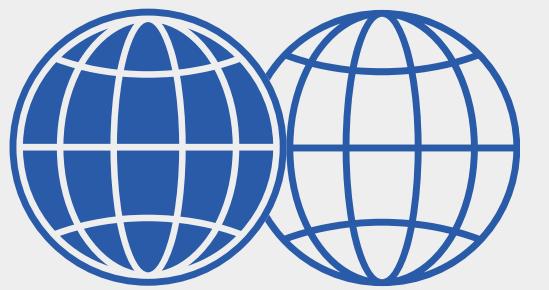


NAME: TRAN THI THU HUONG

SWISS CODING ACADEMY

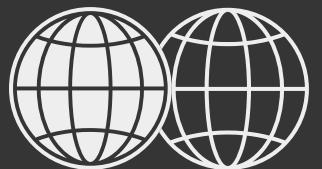
22/08/2025



FINAL PROJECT

TOPIC

TABLE OF **CONTENTS**



- CONTEXT
- RESEARCH OBJECTIVES
- RESEARCH QUESTIONS
- DATASET INTRODUCTION
- EXPLORATORY DATA ANALYST
- MODEL
- KEY INSIGHTS
- RECOMMENDATIONS



5 BILLION

internet users worldwide

CONTEXT

Retail e-commerce sales are estimated to exceed

4.3 TRILLION USD



Data analytics has become increasingly important in today's e-commerce landscape

TOPIC UNDERSTAND BUSINESS PERFORMANCE AND DETERMINE FACTORS INFLUENCE SALES BY USING LINEAR REGRESSION MODEL.

PURPOSES

- Line marketing/Sales managers
- Marketing/Sales team
- R&D line managers
- Other stakeholders

OUTCOMES

- Line marketing/Sales managers
- Marketing/Sales team
- R&D line managers
- Other stakeholders

AUDIENCE

- Line marketing/Sales managers
- Marketing Analysts
- Marketing/Sales team
- R&D line managers
- Other stakeholders

PROJECT OVERVIEW

DATA PROFILING



Data source: public dataset from Kaggle

Original Dataset

- CustomerData.xlsx
- Discount_Coupon.csv
- Online_Sales.csv
- Marketing_Spend.csv

No. of records: 52924 rows

Aggregated metrics:

- **Gross_Revenue:** Avg_Price*Quantity
- **Discount_Amount:** Avg_Price * Quantity * Discount_pct
- **Purchase_Amount (~ Sales):** Avg_Price *Quantity*(1-Discoun

CustomerID	Transaction_ID	Transaction_Date	Product_SKU	Product_Description	Product_Category	Quantity	Avg_Price	Delivery_Charges	Coupon_Status	Gender	Location	Tenure_Months	Month	Discount_pct	Purchase_Amount	Discount_Amount	Gross_Revenue
12567	20436	2019-02-14	GGOENEBB078899	Nest Cam Indoor Security Camera - USA	Nest-USA	2	119.00	6.5	Clicked	F	Washington DC	23	Feb	0.2	244.500	0.000	238.00
13328	43293	2019-11-13	GGOEGDHC018299	Google 22 oz Water Bottle	Drinkware	1	2.39	6.0	Clicked	M	Chicago	18	Nov	0.2	8.390	0.000	2.39
15061	26347	2019-04-24	GGOEGBMJ013399	Sport Bag	Bags	1	3.99	6.0	Not Used	M	California	7	Apr	0.1	9.990	0.000	3.99
17428	35920	2019-08-17	GGOEGFKA022299	Keyboard DOT Sticker	Office	6	1.20	6.0	Used	F	California	14	Aug	0.2	11.760	1.440	7.20
13069	27975	2019-05-13	GGOEGAFB035816	Google Men's Zip Hoodie	Apparel	1	44.79	6.0	Used	F	New York	38	May	0.2	41.832	8.958	44.79

MONTHLY SALES

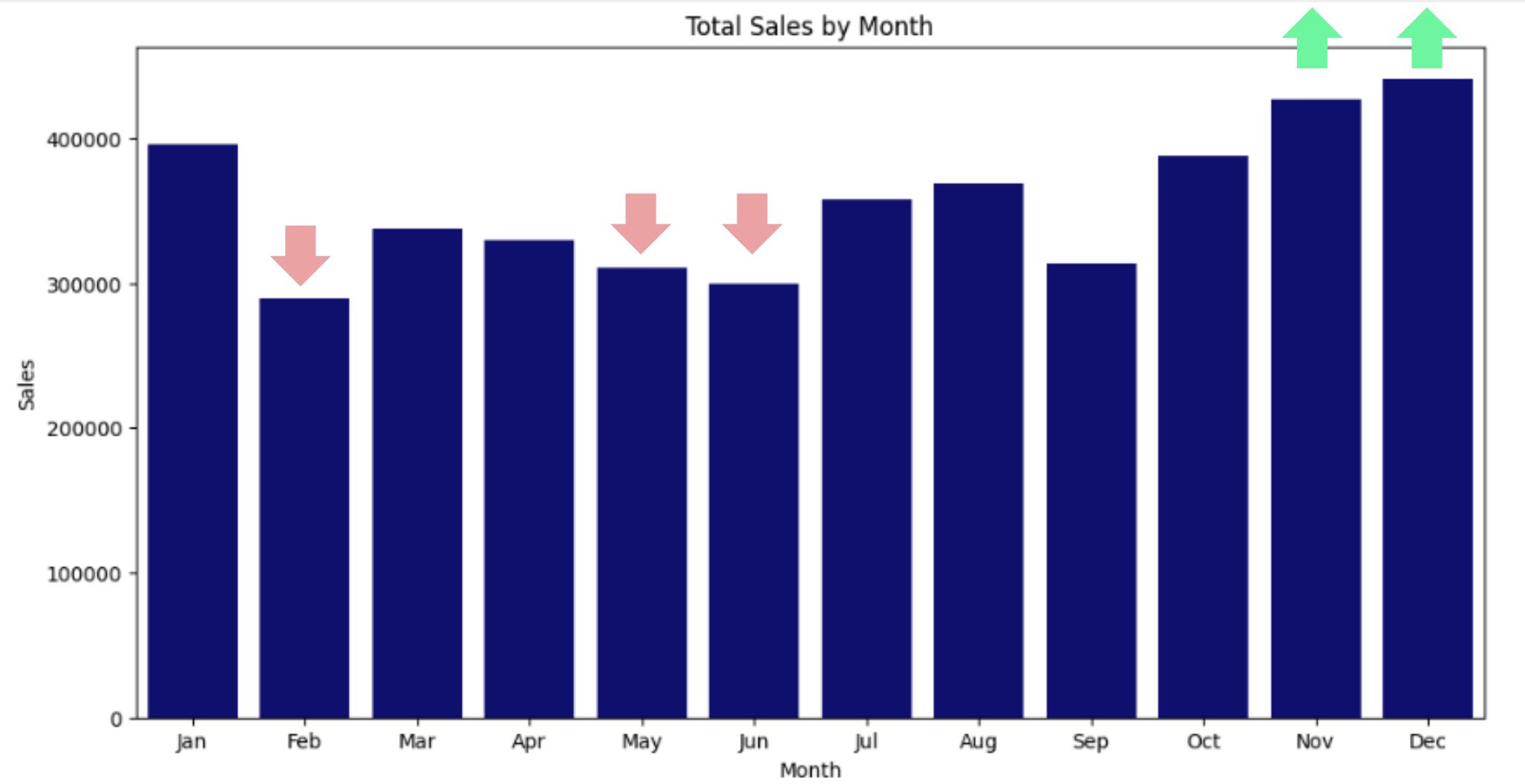
Key Takeaway:

The revenue of the company fluctuated depending on the seasonal factors

- **Peak revenue months** (Jan, Oct, Nov, Dec)
- **Low revenue months** (Feb, May, June).

Questions?

Why there is such difference patterns in each months?



	Month	Total sale (\$)	No. of transaction	Avg basket value (\$)	No. of unique customers	Purchase frequency
Low sales	Feb	289,657	1627	178.03	109	14.93
	May	311178	2007	155.05	200	10.04
	Jun	299311	1914	156.38	259	7.39
High sales	Nov	427380	2174	196.59	188	11.56
	Dec	441363	2610	169.10	236	11.06

Small but highly loyal customer base + frequent and high basket purchases.

Larger number of customers, but spending per purchase was smaller, and purchase frequency dropped. → **Low engagement.**

Highest basket value + frequently loyal customer base

Largest customers

FOCUSED ANALYSIS

PRODUCT TIERS CLASSIFICATION

Context

- **Problem:** significant product outliers due to various price options.
- **Solution:** divide product categories that have homogenous characteristics and analyze each group independently.



Methodology

- **Step 1:** Mathematically divide the product categories by 3 according to their average price to have 3 equal sub-groups.
- **Step 2:** adjust the list in each group by intuitive reasoning.

	Product_Category	Avg_Price
11	Nest-Canada	157.09
10	Nest	148.09
12	Nest-USA	124.33
6	Gift Cards	111.36
3	Bags	30.47
2	Apparel	19.78
7	Headgear	15.91
1	Android	15.90
13	Notebooks & Journals	11.81
5	Drinkware	10.82
0	Accessories	8.29
15	Waze	6.59
9	Lifestyle	3.89
14	Office	3.79
4	Bottles	3.43
8	Housewares	2.07

Divide into
3 equal
groups

	Product_Category	Avg_Price
11	Nest-Canada	157.09
10	Nest	148.09
12	Nest-USA	124.33
6	Gift Cards	111.36
3	Bags	30.47
2	Apparel	19.78
7	Headgear	15.91
1	Android	15.90
13	Notebooks & Journals	11.81
5	Drinkware	10.82
0	Accessories	8.29
15	Waze	6.59
9	Lifestyle	3.89
14	Office	3.79
4	Bottles	3.43
8	Housewares	2.07

Adjust the
price range

FOCUSED ANALYSIS

PRODUCT TIER CLASSIFICATION

HIGH TIER

Dominate revenue but are heavily skewed toward year-end months.

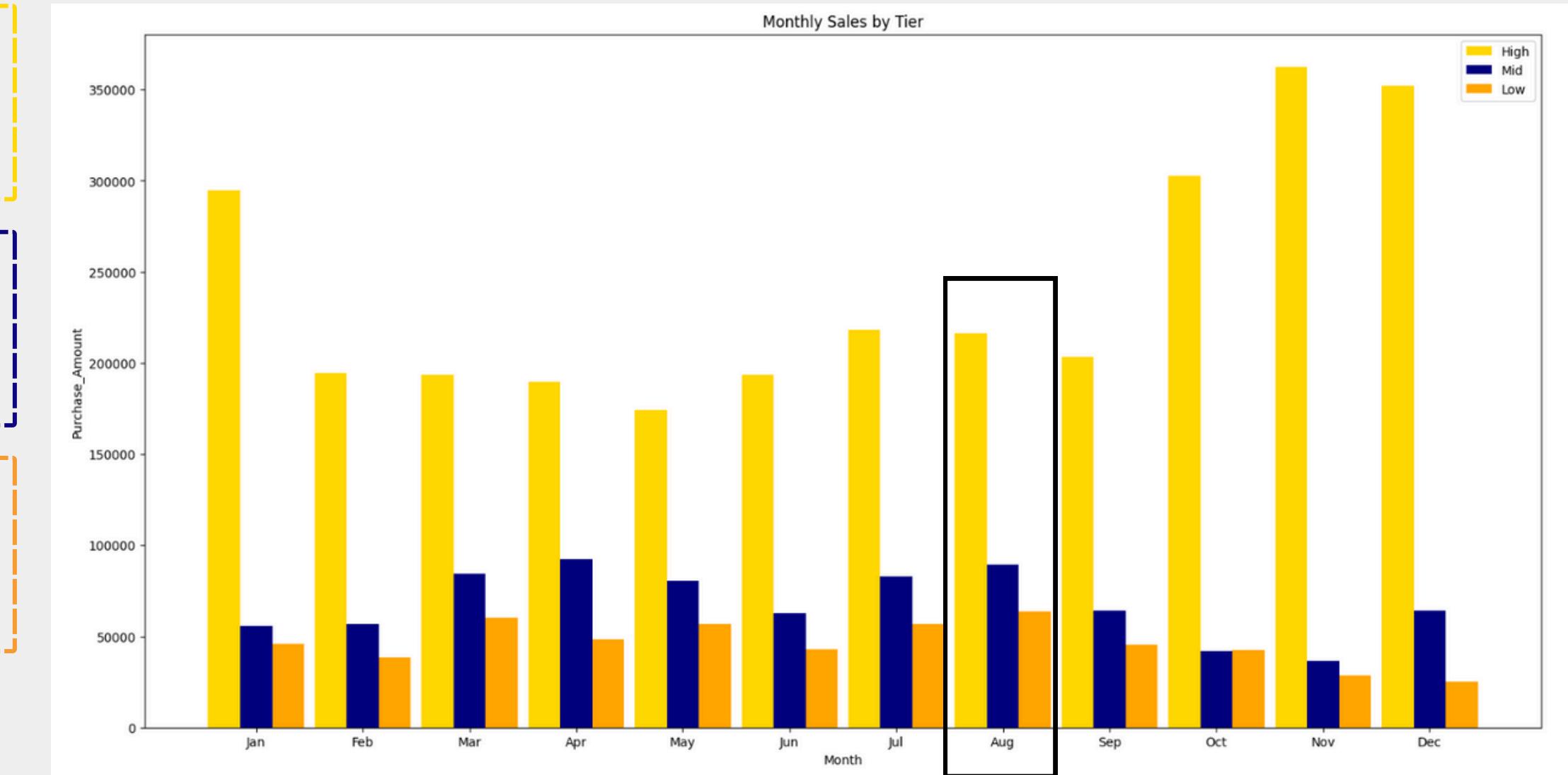
MID TIER

Sales fluctuate, peaking in April and August.

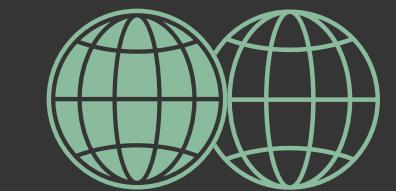
LOW TIER

Sales remain stable but low contribution.

Over-reliance on high-tier products. Sales could grow significantly by boosting all tiers.



FOCUSED ANALYSIS



Top purchased product categories

	Product_Category	Quantity
18	Office	88383
2	Apparel	32438
6	Drinkware	30501
12	Lifestyle	24881
16	Nest-USA	21430
4	Bags	15273
17	Notebooks & Journals	9556
10	Headgear	3533
14	Nest	2837
11	Housewares	2484
5	Bottles	2090
19	Waze	1144
0	Accessories	1097
7	Fun	834
9	Google	575
15	Nest-Canada	469
8	Gift Cards	206
13	More Bags	144
3	Backpacks	113
1	Android	45

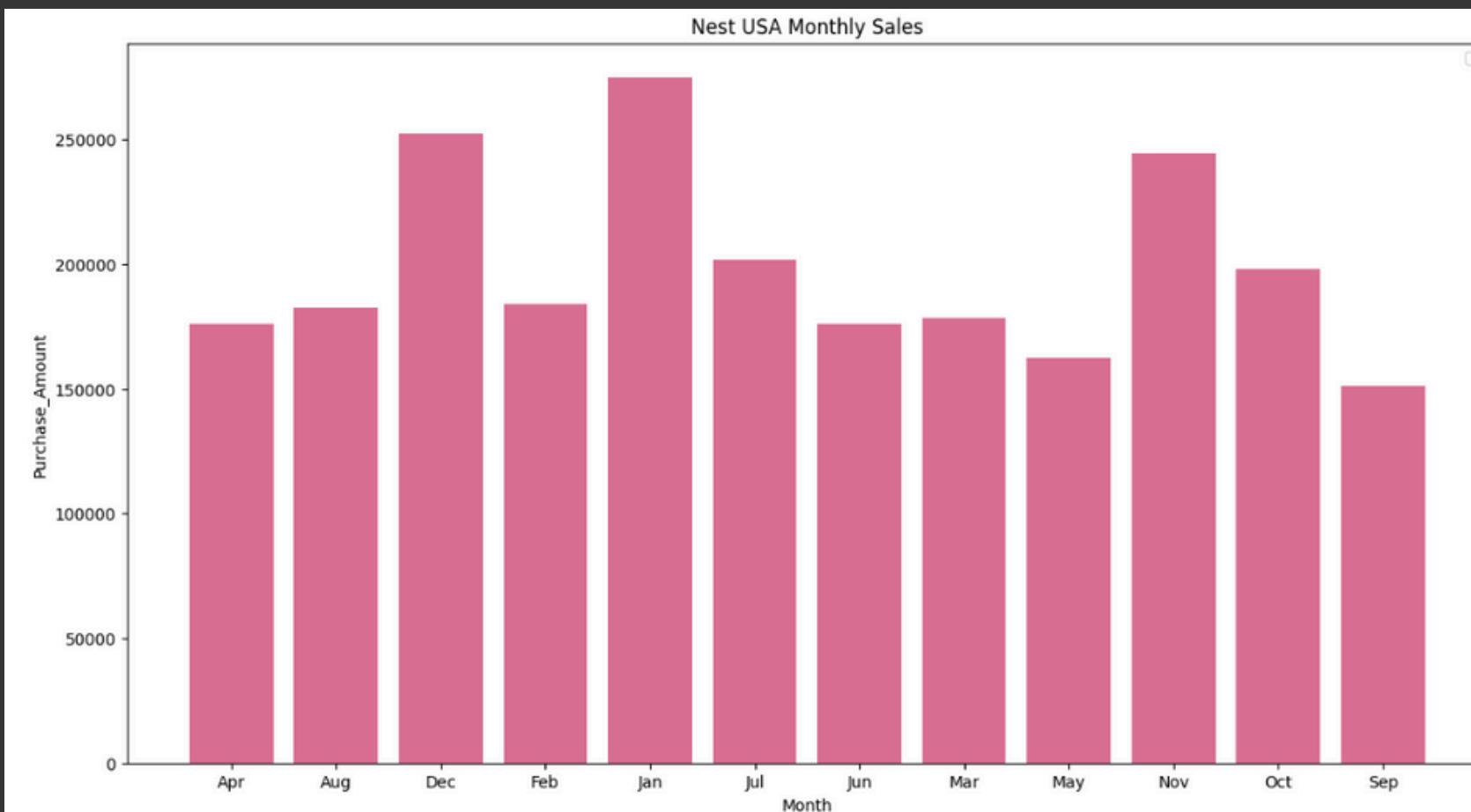
Top sales product categories

	Product_Category	Purchase_Amount
16	Nest-USA	2383641.541
2	Apparel	551665.061
14	Nest	480543.688
18	Office	257724.871
6	Drinkware	188554.055
4	Bags	142357.257
17	Notebooks & Journals	101533.484
12	Lifestyle	69525.287
15	Nest-Canada	66675.282
10	Headgear	47998.043
8	Gift Cards	18521.375
9	Google	9420.470
3	Backpacks	8772.690
0	Accessories	6738.200
5	Bottles	6349.852
7	Fun	6029.010
19	Waze	5892.313
11	Housewares	4451.240
13	More Bags	2946.960
1	Android	681.483

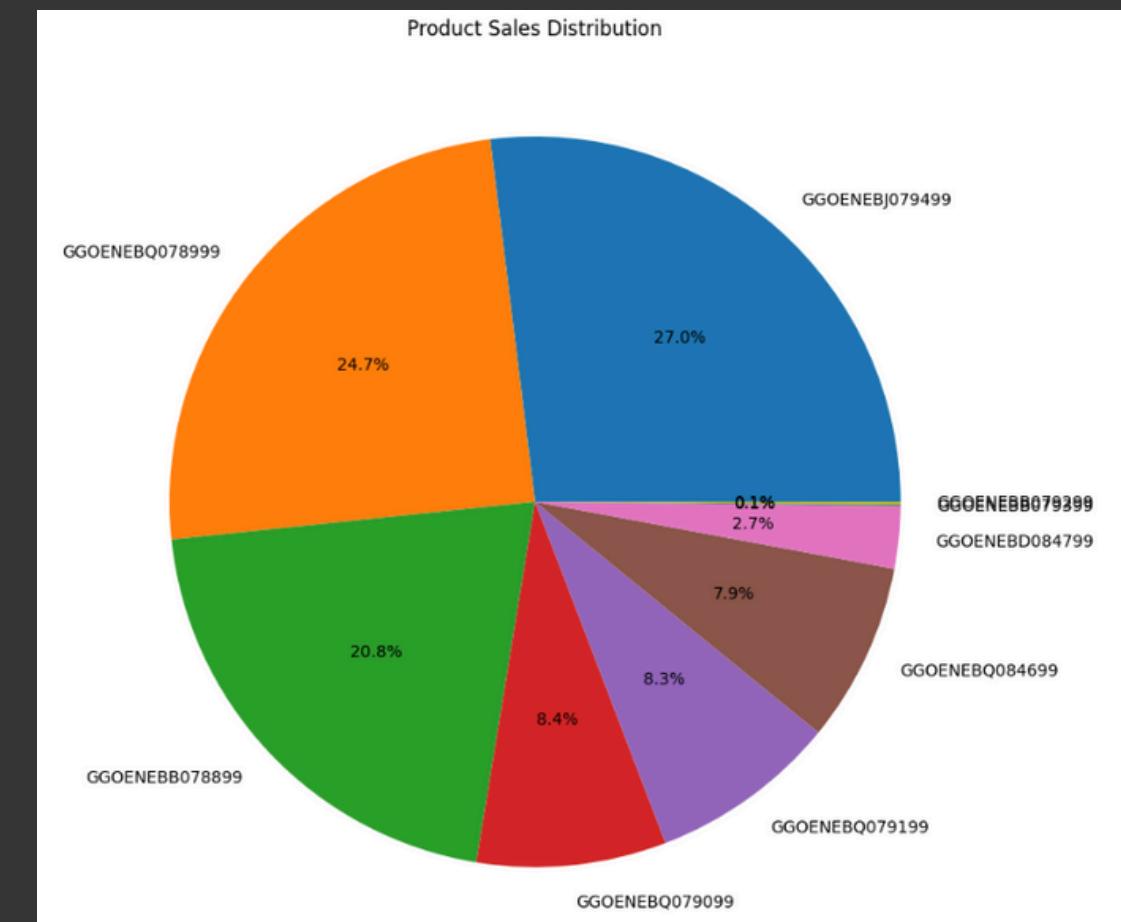
Key takeaway:

NEST USA PERFORMANCE

Product description: household technology devies (e.g. security camera, protect smoke,...)



Key takeaways:
re magna aliqua. Ut enim ad minim
veniam, quis nostrud exercitation
ullamco laboris nisi ut aliquip ex ea
commodo consequat

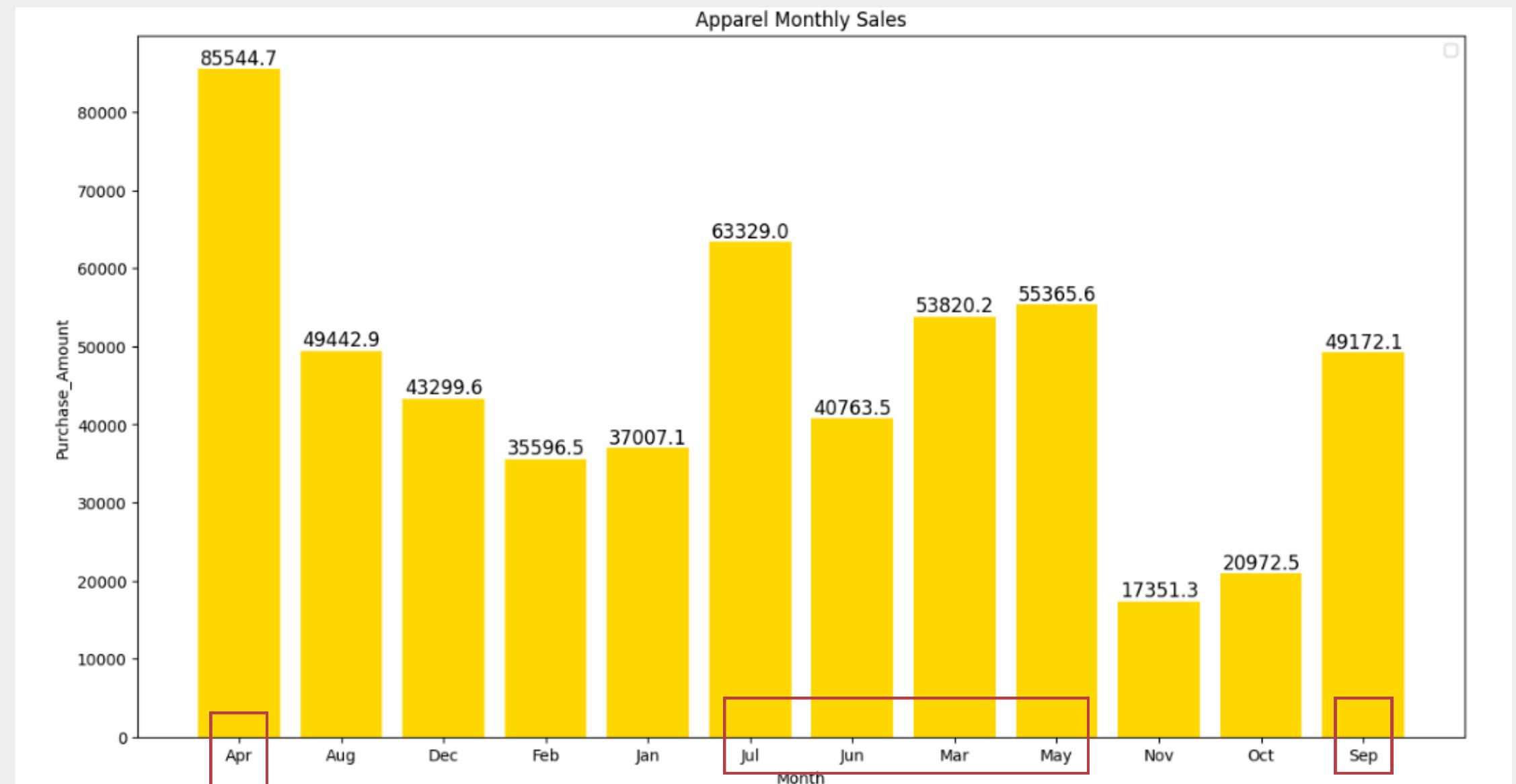


APPAREL USA

Product description: clothes

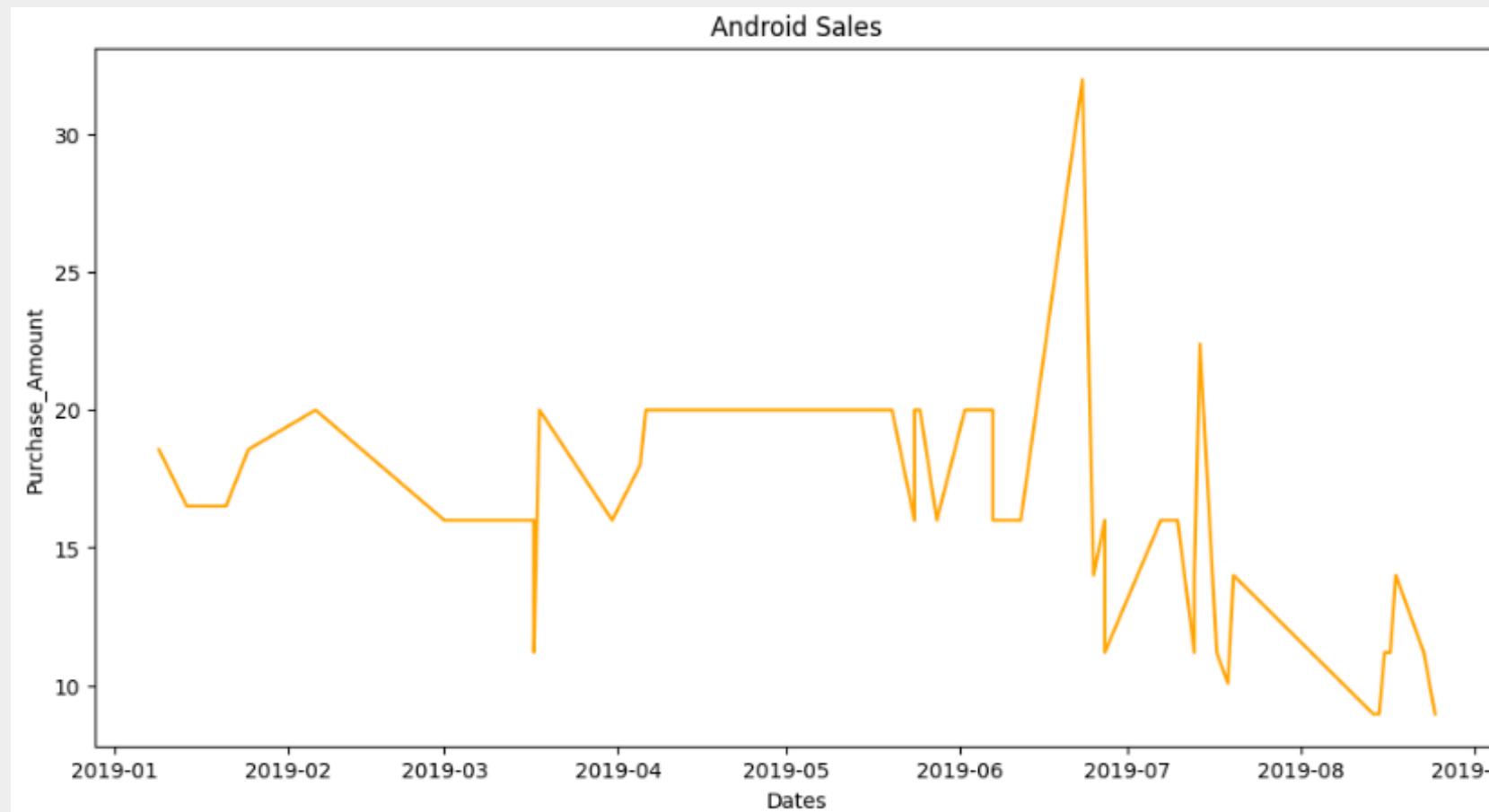
— PERFORMANCE —

- Have high sales volume in mid year (From April to September).
 - Possible reasons: high season of hiring. The B2B customers need to buy uniforms, clothes for their new employees.
- Sales peaked in mid-year.



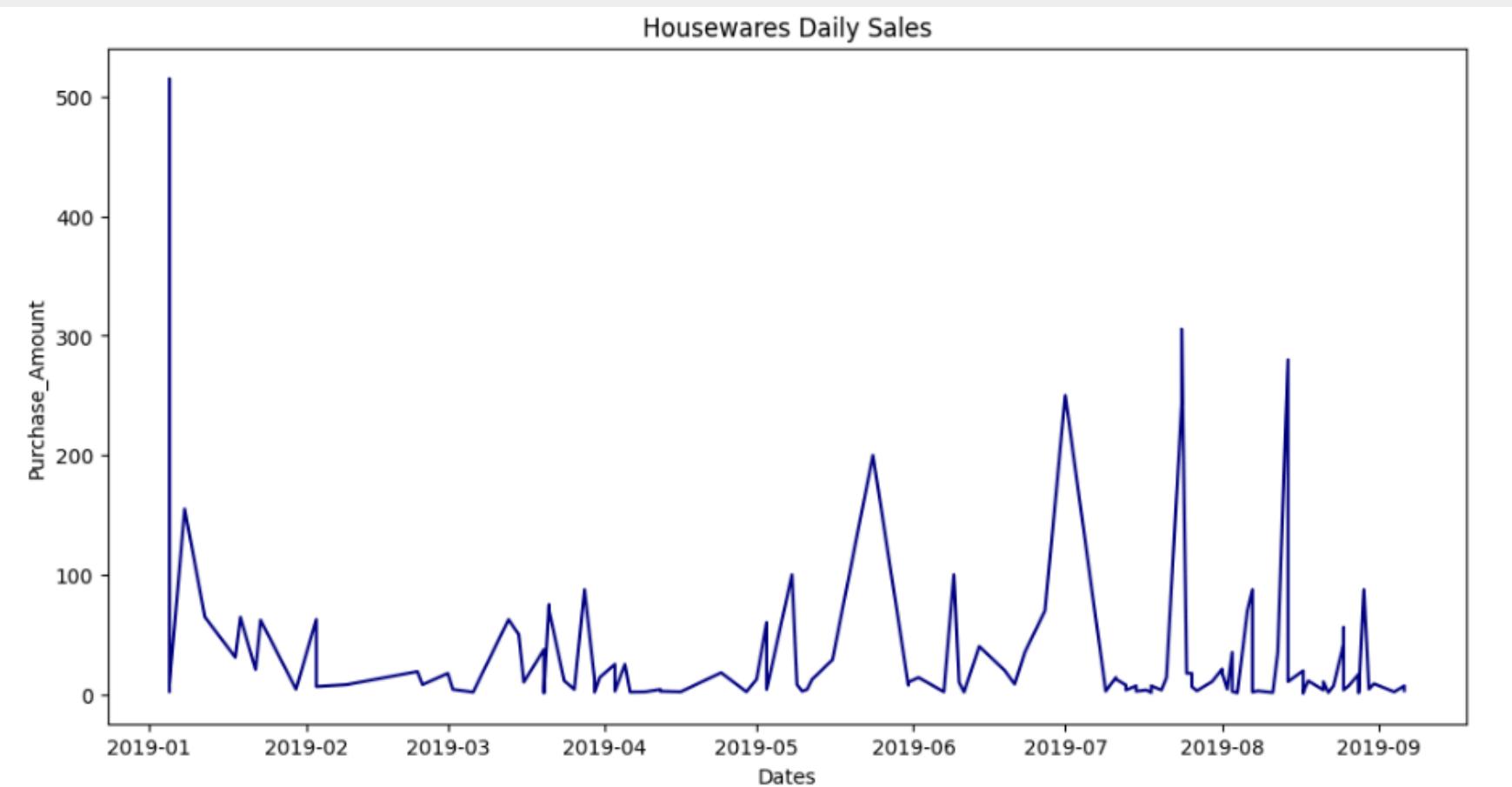
ANDROID

- accounted for only **0.015%** of total sales (the lowest)
- average quantity sold per month was **276 items**

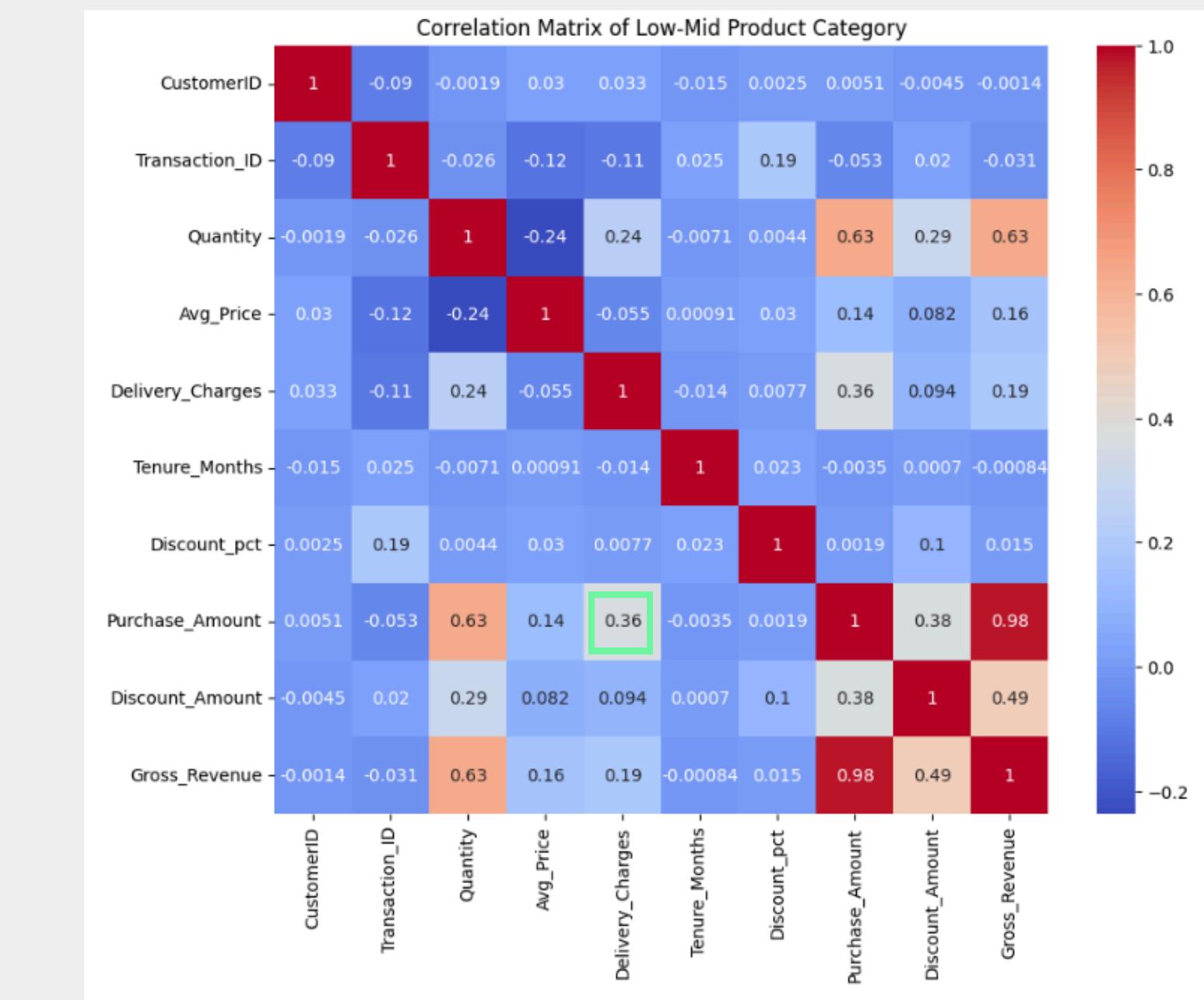
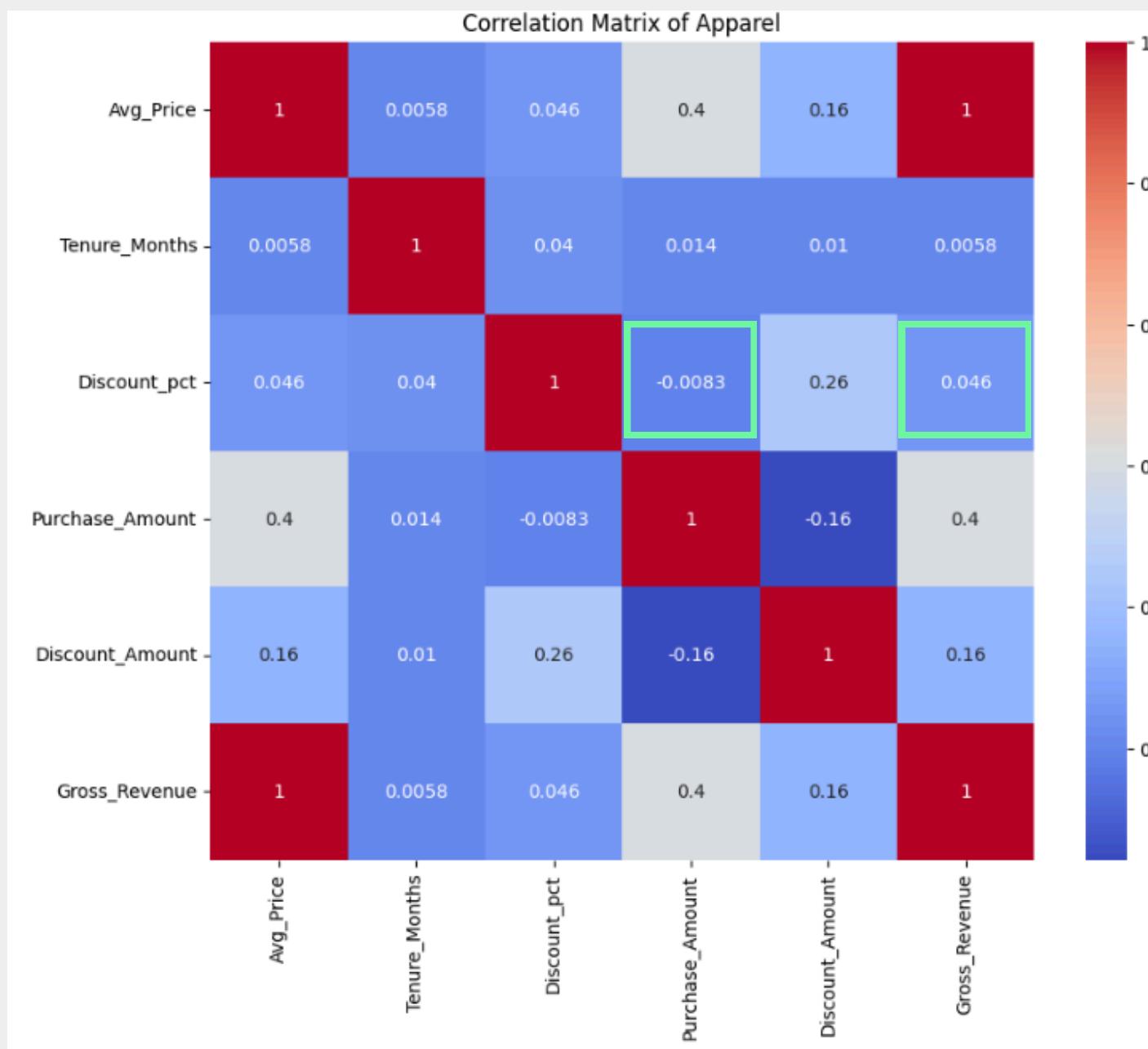


HOUSEWARES

- accounted for only **0.102%** of total sales
- average quantity sold per month was **5 items**



CORRELATION



LINEAR REGRESSION MODEL

```
OLS Regression Results
=====
Dep. Variable: Quantity R-squared: 0.055
Model: OLS Adj. R-squared: 0.055
Method: Least Squares F-statistic: 940.0
Date: Fri, 15 Aug 2025 Prob (F-statistic): 0.00
Time: 10:46:34 Log-Likelihood: -1.1096e+05
No. Observations: 32058 AIC: 2.219e+05
Df Residuals: 32055 BIC: 2.220e+05
Df Model: 2
Covariance Type: nonrobust
=====
            coef    std err        t      P>|t|      [0.025      0.975]
-----
const      2.0582   0.122     16.927     0.000     1.820     2.297
Delivery_Charges  0.1852   0.004     43.351     0.000     0.177     0.194
Discount_pct     0.2554   0.534      0.478     0.632    -0.791     1.302
=====
Omnibus: 25941.236 Durbin-Watson: 1.119
Prob(Omnibus): 0.000 Jarque-Bera (JB): 525664.730
Skew: 3.929 Prob(JB): 0.00
Kurtosis: 21.215 Cond. No. 179.
=====
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Hannah Morales

THANK YOU

www.reallygreatsite.com



