demographic Trends

- **Gender Gap:** Male users dominate the dataset and account for higher average purchases.
- **Age Groups:** The age group 26–35 was the most active and highest-spending segment.
- **Marital Status:** Unmarried individuals showed slightly higher purchase trends than married users.

2. Product & Category Analysis

- **Product Category 1** recorded the highest number of purchases, indicating it as the core focus for promotions.
- **Product Categories 12 and 13** had significantly fewer sales, suggesting low engagement or niche appeal.

3. Occupation-Based Behavior

- Occupations 12–16 demonstrated greater variability and higher median purchase values.

- Frequent high-value outliers were observed across almost all occupational categories.
- Users belonging to occupation 4 contributed the highest in revenue.

4. City-Based Distribution

- **City Category** B had the highest overall purchase density.
- **City Category C** showed relatively lower user engagement and purchasing activity.

5. Multi-Variable Analysis (Python & R)

- Combined analysis across Gender, Age, and Marital Status revealed stronger purchase behavior among married males aged 26–35.
- These behaviors were more prominent in urban regions (City Categories A & B).

6. Statistical Testing (R)

- A Chi-Square Test between Gender and Product Category showed a strong and statistically significant association ($p < 0.001$).
- This confirms that purchasing preferences are meaningfully influenced by gender.

Business Recommendations:

- Target male customers aged **26–35** in **all cities** with focused marketing campaigns.
- Prioritize advertising and discounting **Product Category 1**, which performs consistently across demographics.
- Design loyalty programs or upselling campaigns for **Occupations 12–16**, who display higher average spending behavior.

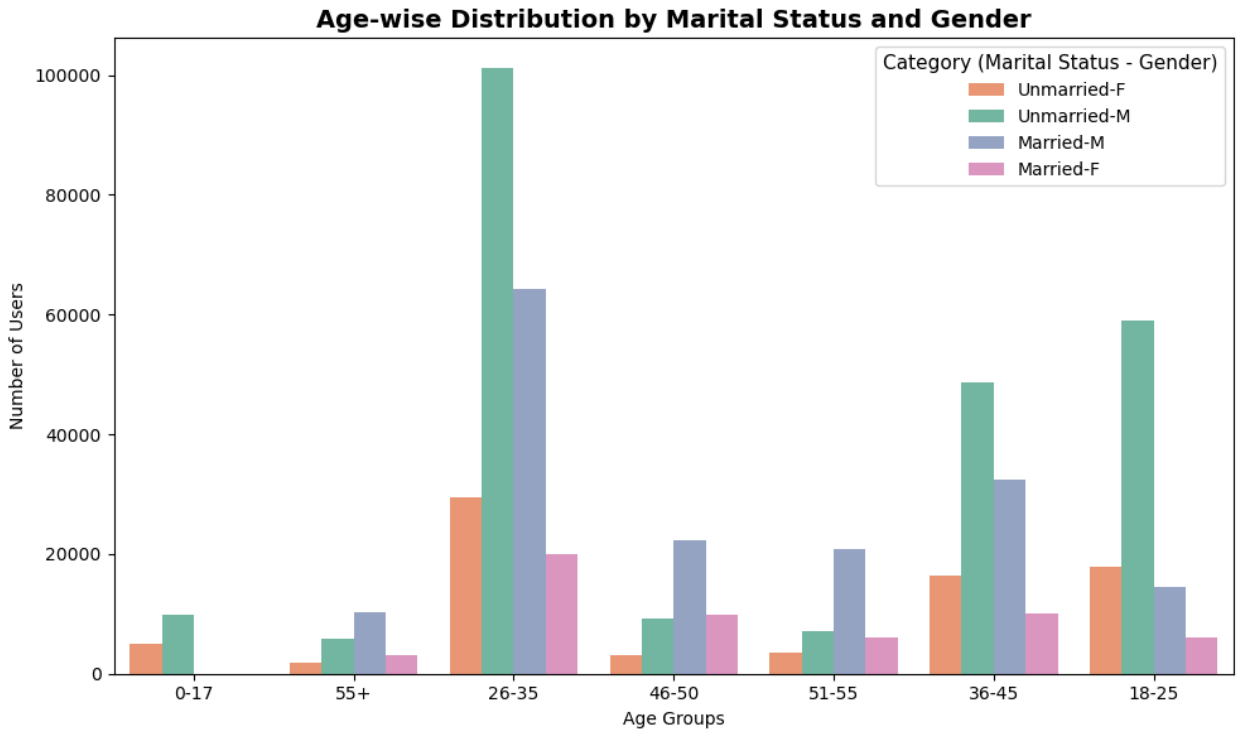
Deliverables: - ✅ Python Jupyter Notebook (EDA)
- ✅ R Markdown Report on RPubs
- ✅ Visualizations: Seaborn & ggplot2
- ✅ PDF Project Report
- ✅ GitHub Repository (with README and project files)

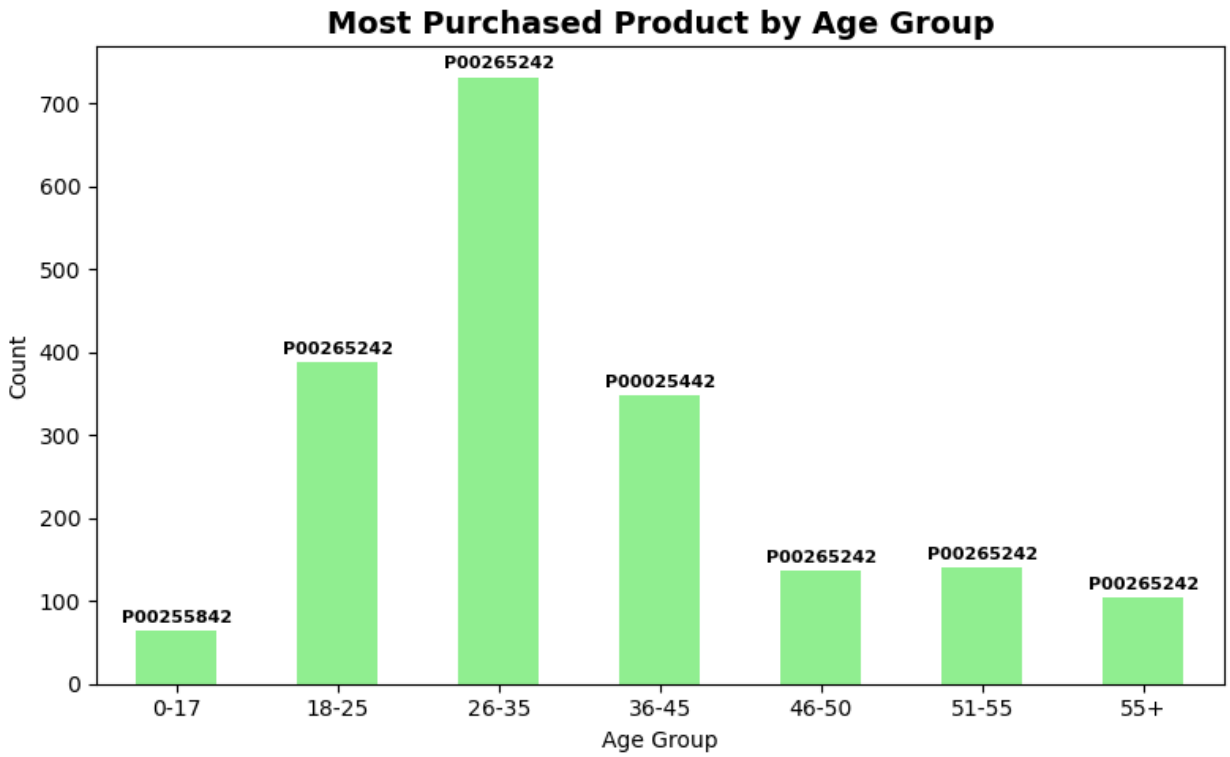
Link to Project:

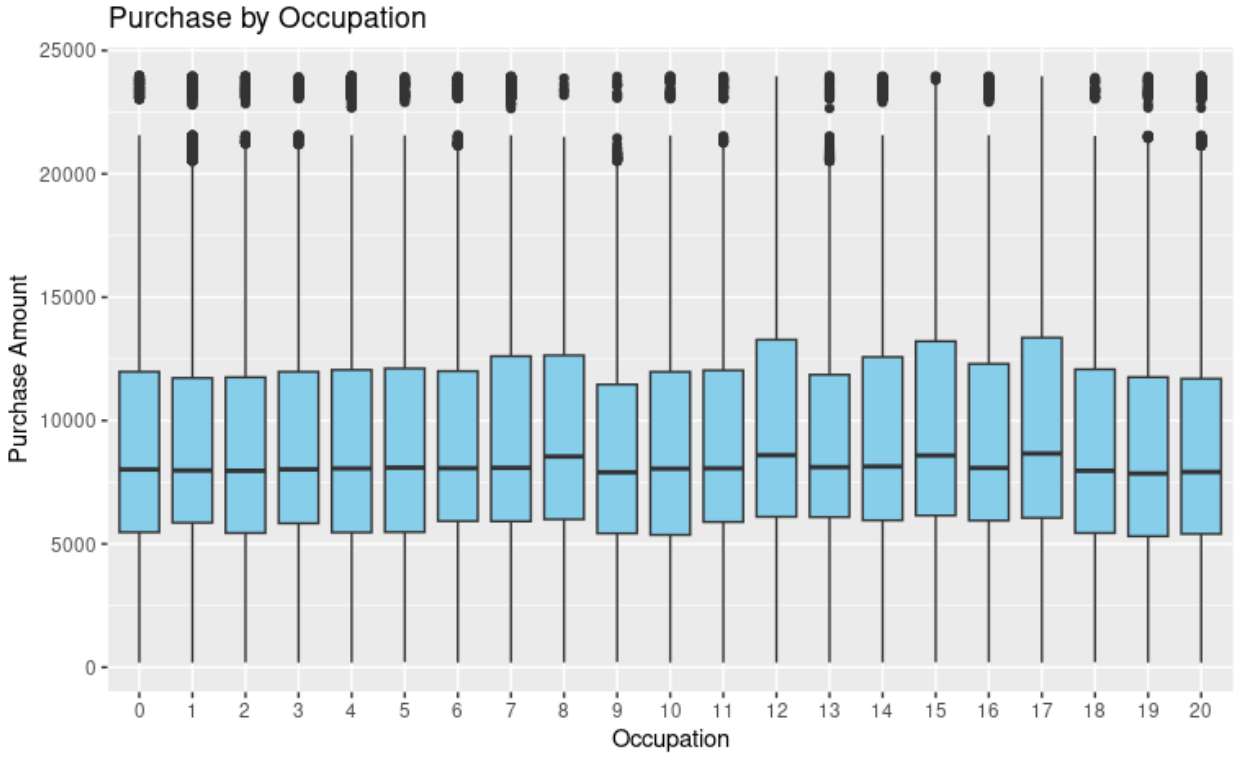
R Report: rpubs.com/shijinramesh/blackfridaysales
GitHub Repo: https://github.com/shijin/BlackFridaySalesDataAnalysis-Python_R

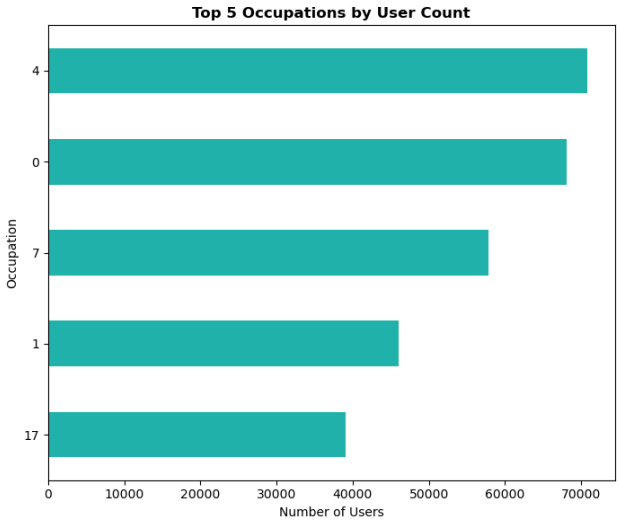
Summary: This cross-platform project demonstrates my ability to perform **end-to-end exploratory analysis**, communicate insights clearly, and combine statistical and visual storytelling techniques — all essential for solving real-world data problems as a Data Analyst.

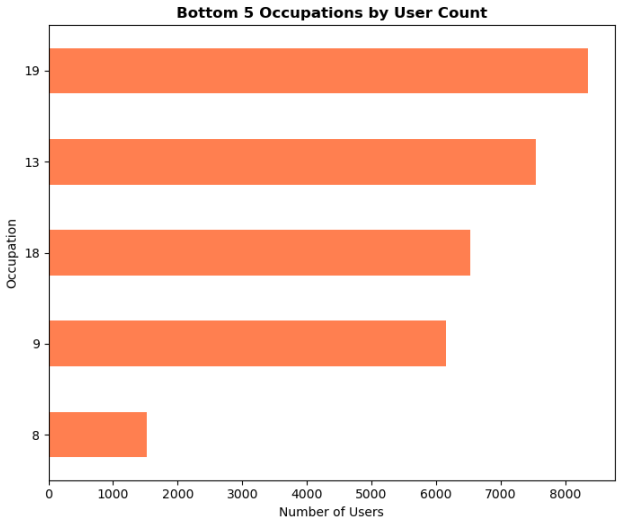
Visuals:

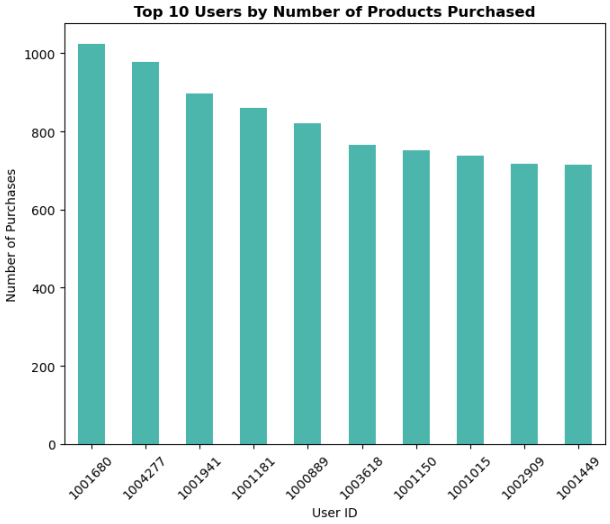


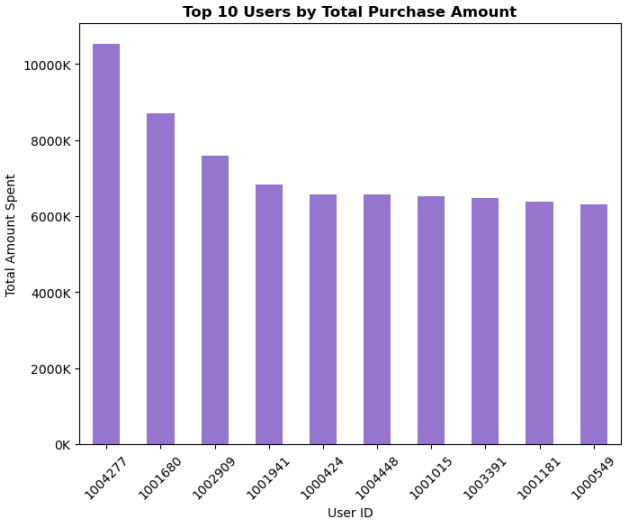

Most Purchased Product by Age Group
P00265242
P00265242
P00025442
P00265242
P00265242
P00265242
P00255842
Count
0
100
200
300
400
500
600
700
0-17
18-25
26-35
36-45
46-50
51-55
55+
Age Group

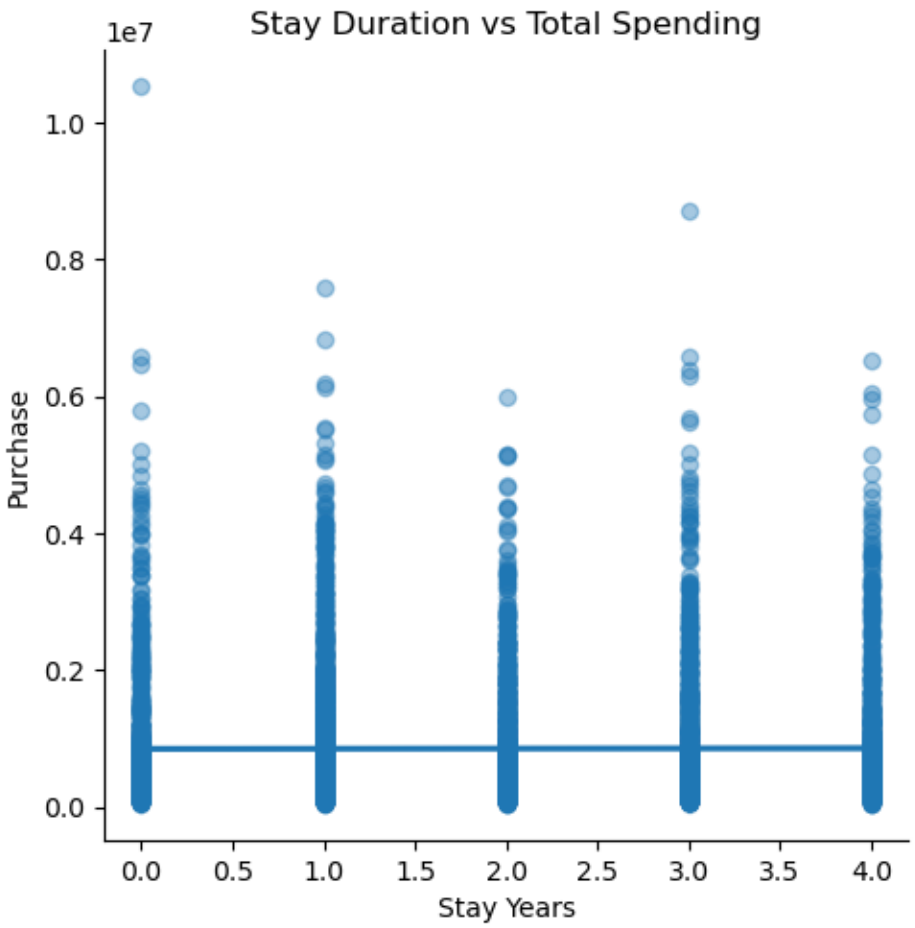

Purchase by Occupation
Purchase Amount
0
5000
10000
15000
20000
25000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Occupation

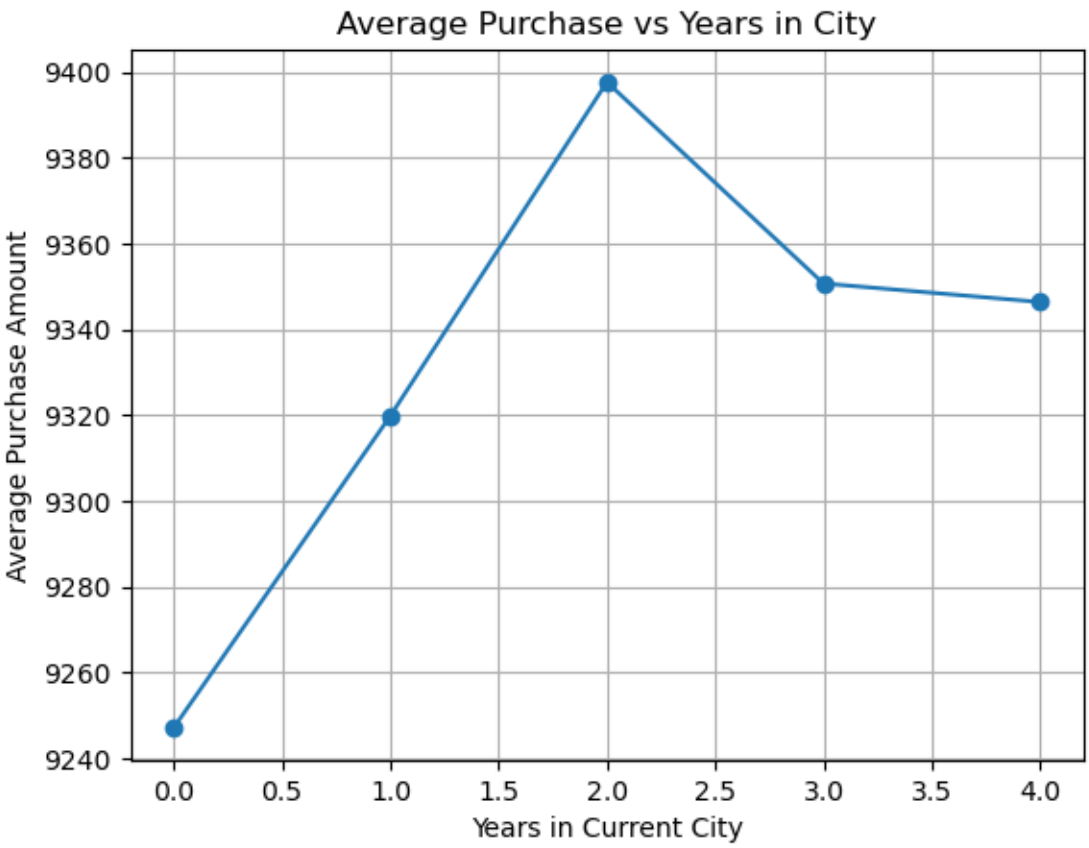

Top 5 Occupations by User Count
Occupation
4
0
7
1
17
0
10000
20000
30000
40000
50000
60000
70000
Number of Users


Bottom 5 Occupations by User Count
Occupation
19
13
18
9
8
0
1000
2000
3000
4000
5000
6000
7000
8000
Number of Users


Top 10 Users by Number of Products Purchased
Number of Purchases
0
200
400
600
800
1000
1001680
1004277
1001941
1001181
1000889
1003618
1001150
1001015
1002909
1001449
User ID


Top 10 Users by Total Purchase Amount
Total Amount Spent
0K
2000K
4000K
6000K
8000K
10000K
1004277
1001680
1002909
1001941
1000424
1004448
1001015
1003391
1001181
1000549
User ID


Stay Duration vs Total Spending
1e7
1.0
0.8
0.6
0.4
0.2
0.0
Purchase
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Stay Years


Average Purchase vs Years in City
Average Purchase Amount
9400
9380
9360
9340
9320
9300
9280
9260
9240
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Years in Current City