Paper Mill Data Analysis

Course: Data Analysis

By: Ghufran Amjad

Student ID: 7056988

Case Study Paper Mill Corporation

PaperMill Corporation is a manufacturer of printing paper located in the USA. The company sells its products worldwide to two types of customers: the newsprint industry and the magazine industry. Paper products are sold either directly to the customer or indirectly through a broker. To develop a marketing plan for the next year, the company conducted a market segmentation study (n = 200) in which respondents included purchasing managers of firms buying from PaperMill. Two types of information were collected in the survey. First, information about how customers perceive the performance of PaperMill in relation to various attributes (e.g., product quality, complaint resolution). These attributes are considered to be most influential in the selection of suppliers in the paper industry. Second, information related to purchase outcomes (e.g., satisfaction with PaperMill, likelihood of future purchase) and measures of business relationships (strategic alliance). A third type of information was taken from PaperMill's data warehouse and included customer attributes such as firm size, length of customer relationship and type of distribution.

Your task is to support PaperMill's management in developing a better understanding of both the characteristics of its customers and the relationships between their perceptions of PaperMill, and their actions toward PaperMill. To help you with this task, PaperMill has prepared several questions that it would like to see answered based on appropriate analyses of the data contained in the *PaperMill.sav* file:

1. Do the following perceptions and purchase outcomes differ between customers served directly and those served by a broker (*x5*)?
   - *product quality* (x*6*)
   - *complaint resolution* (*x9*)
   - *new products* (*x15*)
   - *price flexibility* (*x17*)
   - *likelihood of future purchases* (*x21*)

What can be deduced from the analysis results?

2. Do the various *customer types* (*x1*) differ regarding their likelihood to recommend PaperMill to other firms (*x*20)? If so, use post-hoc tests to analyze if an increasing duration of the relationship with PaperMill leads to higher recommendation levels. Interpret the results.
3. PaperMill's management suspects that the type of *distribution system* (*x5*) influences *customer satisfaction* (*x19*), but that this effect also depends on the *firm size* (*x3*). Is the management right? If the answer is yes, what can be deduced from the analysis? If there is an interaction between *distribution system* and *firm size*, what type of interaction is it and how can this interaction be interpreted? What are the managerial implications?
4. PaperMill's management has long been interested in identifying the factors that lead to increased customer satisfaction for use in their marketing campaigns. The 13 performance perception variables *x6* to *x18* contained in the data set have been developed through focus groups as possible drivers of *customer satisfaction* (*x19*). Based on the estimation of a regression model with all 13 variables, PaperMill's management wants to decide on the main influences on customer satisfaction. How do you assess this approach given the regression results (think about assumptions which might be violated)?
5. Do the variables *x6* to *x18* each represent rather independent dimensions of the customer performance perception or is it possible to reproduce the data structure for these variables by a much smaller number of components? If so, what kind of perceptual dimensions might emerge? Choose an appropriate measure to answer this question. Try to describe these dimensions and assign a meaningful name ("umbrella term") to the different dimensions to help PaperMill develop its marketing campaigns.
6. What are the consequences of the preceding interdependence analysis for a regression analysis used to identify the most important drivers of *customer satisfaction* (*x19*)? Try to develop a valid alternative model to the model used

by PaperMill's management (the model estimated in 4.) including only those performance perception variables which have a strong influence on customer satisfaction. What management implications can be derived from this model?

Answer 1

First step is to formulate the hypothesis such as null hypothesis and alternate hypothesis so in this case the hypothesis are following

- H_0 = perceptions & purchase outcomes(x5) **are different** B/w customers served by direct and served by a broker
- H_1 = perceptions & purchase outcomes **are Not different** B/w customers served by direct and served by a broker

We will find the effect on each customer group, by performing independent sample T test because the question is involve with comparison of two independent groups and T test are suitable to compare the means of two group to find differences.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
X6 - Product Quality	Equal variances assumed	2.587	.109	-5.819	198	<.001	<.001	-1.0578	.1818	-1.4163	-.6993
	Equal variances not assumed			-5.853	196.430	<.001	<.001	-1.0578	.1807	-1.4142	-.7014
X9 - Complaint Resolution	Equal variances assumed	3.449	.065	-2.851	198	.002	.005	-.4809	.1687	-.8135	-.1483
	Equal variances not assumed			-2.895	197.821	.002	.004	-.4809	.1661	-.8085	-.1533
X15 - New Products	Equal variances assumed	.756	.386	-.410	198	.341	.682	-.0872	.2127	-.5067	.3322
	Equal variances not assumed			-.411	194.081	.341	.682	-.0872	.2124	-.5061	.3316
X17 - Price Flexibility	Equal variances assumed	.753	.387	3.122	198	.001	.002	.5171	.1656	.1905	.8437
	Equal variances not assumed			3.137	195.982	<.001	.002	.5171	.1648	.1920	.8421
X21 - Likely to Purchase	Equal variances assumed	2.832	.094	-6.140	198	<.001	<.001	-.7150	.1164	-.9446	-.4853
	Equal variances not assumed			-6.222	197.994	<.001	<.001	-.7150	.1149	-.9416	-.4884

Here we can analyze that F value represents equality of variance assumptions b/w the two groups. If F value is not significant then it indicate that variance is same or in other case significant F value represents equal variance assumption is violated.

In this case the Assumptions is violated in variable x6 x9 x15 x17 x21 because significance is higher than the typical **alpha value of 0.05**. This equal variance test is called **levene's test** But welch's test Also providing the same results so we can assume that equal variance assumption is true and we can go ahead with T test to compare the means differences.

The mean difference in each out comes except **one X15** with its corresponding p value which is less than 0.05 implies that we there are difference B/w customers served by direct and served by a broker hence Rejecting the null hypothesis

While in one out comes X15 new products we fail to reject the null hypothesis because the corresponding p value is more than 0.05 which illustrate new products has no difference on customer served by a broker or direct served.

The one sided p value tells us about the direction of difference in perception and purchase out comes of the customer like it product quality increase or decrease of a customer served by broker or direct. While two sided p value tells us there is significant difference in specific outcome variable in case of broker dealing or direct deals. According to my hypothesis I will consider the two sided p value because we are interested to know if the outcome variables are different or not.

Cohen's d, hedges' correction, glass's delta indicate the effect size or in our case it quantify the difference, which can be interpreted in effect size value typically more than 0.8 designate as high effect 0.5-0.8 considered as moderate and below 0.5 considered small

Independent Samples Effect Sizes

				95% Confidence Interval	
		Standardizer[a]	Point Estimate	Lower	Upper
X6 - Product Quality	Cohen's d	1.2813	-.826	-1.114	-.535
	Hedges' correction	1.2861	-.822	-1.110	-.533
	Glass's delta	1.2309	-.859	-1.162	-.553
X9 - Complaint Resolution	Cohen's d	1.1889	-.404	-.685	-.123
	Hedges' correction	1.1934	-.403	-.682	-.123
	Glass's delta	1.0625	-.453	-.737	-.166
X15 - New Products	Cohen's d	1.4991	-.058	-.336	.220
	Hedges' correction	1.5048	-.058	-.335	.219

	Glass's delta	1.4835	-.059	-.337	.220
X17 - Price Flexibility	Cohen's d	1.1673	.443	.161	.724
	Hedges' correction	1.1717	.441	.160	.721
	Glass's delta	1.1291	.458	.171	.743
X21 - Likely to Purchase	Cohen's d	.8208	-.871	-1.161	-.579
	Hedges' correction	.8239	-.868	-1.157	-.577
	Glass's delta	.7449	-.960	-1.269	-.647

a. The denominator used in estimating the effect sizes.

Cohen's d uses the pooled standard deviation.

Hedges' correction uses the pooled standard deviation, plus a correction factor.

Glass's delta uses the sample standard deviation of the control (i.e., the second) group.

In our case the product quality X6, new product x15 and Complaint Resolution X9 indicate high effects because the values are more than 0.8 which illustrate that customer served directly perceive higher product quality and high likely hood of future purchase. While the effect size value is lowest for the X21 but still the effects comes in the high effect zone.

Cohen's calculation is vague especially if sample size is small so hedges correction uses. But in this case the results are showing almost the same results as of cohen's d.

Answer 2

In this Case we have to find if the customer's types are different in their likely hood to recommend paper mill to other firms so we set our hypothesis on bases of our question.

- H_0 Various customer type differ in their likely hood of recommendation to other firms.
- H_1 is various customer type not differ in their likely hood of recommendation to other firms.

I will select one way Anova to answer the question because the question involve different customer types of customer which mean we have more than two groups to analyze and find the variance b/w more than two customer type groups.

Descriptives

X20 - Likely to Recommend

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Less than 1 year	68	6.141	.9949	.1207	5.900	6.382	4.0	9.4
1 to 5 years	64	7.209	.7144	.0893	7.031	7.388	5.6	9.1
Over 5 years	68	7.522	.9761	.1184	7.286	7.758	5.6	9.9
Total	200	6.953	1.0829	.0766	6.802	7.103	4.0	9.9

As it can be seen in the table that sample size N is nearly similar to each other from its mean difference Comparisons so I will perform post hoch test to find which customer group is different in recommendations to the firms.

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
X20 - Likely to Recommend	Based on Mean	2.951	2	197	.055
	Based on Median	2.720	2	197	.068
	Based on Median and with adjusted df	2.720	2	178.833	.069

	Based on trimmed mean	2.804	2	197	.063

Homogeneity of variance indicated that there is no significance difference in variances between the groups in likely to recommend yes we can see that the values are very close to the typical thresh hold value. The variances are assumed to be equal (homogeneous) for the variable "X20 - Likely to Recommend" based on the Levene's test at the 0.05 significance level.

ANOVA

X20 - Likely to Recommend

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	71.043	2	35.521	43.112	<.001
Within Groups	162.316	197	.824		
Total	233.359	199			

As we can see the variance B/w the groups can assume equall because high F value and corresponding significance value (p) which is less than the alpha value of 0.05. which implies rejecting the null hypothesis.

ANOVA Effect Sizes[a]

		Point Estimate	95% Confidence Interval Lower	Upper
X20 - Likely to Recommend	Eta-squared	.304	.199	.394
	Epsilon-squared	.297	.191	.387
	Omega-squared Fixed-effect	.296	.190	.386
	Omega-squared Random-effect	.174	.105	.239

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

Eta square is the measure of variation or effects by independent variable in this case the Eta squared value is very high illustrating high impact .so we can identify which customer type is more significant in likeliness to recommend by interpreting the post Hoch test. Post Hoch test need to perform to identify significance of variance b/w different types of customer.

Multiple Comparisons

Dependent Variable: X20 - Likely to Recommend

	(I) X1 - Customer Type	(J) X1 - Customer Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Tukey HSD	Less than 1 year	1 to 5 years	-1.0682*	.1581	<.001	-1.442	-.695
		Over 5 years	-1.3809*	.1557	<.001	-1.749	-1.013
	1 to 5 years	Less than 1 year	1.0682*	.1581	<.001	.695	1.442
		Over 5 years	-.3127	.1581	.120	-.686	.061
	Over 5 years	Less than 1 year	1.3809*	.1557	<.001	1.013	1.749
		1 to 5 years	.3127	.1581	.120	-.061	.686
Games-Howell	Less than 1 year	1 to 5 years	-1.0682*	.1501	<.001	-1.424	-.712
		Over 5 years	-1.3809*	.1690	<.001	-1.781	-.980
	1 to 5 years	Less than 1 year	1.0682*	.1501	<.001	.712	1.424

	Over 5 years	-.3127	.1483	.092	-.664	.039
Over 5 years	Less than 1 year	1.3809*	.1690	<.001	.980	1.781
	1 to 5 years	.3127	.1483	.092	-.039	.664

*. The mean difference is significant at the 0.05 level.

Both Tukey HSD and Games-Howell tests reveal significant differences in the "Likely to Recommend" scores across different customer types. Both the test are giving similar results so we can also sure of equality of variance assumption is not violated in this case

Less than 1 year vs. 1 to 5 years:

Customers who have been with the company for less than 1 year are significantly less likely to recommend compared to those who have been customers for 1 to 5 years. On average, the difference in likelihood to recommend between these two groups is around -1.07, meaning those with less than 1 year of experience are less likely to recommend.

Less than 1 year vs. over 5 years:

Similar to the previous comparison, customers with less than 1 year of experience are significantly less likely to recommend compared to those with over 5 years of experience. The difference in likelihood to recommend between these two groups is even larger, around -1.38.

1 to 5 years vs. over 5 years:

There is no significant difference in likelihood to recommend between customers with 1 to 5 years of experience and those with over 5 years of experience. The average difference in likelihood to recommend between these two groups is not statistically significant.

Answer 3

In this case the dependent variable is customer satisfaction and independent variable are distribution systems and the firm size. Because we are interested to find out the effects on Customer Satisfaction.

The null hypothesis we will set as per our question is

- H_0 Distribution System and Firm size influence customer satisfaction.

In this question we need to find out the effect of distribution system and firm size on customer satisfaction. There are three variables which are correlated to each other so we need to find the correlation between the variables to deduce the suitable conclusion. We will also conduct ANOVA to find the interaction B/W the Variables.

Correlations

		X5 - Distribution System	X3 - Firm Size	X19 - Satisfaction
X5 - Distribution System	Pearson Correlation	1	-.159*	.549**
	Sig. (2-tailed)		.025	<.001
	N	200	200	200
X3 - Firm Size	Pearson Correlation	-.159*	1	.194**
	Sig. (2-tailed)	.025		.006
	N	200	200	200
X19 - Satisfaction	Pearson Correlation	.549**	.194**	1
	Sig. (2-tailed)	<.001	.006	
	N	200	200	200

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

As we can see that the dependent variable which is customer satisfaction is in high correlation with distribution and also in correlation with firm size but this correlation is Ruther weak which provide us the answer of our first part of the Question by

rejecting our null hypothesis because in both the case the significance is less than the typical value of 0.05.

The Correlation only suggest that the variables are in relation but to find what kind of interaction it is we need to perform Two way ANOVA by performing univariate because we have only one dependent variable for which we are interested to find the interaction.

Tests of Between-Subjects Effects

Dependent Variable: X19 - Satisfaction

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	132.472[a]	3	44.157	49.721	<.001
Intercept	9623.346	1	9623.346	10835.926	<.001
x5	106.036	1	106.036	119.397	<.001
x3	27.810	1	27.810	31.314	<.001
x5 * x3	15.326	1	15.326	17.258	<.001
Error	174.067	196	.888		
Total	9972.600	200			
Corrected Total	306.539	199			

a. R Squared = .432 (Adjusted R Squared = .423)

The significant **F-test (F = 17.258)** indicates that the interaction effect is statistically significant, suggesting that the relationship between distribution system and customer satisfaction differs depending on the level of firm size, and vice versa. A larger F-value suggests a **stronger interaction effect**. So we can analyze that F values of individual variable and their combine effect both have significance in their interaction. The significant interaction effect implies that the impact of distribution

system on customer satisfaction varies depending on the level of firm size, and vice versa

As it can be conclude with the help of this graph that customer satisfaction is dual function of firm size and distribution system as well but firm size effect is not that strong as of distribution system and the interaction is **ordinal interaction** and the direction of the effects is positive so it implies that both the variables have positive effects towards enhancing customer satisfaction.

Keeping in the sight the above results managerial implication could be

- PaperMill may consider focusing resources on specific combinations of distribution systems and firm sizes that resulted the highest customer satisfaction
- Certain combinations of distribution systems and firm sizes are less effective in improving customer satisfaction, potentially leading to lower overall satisfaction levels because their combine effect is less than the individual effects.

- Paper mill may consider to focus on new distribution channels to increase the customer satisfaction as it have the most individual effect proven above.

Answer 4

In this Question we have a motive to identify the factors the lead to increase Customer satisfaction based on regression analysis. Which have prior assumption like multicollinearity of the variables because it make difficult to explain the individual effect on the dependent variable.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	-2.255	.789		-2.859	.005		
	X6 - Product Quality	.412	.038	.459	10.879	<.001	.592	1.690
	X7 - E-Commerce	-.264	.087	-.163	-3.044	.003	.365	2.739
	X8 - Technical Support	.031	.046	.041	.669	.505	.282	3.551
	X9 - Complaint Resolution	.099	.073	.097	1.366	.173	.210	4.764
	X10 - Advertising	-.034	.047	-.031	-.714	.476	.548	1.825
	X11 - Product Line	.347	.186	.368	1.862	.064	.027	37.180
	X12 - Salesforce Image	.629	.068	.572	9.194	<.001	.272	3.677
	X13 - Competitive Pricing	-.064	.033	-.082	-1.965	.051	.611	1.637
	X14 - Warranty & Claims	.002	.089	.002	.026	.979	.266	3.761
	X15 - New Products	.033	.028	.040	1.185	.237	.915	1.093
	X16 - Order & Billing	.030	.073	.022	.413	.680	.363	2.755
	X17 - Price Flexibility	.263	.193	.252	1.358	.176	.031	32.757
	X18 - Delivery Speed	.008	.375	.005	.022	.982	.021	48.712

a. Dependent Variable: X19 - Satisfaction

Tolerance values close to 1 and VIF values below 10 generally **indicate no multicollinearity concerns**. In this Case the higher VIF values for the variables like X17, X18 ,X11 shows us the problem of multicollinearity .

The Beta values suggest the effect size on dependent variable with increase one degree in independent variables, In case of sales force image, E commerce and product Quality the beta coefficient is relatively large implies the **high effect** of changes with its significance value less than 0.05 explaining the significance of these variable on dependent variable which is also highly significant.

One degree increase in each of these (x6, x12) variables will have **impact of 0.412, 0.629 unit increase in customer satisfaction** while one unit increase in E commerce will result **decrease in customer satisfaction by 0.264**.

Model Summary[b]

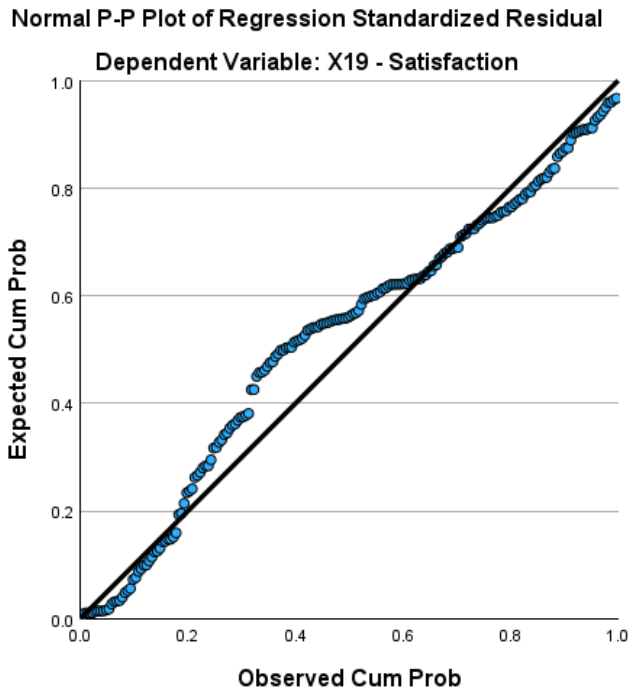
					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.897[a]	.804	.790	.5681	.804	58.756	13	186	<.001

a. Predictors: (Constant), X18 - Delivery Speed, X13 - Competitive Pricing, X8 - Technical Support, X15 - New Products, X7 - E-Commerce, X6 - Product Quality, X10 - Advertising, X16 - Order & Billing, X17 - Price Flexibility, X12 - Salesforce Image, X14 - Warranty & Claims, X9 - Complaint Resolution, X11 - Product Line

b. Dependent Variable: X19 - Satisfaction

R is correlation coefficient which suggest the strength and direction of the relationships between dependent and independent variable in this case the relationship is strong and direction is positive.

Adjusted R^2 value suggest the variation explained by independent variables which in this case around 79%. Higher the R^2 value more the model fit.



we see that the residuals are randomly scattered around the **zero line, and their spread remains relatively constant across the range of predicted values**. This indicates that the model fits the data well.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	246.511	13	18.962	58.756	<.001[b]
	Residual	60.028	186	.323		
	Total	306.539	199			

a. Dependent Variable: X19 - Satisfaction

b. Predictors: (Constant), X18 - Delivery Speed, X13 - Competitive Pricing, X8 - Technical Support, X15 - New Products, X7 - E-Commerce, X6 - Product Quality, X10 - Advertising, X16 - Order & Billing, X17 - Price Flexibility, X12 - Salesforce Image, X14 - Warranty & Claims, X9 - Complaint Resolution, X11 - Product Line

Analysis of Variance is also suggesting us the model is successfully explaining the Variance by its **Higher F value and corresponding lower significance value** ultimately expressing the higher statistical model significance.

Model	Dimension	Eigenvalue	Condition Index
1	1	13.471	1.000
	2	.133	10.066
	3	.098	11.719
	4	.082	12.849
	5	.081	12.915
	6	.050	16.414
	7	.028	21.904
	8	.019	26.852
	9	.013	32.002
	10	.009	39.590
	11	.008	40.842
	12	.006	48.967
	13	.003	63.342
	14	.000	190.246

Larger eigenvalues indicate more variance explained by the corresponding set of variables.**Condition Index** In our case also indicates the extent of multicollinearity in the model. A condition index greater than 30 suggests potential multicollinearity issues. Which can be Analyze in last 6 dimension. As the condition Index value is high with **corresponding less Eigen Values**.

Conclusively the factors like sales force and product quality have high positive impact on customer satisfaction but actually individual effect is difficult to know because of **multicollinearity of different factors**.

Answer 5

We are interested to find if some of these variables have same dimensions and can put to gather under a new factor and this we can check with The KMO measure which evaluates whether the variables x6 to x18 are suitable for factor analysis. It determines whether these variables collectively capture underlying dimensions or factors in the data related to customer performance perception.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.632
Bartlett's Test of Sphericity	Approx. Chi-Square	2020.335
	df	78
	Sig.	<.001

To determine the dimensions of the factor we analyze If the KMO measure indicates that the variables have high intercorrelations (i.e., the data is suitable for factor analysis), which we already know in previous answer that the data is intercorrelated thus we interpret the KMO results as high KMO value (close to 1) suggests that the variables are highly interrelated and can be grouped into fewer dimensions or factors. On the other hand, a low KMO value (closer to 0) indicates that the variables are relatively independent and may not represent underlying dimensions well. In this Case the KMO value is **0.632 which means we can do the grouping of the variables** however the value is not so high still it is significant.

Bartletts test of spheicity compare the data matrix with idealized identity matrix to find the difference and assess that difference is significant or not In our case this difference is **significant (<0.001)** which implies that variables are not independent may group together.

Communalities

	Initial	Extraction
X6 - Product Quality	1.000	.650
X7 - E-Commerce	1.000	.799
X8 - Technical Support	1.000	.915
X9 - Complaint Resolution	1.000	.863
X10 - Advertising	1.000	.636
X11 - Product Line	1.000	.844
X12 - Salesforce Image	1.000	.868
X13 - Competitive Pricing	1.000	.614
X14 - Warranty & Claims	1.000	.911
X15 - New Products	1.000	.097
X16 - Order & Billing	1.000	.767
X17 - Price Flexibility	1.000	.875
X18 - Delivery Speed	1.000	.936

Extraction Method: Principal Component Analysis.

These communalities represent the proportion of variance in each variable that is accounted for by the underlying factors extracted during principal component analysis (PCA).

Initial values are 1.000 for each variable because they represent the total variance of each variable before the extraction of common factors. After extraction these values represent the proportion of variance in each variable that is explained by the extracted factors. Higher values indicate that a larger proportion of the variance in the variable is accounted for by the underlying factors.

• **High Communalities (0.7 - 1.0)** all variables with highest values in highlighted in Red this suggests that the PCA factors explain a significant amount of each variable's variance.

• **Moderate Communalities (0.4 - 0.7)**: The extracted factor explain the variance of original variable up to 70 percent so PCA is also well suited in this case.

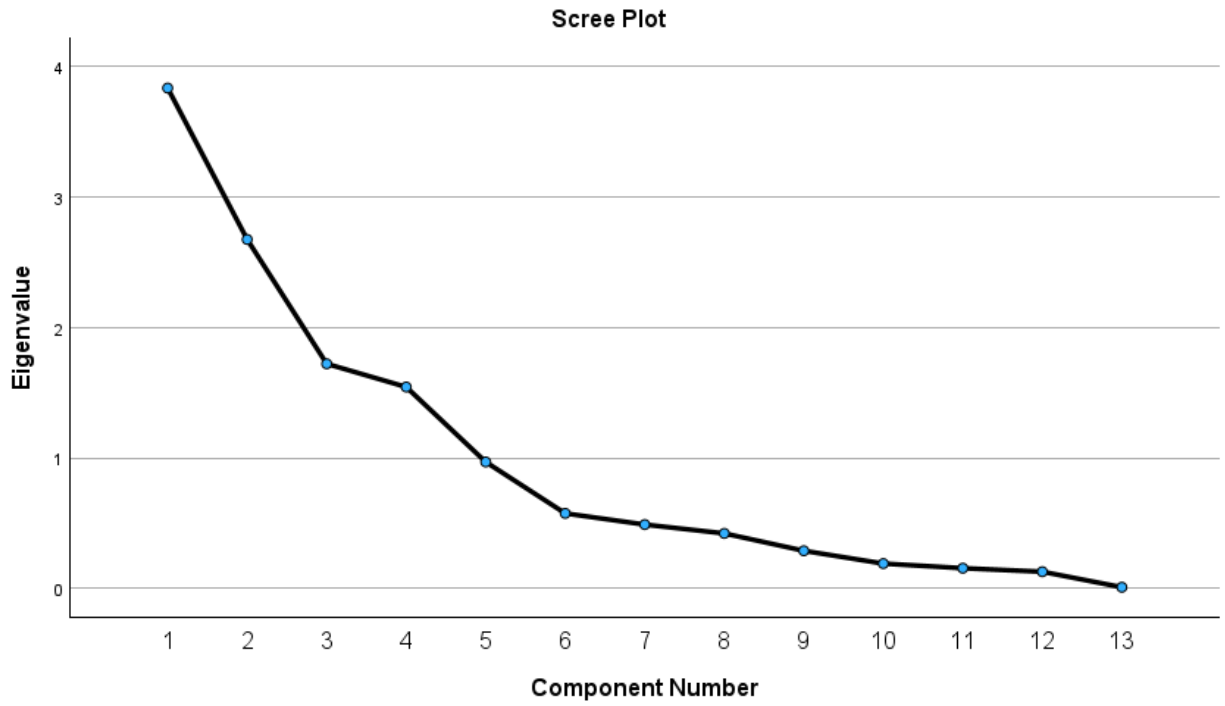
• **Low Communalities (< 0.4)**: X15 - New Products (0.097) has a low communality, indicating that the extracted components do not explain much of the variance.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.835	29.503	29.503	3.835	29.503	29.503
2	2.675	20.577	50.080	2.675	20.577	50.080
3	1.722	13.243	63.322	1.722	13.243	63.322
4	1.544	11.876	75.198	1.544	11.876	75.198
5	.969	7.455	82.653			
6	.575	4.422	87.075			
7	.489	3.763	90.838			
8	.421	3.242	94.080			
9	.288	2.218	96.298			
10	.190	1.462	97.760			
11	.155	1.190	98.950			
12	.128	.983	99.933			
13	.009	.067	100.000			

Extraction Method: Principal Component Analysis.

Based on this Data total four components are extracted which are explaining almost total variances. The Eigen values tells us the variation explained by the extracted components individually and the cumulative illustrate the total variance explained. In our case the four extracted components successfully explaining **the 75 % of total variance**. While principal component **factor 1** alone is explaining **29.50 % of Variance**.



Scree plot indicating that the first three or four are the most significant as their Eigan values are also more than 1.I have selected 4 Components.

By assigning the umbrella term to each of this component to develop the marketing campaigns paper mill Management implication could be

Component 1 might represent an overall **Services** dimension, given its significant contribution to the variance explained.

Component 2 could represent **Product and Value**, emphasizing aspects like pricing and delivery.

Component 3 might focus on **Sales and Channels**, related to new products and product line expansion.

Component 4 could cover **Customer feedback**, relating to aspects like warranties and claims.

For PaperMill's marketing campaigns, instead of focusing all performance variables it might be more helpful to emphasis on these 4 dimensions which already accounted for covering 75 % of variances which explain the influence on Customer's satisfaction.

Answer6

In this question we are interested to apply alternative model with the variables which have been extracted in previous question in principal component analysis so firstly we will identify the factors with high loading in each principal components after the **Varimax rotation** which implies to maximize the variance of the squared loadings within each factor

Rotated Component Matrix[a]

	Component			
	1	2	3	4
X6 - Product Quality	.064	.803	-.008	.039
X7 - E-Commerce	.062	-.047	.889	.047
X8 - Technical Support	.052	.045	.010	.954
X9 - Complaint Resolution	.912	.040	.133	.105
X10 - Advertising	.183	-.041	.774	-.038
X11 - Product Line	.555	.716	.078	.132
X12 - Salesforce Image	.134	-.127	.908	.096
X13 - Competitive Pricing	-.088	-.760	.165	-.038
X14 - Warranty & Claims	.103	.084	.071	.943
X15 - New Products	.232	.172	-.025	-.112
X16 - Order & Billing	.855	.000	.158	.103
X17 - Price Flexibility	.515	-.768	.133	-.046
X18 - Delivery Speed	.943	.015	.198	.084

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Variables with high positive loadings or absolute values close to 1 indicate a strong association with Component 1. In this case, variables such as **X9 (Complaint**

Resolution), **X18 (Delivery Speed)**, and **X16 (Order & Billing)** have the highest loadings on Component 1 **(Services)**.

Similar to Component 1, but different variables may have high loadings. Variables with high positive or negative loadings on Component 2 are strongly associated with this component. In this, variables like **X6 (Product Quality)**, **X11 (Product line)**, and **X17 (price flexibility)** have high loadings on Component. 2 (Sales Channels)

The component 3 which have umbrella term sales and channels have high loading of **X7 (E-Commerce) , X10 (Advertising)**, **X12 (Salesforce Image).**

The component 4 termed as **Customer feedback** have high loading of **X8 - Technical Support and X14 - Warranty & Claims.**

After Conducting the regression Analysis based on our extracted factors it is quite clear that these 4 factors are enough to explain the change or variance in Customer satisfaction as it can be analyze in correlations that all four factors are in good correlations with the parameter X19-Satisfaction and this correlation is also significant because corresponding significance values are below 0.05.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.839[a]	.704	.698	.6816	.704	116.194	4	195	<.001

a. Predictors: (Constant), REGR factor score 4 for analysis 1, REGR factor score 3 for analysis 1, REGR factor score 2 for analysis 1, REGR factor score 1 for analysis 1

b. Dependent Variable: X19 - Satisfaction

Although the model fitness is not showing significantly different than our last analysis as the **R2 value is somehow same** but overall the model fitness is still significant.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	215.940	4	53.985	116.194	<.001[b]
	Residual	90.599	195	.465		
	Total	306.539	199			

a. Dependent Variable: X19 - Satisfaction

b. Predictors: (Constant), REGR factor score 4 for analysis 1, REGR factor score 3 for analysis 1, REGR factor score 2 for analysis 1, REGR factor score 1 for analysis 1

The high **F value** in this model can also indicates that the explained variance by the regression model is significantly higher than the unexplained variance, further supporting the model's significance.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	6.952	.048		144.238	<.001		
	REGR factor score 1 for analysis 1	.690	.048	.556	14.271	<.001	1.000	1.000
	REGR factor score 2 for analysis 1	.558	.048	.449	11.544	<.001	1.000	1.000
	REGR factor score 3 for analysis 1	.511	.048	.412	10.579	<.001	1.000	1.000
	REGR factor score 4 for analysis 1	.193	.048	.155	3.993	<.001	1.000	1.000

a. Dependent Variable: X19 - Satisfaction

VIF values are also clearly describing that new factors have no problem of multicollinearity and Beta Coefficients is interpreting the relationship of dependent and independent variables in our case customer satisfaction and new regression factors. All the new factors are in positive relationships and strength is strongest in 1st component.

So after verification of the new factors the managerial implication could be

- Focusing on new dimensions could lead the paper mill towards achieving the customer satisfaction targets.
- Dimension 1 which I have given the umbrella term 'Services' might have strong connection with the Customer satisfaction as improving the services lead to improve the factors like Delivering speed, complaint resolution, Order and billing.
- The Dimension 4 termed as Customer feedback has relatively less impact in analyzing the Customer Satisfaction but still the impact is significant to improving those factors could also have positive impact on customer satisfaction.
- Regressed factor 2 is very important like factor 1 as it shows that if the products line, product quality and price flexibility improve or change could lead to increase or decrease the customer satisfaction so new channels and new product addition to the portfolio is imminent to keep the customer happy and satisfied.