

POWER BI DETAILED DOCUMENT

Retail Customer Retention Analytics – WALMART

Project overview – Walmart is one of the largest retail chain businesses, so the project was based on developing an interactive Customer Retention Analytics Dashboard in Power BI.

Customer Retention Analytics helps in:

- Identifying Loyal vs at-risk customers
- Identifying the impact of loyalty tiers, promotions applied, seasonal sales, discounts.

TASKS

Task 1: Data Modelling & Cleaning

- Loaded all 5 datasets into power query.
- Checked for inconsistencies, data type, missing values, duplicate values, etc. in power query.
- Created calculated column in power query named “*Membership_Duration*” and Extracted transaction “year” and “month”.
- Calculated the “**“Membership_Duration”** column using the formula:

`DateTime.Date(DateTime.LocalNow()) - [Membership_Since]`



Fig. no - 1

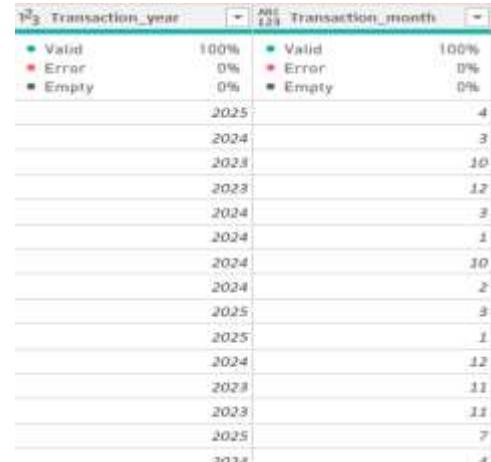


Fig. no -2

- Where `DateTime.LocalNow()` will return the current date and time.

- As we need only today's date, so we extracted only date from `DateTime.LocalNow()` using `DateTime.Date()`.
- And then subtracted `Membership_Since` from current extracted date, so we got the value of `Membership_Duration` in days.
- For **Extracting Month and year** used the `Date.Month()` and `Date.Year()`.
- Made One-to-Many Relations between:
 - o Customer Demographics → Transactions
 - o Customer Demographics → Loyalty Program
 - o Customer Demographics → Churn Labelled Customer
- And Many-to-One between:
 - o Transactions → Store Locations

Task 2: Churn & Retention Metrics

- Task 2.1:

- o Calculated Churn Rate → (Churned Customers / Total Customers) *100

Used the formula:

DIVIDE (Churned Customers, Total Customers) * 100

Where churned Customers is,

VAR

Churned Customers =CALCULATE (DISTINCTCOUNT (Customer Transactions [Customer_ID]), Churn_Labelled_Customers [Churn_Flag] =1)

- Here the VAR defines the variable Churned Customers, Calculate Function takes the expression and filter, like in the above case, the expression is:

DISTINCTCOUNT (Customer Transaction [Customer_ID])

The above expression takes the total unique Customer_ID from transaction table (total- 294 customers).

- And the filter is:

Churn_Labelled_Customers [Churn_Flag] =1)

which filters the total customers (294) and checks for Churn Flag which is in Churn_Labelled_Customers table and this calculates the count of Churned Customers from the total customers.

- Task 2.2: Visualize churn rate by:

- o **Region**
- o **Income Group**
- o **Channel (Store/Online)**

- Loyalty Tier

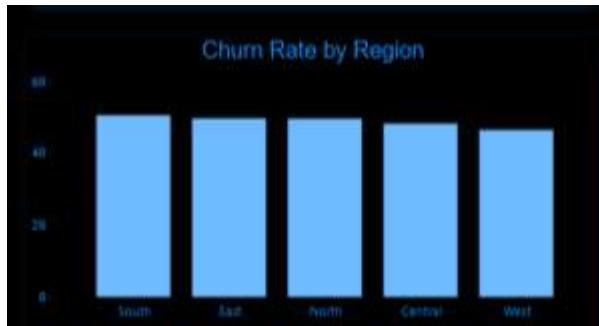


Fig. no – 3



Fig. no- 4



Fig. no – 5



Fig. no – 6

- For Churn Rate by Region, I used the column chart where I placed Region in Y-axis and churn rate in X-axis, as we can see in figure no. 3 that the highest churn rate is in South region.
- And for Churn rate by Income Group I used Donut chart, where I placed Income level in Legend and Churn rate in Values. (figure no. 4)
- For Churn rate by Channel, I used Column chart, where on the X-axis I placed Preferred channel and on Y-axis I placed Churn rate. (figure no. 5)
- And finally for Churn Rate by Loyalty Tier, I used Pie chart, where I placed Loyalty Tier in Legend and Churn rate in values. (figure no. 6)

- **Task 2.3:**

Funnel Chart: Total Customers → Repeat Customers → Churned

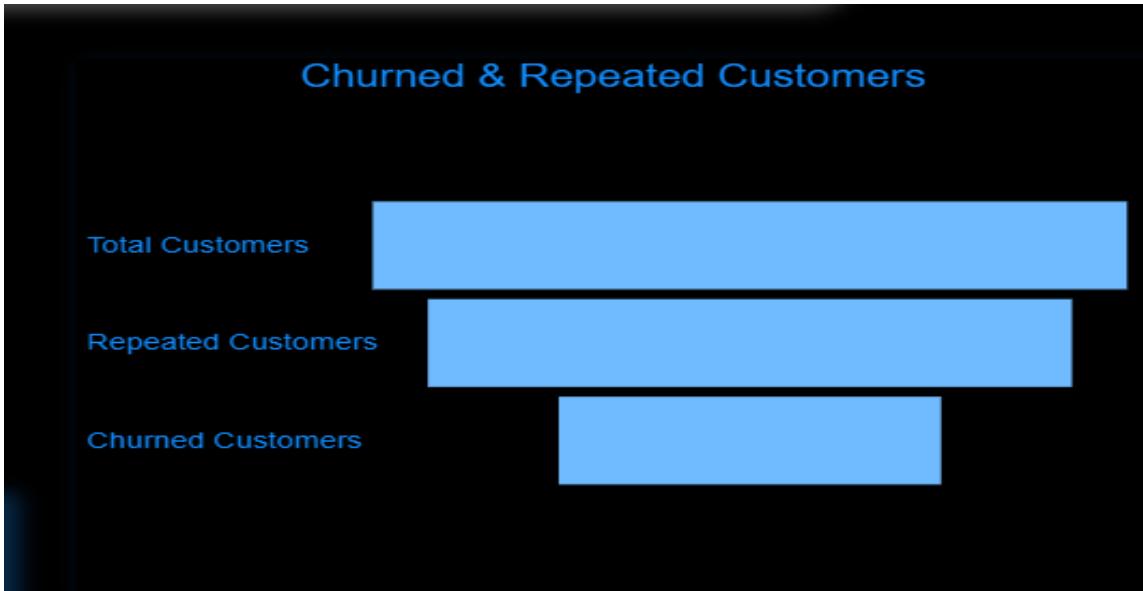


Fig. no – 7

- For this task I used the funnel chart, where I placed Total Customers, Repeated Customers and Churned Customers.
- I created a new column in transaction count table named as Repeated_Customers, considered that customers having more than 1 purchase or at least 2 purchases as repeated customers.
- As in transaction count table, I already had Unique Customer_ID with their specific count of transactions they made, so using the Transaction_Count column, I added a conditional column named as repeated customers and set the criteria on Transaction_Count column that if the value is greater than or equal to 2, then 1, else 0.
- I chose it to be binary and numeric instead of binary and string (text), as it would be easier to sum the number of repeated customers.
- This way when I added Repeated customers column in funnel chart, there I aggregated it to sum and it returned me the value of 251 customers out of 294 customers are repeated customers.

Task 3: Repeat Purchase Analysis

- **Task 3.1: Segment customers:**

Low-Tier: 0–3 purchases

Mid-Tier: 4–8 purchases

High-Tier: 9+ purchases

- For this Specific task, I used Power Query editor, used the duplicate copy of transaction table and named it as Transaction count (a new table having all the columns that were present in transaction table).
 - Then using the copy, I grouped by Customer_ID and named the new column as Transaction_Count, and the operation was Count Rows.
 - It returned me the Unique Customers with the number of Transactions they made, for example: CUST0294 → 6 transactions.
 - Using this Transaction_Count column it became easier to segment the customers into Low, Mid, High tier.
 - Using the Conditional Column, I set the criteria given (0-3 → Low Tier, 4-8 → Medium Tier, 9+ → High Tier) and named the column as Customer_segment.
-
- **Task 3.2:**

Compare avg. purchase frequency by Region, Age Group, Loyalty Tier

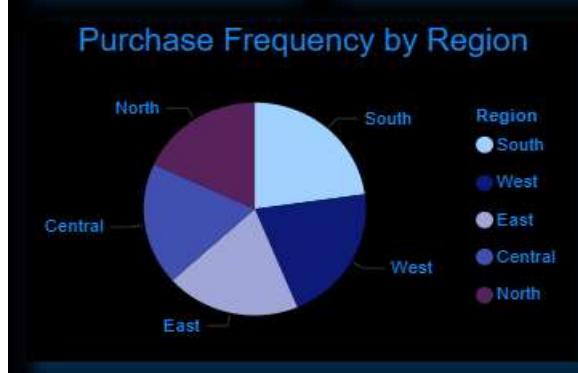


Fig. no – 8

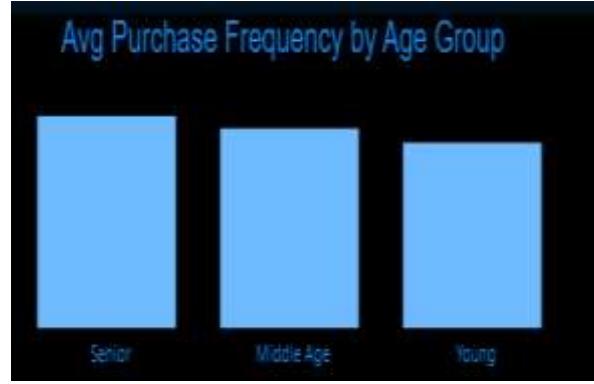


Fig. no - 9

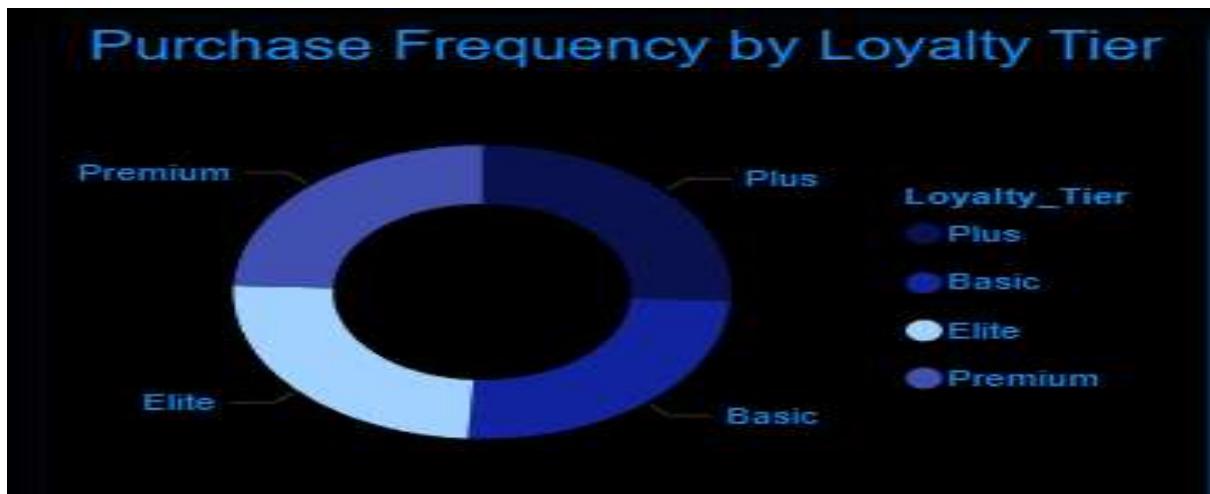


Fig. no – 10

- For Calculating Avg. Purchase Frequency, I used the formula Total Transactions / Unique Customers, in power BI I used:

```
Avg_Purchase_frequency =  

DIVIDE (count (Customer_Transactions [Transaction_ID]),  

DISTINCTCOUNT (Customer_Transactions [Customer_ID]))
```

- The above formula where I used divide function, it calculates first the total count of Transaction_ID and then it calculates the total count of Unique customers and then it divides them.
- The total number of transactions were 1000, and unique customers were 294, so the formula will simply be:

$$\text{Avg_Purchase_frequency} = \frac{1000}{294}$$

- And then using pie chart, where I placed Region in Legend and Avg. Purchase frequency in Values for Avg. Purchase Freq by Region (figure no. 8)
- Similarly, for Avg. Purchase Freq by Age Group, I first segmented the age into “Young”, “Middle-age”, “Senior” using the add conditional column in Power Query and added the new Column named Age Group in Demographics table.
- And then using the column chart I performed this task by placing the Age Group column in X-axis, and Avg. Purchase frequency in Y-axis (figure no. – 9).
- For Avg. Purchase frequency by Loyalty Tier, I used the pie chart, where I placed Loyalty Tier in legend and in values, I place Avg. Purchase Frequency (figure no.-10).

- **Task 3.2:**

Identify most purchased product categories by loyal customers

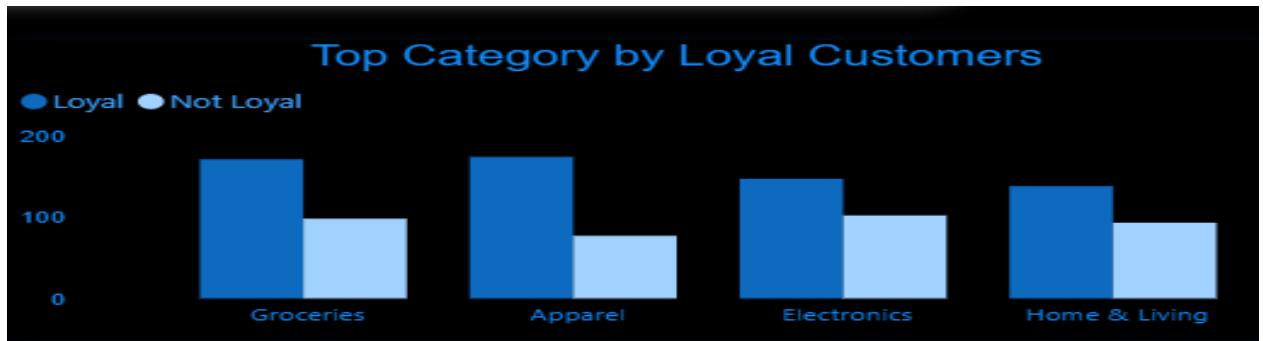


Fig. no – 11

- In this specific task, I used Clustered Column chart to visualize the most purchased product category by loyal customers.
- First, I added a column named loyal customers in the table transaction count where I already had Customer ID, Transaction_Count, Repeated Customers, Customer segment, etc. with the help of Avg. Purchase frequency and transaction I added Loyal_Customers column.
- As the Avg. Purchase frequency is 3.40, so the customers having transactions more than or equal to 4 are the loyal customers and others are not loyal, it's like I added a loyal customer flag.
- Then placed Product categories in X-axis, Transaction_ID in Y-axis with count aggregation, and in legend I placed the newly added column named Loyal_Customers (figure no. 11).
- By this visualization, I got that, for loyal and not loyal customers the total count for each product category is:

Category	Loyal Customers	Not Loyal Customers
Groceries	171	98
Apparel	174	77
Electronics	147	102
Home & living	138	93

- And we can clearly see that Apparel has the highest number of transactions for loyal customers and for not loyal customers its Electronics.

Task 4: Promotion & Loyalty Impact

- **Task 4.1: % of transactions with promotion applied**
- First, I needed to calculate the percentage of transactions with promotion applied, so for calculating that, I have the formula:

$$\% \text{ of transactions with promotion applied} = \frac{\text{Total number of transaction with promotion applied}}{\text{Total Transaction}}$$

- And in Power BI I used:

```
DIVIDE(CALCULATE(COUNT(Customer_Transactions [Transaction_ID]),  
Customer_Transactions [Promotion_Applied] = "Yes"), COUNT  
(Customer_Transactions [Transaction_ID])) *100
```

- Here, I first get the total number of transactions with promotion applied, with the help of calculate function, where the expression is, `Customer_Transactions [Transaction_ID]` and the filter is `Customer_Transactions [Promotion_Applied] = "Yes"`
- And I used the DIVIDE function with it so, I used count of `Transaction_ID` which will return me the total transactions.
- And it gave me the value of 49%, which means 490 transactions were done with promotions applied.
- Here the formula was simple when the numbers were out, it became:

$$\% \text{ of transactions with promotion applied} = \frac{490}{1000}$$

- **Task 4.2: Compare avg. purchase amount with vs without promotions**



Fig. no - 12

- In the above task, I used Clustered Column chart, where I also added Loyal_Customers column, and I placed Promotion Applied in X-axis and Amount in Y-axis with Average aggregation, and in legend I placed Loyal_Customers so that we can differ them (fig. no – 12)
- So, the numbers were:

	Average Amount (Promotion applied = Yes)	Average Amount (Promotion applied= No)
Loyal Customers	502.95	512.77
Not Loyal Customers	536.66	522.79

- **Task 4.3: Churn rate across loyalty tiers**



Fig. no – 13

- So, this task already had been explained in the Task 2, while there I used Pie chart and here I used funnel chart to show that how the churn rate varies across different loyalty tiers which are Elite, Plus, Premium, Basic.
- So basically, here in category I placed Loyalty Tier and in Values I placed Churn rate, and gave it me a awesome funnel chart (figure no. – 13).

- **Task 4.4: Points Earned vs Redeemed by Tier**

- In this task, I had to visualize Total points earned vs Total points redeemed by Loyalty tier.

- So, at first, I calculated the total points earned by using the DAX measure and there using the SUM function and applied it to the column where I had points earned. The formula that I used in DAX is:

Points_earned_sum = SUM (Loyalty_Program [Points_Earned])

- Here, in Loyalty_Program table I had the column Points_Earned, so I did sum of that column and it gave me the value of 758k.
- Similarly, for points redeemed, I used the DAX measure and there also I used the SUM function, the formula that I used in the DAX is:

Points_redeemed_sum = SUM (Loyalty_Program [Points_Redeemed])

- Here, in the same table where we had Points_Earned column we also do have Points_Redeemed column, so again I followed the same process and got the Total Points Redeemed, and it was 624k.



Fig. no – 14

- After, calculating the Total Points Earned and Redeemed, I used the clustered column chart, and then placed the Loyalty Tier in X-axis and in Y-axis placed Total points earned as well as Total points redeemed.
- And the values were:

Loyalty Tier	Points Earned	Points Redeemed
Premium	201468	170504
Plus	200269	177075
Basic	180392	121685
Elite	175954	154708

- **Task 4.5: Recommendations to improve redemption & retention**
- After, analysing and having some search on WALMART, I can say that the value of points earned is more than value of points redeemed.
- As to improve this i.e. **Total Points Redeemed \geq Total Points Earned**, to make this situation, WALMART can change their rule of minimum points wallet.
- Let's say that WALMART has a rule that you need minimum of 200 points to redeem points, so this affects the use of redemption as customers wait and then after when it reaches the goal, they don't use it for small transactions, so if this minimum points wallet rule is changed or negotiated to 100 or 50, then on every transaction customers will be getting points and they can spend it as early as they reach 50 points.

Task 5: Store & Channel Performance vs Retention

- **Task 5.1: Merge store data with transactions**
 - The merging of Store table and transaction table was done using the Store ID (the similar column both the table has).
 - It helped me to move Region, Store type, Opening Year in transaction table.
 - And also, it helped me with some visuals, like Churn Rate by Store type which is in next task (task 5.2).
-
- **Task 5.2: Visualize:**
 - o Avg. transaction amount by Store Type
 - o Churn rate by store type
 - o Correlation between store opening year & retention



Fig. no – 15



Fig. no – 16

- For the task, Avg. transaction amount by store type, I used column chart and placed Store_Type in X-axis and Amount in Y-axis with Average aggregation (figure no. -15)
- For the task, churn rate by Store type, I used Donut Chart, where I placed Store_Type in Legend and Churn rate in values, I formatted it to look good.
- For the next task which is Correlation between store opening year & retention, I first calculated the retention percentage.

- For Calculating retention percentage, I used the DAX measure with the formula:

```
Retention % = VAR retainedcustomers = CALCULATE (DISTINCTCOUNT  
(Customer_Transactions [Customer_ID]), Churn_Labelled_Customers [Churn_Flag]  
= 0)
```

```
VAR Totalcustomers = CALCULATE (DISTINCTCOUNT (Customer_Transactions  
[Customer_ID]))
```

```
RETURN DIVIDE (retainedcustomers, Totalcustomers) *100
```

- Here, VAR defines the variable, and I defined two variables retainedcustomers and Totalcustomers.
- For retainedcustomers the formula is:

```
CALCULATE (DISTINCTCOUNT (Customer_Transactions [Customer_ID]),  
Churn_Labelled_Customers [Churn_Flag] = 0)
```

- The calculate function takes, expression and filter, so the expression is *(DISTINCTCOUNT (Customer_Transactions [Customer_ID])* and the filter is *Churn_Labelled_Customers [Churn_Flag] = 0*, the filter says that in the table churn labelled customers and in churn flag column only count those who are having value of 0. Value of 0 means the customers who are active or not churned.
- And the expression says that, match it with the transaction table and with Customer_ID column.
- Later used DIVIDE function to get the specific retention value and then multiplying it with 100 so it would give us percentage.
- Here, I have used two different tables because, they are connected and have relations, and the main reason is that the unique customers in the churn labelled customers has the count of 300 and here in transaction table it is 294.
- It gives me the retention percentage, also I we go term wise then the retention formula is:

$$\text{Retention \%} = \frac{\text{Retained Customers or Not churned customers}}{\text{Total Customers}}$$

- Then, I used Scatter plot to visualize the correlation between opening year and retention percentage.

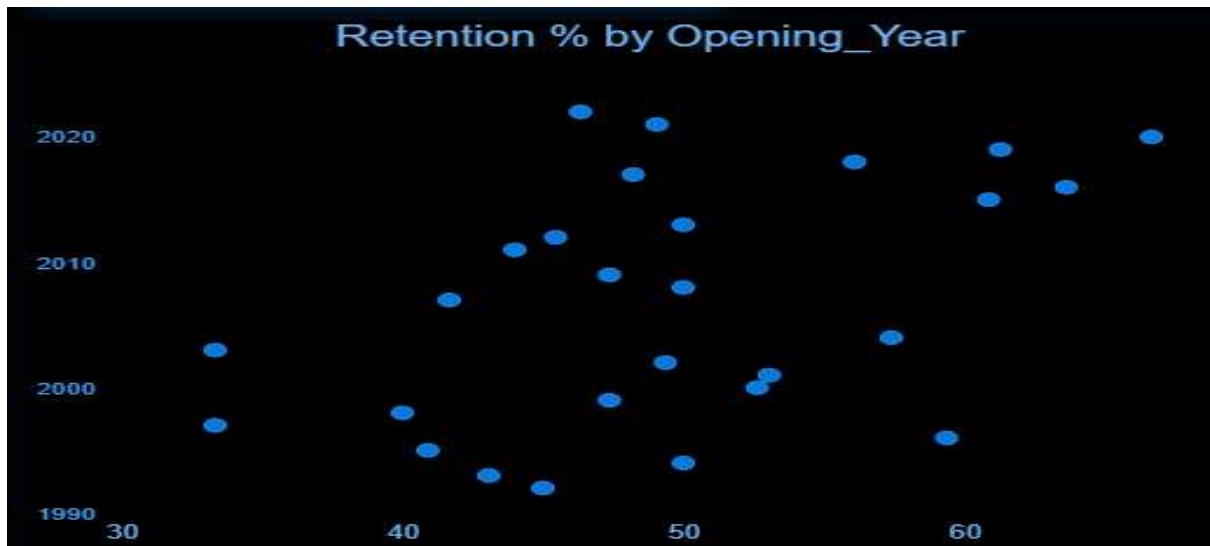


Fig. no – 17

- And after analyzing it, I can say that the retention % is having a positive correlation as it going increasing as the increase in Opening year.

Task 6: Customer Lifetime Value (CLV)

- **Task 6.1: CLV = Total Amount Spent / Membership Duration (Years)**
 - For calculating the customer lifetime value (CLV), I again used the power query editor, this time I created a copy of customers transactions table and then using group by I calculated the Total amount spent by each customer but setting the operation to sum, and applied on column Customer_ID and aggregated on Amount column.
 - This returned me two columns, Customer_ID and Total_Amount (renamed).
 - Now, for membership duration in years, I had membership duration in days in demographics table.
 - So, to bring it into CLV table I merged queries and just expanded Membership duration in days into CLV table.
 - And then calculated the Membership Duration in years column with the help of custom column and formula was Membership duration in days / 365.
 - Then by using custom column added new column CLV with the formula Total Amount / Membership Duration in years.
 - So, now I had CLV column in my CLV table, having the CLV values of each customer.
-
- **Task 6.2: Segment customers into:**
 - o **Low, High CLV**
 - For Segmentation of CLV, I used the CLV table and in the same table I added a new conditional column where I set the Criteria that if the CLV of a customer is below average then its low and if above average then its high.
 - So, the Average of CLV is 745.870,
i.e. $CLV > 745.870 \rightarrow \text{High}$
 $CLV < 745.870 \rightarrow \text{Low}$
 - So, now I had CLV_segment in CLV table.
-
- **Task 6.3: Visualize:**
 - o **CLV vs Days Since Last Purchase**
 - o **CLV by Loyalty Tier & Region**

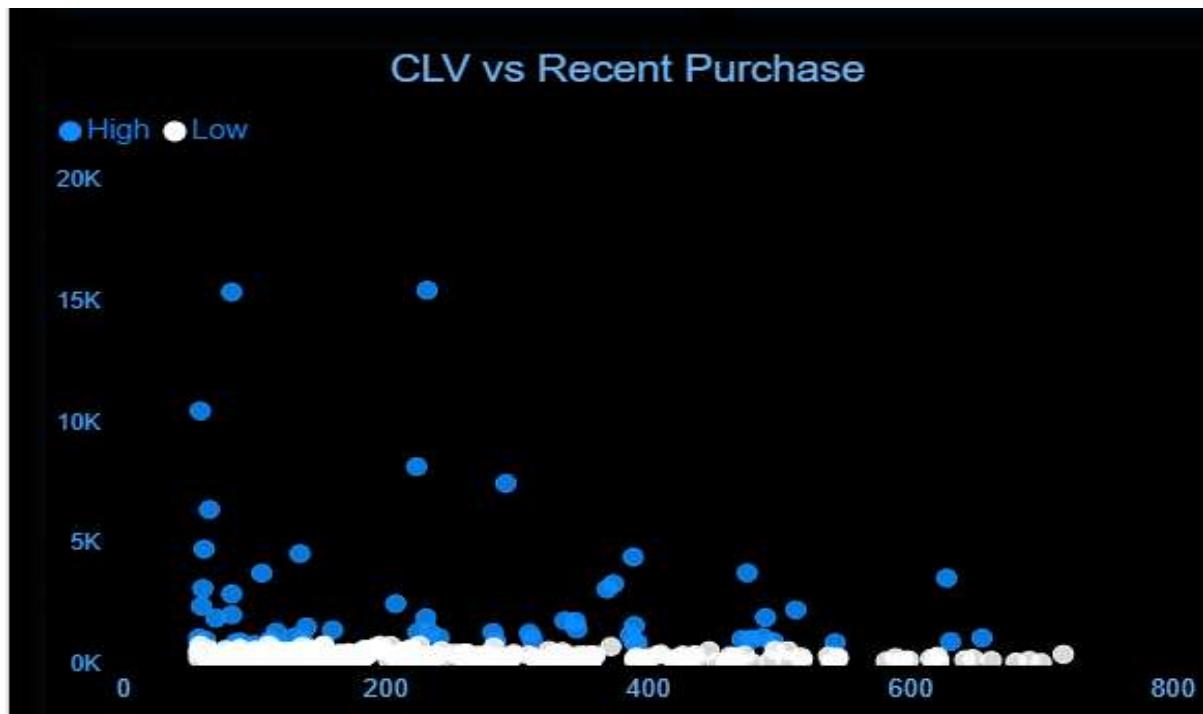


Fig. no – 18

- For **CLV vs Days Since Last Purchase** I first needed to calculate the Days since last purchase, so for this I used Power Query editor.
- I had transaction date in transaction table, so I made a copy of transaction table again and this time I used group by on Customer_ID and now set the new name as Last_transaction (same as table name after copying I renamed it as last transaction) and operation of MAX, and on column of transaction date.
- This gave me the most recent date of each customer and then I changed the datatype to date.
- Then to bring it in CLV table, as I had CLV in CLV table, I also could have done it without doing it but for ease use and for less chances of error I brought last transaction column in CLV table with the help of merge queries.
- Expanded only last transaction column in CLV table, and that is what I needed.
- And then to calculate Days Since Last Purchase I have to subtract it from current date, so using custom column (Add column tab) I added the new column named Days since last transaction and used the formula:

`DateTime.Date(DateTime.LocalNow()) - [Last_transaction]`

- `DateTime.LocalNow()` gives the current date and time and I extracted only date using `DateTime.Date()` and it gave me a value in days for each customer.

- And then using the Scatter plot I visualized CLV vs Days since last transaction and gave the title of CLV vs Recent Purchase.
- And also added CLV segment ad legend, so now we can clearly see that customers having High CLV are having the most recent purchases while customers with Low CLV are having no recent or way too old purchase.
- For **CLV by Loyalty Tier & Region**, I used column chart and placed Region and Loyalty Tier in X-axis and CLV in Y-axis.



Fig. no – 19

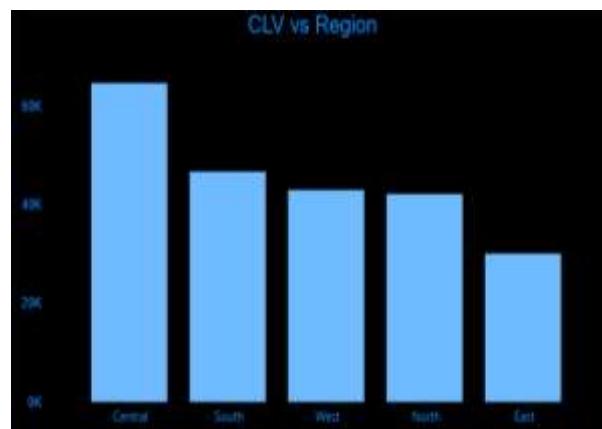


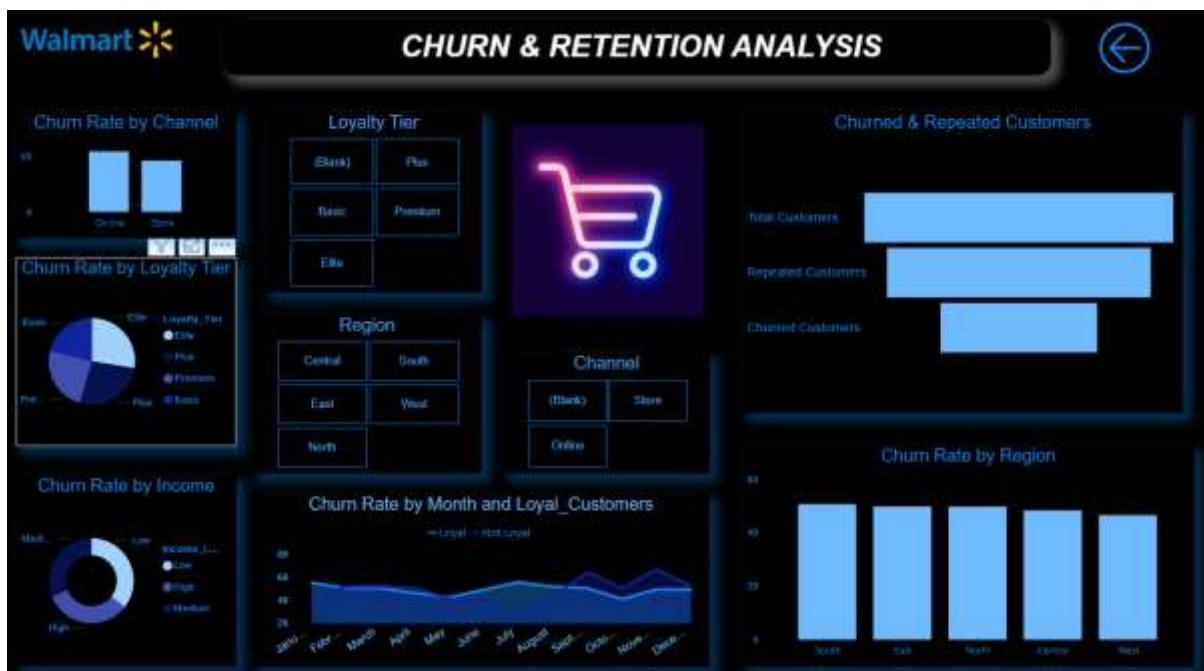
Fig. no – 20

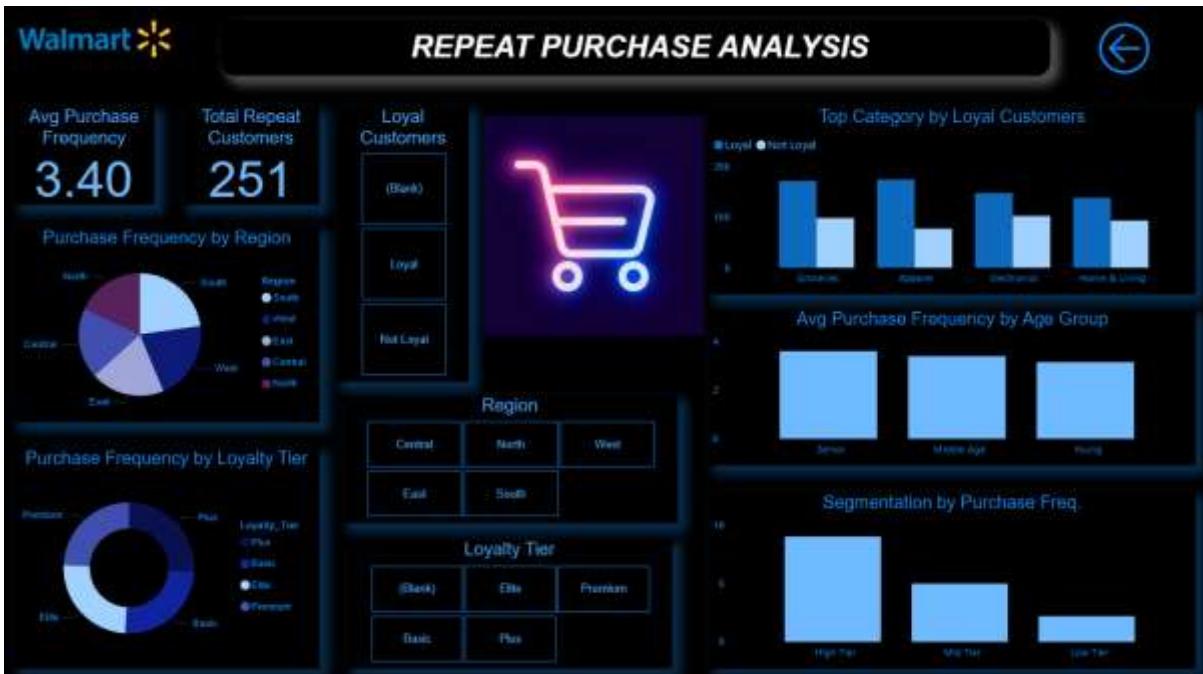
- As we can see that, Basic tier customers are having the highest CLV when its CLV vs Loyalty Tier.
- And in Central region the CLV is highest, I can't say that most of the Basic tier customers live in Central region. It is possible that they live but as we haven't seen the correlation between them so I cannot say that.

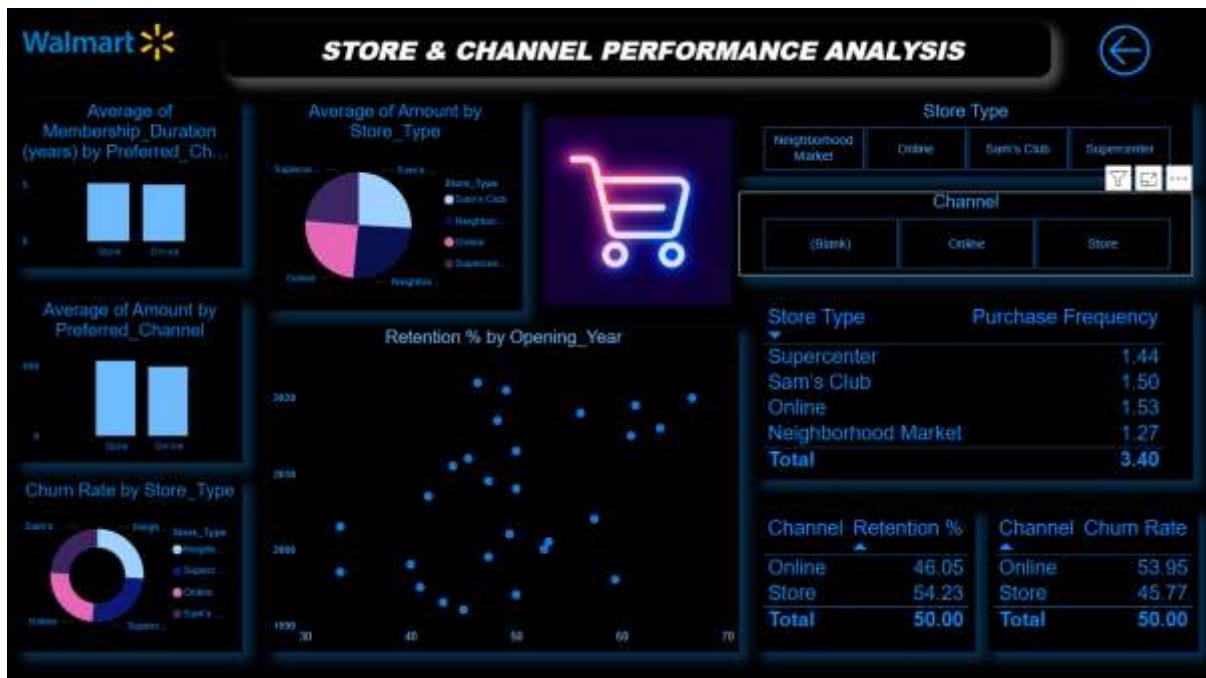
Task 7: Final Dashboard & Executive Summary

- **Task 7.1: Multi-page Power BI Report:**
 - o **Page 1: KPIs (Churn, CLV, Repeat Rate)**
 - o **Page 2: Loyalty & Promotion Impact**
 - o **Page 3: Store/Channel Insights**
 - o **Page 4: Segmentation (Churned, Repeat, High-Value)**
- **Task 7.2: Slicers: Region, Channel, Income, Loyalty Tier**
- **Task 7.3: Top 3 recommendations for Walmart:**
 - o **Which customers to prioritize for retention?**
 - o **Which channels are underperforming?**
 - o **How to strengthen loyalty program engagement?**
- The customers who are having at least 2 transactions or those who we call repeated customers should be prioritize for retention. As people with transaction ≥ 4 are already considered are loyal customers and they are coming again and again, which indicates that they are actually retained customers. So that's why with at least 2 transactions.
- Channels that are underperforming are “Online” as it has the highest churn rate and lowest retention rate among the channels as there are only two channels, Store and Online.
- To strengthen the loyalty program WALMART should focus on their loyal customers, let's say after analyzing the customers purchase history, WALMART can give them a discount, on their purchase.
- Let's say A specific customer is going to the Gym regularly and we know that the person going to the gym needs protein, milk, peanut butter, protein bars, etc. and after analyzing this they can give that person a 10-15% off on next purchase of Protein or peanut butter. This will make that person to come again to WALMART instead of going somewhere else where that person is not getting discounts.
- Also, WALMART can give their Loyal customers a early bird discount which they will not get in other stores or at other competitors, so it will strengthen the Loyalty program.

REPORT IMAGES







NOTE

- All the Pages has slicers linked with each page slicer.
- On every page except the first page there is the button on the top right corner to go to previous page.
- All the unique customers that are used in formulas as count or distinctcount are from the table of transaction, as I considered only those customers who have transacted once with Walmart.

[**WALMART Power BI report \(saved in .pbix\)**](#)

[**Explanation Video**](#)

Student name → Nilesh Bahirgaonkar

-----*END OF REPORT*-----