Data analysis of customer satisfaction & demographics in purchases at Amazon in UAE

Executive Summary

To determine whether a customer will spend more than 1000 AED per purchase or not for the Amazon/souq.com E-Commerce, specific demographics and customer satisfaction factors needed to be analyzed. Explanatory variables such as marital status, monthly household income, number of times purchased in the last 6 months, most common platform to buy on the website, and physical or online store purchase would best determine a customer's spending habits. Based on a logistic regression of these variables, the model predicts that most customers spend less than 1000 AED per purchase with an accuracy of 88 % and a precision of 26 %. This means that the model was good in predicting both groups together of the number of people who actually spend less and more than 1000 AED. However, precision is low, depicting that the model could be improved to predict those who spent more than 1000 AED accurately. Moreover, to increase revenues for Amazon/souq.com, it is recommended to offer promotions, rewards after purchase to increase reviews, improve ease of navigation on website/app, and learn from other E-Commerce competitors in the UAE as they seem to have a higher competitive advantage.

Business problem

The main business problem is trying to determine what factors such as demographics of a customer and their satisfaction ratings of a website, can aid in predicting whether a customer will spend more than 1000 AED per purchase. The dataset is taken from 1600 individuals in UAE to compare the overall satisfaction of various e-Commerce websites. There are 22 e-Commerce companies in total where I am focusing on Amazon/souq.com as the main e-Commerce website compared to other e-Commerce websites. According to the descriptive statistics of the data, most customers buy from Amazon/souq.com. The data shows that most customers who purchase orders from the Amazon/souq.com platform have the following demographics & characteristics:

- Male

- Married

- 31-year-olds

- Have a university/college degree

- Work full-time

- Have a monthly household income of 15,000 - 20,000 AED

- Read reviews of products before purchasing

- Interact with sellers during the shopping experience

- Prefer to pay with credit cards

- Rate between 7 and 8 across several parameters showing high satisfaction overall in webpage, checkout, delivery, tracking of package and purchase information

- Have no preference between physical or online stores. However, customers who purchase more than 7 times within 6 months, tend to prefer online stores.

Variables in the dataset

The dataset has 43 variables describing the various e-Commerce competitors in UAE, various ratings related to the product quality, service quality, offering, app/website, checkout procedure, delivery options, package descriptions, payment preference, number of purchases in the last six months, the average amount spent in three months, overall satisfaction ratings and demographics such as gender, age, employer status, marital status, the highest level of education, monthly personal and household income and number of children dependent on customer's financial support. To have an overview of what factors have an impact on the average amount spent in three months, I ran a correlation table of all the variables and sorted for the dependent variable. The results show that there is a negative correlation of 0.24 with the number of times purchased in the last six months. This means that customers spend less money if they purchase several times in six months. The next variable that has the strongest positive correlation of 0.7 is the most common method used to buy on the website. This shows that mobile app has a significant impact on customers' shopping experience and should be taken into consideration when targeting customer satisfaction to increase sales.

Assumptions were drawn when cleaning the data

When cleaning the data, I discovered that there were 196 missing values in the variable monthly personal income. These missing values were replaced by the mode, the most frequent monthly personal income group equal to 4 = 11,000 AED – 15,000 AED. Three variables contained values corresponding to 97 – 99 meaning it was not applicable. These values were replaced with the mode for the variables "Return and exchange process satisfaction rating" & "Monthly household income".

For the variable “Number of children dependent on your financial support”, the value 99 was replaced with 0 because the mode was 99. The value 0 would better reflect an individual who does not have children. Also, note that the variable “Monthly Household Income” had both 97 and 98 values. Since 97 is not part of the original scale, I am assuming this is a mistake. Therefore, both values 97 and 98 were replaced with the mode. Furthermore, I created dummy variables for the categorical variables such as marital status, gender, education, and the average amount spent per visit last 3 months to be able to predict a customer spending 1000 AED per visit.

Variables of interest

After cleaning the data in Python, I ran multiple regression models (Appendix A-C) to see how much the explanatory variables could predict a customer spending more than 1000 AED. When creating a regression model solely based on the demographics of a customer purchasing for Amazon/souq.com, the variable Age had a significant impact in predicting a customer's spending habits (Appendix B). On average, a customer would spend 10 AED more with age increasing by one year. This suggests that older people have more purchase power than younger segments.

Further analysis revealed that the variables of interest are e-Commerce being Amazon/souq.com, customers who are married, the number of times purchased in the last 6 months, most common platform to buy on website, physical or online store purchase, and monthly household income. This model gave the highest adjusted R-square of 14.2 %. There is a higher chance of a customer purchasing over 1000 AED if the customer is married, uses a mobile app, and has an adequate monthly household income. This suggests that family-oriented individuals might purchase more not only for themselves but for their families and prefer to purchase on their phones for ease of method and convenience (Appendix A). The regression model also showed that customers purchasing orders over 1000 AED purchased from other e-Commerce websites. It is therefore appropriate to do more research on competitors in UAE and implement competitive strategies to gain a competitive advantage in the business environment (Muhannad & Ahmed, 2014).

Moreover, the regression model suggests that customers are indifferent in purchasing from physical or online stores. According to Saxena (2016), customers would prefer to buy online if there are a wide variety of products that they cannot find in the physical stores. E-Commerce can capture great market share when selling in both channels, rather than only focusing on online stores. This is why this variable is important to include in the regression model despite a high p-value (Saxena, 2016). Other factors that also help increase customer satisfaction and thus increase sales can be seen in Appendix E. The single most important factor that explains how to gain customers with a strong purchasing power is targeting individuals with high monthly household income.

Logistic regression model to predict spending over 1000 AED

Logistic regression was conducted with the variables of interest (Appendix D). Note that 0 means that an individual spends less than 1000 AED and 1 means that a person spends more than 1000 AED on average in three months. The confusion matrix created based on the logistic regression (Appendix D) shows that the model predicts that 1385 individuals will spend less than 1000 AED at Amazon/souq.com and do spend less than 1000 AED with an accuracy of 88 %. While only 23 individuals are predicted to spend more than 1000 AED out of 1600 customers with a cut-off at 20 %.

Recommendations

Based on the analysis disseminated above, here are recommendations for Amazon/souq.com to increase revenues:

Data shows that those who make a lot of purchases within six months happen to purchase a lot from competitors 3, 4, and 5.

1. It is recommended to create a competitive analysis to stand out in the market. One of the ways to gain a competitive advantage is to have promotions and offers for the customers.

   Analysis reveals that high customer satisfaction across several parameters is essential for a higher number of purchases in 6 months. Note that customers who have a high number of purchases tend to pay less than 200 AED per purchase. Even though the value is low per purchase, it increases exponentially in the long run and will influence the average amount spent in 3 months.

2. It is therefore recommended to increase customer satisfaction by focusing more on the mobile app as the main platform for purchase. Focus on making the website/app tech-friendly and easy to navigate specifically for the older population who are family-oriented (Abdeldayem, 2010).

3. Give rewards to customers when purchasing from Amazon/souq.com and entice them to post reviews on the company's website (Appendix E).

References

Abdeldayem, M. M. (2010). A study of customer satisfaction with online shopping: Evidence from the UAE. *International Journal of Advanced Media and Communication*, *4*(3), 235. https://doi.org/10.1504/ijamc.2010.034659

Muhannad, K., & Ahmed, G. (2014). Customer Perceptions of e-Commerce in the United Arab Emirate. *International Journal of Global Business*. https://doi.org/https://www.researchgate.net/profile/Gouher-Ahmed/publication/319183264_Customer_Perceptions_of_e-Commerce_in_the_United_Arab_Emirates/links/59a59e66a6fdcc61fcf97df6/Customer-Perceptions-of-e-Commerce-in-the-United-Arab-Emirates.pdf

Saxena, D. R. P. (2016). Shoppers' Attitude towards Online Retailing: An Exploratory Study Performed in the Context of United Arab Emirates (UAE). *Faculty of Business and Management, University of Wollongong in Dubai*. http://archives.marketing-trends-congress.com/2011/Materiali/Paper/Territorial/Saxena.pdf

Appendices

Appendix A – Final regression model with highest R-Square

OLS Regression Results

Dep. Variable:	Y >1000 AED	**R-squared (uncentered):**	0.146
Model:	OLS	**Adj. R-squared (uncentered):**	0.142
Method:	Least Squares	**F-statistic:**	45.30
Date:	Sun, 14 Nov 2021	**Prob (F-statistic):**	2.18e-51
Time:	18:05:54	**Log-Likelihood:**	-250.64
No. Observations:	1600	**AIC:**	513.3
Df Residuals:	1594	**BIC:**	545.6
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
is_Amazon	-0.0409	0.016	-2.516	0.012	-0.073	-0.009
is_married	0.0475	0.015	3.232	0.001	0.019	0.076
Number of times purchased in the last 6 months	-0.0311	0.004	-7.203	0.000	-0.040	-0.023
Most common platform to buy on website	0.0252	0.010	2.445	0.015	0.005	0.045
Physical or online store purchase	-0.0046	0.008	-0.597	0.551	-0.020	0.010
Monthly household income	0.0198	0.004	5.507	0.000	0.013	0.027

Omnibus:	802.064	**Durbin-Watson:**	1.875
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3378.403
Skew:	2.535	**Prob(JB):**	0.00
Kurtosis:	7.997	**Cond. No.**	15.9

Appendix B – Regression with demographic variables

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.1855546
R Square	0.03443051
Adjusted R Square	2.84%
Standard Error	501.262934
Observations	1600

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	10	14236885.2	1423688.52	5.6660943	2.2246E-08
Residual	1589	399259336	251264.529		
Total	1599	413496222			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	
Intercept	92.9073445	308.167444	0.30148332	76%	-511.550165	697.364854	
Where do you usually make most purchases on physical store or an online store?	-22.9992146	15.8451912	-1.45149493	15%	-54.0788921	8.08046296	
Age	10.1419368	1.80977935	5.60396318	0%	6.59213057	13.6917431	significant value
Married	68.3792664	290.712851	0.2352124	81%	-501.841792	638.600325	
Single	1596.22277	1605.7884	0.99404303	32%	-1553.46379	4745.90934	
is_female	6.59777969	25.9282581	0.2544629	80%	-44.2594105	57.4549698	
has_college_degree	53.7123233	34.3180945	1.56513128	12%	-13.601179	121.025826	
# of children that depend on your financial support	-15.8882699	16.58822	-0.95780439	34%	-48.4253673	16.6488274	
Employement Status	-14.6148953	14.0875975	-1.03742993	30%	-42.2471266	13.017336	
Monthly Personal Income	-1.39140401	0.99786731	-1.39437778	16%	-3.34867888	0.56587085	
Monthly Household Income	0.92213384	0.61409264	1.50162009	13%	-0.2823831	2.12665079	

Appendix C – Multiple regressions to find highest R-square and significant explanatory variables

OLS Regression Results

Dep. Variable:	Y >1000 AED	**R-squared (uncentered):**	0.106
Model:	OLS	**Adj. R-squared (uncentered):**	0.101
Method:	Least Squares	**F-statistic:**	23.54
Date:	Sun, 14 Nov 2021	**Prob (F-statistic):**	2.30e-34
Time:	17:53:24	**Log-Likelihood:**	-287.15
No. Observations:	1600	**AIC:**	590.3
Df Residuals:	1592	**BIC:**	633.3
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
is_Amazon	-0.0401	0.017	-2.403	0.016	-0.073	-0.007
Service quality	-0.0014	0.008	-0.188	0.851	-0.016	0.013
Tracking package/order rating	0.0084	0.007	1.171	0.242	-0.006	0.022
Most common platform to buy on website	0.0264	0.011	2.394	0.017	0.005	0.048
Easy to compare products in e-Commerce app/website rating	0.0024	0.007	0.341	0.733	-0.011	0.016
Availability of products rating	0.0040	0.007	0.580	0.562	-0.010	0.018
e-Commerce satisfaction rating	-0.0050	0.011	-0.467	0.640	-0.026	0.016
yes_recommended_friend	-0.0072	0.018	-0.410	0.682	-0.042	0.027

Omnibus:	873.383	**Durbin-Watson:**	1.869
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	4154.057
Skew:	2.750	**Prob(JB):**	0.00
Kurtosis:	8.662	**Cond. No.**	42.8

OLS Regression Results

Dep. Variable:	Y >1000 AED	**R-squared (uncentered):**	0.118
Model:	OLS	**Adj. R-squared (uncentered):**	0.115
Method:	Least Squares	**F-statistic:**	42.63
Date:	Sun, 14 Nov 2021	**Prob (F-statistic):**	2.53e-41
Time:	18:03:01	**Log-Likelihood:**	-276.27
No. Observations:	1600	**AIC:**	562.5
Df Residuals:	1595	**BIC:**	589.4
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
is_Amazon	-0.0407	0.017	-2.462	0.014	-0.073	-0.008
is_married	0.0390	0.015	2.617	0.009	0.010	0.068
Most common platform to buy on website	0.0224	0.010	2.148	0.032	0.002	0.043
Physical or online store purchase	-0.0192	0.007	-2.560	0.011	-0.034	-0.004
Monthly household income	0.0153	0.004	4.243	0.000	0.008	0.022

Omnibus:	854.223	**Durbin-Watson:**	1.869
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3944.550
Skew:	2.690	**Prob(JB):**	0.00
Kurtosis:	8.499	**Cond. No.**	15.0

OLS Regression Results

Dep. Variable:	Y >1000 AED	**R-squared (uncentered):**	0.118
Model:	OLS	**Adj. R-squared (uncentered):**	0.114
Method:	Least Squares	**F-statistic:**	30.33
Date:	Sun, 14 Nov 2021	**Prob (F-statistic):**	1.38e-39
Time:	17:58:08	**Log-Likelihood:**	-276.51
No. Observations:	1600	**AIC:**	567.0
Df Residuals:	1593	**BIC:**	604.7
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
is_Amazon	-0.0415	0.017	-2.503	0.012	-0.074	-0.009
Product quality	0.0014	0.004	0.333	0.739	-0.007	0.009
Most common platform to buy on website	0.0159	0.011	1.415	0.157	-0.006	0.038
is_single_or_divorced	-0.0445	0.016	-2.749	0.006	-0.076	-0.013
is_female	-0.0095	0.014	-0.659	0.510	-0.038	0.019
Monthly household income	0.0119	0.005	2.570	0.010	0.003	0.021
Age	0.0007	0.001	0.764	0.445	-0.001	0.002

Omnibus:	858.821	**Durbin-Watson:**	1.874
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	3994.236
Skew:	2.704	**Prob(JB):**	0.00
Kurtosis:	8.538	**Cond. No.**	87.0

OLS Regression Results

Dep. Variable:	Y >1000 AED	**R-squared (uncentered):**	0.108
Model:	OLS	**Adj. R-squared (uncentered):**	0.105
Method:	Least Squares	**F-statistic:**	38.49
Date:	Sun, 14 Nov 2021	**Prob (F-statistic):**	2.12e-37
Time:	17:46:58	**Log-Likelihood:**	-285.46
No. Observations:	1600	**AIC:**	580.9
Df Residuals:	1595	**BIC:**	607.8
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
is_Amazon	-0.0390	0.017	-2.341	0.019	-0.072	-0.006
Service quality	0.0056	0.003	2.108	0.035	0.000	0.011
is_married	0.0383	0.016	2.357	0.019	0.006	0.070
is_female	-0.0037	0.014	-0.256	0.798	-0.032	0.025
Monthly personal income	0.0109	0.005	2.034	0.042	0.000	0.021

Omnibus:	867.827	**Durbin-Watson:**	1.872
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	4093.134
Skew:	2.732	**Prob(JB):**	0.00
Kurtosis:	8.615	**Cond. No.**	19.7

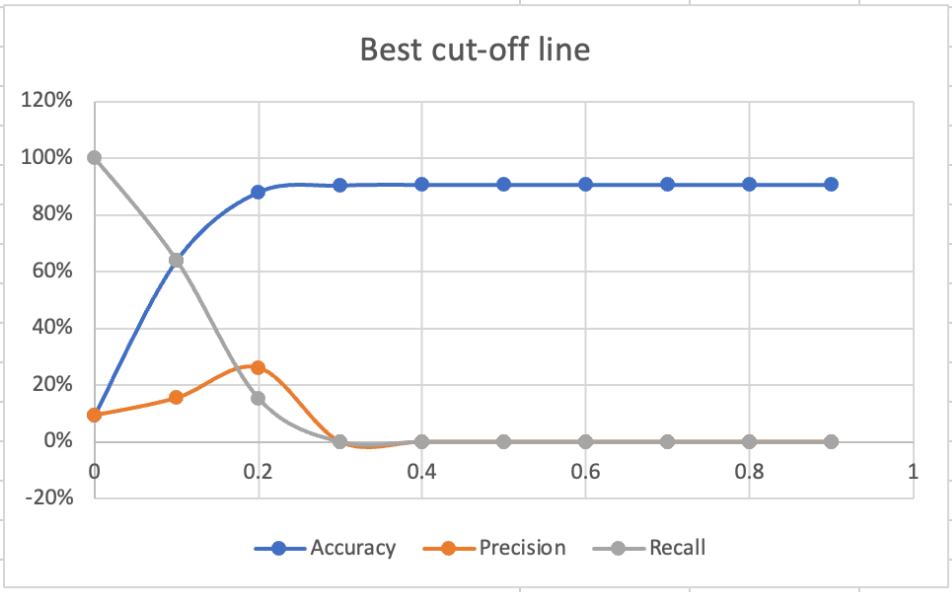
Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Appendix D - Predictive model based on logistic regression model

0 means Y < 1000 AED and 1 means Y > 1000 AED

		Predicted		
		0	1	
Actual	0	1385	65	1450
	1	127	23	150
		1512	88	1600

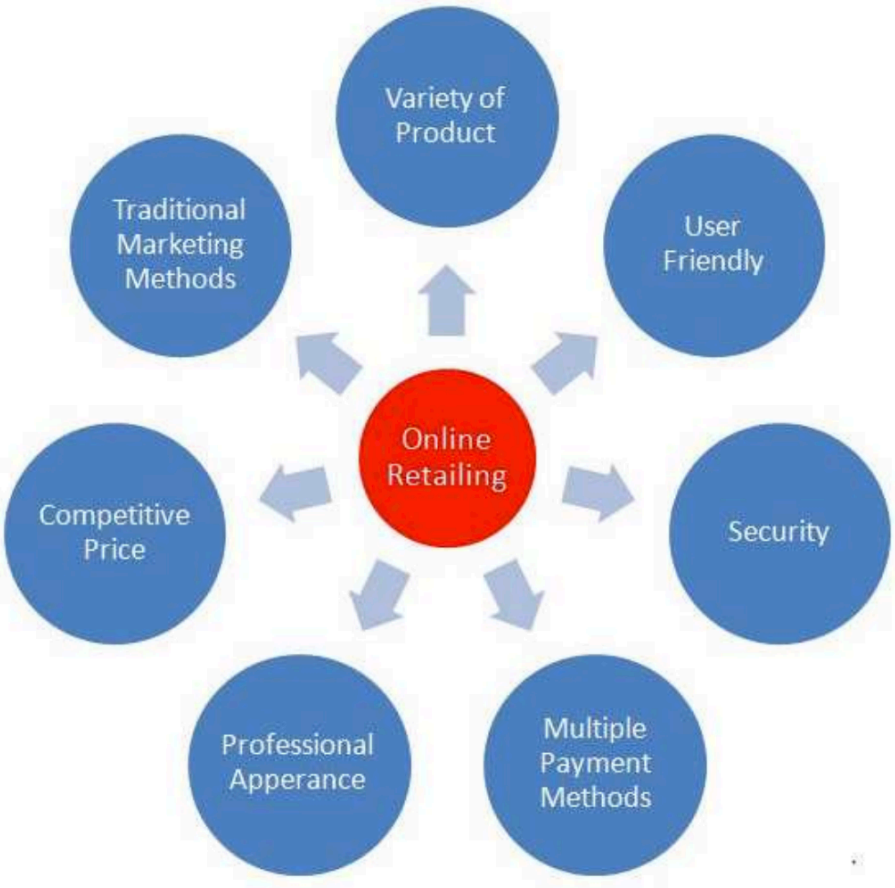
Accuracy	88%
Precision	26%
Recall	15%



Intercept	b1	b2	b3	b4	b5	b6			Cut-off	20%	Sum of the log likehood	-203.8109365
-4.0256387	-0.52999	0.578375	-1.874445973	0.319745144	-0.141426329	0.218701188						

Y >1000 AED	is_Amazon	is_married	# of times purchased bin	Most common platform to buy on website	Physical or online store purchase	Monthly household income	Logit	Odds	Probability of Y> 1000 AED	Probability of a Correct Match	Log Likelihood	Predicted classification
1	0	1	0	2	3	4	-2.3572476	0.0946805	8.649%	0.086491418	-1.063026985	0
0	0	1	0	2	3	7	-1.701144	0.1824747	15.432%	0.845684091	-0.072791839	0
0	0	0	1	1	3	3	-5.348515	0.0047552	0.473%	0.995267298	-0.002060266	0
0	0	1	0	2	2	5	-1.9971201	0.1357256	11.951%	0.880494371	-0.055273416	0
0	0	1	0	2	3	6	-1.9198452	0.1466297	12.788%	0.872121171	-0.059423171	0
0	1	0	1	1	2	8	-4.6435726	0.0096233	0.953%	0.990468469	-0.004159346	0
0	0	1	0	2	3	5	-2.1385464	0.117826	10.541%	0.89459362	-0.048374203	0
0	1	1	0	3	2	4	-2.4260661	0.0883838	8.121%	0.918793497	-0.036782087	0
0	0	0	0	1	3	7	-2.5992642	0.0743282	6.919%	0.930814212	-0.031136995	0
0	0	1	0	3	1	8	-0.879845	0.4148472	29.321%	0.706790106	-0.150709539	1
1	0	1	0	1	2	7	-1.8794628	0.1526721	13.245%	0.132450586	-0.877946116	0

Appendix E – Essential factors to increase customer spending habits online



Essentials for Successful Online Retailing

Figure-13

Source: http://archives.marketing-trends-congress.com/2011/Materiali/Paper/Territorial/Saxena.pdf