FINANCIAL STATEMENT ANALYSIS: PRINCIPAL COMPONET APPROCH CASE STUDY ON INDIAN TELICOM COMPANEY

USING SPSS

A NAGASHESHU

SKOLAR MAJOR PROJECT

FINANCIAL STATEMENT ANALYSIS PRINCIPAL COMPONET APPROCH CASE STUDY ON INDIAN TELICOM COMPANEY

INTRODUCTION

The dynamic landscape of the telecommunications industry in India has witnessed unprecedented growth and evolution over the past decade. As the sector continues to navigate through regulatory changes, technological advancements, and intense market competition, a comprehensive understanding of the financial health and performance of telecom companies becomes imperative for stakeholders, investors, and industry analysts.

This major project has a detailed exploration of financial statement analysis, focusing on a Indian telecom company. The chosen methodology for this analysis is the Principal Component Approach (PCA), a sophisticated technique widely employed in various domains for dimensionality reduction and extracting essential patterns from complex datasets.

Financial statement analysis has evolved beyond traditional approaches, with advanced techniques like PCA gaining prominence. This project stands at the intersection of traditional financial analysis and cutting-edge methodologies, showcasing the evolution in analytical tools and methodologies used to extract meaningful insights from complex financial data.

Stakeholders, including investors, policymakers, and industry analysts, rely on accurate and insightful financial analyses to make informed decisions. This project's findings will provide a valuable resource for these stakeholders, offering perspectives on the financial health of the selected telecom company and its implications for the broader industry.

Introduction to Principal Component Analysis (PCA):

Principal Component Analysis (PCA) stands as a powerful and widely utilized technique in the realm of data analysis and dimensionality reduction. Originating from the field of multivariate statistics, PCA serves as an invaluable tool for extracting essential patterns and reducing the complexity of high-dimensional datasets. The fundamental objective of PCA is to transform a set of correlated variables into a new set of uncorrelated variables, known as principal components, thereby capturing the most critical information while discarding redundant or less relevant aspects.

Key Concepts of PCA:

Dimensionality Reduction:

PCA addresses the challenge of handling datasets with a large number of variables by condensing the information into a smaller set of principal components.This reduction in dimensionality simplifies data interpretation, visualization, and computational complexity.

Uncorrelated Principal Components:

One of PCA's key features is the creation of principal components that are uncorrelated, enabling a clearer understanding of the underlying structure of the data.These components are derived by finding linear combinations of the original variables, emphasizing the directions of maximum variance.

Feature Extraction

PCA identifies the principal components, which are linear combinations of the original features. These components can be considered as new, derived features that capture the most significant patterns in the data. This can help in focusing on the most relevant information and discarding less important details.

Computational Efficiency

In some cases, reducing the dimensionality of the data can lead to significant computational savings, especially when dealing with large datasets. When working with a reduced set of principal components, it can be easier to interpret and understand the underlying patterns in the data.

Eigenvalues and Eigenvectors

Central to PCA are the concepts of eigenvalues and eigenvectors, which determine the principal components.Eigenvectors represent the directions of maximum variance, while eigenvalues quantify the amount of variance captured along these directions.

Applications in Data Analysis

PCA finds applications in various fields, including finance, image processing, biology, and machine learning.In finance, for insance, PCA aids in portfolio optimization by identifying the key factors influencing asset returns.

Challenges and Considerations:

While PCA offers substantial advantages, its application is not without challenges. Understanding these considerations is crucial for making informed decisions when implementing PCA:

Interpretability of Principal Components:

Interpreting the meaning of principal components in real-world terms may sometimes pose challenges. Ensuring a meaningful interpretation requires a deep understanding of the specific context and domain expertise.

Assumption of Linearity:

PCA assumes a linear relationship between variables. In cases where relationships are highly nonlinear, alternative techniques may be more appropriate.

Impact of Outliers:

Outliers in the dataset can significantly influence PCA results. It is essential to handle outliers appropriately, either by robustifying the analysis or addressing outliers directly.

Selecting the Right Number of Components:

Determining the optimal number of principal components to retain is a crucial decision. Balancing the need for dimensionality reduction with the preservation of meaningful information requires careful consideration.

PCA's impact spans various disciplines, offering insights into complex datasets, aiding in decision-making, and serving as a foundation for more advanced analyses. As technology progresses, the continued exploration and refinement of PCA contribute to its enduring relevance and effectiveness in diverse analytical endeavors.

CASE STUDY

The dataset now which we are dealing with contains data about a telecom companey of the services they provide and it encompasses essential customer-centric information, including customer demographics, service utilization patterns, billing details, and interaction history. Each entry is characterized by features such as telecom_partner, Gender, Senior_Citizen status, Partner, Dependents, Tenure, Phone_Service, Multiple_Lines, Internet_Service, Online_Security, Online_Backup, Device_Protection, Tech_Support, Streaming_TV, Streaming_Movies, Contract, Paperless_Billing, Payment_Method, Monthly_Charges, Yearly_Charge, Admin_Tickets, Tech_Tickets, and Churn status.

Partner	Dependents	Tenure	Phone_Service	Multiple_Lines	Internet_Service	Online_Security	Online_Backup	Device_Protection	Tech_Suppor	Streaming_TV	Streaming_Movies	Contrac	Paper_less_Billing	Payment_Method	Monthly_Charges	Yearly_Charg	Admin_Tickets	Tech_Tickets	Churn
No	No	56	Yes	Yes	Fiber optic	Yes	Yes	Yes	No	Yes	Yes	One year	No	Electronic check	1105	13260	0	0	No
No	No	30	Yes	No	DSL	Yes	Yes	No	No	Yes	Yes	Month-to-	Yes	Electronic check	747.5	8970	0	0	No
No	No	1	Yes	No	Fiber optic	No	No	No	Yes	No	No	Month-to-	No	Electronic check	747	8964	0	0	No
Yes	No	25	Yes	No	Fiber optic	No	Yes	Yes	No	Yes	Yes	Month-to-	Yes	Electronic check	985	11820	0	1	Yes
Yes	No	66	No	No phone servic	DSL	No	Yes	Yes	No	Yes	No	One year	No	Bank transfer (autc	455.5	5466	0	0	No
Yes	No	37	Yes	Yes	Fiber optic	No	No	No	No	No	No	Month-to-	Yes	Electronic check	765	9180	0	0	Yes
No	No	27	Yes	Yes	Fiber optic	No	No	Yes	No	No	No	Month-to-	Yes	Electronic check	780.5	9366	0	0	Yes
No	No	71	No	No phone servic	DSL	Yes	Yes	No	No	Yes	Yes	One year	Yes	Electronic check	564.5	6774	0	0	No
Yes	No	1	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-	Yes	Mailed check	704.5	8454	0	0	Yes
Yes	Yes	72	Yes	Yes	DSL	Yes	Yes	Yes	Yes	Yes	Yes	Two year	Yes	Credit card (autom.	880.5	10566	0	0	No

Objectives of the Case Study:

The primary goal of this case study is to employ Principal Component Analysis (PCA) as a powerful tool for dimensionality reduction and pattern extraction. The telecom industry, characterized by rapidly evolving customer needs and technological advancements, benefits immensely from advanced analytical techniques. PCA serves as a valuable instrument in navigating through the complexity of telecom datasets, offering a means to distill critical information and unveil hidden relationships. The methodology involves the step-by-step application of PCA to the telecom customer dataset. The subsequent application of PCA will yield principal components that capture the most significant variance in the data.

Expected Outcomes:

Through the application of PCA, we anticipate revealing dominant patterns in customer behavior, understanding the key drivers behind churn, and gaining insights that can inform strategic decision-making within the telecom industry.

Performing Principal Component Analysis (PCA) in SPSS involves several steps. Here is a step-by-step guide:

Step 1: Importing Data

Step 2: Data Preprocessing

Step 3: Running Principal Component Analysis

Step 4: Variable Selection

Step 5: Reviewing Results

Step 6: Interpreting and Naming Components

Step 7: Saving Scores

Step 9: Interpretation and Reporting

STEPS:

1.Open SPSS and load your dataset.Ensure that your dataset contains numerical variables suitable for PCA. Handle any missing values in your dataset.Standardize or normalize the data if variables are on different scales.

2. Running Principal Component Analysis,Click on "Analyse" in the top menu,Choose "Dimension Reduction" from the dropdown menu.Select "Factor" to initiate the Factor Analysis/Principal Components Analysis dialog.

3. Move the variables you want to include in the PCA to the "Variables" box.Adjust options in the "Extraction" tab, such as the extraction method (usually "Principal Components")

4. go to the "Descriptives" and check the “KMO and Bartlett's Test of sphericy” and also check the “coefficiency” and “Scree plot” click continue. Click "OK" to run the PCA

OUTPUT:

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.838
Bartlett's Test of Sphericity	Approx. Chi-Square	2964.811
	df	45
	Sig.	.000

In this table we can see the scores of **Kaiser-Meyer-Olkin Measure of Sampling Adequacy, the scorces should be more than 0.7 and Bartlett's Test of Sphericity significance should be less than 0.05.** So the results which we got is agrees with the condition so with this we can comprehend that we can apply the PCA for this dataset

Correlation Matrix

		Tenure	Online_Security	Online_Backup	Tech_Support	Streaming_TV	Streaming_Movies	Monthly_Charges	Admin_Tickets	Tech_Tickets	Payment_Method
Correlation	Tenure	1.000	.262	.338	.274	.220	.225	.267	.026	.255	.166
	Online_Security	.262	1.000	.409	.474	.297	.328	.372	.040	.088	.041
	Online_Backup	.338	.409	1.000	.426	.401	.388	.446	.019	.168	-.041
	Tech_Support	.274	.474	.426	1.000	.440	.383	.438	.002	.057	.055
	Streaming_TV	.220	.297	.401	.440	1.000	.595	.665	.013	.256	-.150
	Streaming_Movies	.225	.328	.388	.383	.595	1.000	.663	-.015	.256	-.168
	Monthly_Charges	.267	.372	.446	.438	.665	.663	1.000	.012	.258	-.157
	Admin_Tickets	.026	.040	.019	.002	.013	-.015	.012	1.000	-.029	-.015
	Tech_Tickets	.255	.088	.168	.057	.256	.256	.258	-.029	1.000	-.049
	Payment_Method	.166	.041	-.041	.055	-.150	-.168	-.157	-.015	-.049	1.000

So this is the corelation matrics between the veriables so by this we can see how the features are related to each other.for better understanding look at bellow table.

Correlation Matrix

		Tenure	Online_Security	Online_Backup	Tech_Support	Streaming_TV	Streaming_Movies	Monthly_Charges	Admin_Tickets	Tech_Tickets	Payment_Method
Correlation	Tenure	1.000	0.262	0.338	0.274	0.220	0.225	0.267	0.026	0.255	0.166
	Online_Security	0.262	1.000	0.409	0.474	0.297	0.328	0.372	0.040	0.088	0.041
	Online_Backup	0.338	0.409	1.000	0.426	0.401	0.388	0.446	0.019	0.168	-0.041
	Tech_Support	0.274	0.474	0.426	1.000	0.440	0.383	0.438	0.002	0.057	0.055
	Streaming_TV	0.220	0.297	0.401	0.440	1.000	0.595	0.665	0.013	0.256	-0.150
	Streaming_Movies	0.225	0.328	0.388	0.383	0.595	1.000	0.663	-0.015	0.256	-0.168
	Monthly_Charges	0.267	0.372	0.446	0.438	0.665	0.663	1.000	0.012	0.258	-0.157
	Admin_Tickets	0.026	0.040	0.019	0.002	0.013	-0.015	0.012	1.000	-0.029	-0.015
	Tech_Tickets	0.255	0.088	0.168	0.057	0.256	0.256	0.258	-0.029	1.000	-0.049
	Payment_Method	0.166	0.041	-0.041	0.055	-0.150	-0.168	-0.157	-0.015	-0.049	1.000

By this heatmap we can understand that the the relation between the veriables better so in this image we can visually see a relation between the veriables like online_security, online_backup, tech_support and streaming movies and more but by seeing the "component matrics" we can understand that total factors formed .

Rotated Component Matrix[a]

	Component		
	1	2	3
Monthly_Charges	.796		
Streaming_Movies	.756		
Streaming_TV	.748		
Tech_Tickets	.492		
Device_Protection	.459		
Tenure		.700	
Tech_Support		.646	
Online_Security		.645	
Payment_Method		.543	
Online_Backup		.510	
Admin_Tickets			.961

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

SO there are three factors formed through this analysis:

Factor 1 includes: [with factor loading values]

Monthly_charges	0.796
Streaming_tv	0.756
Streaming_Movies	0.748
Tech_tickets	0.492
Device_protection	0.459

2nd FACTOR includes:

Payment methord	0.743
Tenure	0.700
Tech_support	0.646
Online_backup	0.510

Online_security 0.645

3rd FACTOR include:

Admin_tickets 0.961

Total Variance Explained

Component	Initial Eigenvalues Total	Initial Eigenvalues % of Variance	Initial Eigenvalues Cumulative %	Extraction Sums of Squared Loadings Total	Extraction Sums of Squared Loadings % of Variance	Extraction Sums of Squared Loadings Cumulative %
1	3.543	35.432	35.432	3.543	35.432	35.432
2	1.290	12.897	48.329	1.290	12.897	48.329
3	1.061	10.610	58.939	1.061	10.610	58.939
4	.991	9.914	68.853			
5	.729	7.287	76.141			
6	.636	6.363	82.503			
7	.548	5.483	87.986			
8	.507	5.073	93.059			
9	.387	3.868	96.927			
10	.307	3.073	100.000			

Extraction Method: Principal Component Analysis.

Hear in this table look at the Initial Eigenvalues column and in this column look at Cumulative %. We can also observe that the eiganvalue is greater than 1 .1st factor explains total **35.48%** veriation in the table individually and 2nd and 3rd explains 12% and 10% each. So this table says there are three factors formed and these 3 factors contribute to total **58.939%** we can comprehend that the veriation occurred in the table is caused by these 3 factors by **58 %**

	Monthly_Charg	Yearly_Charg
Monthly_Charges	1	1.000
Yearly_Charge	1.000	1
Tenure	0.267	0.267
Online_Security	0.372	0.372
Online_Backup	0.446	0.446
Tech_Support	0.438	0.438
Streaming_TV	0.665	0.665
Streaming_Movies	0.663	0.663
Admin_Tickets	0.012	0.012
Tech_Tickets	0.258	0.258
Payment_Method	-0.157	-0.157

Coming to the **financial statement analysis** for this companey we can understand it by seeing the relation between the revenue of the companey and other features

By looking at this corelation we can understand that there is a positive corelation between all the features , so any changes done to these positively corelated features would increase the revenue of the companey, and by applying the marketing techniques on the features of 1st factor then there could be a chance of increasing the revenue of the companey.

LETS DO REGRESSION ANALYSIS TO FIND THE RELATION BETWEEN DEPENDENT VERIAVBLE {MONTHLY_REVENUR} AND INDEPENDENT VERIABLE {OTHER FEATURES}

STEPS:

- Goto Analyze->REGRESSION->LINEAR
- Move monthly revenue to depenedent
- Move other features to independent
- Go to statistics check casewise diagnostics and check all the boxes on the right,continue
- Goto options check Exclude cases pairwise, continue
- Click ok

OUTPUT

Descriptive Statistics

	Mean	Std. Deviation	N
Monthly_Charges	798.204	237.6403	1142
Online_Security	.25	.431	1142
Online_Backup	.42	.493	1142
Device_Protection	.41	.492	1142
Tech_Support	.23	.420	1142
Streaming_TV	.50	.500	1142
Streaming_Movies	.52	.500	1142
Tech_Tickets	.68	1.550	1142
Admin_Tickets	.51	1.297	1142

In this table we can see the mean , std deveation of the veriables.

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	.685[a]	.469	.465	173.8533	.469	124.858	8	1133	.000

a. Predictors: (Constant), Admin_Tickets, Device_Protection, Online_Security, Tech_Tickets, Online_Backup, Streaming_Movies, Tech_Support, Streaming_TV

b. Dependent Variable: Monthly_Charges

Hear in this table the RSQUARE value is 0.469 which explains that how much movement is cause in the dependent veriable(monthly returenes) when there is a varience in independent veriables(other features) so almost 47% of varience is explained in dependent veriable by movement in independent veriable.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	30190678.16	8	3773834.769	124.858	.000[b]
	Residual	34244896.76	1133	30224.975		
	Total	64435574.92	1141			

a. Dependent Variable: Monthly_Charges

b. Predictors: (Constant), Admin_Tickets, Device_Protection, Online_Security, Tech_Tickets, Online_Backup, Streaming_Movies, Tech_Support, Streaming_TV

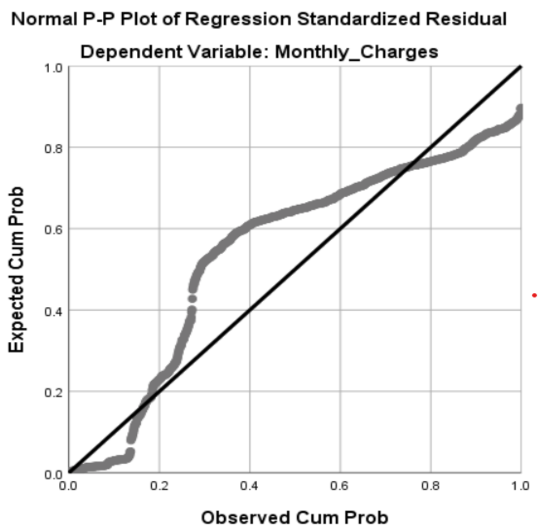
Hear in this table we can see that the signifcance is 0.00 so our result is very significant . we should accept the null hypothesis only when its greater that 0.05.

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Partial	Part	Collinearity Statistics Tolerance	VIF
1	(Constant)	572.968	9.195		62.312	.000					
	Online_Security	45.690	12.637	.083	3.616	.000	.175	.107	.078	.891	1.122
	Online_Backup	53.526	11.147	.111	4.802	.000	.290	.141	.104	.876	1.141
	Device_Protection	48.492	11.598	.100	4.181	.000	.356	.123	.091	.815	1.228
	Tech_Support	25.496	13.537	.045	1.883	.060	.248	.056	.041	.821	1.217
	Streaming_TV	154.173	12.103	.325	12.738	.000	.553	.354	.276	.723	1.384
	Streaming_Movies	157.050	11.908	.330	13.189	.000	.552	.365	.286	.748	1.337
	Tech_Tickets	7.759	3.518	.051	2.206	.028	.258	.065	.048	.891	1.123
	Admin_Tickets	3.295	3.976	.018	.829	.407	.012	.025	.018	.996	1.004

Coefficients[a]

a. Dependent Variable: Monthly_Charges

In this table look at the STANDARDISED COEFFECIENTS it explains the verience in dependent veriable. As we see streamimg_tv and streaming_movies explain more verience in the dependent veriable(monthly revenue) and look at the part column of streamimg_tv and streaming_movies it explains unique contribution of specific veriable in the verience. And in the significance column the significance should be less that 0.001 only then its highly significance but in this case Admin_ticket , tech_ticket ,tech_support are not significance so we can say that Admin_ticket , tech_ticket ,tech_support are not the good predictor of the dependent veriable.



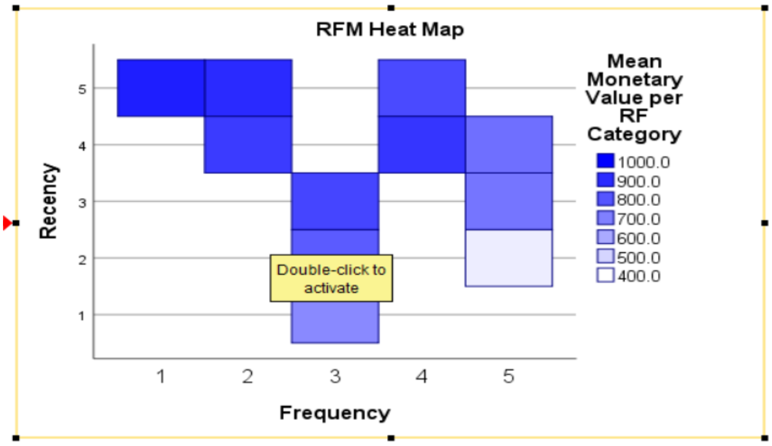
This is probablity plot explains most points are relatively close to the line and there is also deveation but it followsup the line

LETS APPLY RFM ANALYSIS

By taking the TENURE as the RESENCY , CONTRACT as FREQUENCY and MONTHLY_CHARGES as the MONETARY we can prosede to do this analysis

Steps:

- Got to analysis digital marketing
- Select rfm
- Click on customer data
- Place tenure in transaction date,frequency in NUMBER OF TRANSACTIONS, monthly_charges in AMOUNT and customer_id in identifiers click ok

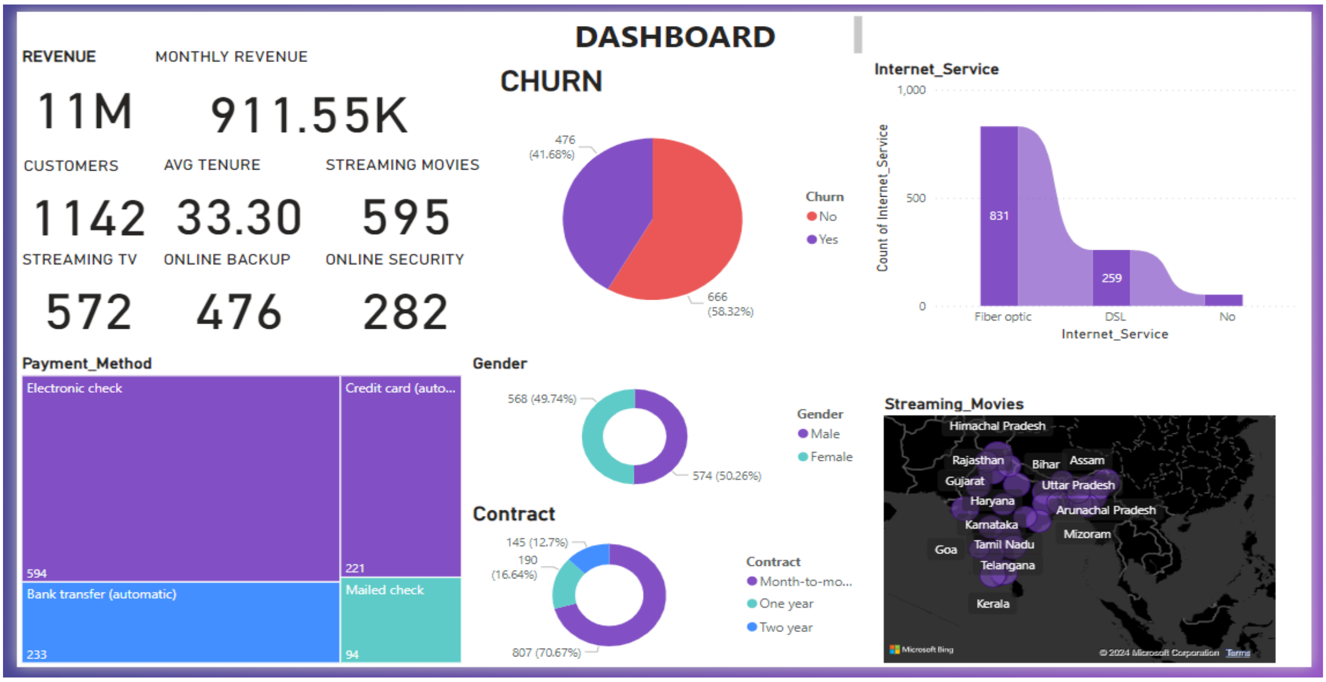


This is the RFM heatmat which show that Darker areas indicate a higher average monetary value. In other words, customers with recency and frequency scores in the darker areas tend to spend more on average than those with recency and frequency scores in the lighter areas. So by doing this analysis we can understand that Customers with longer tenure may be considered more loyal, as they have been with your business for an extended period. Monthly charges as monetary value can help identify high-value customers or those contributing more to your revenue on a regular basis. A higher frequency of contract renewal might indicate active and engaged customers who are more likely to stay with your service.

So by sorting the rfm values of customer this is the output:

customer_id	recency	frequency	monetary	RFM SCORE
1	5	4	5	545
2	5	4	5	545
3	5	4	5	545
4	5	4	5	545
5	5	4	5	545
6	5	4	5	545
7	5	4	5	545
8	5	4	5	545
9	5	4	5	545
10	5	4	5	545
11	5	4	5	545
12	5	4	5	545
13	5	4	5	545
14	5	4	5	545
15	5	4	5	545
16	5	4	5	545
17	5	4	5	545
18	5	4	5	545
19	5	4	5	545
20	5	4	5	545
21	5	4	5	545
22	5	4	5	545
227	5	4	4	544
228	5	4	4	544
229	5	4	4	544

LETS SEE SOME OF THE VISUALIZATION DRAWN FROM THE DATASET



<u>CONCLUSION</u>

In conclusion, the application of PCA to financial statement analysis has unfolded a new dimension of understanding within the intricate financial landscape of an Indian telecom company. As the telecom industry continues its evolution, the integration of advanced analytical methodologies, including PCA, will remain integral to ensuring financial resilience, strategic growth, and sustainable financial health. Through regression analysis we got to know that how the revenue is affected by various other features of the dataset and by the RFM analysis we got to know about the customers which the company can do target marketing for getting better results.