Video Streaming App Subscription Analysis

Understanding Customer Behavior and Revenue Trends

Name: Mohit Rathod

Date: 15th October 2025

GitHub Repo: View Here

INDEX

INTRODUCTION

- **Goal:** To analyze streaming video subscription data to understand customer behavior, identify key drivers of churn, and provide actionable insights for business growth.

- **Dataset:** Subscription Cohort Analysis Data.csv, containing information on customer subscriptions, payment status, cancellation dates, etc.

- **Methodology:** Data cleaning, exploratory data analysis, customer segmentation, and churn prediction using a machine learning model.

IMPORT LIBRARIES

Started by importing essential libraries for data manipulation, analysis, and visualization:

- **Pandas:** For data handling and manipulation using DataFrames.
- **NumPy:** For numerical computations and array operations.
- **Matplotlib:** For creating static, interactive, and animated visualizations.

DATA EXPLORATION

Used subs_df.info() to:

•Understand the structure of the dataset, including the number of rows and columns.

•Check data types of each column (e.g., integer, float, object).

•Identify missing values (null or NaN) in any columns.

```
subs_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3069 entries, 0 to 3068
Data columns (total 6 columns):
 #   Column                Non-Null Count  Dtype 
---  ------                --------------  ----- 
 0   customer_id           3069 non-null   int64 
 1   created_date          3069 non-null   object
 2   canceled_date         2004 non-null   object
 3   subscription_cost     3069 non-null   int64 
 4   subscription_interval 3069 non-null   object
 5   was_subscription_paid 3069 non-null   object
dtypes: int64(2), object(4)
memory usage: 144.0+ KB
```

DATA TRANSFORMATION

	customer_id	created_date	canceled_date	was_subscription_paid
0	154536156	2022-09-01	NaN	1
1	149713408	2022-09-01	2022-09-02	0
2	153756284	2022-09-01	2022-09-02	0

Recoding Categorical Variables: Converted categorical variables like 'was_subscription_paid' (Yes/No) into numerical format (1/0) for memory efficiency and compatibility with machine learning models.

Date Handling: Transformed 'created_date' and 'canceled_date' columns into datetime objects for easier date-based calculations and analysis.

FEATURE ENGINEERING

- Created new columns, such as 'created_month' and 'canceled_month,' by extracting the month from the respective date columns. This enabled monthly analysis of churn and other metrics.

- Calculated the 'subscription_duration' by subtracting the creation date from the cancellation date, providing a measure of how long each subscription lasted.

- Categorized subscription durations into groups *('0-7 days', '8-30 days', '31-60 days', '61-90 days', '91-120 days', '121-150 days', '151-180 days', '181-365 days', '1-2 years')* using the pd.cut function to analyze retention patterns across different duration ranges.

CHURN ANALYSIS

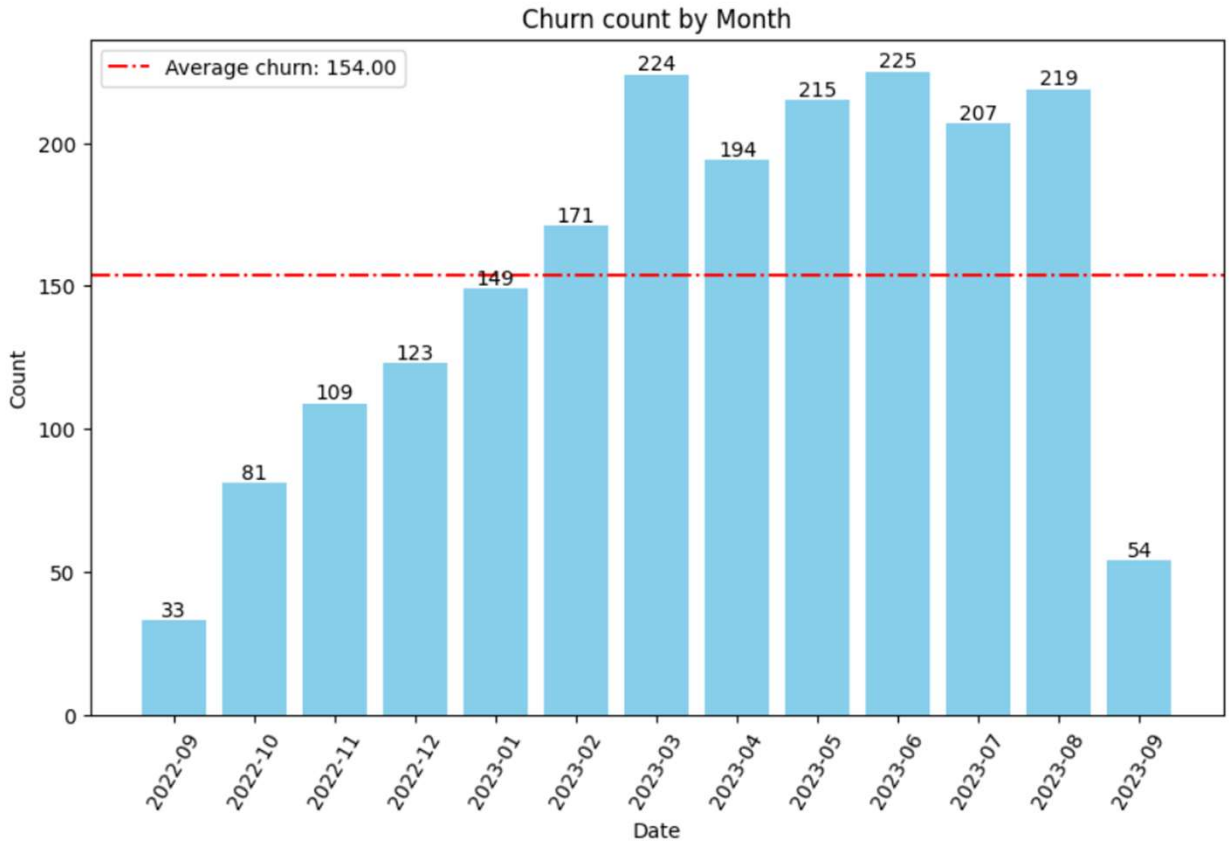
Category	Total	Subscription Paid	Subscription Not Paid
Total Subscriptions	3069 **(100%)**	2936 **(95.66%)**	133 **(4.34%)**
Cancellations	2004 **(65.29%)**	1881 **(93.86%)**	123 **(6.14%)**

Higher Cancellation among Non-Paying Users indicates that free users are more likely to churn after their initial engagement.

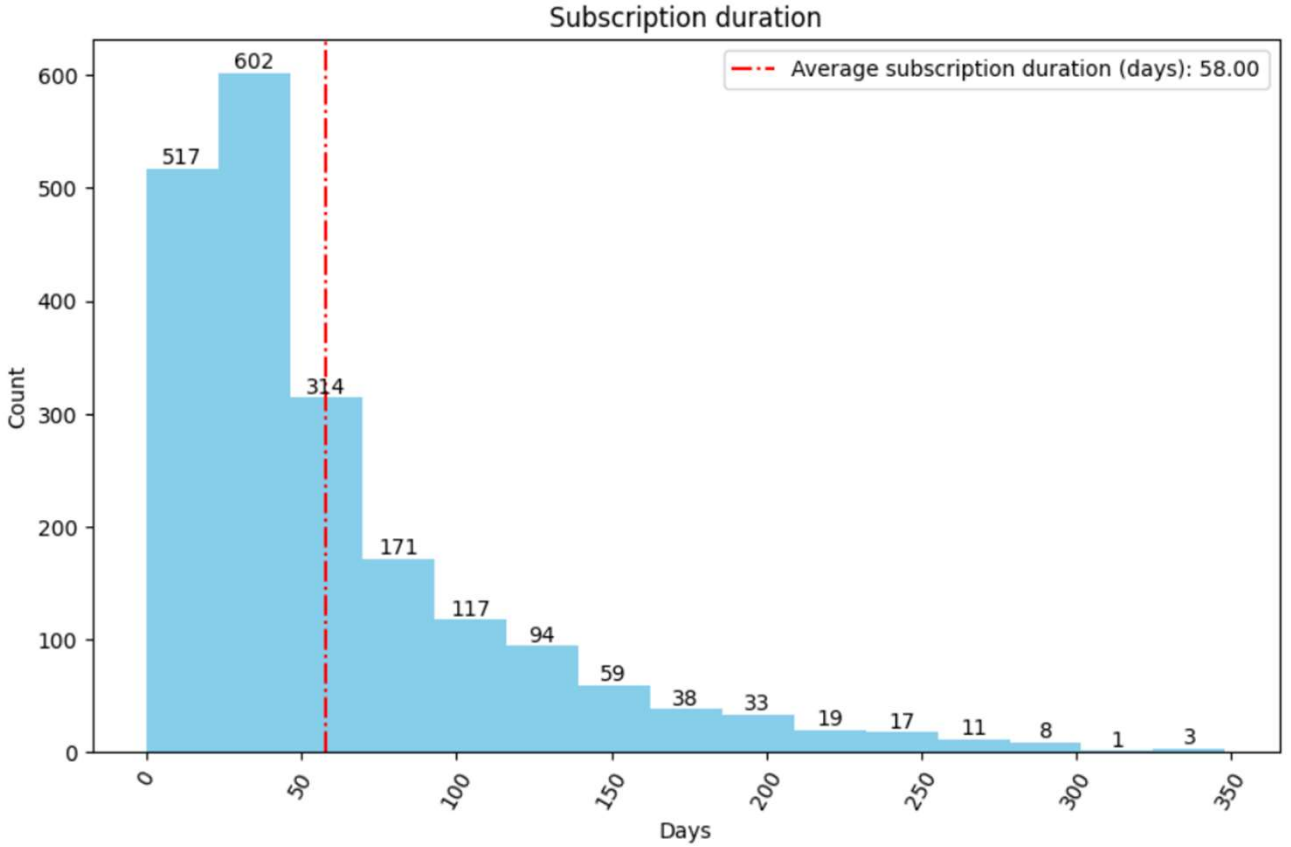
Retention Strategy Opportunity:
Paid users show a stable retention pattern, but non-paying users present an opportunity for improvement. Focusing on converting free users into paying subscribers can reduce cancellations and enhance long-term growth.

CHURN ANALYSIS

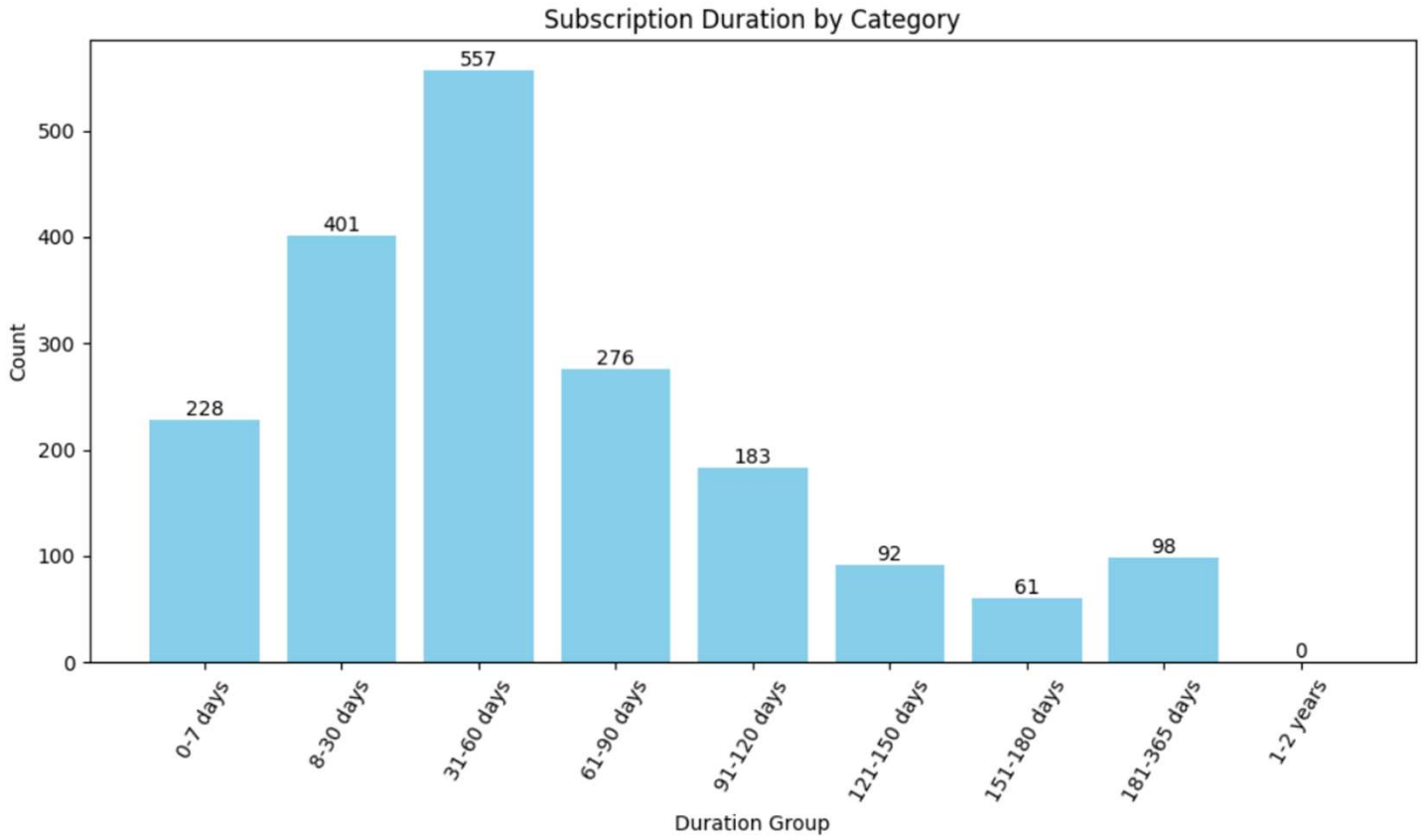
Total Customers Cancelled their Subscription by Month



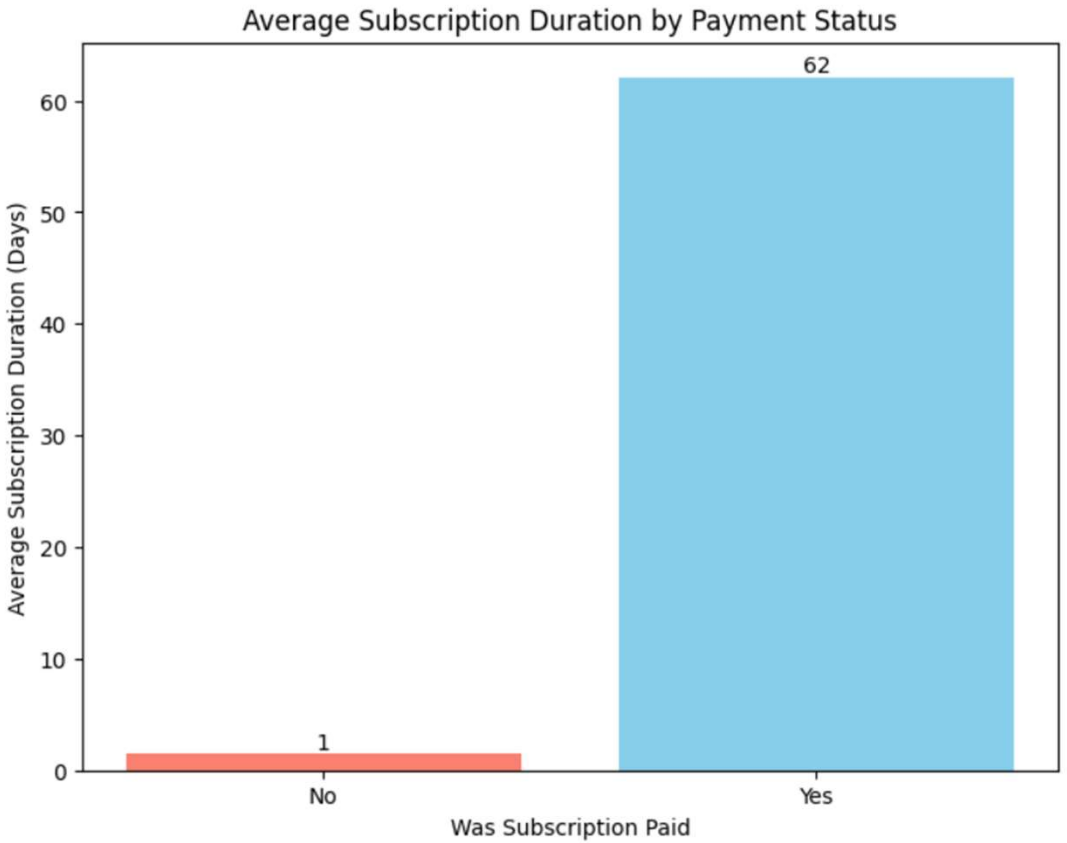
RETENTION ANALYSIS



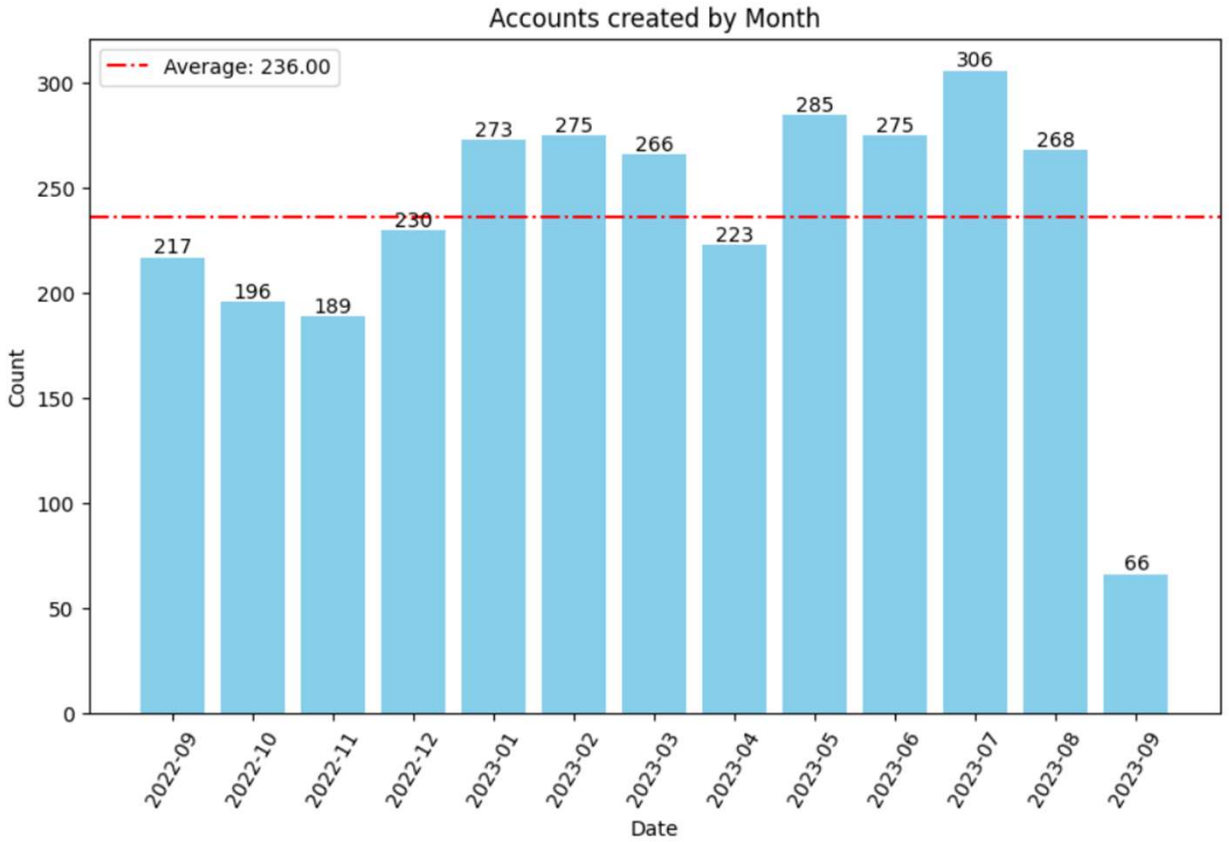
RETENTION ANALYSIS



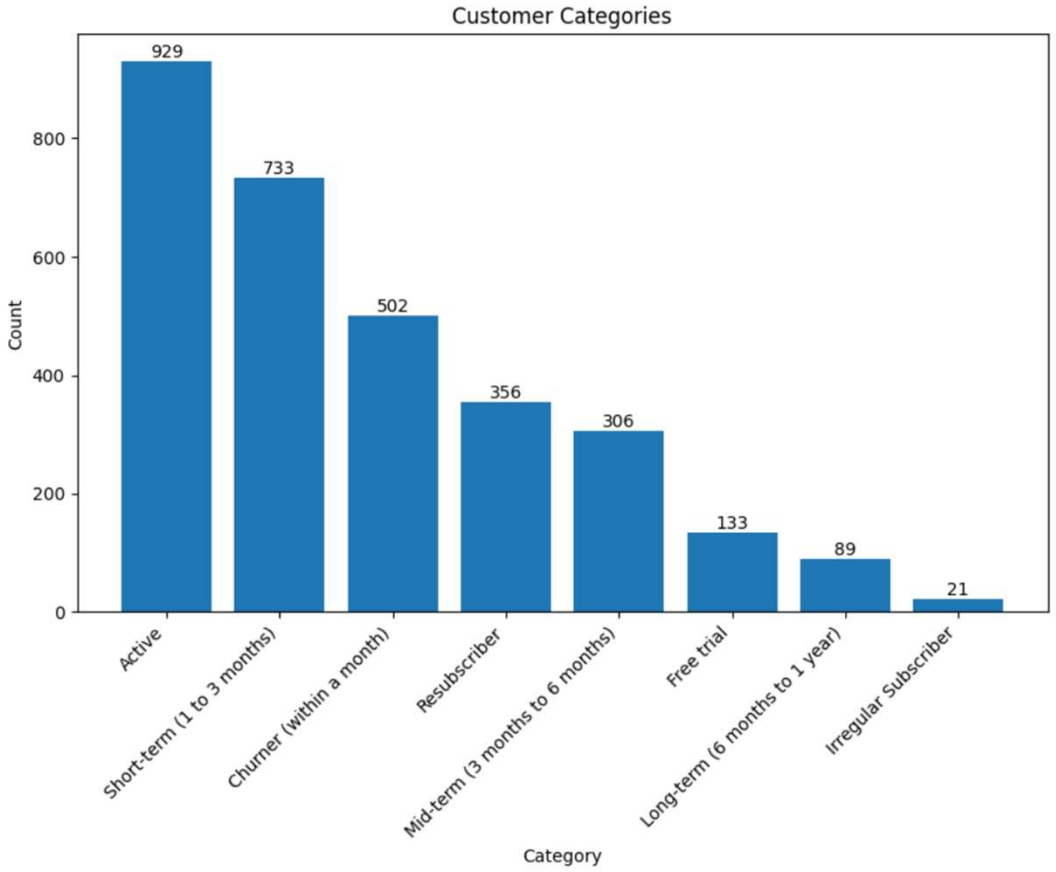
RETENTION ANALYSIS



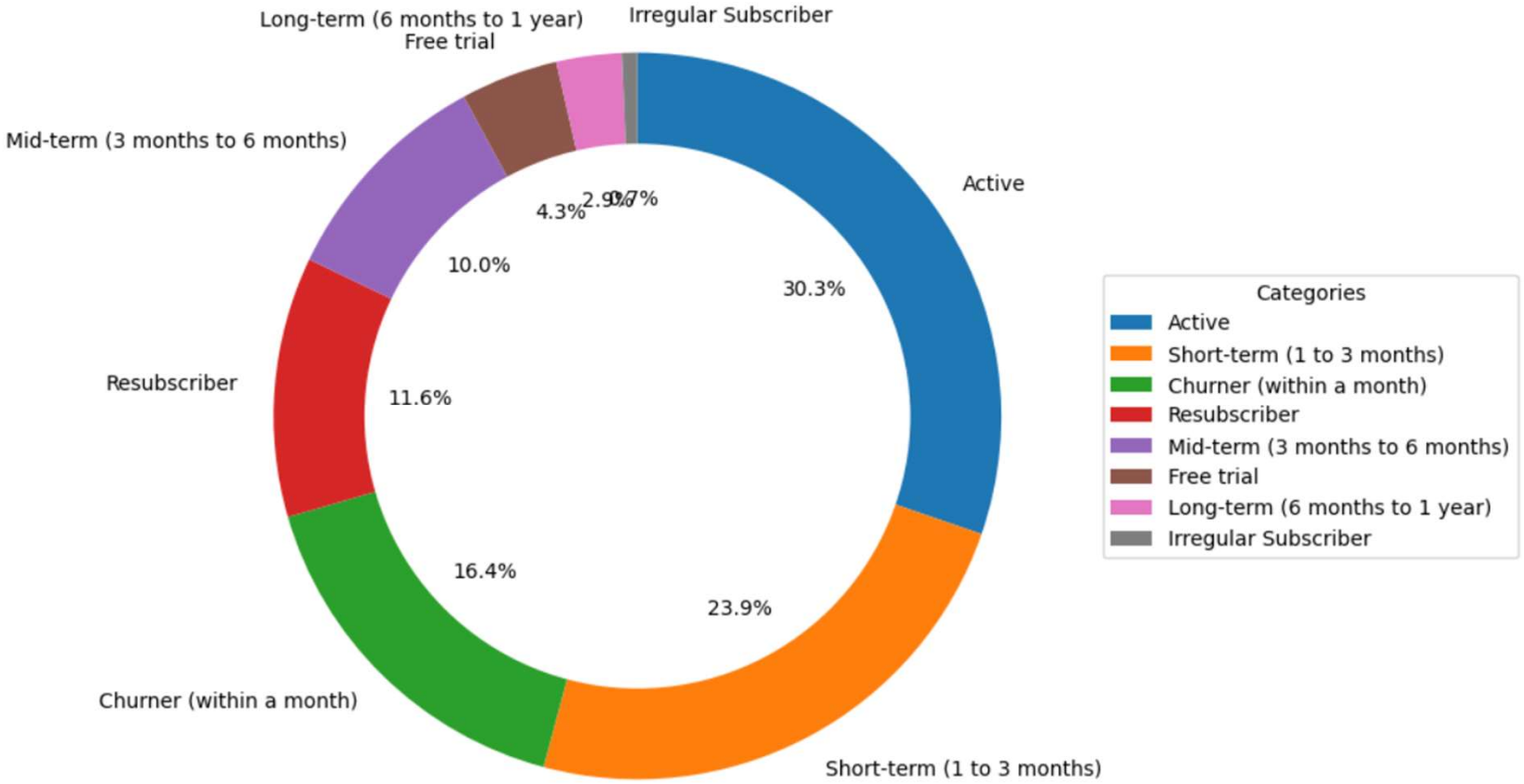
MONTHLY NEW SUBSCRIBERS



CUSTOMER SEGMENTATION



CUSTOMER SEGMENTATION



CONCLUSION

Key Insights: Identified key factors driving churn, including subscription duration, payment status, and customer categories.

Actionable Strategies: Developed data-driven recommendations for reducing churn, such as targeted interventions for high-risk segments and enhancing customer engagement.

Future Work: Outlined areas for future analysis, including investigating external factors and exploring more advanced predictive models.

Overall, this analysis provides a comprehensive understanding of subscription churn and empowers the business to make data-informed decisions for improved customer retention.

THANKYOU

Thank you for your time and attention throughout this presentation. I appreciate your interest and hope my work provided valuable insights.

- Presented by: **Mohit Rathod**

- Feel free to reach out for further inquiries or networking:
  - **Email:** mohitrathod.ca@gmail.com
  - **LinkedIn:** https://www.linkedin.com/in/mohit-rathod/

I look forward to connecting!