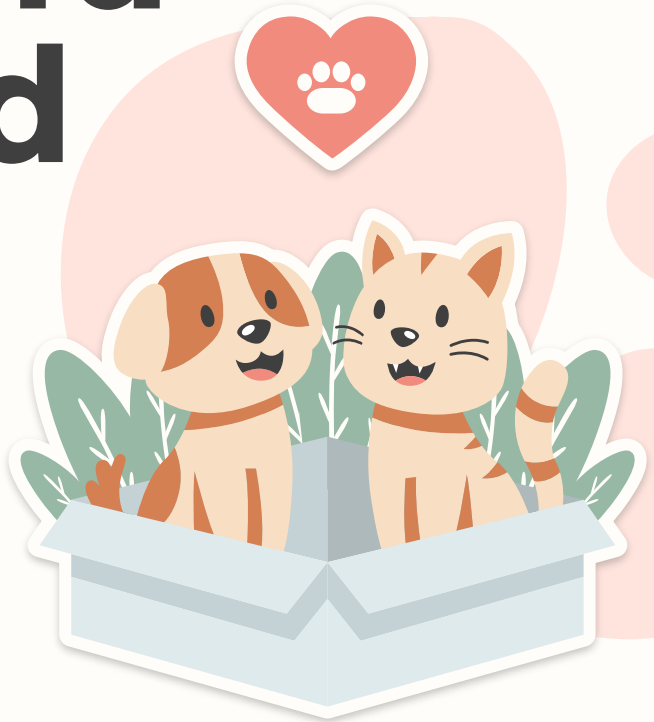
Pet store data cleaning and analysis



Developed by chloewongwy

Table of contents

01.

Introduction

Objectives:

- product selection in the store
- product quality
- product price

Why the pet industry in Hong Kong become stronger?

demographic changes

many elderlies who are single
→ adopting a pet
→ to erase their feeling of isolated

pet owners keep their pet as another housemate

will buy different luxuries and service for their pets as their lifestyle like

- grooming needs
- wellness
- beauty

02.

Data Processing and Cleaning

After cleaning the data, a total of 22 rows were deleted and leaving 994 rows

Spelling mistakes

(Column A, B, D, AQ)

Having Synonyms

(Column A, B, D, AQ, AR, AS, AU, AX)

Replace with the most data option in the same column with the same meaning

Blank cell

(Column AQ, AU)

BBA is not an education level

(Column AS)

Delete the record with missing data

Intentional misreporting

(Column C) Maximum time to visit the pets good shop in a month must not be 0 if (Column A) last time visiting the pet goods shop is this week or last week

Delete the entire row that answers 0 in column C and week in column A

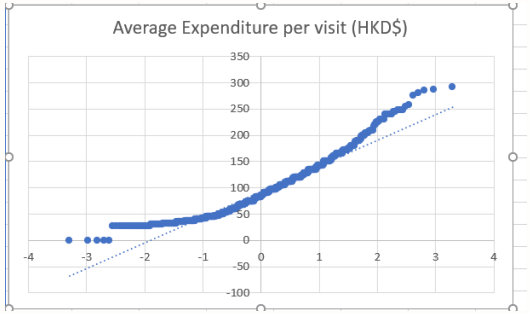
Missing data

(Column E to AP)

Delete the record with missing data if the entire row of data in column E to AP is missing

Mean substitution if there are less than 3 missing data in a row

Outliers
(Column AT)



After deleting the row with the value of 25000, following rules of Thumb, the value range should be below 244.5. Since the value is related to money, the value should be 0 or above.

Extreme value

Column	AV	AX	AW
Solution	Replace numbers greater than 100 with 100 and negative numbers with 0, because when the unit is a percentage, the maximum number should be 100 and the minimum number should be 0		Delete the record with cell is 0 because it is impossible to leave within 1 minute

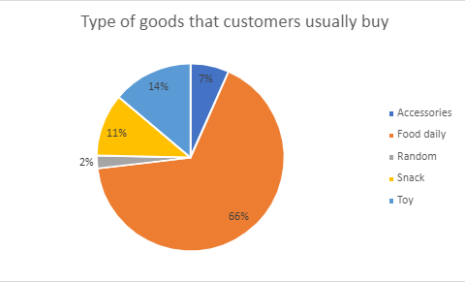
03.

Descriptive Analysis

Qualitative & Quantitative

Qualitative

Objective (Product selection in the store)



Columns: Type of goods usually buy

- Daily food is the most attractive goods in the store (66%)

→ More types of daily foods can be added to the store

Count of Type of goods usually buy	Gender		
Type of goods	Female	Male	Grand Total
Accessories	37	29	66
Food (Daily)	367	293	660
Random	9	13	22
Snack	64	43	107
Toy	74	64	138
Grand Total	**551**	**442**	**993**

Columns: Type of goods usually buy, Gender

- Both females and males usually buy daily food for their pets
- Female buy more goods for their pets (Accessories, Food, Snack, Toy) than male

Count of Type of goods usually buy	Maximum time (hours)							
Type of goods	0	1	2	3	4	5	6	Grand Total
Accessories	0	40	14	5	1	5	1	66
Food (Daily)	8	233	163	153	57	34	12	660
Random	0	16	6	0	0	0	0	22
Snack	1	45	32	17	5	5	2	107
Toy	1	73	29	24	6	4	1	138
Grand Total	**10**	**407**	**244**	**199**	**69**	**48**	**16**	**993**

Columns: Type of goods usually buy Maximum time to visit the pets good shop in a month

- Customers typically spend more time shopping for food, followed by snacks
- Customers with no fixed buying goals usually buy for a shorter period of time

Quantitative

Objective (Product quality)

	Quality (Importance)	Quality (Satisfaction Level)	Hygiene (Importance)	Hygiene (Satisfaction Level)	Freshness (Importance)	Freshness (Satisfaction Level)
Mean	3.464249748	3.435045317	3.5075287	3.347432024	3.996978852	3.694864048
Standard Error	0.03139604	0.030793356	0.032744718	0.028977305	0.03158961	0.034881608
Median	3	4	4	3	4	4
Mode	3	4	4	3	4	4
Standard Deviation	0.989348956	0.970357237	1.031848362	0.913130003	0.995448785	1.099185827
Sample Variance	0.978811357	0.941593168	1.064711042	0.833806403	0.990918283	1.208209483
Kurtosis	0.072417199	0.123968903	-0.059398717	0.193961686	1.239737764	-0.219378659
Skewness	-0.42003701	-0.648764765	-0.467120559	-0.311458175	-1.124463516	-0.661433234
Range	4	4	4	4	4	4
Minimum	1	1	1	1	1	1
Maximum	5	5	5	5	5	5
Sum	3440	3411	3483	3324	3969	3669
Count	993	993	993	993	993	993

Columns: Quality (Importance), Quality (Satisfaction Level), Freshness (Importance), Hygiene (Importance), Hygiene (Satisfaction Level), Freshness (Satisfaction Level)

- the mean of these three types of satisfaction levels is all lower than the importance

→ customers may not be satisfied with the product's quality, hygiene, and freshness

- the mean of freshness importance and satisfaction level are both the highest

→ customers are placing emphasis on the product's freshness

- the quality satisfaction level is lower than the freshness satisfaction level

→ even if customers are satisfied with product freshness, they may not be satisfied with product quality

Objective (Product selection in the store)

	Choice (Satisfaction Level)	New Product (Satisfaction Level)
Mean	3.793762575	3.284708249
Standar Error	0.033575724	0.032931448
Median	4	3
Mode	4	3
Standard Devistion	1.058567563	1.038254968
Sample Variance	1.120565285	1.077973379
Kurtosis	0.236100137	-0.217659316
Skewness	-0.770574445	-0.178713427
Range	4	4
Minimum	1	1
Maximum	5	5
Sum	3771	3265
Count	994	994

Columns: Choice (Satisfaction Level), New Product (Satisfaction Level)

- choice's mean > new product's mean

→customers may think our choice is enough and they are satisfied but it is not enough satisfaction for our new product

- left skewed distribution (Choice & New Product: skewness are a negative number)

	Choice (Importance)	New Product (Importance)
Mean	3.895372233	3.277665996
Standar Error	0.035914236	0.033615574
Median	4	3
Mode	5	3
Standard Devistion	1.132295618	1.059823935
Sample Variance	1.282093366	1.123226773
Kurtosis	0.373027955	-0.2696308
Skewness	-1.011541802	-0.205453063
Range	4	4
Minimum	1	1
Maximum	5	5
Sum	3872	3258
Count	994	994

Columns: Choice (Importance), New Product (Importance)

- choice's mean > new product's mean

→customers may think our number of choices is importance than the new product

- heavy tailed distribution (Choice: positive kurtosis)
- light tailed distribution (New Product: negative kurtosis)

Quantitative

Objective (Product price)

Columns: Price (Importance) (y), Quality (Importance) (x)

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.372724131
R Square	0.138923278
Adjusted R Square	0.138054381
Standard Error	1.025100657
Observations	993

r = 0.3727
weak positive linear correlation

When Quality (Importance) increase, Price (Importance) also increase.

Columns: Price (Satisfaction Level) (y), Quality (Satisfaction Level) (x)

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.48950633
R Square	0.239616447
Adjusted R Square	0.238849158
Standard Error	0.89634245
Observations	993

r = 0.4895
weak positive linear correlation

When Quality (Satisfaction Level) increase, Price (Satisfaction Level) also increase.

Regression models and Chi-square tests

04.

Tasks allocation

Consultant	Group Member Name
A	Foo Chi Hin
B	Tang Jacky
C	Lai Kin Hei
D	Lau Wai Hin
E	Wong Wing Yi

Consultant A

Qualitative--Chi Square independent test

Age x Type of goods usually buy

Count of Age	Type of goods usually buy					
Age	Accessories	Food (Daily)	Random	Snack	Toy	Grand Total
18-24	21	268	12	39	46	386
25-29	22	156	6	29	42	255
30-34	10	100	2	19	29	160
35-39	5	85		8	7	105
40 or above	8	51	2	12	14	87
Grand Total	**66**	**660**	**22**	**107**	**138**	**993**

Observed	Type of goods usually buy					
Age	Accessories	Food (Daily)	Random	Snack	Toy	Grand Total
18-24	21	268	12	39	46	386
25-29	22	156	6	29	42	255
30-34	10	100	2	19	29	160
35-39	5	85	0	8	7	105
40 or above	8	51	2	12	14	87
Grand Total	66	660	22	107	138	993

Expected	Type of goods usually buy					
Age	Accessories	Food (Daily)	Random	Snack	Toy	Grand Total
18-24	25.65558912	256.5558912	8.551863041	41.59315206	53.64350453	386
25-29	16.94864048	169.4864048	5.649546828	27.47734139	35.43806647	255
30-34	10.63444109	106.3444109	3.544813696	17.24068479	22.23564955	160
35-39	6.978851964	69.78851964	2.326283988	11.3141994	14.59214502	105
40 or above	5.782477341	57.82477341	1.927492447	9.374622356	12.09063444	87
Grand Total	66	660	22	107	138	993

Chi-square	Type of goods usually buy					
Age	Accessories	Food (Daily)	Random	Snack	Toy	Grand Total
18-24	0.844826053	0.510483796	1.390299217	0.161671749	1.089100387	3.996381201
25-29	1.505503228	1.073142802	0.021739341	0.084378223	1.215048562	3.899812157
30-34	0.037850179	0.378501785	0.673222787	0.179528251	2.057796286	3.326899288
35-39	0.561103046	3.315575915	2.326283988	0.970808208	3.95011603	11.12388719
40 or above	0.850397927	0.805494417	0.002727557	0.735241113	0.301528894	2.695390007
Grand Total	3.799680432	6.083198715	4.414272889	2.131627543	8.613590258	25.04236984

p-value (1.3081E-24) < level of significance (0.05)

Age and Type of goods usually buy are not independent.

Age x Last time visiting the pet goods shop

Count of Age	Last time visiting the pet goods shop				
Age	This week	Last week	Last month	No visit within last 6 months	Grand Total
18-24	72	109	172	33	386
25-29	60	78	107	10	255
30-34	25	79	56		160
35-39	53	2	49	1	105
40 or above	15	13	59		87
Grand Total	**225**	**281**	**443**	**44**	**993**

Observed	Last time visiting the pet goods shop				
Age	This week	Last week	Last month	No visit within last 6 months	Grand Total
18-24	72	109	172	33	386
25-29	60	78	107	10	255
30-34	25	79	56	0	160
35-39	53	2	49	1	105
40 or above	15	13	59	0	87
Grand Total	225	281	443	44	993

Expected	Last time visiting the pet goods shop				
Age	This week	Last week	Last month	No visit within last 6 months	Grand Total
18-24	87.46223565	109.2306143	172.203424	17.10372608	386
25-29	57.77945619	72.16012085	113.7613293	11.29009366	255
30-34	36.25377644	45.27693857	71.3796576	7.089627392	160
35-39	23.79154079	29.71299094	46.8429003	4.652567976	105
40 or above	19.71299094	24.61933535	38.81268882	3.854984894	87
Grand Total	225	281	443	44	993

Chi-square	Last time visiting the pet goods shop				
Age	This week	Last week	Last month	No visit within last 6 months	Grand Total
18-24	2.733530986	0.000486887	0.000240305	14.77406287	17.50832105
25-29	0.085338546	0.472618229	0.401855132	0.149361035	1.109172942
30-34	3.493359768	25.11743031	3.31743382	7.089627392	39.01426075
35-39	35.85871539	25.84761219	0.099333711	2.867920041	64.67716433
40 or above	1.12678404	5.483858602	10.49985314	3.854984894	20.96548068
Grand Total	43.29772873	56.92210611	14.31502567	28.73553923	143.2703997

p-value (0.0691) > level of significance (0.05)

Age and Last time visiting the pet goods shop are independent.

Consultant A

Quantitative: Maximum time to visit the pets good shop in a month

A regression analysis for the rating on Maximum time to visit the pets good shop in a month (Y) with Average time to shop (in minutes) x_1, Sore convenience (Satisfaction Level) x_2, Probability to visit in next month (%) x_3, Discount level required to attract you to visit (%) x_4, Quality (Satisfaction Level) x_5.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.280718583
R Square	0.078802923
Adjusted R Square	0.074136271
Standard Error	1.220481297
Observations	993

				2.22317062	0.05			
ANOVA								
	df	SS	MS	F	Significance F			
Regression	5	125.7677187	25.15354375	16.8863942	4.85341E-16			
Residual	987	1470.210126	1.489574596					
Total	992	1595.977845						

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1.376316551	0.208750332	6.593122686	7.0043E-11	0.966671078	1.785962	0.966671078	1.785962024
Average time to shop (in minutes)	-0.033717075	0.010255781	-3.28761647	0.00104605	-0.053842716	-0.013591	-0.05384272	-0.01359143
Store convenience (Satisfaction Level)	0.308706187	0.040270052	7.665899928	4.2395E-14	0.22968143	0.3877309	0.22968143	0.387730945
Probability to visit in next month (%)	0.006111734	0.001480688	4.127630553	3.9747E-05	0.003206075	0.0090174	0.003206075	0.009017393
Discount level required to attract you to visit (%)	0.000131753	0.003234374	0.0407353	0.96751516	-0.006215286	0.0064788	-0.00621529	0.006478792
Quality (Satisfaction Level)	-0.102845105	0.047476445	-2.16623434	0.03053207	-0.196011475	-0.009679	-0.19601147	-0.00967874

F-test:

p-value(4.85E-16) < level of significance (0.05) → Reject the hypothesis → there is sufficient evidence that not all the coefficient(s) are equals to zero

T-test:

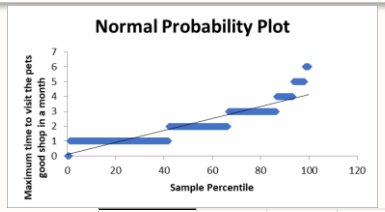
p-value(4.239E-14) < level of significance (0.05) → Reject hypothesis → there is sufficient evidence that the coefficient of Store convenience (Satisfaction Level) is not equal to zero

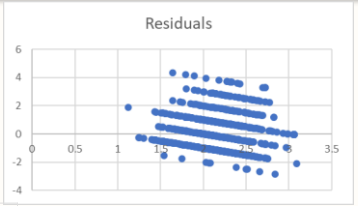
The regression line: $\hat{y}$ = 1.38 - 0.03x_1 + 0.31x_2 + 0.01x_3 - 0.1x_5

Consultant A

Quantitative: Maximum time to visit the pets good shop in a month

Assumptions





SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.471594
R Square	0.222401
Adjusted R	0.219253
Standard E	3.786029
Observatio	993

VIF = 1/(1-r square) = 1.28601

ANOVA

	df	SS	MS	F	gnificance F
Regressio	4	4050.473	1012.618	70.64444	1.2E-52
Residual	988	14162.01	14.33401		
Total	992	18212.48			

	Coefficients	tandard Err	t Stat	P-value	ower 95%	Upper 95%	ower 95.0%	pper 95.0%
Intercept	4.197042	0.633644	6.62366	5.75E-11	2.9536	5.440485	2.9536	5.440485
Store conv	1.098532	0.11993	9.160975	2.9E-19	0.863303	1.334001	0.863303	1.334001
Probabilit	-0.01124	0.004579	-2.45448	0.01428	-0.02023	-0.00225	-0.02023	-0.00225
Discount l	0.045163	0.00993	4.548221	6.08E-06	0.025677	0.064649	0.025677	0.064649
Quality (S	0.961317	0.144065	6.672791	4.18E-11	0.678608	1.244025	0.678608	1.244025

Residuals	DW
-0.94552669	1.892112
-0.79431947	
-1.0586145	
1.459605844	
0.752336418	
-1.4786539	

X	VIF
Average time to shop (in minutes)	1.286009871
Store convenience (Satisfaction Level)	1.252625487
Probability to visit in next month (%)	1.020829112
Discount level required to attract you to visit (%)	1.233600199
Quality (Satisfaction Level)	1.413413032

Durbin-Watson = 1.8921, The DW is in between 1.5 to 2.5, so there is no possible violation in the assumptions over the independence of errors

Because majority of the points are not on the line, there is a possible violation in the assumption over the normality of the residuals.
Because majority of the points are concentration, there is a possible violation in the assumption over the Homoscedasticity.
Because all VIFs are below 10, there is no possible violation over the multicollinearity.

Recommendation

Qualitative:
The pet shop should sell different types of goods. So that it can attract customers from different age groups.

Quantitative:
The pet shop should increase store convenience and reduce customers' time to shop. So that it can increase customers' maximum time to visit the pet's good shop in a month.

Consultant B

Qualitative--Chi Square independent test

Last time visiting the pet goods shop x Channel for knowing the pet goods shop

Count of Channel for knowing the pet goods shop	Column Labels					
Row Labels	No visit within last 6 months	Last month	1 week before	Last week	This week	Grand Total
Classmate		1		3	2	6
Friends		39		28	34	101
Promotion		93	5	106	113	317
Relative		11		3	6	20
Social Media	44	299	10	126	70	549
Grand Total	**44**	**443**	**15**	**266**	**225**	**993**

Observed	Column Labels					
Row Labels	No visit within last 6 months	Last month	1 week before	Last week	This week	Grand Total
Classmate	0	1	0	3	2	6
Friends	0	39	0	28	34	101
Promotion	0	93	5	106	113	317
Relative	0	11	0	3	6	20
Social Media	44	299	10	126	70	549
Grand Total	44	443	15	266	225	993

Expected	Last time visiting the pet goods shop					
Channel for knowing the pet goods shop	**No visit within last 6 months**	**Last month**	**1 week before**	**Last week**	**This week**	**Grand Total**
Classmate	0.265861027	2.67673716	0.09063441	1.60725076	1.35951662	6
Friends	4.475327291	45.05840886	1.525679758	27.0553877	22.8851964	101
Promotion	14.04632427	141.4209466	4.788519637	84.9164149	71.8277946	317
Relative	0.886203424	8.9224572	0.302114804	5.35750252	4.53172205	20
Social Media	24.32628399	244.9214502	8.293055136	147.063444	124.39577	549
Grand Total	**44**	**443**	**15**	**266**	**225**	**993**

Chi-Square Test	Last time visiting the pet goods shop					
Channel for knowing the pet goods shop	**No visit within last 6 months**	**Last month**	**1 week before**	**Last week**	**This week**	**Grand Total**
Classmate	0.070682086	2.811447504	0.008214602	1.939750459	0.410218965	5.169631529
Friends	20.02855436	36.70431794	2.327698725	0.892282371	123.5388596	163.4631687
Promotion	197.2992255	2344.588072	0.044723944	444.5175605	1695.150501	4681.300857
Relative	0.785356509	4.316184084	0.091273355	5.557818121	2.155840126	12.12113568
Social Media	307.0551017	2924.489554	2.913673661	443.6686777	2958.899837	6329.977742
Grand Total	605.2389202	5312.906575	5.385584286	896.5766992	4780.155256	10995.02651

critical value (26.2962) < test statistics (11600.26544)

Last time visiting the pet goods shop and Channel for knowing the pet goods shop are independent.

Type of goods usually buy x Channel for knowing the pet goods shop

Count of Type of goods usually buy	Column Labels					
Row Labels	Classmate	Friends	Promotion	Relative	Social Media	Grand Total
Accessories			16	17		33
Food (Daily)		5	56	212	13	374
Random			4	13	1	4
Snack		1	14	31	4	57
Toy			11	44	2	81
Grand Total		**6**	**101**	**317**	**20**	**549**

Observed	Channel for knowing the pet goods shop					
Count of Type of goods usually buy	**Classmate**	**Friends**	**Promotion**	**Relative**	**Social Media**	**Grand Total**
Accessories		0	16	17	0	33
Food (Daily)		5	56	212	13	374
Random		0	4	13	1	4
Snack		1	14	31	4	57
Toy		0	11	44	2	81
Grand Total		**6**	**101**	**317**	**20**	**549**

Expected	Channel for knowing the pet goods shop					
Count of Type of goods usually buy	**Classmate**	**Friends**	**Promotion**	**Relative**	**Social Media**	**Grand Total**
Accessories		0.398791541	6.712991	21.0694864	1.329305	36.48942598
Food (Daily)		3.987915408	67.12991	210.694864	13.29305	364.8942598
Random		0.132930514	2.237664	7.023162135	0.443102	12.16314199
Snack		0.64652568	10.88318	34.15810675	2.155086	59.1570997
Toy		0.833836858	14.03625	44.05438066	2.779456	76.29607251
Grand Total		6	101	317	20	549	993

Chi-Square Test	Channel for knowing the pet goods shop					
Count of Type of goods usually buy	**Classmate**	**Friends**	**Promotion**	**Relative**	**Social Media**	**Grand Total**
Accessories		0.159034693	86.24854	16.5607196	1.767052144	12.17609368	116.9114375
Food (Daily)		1.024315222	123.8749	1.703379852	0.085879099	82.91450425	209.6029609
Random		0.017670521	3.105829	35.72259087	0.310135703	66.63688721	105.7931137
Snack		0.124944095	9.714553	9.973638227	3.403709146	4.653079107	27.8699233
Toy		0.695283906	9.218837	0.002957257	0.607551957	22.12693385	32.65156397
Grand Total		2.021248437	232.1626	63.9632858	6.174328051	188.5074981	492.8289994

p-value (0.003025065) < level of significance (0.05)

Type of goods usually buy and Channel for knowing the pet goods shop are independent.

Consultant B

Quantitative: Average time to shop (in minutes)

A regression analysis for the Average time to shop (in minutes) (Y) with Service of the staff (Satisfaction Level) X_1 , Store convenience (Satisfaction Level) X_2 , Staff knowledge (Satisfaction Level) X_3 , Promotion (Satisfaction Level) X_4 , Decoration (Satisfaction Level) X_5.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.404336611
R Square	0.163488095
Adjusted R Square	0.145689969
Standard Error	4.030436172
Observations	241

ANOVA

	df	SS	MS	F	Significance F
Regression	5	746.0809741	149.2161948	9.185691701	5.3312E-08
Residual	235	3817.437698	16.24441574		
Total	240	4563.518672			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	4.296585169	1.440590246	2.982517188	0.003160039	1.45846383	7.134706508	1.45846383	7.134706508
Service of the staff (Satisfactio	0.434039509	0.305440977	1.421025799	0.156634824	-0.167712834	1.035791852	-0.167712834	1.035791852
Store convenience (Satisfaction	0.92551606	0.304683078	3.037635263	0.002653656	0.325256863	1.525775257	0.325256863	1.525775257
Staff knowledge (Satisfaction L	-0.03881444	0.312241766	-0.124308938	0.901176934	-0.653965088	0.576336203	-0.653965088	0.576336203
Promotion (Satisfaction Level)	-0.0754701	0.327444953	-0.230481793	0.817917926	-0.720572697	0.569632497	-0.720572697	0.569632497
Decoration (Satisfaction Level)	1.113014967	0.324821623	3.426542098	0.000721493	0.473080618	1.752949315	0.473080618	1.752949315

F-test:

test statistics(9.18569170095076) > critical value (2.252455) → Reject the hypothesis → there is sufficient evidence that not all the coefficients are equals to zero.

T-test:

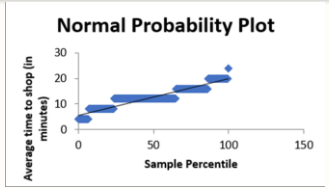
critical value (1.97011) > level of significance(0.05) → Reject the hypothesis → there is sufficient evidence that the coefficient of Store convenience (Satisfaction Level) is not equal to zero.

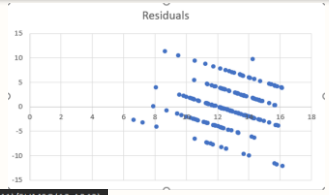
The regression line: $\hat{y} = 4.30 + 0.43x_1 + 0.93x_2 + 1.11x_5$

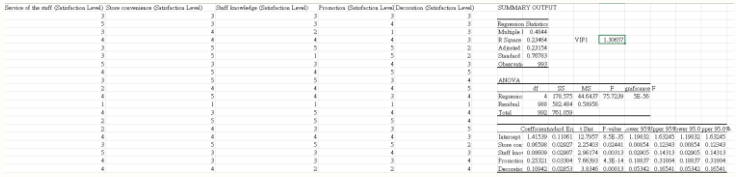
Consultant B

Quantitative: Average time to shop (in minutes)

Assumptions







=SUMXMY2(A3:A242,A2:A241)/SUMSQ(A2:A242)

Residuals	Durbin-Watson
-3.37144	1.937798
-0.16405	
3.513286	
-0.61671	
-3.88089	
-4.03615	
-0.16405	

X	VIF
Service of the staff (Satisfaction Level)	1.306575
Store convenience (Satisfaction Level)	1.68092
Staff knowledge (Satisfaction Level)	1.685127
Promotion (Satisfaction Level)	1.865889
Decoration (Satisfaction Level)	1.447987

Durbin-Watson = 1.937798, The DW is in between 1.5 to 2.5, so there is no possible violation in the assumptions over the independence of errors

Because majority of the points are not on the line, there is a possible violation in the assumption over the normality of the residuals.
Because the points are scattered around the plot ad there is concentration, there is possible violation over the Homoscedasticity.
Because all VIFs are below 10, there is no possible violation over the multicollinearity

Recommendation

Qualitative:
We can conclude that daily food is the most attractive goods in the store and social media is the fastest way to let customers know our shop.
Quantitative:
The pet stop should increase staff service to let customers stay longer.

Consultant C

Qualitative--Chi Square independent test

Payment x Education Level

Count of Payment	Column Labels					
Row Labels	Degree	Doctorate	Junior High School	Master	Senior High School	Grand Total
Card	17	0	1	0	3	21
Cash	171	0	9	13	67	260
Credit Card	388	0	47	14	133	582
E-Money	59	1	2	0	68	130
Grand Total	**635**	**1**	**59**	**27**	**271**	**993**

Observed	Education level					
Payment	Degree	Doctorate	Junior High School	Master	Senior High School	Grand Total
Card	17	0	1	0	3	21
Cash	171	0	9	13	67	260
Credit Card	388	0	47	14	133	582
E-Money	59	1	2	0	68	130
Grand Total	635	1	59	27	271	993

Expected	Education level					
Payment	Degree	Doctorate	Junior High School	Master	Senior High School	Grand Total
Card	13.42900302	0.021148036	1.247734139	0.570096979	5.73117825	21
Cash	166.2638469	0.26183283	15.44813996	7.069486405	70.95669688	260
Credit Card	372.1752266	0.586102719	34.58006042	15.82477341	158.8338369	582
E-Money	83.13192346	0.130916415	7.724068479	3.534743202	35.47834844	130
Grand Total	635	1	59	27	271	993

Chi-square test	Education level						
Payment	Degree	Doctorate	Junior High School	Master	Senior High School	Grand Total	
Card	0.949588	0.021148	0.049186923	0.570096979	1.3014921	2.892411987	
Cash	0.134913	0.2618328	2.691487676	4.97504196	0.220633864	8.283009279	
Credit Card	0.672864	0.5861027	4.460804788	0.210416789	4.201794404	10.13198308	
E-Money	7.005128	5.769378	4.241930279	3.534743202	29.81136008	50.36253909	
Grand Total	8.762493	6.6384615	11.44340967	9.291198931	35.53528044	71.67084344	test statistics

critical value (21.0261) < test statistics (71.6708)

Payment and Education Level of the customers are not independent

Payment x Age

Count of Payment	Column Labels					
Row Labels	18-24	25-29	30-34	35-39	40 or above	Grand Total
Card	10	3	5	1	2	21
Cash	106	60	43	27	24	260
Credit Card	222	169	97	62	32	582
E-Money	48	23	15	15	29	130
Grand Total	**386**	**255**	**160**	**105**	**87**	**993**

Observed	Age					
Payment	18-24	25-29	30-34	35-39	40 or above	Grand Total
Card	10	3	5	1	2	21
Cash	106	60	43	27	24	260
Credit Card	222	169	97	62	32	582
E-Money	48	23	15	15	29	130
Grand Total	386	255	160	105	87	993

Expected	Age					
Payment	18-24	25-29	30-34	35-39	40 or above	Grand Total
Card	8.163141994	5.392749245	3.383685801	2.220543807	1.839879154	21
Cash	101.0674723	66.7673716	41.89325277	27.49244713	22.77945619	260
Credit Card	226.2356495	149.4561934	93.77643505	61.5407855	50.90093656	582
E-Money	50.53373615	33.3836858	20.94662638	13.74622356	11.3897281	130
Grand Total	386	255	160	105	87	993

Chi-square test	Age						
Payment	18-24	25-29	30-34	35-39	40 or above	Grand Total	
Card	0.413327	1.0616568	0.772078658	0.670883943	0.013934983	2.931881434	
Cash	0.240729	0.6859236	0.029238346	0.008820756	0.065397838	1.030109161	
Credit Card	0.079301	2.5556678	0.110810045	0.003426637	7.072936793	9.822142319	
E-Money	0.12704	3.2297491	1.688212923	0.114355433	27.22818964	32.3875474	
Grand Total	0.860397	7.5329974	2.600339973	0.797486769	34.38045925	46.17168031	test statistics

p-value (0.000006481) < level of significance (0.05)

Payment and Age of the customers are not independent

Consultant C

Quantitative: Average expenditure (in dollars)

A regression analysis for the Average expenditure (in dollars) of customers (Y) with Probability to visit in next month (%)x_1, Average time to shop (in minutes)x_2, Price (Satisfaction Level)x_3, Choice (Satisfaction Level)x_4, Product Packing (Satisfaction Level) x_5.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.112544458
R Square	0.012666255
Adjusted R Square	0.007664564
Standard Error	49.65010364
Observations	993

ANOVA

	df	SS	MS	F	Significance F
Regression	5	31213.44632	6242.689264	2.53239472	0.027424092
Residual	987	2433086.065	2465.132791		
Total	992	2464299.511			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	83.72994122	7.827171134	10.69734388	2.40899E-25	68.37013226	99.08975018	68.37013226	99.08975018
Probability to visit in next month (%)	0.141528338	0.060147069	2.353002769	0.018817936	0.023495744	0.259560933	0.023495744	0.259560933
Average time to shop (in minutes)	-0.632730586	0.43377627	-1.458656524	0.144977811	-1.483960298	0.218499127	-1.4839603	0.218499127
Price(Satisfaction Level)	5.207234564	2.087679757	2.494268839	0.012784331	1.1104336	9.304035527	1.1104336	9.304035527
Choice (Satisfaction Level)	-0.086917332	1.711475656	-0.050785024	0.9595071	-3.445466502	3.271631838	-3.4454665	3.271631838
Product Packing (Satisfaction Level)	-2.144672102	1.947488179	-1.101250383	0.271056184	-5.966365262	1.677021058	-5.96636526	1.677021058

F-test:

test statistics (2.5324) > critical value (2.2232) → Reject the hypothesis → there is sufficient evidence that not all the coefficient(s) are equals to zero

T-test:

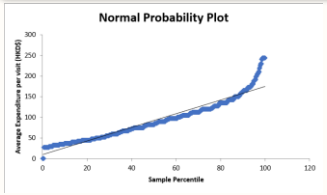
p-value (0.0188) < level of significance (0.05) → Reject hypothesis → there is sufficient evidence that the coefficient of Probability to visit in next month (%) is not equal to zero

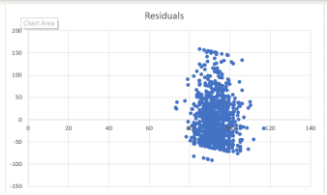
The regression line: $\hat{y}$ = 83.73 + 0.14x_1 + 5.21x_3

Consultant C

Quantitative: Average expenditure (in dollars)

Assumptions





	A	B	C	D	E	F	G
1	Residuals						
2	63.34715442	1.529406923	Durbin-Watson				
3	151.4734823						
4	-82.30875847						
5	-89.20484539						

X	VIF
Probability to visit in next month (%)	1.017864
Average time to shop (in minutes)	1.390146
Price(Satisfaction Level)	1.851298
Choice (Satisfaction Level)	1.322120
Product Packing (Satisfaction Level)	2.018254

Durbin-Watson = 1.5294, The DW is in between 1.5 to 2.5, so there is no possible violation in the assumptions over the independence of errors

Because majority of the points are on the line, the residuals are normally distributed.
Because majority of the points are concentration, there is a possible violation over the Homoscedasticity.
Because all VIFs are below 10, there is no possible violation over the multicollinearity.

Recommendation

Qualitative:
The most common way for customers to pay is by using credit card. Therefore, the pet shop should give some discounts to the customers if the customers are using credit card for the payment.

Quantitative:
The pet stop should make acceptable and reasonable goods price so that the customers can afford it to increase their average expenditure.

Consultant D

Qualitative--Chi Square independent test

Type of goods usually buy x Last time visiting the pet goods shop

count of Last time visiting the pet goods shop	Column Labe					
Row Labels	Accessories	Food (Daily)	Random	Snack	Toy	Grand total
1 week before	2	9		1	3	15
Last month	32	288	5	54	64	443
Last week	14	188	8	26	30	266
No visit within last 6 months	3	32		5	4	44
This week	15	143	9	21	37	225
Grand total	**66**	**660**	**22**	**107**	**138**	**993**

observed	Tyoe of goods usually buy					
Last time visiting the pet goods shop	Accessories	Food (Daily)	Random	Snack	Toy	Grand total
1 week before	2	9		1	3	15
Last month	32	288	5	54	64	443
Last week	14	188	8	26	30	266
No visit within last 6 months	3	32		5	4	44
This week	15	143	9	21	37	225
Grand total	66	660	22	107	138	993

expeced	Tyoe of goods usually buy					
Last time visiting the pet goods shop	Accessories	Food (Daily)	Random	Snack	Toy	Grand total
1 week before	0.996978852	9.96978852	0.332326	1.6163	2.08	15
Last month	29.44410876	294.4410876	9.814703	47.735	61.6	443
Last week	17.67975831	176.7975831	5.893253	28.663	37	266
No visit within last 6 months	2.924471299	29.24471299	0.974824	4.7412	6.11	44
This week	14.95468278	149.5468278	4.984894	24.245	31.3	225
Grand total	66	660	22	107	138	993

chi-square test	Tyoe of goods usually buy					
Last time visiting the pet goods shop	Accessories	Food (Daily)	Random	Snack	Toy	Grand total
1 week before	1.009100064	0.094333974	0.332326	0.235	0.4	2.07274956
Last month	0.221865738	0.140902922	2.361902	0.8222	0.1	3.64319216
Last week	0.76588271	0.709818215	0.75313	0.2473	1.31	3.7891375
No visit within last 6 months	0.001950638	0.259589024	0.974824	0.0141	0.73	1.98189584
This week	0.000137325	0.286605572	3.233985	0.4342	1.05	5.00540184
Grand total	1.998934475	1.491249708	7.656167	1.7529	3.59	16.4923769

critical value (26.2962) < test statistics (16.4923769)

Type of goods usually buy and Last time visiting the pet goods shop are independent.

Type of goods usually buy x Payment

count of Last time visiting the pet goods shop	Column Labels					
Row Labels	Accessories	Food (Daily)	Random	Snack	Toy	Grand total
Card		16		2	3	21
Cash	13	187	7	23	30	260
Credit Card	36	398	12	62	74	582
E-Money	17	59	3	20	31	130
Grand total	**66**	**660**	**22**	**107**	**138**	**993**

observed	Tyoe of goods usually buy					
Payment	Accessories	Food (Daily)	Random	Snack	Toy	Grand total
Card		16		2	3	21
Cash	13	187	7	23	30	260
Credit Card	36	398	12	62	74	582
E-Money	17	59	3	20	31	130
Grand total	66	660	22	107	138	993

expeced	Tyoe of goods usually buy					
Payment	Accessories	Food (Daily)	Random	Snack	Toy	Grand total
Card	1.395770393	13.95770393	0.465257	2.2628	2.92	21
Cash	17.28096677	172.8096677	5.760322	28.016	36.1	260
Credit Card	38.68277946	386.8277946	12.89426	62.713	80.9	582
E-Money	8.640483384	86.40483384	2.880161	14.008	18.1	130
Grand total	66	660	22	107	138	993

chi-square test	Tyoe of goods usually buy					
Payment	Accessories	Food (Daily)	Random	Snack	Toy	Grand total
Card	1.395770393	0.298829469	0.465257	0.0305	0	2.1927
Cash	1.060512222	1.165244597	0.266791	0.8981	1.04	4.4316
Credit Card	0.186059681	0.322671163	0.06202	0.0081	0.59	1.1645
E-Money	8.087686181	8.691931739	0.004986	2.5631	9.26	28.607
Grand total	10.73002848	10.47867697	0.799054	3.4998	10.9	36.395

p-value (0.000002645) < level of significance (0.05)

Type of goods usually buy and Payment are not independent.

Consultant D

Quantitative: Probability to visit in the next month

A regression analysis for the required to age(Y) with Gender x_1, Average time to shop(in minutes)x_2, Education Level x_3, Probability to visit in next month(%) x_4.

SUMMAR OUTPUT

Regression Statistics	
Multiple	0.265433007
R Square	0.070454681
Adjusted R Square	0.066687527
Standard Error	1.261031901
Observations	992

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	118.9620904	29.74052	18.70236	7.85302E-15
Ressidual	987	1569.528837	1.590201		
Total	991	1688.490927			

	Coefficiemts	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2.057534287	0.242695866	8.47783	8.28E-17	1.581275102	2.533793472	1.581275102	2.533793472
Gender	-0.24781452	0.082380088	-3.00818	0.002695	-0.409474767	-0.08615427	-0.40947477	-0.08615427
Average time to shop (in minutes)	-0.000359454	0.038294719	-0.00939	0.992513	-0.075507876	0.074788969	-0.07550788	0.074788969
Education Level	0.180713032	0.022767796	7.937221	5.59E-15	0.136034184	0.22539188	0.136034184	0.22539188
Probability to visit in next month (%)	0.038414062	0.035267736	1.089213	0.276326	-0.0307943	0.107622425	-0.0307943	0.107622425

F-test:

test statistics (18.70236) > critical value (2.38089) → Reject the hypothesis → there is sufficient evidence that not all the coefficient(s) are equals to zero

T-test:

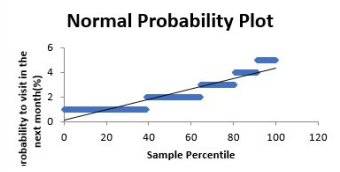
p-value (8.47783) < level of significance (0.05) → Reject hypothesis → there is sufficient evidence that the coefficient of age is not equal to zero.

The regression line: $\hat{y}$ = 2.058 $-0.2478\ x_1\ -0.0004x_2$ + $0.1807x_3$ + $0.038x_4$

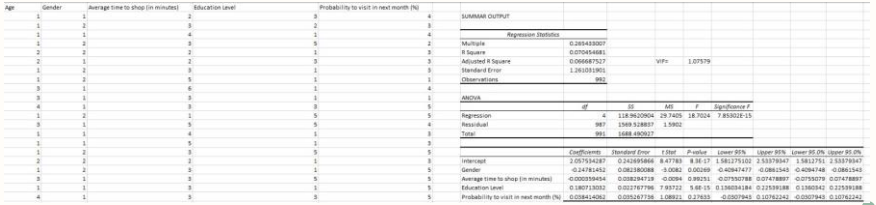
Consultant D

Quantitative: Probability to visit in the next month

Assumptions



Residuals	DW
-1.0375	1.791172
-1.14265	
-1.54122	
0.142858	
-0.46638	
-0.85678	
-0.77924	
0.858068	
0.972232	
1.457149	
-1.65718	
0.134856	
-1.10424	
-1.10388	
-1.65646	
0.142858	



X	VIF
Gender	2.159614
Average time to shop (in minutes)	2.47638
Fducation Level	2.38215
Quality (Satisfaction Level)	2.632947

Durbin-Watson = 1.7912, The DW is in between 1.5 to 2.5, so there is no possible violation in the assumptions overindependence of errors

There is a possible violation of the normality assumption on the residuals. This is because most of the points are not on the line.
Because majority of the points are concentration, there is a possible violation over the Homoscedasticity.
Because all VIFs are below 10, there is no possible violation over the multicollinearity.

Recommendation

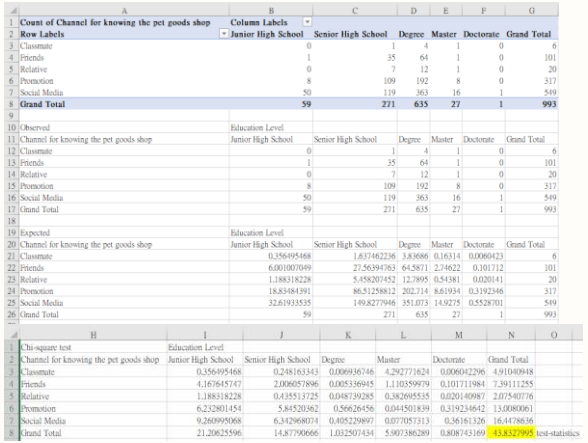
Qualitative:
We can see that not many buy random item in the pet shop that mean we can do more promote to our new products or we can make the new product better therefore we can let more people want to buy it.

Quantitative:
The pet stop should do more promote So that it can increase new customers' time to visit the pet's good shop in next month.

Consultant E

Qualitative--Chi Square independent test

Channel for knowing the pet goods shop x Education Level



Count of Channel for knowing the pet goods shop	Column Labels					
Row Labels	Junior High School	Senior High School	Degree	Master	Doctorate	Grand Total
Classmate	0	1	4	1	0	6
Friends	1	35	64	1	0	101
Relative	0	7	12	1	0	20
Promotion	8	109	192	8	0	317
Social Media	50	119	363	16	1	549
Grand Total	59	271	635	27	1	993

Observed	Education Level					
Channel for knowing the pet goods shop	Junior High School	Senior High School	Degree	Master	Doctorate	Grand Total
Classmate	0	1	4	1	0	6
Friends	1	35	64	1	0	101
Relative	0	7	12	1	0	20
Promotion	8	109	192	8	0	317
Social Media	50	119	363	16	1	549
Grand Total	59	271	635	27	1	993

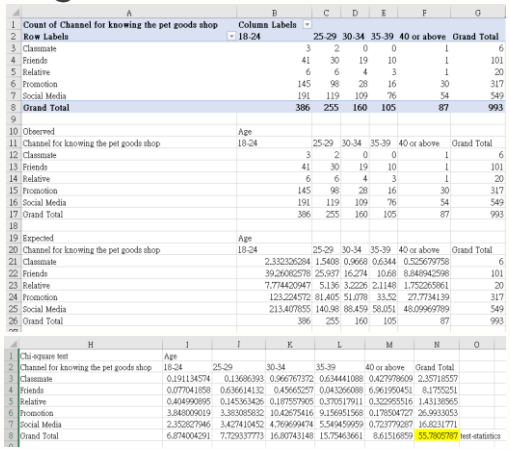
Expected	Education Level					
Channel for knowing the pet goods shop	Junior High School	Senior High School	Degree	Master	Doctorate	Grand Total
Classmate	0.356695468	1.637462236	3.83686	0.16314	0.0060423	6
Friends	6.001007049	27.56394763	64.5871	2.74622	0.101712	101
Relative	1.188318228	5.458207452	12.7895	0.54381	0.020141	20
Promotion	18.83484391	86.51258812	202.714	8.61934	0.3192346	317
Social Media	32.61933535	149.8277946	351.073	14.9275	0.5528701	549
Grand Total	59	271	635	27	1	993

Chi-square test	Education Level					
Channel for knowing the pet goods shop	Junior High School	Senior High School	Degree	Master	Doctorate	Grand Total
Classmate	0.356695468	0.248163343	0.006936746	4.292771624	0.006042296	4.91040948
Friends	4.167845747	2.006057896	0.005336945	1.110359979	0.101711984	7.39111255
Relative	1.188318228	0.435513725	0.048739285	0.382695535	0.020140987	2.07540776
Promotion	6.232801454	5.84520362	0.56626456	0.044501839	0.319234642	13.0080061
Social Media	9.260995068	6.342968074	0.405229897	0.077057313	0.36161326	16.4478836
Grand Total	21.20625596	14.87790666	1.032507434	5.907386289	0.808743169	43.8327995 test-statistics

critical value (26.2962) < test statistics (43.8328)

Channel for knowing the pet goods shop and Education Level of the customers are not independent

Channel for knowing the pet goods shop x Age



Count of Channel for knowing the pet goods shop	Column Labels					
Row Labels	18-24	25-29	30-34	35-39	40 or above	Grand Total
Classmate	3	2	0	0	1	6
Friends	41	30	19	10	1	101
Relative	6	6	4	3	1	20
Promotion	145	98	28	16	30	317
Social Media	191	119	109	76	54	549
Grand Total	386	255	160	105	87	993

Observed	Age					
Channel for knowing the pet goods shop	18-24	25-29	30-34	35-39	40 or above	Grand Total
Classmate	3	2	0	0	1	6
Friends	41	30	19	10	1	101
Relative	6	6	4	3	1	20
Promotion	145	98	28	16	30	317
Social Media	191	119	109	76	54	549
Grand Total	386	255	160	105	87	993

Expected	Age					
Channel for knowing the pet goods shop	18-24	25-29	30-34	35-39	40 or above	Grand Total
Classmate	2.332326284	1.5408	0.9668	0.6344	0.525679758	6
Friends	39.26082578	25.937	16.274	10.68	8.848942598	101
Relative	7.774420947	5.136	3.2226	2.1148	1.752265861	20
Promotion	123.2245721	81.405	51.078	33.52	27.7734139	317
Social Media	213.4078551	140.98	88.459	58.051	48.09969891	549
Grand Total	386	255	160	105	87	993

Chi-square test	Age					
Channel for knowing the pet goods shop	18-24	25-29	30-34	35-39	40 or above	Grand Total
Classmate	0.191134574	0.13686393	0.966767372	0.634441088	0.427978609	2.35718557
Friends	0.077041858	0.636614132	0.45665257	0.043266088	6.961950451	8.1755251
Relative	0.404990895	0.145363426	0.187557905	0.370517911	0.332955516	1.43138565
Promotion	3.848009019	3.383085832	10.42675416	9.156951568	0.178504727	26.9933053
Social Media	2.352827946	3.427410452	4.769699474	5.549459959	0.723779287	16.8231771
Grand Total	6.874004291	7.729377773	16.80743148	15.75463661	8.61516859	55.7805787 test-statistics

p-value (0.000002645) < level of significance (0.05)

Channel for knowing the pet goods shop and Age of the customers are not independent

Consultant E

Quantitative: Discount level (%) required to attract you to buy

A regression analysis for the rating on Discount level (%) required to attract you to buy (Y) with Price (Importance) x_1, Price (Satisfaction Level) x_2, Quality (Importance) x_3, Quality (Satisfaction Level) x_4, Product Availability (Satisfaction Level) x_5.

	I	J	K	L	M	N	O	P	Q
1	SUMMARY OUTPUT								
2									
3	Regression Statistics								
4	Multiple R	0.44370773							
5	R Square	0.19687655							
6	Adjusted R Square	0.192808042							
7	Standard Error	11.95530727							
8	Observations	993							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	5	34582.01706	6916.403411	48.39035754	7.49315E-45			
13	Residual	987	141071.2901	142.9293719					
14	Total	992	175653.3072						
15									
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	23.00072361	1.770571733	12.99056298	9.80048E-36	19.52620604	26.47524117	19.52620604	26.47524117
18	Price (Importance)	0.389083717	0.505203504	0.77015245	0.441393757	-0.602312687	1.38048012	-0.602312687	1.38048012
19	Price(Satisfaction Level)	1.012793237	0.581399894	1.741990748	0.081821569	-0.128128707	2.153715181	-0.128128707	2.153715181
20	Quality (Importance)	1.764520923	0.592161248	2.979798032	0.002955042	0.602481216	2.926560631	0.602481216	2.926560631
21	Quality (Satisfaction Level)	3.229926047	0.634737714	5.088599552	4.31716E-07	1.984335544	4.475516549	1.984335544	4.475516549
22	Product Availability (Satisfaction Level)	0.955841886	0.428081741	2.232848996	0.025782486	0.115786948	1.795896824	0.115786948	1.795896824

F-test:

test statistics (48.3904) > critical value (2.2232) → Reject the hypothesis → there is sufficient evidence that not all the coefficient(s) are equals to zero

T-test:

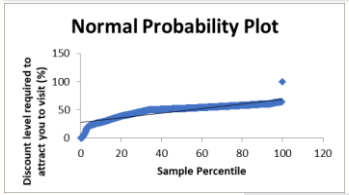
p-value (0.0258) < level of significance (0.05) → Reject hypothesis → there is sufficient evidence that the coefficient of Product Availability (Satisfaction Level) is not equal to zero

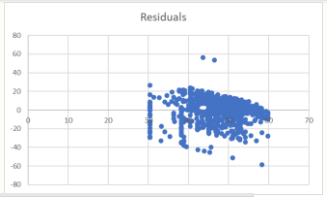
The regression line: $\hat{y} = 23 + 1.76x_3 + 3.23x_4 + 0.96x_5$

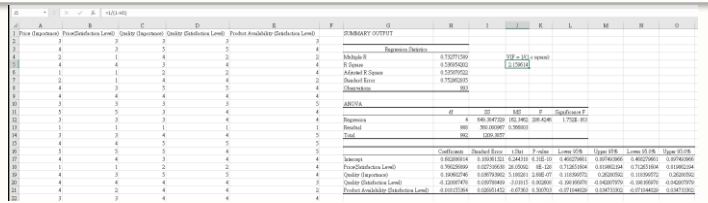
Consultant E

Quantitative: Discount level (%) required to attract you to buy

Assumptions







B2 =SUMXMY2(A3:A994,A2:A993)/SUMSQ(A2:A994)

	A	B	C	D	E	F	G
1	Residuals						
2	-23.05722103	1.981144358	Durbin-Watson				
3	3.608959427						
4	2.778696168						
5	-22.64486592						

X	VIF
Price (Importance)	2.159614
Price (Satisfaction Level)	2.47638
Quality (Importance)	2.38215
Quality (Satisfaction Level)	2.632947
Product Availability (Satisfaction Level)	1.280433

Durbin-Watson = 1.9811, The DW is in between 1.5 to 2.5, so there is no possible violation in the assumptions over the independence of errors

Because majority of the points are on the line, the residuals are normally distributed.
Because majority of the points are concentration, there is a possible violation over the Homoscedasticity.
Because all VIFs are below 10, there is no possible violation over the multicollinearity.

Recommendation

Qualitative:
The most common way customers of all ages and education levels get to know pet goods shop is through social media. Therefore, the pet shop should increase their advertising on social media to increase customer sources.

Quantitative:
The pet stop should strengthen product quality and ensure that each product has sufficient inventory to increase the attractiveness of discounts.

Thanks!