STRATEGY STORM 2025 IIT GUWAHATI

ANALYTICS REPORT OF BLINKIT CUSTOMER CHURN

TEAM SOBER STRATEGISTS

Introduction

This report has been made by Team Sober Strategists to understand the root cause behind the problem of high customer churn faced by Blinkit and frame retention strategy for the same.

Namely three datasets were used:

- Strategy Storm Round-2 Dataset
- Sober Strategists Sheet : Modified dataset made by algorithmic manipulation of base dataset
- Product Commodity Sheet:Self made by scraping data from Blinkit's Website.

Techstack Used

- Python - Beautiful Soup, Matplotlib, Seaborn, Plotly, Numpy & Pandas
- Power BI
- Excel & Google Sheets

Disclaimer

After required manipulation made by algorithmic usage of columns provided in the base dataset the final dataset used was named as Sober Strategists Sheet and all the analytics has been performed using the same dataset.

Dataset Information

Columns: 19

Shape: **16,093 rows × 19 columns**

Data Types: Mixed (int, float, object)

Names of all the columns in the dataset:

1. **UserID :** Unique identifier for each customer.
2. **CustomerAge :** Age of the customer.
3. **Sex :** Gender of the customer.
4. **Tenure :** Time the consumer has been with company
5. **ServiceUsageRate :** Frequency of service usage.
6. **SupportCalls :** Number of calls to customer support.
7. **BillingDelay :** Refers mainly to payment failures
8. **PlanType :** Type of plan
9. **AgreementDuration :** Duration of Plan
10. **TotalExpenditure :** Amount spent so far on purchase

11. **RecentActivity :** Most recent interaction or activity level of the customer on the platform (measured in days since last interaction).
12. **Churn :** Whether churn or not
13. **Call-Tenure Ratio :** Ratio of Support Calls and Tenure
14. **Activity Decay Rate :** Inference of how activity declines over time measured as **(Recent Activity - Average Activity)/Tenure**
15. **Interaction Index :** Inference of engagement of user with the platform measured as **(ServiceUsageRate +Recent Activity - Support Calls)/3**
16. **CLV :** Refers to Customer Lifetime Value and measured as **(Total Expenditure/Tenure)*Average Tenure**
17. **Tenure to Customer age Ratio :** Ratio of Tenure and age of consumer
18. **Churn Probability Score :** Churn Prediction Metric measured **w1*(Billing Delay) + w2*(Support Calls) + w3*(Tenure) [**Based on correlation matrix made from base sheet**]**
19. **Churn Risk Index :** Another Churn Prediction Metric measured as **(churn probability score + (call-tenure ratio×10))/2**

Information of the dataset

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16093 entries, 0 to 16092
Data columns (total 19 columns):
 #   Column                                                                Non-Null Count  Dtype  
---  ------                                                                --------------  -----  
 0   UserID                                                                16093 non-null  int64  
 1   CustomerAge                                                           16093 non-null  int64  
 2   Sex                                                                   16093 non-null  object 
 3   Tenure                                                                16093 non-null  int64  
 4   ServiceUsageRate                                                      16093 non-null  int64  
 5   SupportCalls                                                          16093 non-null  int64  
 6   BillingDelay                                                          16093 non-null  int64  
 7   PlanType                                                              16093 non-null  object 
 8   AgreementDuration                                                     16093 non-null  object 
 9   TotalExpenditure                                                      16093 non-null  int64  
 10  RecentActivity                                                        16093 non-null  int64  
 11  Churn                                                                 16093 non-null  int64  
 12  Call-Tenure Ratio                                                     16092 non-null  float64
 13  Activity Decay Rate (Recent Activity - Average Activity)/Tenure       16093 non-null  float64
 14  Interaction Index (ServiceUsagzeRate +Recent Activity - Support Calls)/3  16093 non-null  float64
 15  CLV (Total Expenditure/Tenure)*Average Tenure                         16093 non-null  float64
 16  Tenure to Customer age Ratio                                          16093 non-null  float64
 17  Churn Probablity Score                                                16093 non-null  float64
 18  Churn Risk Index                                                      16093 non-null  float64
dtypes: float64(7), int64(9), object(3)
memory usage: 2.3+ MB
```

Analytics Insights

1. Correlation Matrix

1. **Strong Positive Correlations**:

    - **CustomerAge and Tenure (0.77)**: Older customers tend to have longer tenure with the service.
    - **Churn Probability Score and Churn Risk Index (0.91)**: These two metrics are highly related, suggesting they are designed to measure similar aspects of churn risk.
    - **Interaction Index and ServiceUsageRate (0.69):** Higher service usage correlates strongly with a better interaction index.
    - **BillingDelay**: Strongest correlation (+0.57), indicating that delayed payments significantly influence churn.

2. **Strong Negative Correlations:**

    - **Call-Tenure Ratio and Tenure (-0.45)**: Customers with longer tenure tend to make fewer calls per unit of tenure.
    - **Tenure to Customer Age Ratio and CustomerAge (-0.54)**: As customers age, their tenure as a proportion of age decreases.

3. **Churn Correlations:**

    - **Positive:**
        - Churn is positively correlated with SupportCalls (0.31): Customers making more support calls are more likely to churn.
        - Churn is moderately correlated with Churn Probability Score (0.25) and Churn Risk Index (0.36), as expected.
    - **Negative:**

- Churn is negatively correlated with Tenure (-0.20): Customers with longer tenure are less likely to churn.
- Churn shows a weak negative correlation with TotalExpenditure (-0.08).

4. **Feature Relationships:**

   - CLV (Customer Lifetime Value) is moderately correlated with Tenure (0.48), showing that longer tenure contributes to higher lifetime value.
   - Activity Decay Rate shows weak correlations with other variables, indicating limited direct influence.

5. **Independent Variables:**

   - UserID and most other variables show negligible correlations, confirming it as an identifier column with no analytical significance.

6. **Churn Predictors:**

   - Features like SupportCalls, Churn Risk Index, and Churn Probability Score have notable correlations with churn, making them strong candidates for predictive modeling.

IMAGE 1: CORRELATION MATRIX FROM THE BASE DATASHEET

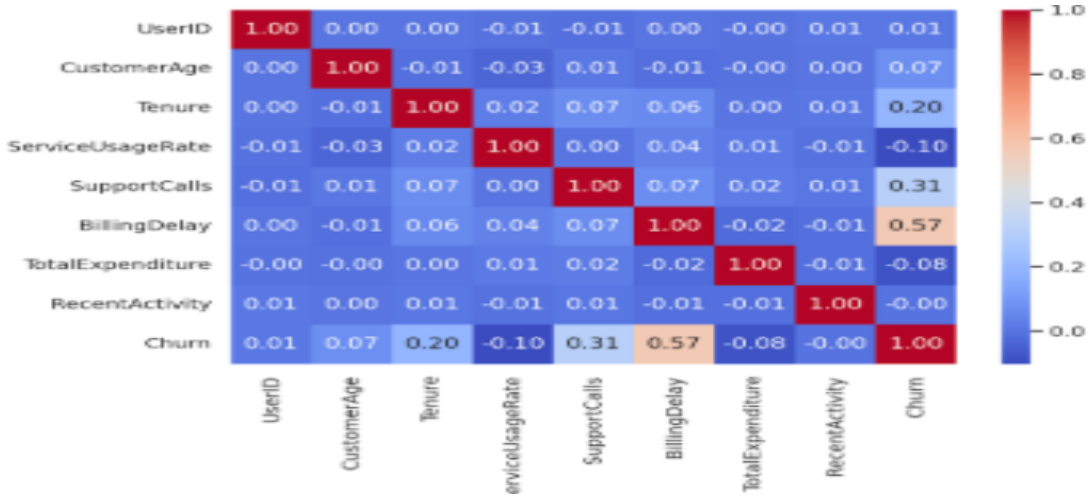
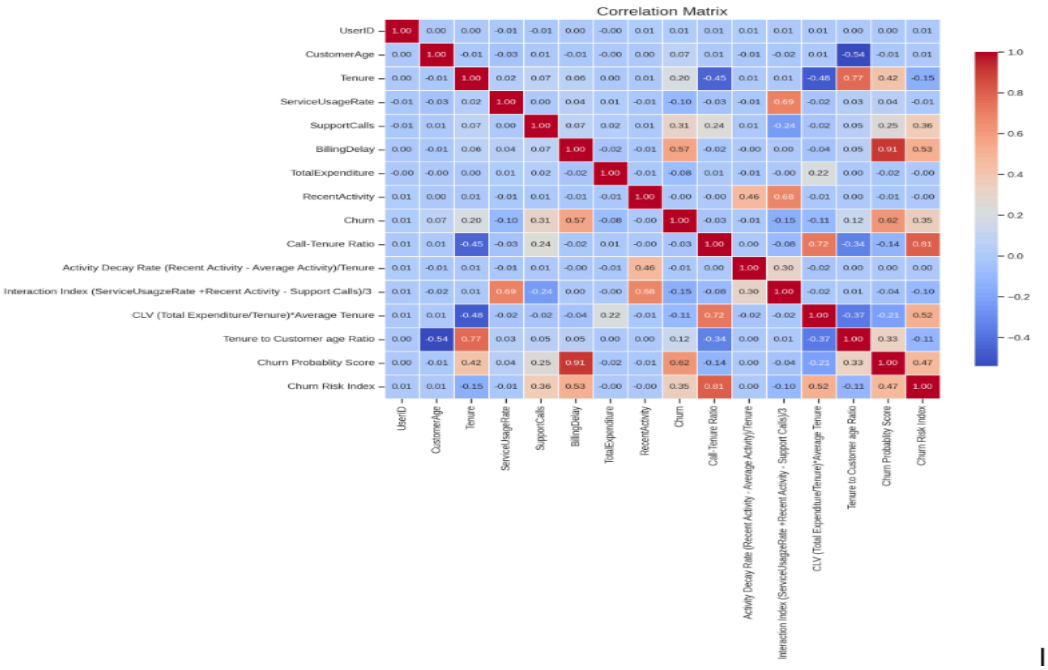


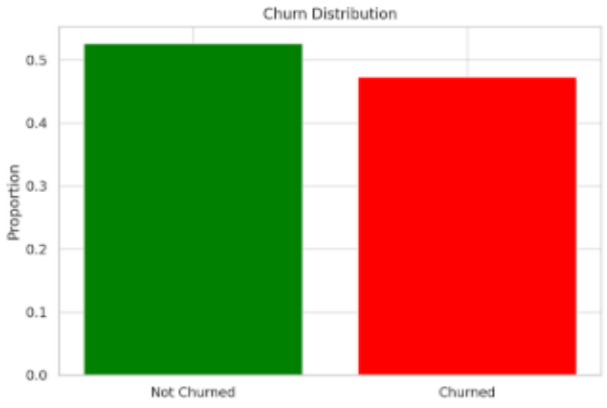
IMAGE 2 : CORRELATION MATRIX FORMED FROM MANIPULATED DATASET



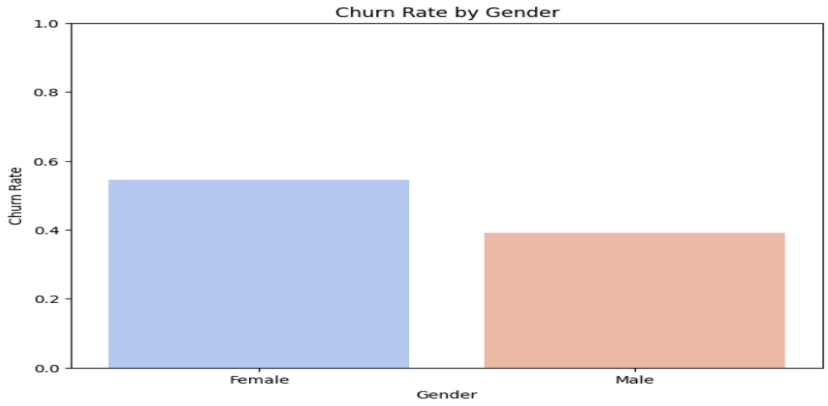
2.Charts & Graphs with their inferences

INSIGHT: Churn Distribution

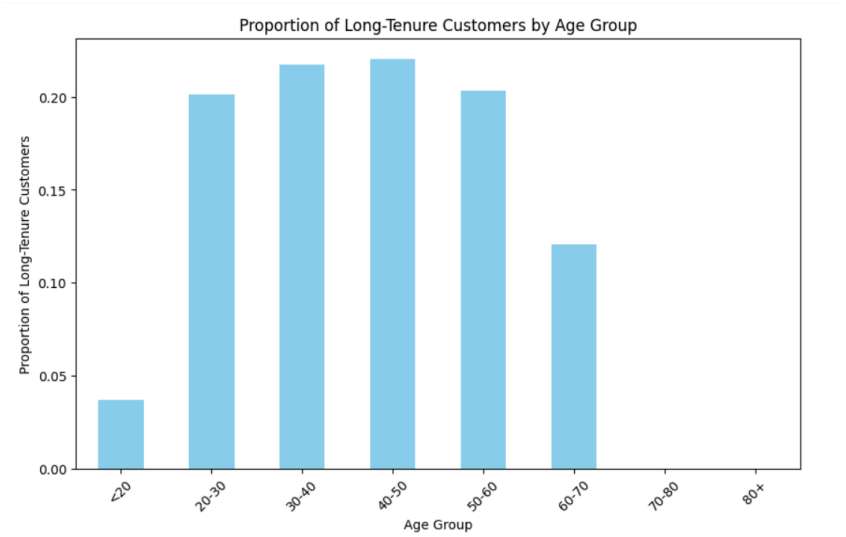
52.7% have not churned whereas 47.3% have churned indicating a significant churn rate



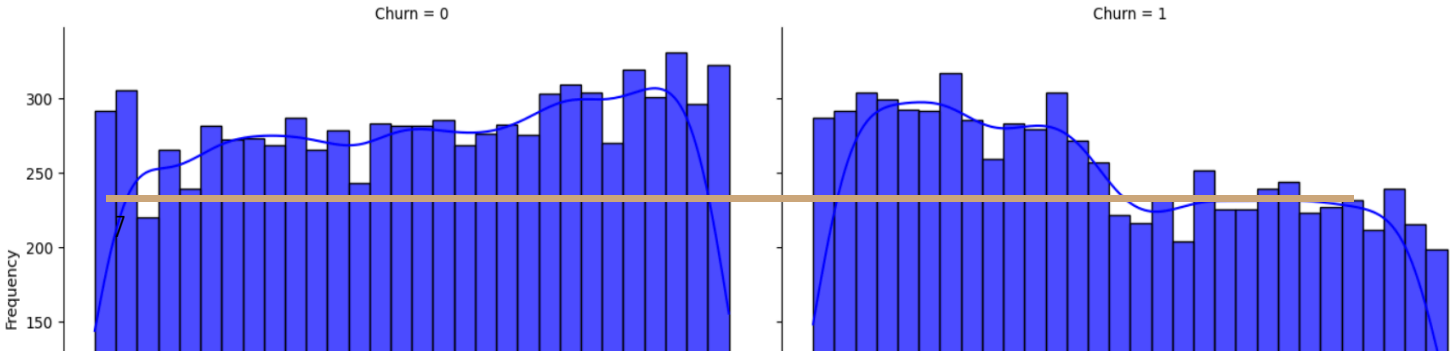
INSIGHT: Based on this column chart it can be clearly seen that the churn rate of female customers are comparatively much more than male customers.



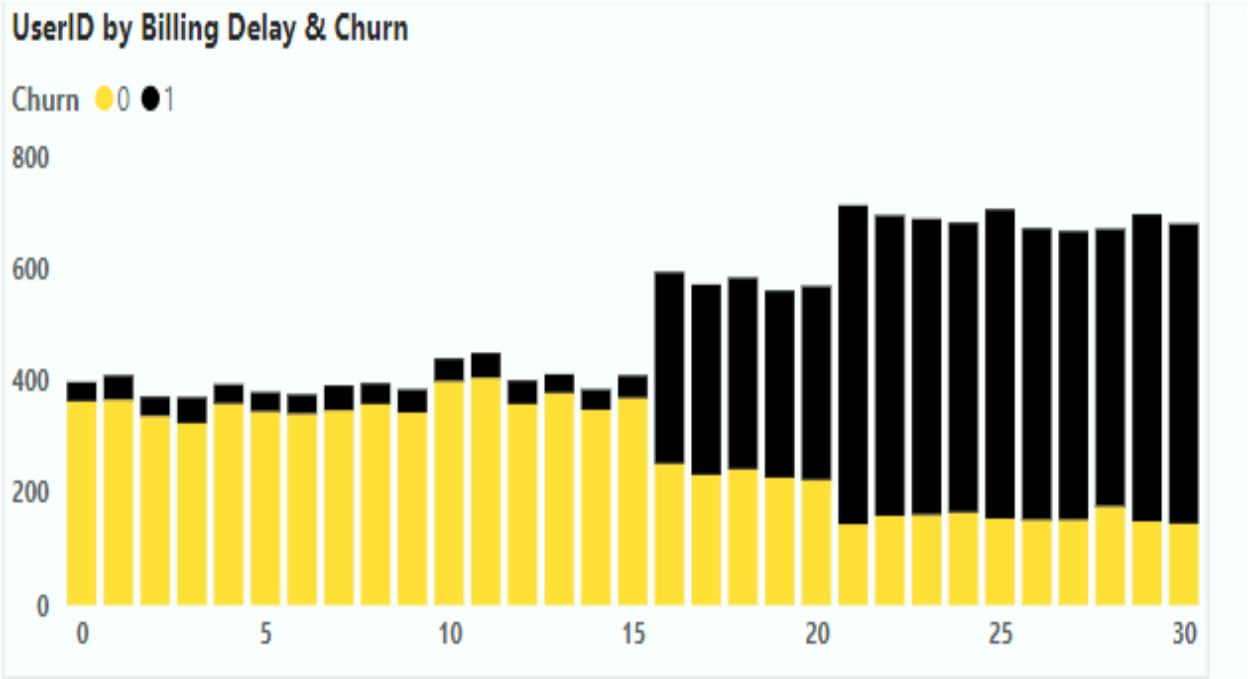
INSIGHT : Majorly the age group of 30 to 50 years spend the highest time on the platform and age segment less than 20 and 60 to 70 years also spend comparatively less time.



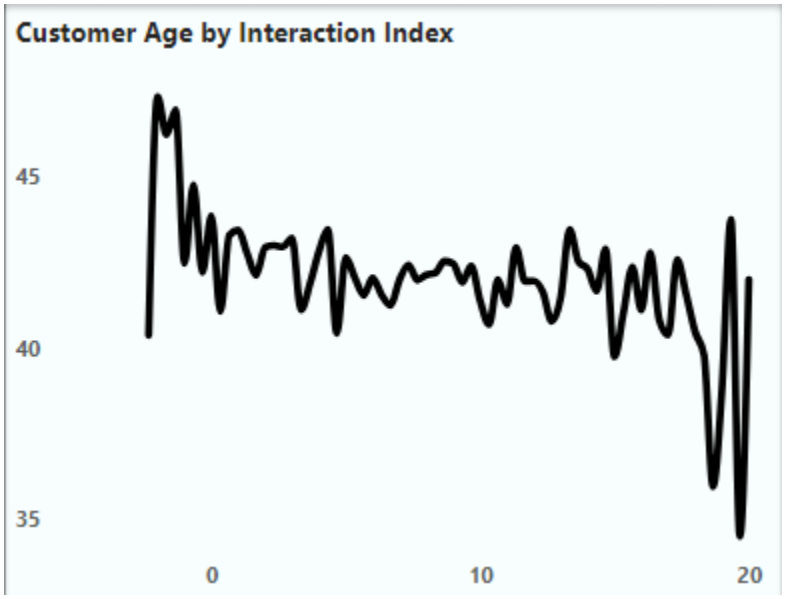
INSIGHT : It can be clearly noticed that the customers who are churning do not have a very high AOV and usually order majorly in the range of 200 to 500 rupees.



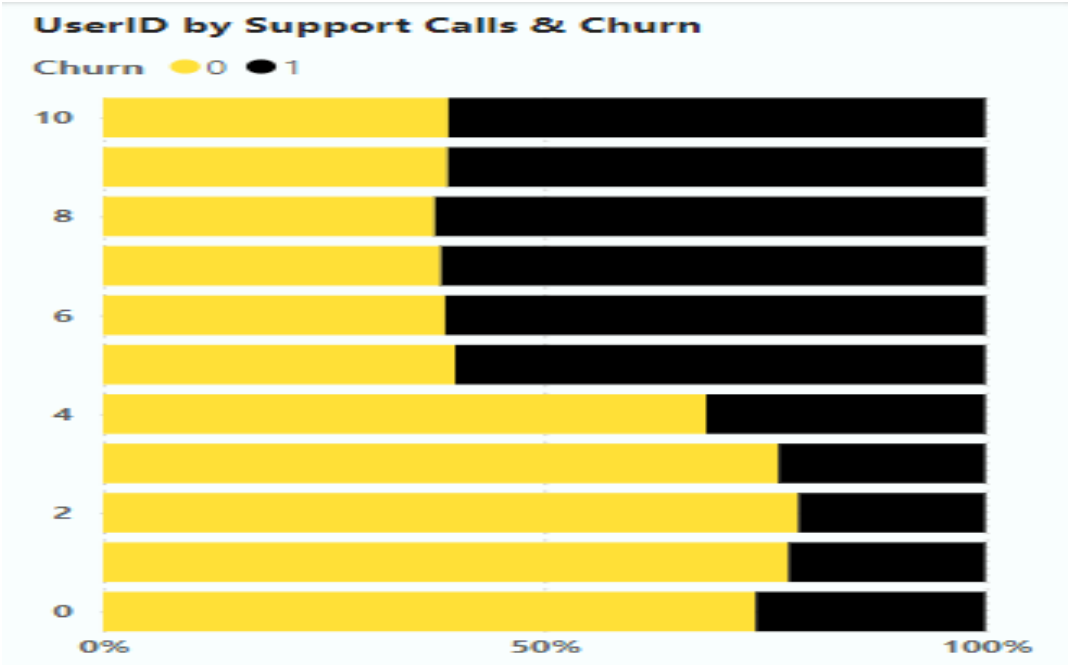
INSIGHT: As the delay in billing is increasing and crossing the x=15 (on x-axis) churn proportion is having a spike.



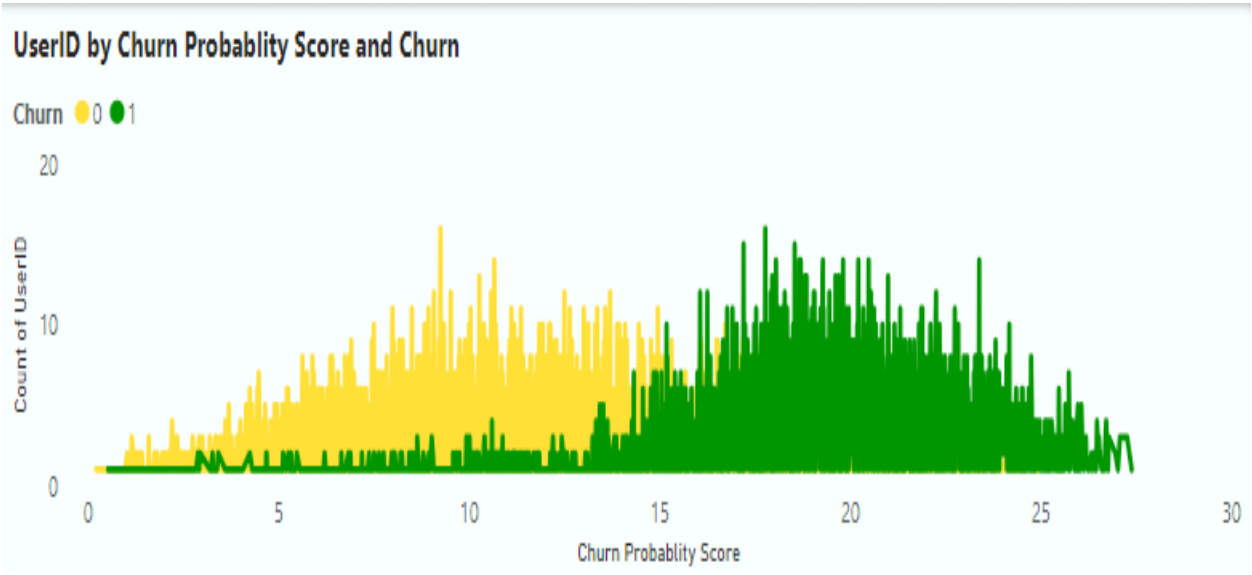
INSIGHT: Customers in the age category of 40-50 almost have a negative interaction index.



INSIGHT: **As the support calls are increasing by 5, the churn of customers is taking a sudden spike.**



INSIGHT: After crossing the CPS of 15 there is a very high spike in the customers who are churning.



INSIGHT: In Spite of taking aid subscription plans like standard & premium the average of

- **Support calls**
- **Churn Probability Score**
- **Churn Risk Index**
- **Customer Lifetime Value**
- **Billing Delay (Payment Failure)**

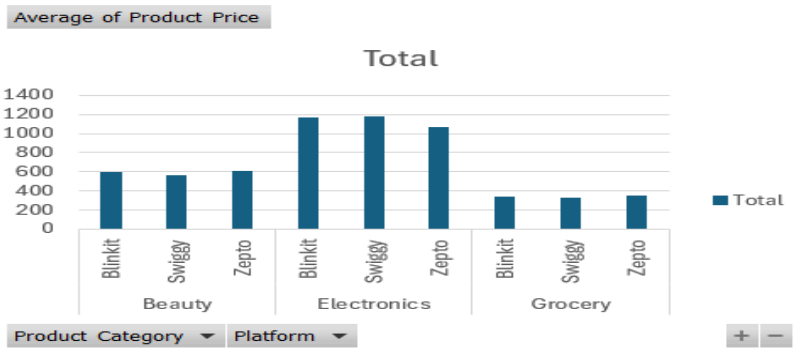
Has been the same for each plan type.

AVERAGE OF ESSENTIAL METRICS

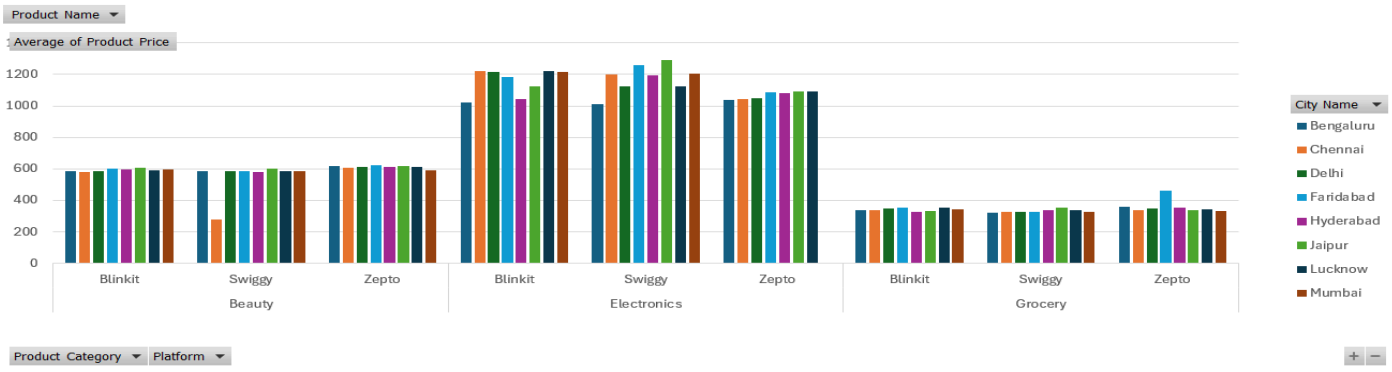
PlanType	Average of SupportCalls	Average of Churn Probablity Score	Average of Tenure to Customer age Ratio	Average of Churn Risk Index	Average of CLV	Average of BillingDelay
Basic	5.40	15.24	0.88	9.50	1,267.41	17.06
Premium	5.39	15.25	0.86	9.48	1,231.22	17.10
Standard	5.41	15.29	0.88	9.47	1,236.41	17.06
Total	5.40	15.26	0.87	9.49	1,244.94	17.07

INSIGHT: For three essential product commodities namely Beauty, Electronics & Grocery in electronics segment Blinkit is offering the highest price segment and second highest in beauty and grocery. Also in 8 major operations it fails to provide the lowest price point option of these three product commodities.

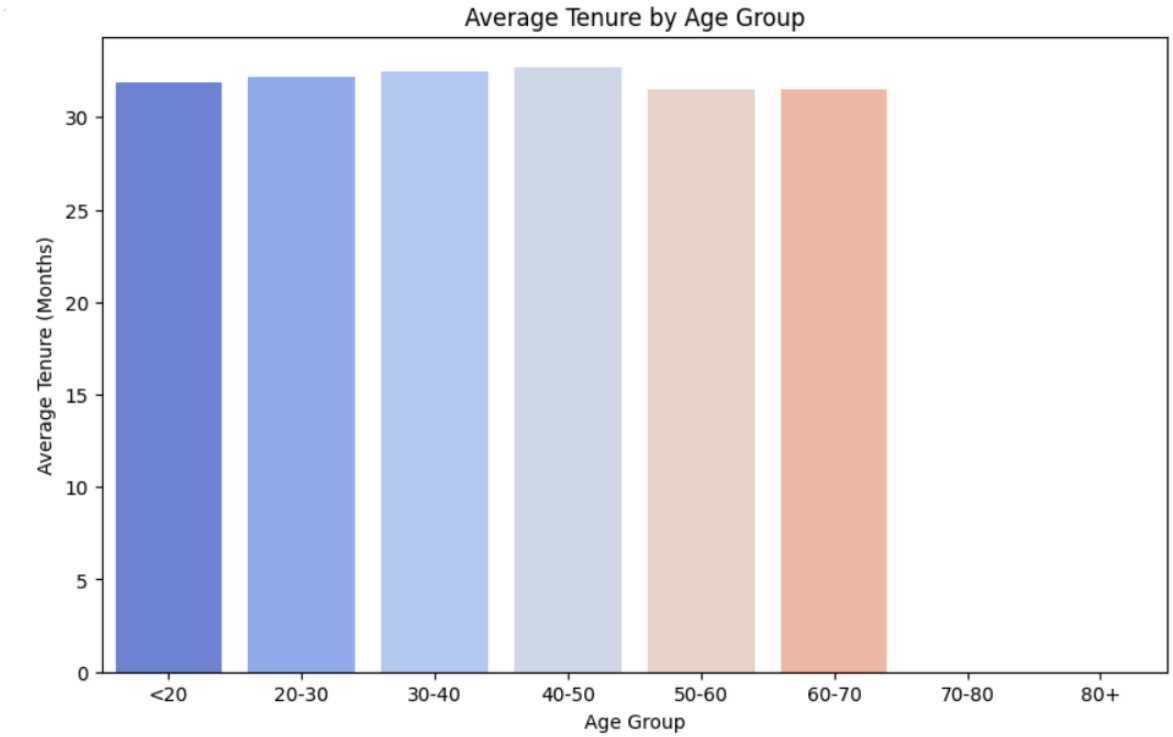
Row Labels	Average of Product Price
Beauty	**588.6571831**
Blinkit	592.875
Swiggy	559.6521739
Zepto	612.2358333
Electronics	**1138.189701**
Blinkit	1163.227273
Swiggy	1176.25
Zepto	1068.462381
Grocery	**342.9109859**
Blinkit	341.375
Swiggy	332.75
Zepto	355.1165217
Grand Total	**681.3399522**



Average of Product Price	Column Labels								
Row Labels	Bengaluru	Chennai	Delhi	Faridabad	Hyderabad	Jaipur	Lucknow	Mumbai	Grand Total
Beauty	**595.5488889**	**514.46375**	**593.5955556**	**602.7777778**	**598.1111111**	**608.5555556**	**596.3788889**	**591.5822222**	**588.6571831**
Blinkit	584.6666667	580.3333333	583.6666667	599.6666667	596.6666667	609.3333333	591.3333333	597.3333333	592.875
Swiggy	583.3333333	277.5	583.6666667	584.6666667	582.6666667	600.3333333	587	584	559.6521739
Zepto	618.6466667	606.57	613.4533333	624	615	616	610.8033333	593.4133333	612.2358333
Electronics	**1023.674444**	**1154.425556**	**1129.484444**	**1177**	**1114.875**	**1174.375**	**1145.827778**	**1211.833333**	**1138.189701**
Blinkit	1022.666667	1224	1216.333333	1183.666667	1045.5	1126	1219.666667	1216.333333	1163.227273
Swiggy	1010.666667	1198	1125.666667	1258	1194.333333	1291.333333	1124.666667	1207.333333	1176.25
Zepto	1037.69	1041.276667	1046.453333	1089.333333	1081.666667	1089.666667	1093.15	#DIV/0!	1068.462381
Grocery	**338.2955556**	**334.3844444**	**342.8266667**	**371.25**	**338**	**341**	**346.4066667**	**334.2733333**	**342.9109859**
Blinkit	336	338.6666667	351	352	325.3333333	330.6666667	356.6666667	340.6666667	341.375
Swiggy	319.3333333	327.3333333	329.6666667	329	335.6666667	355.3333333	336.6666667	329	332.75
Zepto	359.5533333	337.1533333	347.8133333	463.5	353	337	345.8866667	333.1533333	355.1165217
Grand Total	**652.5062963**	**673.6538462**	**688.6355556**	**730.3076923**	**667.0769231**	**690.0384615**	**696.2044444**	**650.1541667**	**681.3399522**



INSIGHT: Age group as per tenure spent is uniformly distributed across the dataset.



INSIGHT: In all the categories almost 47-48% consumers are churning which implies premium, standard are not getting any special benefits and all the plan types are equally risk potent to churning of customers.