



Online Shopping Data Analysis

Group 1: Anqi Li, Chi-Fang Cheng, Isaac Chen, Wanqing Yu



EDA Report



DESCRIPTION

The dataset is composed of individual product purchases of customers from a specific online retail company of Nest products from Google, attributes in this dataset include some basic customer demographics, transaction details, promotion and related details, and product details.

- > 52955 Entities with 20 Attributes
- > **Data Collection Period:** 2019/1/1 - 2019/12/31
- > **Unique Identifier:** TransactionID + ProductSKU



1. **CustomerID:** Unique identifier for each customer.
2. **Gender:** Gender of the customer (e.g., Male, Female).
3. **Location:** Location information of the customer in terms of city or state.
4. **Tenure_Months:** Number of months the customer has been associated with the platform.
5. **Transaction_ID:** Unique identifier for each transaction.
6. **Transaction_Date:** Date of the transaction.
7. **Product_SKU:** Stock Keeping Unit (SKU) identifier for the product.
8. **Product_Description:** Description of the product.
9. **Product_Category:** Category to which the product belongs.
10. **Quantity:** Quantity of the product purchased in the transaction.
11. **Avg_Price:** Average price of the product.
12. **Delivery_Charges:** Charges associated with the delivery of the product.
13. **Coupon_Status:** Status of the coupon usage associated with the transaction.
14. **GST:** Goods and Services Tax associated with the transaction.
15. **Date:** Date of the transaction (potentially redundant with Transaction_Date).
16. **Offline_Spend:** Amount spent offline by the customer.
17. **Online_Spend:** Amount spent online by the customer.
18. **Month:** Month of the transaction.
19. **Coupon_Code:** Code associated with a coupon, if applicable.
20. **Discount_pct:** Percentage of discount applied to the transaction



Customer Info (Categories)

- > CustomerID
- > Gender
- > Location

Customer Info - Activity (Measures)

- > Tenure_Months
- > Offline_Spend
- > Online_Spend

Product Info (Categories)

- > Product_Desription

Transaction Details (Categories)

- > Transaction_ID
- > Transaction_Date
- > Product_SKU
- > Product_Category
- > Coupon_Status
- > Coupon_Code
- > Date
- > Month

Transaction Details (Measures)

- > Quantity
- > Avg_Price
- > Delivery_Charges
- > GST
- > Discount_pct



Missing values:

- > TransactionID + ProductSKU: There are 31 rows of null values in this combination, since we are accessing customer purchasing behaviors, we would remove or filter out these data.

Redundant Columns:

- > The values in Date and TransactionDate were exactly the same, which we would be using only one of them.

Outliers:

- > There are extreme outliers in AvgPrice and Quantity, we might isolate them after setting thresholds. We would be mindful when using aggregations like average, so that we take into consideration of these outliers.

Regional Variability:

- > The data only includes 5 states (California, Chicago, New Jersey, New York, Washington DC), which could lead to lack of representativeness that may not accurately reflect the broader population or geographic diversity.

Dataset Screenshot



	CustomerID	Gender	Location	Tenure_Months	Transaction_ID	Transaction_Date	Product_SKU	Product_Description
0	17850.0	M	Chicago	12.0	16679.0	2019-01-01	GGOENEBJ079499	Nest Learning Thermostat 3rd Gen-USA - Stainless Steel
1	17850.0	M	Chicago	12.0	16680.0	2019-01-01	GGOENEBJ079499	Nest Learning Thermostat 3rd Gen-USA - Stainless Steel
2	17850.0	M	Chicago	12.0	16696.0	2019-01-01	GGOENEQBQ078999	Nest Cam Outdoor Security Camera - USA
3	17850.0	M	Chicago	12.0	16699.0	2019-01-01	GGOENEQBQ079099	Nest Protect Smoke + CO White Battery Alarm-USA
4	17850.0	M	Chicago	12.0	16700.0	2019-01-01	GGOENEBJ079499	Nest Learning Thermostat 3rd Gen-USA - Stainless Steel
5	17850.0	M	Chicago	12.0	16701.0	2019-01-01	GGOENEBJ079499	Nest Learning Thermostat 3rd Gen-USA - Stainless Steel

Product_Category	Quantity	Avg_Price	Delivery_Charges	Coupon_Status	GST	Date	Offline_Spend	Online_Spend	Month	Coupon_Code	Discount_pct
Nest-USA	1.0	153.71	6.5	Used	0.1	1/1/2019	4500.0	2424.5	1	ELEC10	10.0
Nest-USA	1.0	153.71	6.5	Used	0.1	1/1/2019	4500.0	2424.5	1	ELEC10	10.0
Nest-USA	2.0	122.77	6.5	Not Used	0.1	1/1/2019	4500.0	2424.5	1	ELEC10	10.0
Nest-USA	1.0	81.5	6.5	Clicked	0.1	1/1/2019	4500.0	2424.5	1	ELEC10	10.0
Nest-USA	1.0	153.71	6.5	Clicked	0.1	1/1/2019	4500.0	2424.5	1	ELEC10	10.0
Nest-USA	1.0	153.71	6.5	Clicked	0.1	1/1/2019	4500.0	2424.5	1	ELEC10	10.0



Statistical Summary

12 Dimensions

	Null Count	Non-Null Count	Distinct Count
CustomerID	31	52924	1468
Gender	31	52924	2
Location	31	52924	5
Transaction_ID	31	52924	25061
Transaction_Date	31	52924	365
Product_SKU	31	52924	1145
Product_Description	31	52924	404
Product_Category	0	52955	21
Coupon_Status	31	52924	3
Date	31	52924	365
Month	0	52955	12
Coupon_Code	400	52555	48



Statistical Summary

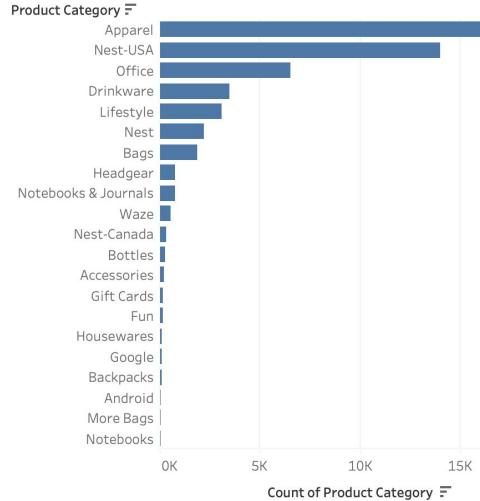
8 Measures

	Tenure_Months	Quantity	Avg_Price	Delivery_Charges	GST	Offline_Spend	Online_Spend	Discount_pct
count	52924.000000	52924.000000	52924.000000	52924.000000	52924.000000	52924.000000	52924.000000	52555.000000
mean	26.127995	4.497638	52.237646	10.517630	0.137462	2830.914141	1893.109119	19.953382
std	13.478285	20.104711	64.006882	19.475613	0.045825	936.154247	807.014092	8.127108
min	2.000000	1.000000	0.390000	0.000000	0.050000	500.000000	320.250000	10.000000
25%	15.000000	1.000000	5.700000	6.000000	0.100000	2500.000000	1252.630000	10.000000
50%	27.000000	1.000000	16.990000	6.000000	0.180000	3000.000000	1837.870000	20.000000
75%	37.000000	2.000000	102.130000	6.500000	0.180000	3500.000000	2425.350000	30.000000
max	50.000000	900.000000	355.740000	521.360000	0.180000	5000.000000	4556.930000	30.000000

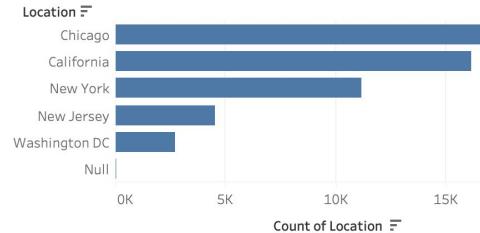


Univariate Visualization: Categories

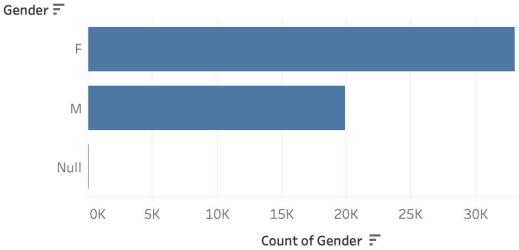
Product Category



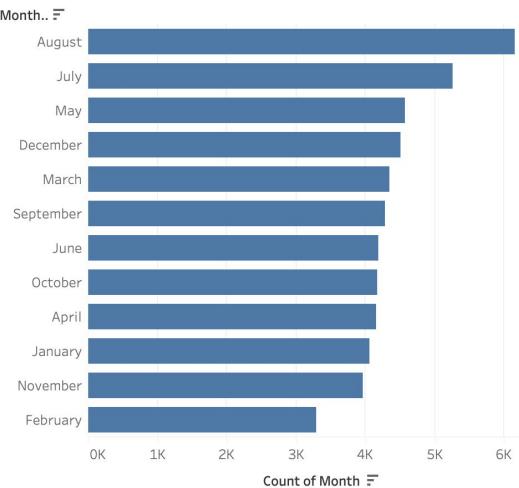
Location



Gender

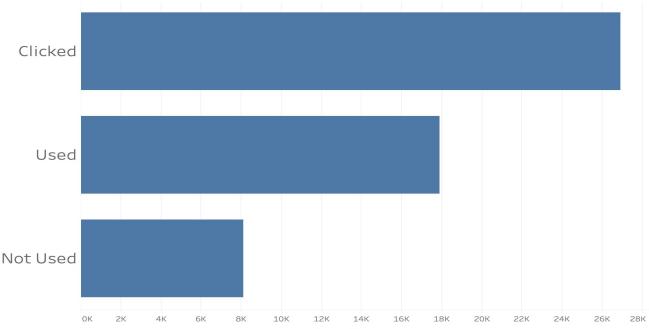


Month



- 7 main products
- Customer base is mostly female
- More sales are made in Chicago
- During the summer months the most sales are made
- In terms of coupon status, a majority of coupons are used

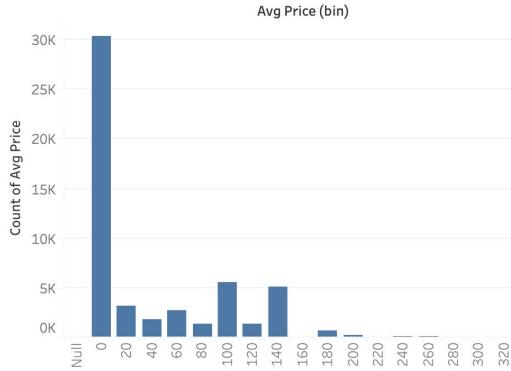
Coupon Status



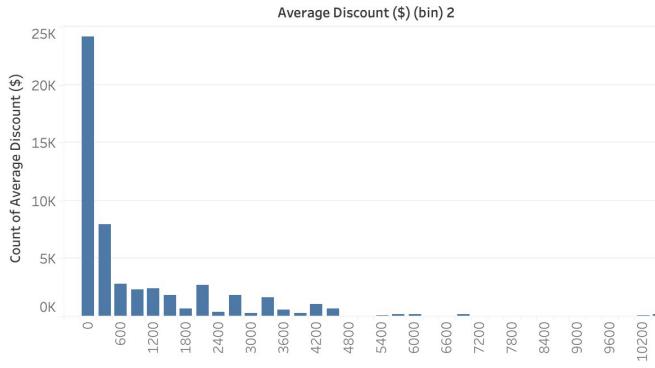


Univariate Visualization: Measures

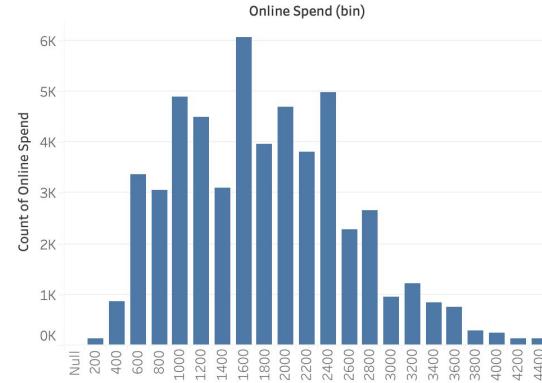
Avg Price



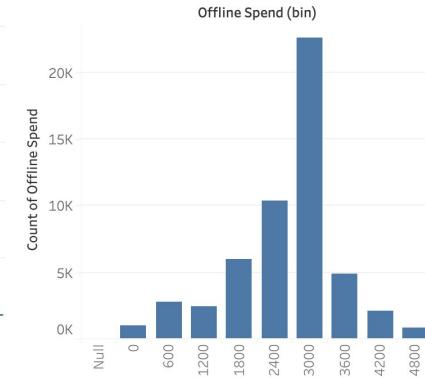
Avg Discount



Online Spend

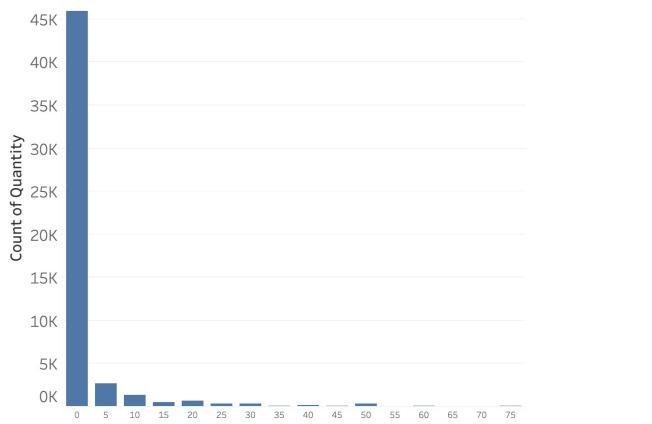


Offline Spend



- Avg price, sales quantity, and avg discount is heavily skewed to the right
- Online spend and offline spend is mostly normally distributed

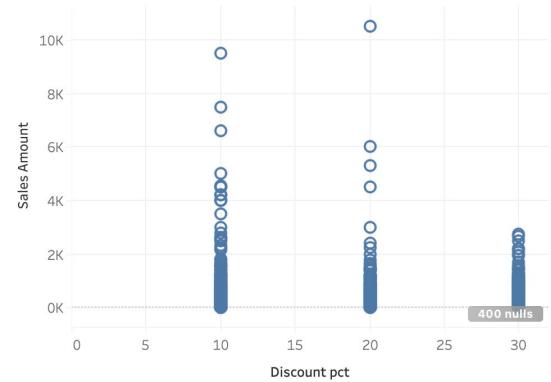
Quantity



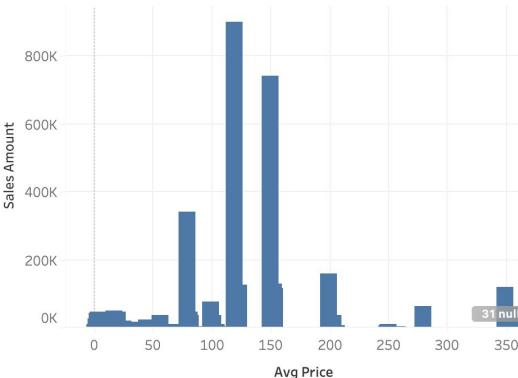


Bivariate Visualizations: Measure

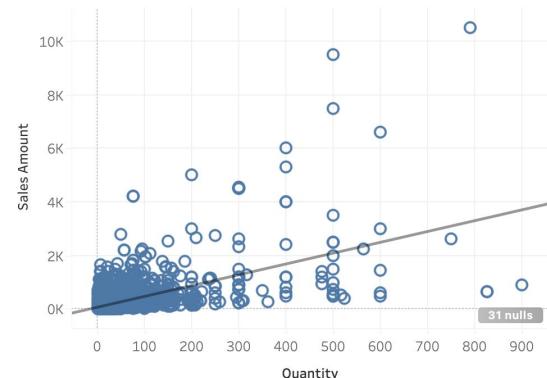
Sales Amount vs Discount Pct



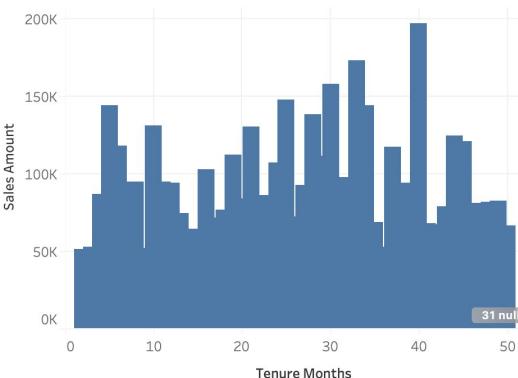
Sales Amount vs Avg Price



Sales Amount vs Quantity



Sales Amount vs Tenure Months

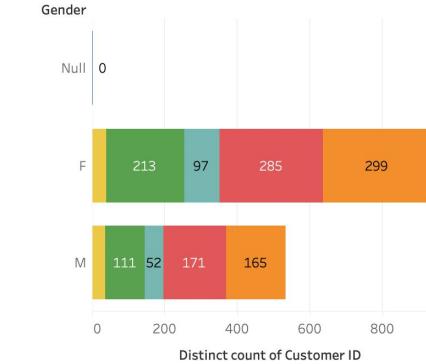


- Explores the relationship between the target variable (the profit) and the various measures of interest.
- As discount percentage increases, there seems to be a generally downward trend with profit
- Quantity and profit has a positive relationship, the more quantity sold, the more profit the company makes.
- Tenure month seems to be evenly distributed, while average price seems to be mostly normally distributed with a little skew to the right



Multivariate Analysis Visualizations

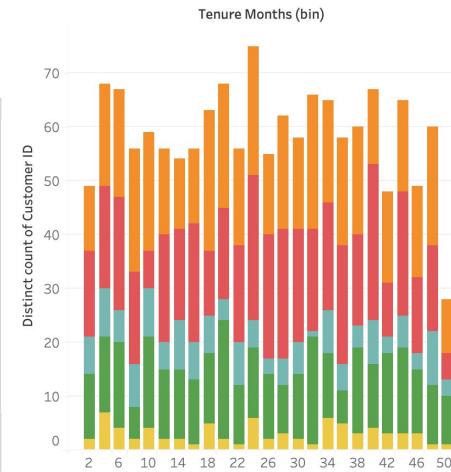
Distribution of Demographics



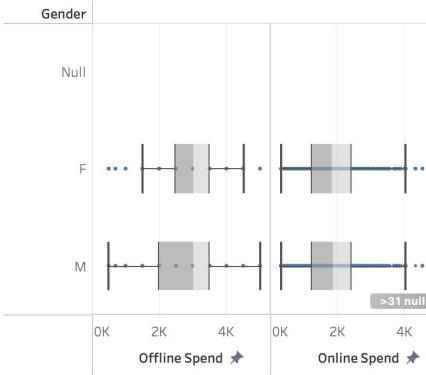
Count of Purchases by Location



Customer Tenure with the Platform



Distribution of Spending by Gender





Contingency table of dimensions

Gender/Product category

Product Category	Gender		Coupon Sta..	Gender		Month of M..	Gender	
	F	M		F	M		F	M
Accessories	143	91	Clicked	16,788	10,138	January	2,050	2,013
Android	26	17	Not Used	5,067	3,027	February	2,196	1,088
Apparel	11,355	6,771	Used	11,152	6,752	March	2,783	1,563
Backpacks	46	43				April	2,788	1,362
Bags	1,182	700				May	3,043	1,529
Bottles	174	94				June	2,675	1,518
Drinkware	2,181	1,302				July	3,397	1,854
Fun	99	61				August	3,741	2,409
Gift Cards	119	40				September	2,298	1,990
Google	63	42				October	2,628	1,536
Headgear	504	267				November	2,406	1,555
Housewares	73	49				December	3,002	1,500
Lifestyle	1,908	1,184						
More Bags	32	14						
Nest	1,326	872						
Nest-Canada	197	120						
Nest-USA	8,642	5,371						
Notebooks & Journals	432	317						
Office	4,178	2,335						
Waze	327	227						

Coupon status/Location

Location	Coupon Status		
	Clicked	Not Used	Used
California	8,325	2,391	5,420
Chicago	9,276	2,888	6,216
New Jersey	2,303	712	1,488
New York	5,640	1,679	3,854
Washington DC	1,382	424	926

Gender/Location

Location	Gender	
	F	M
California	9,875	6,261
Chicago	11,484	6,896
New Jersey	3,067	1,436
New York	7,315	3,858
Washington DC	1,266	1,466

Gender/Month

Month of M..	Gender		
	F	M	
January	2,050	2,013	
February	2,196	1,088	
March	2,783	1,563	
April	2,788	1,362	
May	3,043	1,529	
June	2,675	1,518	
July	3,397	1,854	
August	3,741	2,409	
September	2,298	1,990	
October	2,628	1,536	
November	2,406	1,555	
December	3,002	1,500	

Coupon Status/Quarter

Quarter of ..	Coupon Status		
	Clicked	Not Used	Used
Q1	5,933	1,803	3,957
Q2	6,596	1,978	4,341
Q3	7,957	2,392	5,340
Q4	6,440	1,921	4,266

Location/Weekday

Weekday of Month	Location				
	Californ...	Chicago	New Jersey	New York	Washington DC
Sunday	2,993	3,186	749	2,000	473
Monday	2,692	2,439	722	1,796	578
Tuesday	1,430	1,548	355	1,152	87
Wednesday	1,607	2,272	453	1,478	340
Thursday	3,245	4,322	935	2,394	695
Friday	1,268	1,364	509	879	173
Saturday	2,901	3,249	780	1,474	386



Contingency table of dimensions

Coupon status/Product Category Coupon status/Weekday Location/Product Category

Product Category	Coupon Status			Weekday of..	Coupon Status			Product Category	Location				
	Clicked	Not Used	Used		Clicked	Not Used	Used		Califor..	Chicago	New Jersey	New York	Washington DC
Accessories	125	32	77	Sunday	4,296	1,349	2,874	Accessories	80	86	17	41	10
Android	23	10	10	Monday	2,225	719	1,520	Android	13	18	4	7	1
Apparel	9,223	2,747	6,156	Tuesday	2,452	625	1,534	Apparel	5,491	6,158	1,627	3,902	948
Backpacks	47	15	27	Wednesday	4,516	1,358	3,013	Backpacks	29	34	8	14	4
Bags	972	285	625	Thursday	4,642	1,367	2,991	Bags	539	731	151	383	78
Bottles	138	48	82	Friday	4,620	1,464	3,182	Bottles	91	85	27	54	11
Drinkware	1,752	570	1,161	Saturday	4,175	1,212	2,790	Drinkware	1,117	1,252	284	675	155
Fun	83	25	52					Fun	44	56	14	32	14
Gift Cards	81	22	56					Gift Cards	27	89	18	24	1
Google	51	25	29					Google	40	33	6	22	4
Headgear	401	114	256					Headgear	221	264	62	191	33
Housewares	68	18	36					Housewares	38	50	4	23	7
Lifestyle	1,529	462	1,101					Lifestyle	977	1,086	251	624	154
More Bags	25	3	18					More Bags	15	17	5	8	1
Nest	1,127	351	720					Nest	762	710	188	421	117
Nest-Canada	175	47	95					Nest-Canada	91	120	23	63	20
Nest-USA	7,141	2,160	4,712					Nest-USA	4,184	4,855	1,203	2,975	796
Notebooks & Journ..	389	103	257					Notebooks ..	238	260	41	181	29
Office	3,295	968	2,250					Office	1,993	2,273	522	1,409	316
Waze	281	89	184					Waze	146	203	48	124	33

Gender/Weekday

Weekday of..	Gender		Quarter of ..	Coupon Status			Quarter of Transaction..	Location				
	F	M		Clicked	Not Used	Used		Califor..	Chicago	New Jersey	New York	Washington DC
Sunday	5,400	3,119	Q1	5,933	1,803	3,957	Q1	3,677	4,020	866	2,314	816
Monday	2,760	1,704	Q2	6,596	1,978	4,341	Q2	4,078	4,271	1,224	2,877	465
Tuesday	2,464	2,147	Q3	7,957	2,392	5,340	Q3	4,576	5,702	1,177	3,413	821
Wednesday	5,531	3,356	Q4	6,440	1,921	4,266	Q4	3,805	4,387	1,236	2,569	630
Thursday	5,734	3,266										
Friday	5,960	3,306										
Saturday	5,158	3,019										



Contingency table of dimensions

Location/Month					Quarter/Weekday															
Month of Transaction..	Location				Washington DC	Weekday of..	Transaction Date				Q1	Q2								
	California	Chicago	New Jersey	New York			Q1	Q2	Q3	Q4										
January	1,491	1,058	315	845	354	Sunday	1,818	2,078	2,700	1,923										
February	958	1,368	263	460	235	Monday	962	977	1,395	1,130										
March	1,228	1,594	288	1,009	227	Tuesday	1,101	1,097	1,179	1,234										
April	1,380	1,359	360	846	205															
May	1,430	1,548	355	1,152	87															
June	1,268	1,364	509	879	173	Wednesday	2,010	2,167	2,454	2,256										
July	1,613	1,827	389	1,154	268															
August	1,607	2,272	453	1,478	340	Thursday	2,093	2,372	2,484	2,051										
September	1,356	1,603	335	781	213	Friday	1,935	2,340	2,877	2,114										
October	1,201	1,381	407	951	224															
November	1,059	1,360	384	925	233	Saturday	1,774	1,884	2,600	1,919										
December	1,545	1,646	445	693	173															
Product Category/Quarter																				
Quarter of Transaction..	Product Category																			
	Acce..	Andr..	Appa..	Back..	Bags	Bottl..	Drin..	Fun	Gift Cards	Goog..	Head..	Hous..								
Q1	2	12	3,302	16	489	59	980	42	16	53	186	32	644	24	89	3,717	162	1,733	135	
Q2	17	16	5,104	36	456	68	793	50	42	34	201	32	784	22	89	3,038	183	1,824	126	
Q3	27	15	6,497	25	595	119	1,081	65	23	18	289	58	1,152		377	66	3,191	392	1,575	124
Q4	188		3,223	12	342	22	629	3	78		95		512		1,821	73	4,067	12	1,381	169



Contingency table of dimensions

Product Category/Month

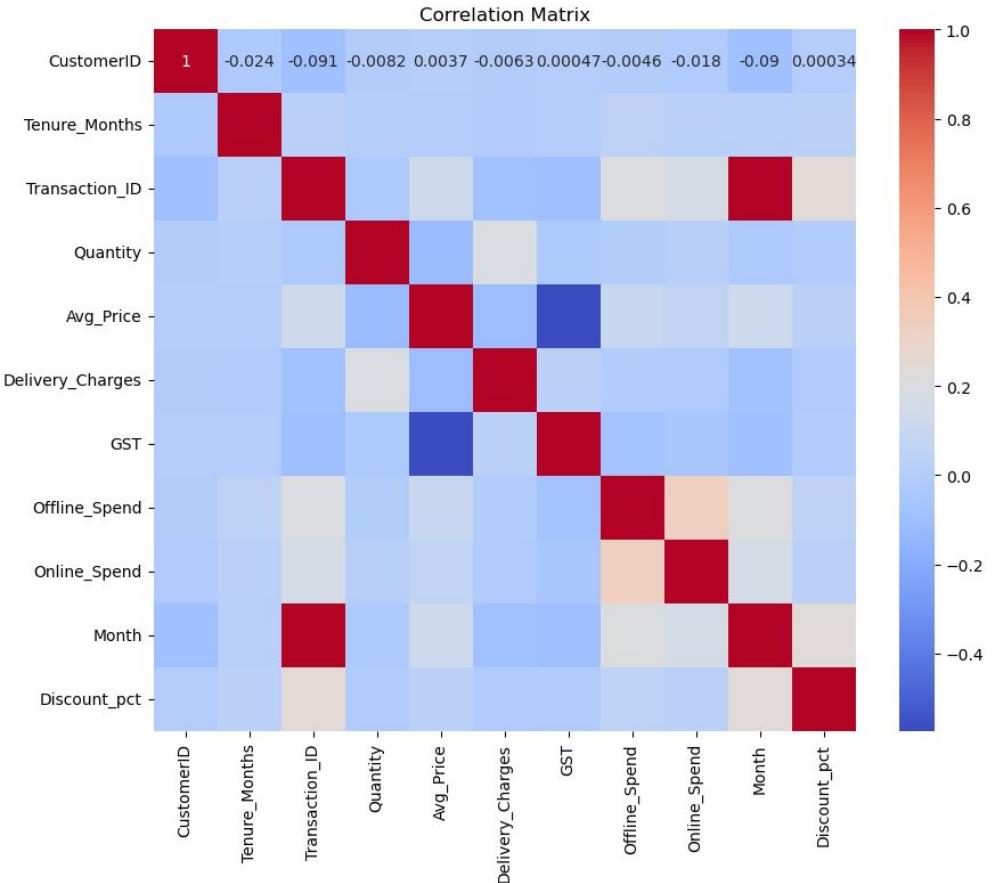
Month of Transaction..	Product Category																		
	Acce..	Andr..	Appa..	Back..	Bags	Bottl..	Drin..	Fun	Gift Cards	Goog..	Head..	Hous..	Lifes..	More Bags	Nest	Nest..	Noteb ooks..	Office	Waze
January	1	4	960	3	155	16	295	11	4	21	54	10	200	5	39	1,563	54	607	61
February		1	905	6	139	19	284	14	6	15	66	5	199	4	22	1,055	51	468	25
March	1	7	1,437	7	195	24	401	17	6	17	66	17	245	15	28	1,099	57	658	49
April	4	2	1,582	11	145	20	267	18	7	10	72	10	252	13	27	996	73	606	35
May	6	5	1,876	21	163	26	287	15	25	11	65	12	262	9	21	970	57	685	56
June	7	9	1,646	4	148	22	239	17	10	13	64	10	270		41	1,072	53	533	35
July	6	8	2,182	8	216	40	298	35	10	12	119	22	383		27	1,161	150	549	25
August	8	7	2,820	11	201	38	462	21	10	6	112	33	429	86	24	1,088	191	541	62
September	13		1,495	6	178	41	321	9	3		58	3	340	291	15	942	51	485	37
October	49		1,114	8	138	22	305	3	68		22		317	455	23	1,106	3	488	43
November	83		772	4	115		164				24		124	654	24	1,427		508	62
December	56		1,337		89		160		10		49		71	712	26	1,534	9	385	64

Product Category/Weekday

Weekday of Transaction..	Product Category																			
	Acce..	Andr..	Appa..	Back..	Bags	Bottl..	Drin..	Fun	Gift Cards	Goog..	Head..	Hous..	Lifes..	More Bags	Nest	Nest..	Noteb ooks..	Office	Waze	
Sunday	43	10	3,144	12	338	40	571	32	16	22	146	18	518	5	328	49	1,970	134	1,047	76
Monday	21	4	1,540	11	114	28	218	16	5	7	68	7	246	1	209	22	1,349	39	504	55
Tuesday	11	2	1,476	11	163	12	228	16	9	5	66	9	224	1	215	25	1,533	41	517	47
Wednesday	57	7	2,793	15	278	33	598	14	11	17	121	26	503	11	421	70	2,636	110	1,066	100
Thursday	42	3	3,088	16	324	52	581	25	34	23	138	16	494	12	369	49	2,353	149	1,118	114
Friday	25	11	3,281	6	357	51	643	32	53	18	122	23	565	8	354	61	2,195	141	1,237	83
Saturday	35	6	2,804	18	308	52	644	25	31	13	110	23	542	8	302	41	1,977	135	1,024	79



Correlation matrix of measures



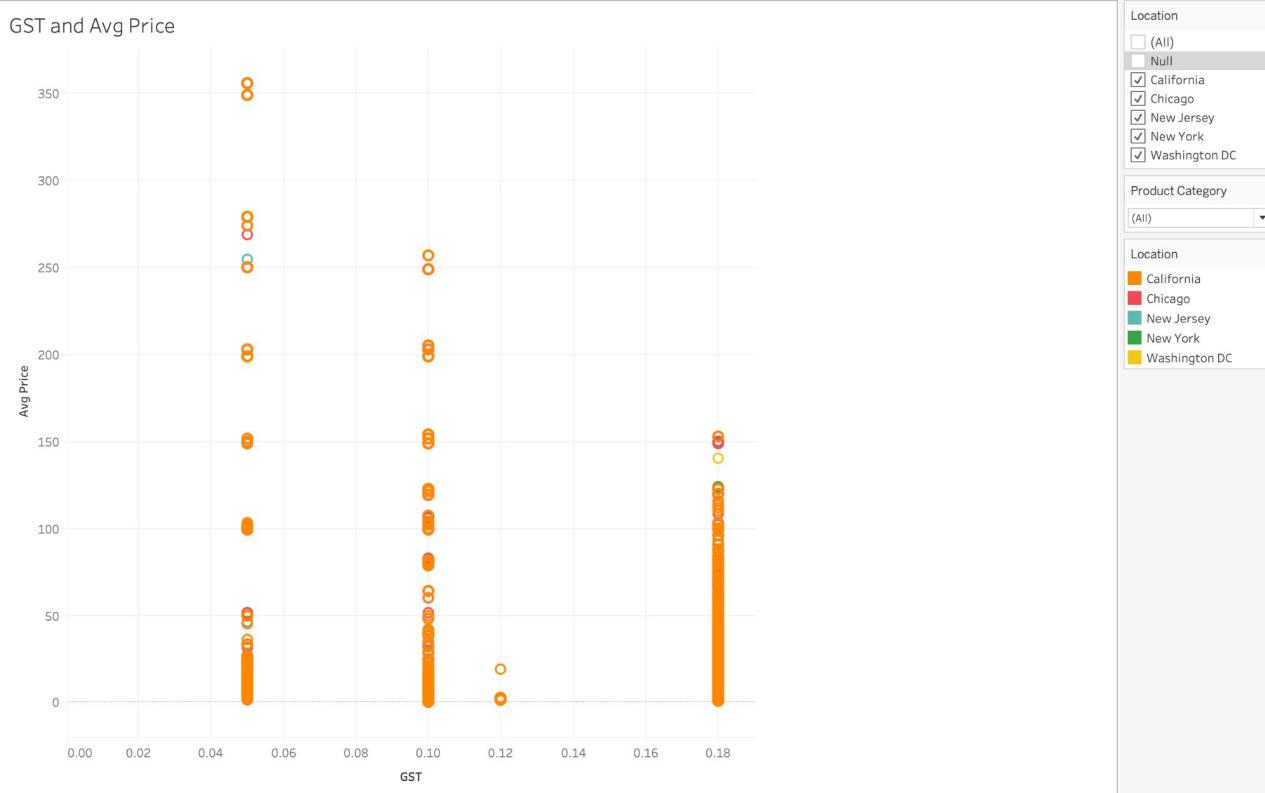
The data seems to lack correlation among the measures, there are only two notable combinations, month and transaction ID, and GST and average price.

- Month/Transaction ID: The transaction ID is given chronologically according to the time of sales, and since this dataset only contains data for 2019, there is a strong positive correlation between month and transaction ID
- GST/Average Price: there is generally a negative correlation between goods and services tax and average price, as explored in the visualization on the next page



Avg Price vs GST

GST and Avg Price



Again, generally there is a downward trend for the different locations. This seems anti-intuitive given that GST is applied as a percentage price, and is hard to analyze because we don't know everything about this dataset, but it might be due to different products being taxed differently



Dashboard Design Document



Target Audience and Purpose

Target Audience

The target audience of this report is the BD (Business Development) teams within the company.

Purpose

The purpose of this dashboard is to improve sales and revenue for the company by generating insightful analysis on product sales information. This dashboard would be beneficial for various BD teams such as the Public Relations Team and the Marketing Team.

- By analyzing the most popular products, the PR Team could post these insights on social media for community building.
- The Marketing team would be able to evaluate campaign performances based on how each promotion affect sales trends.



Dashboard Type: Analytical

- We will make EDA reports and relevant analysis on market trends, customers' behaviors, time series analysis, etc.. based on our dataset. Our analysis will focus on products sold by the retailer such as Nest products from Google. Some related graphs and table calculations could be monthly price trend for the product, sales revenue by area, customer's gender ratio, how promotions affect customer behavior, etc.
- The main goal of delivering analytical dashboard is to better present our findings from the online shopping data to improve business strategies, meanwhile to provide insights for the stakeholders, collaborators and even the whole industry.

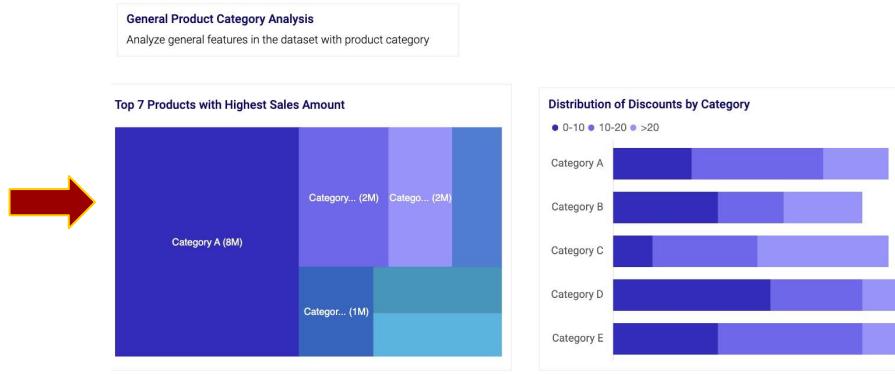


- Product Analysis Pages (for management)
 - Aggregate with product category
 - Sales amount by product category
 - Distribution of discounts by category
 - Sales quantity by time and category
 - Average price distribution by category
 - ...
- Customer Analysis Pages (for promotion)
 - Demographics
 - Aggregate with customer location
 - Sales amount by location
 - Amount spent by customer
 - Number of orders by tenure
 - Sales trend by time
 - ...



Product Analysis: Layout

Selecting any category filters the charts to show the related status.

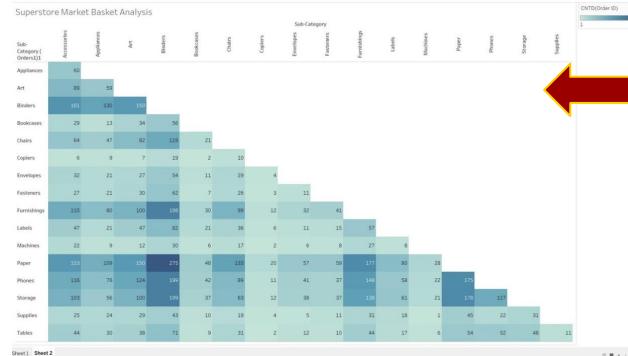


Helps to navigate between stories



Which category has the highest overall prices and does it align with its sales amount

Market Basket Analysis



Example market basket analysis graph: helps the retailers to see how each products could be managed together, i.e., if most people purchased Coke also purchased cups, then they would be discounted together.



See specific category and specific time

Sales trend for each specific product and forecasting

Helps to navigate between stories

See which season has the highest sales amount for specific product

Show which product has the highest sales amount each month

See if the relationship between discount and sales is uniform overtime

Time Series Analysis with Product Category
Analyze the related changes of the sales amount by product category due to time variance

Month (2019)
All

Product Category
All

Sales Amount Over the Seasons

Overall Sales Trend

Total Sales Amount
25.2K
vs prev 11.6K

Product Category Sales
25.2K **11.6K** +10%
Current Previous % change

Monthly Sale Analysis
Title 1 Metric 1 Title 2 Metric 1

Relationships with Discount and Sales Amount (over Time)

Sales Quantity
25.2K
Previous 11.6K (+10%)
Current Previous

Product Category Analysis
Product Category Analysis with Time
Customer Demographics
Customer Sales Amount Analysis

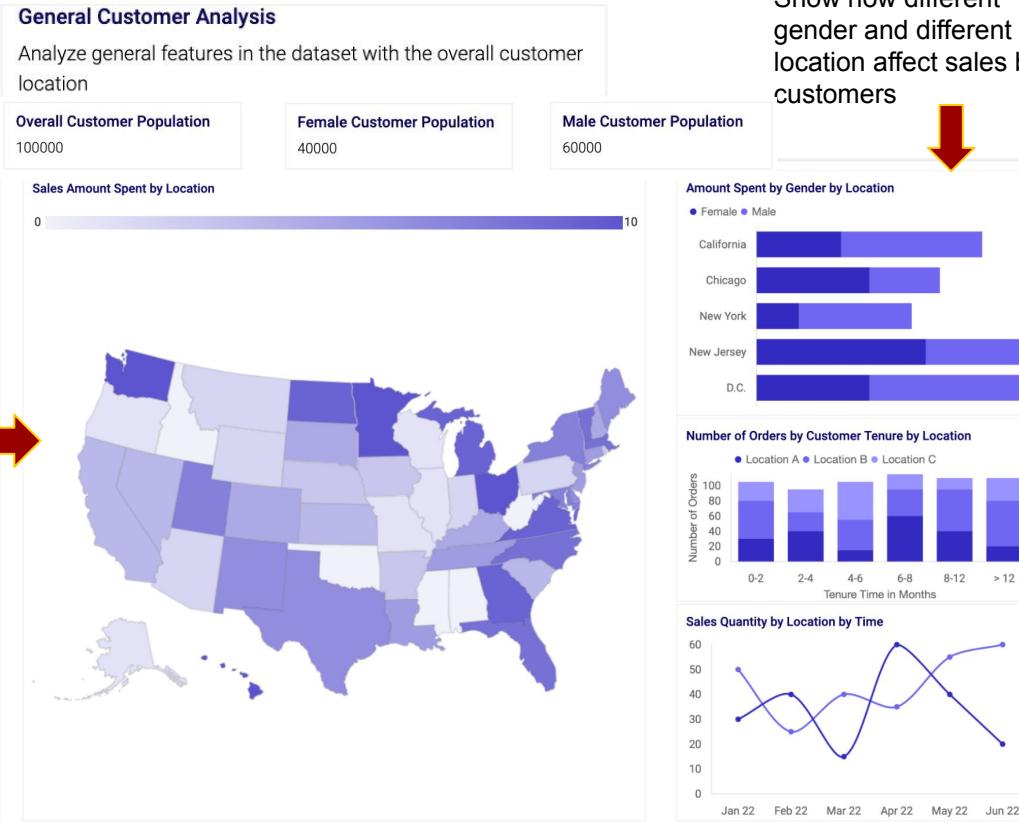


- > Which product has the highest sales amount (over time)?
- > Product management:
 - » Which product should we increase production and promotion
 - » Substitutes & complements
- > How can the management team change plans varied by month/season?
- > Discounts varied by product and by time

Customer Analysis: Layout



General demographies of the customers



Show how different gender and different location affect sales by customers

Selecting any location helps to filter the charts

Product Category Analysis

Product Category Analysis with Time

Customer Demographics

Customer Sales Amount Analysis

Sales Amount (K)

Time

Helps to navigate between stories

Drill down the results

Show results on customer loyalty

See how customer in different region varies with their sales behavior



Customer Analysis: Map Section

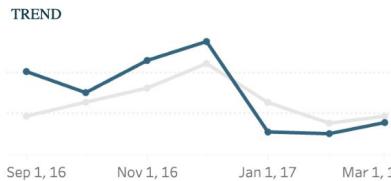
START DATE
9/1/2016

END DATE
3/31/2017

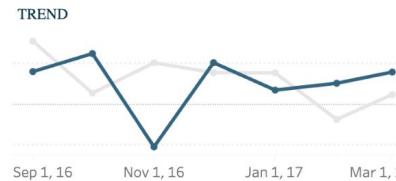
SELECT REGION
East

DATE COMPARISON
Prior Period

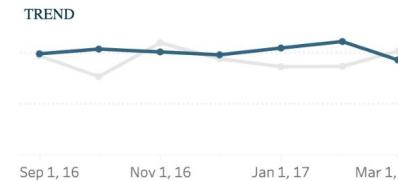
DATE GRANULARITY
Month



REGION COMPARISON



REGION COMPARISON



REGION COMPARISON





- > Who are our customers?
- > Customer Loyalty:
 - >> What ‘category’ of customers spend the most
 - >> Customer tenure
- > Which customers are most sensitive to promotions?
- > Locations where customers prefer online shopping to offline shopping or vice versa



1. Data source:
<https://www.kaggle.com/datasets/jacksondivakarr/online-shopping-dataset>
2. <https://www.linkedin.com/pulse/perform-market-basket-analysis-tableau-anthony-b-smoak/>
3. <https://www.quanthub.com/what-is-a-box-plot/>