

# Rochester Museum & Science Center Team 1

## Final Presentation

Ruqin Chang, Valerie Tam, Shawn Wu, Ruiyu Zhou, Ziyi You



Science  
Museum  
+

Strasensburgh  
Planetarium  
+

Cumming  
Nature Center  
+

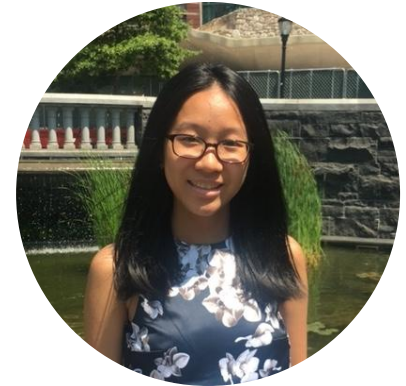
- **Section 1: Project Overview**
  - Team Member Introduction
  - Project Vision and Objectives
  - Business Understanding
  - Workflow Summary
- **Section 2: Data Analysis**
  - Membership Transaction Table
  - Earned Revenue Transaction Table
  - Complete Journey Table
  - Donation Table
- **Section 3: Project Management**
  - Challenges
  - Milestones Achieved
  - Next Steps
- **Questions**

# Section 1: Project Overview

# Team Members



**Product Manager**  
Ziyi You



**Data Analyst**  
Valerie Tam



**Visualization Manager**  
Ruiyu Zhou



**Project Manager**  
Shengyang (Shawn) Wu



**Data Analyst**  
Ruqin Chang

# Project Vision

To spur museum membership growth, encourage donations from members, and increase overall museum revenue

# Project Objectives

- 1) Data visualization showing patterns of membership transactions and renewals
- 2) Create a program that can use new CSV transaction file inputs to output a list of members who may be likely to renew or donate
- 3) An education module to explain to non-data-scientists how the conclusions were achieved

# Section 2: Data Analysis

# Membership Transactions Table



# Data Description

- One transaction each time the member joins, renews, upgrades, or downgrades
- 64,439 records, 8 features

Constituent ID	Member number
Transaction Date	Date of membership transaction
Action	Join, renewal, upgrade, downgrade
Expiration Date	Expiration date of membership
Is Gift	If the membership was a gift
Membership Program	Type of membership program (upper level or general membership)
Membership Level	Level of membership
Amount	Amount paid for membership

# Earned Revenue Transactions Table

# Summary

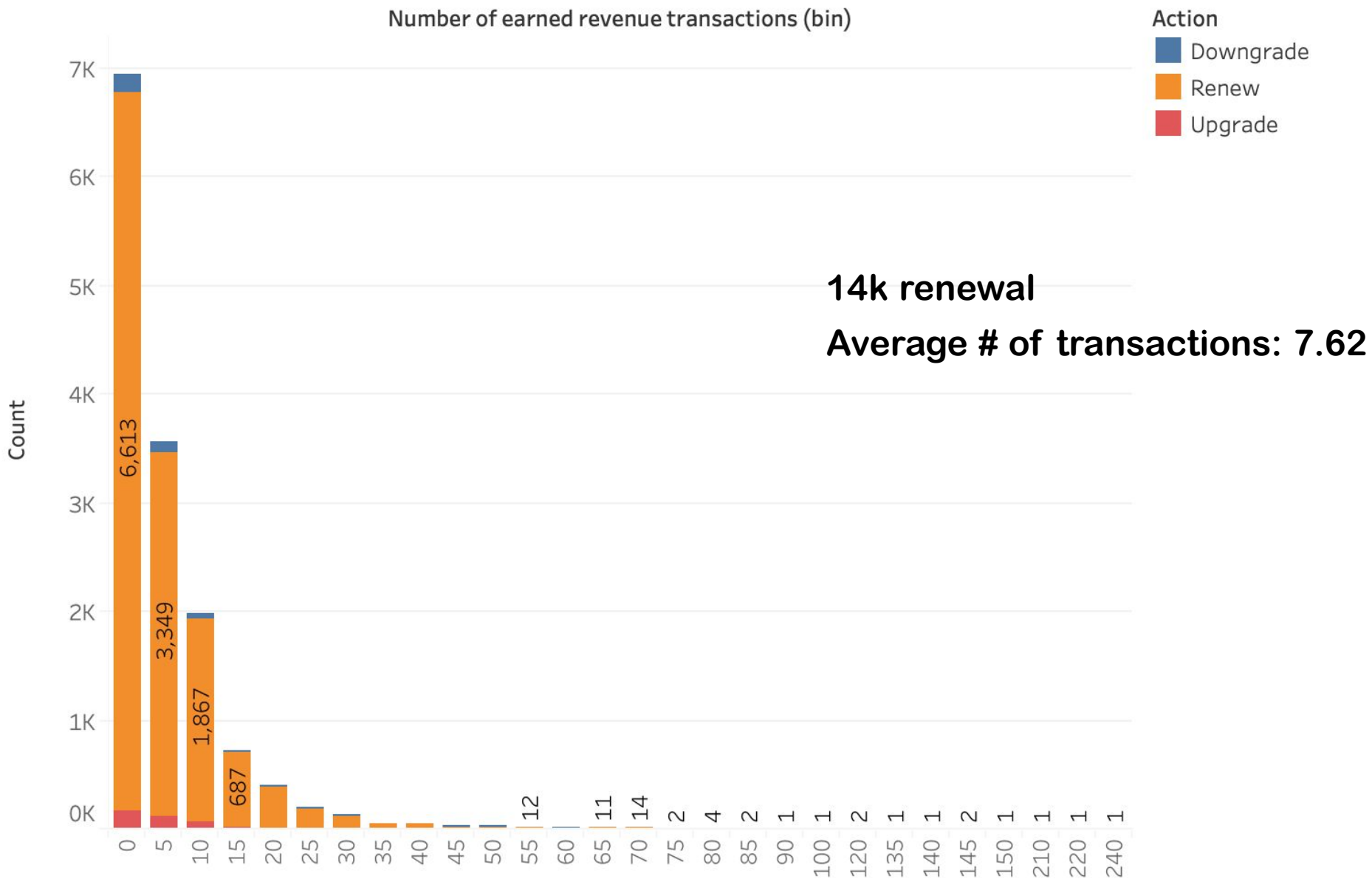
- a. Data description
- b. Data cleaning
- c. Data exploration and visualization:
  - i. Number of earned revenue transactions made within one membership period
  - ii. Number of visits made within one membership period
  - iii. Days to renew and number of visits
  - iv. Adult membership analysis update

# Data Description and Cleaning

- 332066 records, 13 features
- Cropped ignored columns based on feature description

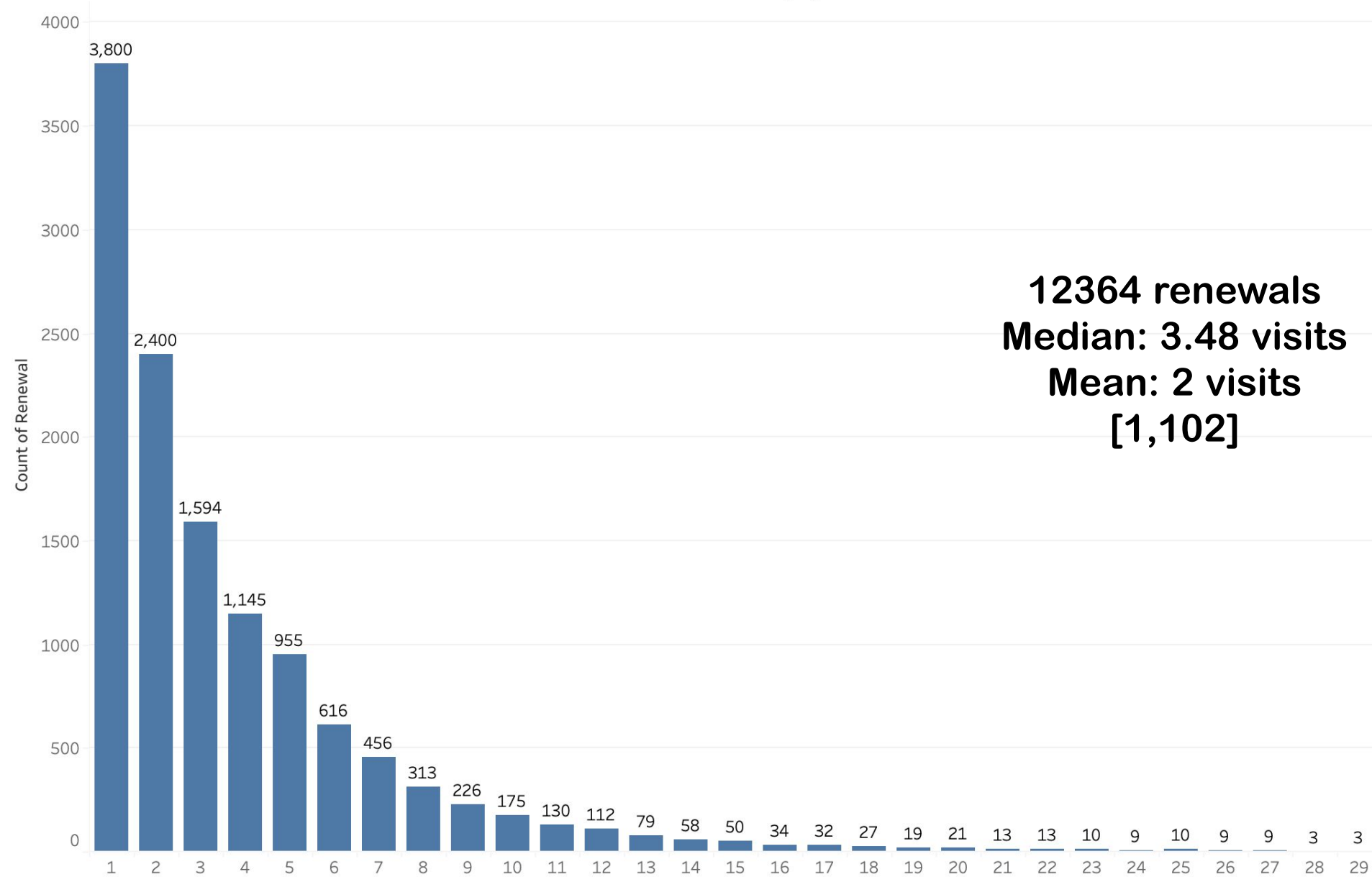
BUSINESSPR	Record counter	PROGRAM_NA	Program Name
ORDER_ID	Order number in system of earned revenue transaction	PROGRAM_DA	Date of program
TRANSACTION	Date of Earned Revenue Transaction	CATEGORY	Type of transaction
ORDER_TOTA	Amount paid for Earned Revenue Transaction	PRICE_TYPE	Price type
CONSTITUEN	Member number	LIST_PRICE	Price
MEMBER_LEV	Membership level	QTY	Quantity of tickets
ZIP	Zip code of member		

# 1. Number of earned revenue transactions made within one membership period (Ruqin)

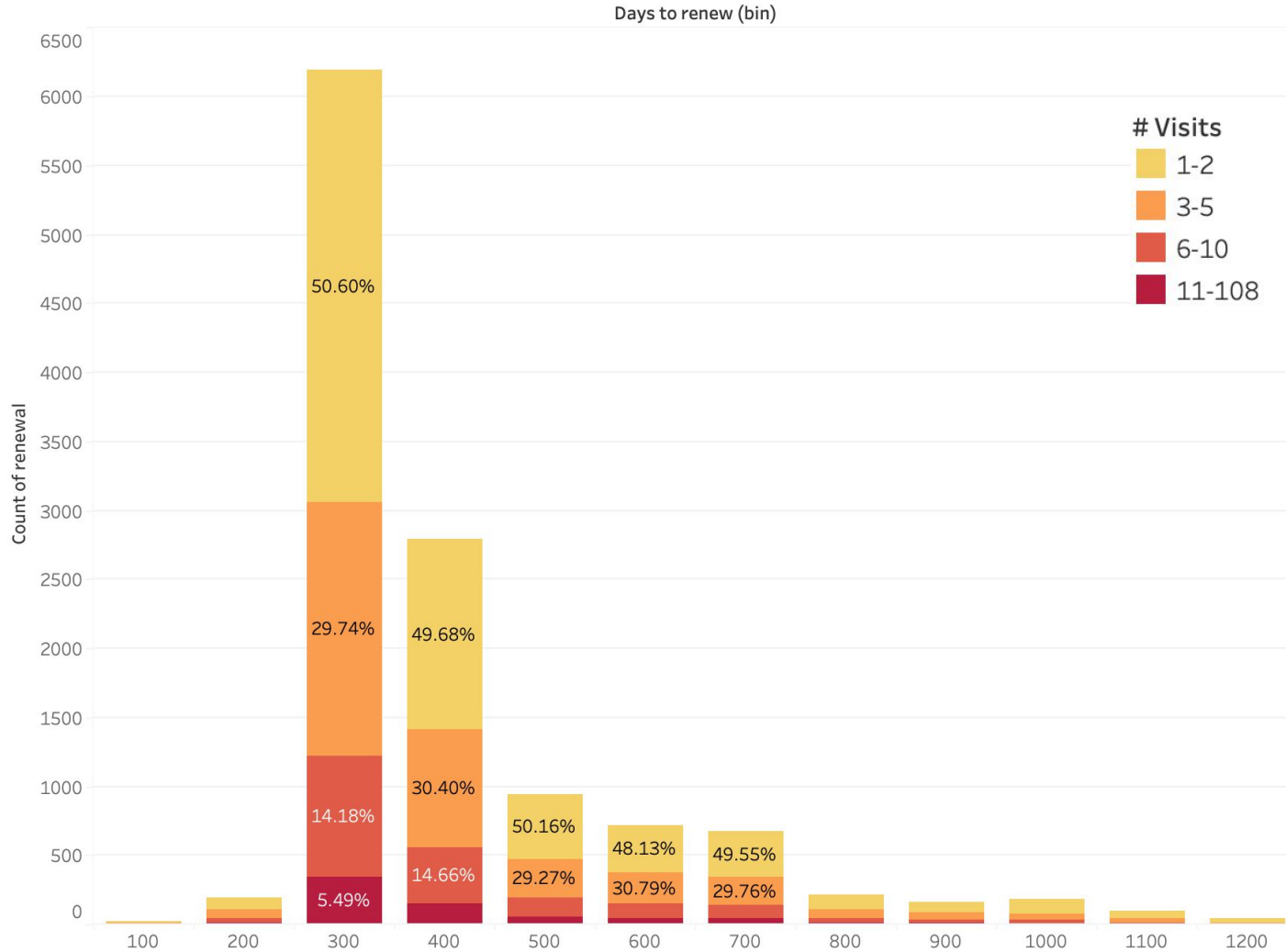


## 2. Number of visits made within one membership period (Ruqin)

Number of Visits (bin)

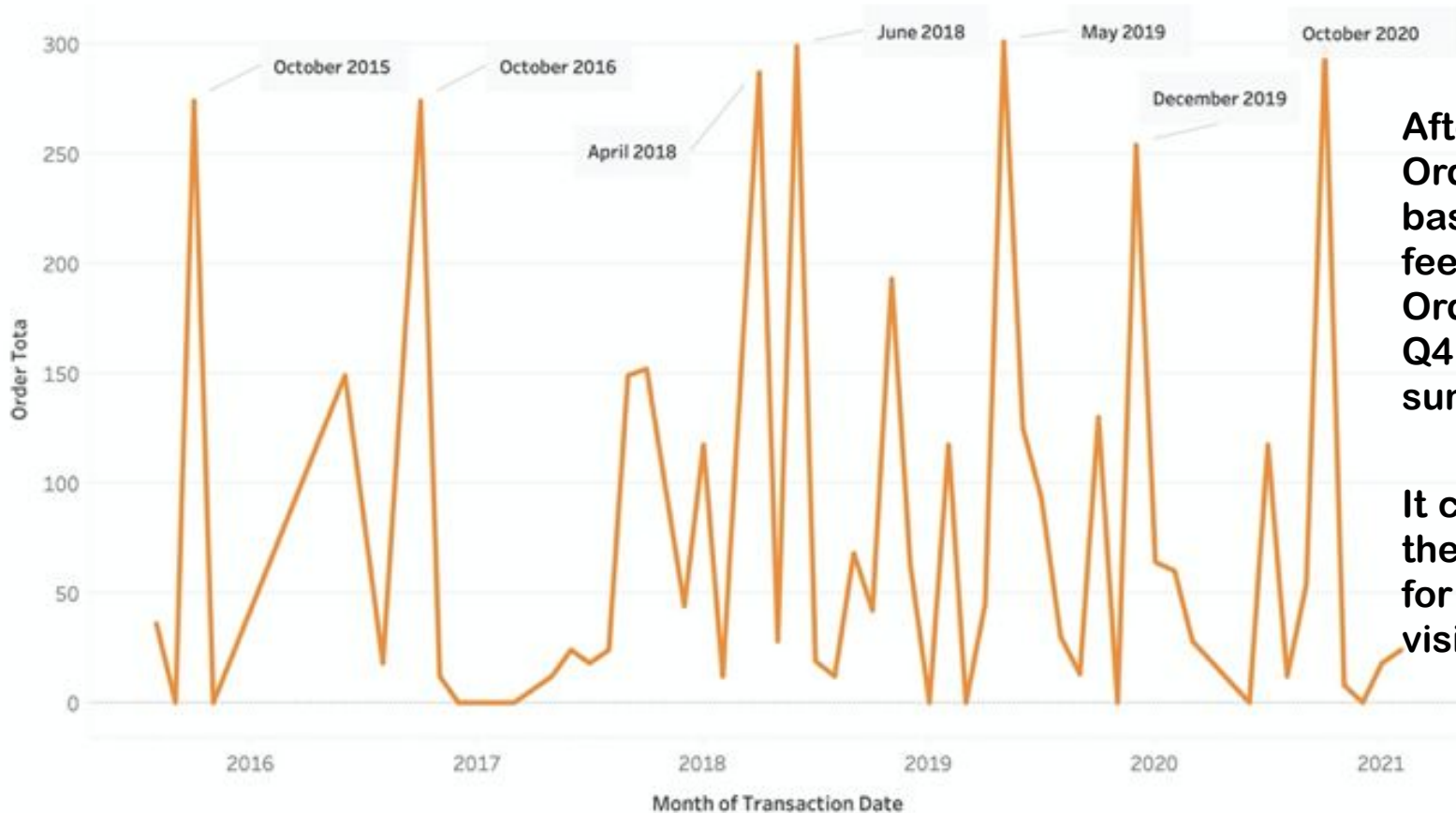


# 3. Days to renew and number of visits (Ruqin)



# 4. Adult Membership Analysis Update (Shawn)

Order Total vs. Month for Adult Members



After filtering out Order total > 200 based on sponsor feedback, we see a Order Total peak at Q4 and during the summer

It corresponds to the popular times for adult visitors to visit the museum

The trend of sum of Order Tota for Transaction Date Month. The data is filtered on Member Lev and Order Tota. The Member Lev filter keeps Adult Individual with Caregiver. The Order Tota filter ranges from 0 to 200.



# Complete Journey Table

# Construct a Complete Journey Table (Ruiyu)

- **Motivation:**
  1. To analyze behaviors of new members who join between 2015-2019
  2. To calculate the number of the total active, new-join, and returning members during certain time periods
  
- **Data Cleaning and Preprocessing of Membership Transaction Table:**
  1. Removed the duplicated “Join”: 88 members
  2. Restricted the years to be within 2015-2019
  3. Removed members without the initial Join record: ~9000 members
  
- **Created new columns:**
  1. Cohort Period/Date
  2. Action\_count
  3. Join\_date
  4. Final\_expiration\_date

# Data Description

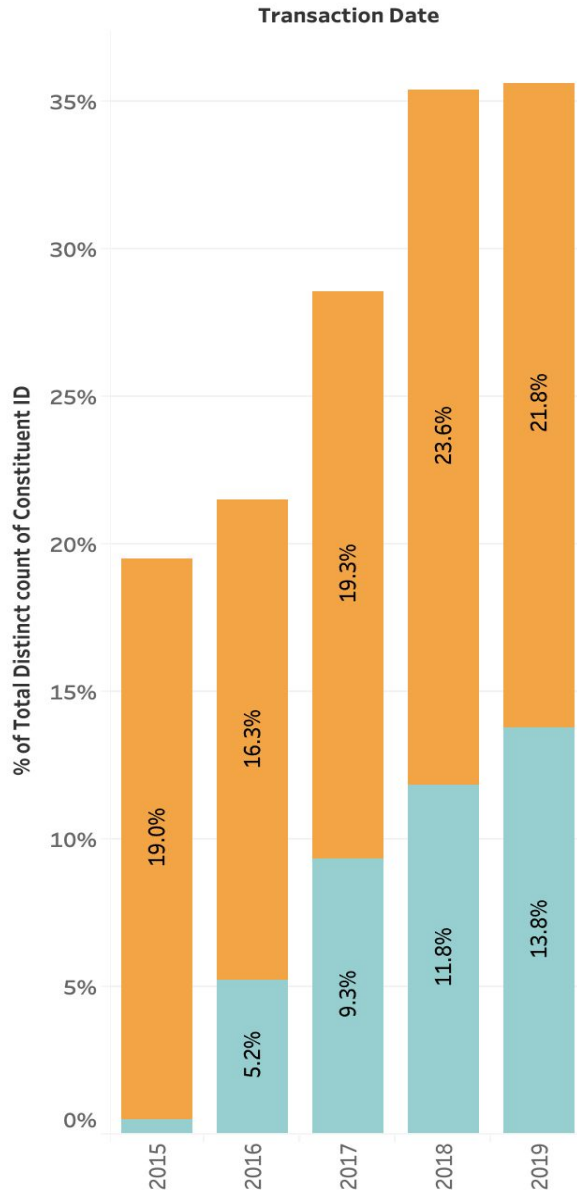
- o Number of records: 30328
- o Number of distinct members: 20900
- o Number of features: 12

Four more columns are created:

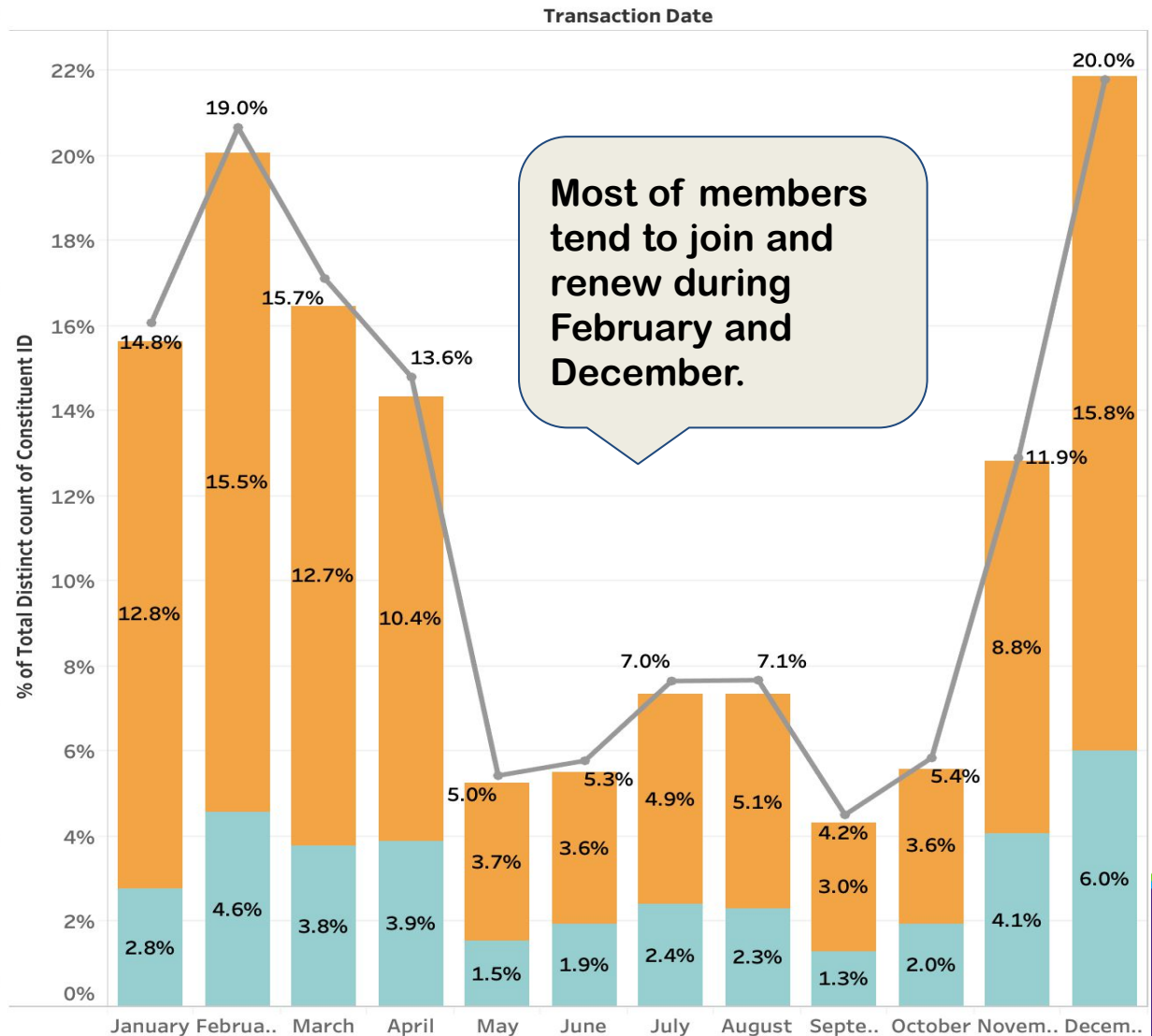
Membership Duration	Date difference between the Transaction Date and Expiration Date of each row
join_date	Date of members' initial join
final_expiration_date	Date of memberships' final expiration date
action_count	The cumulative count of renew/join/upgrade/downgrade action

# 1. Join and Renew within 2015-2019 (Ruiyu)

Join & Renew within 2015-2019

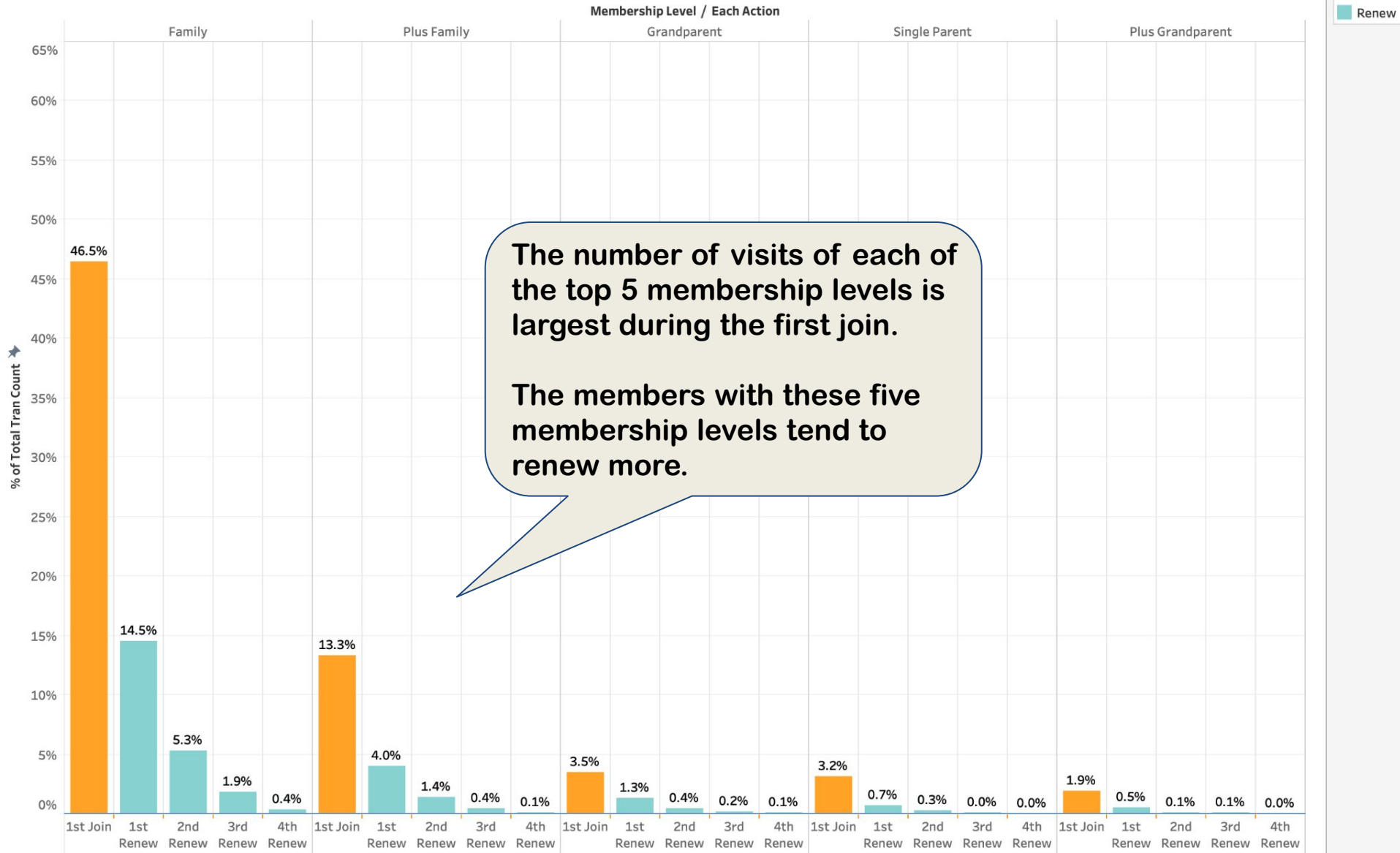


Join & Renew within 2015-2019 by Month



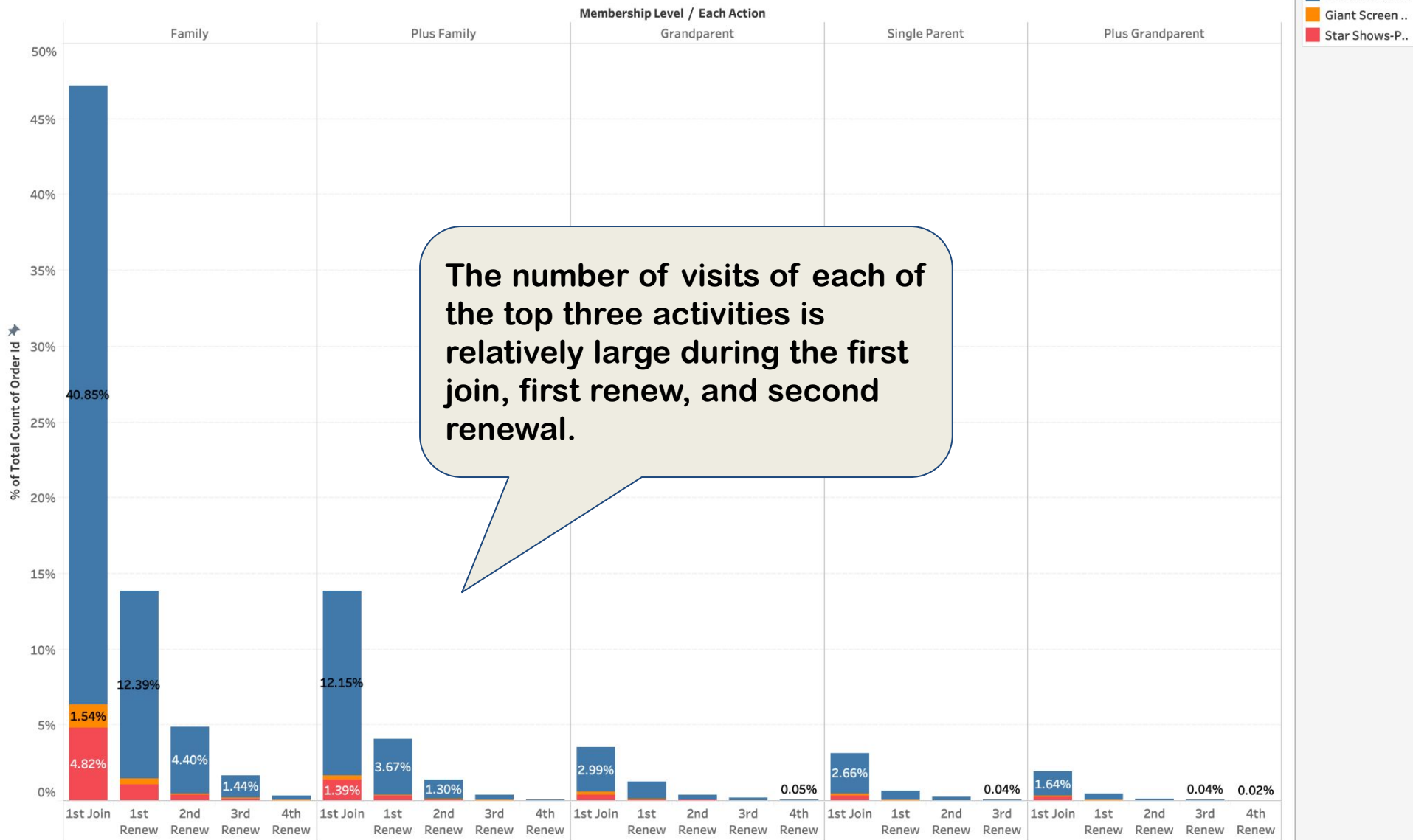
## 2. #Visit v.s. Different Times of Action for Top 5 Membership Levels (Ruiyu) 21

# Visits V.S. Different Times of Action



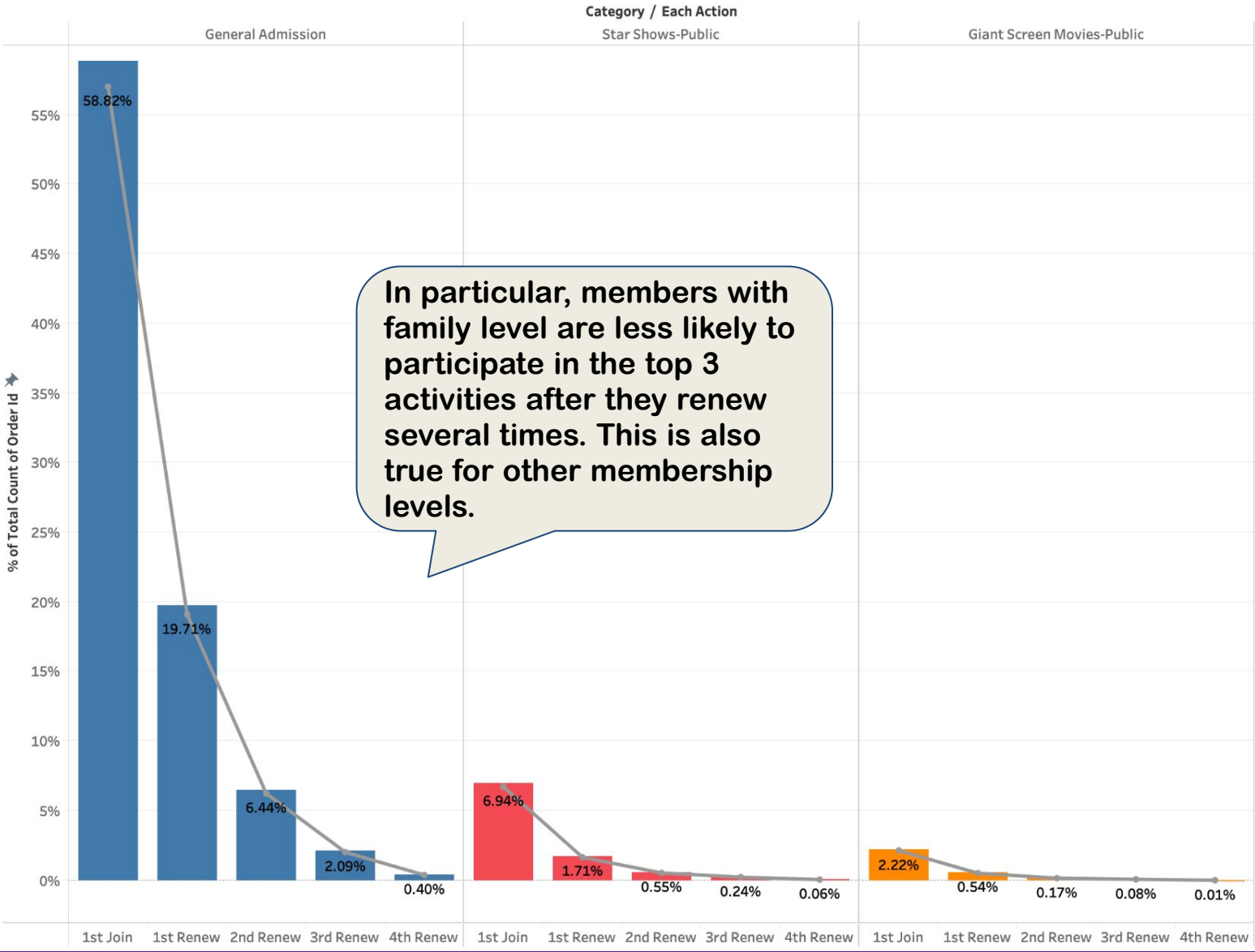
# 3. Number of Visit for Top 3 Programs in Top 5 Membership Levels (Ruiyu) 22

# Visits of Top 3 Activity Categories for Top 5 Membership Levels



# 4. Number of Visits for Top 3 Programs of Family Level (Ruiyu)

# Visits of Top 3 Activity Categories of the Family Level



In particular, members with family level are less likely to participate in the top 3 activities after they renew several times. This is also true for other membership levels.

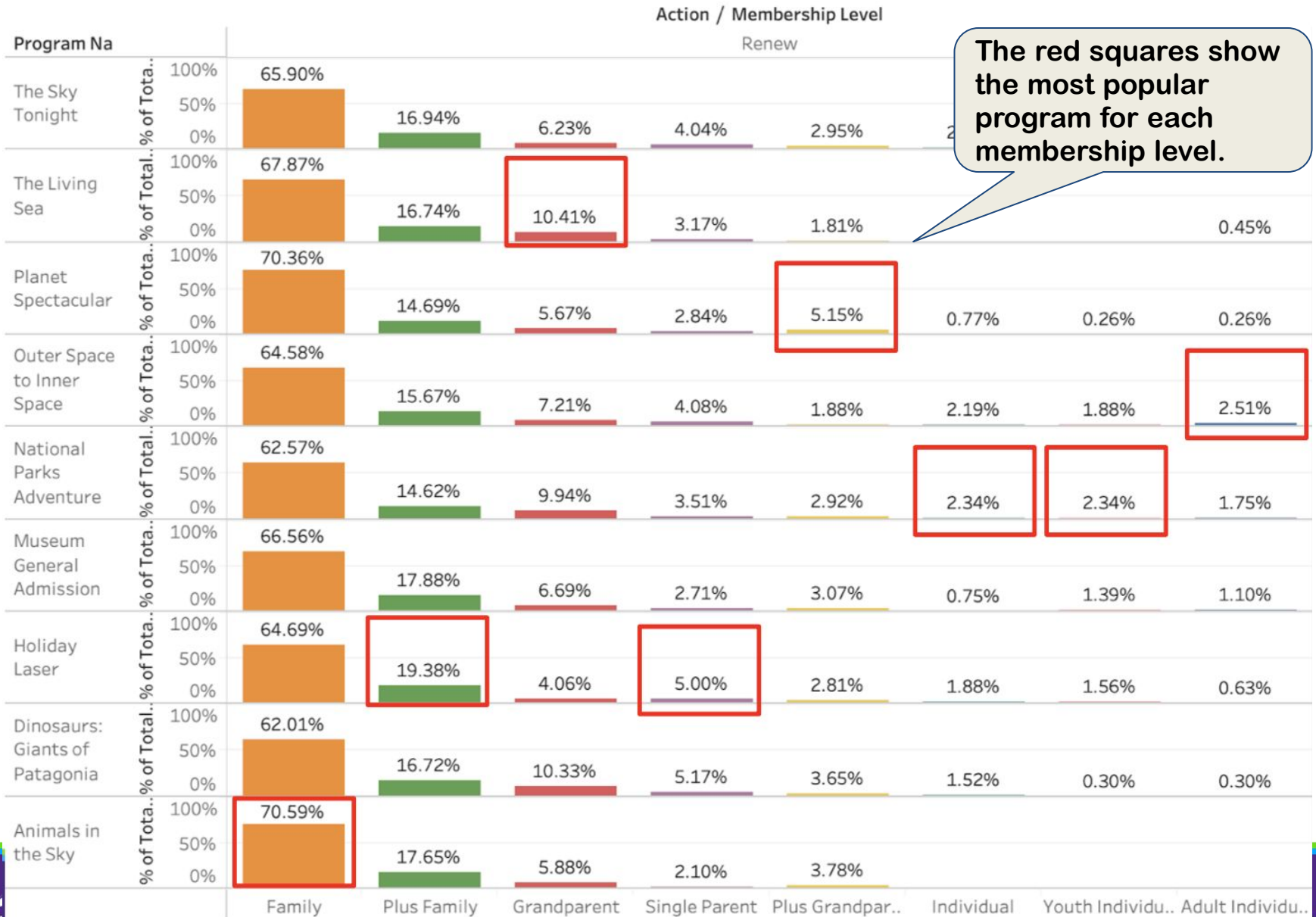
Member Lev

- (All)
- Adult Individual with Caregiver
- Community Partner Pass
- Complimentary Family
- Family
- Grandparent
- Heritage Education
- Heritage Innovation
- Individual
- Inventor
- Investigator
- Patron
- Pines Pass
- Plus Family
- Plus Grandparent
- Single Parent
- Youth Individual with Caregiver

Category

- General Admission
- Star Shows-Public
- Giant Screen Movies-Public

# 5. Membership Level vs. Program Name (Ziyi)

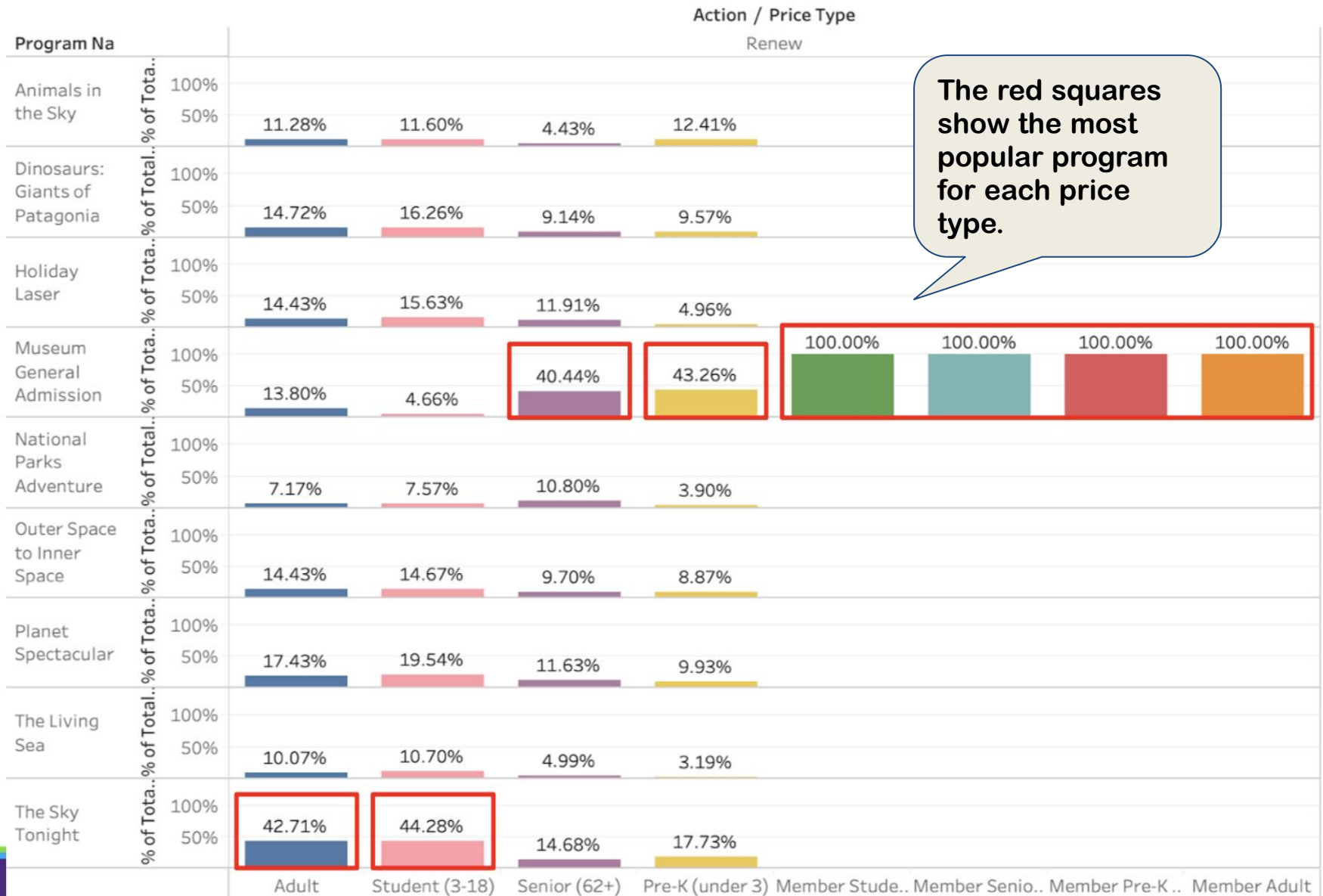


The red squares show the most popular program for each membership level.

% of Total Distinct count of Constituent ID for each Membership Level broken down by Action vs. Program Na. Color shows details about Membership Level. The marks are labeled by % of Total Distinct count of Constituent ID. The view is filtered on Action and Program Na. The Action filter keeps Renew. The Program Na filter keeps 9 members.



# 6. Price Type vs. Program Names (Ziyi)



The red squares show the most popular program for each price type.

% of Total Distinct count of Constituent ID for each Price Type broken down by Action vs. Program Na. Color shows details about Price Type. The marks are labeled by % of Total Distinct count of Constituent ID. The view is filtered on Action, Price Type and Program Na. The Action filter keeps Renew. The Price Type filter keeps 10 of 65 members. The Program Na filter keeps 9 members.

# Activity Participation Rate for Renew Members

# Activity Participation Rate for Renew Members

## Deduplication

Deduplicated the transaction records

## Categorization

Clustered activities into 5 categories

## Yearly Activity Participation Rate

Calculated activity participation rate for each membership level and each year (2016-2019)

## Average Participation Rate

Averaged the participation rate of the 4 years



**RMSC**  
Rochester Museum & Science Center

reward  
your  
curiosity.

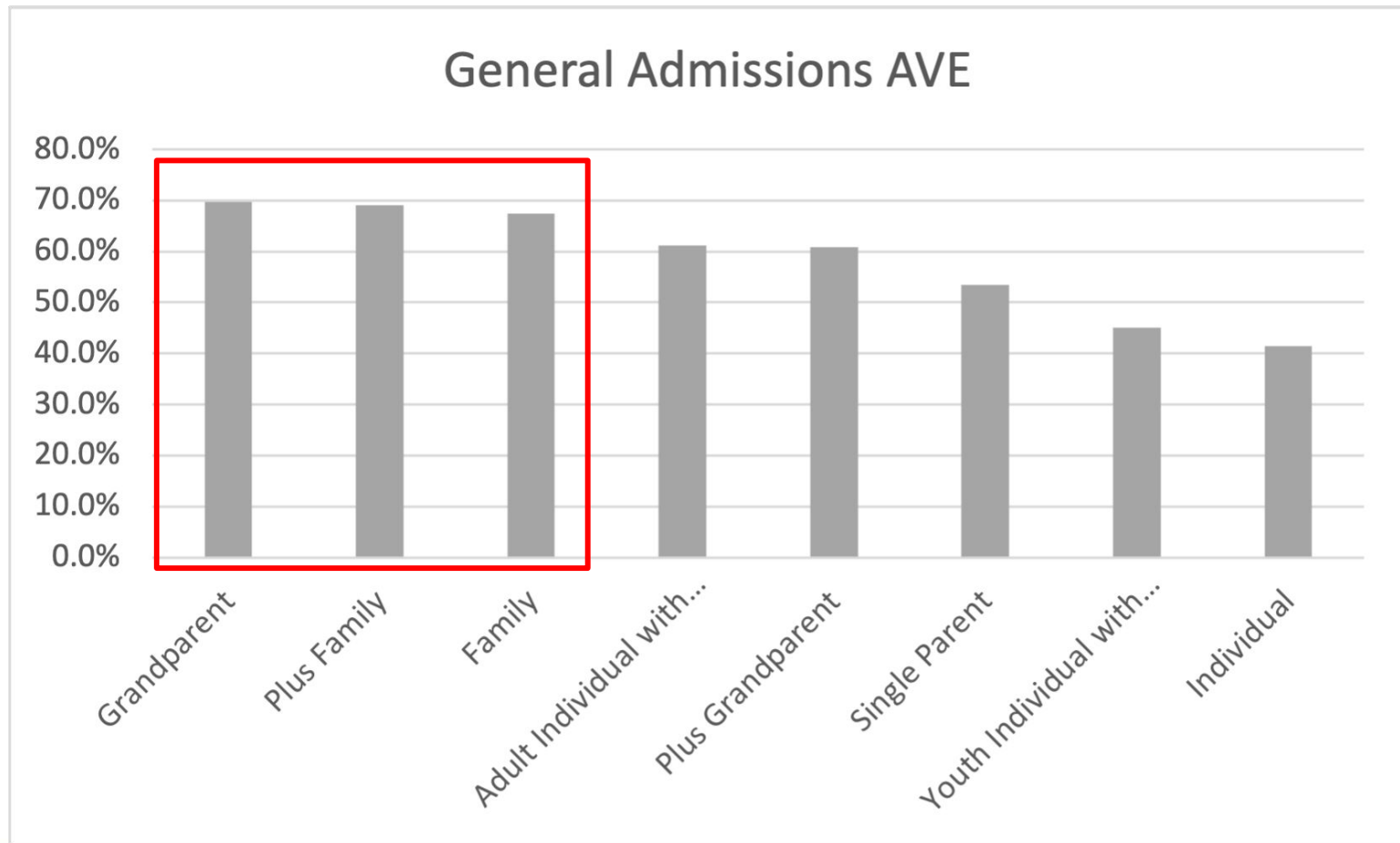
Science  
Museum  
+

Strasenburgh  
Planetarium  
+

Cumming  
Nature Center  
+

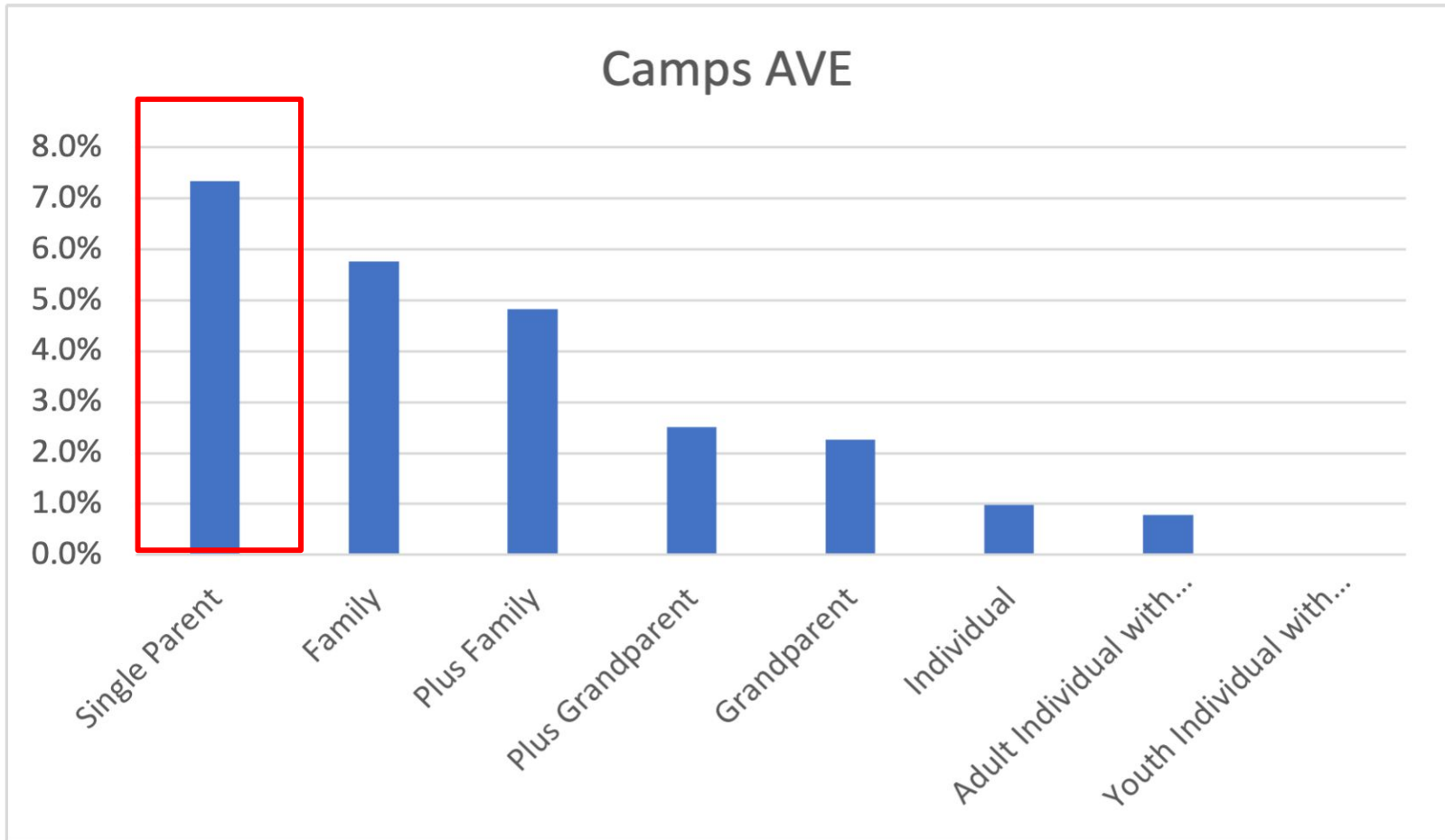
# Activity Participation Rate for Renew Members

## General Admissions



# Activity Participation Rate for Renew Members

## Camps



**RMSC**  
Rochester Museum & Science Center

reward  
your  
curiosity.

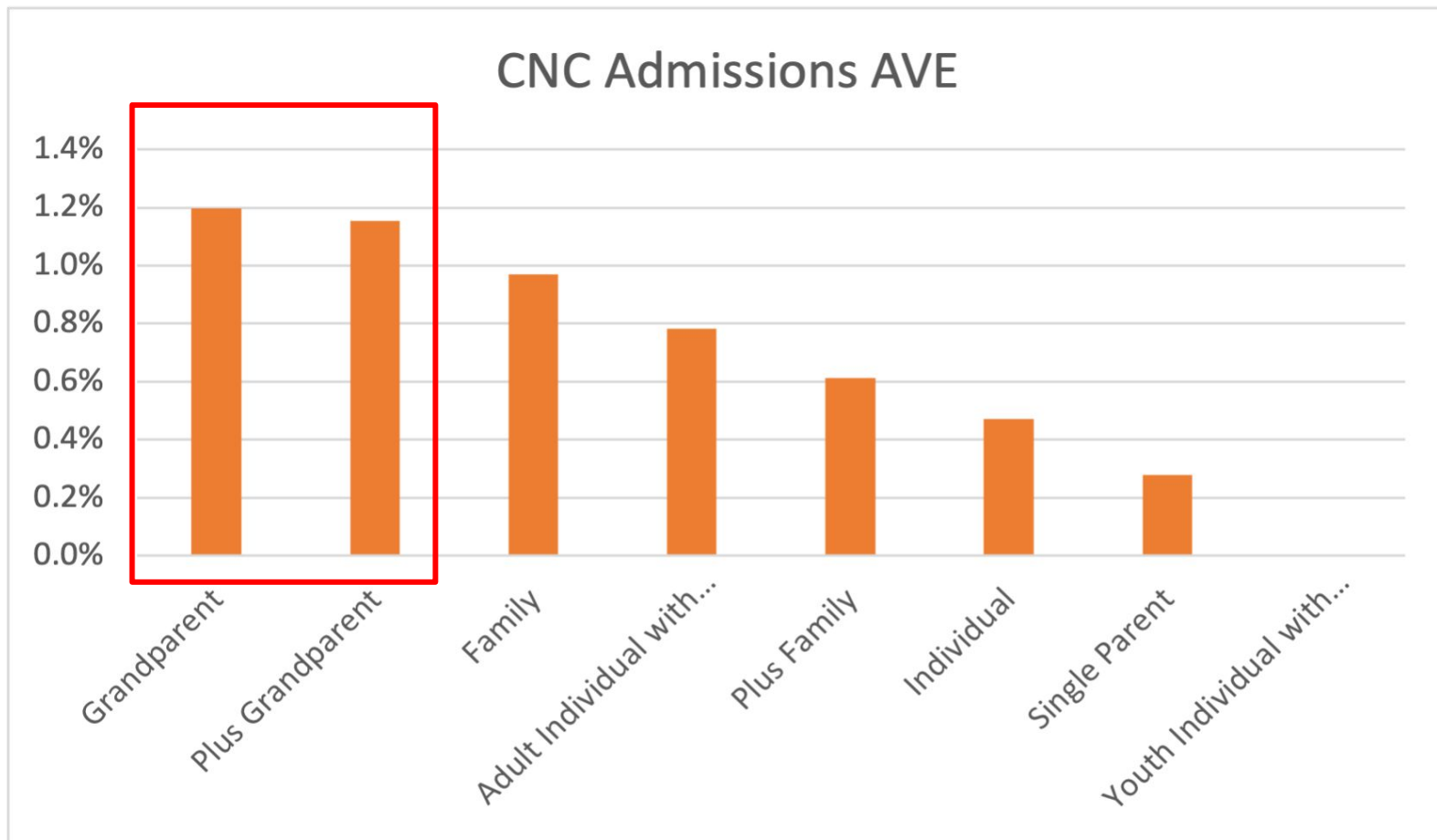
Science  
Museum  
+

Strasenburgh  
Planetarium  
+

Cumming  
Nature Center  
+

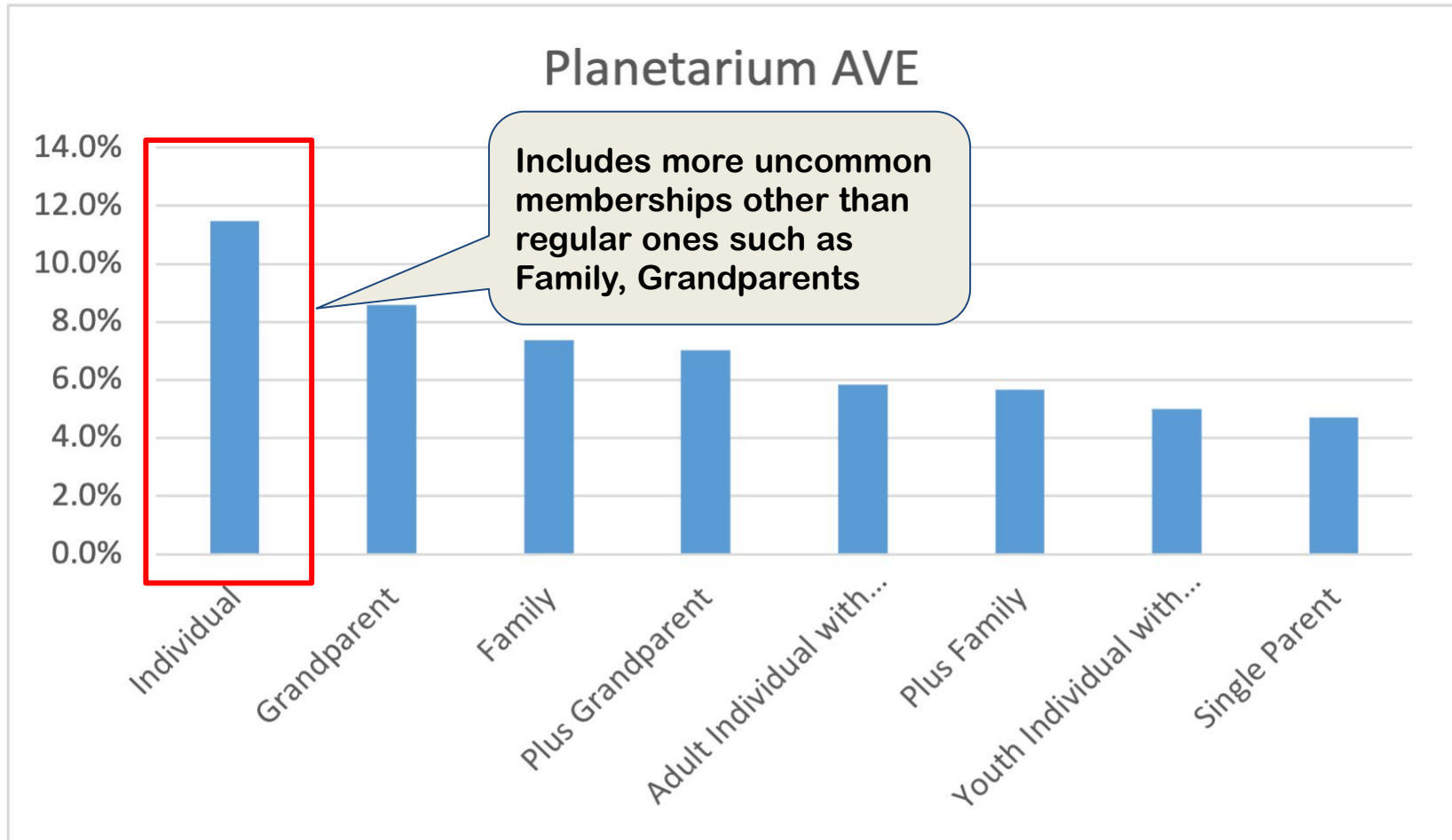
# Activity Participation Rate for Renew Members

## Cumming Nature Center Admissions

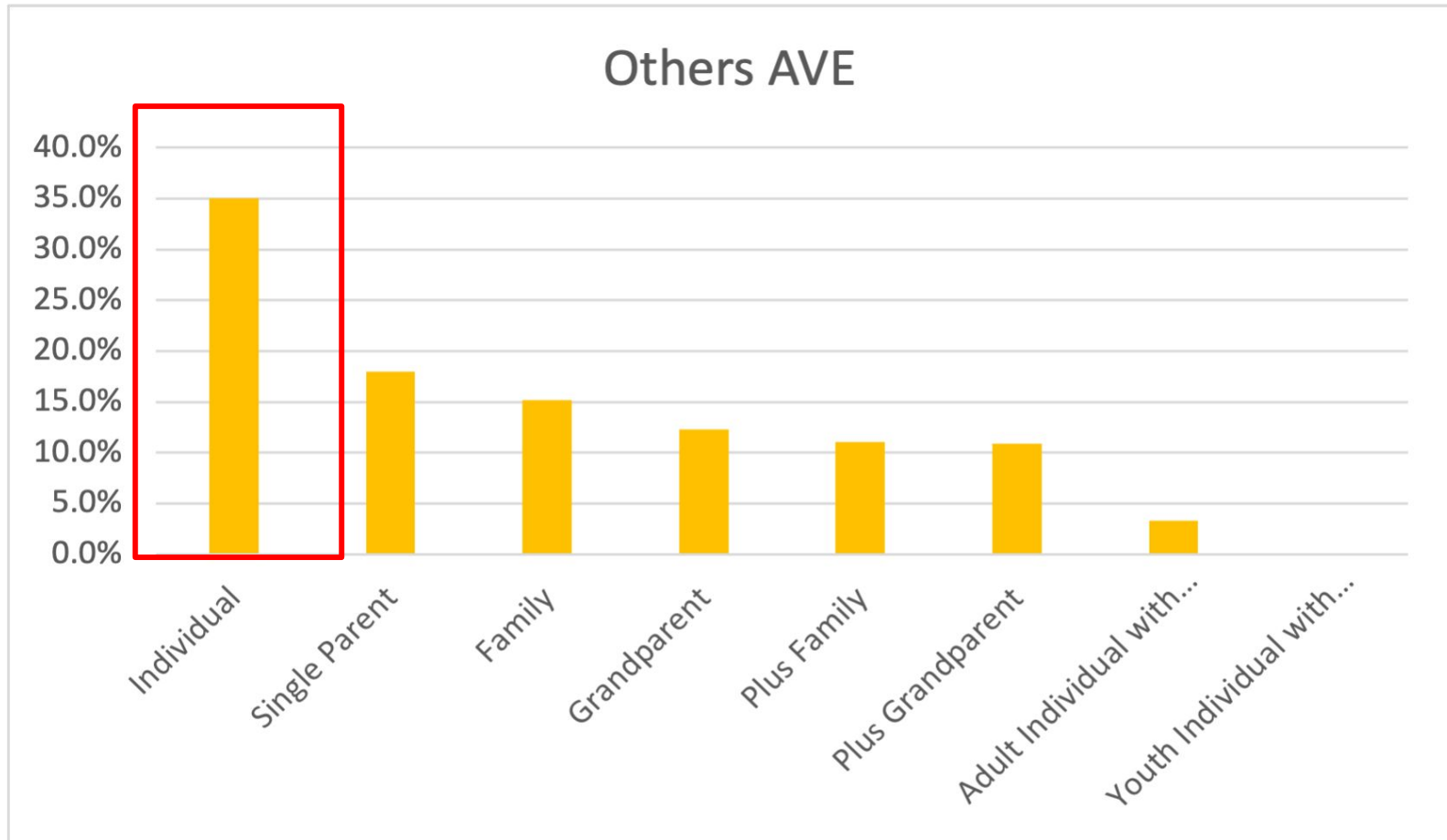


## Activity Participation Rate for Renew Members

# Planetarium



# Activity Participation Rate for Renew Members Others





# Conclusions on Complete Journey Table

- a. The renewal and join count both increase during 2015-2019. People are mostly like to join and renew in February and December, which are around the spring break and winter break.
- b. People would like to visit the museum most frequently in their first join and first renew.
- c. Members of different memberships have varying preferences when choosing activities to participate in
- d. RMSC could target specific membership level visitors to their most frequent program

# Methods: Predictive Models

# Predictive Modeling Process

## - Feature Selection: Logistic Regression

1. Performs higher accuracy than random forest
2. Large sample size, less than 5 classes of the predictor variable
3. Improves the precision and recall by 20% after feature reduction

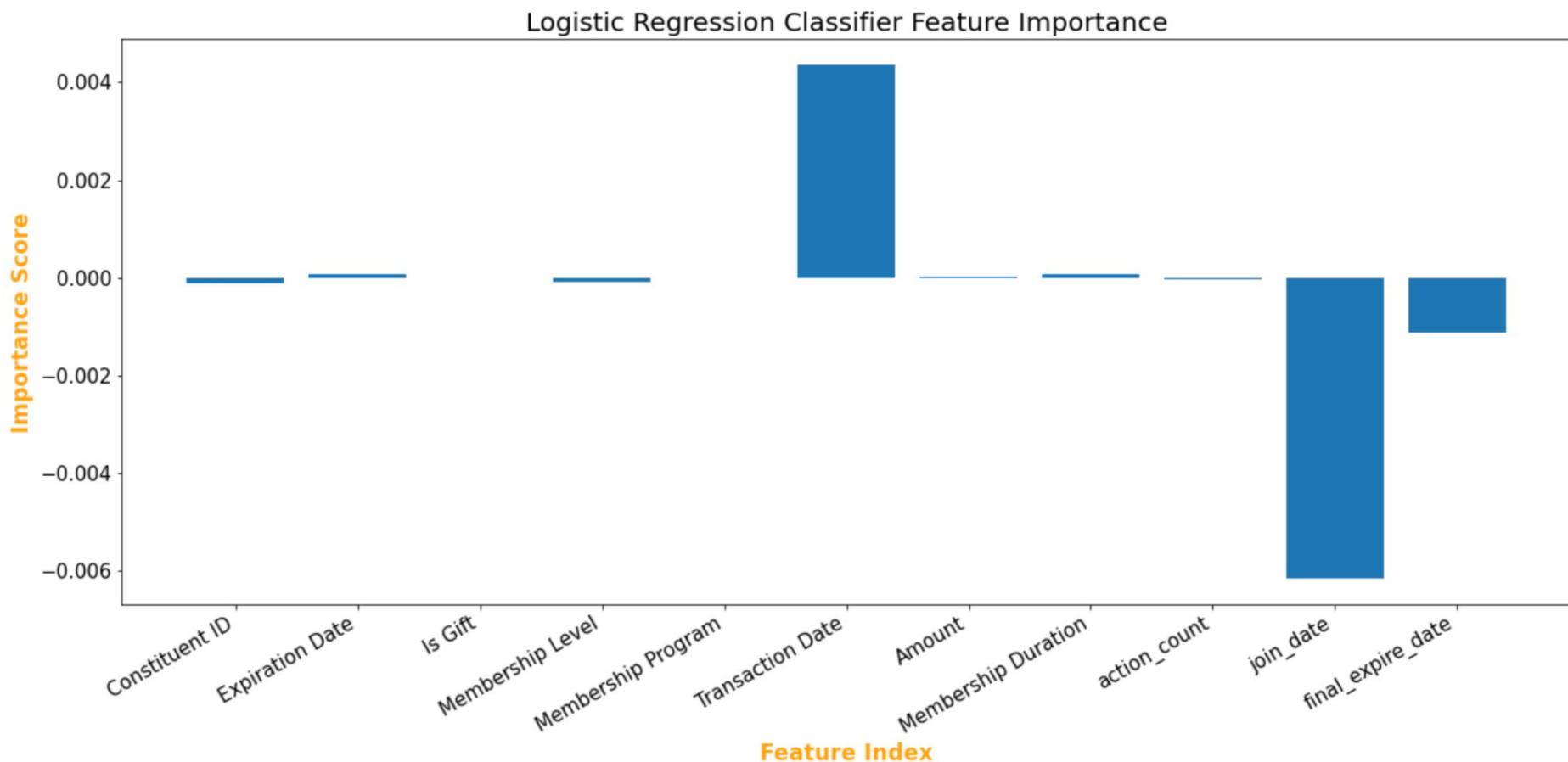
## - Models (Training: 70%, Testing: 30%):

1. Complete Journey Table: Multinomial, Binary LR Classifier
2. Complete Journey & Revenue: KNN Classifier

## - Results:

All classification models achieve accuracy higher than 85%, and we can apply them to determine/predict members' action.

# Feature Selection

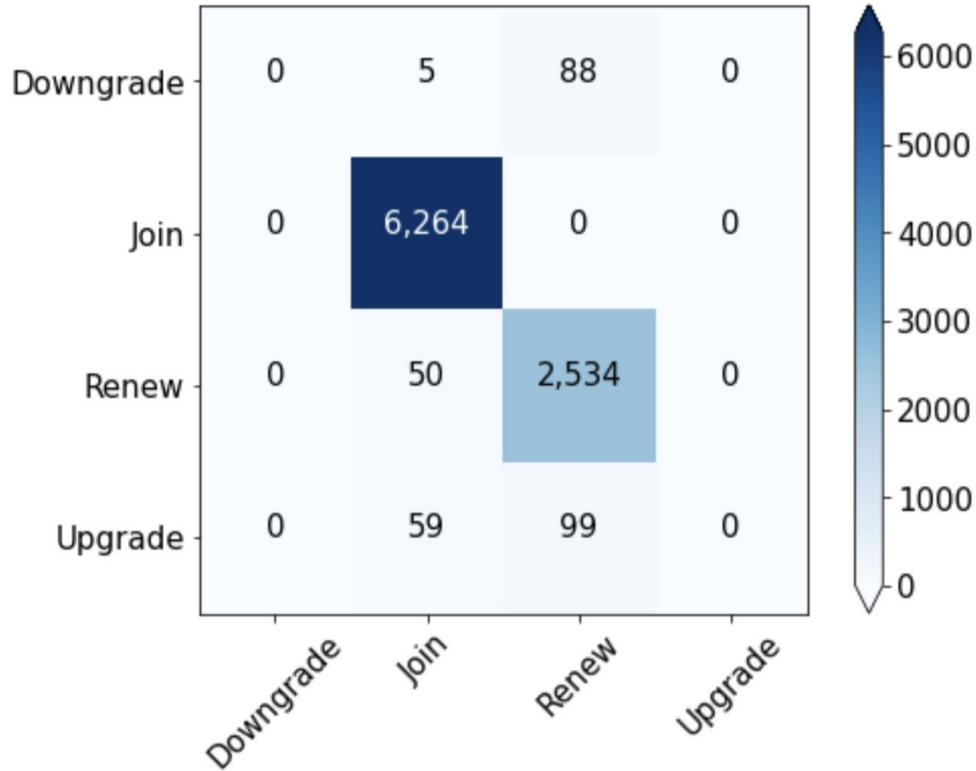


Top 4 important features: Constituent ID, Transaction Date, join\_date, final\_expire\_date

# A Multinomial Logistic Regression Classifier

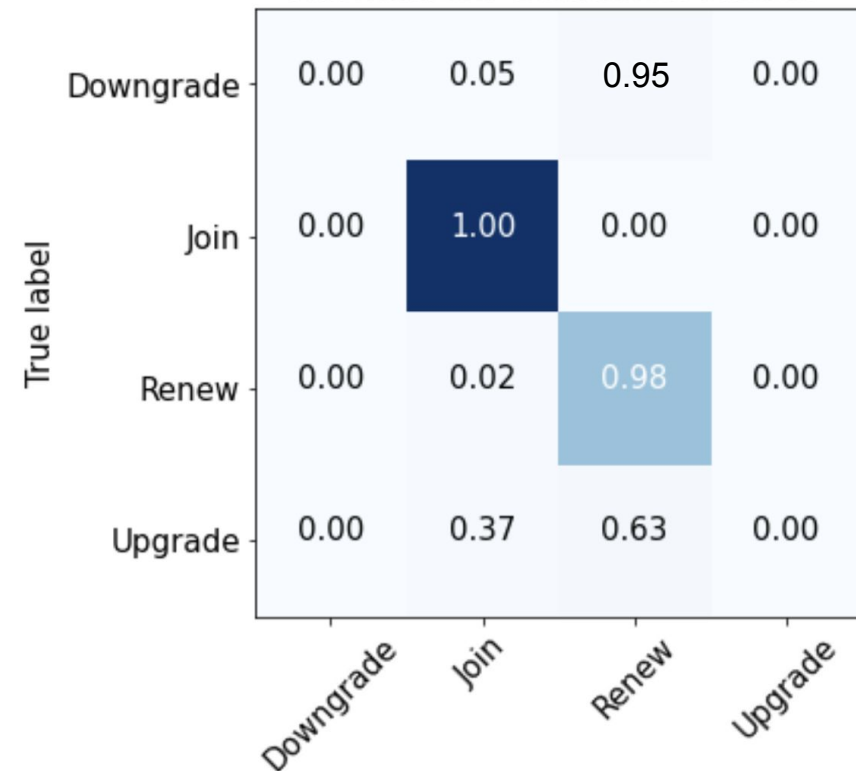
- Classes: Downgrade, Join, Renew, Upgrade

Confusion Matrix



Predicted label  
accuracy=0.97; misclass=0.03

Normalized Confusion Matrix



Predicted label  
accuracy=0.97; misclass=0.03

## Detailed Results of the Multinomial Logistic Regression Classifier

Training-set accuracy score: 96.75%  
 Test-set (Model) accuracy score: 96.69%  
 Classification Report:

	precision	recall	f1-score	support	5-fold Cross Validation:
0	0.00	0.00	0.00	93	Accuracy of Training: 0.97 +/- 0.00
1	0.98	1.00	0.99	6264	Accuracy of Testing: 0.97 +/- 0.01
2	0.93	0.98	0.96	2584	
3	0.00	0.00	0.00	158	
accuracy			0.97	9099	
macro avg	0.48	0.50	0.49	9099	
weighted avg	0.94	0.97	0.95	9099	

Testing class = Join, predicted class = Join, and probability = 97.16%

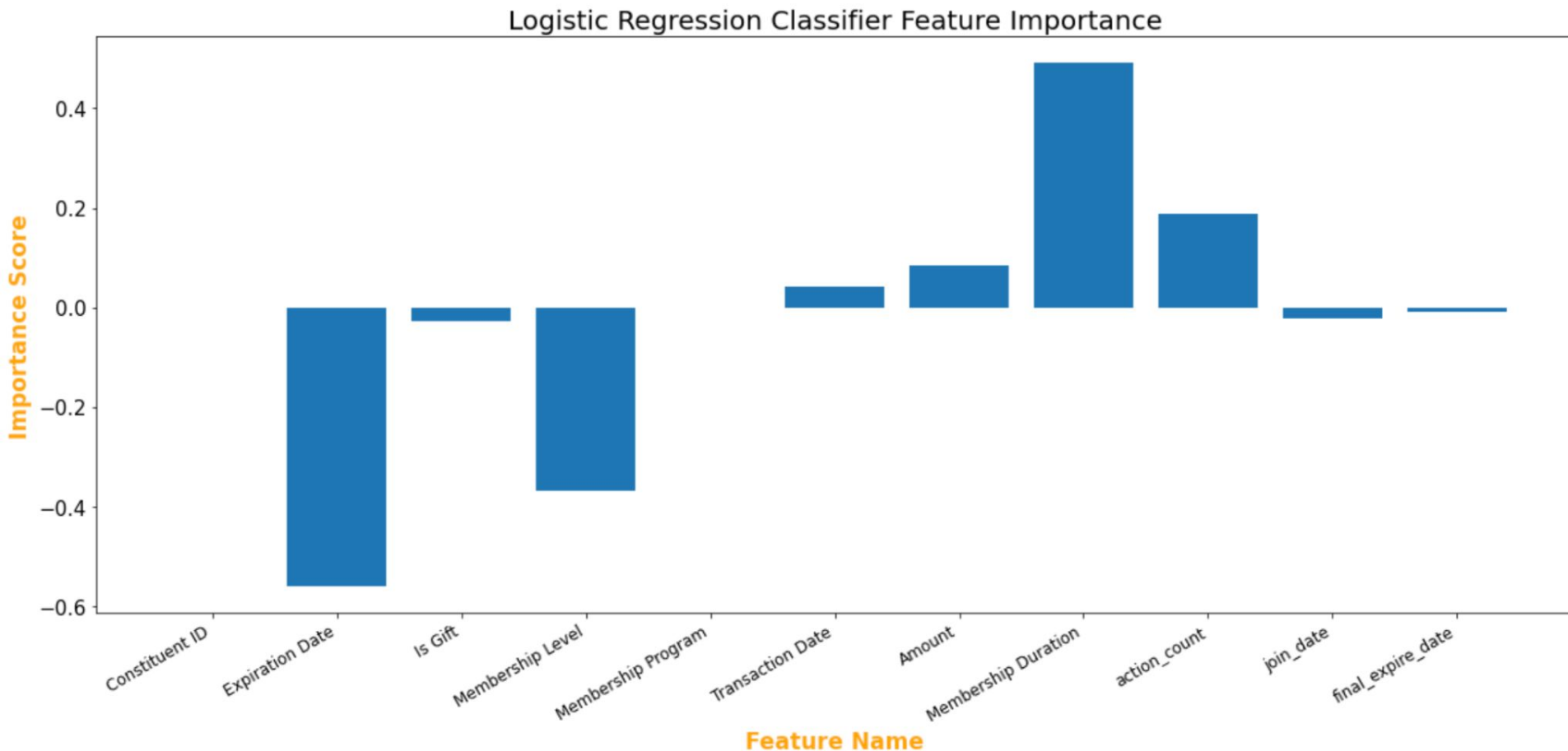
Testing class = Join, predicted class = Join, and probability = 99.69%

Testing class = Join, predicted class = Join, and probability = 68.82%

Testing class = Renew, predicted class = Renew, and probability = 78.98%

Testing class = Upgrade, predicted class = Join, and probability = 95.52%

# Feature Selection

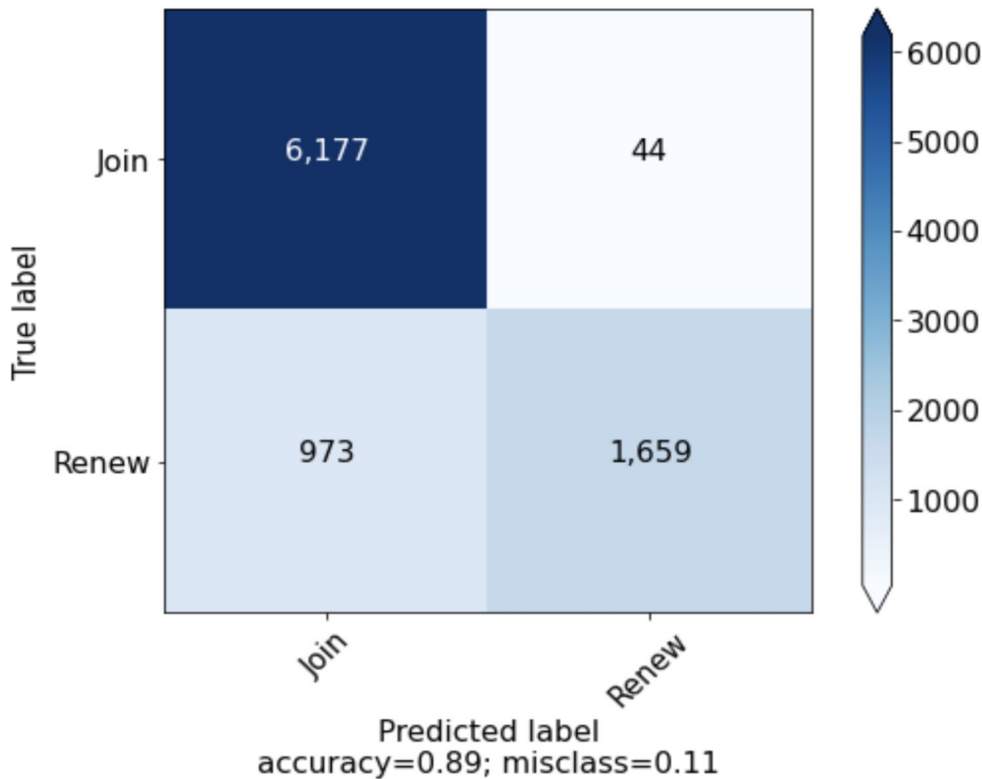


Top 4 important features: Membership Level, Expiration Date, Membership Duration, action\_count

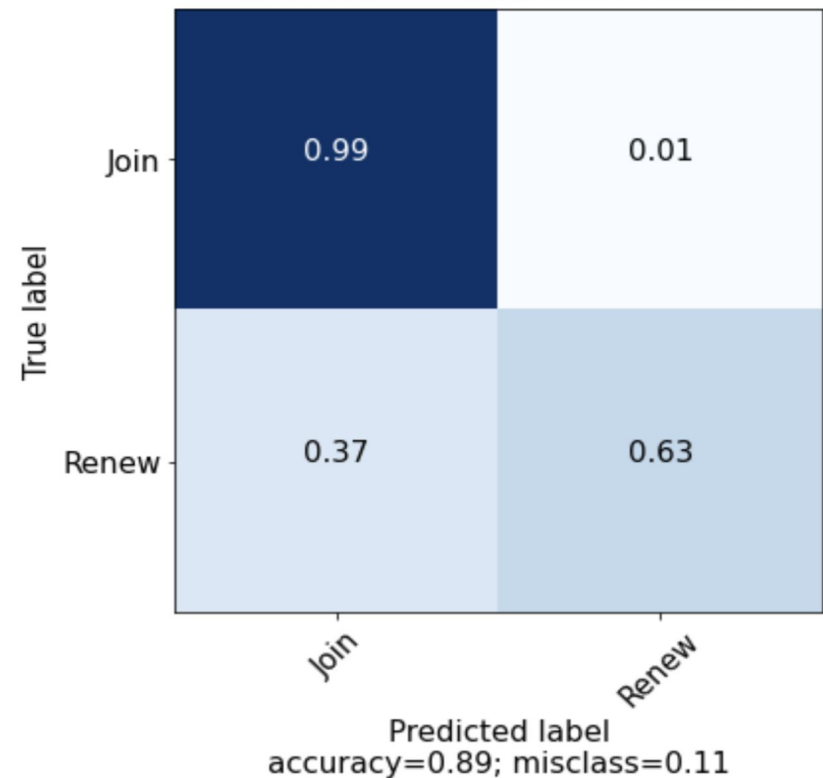
# A Binary Logistic Regression Classifier (Valerie)

- Classes: Join, Renew

Confusion Matrix



Normalized Confusion Matrix



**RMSC**  
Rochester Museum & Science Center

reward  
your  
curiosity.

Science  
Museum  
+

Strasenburgh  
Planetarium  
+

Cumming  
Nature Center  
+



## Detailed Results of the Binary Logistic Regression Classifier (Valerie)

Training-set accuracy score: 89.34%

Test-set (Model) accuracy score: 88.51%

Classification Report:

	precision	recall	f1-score	support	5-fold Cross Validation:
0	0.86	0.99	0.92	6221	Accuracy of Training: 0.89 +/- 0.00
1	0.97	0.63	0.77	2632	Accuracy of Testing: 0.89 +/- 0.01
accuracy			0.89	8853	
macro avg	0.92	0.81	0.84	8853	
weighted avg	0.90	0.89	0.88	8853	

Testing class = Join, predicted class = Join, and probability = 87.79%

Testing class = Join, predicted class = Join, and probability = 91.60%

Testing class = Renew, predicted class = Renew, and probability = 100.00%

Testing class = Join, predicted class = Join, and probability = 95.64%

Testing class = Join, predicted class = Join, and probability = 93.07%

## Complete Journey Table Merged with Revenue Table: K-Nearest Neighbors Classifier: Type of Action (Valerie)

### Data Preprocessing & Exploration:

- Deleted the classes with low occurrence (<0.01%), deleted the NaN rows
- The predictor variable (y) is highly skewed
- Class Distribution of the predictor variable:

1st Join 63.8%, 1st Renew 21.5%, 2nd Renew 8.2%, 3rd Renew 2.8%, 1st Upgrade 1.5%, 1st Downgrade 1.4%, 4th Renew 0.8%

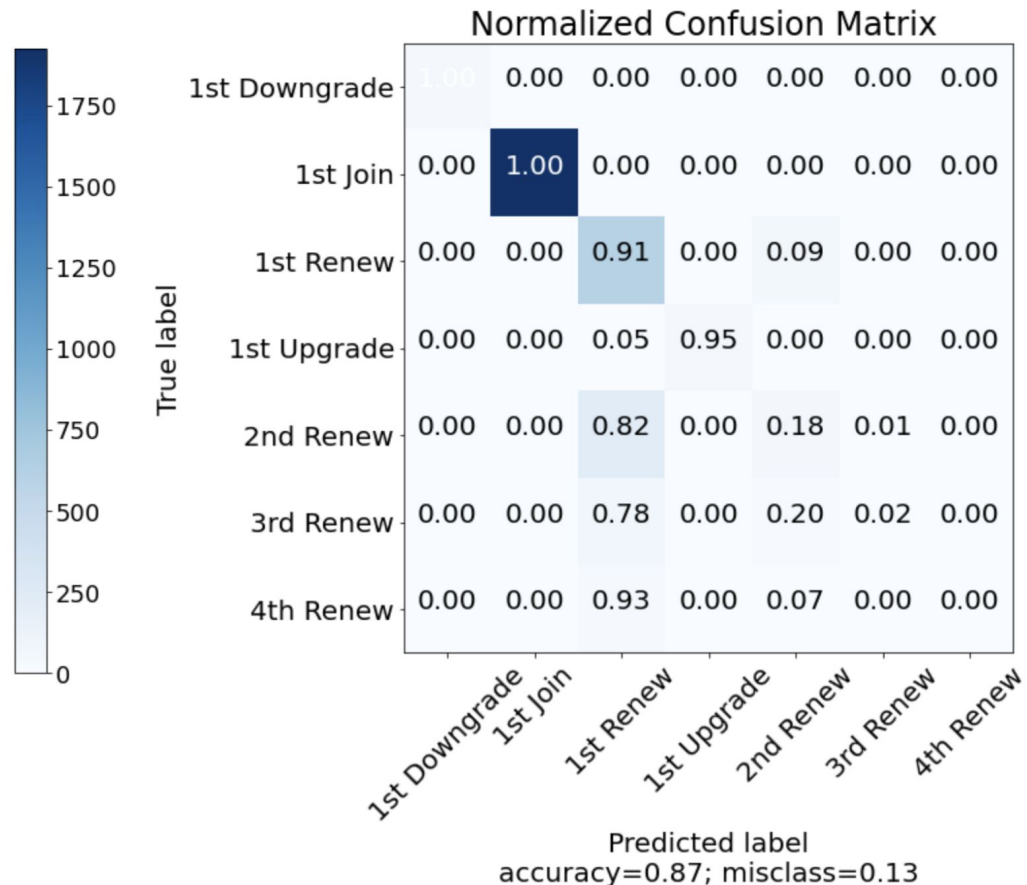
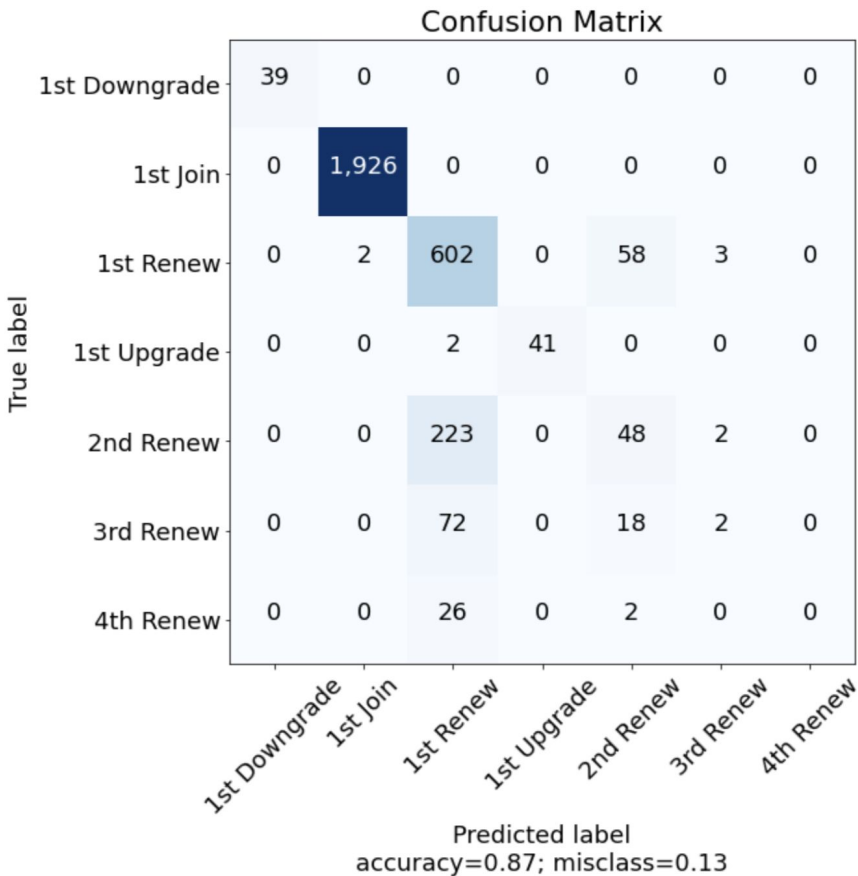
# K-Nearest Neighbors Classifier: Type of Action (Valerie)



**Top 3 important features: Action, Membership Level, Amount**

# K-Nearest Neighbors Classifier: Type of Action (Valerie)

- Classes: 1st Join, 1st Renew, 2nd Renew, 3rd Renew, 1st Upgrade, 1st Downgrade, 4th Renew



# Detailed Results of K-Nearest Neighbors Classifier: Type of Action (Valerie)

Training-set accuracy score: 87.82%

Test-set (Model) accuracy score: 86.69%

Classification Report:

	precision	recall	f1-score	support
0	1.00	1.00	1.00	39
1	1.00	1.00	1.00	1926
2	0.65	0.91	0.76	665
3	1.00	0.95	0.98	43
4	0.38	0.18	0.24	273
5	0.29	0.02	0.04	92
6	0.00	0.00	0.00	28
accuracy			0.87	3066
macro avg	0.62	0.58	0.57	3066
weighted avg	0.84	0.87	0.84	3066

5-fold Cross Validation:

Accuracy of Training: 0.86 +/- 0.01

Accuracy of Testing: 0.85 +/- 0.01

Testing class = 1st Join, predicted class = 1st Join, and probability = 100.00%

Testing class = 1st Join, predicted class = 1st Join, and probability = 100.00%

Testing class = 1st Renew, predicted class = 1st Renew, and probability = 66.67%

Testing class = 1st Join, predicted class = 1st Join, and probability = 100.00%

Testing class = 1st Join, predicted class = 1st Join, and probability = 100.00%

# Time Series and Predictive Modeling

## Calculations of Total Active/New Join/Returning Members (Ruiyu)

**join\_date**: date of members' initial join

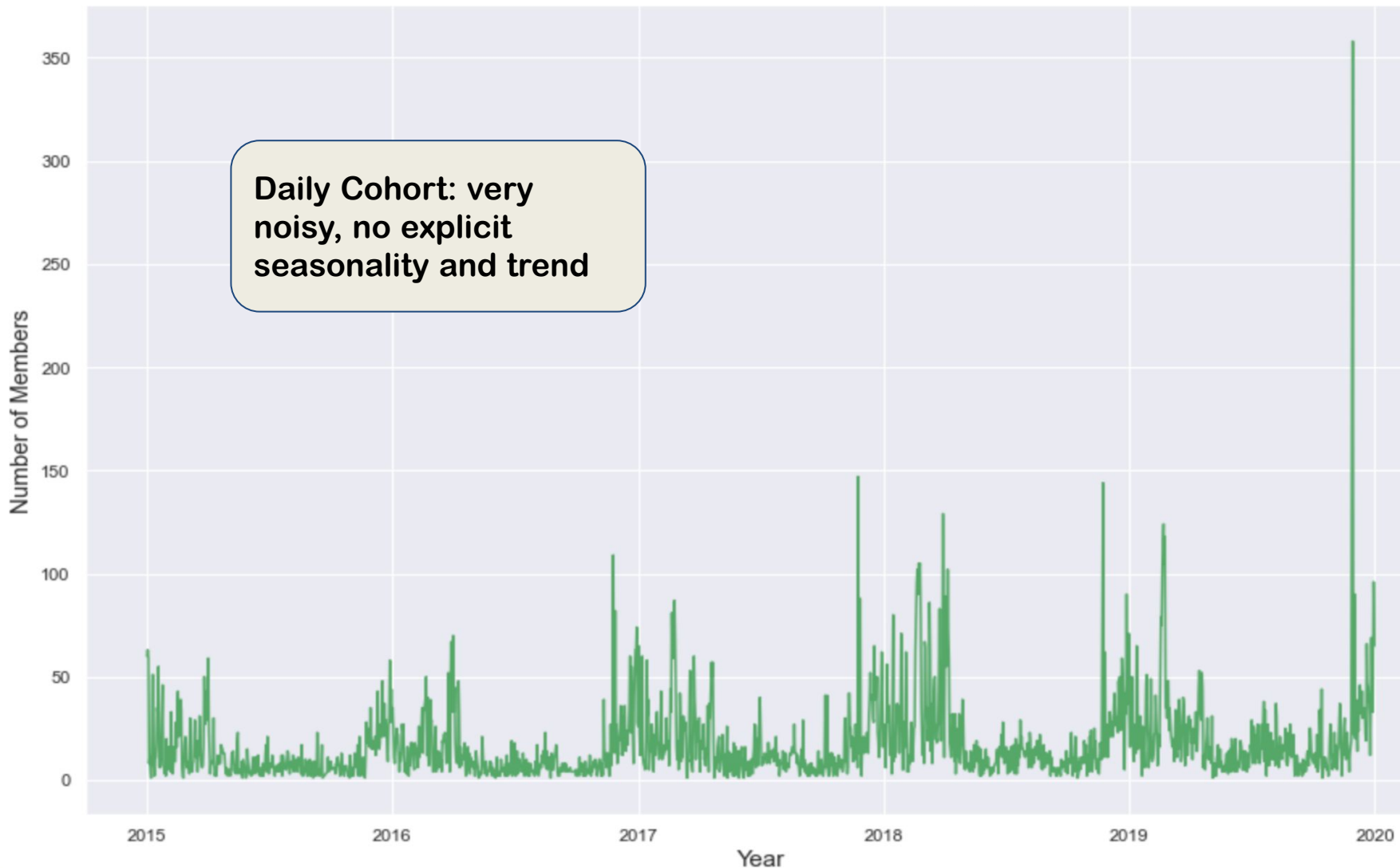
**final\_expiration\_date**: the expiration date of the last transaction made by each member

**Cohort Period/Date**: daily, weekly, monthly, e.g. 01/01/2015  $\Rightarrow$  monthly 01/2015

**Total Active Members**: during join\_date and final\_expiration\_date

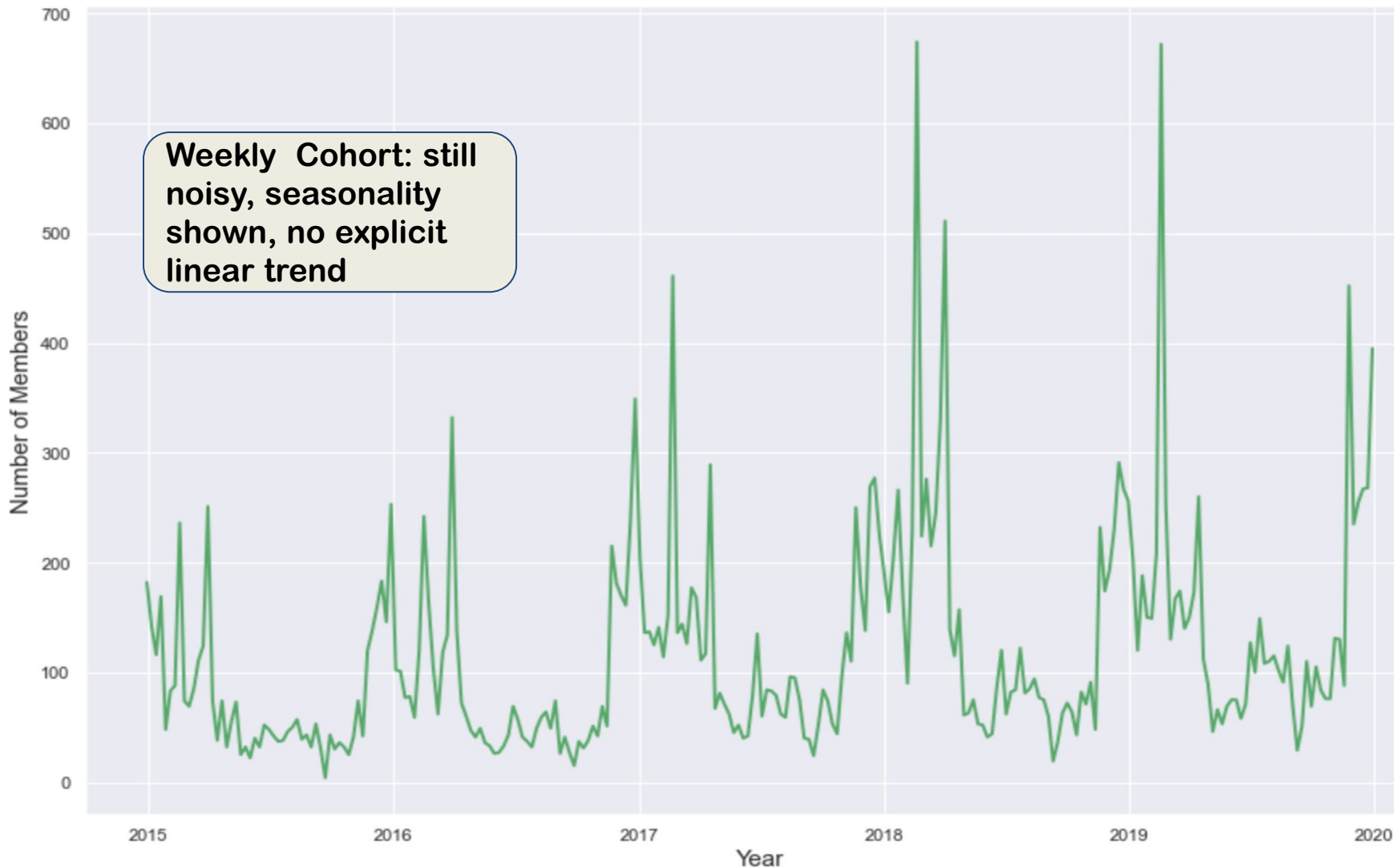
**New Join Members**: initial join

**Returning Members**: have transactions after initial join

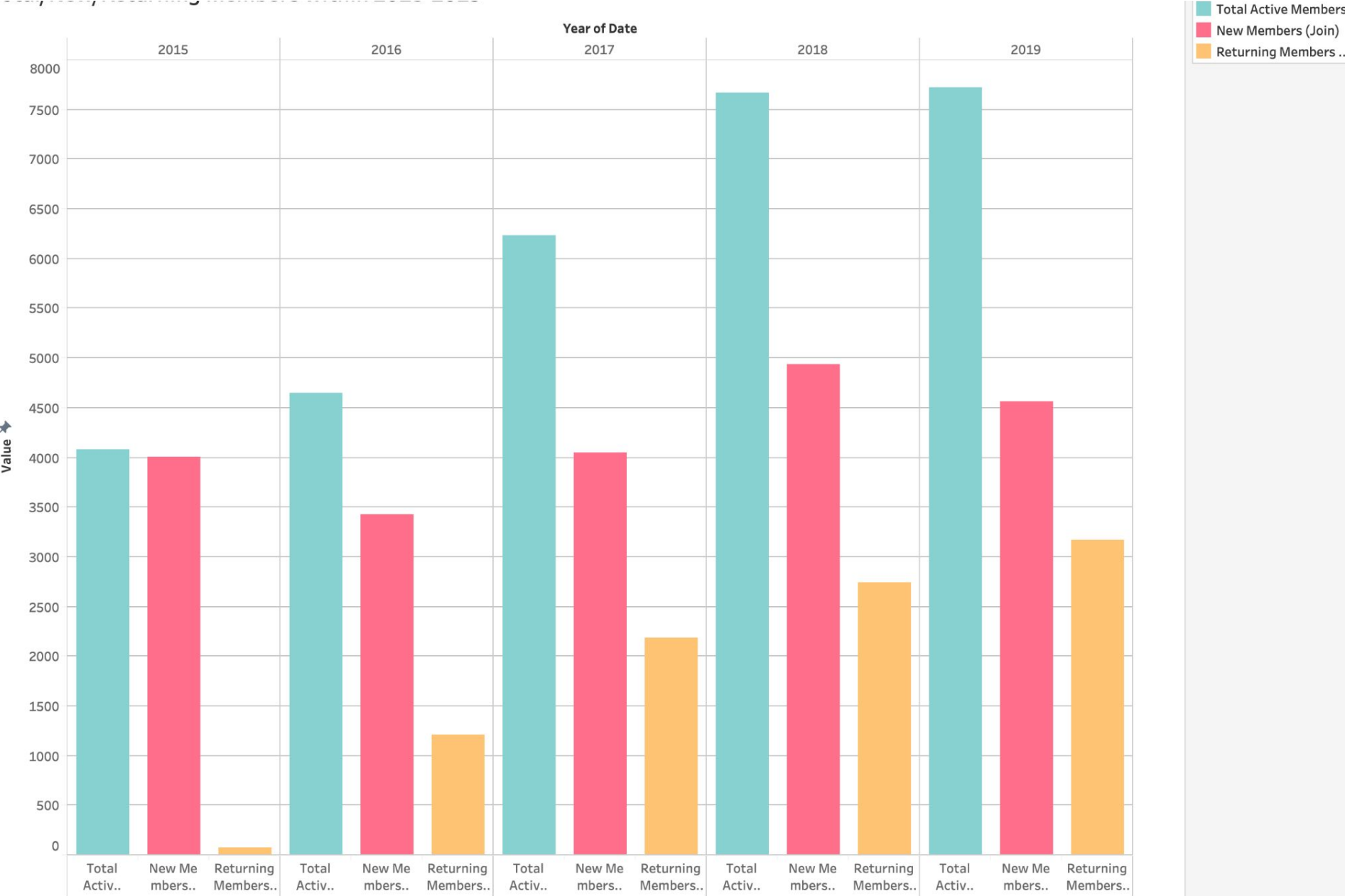




# Cohort Analysis: The Number of Total Active Members v.s. Date

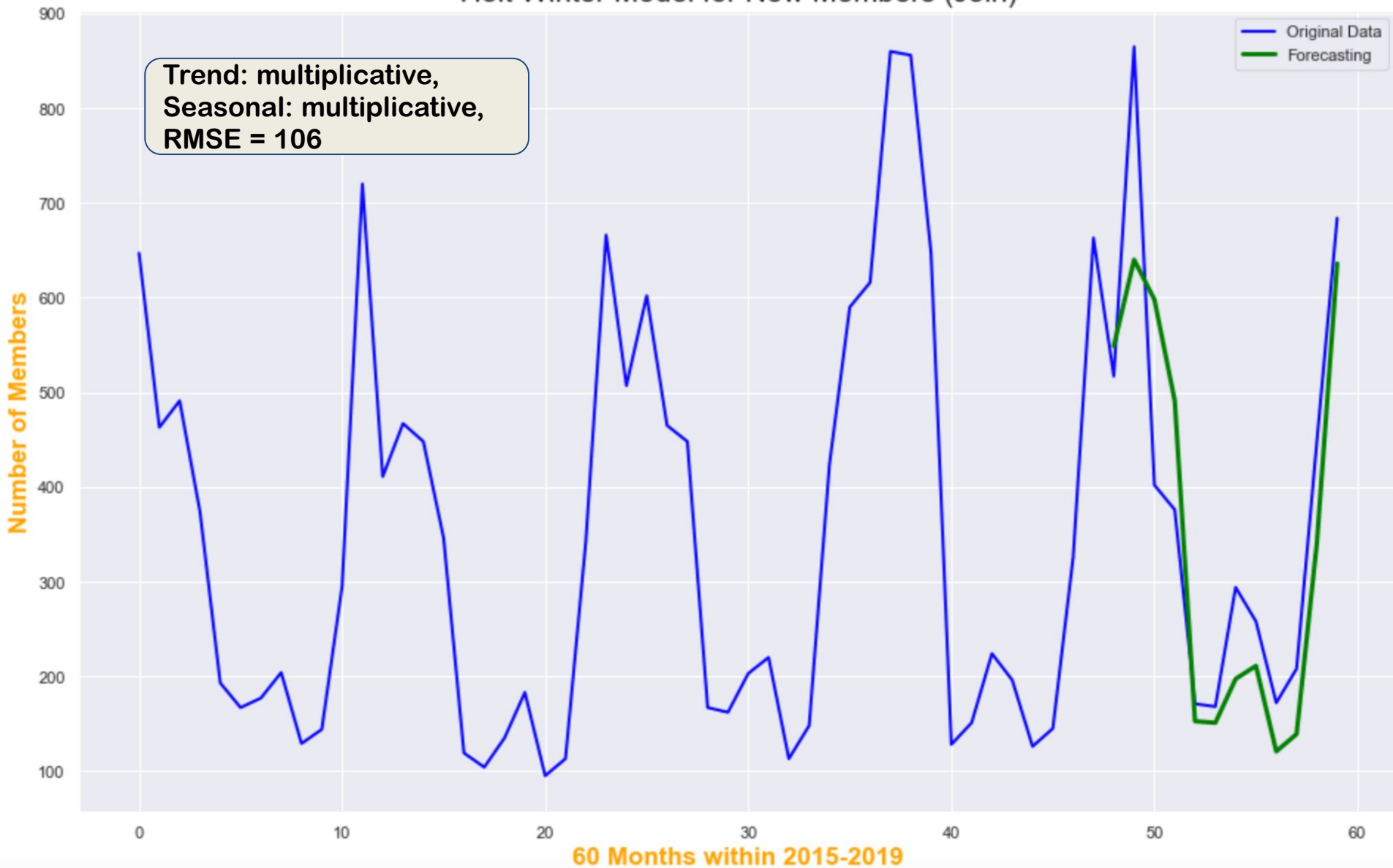


# 1. Total/New/Returning Members within 2015-2019 (Ruiyu)



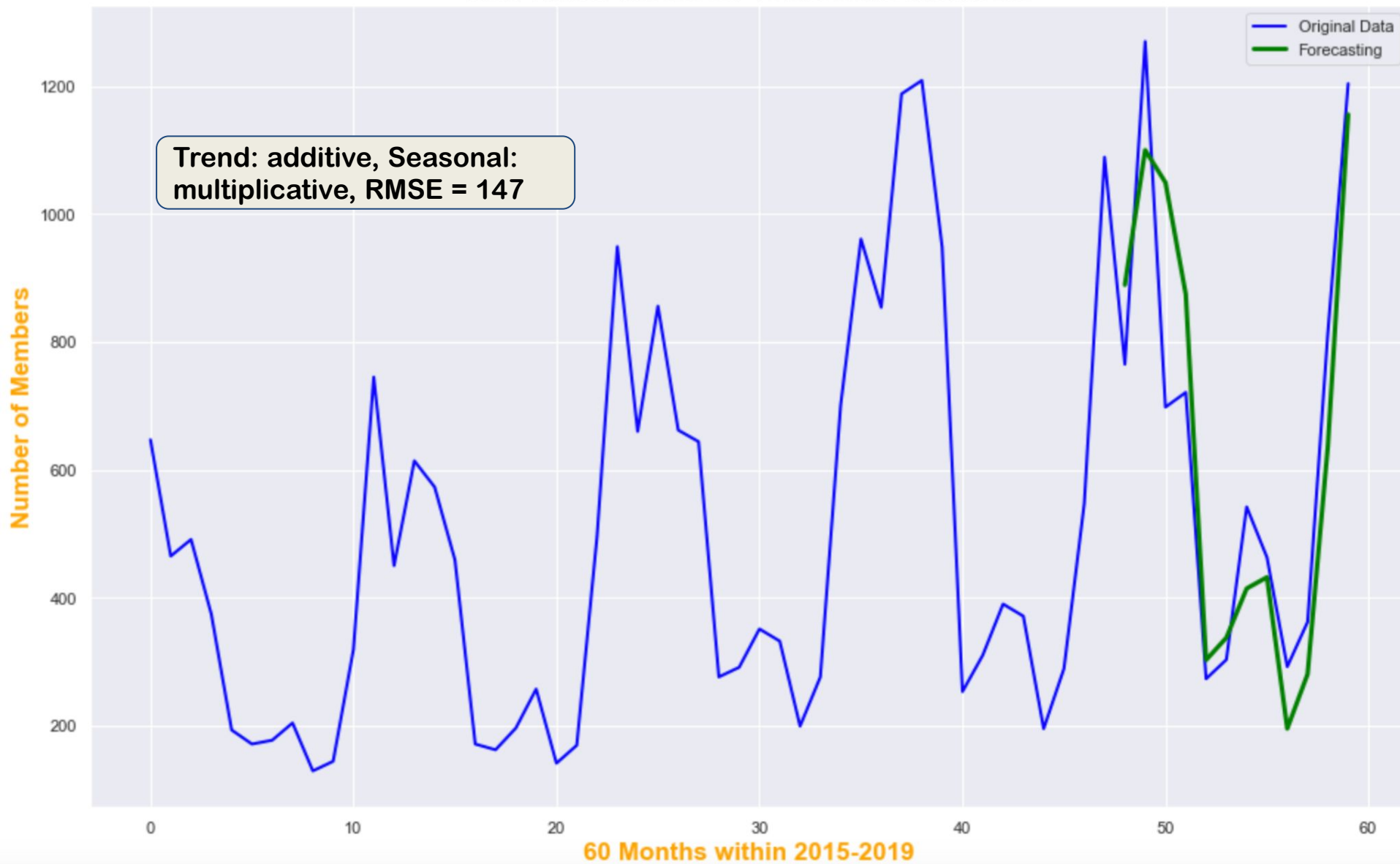
## 2. Complete Journey Table: New, Returning, and Total Active members

Holt Winter Model for New Members (Join)



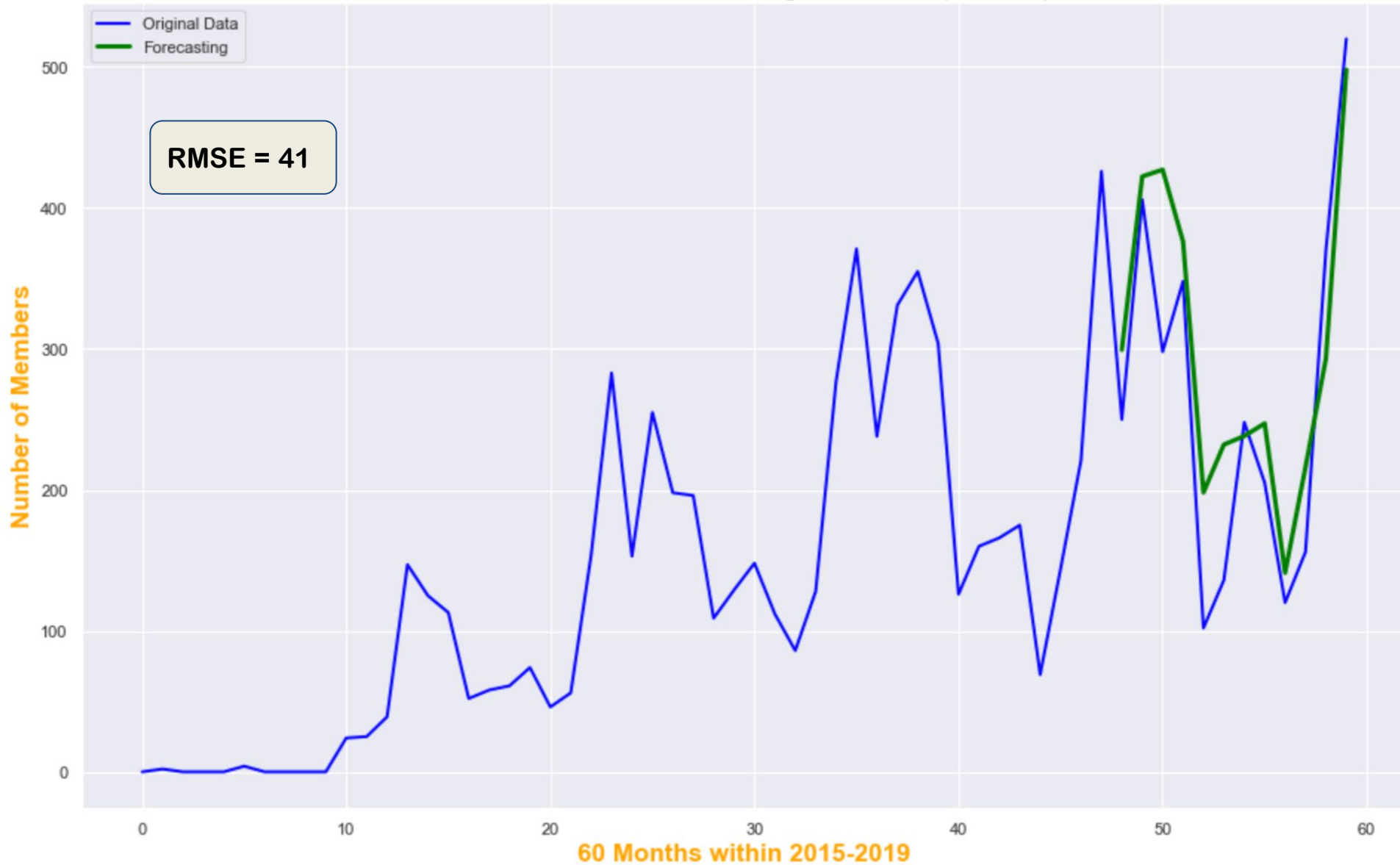
### 3. Complete Journey Table: New, Returning, and Total Active members

Holt Winter Model for Total Active Members



## 4. Complete Journey Table: New, Returning, and Total Active members

ARIMA Model for Returning Members (Renew)



# Donation Table

# Summary

- a. **Data Description**
- b. **Data Analysis**
  - i. **Number of donation records for each action**
  - ii. **Number of Donation Records for Join/ Renew Action Breakdown**
  - iii. **Average Donation Amount for Each Action**
  - iv. **Average Donation Amount for Each Membership Level**
- c. **Conclusions**

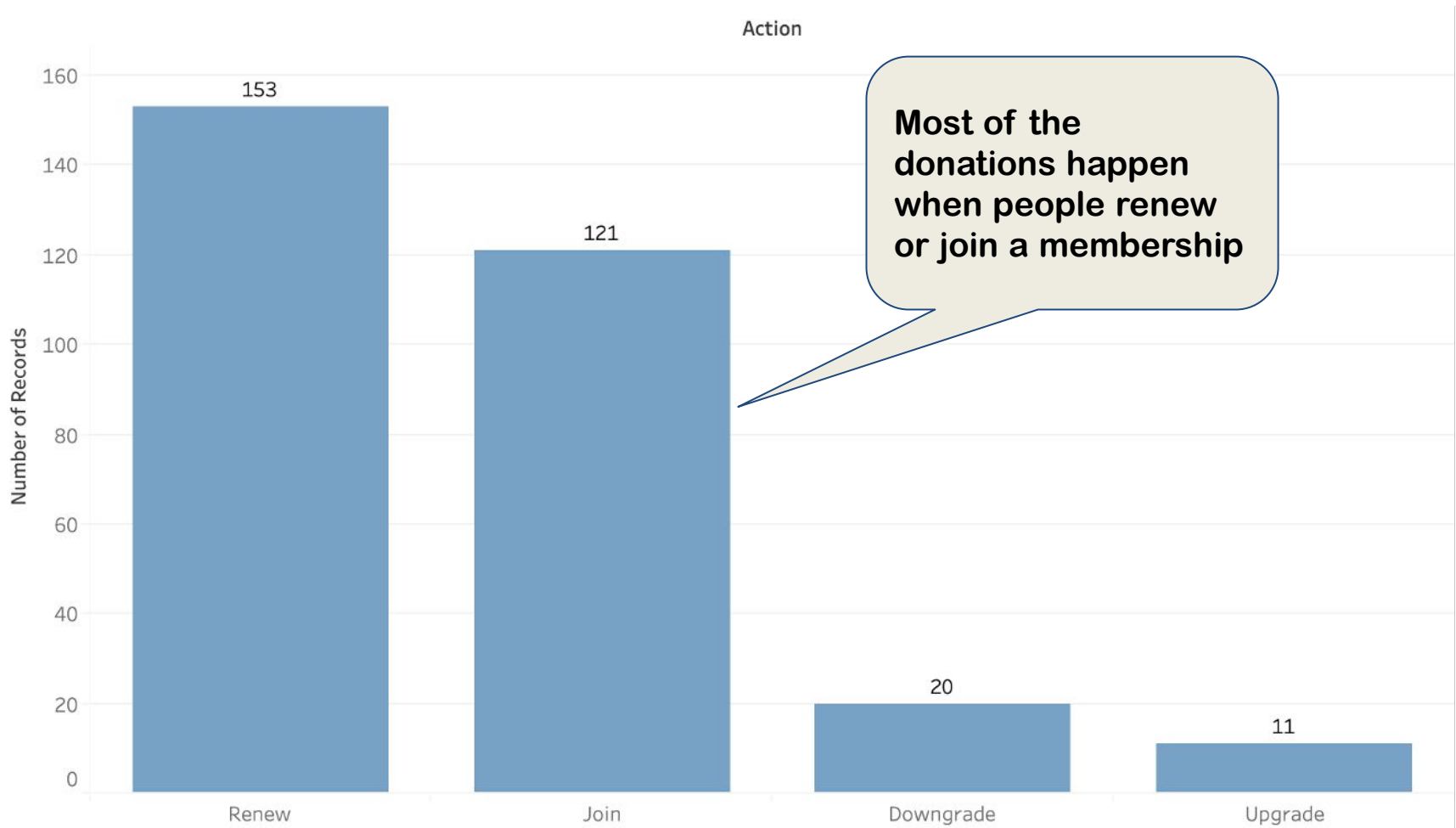
# Data Description

- Original data contains 10k records and 3 features
- Combined with complete journey table to obtain information about members who donated
- Resulted in around 300 records

Constituent ID	Member number
Amount	Amount of Donation
Date	Date of Donation



# 1. Number of Donation Records for Each Action (Shawn)



**RMSC**  
Rochester Museum & Science Center

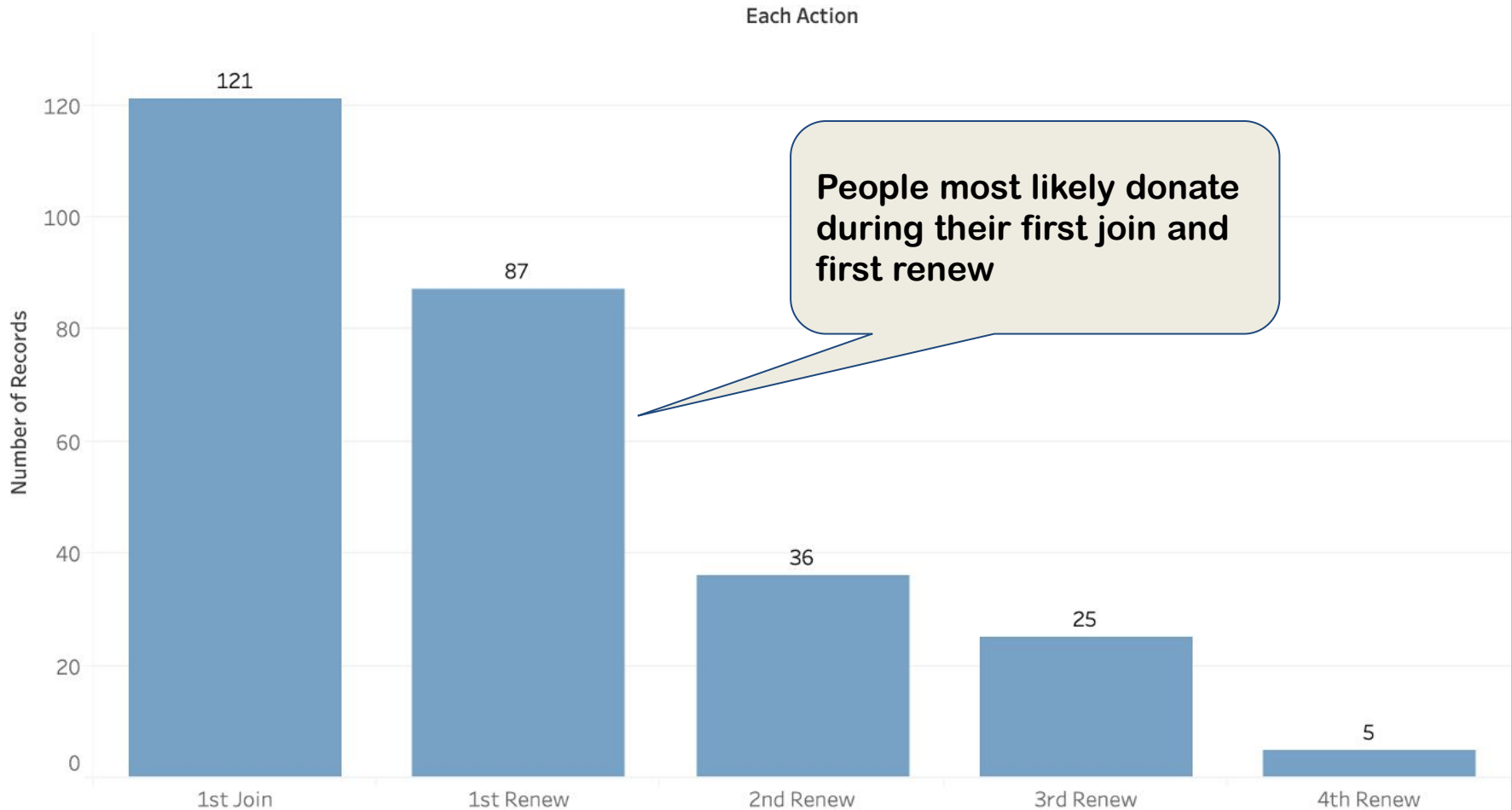
reward  
your  
curiosity.

Science  
Museum  
+

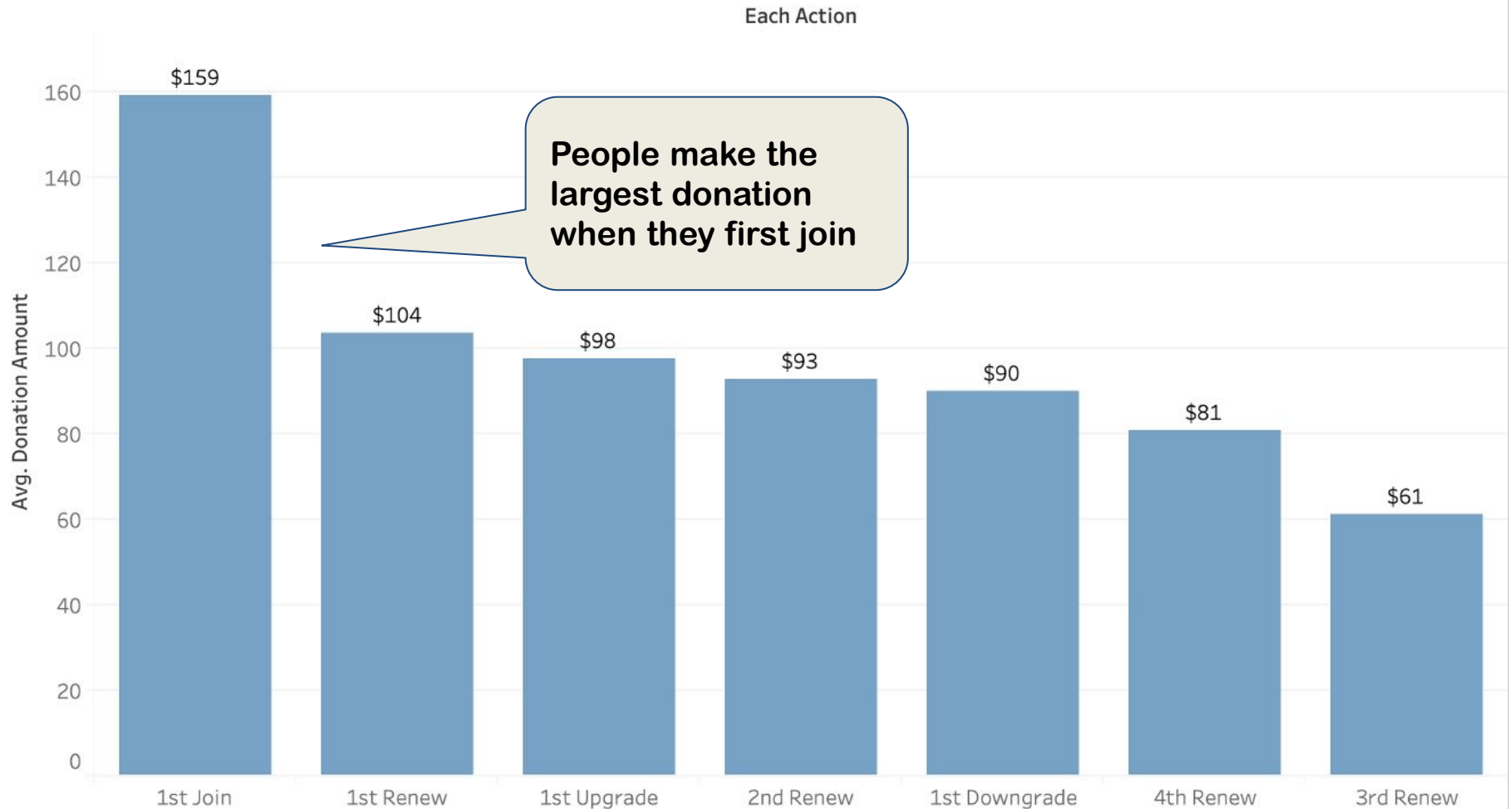
Strasenburgh  
Planetarium  
+

Cumming  
Nature Center  
+

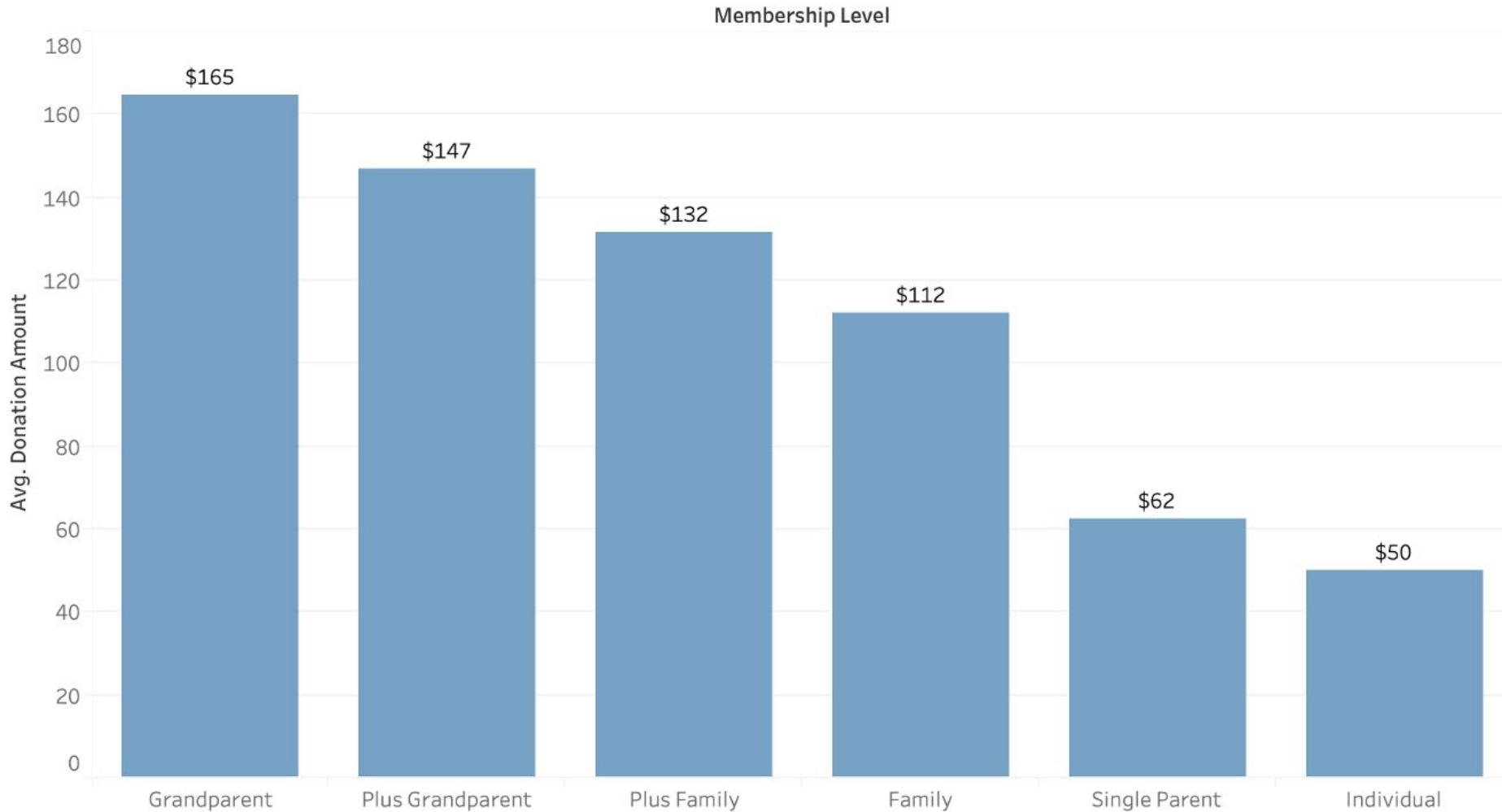
## 2. Number of Donation Records for Join/ Renew Action Breakdown (Shawn)



### 3. Average Donation Amount for Each Action (Shawn)



# 4. Average Donation Amount for Each Membership Level (Shawn)



# Conclusions on Donation Table

- People usually make donations at the beginning of their membership journey
- Member donation amount goes down as time goes on
- The average donation amount of grandparents is the highest among all membership levels

# Section 4: Project Management

# Challenges

1. We have very limited donation records (~300 records), so we went for visualization graphs instead of predictive modeling
2. Donation cannot be well predicted by a linear model
3. There is significant class imbalance for people who donated vs. those who didn't

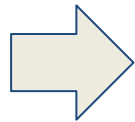
# Milestones Achieved (after midterm)

Date	Milestone	Dataset	Description
3/30	Milestone 7	[Earned Revenue Table]	Updated earned revenue analysis based on feedback
3/30	Milestone 8	[Membership Transaction Table]	Reduced predictive error and improved member renewal model
4/02	Milestone 9	[Complete Journey Table]	Conducted analysis on relationship between member actions and visits
4/09	Milestone 10	[Complete Journey Table]	Updated analysis on membership levels and their popular programs
4/23	Milestone 11	[Donation Table]	Conducted visualization analysis on member information



# Next Steps

Date	Description
~ 5/5	To present to the RMSC board committee
~ 5/11	To finish and submit final report
...	...



**Expected Project Finish Time (currently on time)**

**5/11**

# Questions?



**RMSC**  
Rochester Museum & Science Center

reward  
your  
curiosity.

Science  
Museum  
+

Strasenburgh  
Planetarium  
+

Cumming  
Nature Center  
+