

Customer Demographics and Spending Behavior Analysis for Optimized Marketing Strategies

Business Problem:

The objective of this SQL project is to conduct a comprehensive cluster analysis of customer demographics and purchasing behavior for a specific company. The primary goal is to gain deep insights into the company's ideal customers, enabling the business to better understand their specific needs, behaviors, and concerns. By analyzing different customer segments, the company can tailor its products and marketing strategies accordingly. The project aims to identify the most dominant customer segments in terms of numbers and spending habits, optimizing marketing efforts and maximizing sales potential.

Research Questions:

1. What are the customer segments that represent the majority of the company's customers in terms of numbers and spending habits?
2. What are the most popular products for each of these customer segments?
3. What are the preferred shopping mediums (web, catalog, or store) for the customers of each of these segments?
4. What is the most successful campaign that returned the highest offer acceptance rate among customers of each segment?
5. What is the complaint rate of the customers in these segments?

Processing the Data:

To begin the analysis, the dataset was downloaded from Kaggle.com and loaded into MS SQL Server for cleaning and analysis. All work on this dataset, for this project, will be done in MS SQL Server. The full SQL code of this project can be found through this link:

<https://bit.ly/42zgOkA>

After checking the credibility of the dataset source, downloading the dataset and choosing the work medium, the data processing also involved updating column names to reflect their content more clearly, rearranging education and marital status columns and organizing the overall dataset into a format suitable for analysis.

Data Cleaning: (Project Full SQL Code Link: <https://bit.ly/42zgOkA>)

Several data cleaning steps were performed on the dataset. Null values in the income column were deleted. Duplicate values in the customer ID column were checked and removed. Outliers in the income and age columns were identified and deleted. The SQL code that I wrote to clean the data can be found in the project code file, whose link is provided in the title above.

Data Analysis: (Project Full SQL Code Link: <https://bit.ly/42zgOkA>)

The analysis phase began by adding columns to the dataset, such as age, age category, age group, economic class, parental status, total spending for the last 2 years, and total number of purchases.

The next analysis step consisted of examining the correlation between customer education level and their overall average spending in the past two years. The results shown in Table 1, showed that customers with the highest overall average spending are high-income, Master's degree holders, with a 2-year average of \$274.50 per customer, followed by high-income, undergrad degree holders, with a 2-year average of \$255.00 per customer, followed by high-income, PhD holders, with a 2-year average of \$240.17 per customer. These are highlighted by the green rectangle in Table 1. However, the majority of the company's customers turned out to be middle-class, undergrad degree holders as shown in the red rectangle in Table 1, being 632 customers out of the 2,166 total customers of the company.

	education	Average_Income	income_class	Avg_Wine_Spending	Avg_Fruits_Spending	Avg_Meat_Spending	Avg_Fish_Spending	Avg_Gold_Spending	Avg_Sweets_Spending	Overall_Avg_Spending	Total_Customers
1	Masters	88083	High	816	74	533	87	62	75	274.50	37
2	Undergrad	87177	High	643	74	543	103	89	78	255.00	104
3	PhD	91686	High	745	49	486	63	39	59	240.17	52
4	High School	83273	High	330	32	433	119	39	81	229.33	17
5	Undergrad	60086	Middle	374	38	206	52	63	38	128.50	632
6	PhD	59362	Middle	485	21	171	28	38	20	127.17	309
7	High School	59124	Middle	295	41	164	66	69	48	113.83	96
8	Masters	58019	Middle	395	20	165	33	45	19	112.83	220
9	Masters	30800	Low	39	5	27	6	18	4	16.50	104
10	Undergrad	28966	Low	28	5	24	9	17	6	14.83	355
11	PhD	32664	Low	48	2	23	4	9	2	14.67	113
12	High School	26713	Low	14	10	21	13	18	8	14.00	79
13	Primary	21643	Low	7	10	11	18	22	12	13.33	48

Table1. Correlation between customer's education level, income class and their overall average spending for the past two years

The correlation between customer marital status, parental status, and their overall average spending in the past two years was also analyzed. Customers without children had the highest overall spending averages, with singles having the highest spending among different marital status groups, with a 2-year average of \$1585 per customer.

On the other hand, when it comes to majority by numbers, middle-class, married customers, having at least one kid, represented the company's top customers in this segment, representing a total of 623 customers out of the overall 2,166 customers. Both of these customer categories -- top spenders and majority by numbers -- are highlighted in green and red respectively, in Table 2, shown below.

	marital_status	have_kids	Average_Income	income_class	Total_Customers	Avg_Wine_Spending	Avg_Fruits_Spending	Avg_Meat_Spending	Avg_Fish_Spending	Avg_Gold_Spending	Avg_Sweets_Spending	Overall_Avg_Spending_2Y
1	Single	No	87738	High	50	737	63	553	87	58	84	1585
2	Married	No	87470	High	108	700	71	548	94	74	71	1560
3	Widow	Yes	80995	High	1	860	28	409	73	56	56	1482
4	Single	Yes	86245	High	6	635	63	463	123	77	54	1417
5	Divorced	Yes	85111	High	3	991	41	182	81	51	62	1411
6	Married	Yes	97302	High	18	742	39	404	83	59	50	1379
7	Widow	No	83441	High	5	421	102	567	96	62	104	1355
8	Widow	No	69258	Middle	16	642	51	395	93	86	55	1324
9	Divorced	No	88771	High	19	508	71	475	88	85	80	1309
10	Single	No	66105	Middle	79	547	57	402	74	72	46	1201
11	Married	No	68450	Middle	199	528	54	392	87	70	57	1191
12	Divorced	No	65073	Middle	29	547	58	330	89	66	49	1141
13	Divorced	Yes	57578	Middle	114	397	26	128	29	54	24	661
14	Married	Yes	57099	Middle	623	358	22	119	30	47	24	603
15	Widow	Yes	55409	Middle	42	340	22	93	34	49	28	569
16	Single	Yes	54546	Middle	155	284	20	104	30	46	18	505
17	Widow	No	27618	Low	5	34	22	99	53	66	38	314
18	Married	No	25779	Low	67	34	15	34	27	29	13	155
19	Single	No	25348	Low	25	29	12	33	18	22	13	130
20	Single	Yes	29770	Low	138	36	5	25	7	16	5	95
21	Married	Yes	29661	Low	394	29	4	20	6	14	4	80
22	Divorced	No	22115	Low	8	36	2	15	2	10	3	71
23	Divorced	Yes	29723	Low	55	21	2	16	4	11	3	60
24	Widow	Yes	31567	Low	7	27	0	12	3	8	1	55

Table 2. Correlation between customer's marital status, parental status, income class and their overall average spending in the past two years

The distribution of customers by age was analyzed to determine the age group and category that represented the majority of customers. The results showed that the majority of customers were adults between the ages of 30-55, representing 56.93% of the total customers. However, the highest overall 2-year average spending were held by customers aging 55 years and up, with an average of \$696 per 2-year. Results are shown below in Table 3.

	age_group	age_category	Avg_Total_Spending_2Y	Total_Customers	Customer_Percentage
1	30 - 55	ADULT	546	1233	56.93
2	55+	OLD	696	933	43.07

Table 3. Distribution of Company's Customers by Age

Based on the analysis, two customer groups were identified: Customer Group 1 (Majority by Numbers) and Customer Group 2 (Majority by Spending or Top Spenders). Customer Group 1 consisted of customers who are middle-class, married, had at least one child, are undergraduate degree holders, and are adults between the ages of 30 and 55. Customer Group 2 consisted of customers who were high-income class, are single or married, have no children, being at least 30, and hold an undergraduate degree or higher.

The next step in the analysis was to determine the most popular product(s), most preferred shopping medium, campaign with the highest acceptance rate and complaint rate across the two customer segments.

For the most popular product(s), customers of Group 1 were big fans of the company's wine, making it the most popular product by far for this customer segment, with an overall total spending of \$52,773 in the past two years on wine alone, making a 51.78% of the total spending on products for customers of Group 1. These results can be seen below in Table 4.

	income_class	marital_status	education	children	age_category	Total_Wine_Spending	Wine_Spending_Percentage	Total_Fruit_Spending	Fruit_Spending_Percentage	Total_Meat_Spending	Meat_Spending_Percentage	Total_Fish_Spending	
1	Middle	Married	Undergrad	Yes	ADULT	52773	51.78	4855	4.76	23154	22.72	6388	
								Fish_Spending_Percentage	Total_Sweets_Spending	Sweets_Spending_Percentage	Total_Gold_Spending	Gold_Spending_Percentage	Total_2Y_Spending
								6.27	5638	5.53	9118	8.95	101926

Table 4. Group 1 Customers' Spending on Different Products for the Last 2 Years

Wine has also proven to be Group 2 customers' favorite product, representing the highest average percentage of 48.38% of the overall spending for these customers on all products for the past two years, as it can be seen in Table 5.

The company should, therefore, tailor its products line and campaign to better and further promote its wine products, since it is by far the favorite product for the customers of both segments.

	income_class	education	marital_status	children	Total_Wine_Spending	Wine_Spending_Percentage	Total_Fruit_Spending	Fruit_Spending_Percentage	Total_Meat_Spending	Meat_Spending_Percentage	Total_Fish_Spending
1	High	Masters	Married	No	17143	51.51	1986	5.97	9395	28.23	1854
2	High	PhD	Married	No	20475	51.86	1117	2.83	13858	35.10	1528
3	High	Undergrad	Married	No	32691	39.84	4014	4.89	30780	37.51	5553
4	High	Masters	Single	No	10242	48.70	558	2.65	7926	37.69	919
5	High	PhD	Single	No	9597	57.05	627	3.73	4509	26.80	725
6	High	Undergrad	Single	No	14934	41.32	1832	5.07	13176	36.46	2407
					Fish_Spending_Percentage	Total_Sweets_Spending	Sweets_Spending_Percentage	Total_Gold_Spending	Gold_Spending_Percentage	Total_2Y_Spending	
					5.57	1580	4.75	1322	3.97	33280	
					3.87	1521	3.85	985	2.49	39484	
					6.77	3911	4.77	5100	6.22	82049	
					4.37	869	4.13	518	2.46	21032	
					4.31	961	5.71	403	2.40	16822	
					6.66	1981	5.48	1809	5.01	36139	

Table 5. Group 2 Customers' Spending on Different Products for the Last 2 Years

The analysis for the most preferred shopping medium for customers of both segments showed that the Store is the most preferred medium for these customers based on the percentage of total purchases that the customers made across all three shopping mediums that the company possesses (Web, Catalog and Store). This can be seen in both Tables 6 and 7.

Store purchases of customers of Group 1 consisted a 46.02% of all purchases, while store purchases of customers of Group 2 consisted a 42.72% of all purchases.

	income_class	marital_status	education	children	age_category	Total_Web_Purchases	Percentage_Web_Purchases	Total_Catalog_Purchases	Percentage_Catalog_Purchases	Total_Store_Purchases	Percentage_Store_Purchases	Total
1	Middle	Married	Undergrad	Yes	ADULT	859	35.78	437	18.20	1105	46.02	2401

Table 6. Purchase Total and Percentages for each Shopping Medium for Group 1 Customers

	Web_Purchases_Perc_Avg	Catalog_Purchases_Perc_Avg	Store_Purchases_Perc_Avg
1	26.25	31.03	42.72

Table 7. Purchase Total and Percentages for each Shopping Medium for Group 1 Customers

Analyzing the acceptance (success) rate of each campaign made by the company, it is concluded that, for customers of Group 1, the campaign that resonated the most with them was campaign 4, with an acceptance rate of 7.45%, as shown in Table 8.

For customers of Group 2, the most successful campaign with the highest acceptance was campaign 5, with a percentage of 50.79% of all customers responding positively to it. These results can be seen in detail in Table 9.

	income_class	marital_status	education	children	age_category	Cmp1_Acceptance_Rate	Cmp2_Acceptance_Rate	Cmp3_Acceptance_Rate	Cmp4_Acceptance_Rate	Cmp5_Acceptance_Rate	Cmp6_Acceptance_Rate
1	Middle	Married	Undergrad	Yes	ADULT	1.24	1.24	6.83	7.45	3.73	4.97

Table 8. Campaign Acceptance Rate for Group 1 Customers

	Cmp1_Acc_Rate_Avg	Cmp2_Acc_Rate_Avg	Cmp3_Acc_Rate_Avg	Cmp4_Acc_Rate_Avg	Cmp5_Acc_Rate_Avg	Cmp6_Acc_Rate_Avg
1	30.83	5.65	8.42	20.82	50.79	44.18

Table 9. Campaign Acceptance Rate for Group 2 Customers

The analysis, of the provided company customers' dataset, that was done to determine the complaint rate of each segments' customers showed that both customer groups did not express any complaint or dissatisfaction with the company's products and services. This was concluded after the calculated complaint rates for each group turned out to be zero, as shown in Table 10. This means that the majority of company's customers, both by numbers and spending power, are satisfied with what they are presented with in terms of products and service quality from this company.

	income_class	marital_status	education	children	age_category	Number_Of_Complaints	Complaint_Rate
1	Middle	Married	Undergrad	Yes	ADULT	0	0.000

	income_class	education	marital_status	children	Number_Of_Complaints	Complaint_Rate
1	High	Masters	Married	No	0	0.000
2	High	PhD	Married	No	0	0.000
3	High	Undergrad	Married	No	0	0.000
4	High	Masters	Single	No	0	0.000
5	High	PhD	Single	No	0	0.000
6	High	Undergrad	Single	No	0	0.000

Table 10. Complaint Rate for Group 1 Customers (Up) and Group 2 Customers (Down)

Research Insights:

1. Customer Segments: In Group 1, the majority of customers belong to the middle class, are married, have at least one child, hold undergraduate degrees, and fall within the age range of 30 to 55. On the other hand, in Group 2, the top spending customers with the highest spending in the past two years are primarily individuals who are either single or married without children, hold undergraduate degrees or higher (such as Masters or PhD), belong to a high-income class, and are aged 30 and above. These findings provide a clear understanding of the demographic characteristics that differentiate the two customer groups, which can guide targeted marketing strategies and product offerings to cater to their specific needs and preferences.

2. Product Popularity: Customers in Group 1 have shown a strong preference for the company's wine, spending a significant amount of \$52,773 on wine alone in the past two years. This accounts for 51.78% of their total spending on products. Similarly, customers in Group 2 also favor wine, with an average of 48.38% of their overall spending allocated to this product. These findings highlight the importance of the company focusing on promoting and customizing its wine offerings in product lines and marketing campaigns, as it is the top choice for customers in both segments.

3. Shopping Preferences: Both Group 1 and Group 2 customers show a strong preference for shopping at the physical store compared to other shopping mediums. In Group 1, store purchases account for 46.02% of all purchase operations across different mediums. Similarly, in Group 2, the store is the preferred shopping medium, representing 42.72% of all purchase operations. These findings emphasize the significance of the physical store as the primary channel for both customer groups, highlighting the importance of maintaining and optimizing the in-store experience to cater to their preferences and maximize sales opportunities.

4. Campaign Success: Among customers in Group 1, Campaign 4 stood out as the most successful campaign, achieving an acceptance rate of 7.45%. On the other hand, among customers in Group 2, Campaign 5 emerged as the most successful, boasting the highest acceptance rate of 50.79%. These findings highlight the varying effectiveness of different campaigns across the two customer groups, indicating the need to tailor marketing strategies and messaging to better resonate with the preferences and interests of each group.

5. Complaint Rate: The analysis of the dataset reveals that both customer groups, representing the majority of the company's customers and top spenders, have recorded a complaint rate of zero percent. This indicates a high level of satisfaction among these customers with the company's services. It is encouraging to see that the efforts put into meeting customer needs and delivering quality service have yielded positive results, as reflected in the absence of complaints from these valuable customer segments.

Conclusion:

In conclusion, this SQL analysis provided valuable insights into the customer demographics and spending habits of the company. By identifying the most dominant customer segments and their preferences, the company can tailor its products and marketing strategies to optimize sales potential. The analysis highlighted the importance of understanding customer characteristics and behavior for effective targeting and personalized marketing. The findings can guide the company in developing strategies to attract and retain customers, maximize offer acceptance rates, and address customer complaints effectively.