BISM2202

Data analytics and information management

Telecom customer churn report provided by analytics

Student Name	MUHAMMAD AL-AMIN BIN ZULKIPLI
Student Number	46748104

Semester 2, 2021

Telecom Customer Churn Report Provided by AnalyticUs

1. **Executive Summary:**

The telecommunications business must simply determine why customers have chosen to quit. The simplest method is to speak with the consumer.

In the business perspectives, power Bi and RapidMiner software indicates customer churn were most contribute by complaint attributes. If telco company review the data in the perspective of each customer from different data tariff plan, tariff plan 4 have the highest cost loss when the customer churn from the service which around $ 2000. Acquiring new customer are far more expensive than it is to retain the existing one, this will cost the company a lot. The company should keep valuable customer which they want to keep, valuable customer should keep in tight for increase the company revenue. To significantly address customer churn is to provide better customer service which give far better resolution in telco industry service. Another method to avoid churn is to keep your consumers actively engaged with your offering. Give your consumers reasons to keep coming back by demonstrating the day-to-day value of utilizing your services by incorporating your services into their daily routine.

Other than that, customer count in tariff plan 2 or below $ 1250 net customer spend have the biggest impact in company revenue loss with total around $ 200,000. Even though, the amount per person are lowest than Tariff plan 3 and 4 but the customer subscription amount were the highest in the service. So, Telco company should provide more cheaper price plan and increasing in the number data usage available and the quality of data service to the customer which result in less customer churn.

2. **Data transformations via SSIS**

2.1. There are certain issues with the current data flow at its current state. Please fix all current errors in the data transformation process and expand the process to exporting three csv files that are ready for import into a Datawarehouse:

(1) **customer dimension** table (contains all available customer attributes)

(2) **product dimension** table (contains all available mobile plan attributes)

(3) **fact table** (contains all available measures)

2.1.1. Data Flow screenshot:

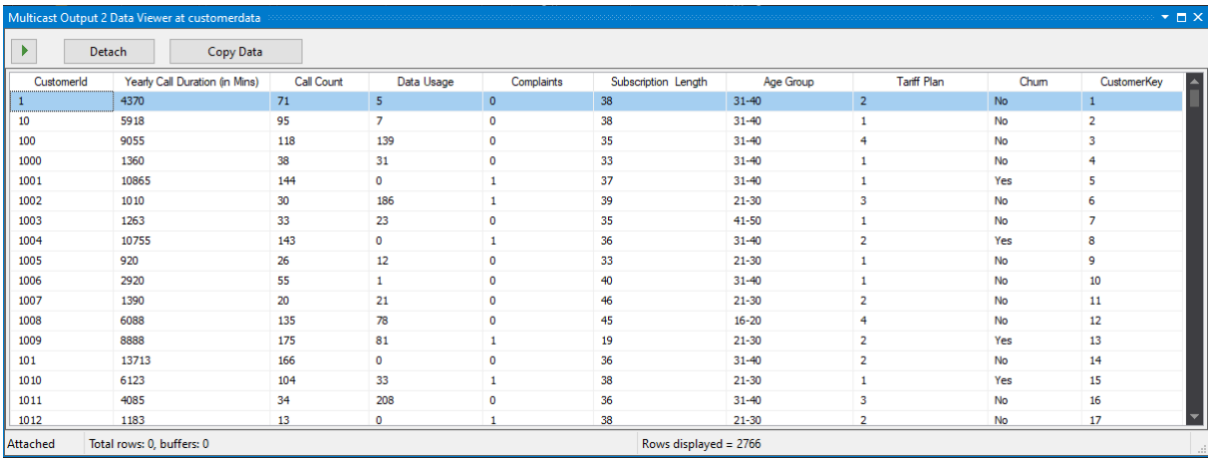
2.1.2. Description of Data Flow:

1.

1	Add “connection managers” rename it Cust 1
2	Add “connection managers” rename it Cust 2
3	Add “sort” after Cust 1 Flat File Source (Ascending: Customer Id)
4	Add “Derived columns” after Cust 2 flat file source Expression:(DT_STR,30,1252) RIGHT([CountryCode&CustomerId],LEN([CountryCode&CustomerId]) FINDSTRING([CountryCode&CustomerId],"+",1)) string [DT_STR]
5	Add “sort 3” after Derived columns (Ascending: Customer Id)
6	Add “union All” for Cust 1 and 2 Flat File Source (customerID)
7	Add “sort” after Union All (Ascending: Customer Id)
8	Add “script component”
9	Add column “CustomerKey”
10	Add “multicast” after Script component
11	Add “ FlatFileDestination” after multicast for customer output table
12	Add “sort 4” after multicast (ascending: Plan ID)
13	Add “connection managers” rename it MobilePlans
14	Add “sort” after mobileplans Flat File Source (Ascending: Plan Id)
15	Add “script component” after Sort 1 for MobilePlans
16	Add output column - Add: product Key
17	Add “multicast” after MobilePlansScript component
18	Add Flat File destination after multicast
19	Add “Merge Join” link by Plans Id(Mobile Plans) and TariffPlan (Customer Consumption Data)
20	Add “Derived Column” after Merge Join - PlanName==IPTeleConnect Plan XL “?”Yes”:”No”
21	Add “Flat File Destination” for Customer Churn Fact Table”

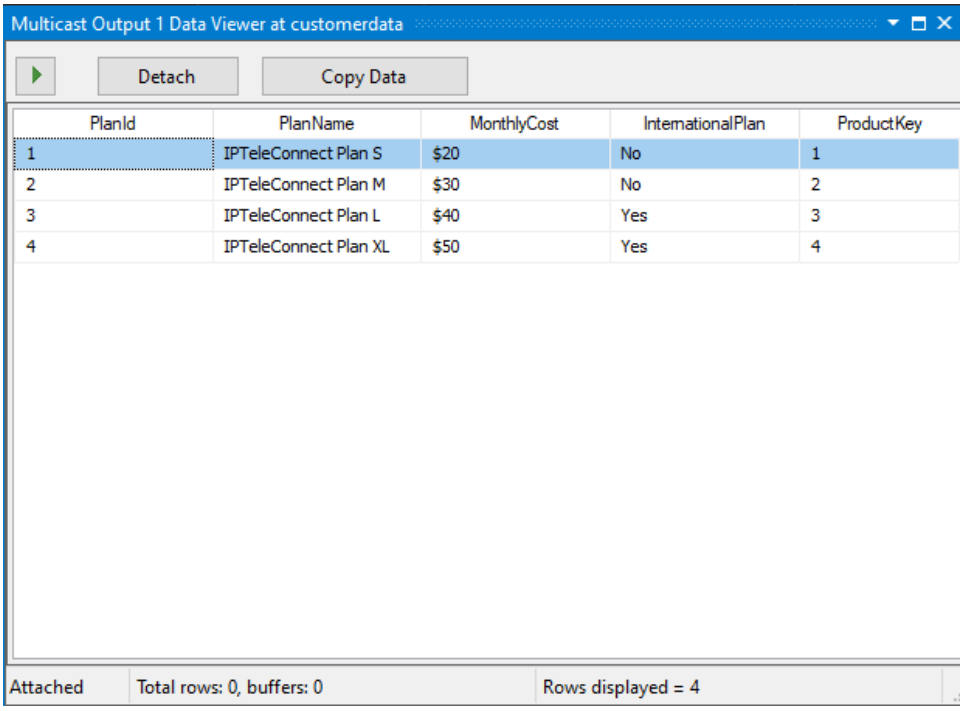
2.1.3. Deliverable:

(1) **Customer Table (Dimension 1**)



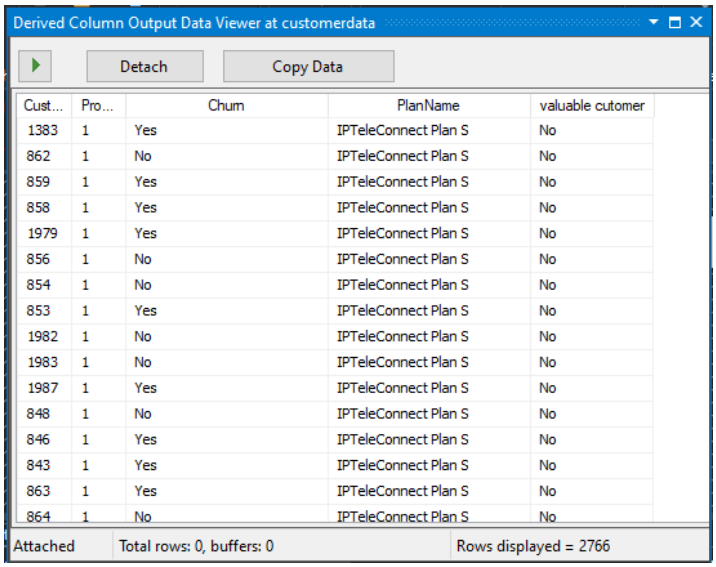
CustomerId	Yearly Call Duration (in Mins)	Call Count	Data Usage	Complaints	Subscription Length	Age Group	Tariff Plan	Churn	CustomerKey
1	4370	71	5	0	38	31-40	2	No	1
10	5918	95	7	0	38	31-40	1	No	2
100	9055	118	139	0	35	31-40	4	No	3
1000	1360	38	31	0	33	31-40	1	No	4
1001	10865	144	0	1	37	31-40	1	Yes	5
1002	1010	30	186	1	39	21-30	3	No	6
1003	1263	33	23	0	35	41-50	1	No	7
1004	10755	143	0	1	36	31-40	2	Yes	8
1005	920	26	12	0	33	21-30	1	No	9
1006	2920	55	1	0	40	31-40	1	No	10
1007	1390	20	21	0	46	21-30	2	No	11
1008	6088	135	78	0	45	16-20	4	No	12
1009	8888	175	81	1	19	21-30	2	Yes	13
101	13713	166	0	0	36	31-40	2	No	14
1010	6123	104	33	1	38	21-30	1	Yes	15
1011	4085	34	208	0	36	31-40	3	No	16
1012	1183	13	0	1	38	21-30	2	No	17

(2) **Mobile Plan Product Table** (Dimension 2)



PlanId	PlanName	MonthlyCost	InternationalPlan	ProductKey
1	IPTeleConnect Plan S	$20	No	1
2	IPTeleConnect Plan M	$30	No	2
3	IPTeleConnect Plan L	$40	Yes	3
4	IPTeleConnect Plan XL	$50	Yes	4

(3) **Fact Table** (Fact Table)

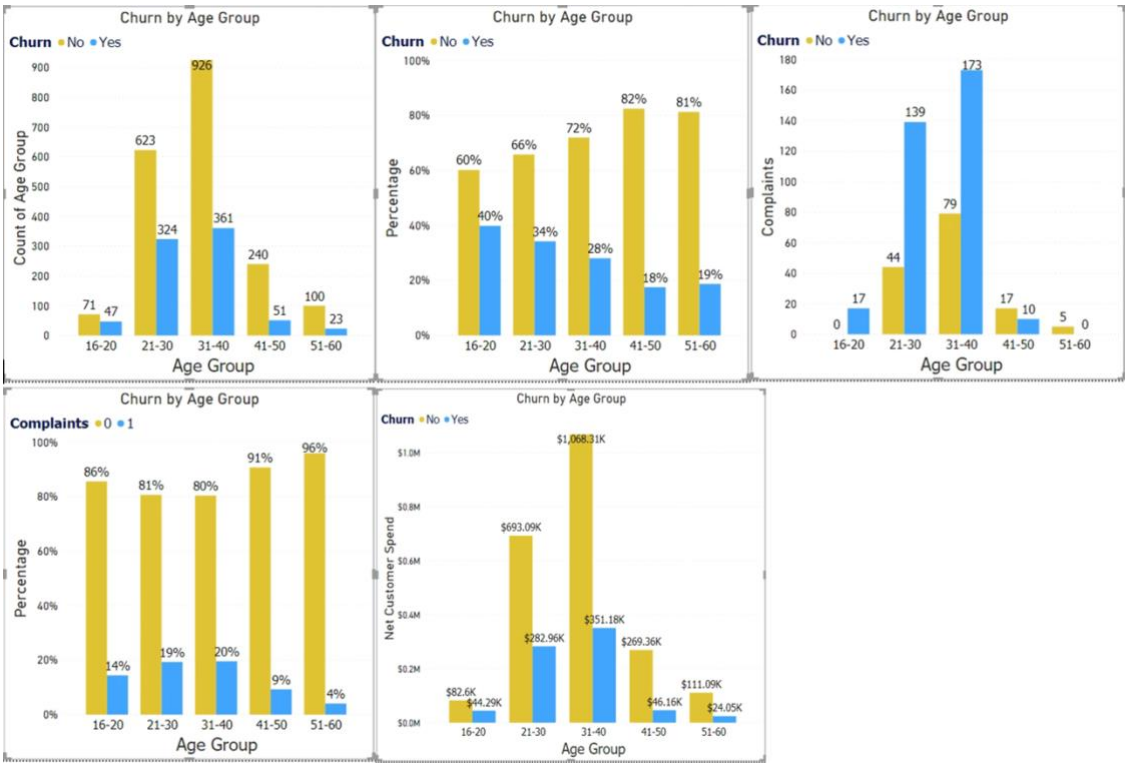


Cust...	Pro...	Churn	PlanName	valuable cutomer
1383	1	Yes	IPTeleConnect Plan S	No
862	1	No	IPTeleConnect Plan S	No
859	1	Yes	IPTeleConnect Plan S	No
858	1	Yes	IPTeleConnect Plan S	No
1979	1	Yes	IPTeleConnect Plan S	No
856	1	No	IPTeleConnect Plan S	No
854	1	No	IPTeleConnect Plan S	No
853	1	Yes	IPTeleConnect Plan S	No
1982	1	No	IPTeleConnect Plan S	No
1983	1	No	IPTeleConnect Plan S	No
1987	1	Yes	IPTeleConnect Plan S	No
848	1	No	IPTeleConnect Plan S	No
846	1	Yes	IPTeleConnect Plan S	No
843	1	Yes	IPTeleConnect Plan S	No
863	1	Yes	IPTeleConnect Plan S	No
864	1	No	IPTeleConnect Plan S	No

3. **<u>Descriptive analysis of customer churn via PBI</u>**

3.1. The marketing team argues about whether customer age is a driver of churn. The head of the marketing department suggests that particularly old people like to file complaints as they become grumpy and then churn.

3.1.1. Screenshot(s) of relevant visualizations regarding churn predictors (Age):

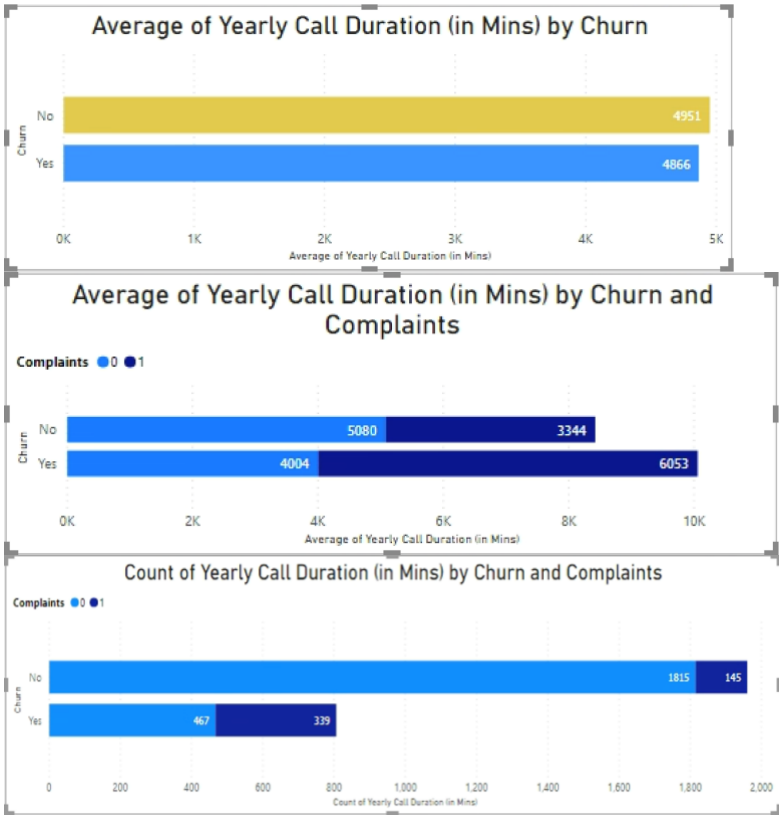


3.1.2. **Discussion of data churn predictors:**

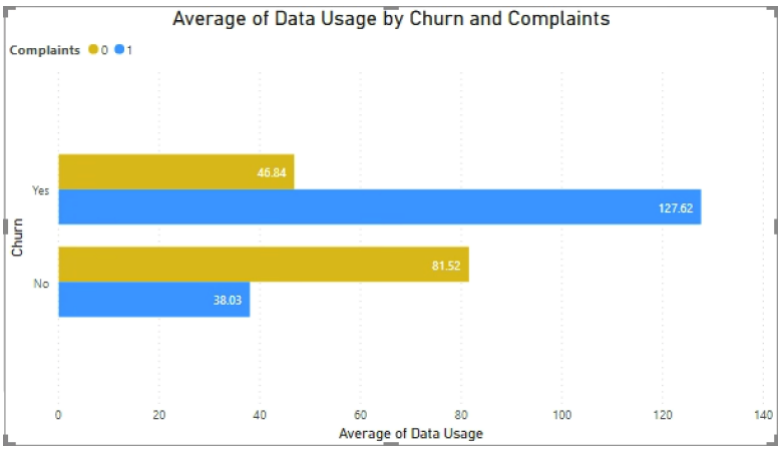
I am totally disagree with this assumption it is because the people in the age 31- 40 have the higher complaint percentage of complaint (20%) as compared to age group 41 above (9% + 4% = 13%). So, the older people from age 41 and above are less likely to complaint and churn as compared to the age group 16 to 40. The assumption was inverse with data result given in the power Bi. As shown, the lower the age (16-20), the higher percentage of churn by age group (40%). For example, age group 16-20 has the highest percentage of churn among other age group.

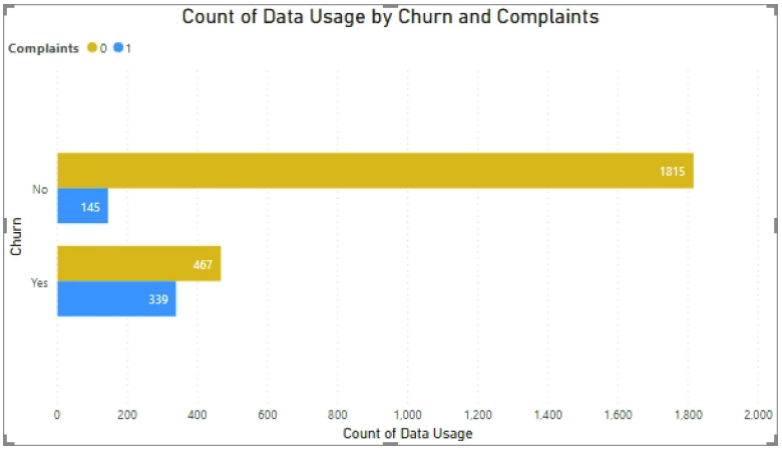
3.2. The marketing team also believes that the customer behaviour will give strong indication for whether or not a customer usually churns.

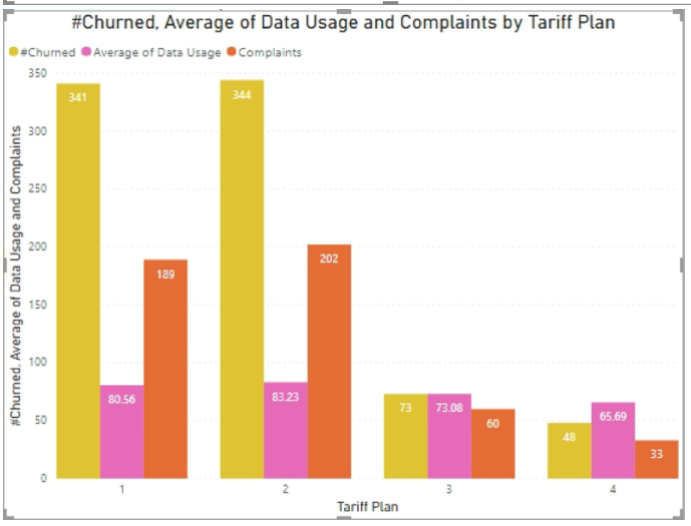
3.2.1. Screenshot(s) of relevant visualizations regarding churn predictors:



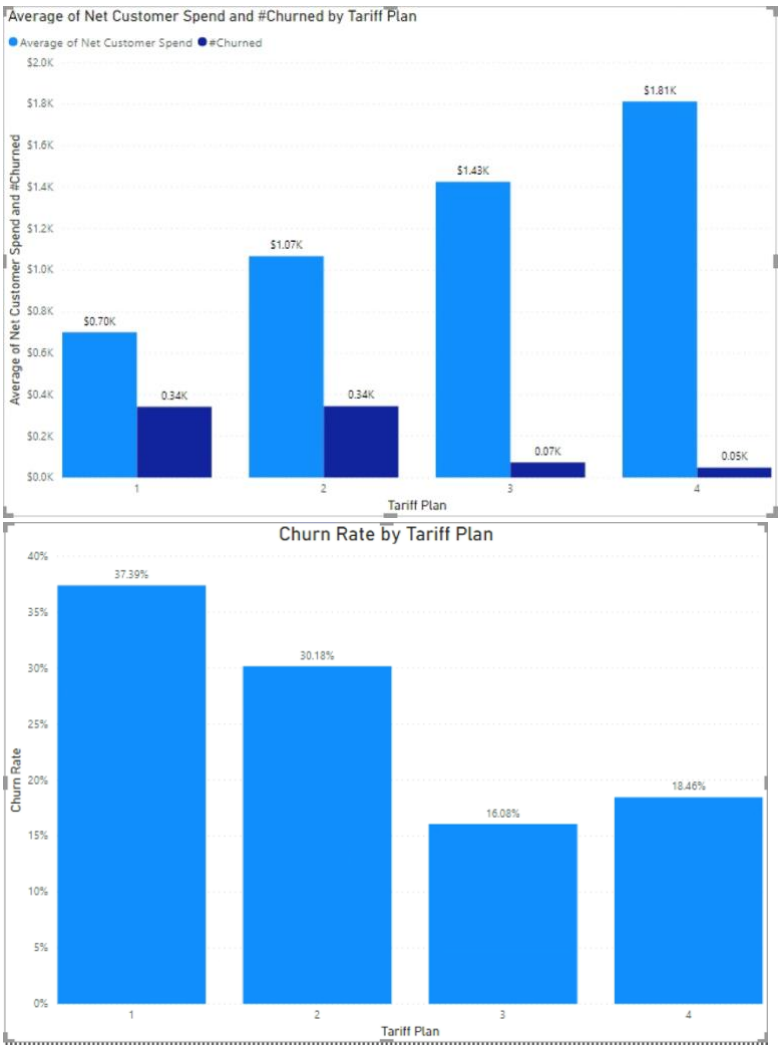
Assumption 1: People who <u>speak longer on the phone</u> will have a better chance at experiencing how good the company's service is and are less likely to churn







Assumption 2: people who use more data are able to experience the full benefit of the company and are less likely to churn.



Assumption 3: people who spend more money with the company receive the best service and are less likely to churn.

3.2.2. Discussion of data churn predictors:

I totally disagree and the assumption given should look over by the company management. Overall customer average of yearly call duration will complain and churn as they consumed more duration of call within the year as compared to the customer with lower average of yearly call and did not churn. There is only 145 customer who didn't churn with lower average of yearly call duration (3344 minutes) as compared to the customer churn (339) with higher average call duration (6053 minutes).

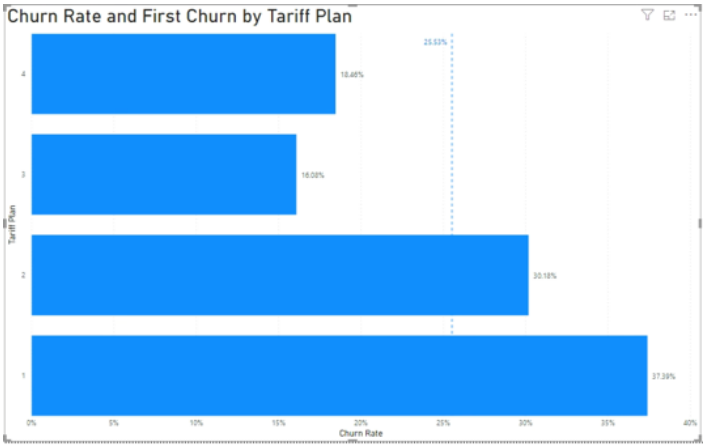
I totally disagree with the statement and the assumption given are not valid to use because it inverses with the data visualization summary. The overall customer average weekly data usage will churn and complaints more if they consumed more data usage in the period of week. There is 339 customer churns

with higher average data usage (127.62 data) as compared to the customer who didn't churn (145) with lower weekly data usage (38.03 data).

The assumption given were valid and logic as the visual show more customers did not churn if they spend more to the service they choose. From the overall perspective, the customer on each tariff plan will not churn if they spend more compared to spend less in the same tariff plan. Tariff 4 average net customer spend ($ 1,842) are less likely to churn compared to customer in tariff plan 1 with average net customer spend is $ 700 have bigger churned customer (340).

3.3. The product management teams that are responsible for the different mobile plans (Tariff Plan 1-4) are concerned about their upcoming performance evaluation. The product teams (for Tariff Plans 1-4) suffer bonus deductions if their performance is below the company average. The company leadership has decided to use the churn rate as the measure for the team's performance. Please explain which of the four Tariff Plan teams need to be concerned about performance bonus deductions.

3.3.1. Deliverable: Screenshot(s) of relevant visualizations regarding tariff plan performance



1.1. Please explain which of the four Tariff Plan teams need to be concerned about performance bonus deductions.

1.1.1. Discussion of tariff plan performance:

Tariff plan 1 and 2 teams should give more attention regerding employee bonus deduction it is because tariff plan 1 and 2 were passed the limit given by the company leadership executive. Tariff plan 2 have 30.18% percent of customer churn which below the company average churn rate (25.53%). However, company should give more concerned to tariff plan 1 because highest churn rate (37.39%) compared to tariff plan 2.

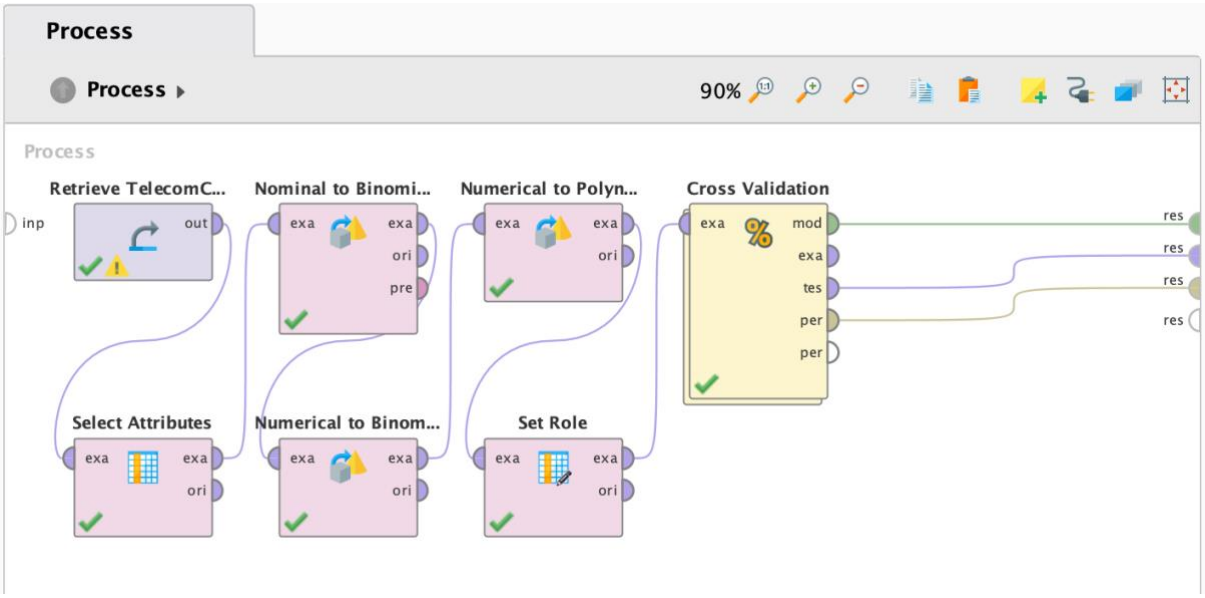
Nevertheless, company churn should be churn according to the specific tariff plan, for example, tariff plans 3 and 4 should have lower churn rate average because the value of net customer spend were bigger as compared to the tariff plan 1 and 2. So, if the company set the customer churn rate bigger in tariff plan 3 and 4, the company will get loss on the profit income.

Churn rate

4. Predictive analysis of customer churn via RapidMiner:

1.2. The team know that you are interested in classifying whether or not customer churn.

1.2.1. Deliverable: **Screenshot of Analytical Process**



1.2.2. Deliverable: **Description of Process**

- ID: Data Usage
- Label: Tariff Plan
- Add “Filter Examples”: Condition “CustomerID = 1”
- Put ‘select Attributes’: attributes filter type = all
- Add “Nominal to binominal” attributes filter type = single, attribute = churn
- Add “Numerical to polynomial” attributes filter type = subset, attribute = complaint & Tariff plan
- Add “Set Role” attributes name = customer Id, Target role= ID & set additional role= churn, Ta
- Add cross validation, output model, testing and performance.
- In cross validation sub process put logistic regression & decision tree into training area.
- In testing set area, add “apply model” & “performance”

1.3. Now that the process is properly configured run the logistic regression and explain

1.3.1. Deliverable: Screenshot of Results of Logistic Regression

Attribute	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
Age Group.21-30	0.218	0.218	0.110	1.983	0.047
Age Group.41-50	-0.580	-0.580	0.185	-3.130	0.002
Age Group.51-60	-0.220	-0.220	0.263	-0.837	0.402
Age Group.16-20	0.858	0.858	0.232	3.702	0.000
Tariff Plan.1	0.277	0.277	0.291	0.951	0.342
Tariff Plan.3	-0.809	-0.809	0.308	-2.627	0.009
Tariff Plan.4	-0.537	-0.537	0.575	-0.935	0.350
International Plan.No	-0.806	-0.806	0.103	-7.851	0.000
Complaints.true	2.252	2.252	0.123	18.295	0
Yearly Call Duration (in Mins)	0.000	0.442	0.000	2.904	0.004
Call Count	-0.009	-0.527	0.003	-3.430	0.001
Data Usage	-0.001	-0.077	0.000	-1.478	0.139
Subscription Length	-0.073	-0.484	0.024	-3.021	0.003
Net Customer Spend	-0.000	-0.035	0.001	-0.112	0.911
Intercept	2.056	-0.883	0.295	6.958	0.000

1.3.2. Deliverable: Interpretation of the results of the logistic regression

(1) Most decisive predictor:

Most decisive factor are customer complaint, the bigger positive coefficient (2.252), the more related the variable to the target label (churn). The figure shows the highest number coefficient to the customer who making complaint about their service the company. The higher the complaint (338), the more customer will churn to the other service product. The higher customer not making a complaint (1,815), the more they will not churn the service provided by the company.

(2) **Tariff Plan**:

Tariff Plan 3 have the higher negative coefficient and standard coefficient (-0.809) as compared to Tariff Plan 4 (-0.537). This show the more customer in Tariff plan 3 were not churn over churn (381:73) as compared to the Tariff plan 4 customer (212:48). The higher positive coefficient is from customer in Tariff Plan 1 (0.277) which show the highest churn ratio over not churn (796:344).

(3) Data Usage:

The level coefficient for Data usage were -0.001 which around 0. This shows the variable have weak correlation between true churn prediction which result the churn prediction. This show this attribute can be discard from the model so it will not disturb the decision making for analyst.

1.4. Conclusions,

1.4.1. Deliverable: Screenshots of Results of Logistic Regression and Decision Tree Performance

Logistic Regression

accuracy: 78.34% +/- 1.63% (micro average: 78.34%)

	true No	true Yes	class precision
pred. No	1813	452	80.04%
pred. Yes	147	354	70.66%
class recall	92.50%	43.92%	

Decision trees

accuracy: 78.67% +/- 1.25% (micro average: 78.67%)

	true No	true Yes	class precision
pred. No	1903	533	78.12%
pred. Yes	57	273	82.73%
class recall	97.09%	33.87%	

1.4.2. Deliverable: Interpretation and justification of model performances

Overall accuracy in using logistic regression model is 78.34%, this were supported by the calculation of sum of class recall and class precision. As compared to decision tree model the overall accuracy are 78.67% which quite better than using logistic regression model. There is higher percentage in class recall for logistic regression (43.92%) in true Yes label as compared to decision trees (33.87%) and it also higher in class precision for prediction No (80.04%) compared to decision tree (78.73%).

As a Data analysis, the logistic regression model should be applied in searching statistical approach for predicting the connection between predictors and a predicted variable when the dependent variable is binary. A logistic regression requires two or more independent variables, or predictors. The IVs, also known as predictors, can be either continuous or categorical. The higher the confidence the most it affects the prediction.

5. Generate Business Insights and Strategize:

(1) Offer compelling explanations for the observed findings and means to investigate them:

First, customer behavior in Telecom company is quite common with all other telco company around the globe. The customer tends to use more data and spend less for the service. In the figure 3.2.1 customer in tariff plan 1 and 2 used more data and spend less in the amount spend for subscription (81 & 83 data) compared to the other plan. This will affect with the higher amount of complaint in plan 1 and 2 (189 & 202).

The cornerstone for forecasting churn is to look at both operational and experiential information along the customer experience (subscription length, average data usage). A client who has refused recent service plans and filed a complaint after their most recent service plan,

may be more likely to churn. Management could start by putting up feedback requests to enable analysts assess and diagnose key determinants of customer satisfaction to better understand the customer's experiences. After that, they may go on to evaluating transactional experiences, such as purchase or post-support follow-up, so Telecom can better understand where their critics are and devise a strategy for following up with them. Finally, AnalyticUs can map out the entire customer journey and assess the important experiences along the way, from the moment a need arises (to buy something, seek support, or submit a claim) until it is met for example low internet speed and limited area usage.

Other that, using yearly call duration and net customer spend in using RapidMiner Logistic regression as an attribute as the variable will have impact in the coefficient correlation. The software will indicate zero correlation between the attribute and target variable.

In my opinion, AnalyticUs should seek more data from Telecom company which are more effective and efficient indicates the reason why customer churn. We can take data such customer annual income, the attractiveness of our tariff plan service.

Telco Company should add one data driven which is customer services experience, customer happiness has always been influenced by customer service which can be explained when the most significant contribute factor was complaints attributes. When majority of those who were complaints will likely to churn, so by finding resolution of the customer problems, will help minimizes your service issues to the minimum.

(2) Describe factors that contribute to churn and propose actions on how to address them:

Overall perspective from all database tool used by the company, the most factors that contribute to churn are complain attributes. This can be review with the result from decision tree in rapid miner where most 80% customers who are not complain will not churn from the service provided by the company while 70% will churn if that customer making complain to the company. So telecom company management should take an action for example, always pay attention to your consumers. Customers complain for a reason, therefore it's critical to understand why they're complaining. Apologize for the error Customers just want an apology and recognition of their complaint, yet many organisations are unwilling to recognise when they have made a mistake. Lastly, Customers have a legal right to complain, therefore corporate management must identify and address the core problem. To prevent transferring your customer across a series of individuals and supervisors, don't give your customer care staff the ability to address most client concerns. If the problem has occurred or is likely to occur again, make the appropriate adjustments to avoid receiving another complain.

Other than that net customer spend were the second most contribute factor if the customer did not make complaint. This analysis can be analyst by rapid miner decision tree after the false customer complaints result node. Customers who spend less than $ 385 net customer spend are less likely to churn (No=1801). However, the story is little bit different if it compared to the result from power Bi which is the less average net customer spend the more customer will churn this is because the different method uses. In decision tree edge classified net customer spend below 385 but in power Bi it was arrange according to the tariff plan.

To attract more customers and less churn, the company should provide more data usage in densely populated location for example school, universities etc. So, this type of customer will make less

complaint about the service because the demand of cheaper plan was higher in such places. So the company will attract more customer who can spend less with bigger data usage and less complaint. This is because most people who live in high density area have income less than average, so this type of customer will stay longer with the company and reduce the number of churns.