Strategic Integration: A Comprehensive Dashboard with Customer Segmentation and Recommender System for Telco Business Enhancement

By: Sammi Chia Li

INTRODUCTION

Customer Segmentations and Recommender System sets a baseline for businesses to plan out their marketing strategies for their products or services. Long-term or short-term marketing campaigns done by marketers often require an achievable goal and mainly an identified targeted audience before starting out.

The key point here is to catch the attention and interest of these targeted customers and this can be achieved by performing customer segmentation to understand them better and provide them with relevant telco packages which they might be interested in. Marketing irrelevant products or services to customers will result in a failed marketing campaign with no significant benefits towards the business. It was found that campaigns done towards segmented customers have 74% higher click rates compared to non-segmented campaigns. Customers are more likely to view and take an interest towards a new promotion or services provided to them only if it resonates with or relates to them. Successful marketing campaigns reduce the market spend because it is being targeted to customer segments that are more likely to buy or subscribe rather than those who are uninterested.

High customer churning rate is a concerning issue because it is found that 65% of a company's purchases come from existing customers rather than new customers. It is seven times easier to retain existing customers than to acquire new customers. The advanced use of machine learning algorithms to detect customer segments and recommend the relevant telco packages to customers, helps companies to gain competitive advantage over other service providers. This does not only create more new market opportunities but differentiate themselves from others through their fast adaptability towards growing customers changing needs.

Valuable insights through data-driven customers can bring innovation to telecommunication services to plan out their package deals to suit customers' preference. Boring unchanging telecom plans will cause existing customers to opt for better choices from another service provider. Existing customers are 50% more likely to try out new services provided to them when compared to new customers.

In this paper, K-Means Clustering with Principal Component Analysis (PCA), Agglomerative Hierarchical Clustering and DBSCAN Clustering was proposed to build the customer segmentation model based on IBM Telco Customer Churn Dataset. Meanwhile, content-based filtering using cosine similarity was implemented to build the Telco Packages Recommender System by finding the most relevant package for each customer.

This paper contributes to the growth of the businesses by segmenting their customers into unique sub-groups in order to gain insights and provide them with the best personalised marketing materials for increased customer retention and customer purchases. Segmented customers are easier to manage and satisfy their needs and expectations towards the telecom services hence creating competitive advantage. Previous research work [3] focused on analysing and visualising customer segments rather than interpreting them for future actions. It is focused on descriptive analysis meanwhile, this paper explores into predictive and prescriptive analysis through telco packages recommendations and marketing strategies suggestions.

The descriptive analysis in this paper explores the use of Customer Lifetime Value (CLTV) Calculation for each segment to identify the highest revenue generating customer segments so marketers can focus on retaining them, especially those who belong to the churned groups. Better marketing strategies can be performed on these high CLTV customers to maintain their loyalty and purchasing power. Not only does CLTV describe the segments but it also provides predictive insights of these segments' future value to the business. These predicted future CLTV can be a guiding line for businesses to plan their future business outlook and help in acquiring new customers. General thumb rule spends one third of these CLTV into acquiring new customers therefore, future actionable plannings can be done.

Based on three different judging criterias, Segment CLTV, Tenure Months and Monthly Charges of each customer, the segments are defined with their own unique characteristics. From here, prescriptive analysis is performed to suggest possible marketing strategies to be applied onto each segment. Customer Segments' CLTV and Marketing Strategies suggestions are both lacking in other research papers and this paper aims to fill up the gap by providing further actionable insights especially on the use of CLTV.

Not only that, this paper also explores the use of content-based filtering to build a Telco Packages Recommender System that recommends the most relevant package to customers based on their historical subscription or preferences. Recommending the most suitable package to customers can be a difficult task and some customers might lose interest when the package does not meet their expectations therefore, the recommender system that compares different packages' offerings to customers' profile can be accurate in finding the best package for them.

LITERATURE REVIEW

A. CUSTOMER SEGMENTATION

The telecommunication industry has the forefront of customer segmentation models due to its subscription-based business model which is prone to high customer churn rates. When constructing the relationship between customer satisfaction and repurchase intention, it was found that there is a positive relationship between them. Customer's satisfaction level influences their repurchase intention [1], [2]. This highlights the need of having an effective segmentation model to retain customers within the business.

A customer segmentation method was proposed by [3] for an artificially generated Telco dataset by using K-Means Clustering with PCA technique to segment the customers in a lower dimensional space. Through the behavioural segmentation, an interactive dashboard was created to visualise clusters and provide analytical insights. [3] leveraged the use of Business Intelligence tools to boost decision-making process using customer segments dashboarding.

Wang et al [4] also proposed the use of K-means clustering algorithm for a Chinese Telco Company and was proven to be the simplest and most effective algorithm. The clustering model was evaluated using silhouette score with comparison to RFM analysis and demographic segmentation.

Customer Lifetime Value calculation as a method to evaluate customer's future value was proposed by [5] after using K-Means Clustering to segment Small and Medium Enterprise (SMEs) customers. The clustering was not performed on Telco dataset but it highlighted the use of measuring customer profit for a company and this method can give valuable meaning to segments.

B. CONTENT-BASED FILTERING ON TELCO SERVICE PACKAGES FOR RECOMMENDATIONS

Telco packages recommender system is crucial in the telecommunication industry to increase customers satisfaction and loyalty to their choice of company. The rapidly changing and wide variety of telecommunication services offered in the domain can be intimidating for customers to find the most relevant and suitable package. Churn rate increases linearly with the relevant recommended packages because it satisfies them to find quality and personalised package deals for them. Recommender systems are based on ratings of a particular item to identify how much a user likes the item. Most recommender systems adoptContent-Based (CB) methods and Collaborative Filtering (CF) as their training techniques [7] and some use a hybrid approach to the recommender system.

Collaborative Filtering Method was proposed by [8] for a European Telecommunication company with 6 months of historical records for 10 million customers to train the recommender system. The research focused on the use of the Alternative Least Squares method to determine the candidate offers.

A hybrid approach of both item-based and user-based CF method was proposed by [9] to build the recommender system and although the research is still ongoing, the proposed system is already being used in a telecommunication company in Australia. This proves its real-life adaptability and implementation for recommending telecom packages.

METHODOLOGY

Figure 1 shows the entire project methodology of this paper from data collection to model training, dashboard creation and web deployment.

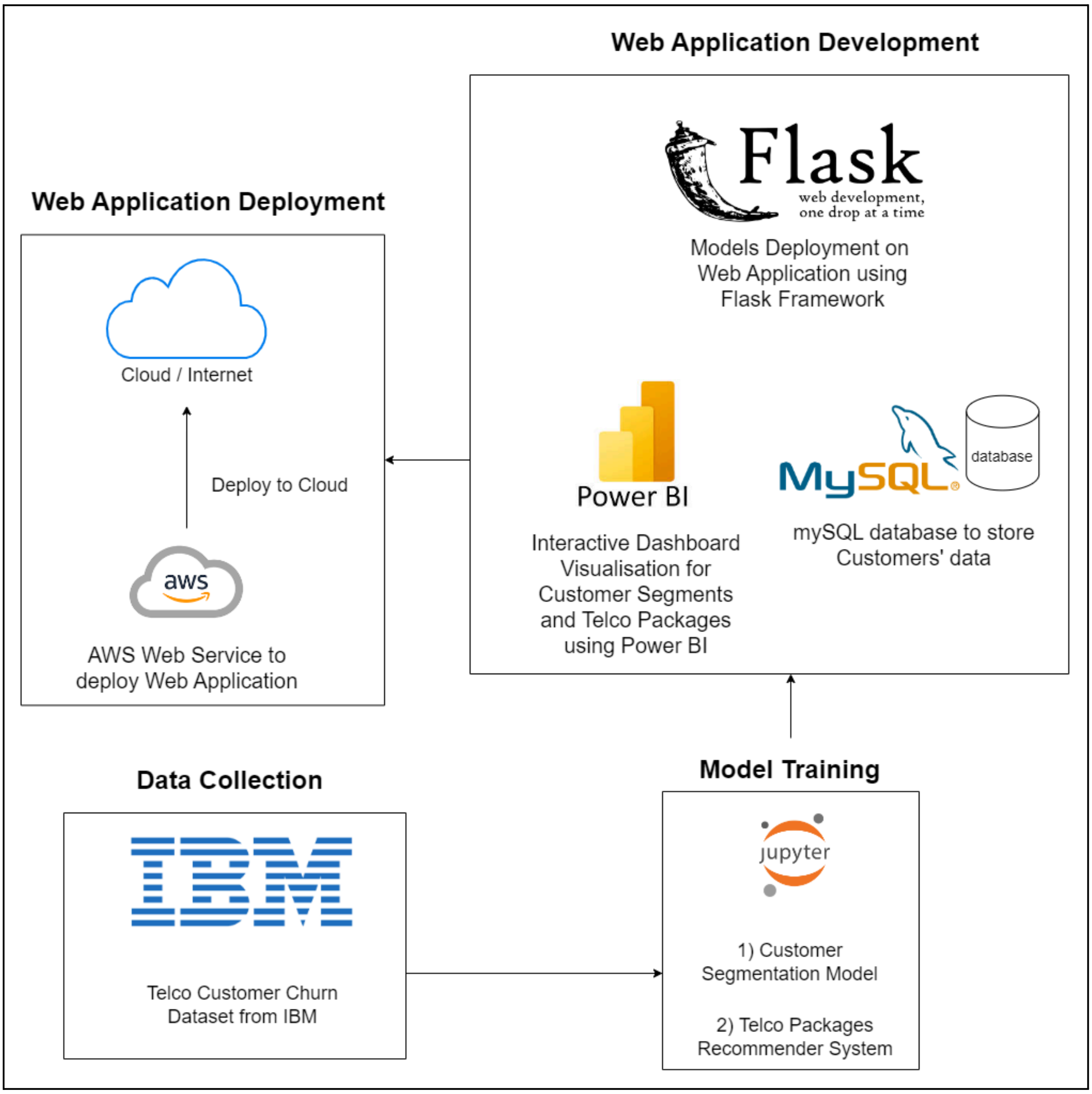

Figure 1 : Project Methodology

Figure 2 shows the stages of data preprocessing on the IBM Telco Dataset

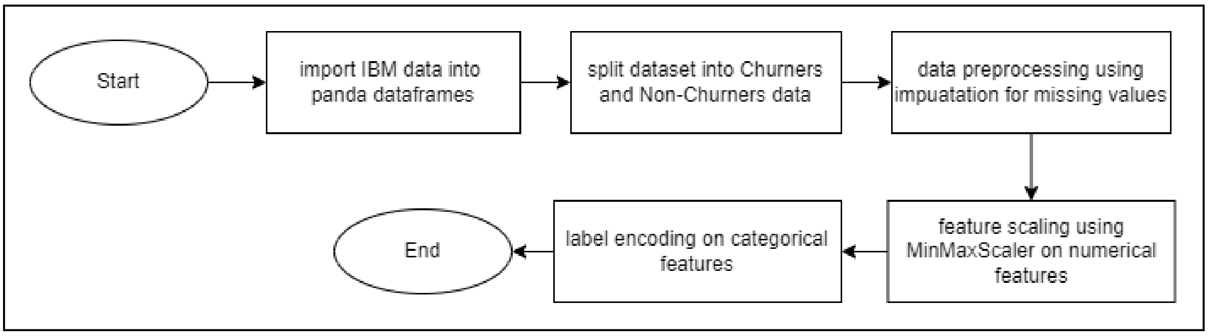

Figure 2: Flowchart for Data Preprocessing Stages

Figure 3 shows the flowchart for training the recommender system for telco packages

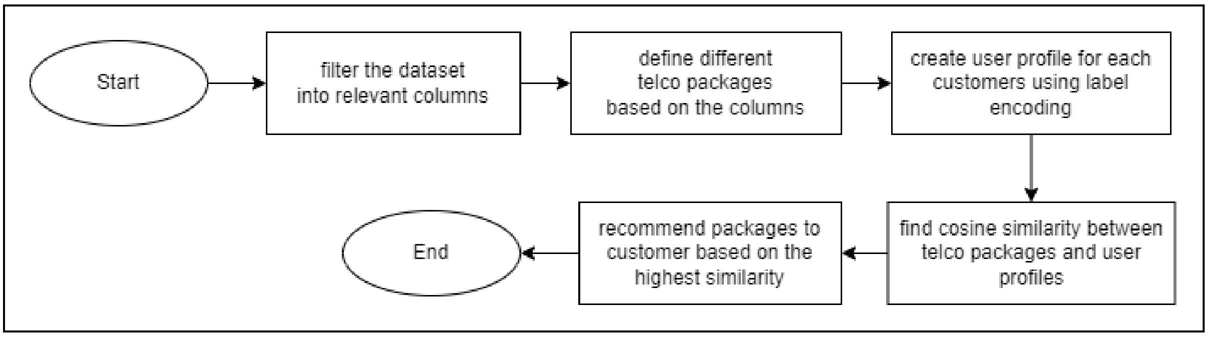

Figure 3: Flowchart for Telco Packages Recommender

Figure 4 shows the flowchart for training the K-Means with PCA model.

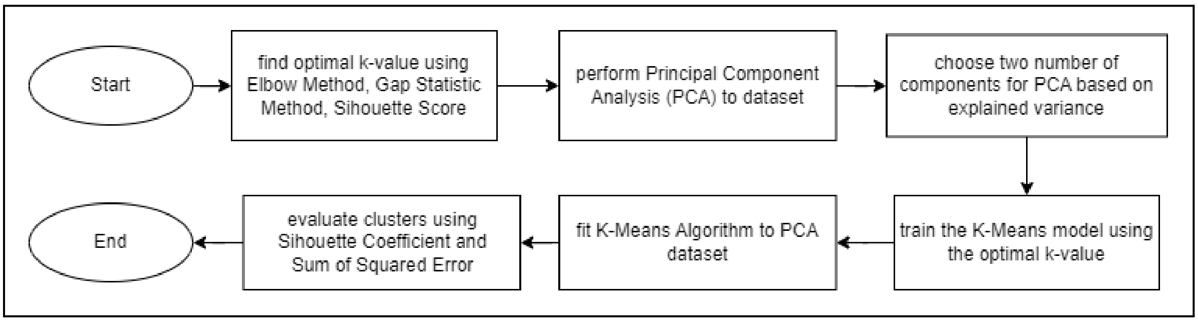

Figure 4: Flowchart for K-Means with PCA Clustering

Figure 5 shows the flowchart for training the Agglomerative Hierarchical Clustering model.

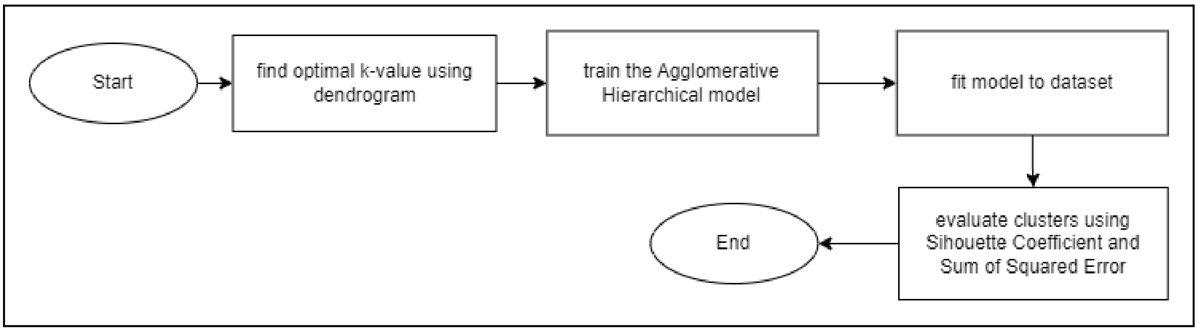

Figure 5: Flowchart for Agglomerative Hierarchical Clustering

Figure 6 shows the flowchart for training the DBSCAN Clustering model.

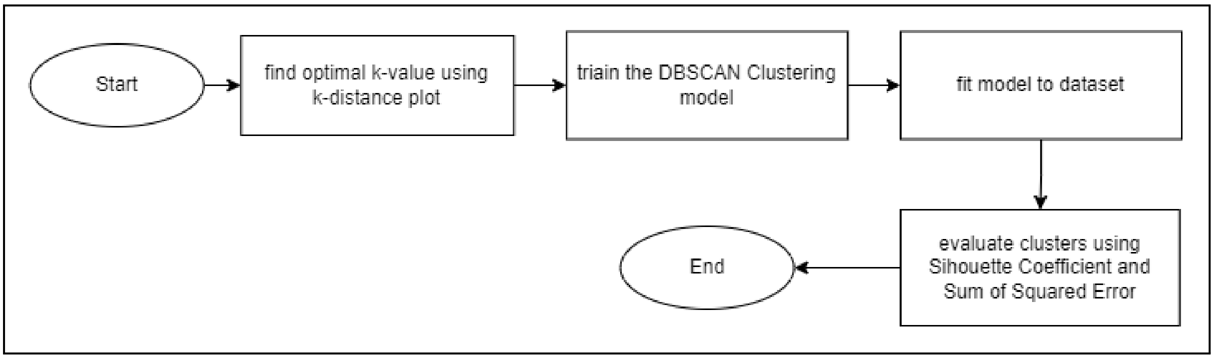

Figure 6: Flowchart for DBSCAN Clustering

Figure 7 shows the process of the project after choosing the best model.

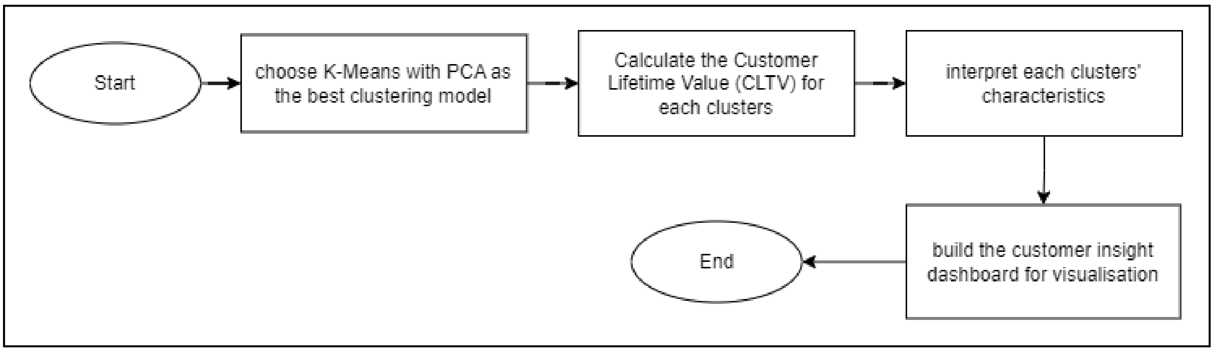

Figure 7: Process after choosing the best model

A. DATA COLLECTION

The data is collected from IBM Telco Customer Churn dataset which contains 7043 rows and 33 columns where each column represents customers' demographic information, account information and service usage.

Table 1 shows the description of all the features within the dataset.

No	Columns	Data Format	Description
1	CustomerID	string	Unique identifier
2	Count	integer64	Customer count
3	Country	string	The country the customer resides in.
4	State	string	The state location the customer resides in.
5	City	string	The city the customer resides in.
6	Zip Code	integer64	The zip code of the place the customer resides in.
7	Lat Long	string	Precise geographical location of the customer.
8	Latitude	string	Precise geographical location of the customer
9	Longitude	string	Precise geographical location of the customer
10	Gender	(Male/Female)	Whether the customer is a male or female.
11	Senior Citizen	(Yes/No)	Whether the customer is a senior citizen or not.
12	Partner	(Yes/No)	Whether the customer has a partner or not.
13	Dependents	(Yes/No)	Whether the customer has dependents or not.
14	Tenure Months	(Number of Months of Tenure; 73 different tenure months contracts)	Number of months the customer has stayed with the company.

15	Phone Service	(Yes/No)	Whether the customer subscribes to a phone service or not.
16	Multiple Lines	(Yes/ No/ No phone service)	Whether the customer subscribes multiple lines or not.
17	Internet Service	(No/ DSL/ Fiber Optic)	Customer's internet service provider
18	Online Security	(Yes/ No/ No internet service)	Whether the customer subscribes to online security or not.
19	Online Backup	(Yes/ No/ No internet service)	Whether the customer subscribes to online backup or not.
20	Device Protection	(Yes/ No/ No internet service)	Whether the customer subscribes to device protection or not.
21	Tech Support	(Yes/ No/ No internet service)	Whether the customer subscribes to tech support or not.
22	Streaming TV	(Yes/ No/ No internet service)	Whether the customer subscribes to streaming TV or not.
23	Streaming Movies	(Yes/ No/ No internet service)	Whether the customer subscribes to streaming movies or not.
24	Contract	(Month-to-Month/ One Year/ Two Year)	The contract term of the customers.
25	Paperless Billing	(Yes/No)	Whether the customer uses paperless billing or not.
26	Payment Method	(Bank Withdrawal/ Credit Card/ Mailed Check)	The customer's payment method.
27	Monthly Charges	float64	The amount charged to the customer monthly.
28	Total Charges	string	The total charges by the end of a quarter.
29	Churn Label	(Yes/No)	Whether the customer churn or not.
30	Churn Value	(0 = remained, 1 = churn)	Whether the customer churn or

			not.
31	Churn Score	(0 to 100; higher value, likely to churn)	The likelihood of the customer churning.
32	CLTV	integer64	Precalculated Customer Lifetime Value
33	Churn Reason	string	Reasons why the customer churn.

Table 1: Features in IBM Telco dataset without pre-processing.

B. DATA PREPROCESSING

Eight unnecessary and redundant columns in the dataset were removed from the total of 33 columns hence, leaving the dataset with 25 columns. The columns that were removed are CustomerID, Count, Country, State, Lat Long, Latitude, Longitude and Churn Label.

‘Total Charges’ columns were converted from ‘object’ data type to ‘float’ data type and missing values were handled in two columns. 11 null values in ‘Total Charges’ were imputed using KNNImputer while 5174 null values in ‘Churn Reasons’ were imputed using mode of the data. No outliers were found in the dataset.

Label Encoding methods were used on all categorical features instead of one hot encoding so that the Curse of Dimensionality can be avoided. Then, MinMaxScaler scaling was used to normalise the data distribution on all numerical data.

C. CUSTOMER SEGMENTATION ALGORITHMS

This research aims to fill the research gap by exploring other different segmentation algorithms such as, K-Means Clustering with Principal Component Analysis (PCA), Agglomerative Hierarchical Clustering and DBSCAN Clustering.

Table 2 shows all the clustering algorithms to be used in this paper as well as the respective method used to initialise the k-value for model training.

Clustering Algorithm	Initialization Value
K-Means Clustering with Principal Component Analysis (PCA)	K-value determined by: ● Elbow Method ● Silhouette Score ● Gap Static Method
Agglomerative Hierarchical Clustering	K-value determined by: ● Dendrogram
DBSCAN Clustering	Epsilon (Eps) Value determined by: ● K-distance plot

Table 2: Initialization Value for different clustering algorithms

Based on previous research work, K-Means Clustering with PCA performs the best therefore it is included in this research. Meanwhile, Agglomerative Hierarchical and DBSCAN are both suggested algorithms by previous researchers.

In this paper, it was found that when PCA is applied with K-Means, it helps with noise reduction and overall clearer defined clusters. PCA reduces the dimensionality of the dataset by combining relevant features together into Principal Component 1 and Principal Component 2 and this prevents the clusters from overlapping each other when visualised in a scatter plot. Clusters are seen to be more closely dense and packed together within their own clusters rather than widely spreaded away from clusters.

D. SEGMENTATION MODEL EVALUATION METHODS

Silhouette Coefficient

- ranges between -1 to +1 where 0 indicates overlapping clusters, 1 indicates that object i is very similar to their own cluster but dissimilar to other clusters.
- higher value indicates good cluster and lower value indicates poor cluster.

$$s(i) = \frac{(b(i) - a(i))}{max(a(i), b(i))}$$

where:

$s(i)$ = silhouette coefficient for object i

$a(i)$ = average distance between object i and all other objects in the same cluster as i

$b(i)$ = minimum average distance between object i and all objects in any other cluster

Sum of Squared Errors (SSE)

- measures the total within-cluster SSE from each data point to the centroid of the assigned cluster.
- lower SSE indicates data points are closer to cluster centroid which is a good indicator and vice versa.

$$SSE = \sum i \sum j \, (x_j - c_i)^2$$

where:

$\sum i$ = sum of all clusters

$\sum j$ = sum of all data points in a cluster

x_j = the j^{th} data point

c_i = the centroid of i^{th} cluster

E. CUSTOMER LIFETIME VALUE (CLTV) FOR EACH SEGMENTS

Upon getting segmented into its own segment, CLTV of that particular segment will also be calculated to see which segment has a higher CTLV and its importance to the business is rated.

The formula to calculate Customer Lifetime Value (CLTV) is as follows:

$$\boxed{CLTV = Customer\ Value \times Customer\ Lifespan \times Retention\ Rate}$$

where:

- $Customer\ Value = Customer\ Revenue$
- $Customer\ Lifespan = \frac{Average\ number\ of\ years\ a\ customer\ stays\ active}{Total\ number\ of\ customers}$
- $Retention\ Rate = \frac{Total\ number\ of\ customers\ at\ the\ start}{Total\ number\ of\ customers\ remained}$

The IBM Telco Customer Churn Dataset is collected as quarterly data. There are no date columns in the dataset therefore, it is assumed that this data is collected for a duration of one quarter which is 3 months.

Based on the above CLTV formula, the following formula is derived to match the dataset we have:

$$\boxed{CLTV\ for\ each\ cluster = \Sigma\,(Customer\ Value \times Customer\ Lifespan \times Retention\ Rate)}$$

where:

- $Customer\ Value = Average\ Customer\ Quarterly\ Revenue$
  $= \frac{Monthly\ Charges \times 3}{Total\ number\ of\ Customers\ within\ the\ cluster}$
- $Customer\ Lifespan = Average\ Lifespan \approx \frac{\Sigma\,Tenure\ Months}{Total\ number\ of\ Customers\ within\ the\ cluster}$
- $Retention\ Rate = \frac{Total\ number\ of\ customers}{Total\ number\ of\ customers\ who\ did\ not\ churn}$

F. CUSTOMER SEGMENTS INTERPRETATION AND DECISION-MAKING

Each customer segment will be interpreted based on three different criterias namely, Cluster's Customer Lifetime Value (CLTV), Customers' Monthly Charges and Customers' Tenure Months with the Telco Company. Based on the findings and interpretation of the clusters, a few marketing strategies will be suggested so that marketing experts can implement the suitable technique to retain customers and improve satisfactions.

G. CONTENT-BASED FILTERING RECOMMEDER FOR TELCO SERVICE PACKAGES

Based on the provided dataset, it is identified that there are 8 columns representing the customers' telco services subscriptions:

Telco Services Columns	Label encoded values
Phone service	(0=No, 1=Yes)
Internet service	(0=DSL, 1=Fiber optic, 2=No)
Online security, Online backup, Device protection, Tech support, Streaming TV, Streaming Movies	(0=No, 1=No internet service, 2=Yes)

Figure 3: Columns representing Telco Services

Telco services bundle packages will be determined by these 8 columns and a recommender system is needed to match the right customer to the right telco package. Therefore, a content-based filtering algorithm is used to recommend a more suitable telco bundle package to the customer based on their subscription history. Similarity Measure between user's telco subscription and available packages is done by using Cosine Similarity.

The formula to calculate Cosine Similarity is as follows:

$$Cosine\ Similarity\ = \frac{(A \bullet B)}{||A|| \bullet ||\ B||}$$

where:

- $(A\ \bullet\ B)$ = dot product of A and B
- $||A||\ = L2\ norm\ of\ A$ = square root of the sum of squares of elements of vector A

The IBM Telco Dataset does not have a predefined Telco Services Packages therefore, a user profile is made using an array of eight binary values representing eight different telco services columns.

Based on the above Cosine Similarity formula, the following formula is derived to match the dataset to calculate the similarity between user and telco packages:

$$\boxed{Relevance\ of\ Telco\ Packages\ = \frac{(Telco\ Packages\ \bullet User\ Profile)}{||\ Telco\ Packages\ ||\bullet||\ User\ Profile\ ||}}$$

where:

- $Telco\ Packages$ = an array of eight binary values separated by commas representing eight different telco services columns.
  Telco Packages = [Phone Service, Internet Service, Online Security, Online Backup, Device Protection, Tech Support, Streaming TV, Streaming Movies]

- $User\ Profile =$ an array of eight binary values separated by commas representing the user's current subscriptions to the eight different telco services.

Example of the calculation of cosine similarity is as follows:

$$TelcoPackages\ = [\ [1,2,1,1,1,1,1,1], \quad sampleUserProfile\ = [\ 1,1,0,0,0,0,0,0\ ]$$
$$[1,1,0,0,0,0,0,0],$$
$$[1,0,0,0,0,0,0,0],$$
$$[1,1,0,0,0,0,2,2]$$
$$[1,1,0,2,2,0,2,2],$$
$$[1,0,2,2,2,2,2,2]\ ]$$

Cosine Similarity between TelcoPackages and sampleUserProfile
= [0.25916053 0.17277369 0.08638684 0.17277369 0.17277369 0.08638684]

The highest similarity value is the most recommended package in this case which is, Package 1.

RESULT ANALYSIS

A. K-MEANS CLUSTERING WITH PRINCIPAL COMPONENT ANALYSIS (PCA)

Finding optimal k-value for K-Means Clustering

Figure 8 shows one of the methods to determine optimal k using the Elbow Method by plotting the Within-Cluster Sum of Square (WCSS) Value against the number of clusters, k. It is shown that the “elbow” can be seen at k=3 for churned customers while k=4 for non-churned customers.

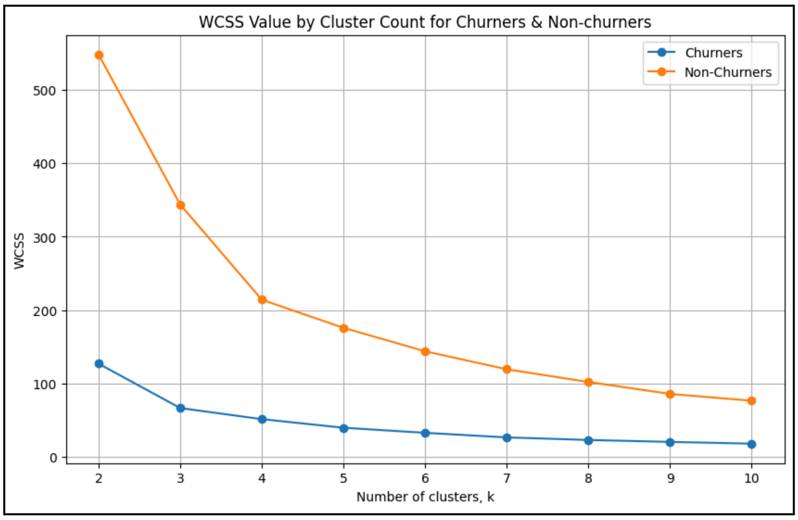

Figure 8: Elbow Method for Churners and Non-Churners Data

Figure 9 shows another method to determine the best k-value using Gap Statistics Method by plotting Gap Value against number of clusters, k. It is shown that the optimal k-value at bending point is also k=3 for churned customers and k=4 for non-churned customers. .

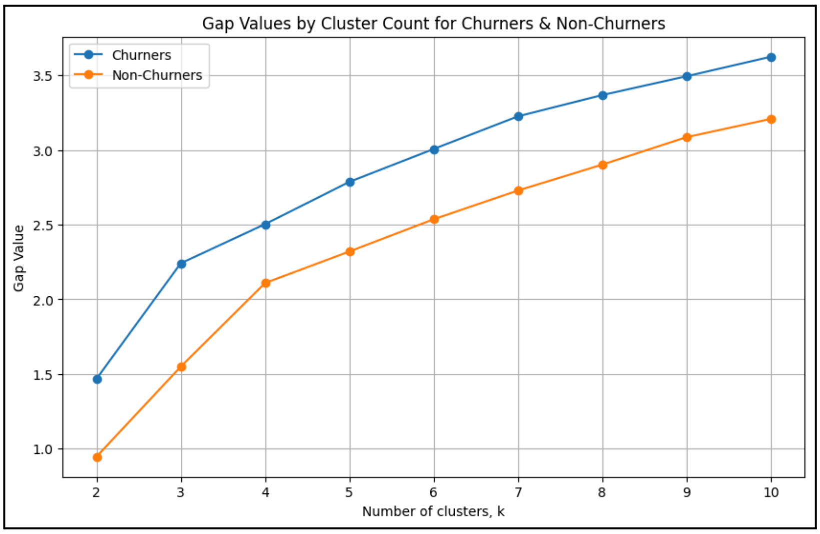

Figure 9: Gap Statistics Method for Churners and Non-Churners Data

Figure 10 shows a plot of silhouette scores against the number of clusters, k to determine the best k-value for the clustering algorithm. It is shown that the optimal k-value is the highest value of silhouette score which defines a well-separated cluster. K-value is found to be 3 for churned customers and k=4 for non-churned customers.

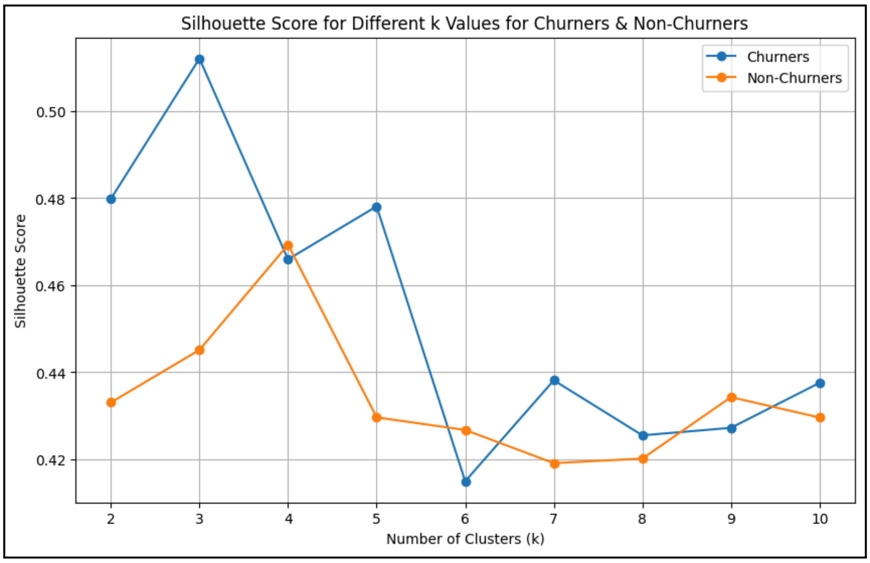

Figure 10: Silhouette Score for Churners and Non-Churners Data

Based on the three different methods above in Figure 8, 9 and 10, it is identified that the best number of clusters or k-value for churners data is 3 clusters while, for non-churners data is 4 clusters.

Performing Principal Component Analysis (PCA) on Dataset

PCA is used on the dataset and the number of components chosen is 2 because the explained variance ratio found on the dataset is [0.70577676, 0.29422324] which shows a significant factor in explaining the variance.

Table 2 below shows the sample data frame of churners data after PCA has been applied.

	Monthly Charges	Tenure Months	Component 1	Component2	Clusters
0	53.85	2	-0.301022	0.024886	0
1	70.70	2	-0.197828	-0.107256	2
2	99.65	8	0.045148	-0.282999	2
3	104.80	28	0.295617	-0.152418	1
4	103.70	49	0.518757	0.035726	1

Table 2: Sample Data with PCA applied for Churners Data.

Figure 11 below shows a scatter plot of churned customers data after PCA has been applied to the dataset. It can be seen that the clusters are more densely clustered and not spreaded out.

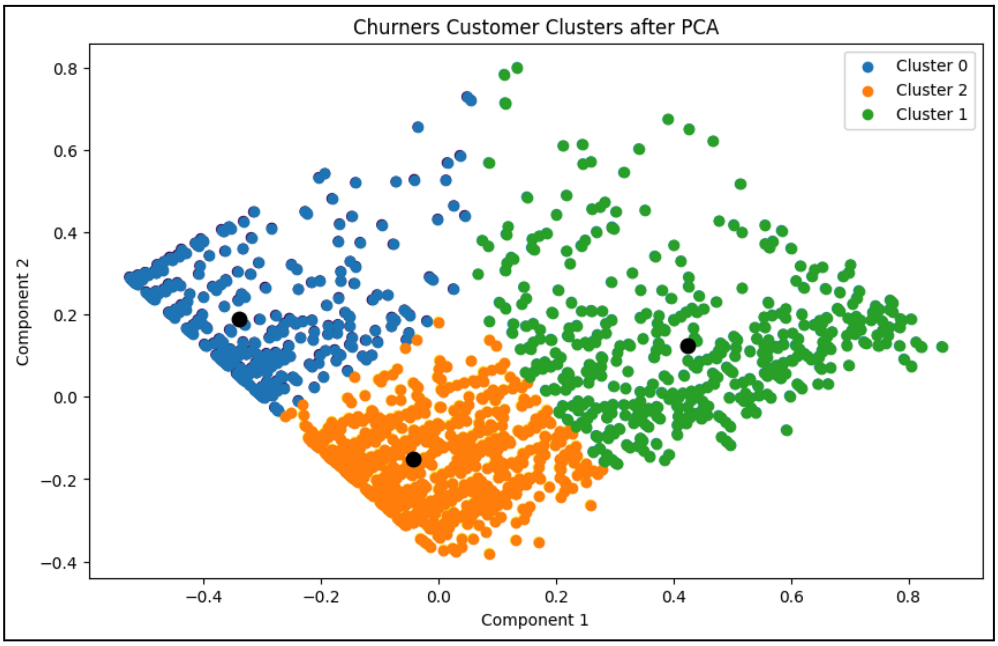

Figure 11: Scatter Plot of Churners' Clusters after PCA

Table 3 below shows the sample data frame of non-churners data after PCA has been applied.

	Monthly Charges	Tenure Months	Component 1	Component2	Clusters
0	29.85	1	-0.592391	0.069012	1
1	56.95	34	-0.065520	-0.003017	2
2	42.30	45	-0.035803	-0.212080	3
3	89.10	22	0.001669	0.351383	2
4	29.75	10	-0.494849	-0.009163	1

Table 3: Sample Data with PCA applied for Non-churners Data.

Figure 12 below shows a scatter plot of non-churned customers data after PCA has been applied to the dataset. It can be seen that the clusters are more densely clustered and well separated.

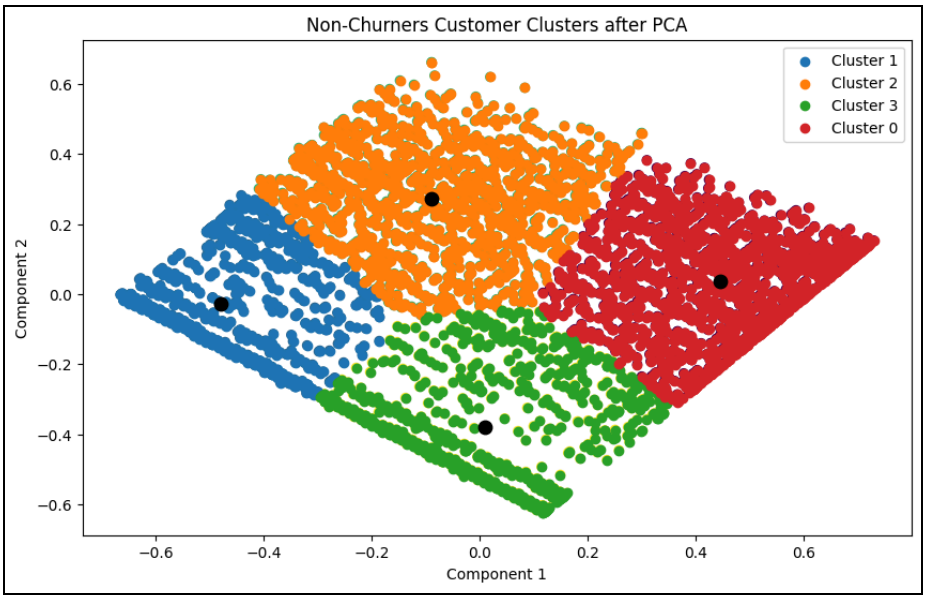

Figure 12: Scatter Plot of Non-Churners' Clusters after PCA

B. AGGLOMERATIVE HIERARCHICAL CLUSTERING

Finding optimal number of clusters using Dendrogram

Figure 13 shows a plotted dendrogram to identify the optimal k-value for agglomerative hierarchical clustering in churned customers. It is identified that the optimal number of clusters for churn customers is 3 when a y-axis threshold value is set to be 10.

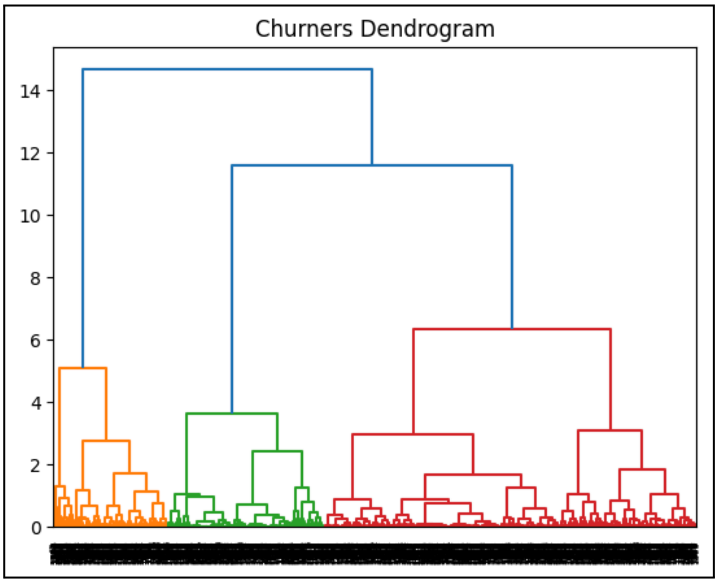

Figure 13: Churners' Data Dendrogram

Figure 14 shows a scatter plot of churned customer data after clustering using agglomerative hierarchical clustering with 3 different clusters.

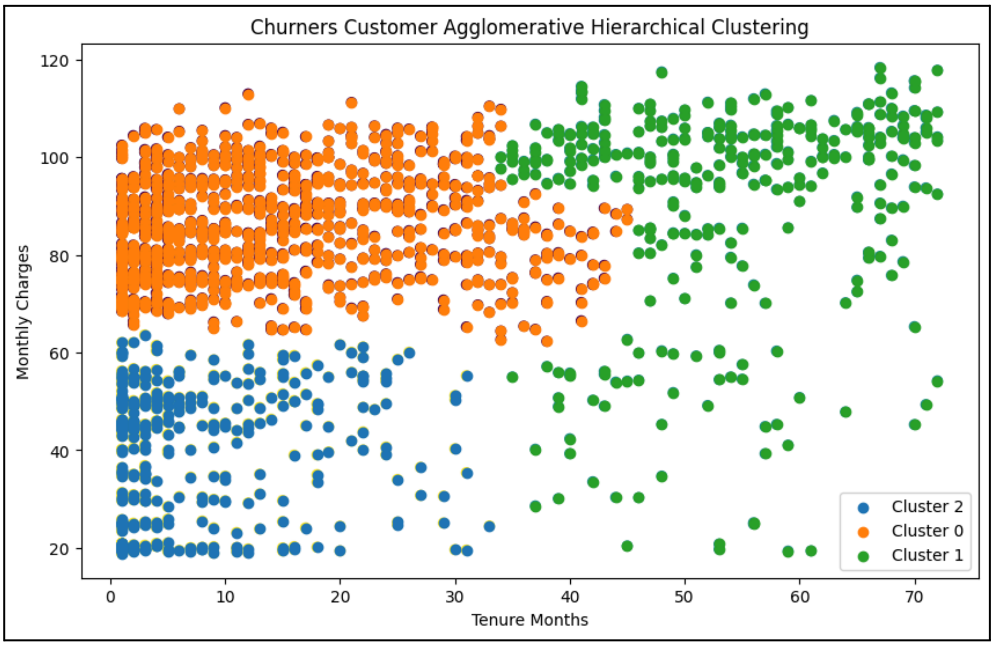

Figure 14: Churners' Customer Clusters using Agglomerative Hierarchical Clustering

Figure 15 shows a plotted dendrogram to identify the optimal k-value for agglomerative hierarchical clustering in non-churned customers. It is identified that the optimal number of clusters for churn customers is 2 when a y-axis threshold value is set to be 25.

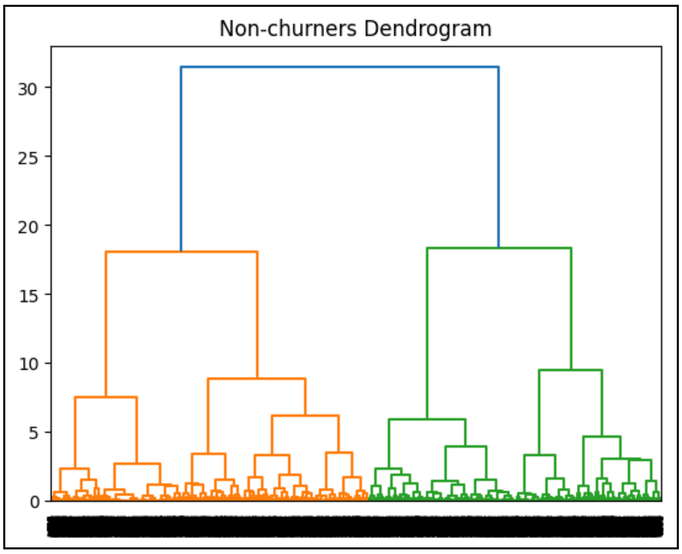

Figure 15: Non-churners' Data Dendrogram

Figure 16 shows a scatter plot of non-churned customer data after clustering using agglomerative hierarchical clustering with 2 different clusters.

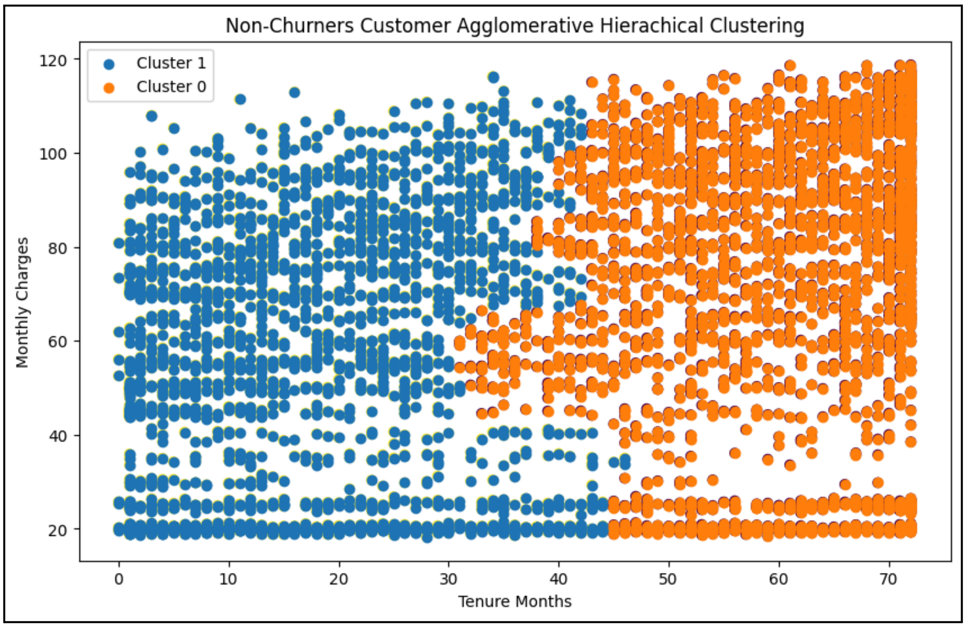

Figure 16: Non-churners' Customer Clusters using Hierarchical Clustering

C. DBSCAN CLUSTERING

Finding optimal Epsilon Value (Eps) for DBSCAN

Figure 17 shows a K-distance plot for churned customers data to find the optimal Eps for DBSCAN Clustering. It is identified that the maximum curvature of the graph is at 0.03. The dataset has two dimensions so, the minimum sample per cluster is initialised as 4. Therefore, eps=0.03 and min_samples=4.

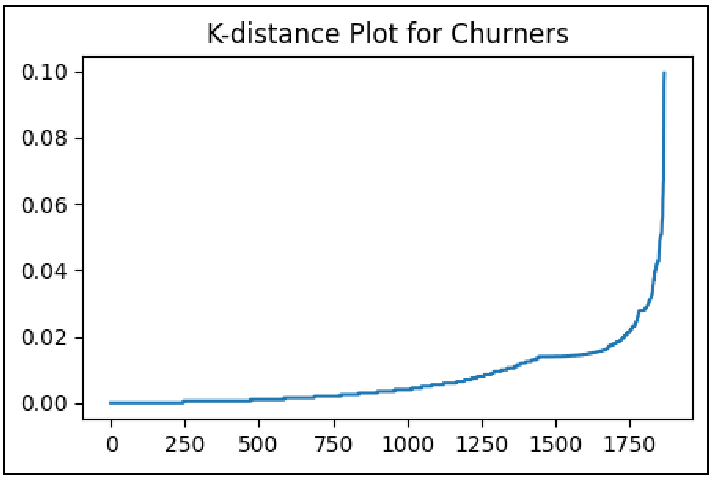

Figure 17: Graph to find optimal Eps Value for churners

Figure 18 shows a scatter plot of churned customer data after clustering and it can be seen that the clusters are poorly separated with the majority of the data points only belonging to only one cluster.

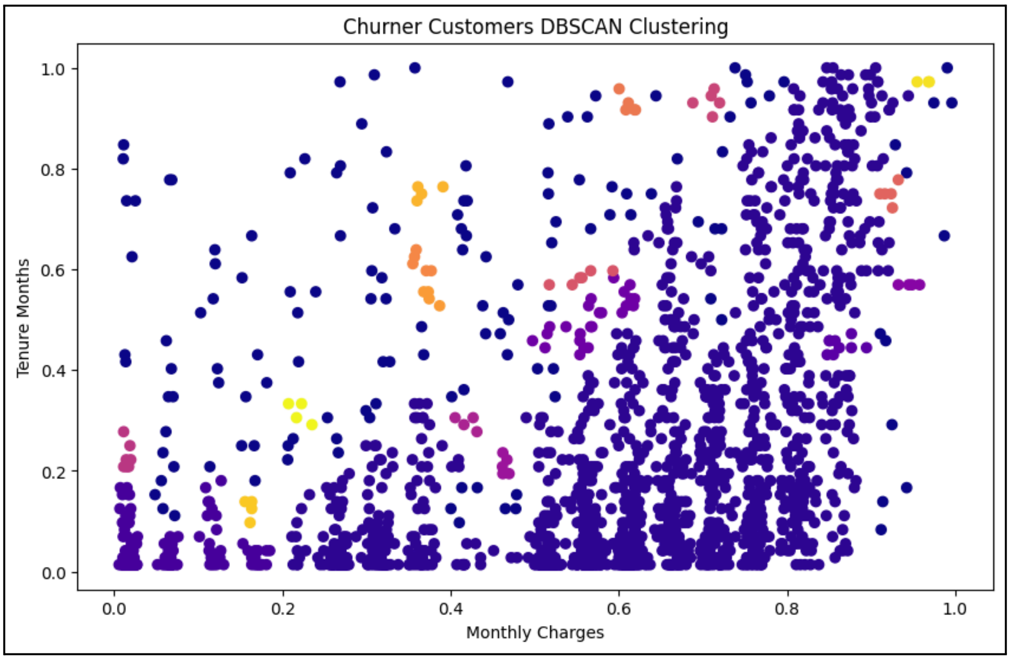

Figure 18: Clustering for churners using DBSCAN

Figure 19 shows a K-distance plot for non-churned customers data to find the optimal Eps. It is identified that the maximum curvature of the graph is at 0.015. The dataset has two dimensions so, the minimum sample per cluster is initialised as 4. Therefore, eps=0.015 and min_samples=4.

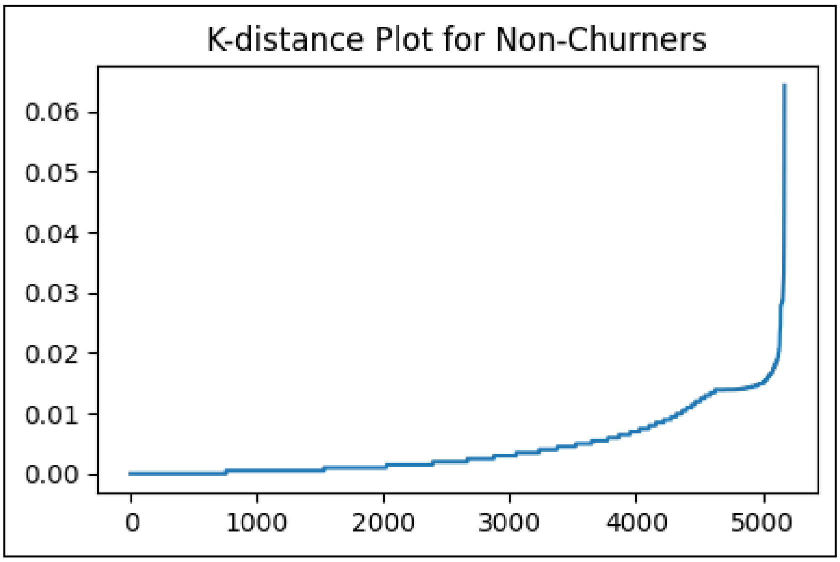

Figure 19: Graph to find optimal Eps Value for non-churners

Figure 20 shows a scatter plot of non-churned customer data after clustering and it can be seen that all the clusters are jumbled up among each other with no clear separations.

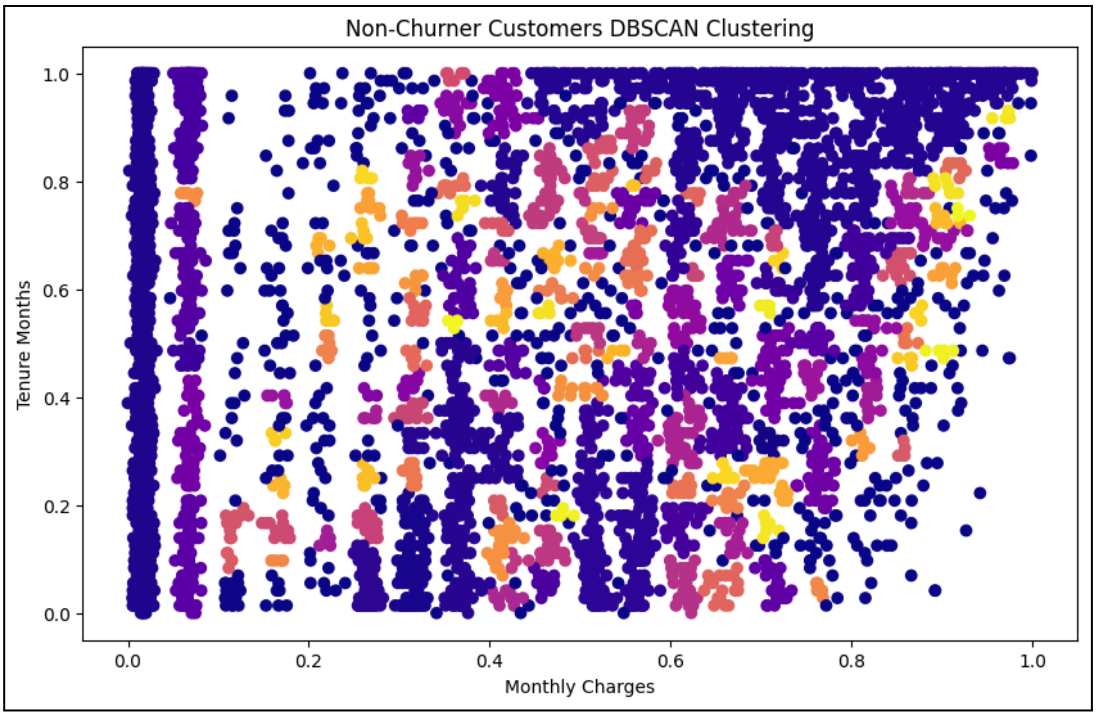

Figure 20: Clustering for non-churners using DBSCAN

Based on Figure 18 and 20, it is found that DBSCAN is not effective at all in identifying any clear clusters within the dataset. Therefore, it is not suitable to be used in this dataset in clustering the customers.

D. SEGMENTATION MODEL EVALUATION AND RESULTS

Table 4 shows the evaluation result for both clustering algorithms on churned customers.

For Churners Clustering (3 clusters)		
	K-Means Clustering with PCA	**Agglomerative Hierarchical Clustering**
Silhouette Coefficient	(0.51)	0.48
Sum of Squared Error (SSE)	(66.404)	75.26
** value that is underlined within brackets ( ) is the better model		

Table 4: Evaluation for churners clustering models

Table 5 shows the evaluation result for both clustering algorithms on non-churned customers.

For Non-Churners Clustering		
	K-Means Clustering with PCA (4 clusters)	**Agglomerative Hierarchical Clustering (2 clusters)**
Silhouette Coefficient	(0.47)	0.41
Sum of Squared Error (SSE)	(214.10)	580.95
** value that is underlined within brackets ( ) is the better model		

Table 5: Evaluation for non-churners clustering models

Based on the model evaluation above in Table 4 and 5, it is shown that the K-Means Clustering with PCA model performs better than the Agglomerative Hierarchical Model when it is evaluated with Silhouette Coefficient and Sum of Squared Error (SSE).

Meanwhile, DBSCAN Clustering Algorithm was also used to cluster the data as suggested by previous research works but it is proven to be not effective with this IBM Telco Dataset due to the nature of the dataset. The dataset is highly dense without any low-density areas hence, it is difficult for DBSCAN to properly cluster the data.

E. CONTENT-BASED FILTERING ON TELCO SERVICE PACKAGE RECOMMENDATIONS

The recommendation algorithm is trained using Cosine Similarity between User Profile and Telco Services Packages. The higher the similarity, the more relevant the package is to the customer. For this, IBM dataset does not have a predefined telco package offers therefore, the following table is the self defined packages based on the most common subscription used by customers.

Table 10 shows six different Telco packages with its different offerings.

Telco Packages	Offers
Package 1	Phone service + No Internet Service + Not eligible Internet Subscriptions
Package 2	Phone service + Fiber optic + No additional Internet Subscriptions
Package 3	Phone service + DSL + No additional Internet Subscriptions
Package 4	Phone service + Fiber optic + Streaming TV + Streaming Movies
Package 5	Phone service + Fiber optic + Online Backup + Device Protection + Streaming TV + Streaming Movies
Package 6	Phone service + DSL + All Internet Subscriptions

Table 10: Telco Service Packages

Based on the IBM Dataset, it is found that the most relevant and similar packages to the existing customers' subscription are Package 1, 4, 5, and 6. Therefore, only these four different Telco packages are recommended to customers in the Recommender System.

DISCUSSION

A. CUSTOMER SEGMENTATION MODEL

This paper has proposed the use of K-Means Clustering with PCA, Agglomerative Hierarchical Clustering and DBSCAN Clustering to create and train a customer segmentation model for Telecommunication companies.

Based on the findings, DBSCAN could not identify any clear clusters due to the nature of the IBM dataset. The dataset is highly dense without any low-density areas hence, it is difficult for DBSCAN to properly cluster the data. DBSCAN is most efficient when handling outliers and noises in data but the dataset in this paper does not have very minimal outliers and noises. This clustering method would be best applied for real-world noisy data.

Meanwhile, when comparing PCA K-Means Clustering with Agglomerative Hierarchical Clustering, it is proven by both evaluation methods, Silhouette Coefficient and Sum of Squared Errors (SSE) that, K-Means Clustering with PCA performs better in finding clearer segments between customers.

Therefore, K-Means Clustering with PCA is chosen as the final best model to be used in clustering both churners and non-churners customers within the Telecommunication Company.

B. CUSTOMER CLUSTER CHARACTERISTICS AND SUGGESTED MARKETING STRATEGIES FOR ACTIONABLE INSIGHTS

Descriptive Analytics is performed on all the identified customer segments through its unique characteristics so that marketers or business users can understand each segment easily and thoroughly.

Then, prescriptive analytics is performed through various suggestions of marketing strategies on each segment so that decision makers can take action to help improve customer retention rate and satisfaction.

Churners Customers Clusters

Table 6 shows the characteristics and statistics of each different identified cluster in churned customers data.

<table>
<tr><th rowspan="2">Churners Clusters</th><th rowspan="2">CLTV</th><th colspan="2">Monthly Charges</th><th colspan="2">Tenure Months</th></tr>
<tr><th>Average</th><th>Range</th><th>Average</th><th>Range</th></tr>
<tr><td>The Short-Lived Moderate Spenders</td><td>$3,003.88</td><td>$83.82</td><td>$60.05 to $112.95</td><td>9 months</td><td>1 month to 29 months</td></tr>
<tr><td colspan="6">● Moderate CLTV Cluster
● Relatively high monthly spending with short tenure months.
● Highest likelihood to churn due to their short tenure with the telco company.
● They may have joined with high expectations but found no long-term value with the telco service.
● They might be potential high-valued customers if they stayed loyal due to their high monthly spending.</td></tr>
<tr><td>The Economical Explorer</td><td>$1,102.96</td><td>$38.34</td><td>$18.85 to $60.40</td><td>7 months</td><td>1 month to 53 months</td></tr>
<tr><td colspan="6">● Lowest CLTV Cluster
● Their low monthly spending suggests their mindful trait in spending.
● Their average tenure suggests that they are still exploring different telco subscriptions.
● They are likely to seek cost-effective telco subscriptions.</td></tr>
<tr><td>The Loyal High-Value Enthusiasts</td><td>$17,617.98</td><td>$90.73</td><td>$19.35 to $118.35</td><td>47 months</td><td>27 months to 72 months</td></tr>
</table>

- Significantly high CLTV Cluster which can contribute to the company's revenue
- They have relatively high spending showing that they are willing to and enthusiastic about premium telco services.
- Despite having high tenure months with the telco service, they still churn due to several reasons.
- A significant cluster that requires more attention to keep them from churning.
- They seek and appreciate exclusive benefits and premium experience from the company.

Table 6:Churner Clusters' Statistics

Table 7 suggests different actionable marketing strategies that can be used on each different cluster found in churned customers.

Churners Clusters	**Suggested Marketing Strategy**
The Short-Lived Moderate Spenders	Educational and Community Marketing ● Highlights the need to educate them on the benefits they may have overlooked during their short tenure ● Educational Marketing ○ provide tutorial contents to educate customers on how to maximize their usage and introduce them to more hidden benefits. ○ highlight long-term benefits and offer incentives to extend their subscription such as loyalty discounts. ● Community Marketing ○ engaging long-term customers into the community to share their stories and testimonials with the telco services. ○ organize engagement events to promote loyalty activities and rewards collections.
The Economical Explorer	Promotional and Freebie Marketing ● Highlights the need to promote more cost-effective telco service packages and exploration of new subscription deals. ● Promotional Marketing ○ offers affordable plans with promotional discounts for those who subscribe within that period of time. ○ emphasize on competitive pricing between different telco

	subscriptions and the current one. ● Freebie Marketing ○ offers them free trials on new telco services packages to encourage them to explore with confidence. ○ include free gifts to them when they sign up for a subscription plan.
The Loyal High-Value Enthusiasts	Loyalty Enhancement and Community Marketing ● Highlights the need to reward their continued loyalty. ● Loyalty Enhancement ○ provide exclusive benefits and top-tier telco deals to enhance their overall experience. ○ offer loyalty programs, VIP access and early access to new telco features so that they continue to stay loyal and not churn. ● Community Marketing ○ organize engagements to seek feedback on improving the services and show appreciation to them if they continue to stay loyal.

Table 7:Churner Clusters' Suggested Marketing Strategies

Non-Churners Customers Clusters

Table 8 shows the characteristics and statistics of each different identified cluster in non-churned customers data.

<table>
<tr><th rowspan="2">Non- Churners Clusters</th><th rowspan="2">CLTV</th><th colspan="2">Monthly Charges</th><th colspan="2">Tenure Months</th></tr>
<tr><th>Average</th><th>Range</th><th>Average</th><th>Range</th></tr>
<tr><td>The Budget Conscious</td><td>$1,384.40</td><td>$29.20</td><td>$18.25 to $57.55</td><td>12 months</td><td>0 months to 33 months</td></tr>
<tr><td colspan="6">Lowest CLTV Cluster
Low to moderate monthly spending and short to medium tenure.
They are price-sensitive and prioritize cost-saving over long-term loyalty.
They have a high likelihood of churning if they find better telco service deals elsewhere.</td></tr>
<tr><td>The Moderate Users</td><td>$6,416.63</td><td>$77.30</td><td>$48.80 to $116.25</td><td>20 months</td><td>0 to 44 months</td></tr>
<tr><td colspan="6">Moderate CLTV Cluster
Moderate monthly spending and medium tenure.
They have a balanced approach in utilizing their subscription package by not being excessively high spenders.</td></tr>
<tr><td>The Seasoned Explorers</td><td>$7,169.50</td><td>$31.91</td><td>$18.40 to $64.20</td><td>55 months</td><td>33 months to 72 months</td></tr>
<tr><td colspan="6">Relatively high CLTV Cluster
Low average monthly spending with long tenure.
They have been subscribing to the telco service for a white but their lower spending suggests that they are exploring different deals.
They may have diverse interests to try different telco deals.</td></tr>
<tr><td>The Premium Lifers</td><td>$22,943.45</td><td>$91.76</td><td>$59.50 to $118.75</td><td>61 months</td><td>38 months to 72 months</td></tr>
</table>

● Highest CLTV Cluster ● High monthly spending and long tenure with the company. ● They are a valuable asset to the company due to their consistent loyalty in subscribing to the telco service as well as having high spending.

Table 8:Non-Churner Clusters' Statistics

Table 9 suggests different actionable marketing strategies that can be used on each different cluster found in non-churned customers.

Non-Churners Clusters	Suggested Marketing Strategy
The Budget Conscious	Promotional and Community Marketing ● Highlights the need to promote affordability and long-term benefits. ● Promotional Marketing ○ offer budget-friendly telco services plans such as discounted telco bundles. ○ introduce flexible payment options for the telco subscription. ● Community Marketing ○ launch referral programs to provide additional discounts on each successful referral. ○ engage high-valued customers as testimonials to showcase their loyalty benefits.
The Moderate Users	Upselling and Cross-Selling Marketing ● Highlights the need to encourage them to explore different subscription plans. ● Upselling Marketing ○ encourage them to experience better telco subscriptions to cater to their increasing usage and needs. ○ this can be a stepping stone for them to start subscribing to higher-end plans. ● Cross-Selling Marketing ○ requires market basket analysis to suggest customers to subscribe to additional services that are similar to their

	current one so that they can enhance their experience and usage.
The Seasoned Explorers	Promotional Marketing and Diverse Offerings ● Highlights the need to emphasize on the variety of telco services packages and promote new features. ● Offer different special discounts to try and explore new packages to foster their interest in other packages. ● Create a plan to come out with different varieties of packages as well as personalized recommendations based on usage history and preferences.
The Premium Lifers	Exceptional Customer Service ● Highlights the need to prioritize their needs and wants with the telco service. ● Crucial to reinforce their perception of the company as a premium telco service provider. ● Exclusive offers as appreciation for loyalty and premium upgrade services.

Table 9:Non-Churner Clusters' Suggested Marketing Strategies

C. TELCO PACKAGES RECOMMENDER SYSTEM

Cosine Similarity between each customers' User Profile and six different Telco Services Packages, has found to be applicable for this IBM Telco Dataset. Based on customers' historical telco subscriptions, the company will be able to recommend them with the most relevant package for them to subscribe to. As a result, the company will be able to save time in finding the best package deal to attract customers to continue their subscription hence improving customer satisfaction and loyalty.

From the data findings, it is known that Package 6 with Phone service, DSL and All Internet Subscriptions Offerings, is the most popular package subscribed by customers from both churners and non-churners category.

Based on this recommendation and data about Telco packages, the Telecommunication Company can further understand their customers' preferences and make an effort to improve their packages with the goal to attract more customers while creating competitive advantage.

D. INTERACTIVE DASHBOARDING USING MICROSOFT POWER BI

Interactive Dashboard helps non-technical business users within the Telecommunication Company to easily understand the analysed data in an overview without having to use all the trained models. In this paper, Microsoft Power BI is used and visualisation regarding churners and non-churners data in each segment has been done.

Figure 21 shows the Dashboard for non-churned customers within the Telecommunication company. It shows the number of customers in each cluster and their respective Telco Packages, genders, dependents and cities. These different clusters can be selected as filter to view it detailedly.

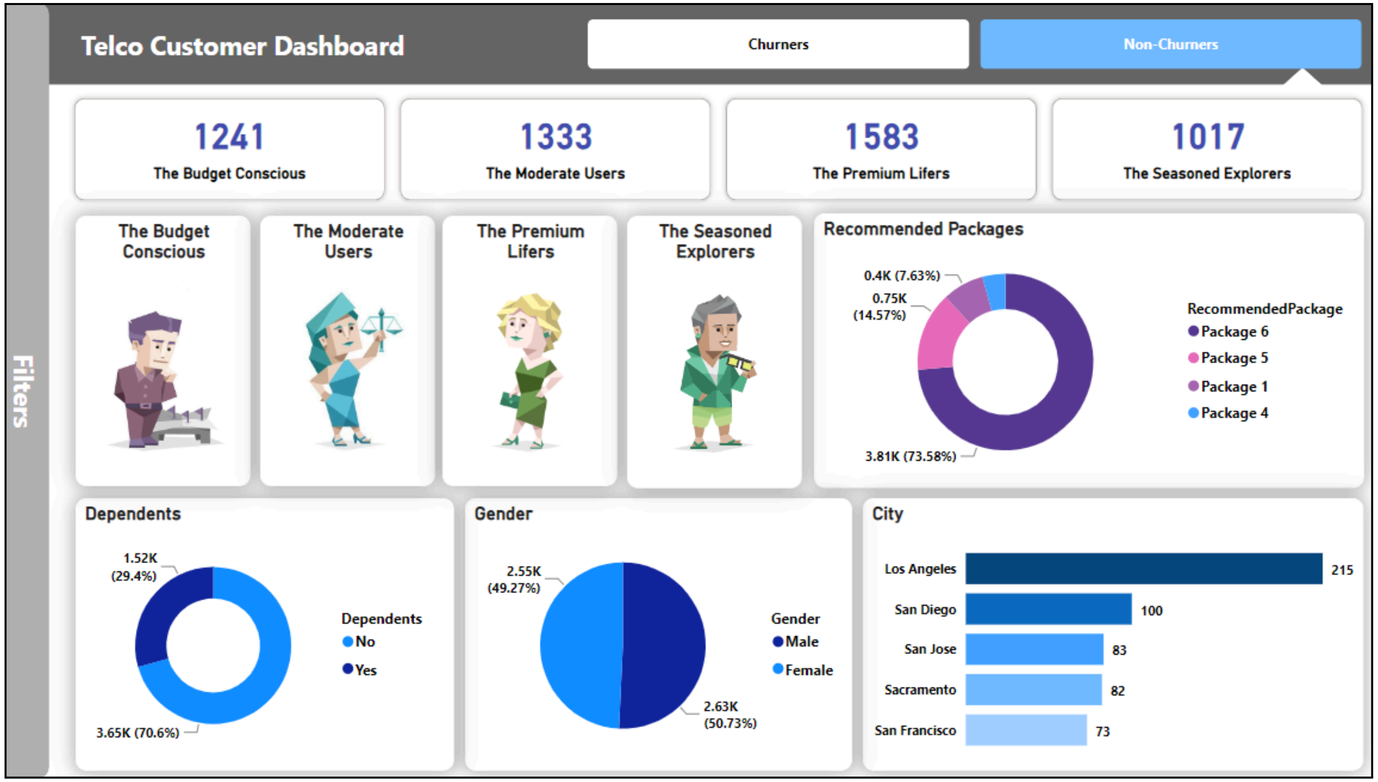

Figure 21: Non-Churner Customers' Dashboard

CONCLUSION

In this paper, K-Means Clustering with PCA on continuous variables have been proposed to train a customer segmentation model to segment telecommunication customers into different clusters in order to mitigate the decreasing customer retention rate.

Descriptive analysis is performed on each customer segment to interpret its unique characteristics based on three criterias, Segment's Customer Lifetime Value (CLTV), Monthly Charges and Tenure Months. Segment CLTV is calculated based on a modified version of the formula so that it applies to the dataset. From this value, telecommunication companies can understand which segments contribute the most to the company's revenue. Then, prescriptive analytics was used to suggest different marketing strategies to each segment so that marketers can keep attracting these customers to continue their subscription.

Meanwhile, content-based filtering was used to build the recommender system using cosine similarity between customer's user profile against all the six available telco packages. Recommending the relevant package to customers can create competitive advantage and make them stay loyal and satisfied with their subscription.

Lastly, interactive dashboarding provides non-technical users an easy-to-understand interface to get fast valuable insights into their customer base. This allows better understanding of their customers' statistics and helps them plan out their business outlook for future improvements.

There are also several suggestions for future improvement works:

a. Automatic interpretation of customer segments

- Customer segments and needs may change overtime as the customer base grows and manual interpretation done once cannot always represent the segments therefore, future work can look into methods on creating an automated interpreter for the segments so that it can always update the customers' needs accordingly.

b. Automatic suggestion of marketing strategies on each segments

- Based on each segments' characteristics and needs, marketing strategies are manually suggested to the marketers and there might be experience bias from the expert suggesting the marketing strategy. Different experts might have different approaches to different customer characteristics. Future work can work on training an automated model to predict and suggest marketing strategies to different segments.

REFERENCES

[1] Y.-C. Tsai et al. (2015) 'A Study of the Relationship among Brand Experiences Self-Concecpt Congruence Customer Satisfaction and Brand Preference', Contemporary Management Research, pp. 97-116.

[2] P. K. Hellier (2003) 'Customer Repurchase Intention: A General Structural Equation Model', European Journal of Marketing, pp. 1762-1800.

[3] Eman Hussein Sharaf Addin, Novia Admodisastro, Siti Nur Syahirah Mohd Ashri, Azrina Kamaruddin & Yew Chew Chong (2022) Customer Mobile Behavioral Segmentation and Analysis in Telecom Using Machine Learning, Applied Artificial Intelligence, 36:1, DOI: 10.1080/08839514.2021.2009223

[4] Wang, Y., Li, J., Li, Y., & Li, Q. (2021) 'A customer segmentation model for telecom industry based on K-means clustering algorithm'. IEEE Access, 9, 16207-16215

[5] Fitri Marisa, Sarifah Shakinah Syed Ahmad, Zeratul Izzah Mohd Yusof, Fachrudin Hunaini, & Tubagus Mohammad Akhriza Aziz. (2019). Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM Model. International Journal of Integrated Engineering, 11(3).

[6] IBM (2022) 'Telco customer churn'

[7] Iaquinta, L., Gentile, A. L., Lops, P., de Gemmis, M. and Semeraro, G., (2007). A Hybrid Content-Collaborative Recommender System Integrated into an Electronic Performance Support System, 7th International Conference on Hybrid Intelligent Systems, Kaiserlautern, pp. 47 – 52.

[8] C. D. Pham et al., "A recommendation system for offers in telecommunications," 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), Phu Quoc Island, Vietnam, 2021, pp. 302-306, doi: 10.1109/ICCE48956.2021.9352111.

[9] Zui Zhang, Kun Liu, William Wang, Tai Zhang and Jie Lu, (2007). A Personalized Recommender System for TelecomProducts and Services. Decision Systems & e-Service Intelligence Lab, Centre for Quantum Computation & Intelligent Systems Faculty of

Engineering and Information Technology, University of Technology PO Box 123, Broadway, NSW 2007, Sydney, Australia.