**Final Project**

**1. Data Warehouse**

**Special events table**

USE club;

DROP VIEW IF EXISTS temp ;

CREATE VIEW temp AS

(SELECT substr(Invoice\_Account, 5) AS Member\_Number, Event

FROM special);

DROP TABLE IF EXISTS specialevents;

CREATE TABLE specialevents AS

(SELECT

Member\_Number,

COUNT(CASE WHEN Event = "Easter Brunch" THEN Member\_Number END) AS Easter\_Brunch,

COUNT(CASE WHEN Event = "4th of July" THEN Member\_Number END) AS Fourth\_of\_July,

COUNT(CASE WHEN Event = "Thanksgiving" THEN Member\_Number END) AS Thanksgiving,

COUNT(CASE WHEN Event = "Private Function" THEN Member\_Number END) AS Private\_Function

FROM

temp

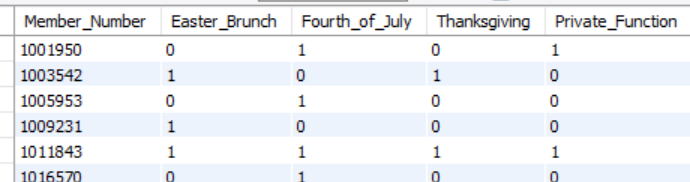
GROUP BY

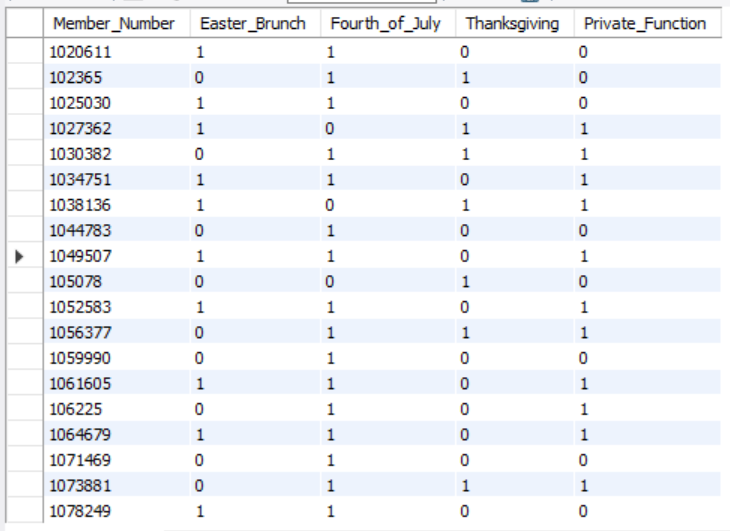
Member\_Number

ORDER BY

Member\_Number);

**Output :**





**Membership info & promotion**

drop table if exists MemberPromo;

Create table MemberPromo as

select mi.Member\_Number, Membership\_Type, Year\_Joined, Number\_of\_Children, Partner, ifnull(p1.promo1,0) promotion1, ifnull(p2.promo2,0) promotion2

from memberships mi

left join

(select Member\_Number, 1 as promo1

from promoone) as p1

on mi.Member\_Number = p1.Member\_Number

left join

(select Member\_Number, 1 as promo2

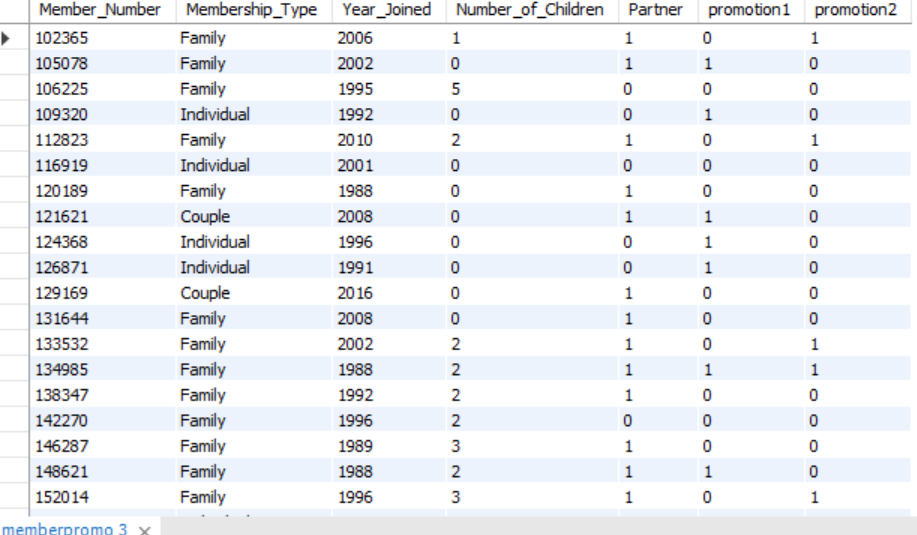
from promotwo) as p2

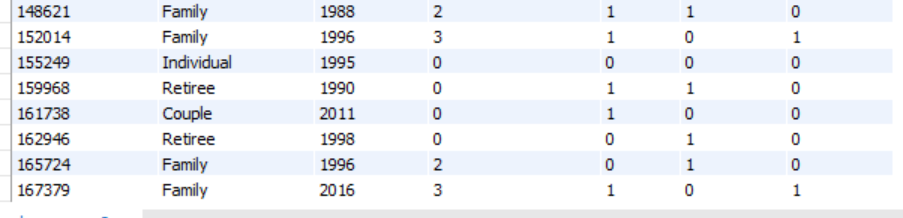
on mi.Member\_Number = p2.Member\_Number left join (

select Member\_Number,sum(Relationship\_to\_Member="Child") as Number\_of\_Children,sum(Relationship\_to\_Member="Partner") as Partner from members group by Member\_number) as j on mi.Member\_Number =j.Member\_Number ;

**------------------------------------------------------------------------------------------------------------------------**

**Output :**





**Frequency**

drop table if exists Frequency;

create table Frequency as (

select df.Member\_Number, COALESCE(dining\_Fre,0) as dining, COALESCE(golf\_Fre,0) as golf, COALESCE(pool\_Fre,0) as pool, COALESCE(tennis\_Fre,0) as tennis, COALESCE(other\_Fre,0) as other

from memberships as df left join

(select Member\_Number, count(\*) as dining\_Fre

from(select Member\_Number,count(\*) as dining\_Fre

from dining

group by Member\_Number, Date) as a group by Member\_Number ) as dine on df.Member\_Number = dine.Member\_number

left join

(select Member\_Number, count(\*) as golf\_Fre

from(select Member\_Number,count(\*) as golf\_Fre

from golf

group by Member\_Number, Date) as aa group by Member\_Number )as gf on df.Member\_Number = gf.Member\_Number

left join (

select Member\_Number, count(\*) as pool\_Fre

from(select Member\_Number, count(\*) as pool\_Fre

from (select p1.Pool\_Account,Member\_Number, Date

from pool p1 left join poolaccounts p2 on p1.Pool\_Account = p2.Pool\_Account

order by Pool\_Account) as new\_pool

group by Member\_Number, Date) as aa group by Member\_Number) as pf on df.Member\_Number = pf.Member\_Number

left join(

select Member\_Number, count(\*) as tennis\_Fre

from(select Member\_Number,count(\*) as tennis\_Fre, Date

from tennis

group by Member\_Number, Date)as aa group by Member\_Number) as tf on df.Member\_Number = tf.Member\_Number

left join(

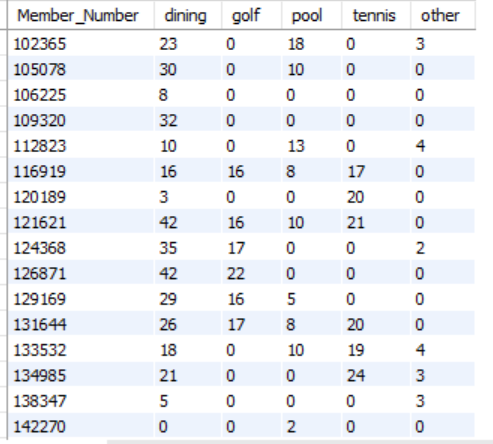
select Member\_Number, count(\*) as other\_Fre

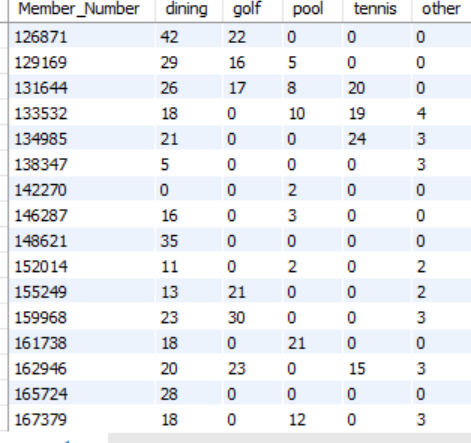
from(select Member\_Number,count(\*) as other\_Fre, Code

from other

group by Member\_Number, Code) as aa group by Member\_Number) as otf on df.Member\_Number = otf.Member\_Number);

**Output :**

****

****

**Revenue:**

CREATE view poolf AS

SELECT

pa.Member\_Number, pa.pool\_Account, p.Amount, p.Description, p.Date

FROM pool p

right join poolaccounts pa

on p.Pool\_Account = pa.Pool\_Account;

DROP table IF EXISTS `revenue`;

CREATE table revenue AS

SELECT i.Member\_Number, COALESCE(Dining,0) as Dining, COALESCE(Golf,0) as Golf,COALESCE(tennis,0) as tennis ,COALESCE(Pool,0) as Pool,COALESCE(Other,0) as Other

FROM

( select Member\_Number from memberships ) as i left join

(SELECT Member\_Number, Round(SUM(Amount),2) AS "Golf"

FROM golf GROUP BY Member\_Number) as j

on i.Member\_Number = j.Member\_Number left join

( SELECT Member\_Number, ROUND(SUM(Amount),2) AS "Tennis"

FROM tennis GROUP BY Member\_Number) as k

on i.Member\_Number = k.Member\_Number left join

(SELECT Member\_Number, ROUND(SUM(Amount),2) AS "Other"

FROM other GROUP BY Member\_Number) as l

on i.Member\_Number = l.Member\_Number

left join (SELECT Member\_Number,ROUND(SUM(Total),2) AS "Dining"

FROM dining GROUP BY Member\_Number) as m

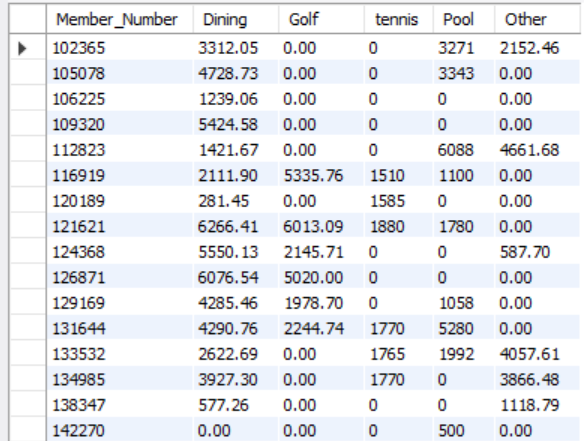
on i.Member\_Number = m.Member\_Number

left join (SELECT Member\_Number, ROUND(SUM(Amount),2) AS "Pool"

FROM poolf GROUP BY Member\_Number ) as n

on i.Member\_Number = n.Member\_Number group by i.Member\_Number;

**Output :**





**2. Queries from data warehouse**

**Query 1**

select member\_type, avg(ave\_dining)as ave\_dining, avg(ave\_golf)as ave\_golf, avg(ave\_tennis)as ave\_tennis, avg(ave\_pool)as ave\_pool,

avg(ave\_other)as ave\_other, avg(ave\_consumption)as ave\_consumption

from (select m.Member\_Number, Year\_Joined, r.dining as ave\_dining , r.Golf as ave\_golf, r.tennis as ave\_tennis, r.Pool as ave\_pool , r.Other as ave\_other,

(r.dining+ r.Golf+ r.tennis+ r.Pool+ r.Other) as ave\_consumption,

'old member' as member\_type

from revenue r

join MemberPromo m on r.Member\_Number = m.Member\_Number

where Year\_Joined < 2000

group by Member\_Number, Year\_Joined) as old\_member

union

select member\_type, avg(ave\_dining), avg(ave\_golf), avg(ave\_tennis), avg(ave\_pool), avg(ave\_other), avg(ave\_consumption)

from (select m.Member\_Number, Year\_Joined, r.dining as ave\_dining , r.Golf as ave\_golf, r.tennis as ave\_tennis, r.Pool as ave\_pool , r.Other as ave\_other,

(r.dining+ r.Golf+ r.tennis+ r.Pool+ r.Other) as ave\_consumption,

'new member' as member\_type

from revenue r

join MemberPromo m on r.Member\_Number = m.Member\_Number

where Year\_Joined >= 2000

group by Member\_Number, Year\_Joined) as new\_member;

**--------------------------------------------------------------------------**

****

Since 2000 is the beginning of the 21st Century, we defined new members as those who joined the club after 2000, and we defined old members as those who joined before 2000.

As we can see in the chart, in 2018, new member's average consumption on dining and pool are significantly higher than the old members'; this shows us that old members consumed more on ‘golf’, ‘tennis’, and ‘other’ relatively more on average. Also, in 2018 new members consumed more than old members overall on average.

**Query 2**

select case when Number\_of\_Children > 0 then 'Y' else 'N' end as have\_children, case when Partner > 0 then 'Y'else 'N' end as have\_partner,

sum(aa.a) / sum(aa.num)as Easter\_Brunch, sum(aa.b)/sum(aa.num)as Fourth\_of\_July, sum(aa.c)/sum(aa.num)as Thanksgiving, sum(aa.d)/sum(aa.num)as Private\_Function, sum(aa.aa) as total\_Frequency, sum(aa.num)as num\_people

from (select Number\_of\_Children, Partner, sum(Easter\_Brunch)as a , sum(Fourth\_of\_July)as b, sum(Thanksgiving)as c, sum(Private\_Function)as d,

sum(Easter\_Brunch)+sum(Fourth\_of\_July)+ sum(Thanksgiving)+sum(Private\_Function) as aa, count(\*)as num

from MemberPromo m join specialevents s on m.Member\_Number = s.Member\_Number

group by Number\_of\_Children, Partner) as aa

group by have\_children, have\_partner

order by have\_children

;

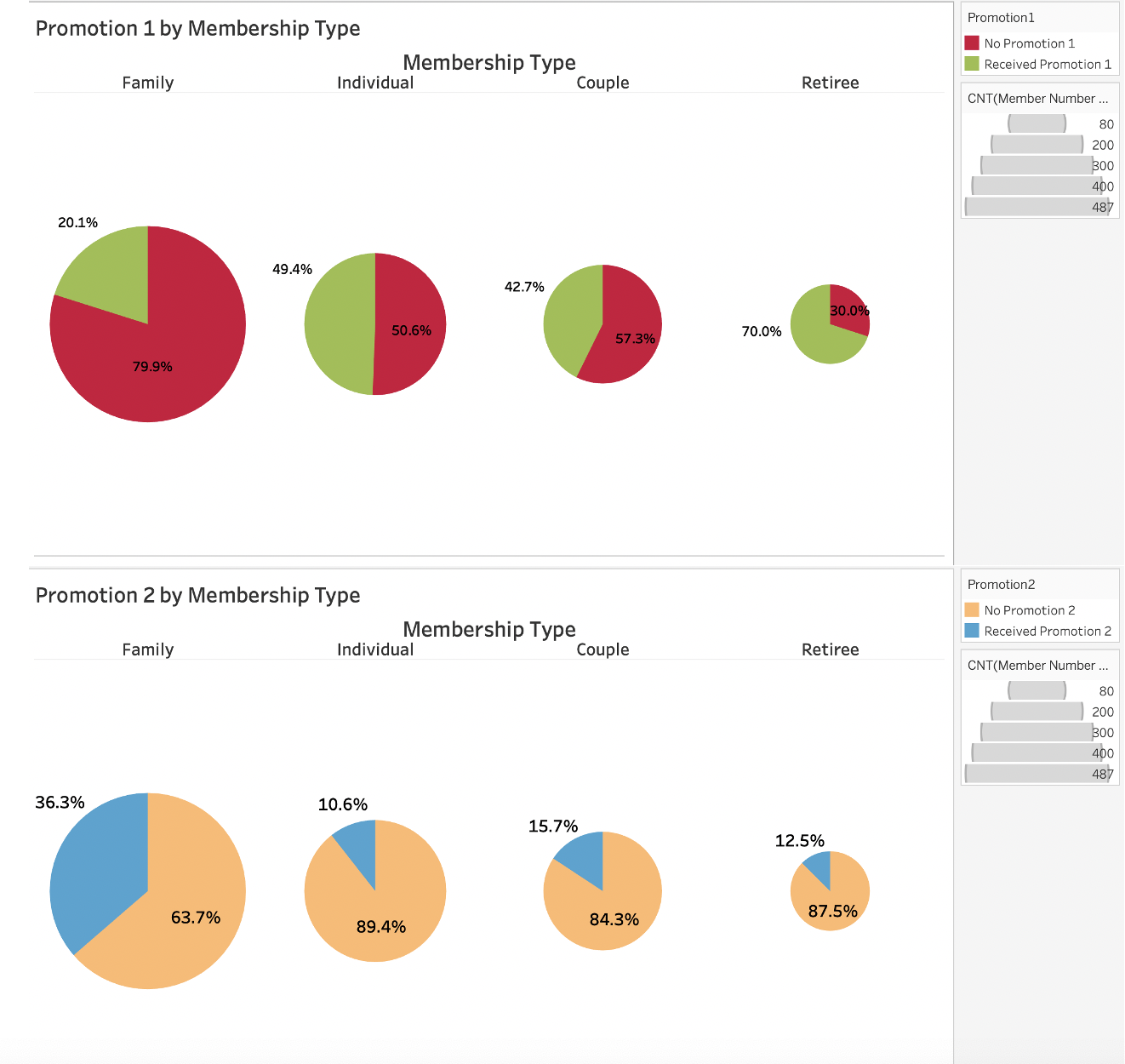
**-----------------------------------------------------------------------------------------------------------------**



\*\* From this table, we use numbers of times/total category people. However, since from our ‘special events’, each member would go to the event once, so we can regard that number of times as the number of people. Therefore, ‘numbers of times/total category people’ becomes ‘number of people go to the event/total category people ’.

We can see that all people have higher chances to go to Easter\_Brunch and Fourth\_of\_July compared to the other 2 special events, and people with children have higher chances to go to special events. Also for people who don’t have children, even though the number of people having a partner is close to people who don't have a partner (298 to 257), chances of going to the club are significantly larger as seen from total\_Frequency.

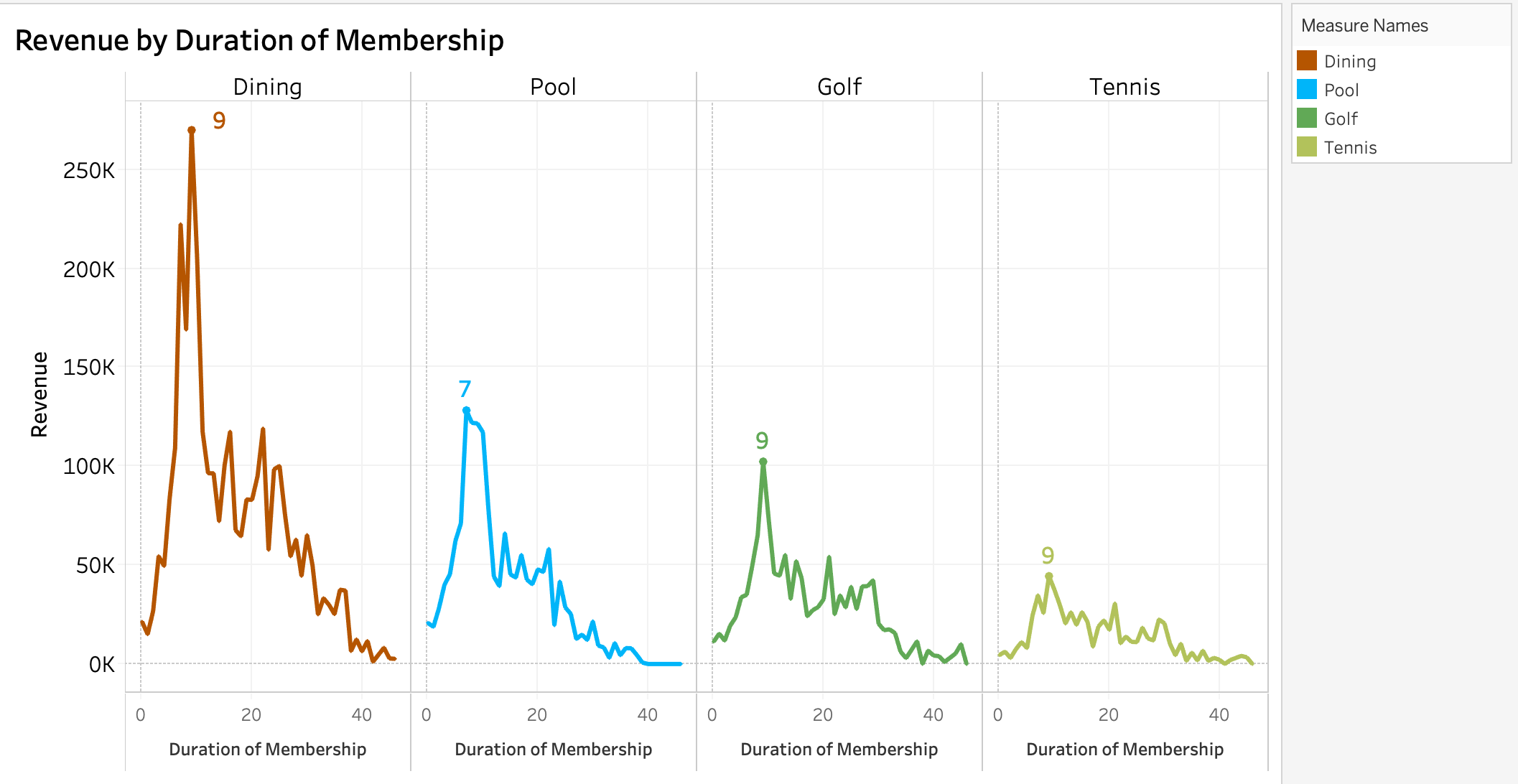
**3. Data Visualization**

****

A higher proportion of members in each membership type received Promotion 1 instead of Promotion 2

Promotion 1 was most commonly received by retirees (70%)

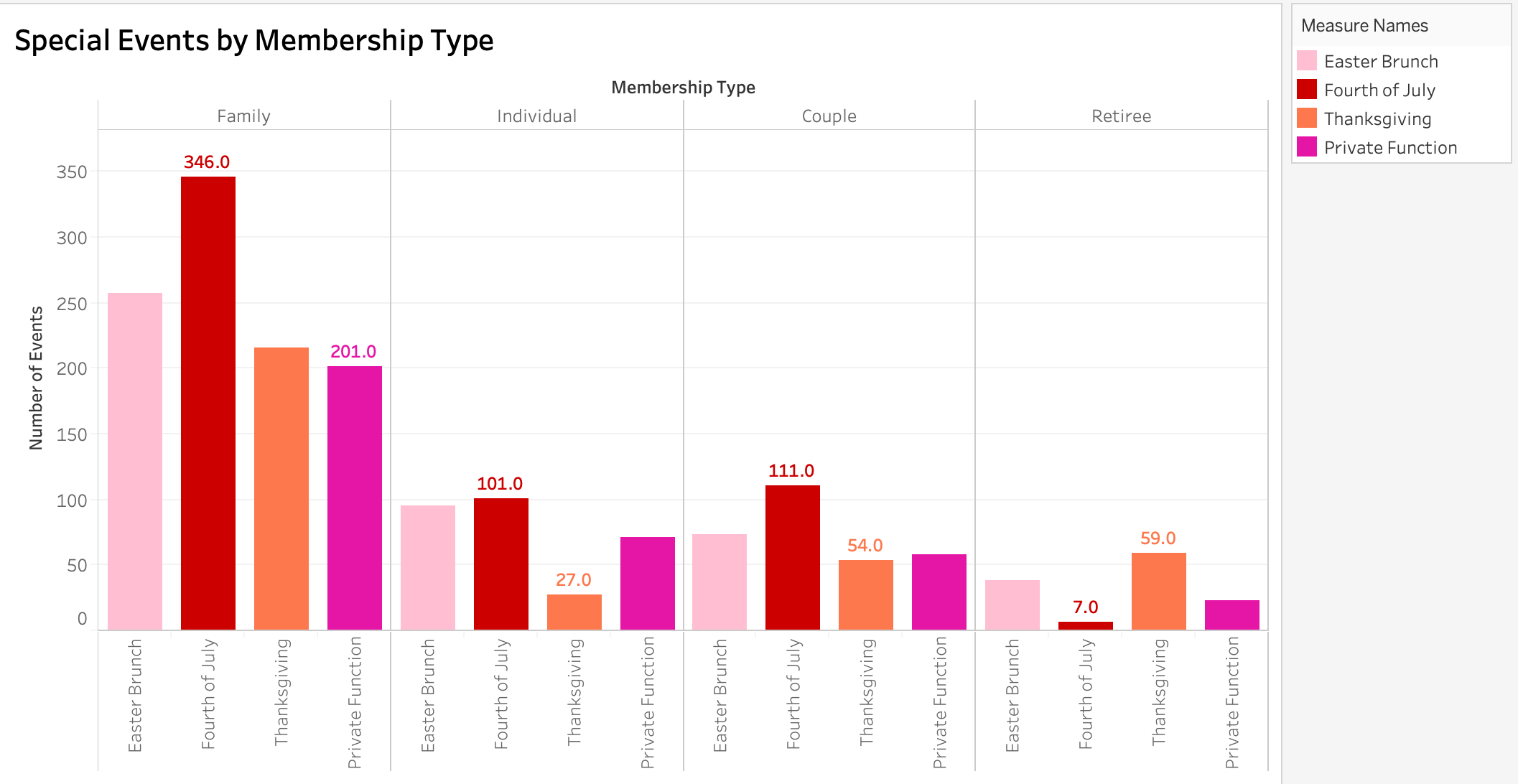
Promotion 2 was most commonly received by families (36.3%)

****

Dining gave the club the maximum revenue compared to Pool, Golf, and Tennis.

People who have been members for 9 years contributed the most towards the Dining, Golf and Tennis revenue. And those who have been members for 7 years contributed the most towards the Pool revenue.

Older members are spending less and less.



Fourth of July was the most celebrated event among Families, Individuals and Couples.

Among families, Private functions were the least common.

Among Individuals and couples, Thanksgiving events were the least common.

Among retirees, Fourth of July events were the least common.

**4. Potential Opportunities**

**Cross shopping opportunity**

/\*members who spend a lot of money on XXX are most likely to also spend money on XXX\*/

select sum(Dining!=0)/sum(Dining!=0) CroDining,sum(Golf!=0)/sum(Dining!=0) CroGolf,

sum(tennis!=0)/sum(Dining!=0) CroTennis,sum(pool!=0)/sum(Dining!=0) CroPool,

sum(Other!=0)/sum(Dining!=0) CroOther

from Revenue where Dining!=0

union

select sum(Dining!=0)/sum(Golf!=0) CroDining,sum(Golf!=0)/sum(Golf!=0) CroGolf,

sum(tennis!=0)/sum(Golf!=0) CroTennis,sum(pool!=0)/sum(Golf!=0) CroPool,

sum(other!=0)/sum(Golf!=0) CroOther

from Revenue

where Golf!=0

union

select sum(Dining!=0)/sum(tennis!=0) CroDining,sum(Golf!=0)/sum(tennis!=0) CroGolf,

sum(tennis!=0)/sum(tennis!=0) CroTennis,sum(pool!=0)/sum(tennis!=0) CroPool,

sum(other!=0)/sum(tennis!=0) CroOther

from Revenue where tennis!=0

union

select sum(Dining!=0)/sum(pool!=0) CroDining,sum(Golf!=0)/sum(pool!=0) CroGolf,

sum(tennis!=0)/sum(pool!=0) CroTennis,sum(pool!=0)/sum(pool!=0) CroPool,

sum(other!=0)/sum(pool!=0) CroOther

from Revenue where pool!=0

union

select sum(Dining!=0)/sum(other!=0) CroDining,sum(Golf!=0)/sum(other!=0) CroGolf,

sum(tennis!=0)/sum(other!=0) CroTennis,sum(pool!=0)/sum(other!=0) CroPool,

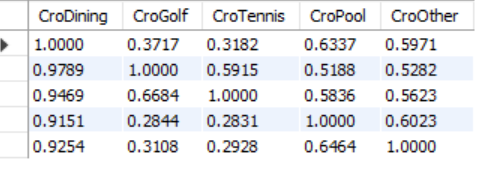
sum(other!=0)/sum(other!=0) CroOther

from Revenue where Other!=0;

**---------------------------------------------------------------------------------------------------------------------------**

**Recommendation:**

Members who spend a lot of money on golf,tennis,pool and other activities are most likely to also spend money on dining. For those who joined golf or tennis, they would have more than half probability to join the other 4 activities.



**Children and shopping behavior**

/\*members with more children are most likely to purchase XXX\*/

/\*Frequency\*/

select Number\_of\_Children, d/total ToDining, g/total ToGolf, p/total ToPool, t/total ToTennis, o/total ToOther

from

(select Number\_of\_Children, d, g, p, t, o, d+g+p+t+o as total

from

(select Number\_of\_Children, sum(dining) d,sum(golf) g,sum(pool) p,sum(tennis) t,sum(other) o

from

(select MP.Member\_Number,Number\_of\_Children,dining, golf, pool, tennis, other

from Frequency F

right join MemberPromo MP

on MP.Member\_Number = F.Member\_Number) as A

group by Number\_of\_Children

order by Number\_of\_Children) as B) as C;

/\*Revenue\*/

select Number\_of\_Children, D/Total ToDining, G/Total ToGolf, P/Total ToPool, T/Total ToTennis, O/Total ToOther

from

(select Number\_of\_Children, D, G, P, T, O, D+G+P+T+O as Total

from

(select Number\_of\_Children, sum(Dining) D,sum(Golf) G,sum(Pool) P,sum(tennis) T,sum(Other) O

from

(select MP.Member\_Number,Number\_of\_Children,Dining, Golf, Pool, tennis, Other

from Revenue R

right join MemberPromo MP

on MP.Member\_Number = R.Member\_Number) as AA

group by Number\_of\_Children

order by Number\_of\_Children) as BB) as CC;

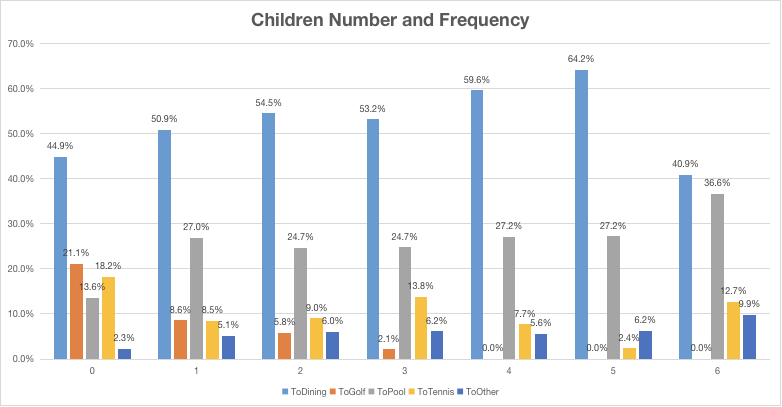
------------------------------------------------------------------------------------------------------------------------

**Recommendation:**

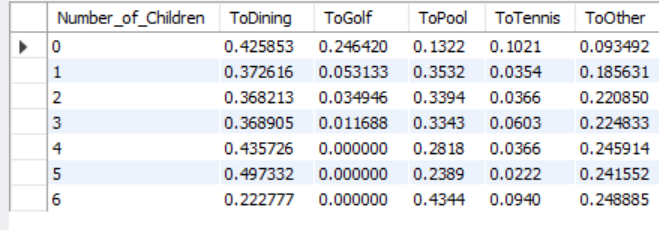
As the number of children increases for the member, it is less likely that the members will go to golf and members with more than 3 children never go to golf. Members with more children are more likely to go to the pool and other activities, especially members with 6 children.

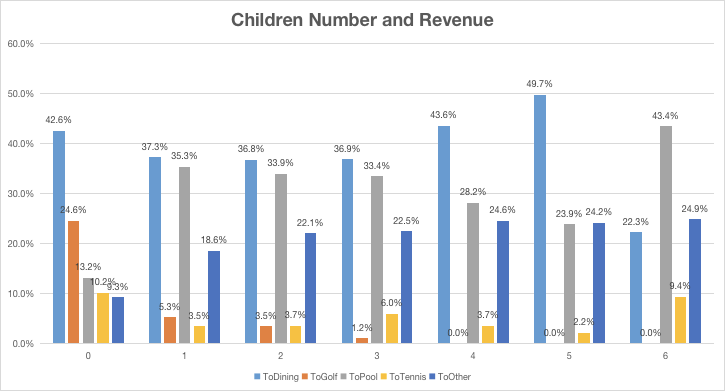
**Frequency:**





**Revenue:**





**5. Promotion 1 &2**

**Promotion analysis by Revenue**

select case when promotion1=1 and promotion2=0 then "Pro1"

when promotion2=1 and promotion1=0 then "Pro2"

when promotion1=1 and promotion2=1 then "Pro1&Pro2"

when promotion1=0 and promotion2=0 then "none"

end as promotion,

avg(Dining) as Dining, avg(Golf) as Golf,avg(tennis) as tennis, avg(Pool) as Pool, avg(Other) as Other, avg(TotalRevenue) as Total

from

(select Dining,Golf,tennis, Pool, Other, Dining+Golf+tennis+Pool+Other as TotalRevenue,promotion1, promotion2

from revenue r

right join MemberPromo m

on r.Member\_Number = m.Member\_Number) as RM

group by promotion

order by promotion;

**Promotion analysis by Frequency**

select case when promotion1=1 and promotion2=0 then "Pro1"

when promotion2=1 and promotion1=0 then "Pro2"

when promotion1=1 and promotion2=1 then "Pro1&Pro2"

when promotion1=0 and promotion2=0 then "none"

end as promotion,

avg(Dining)as Dining, avg(Golf) as Golf,avg(tennis) as tennis, avg(Pool) as Pool, avg(Other) as Other, avg(TotalRevenue) as total

from

(select Dining,Golf,tennis, Pool, Other, Dining+Golf+tennis+Pool+Other as TotalRevenue,promotion1, promotion2

from frequency r

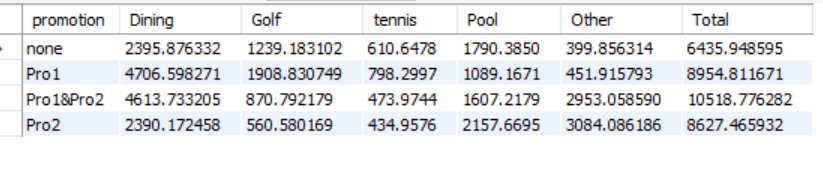
right join MemberPromo m

on r.Member\_Number = m.Member\_Number) as RM

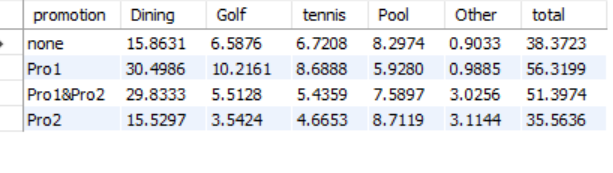
group by promotion

order by promotion;

**Revenue:**



**Frequency:**



A potential explanation for promotion 1 is that it was focused on the activities tennis and golf. It's possible that the promotion focused on an offer of discounted dining contingent on whether the consumer participated in golf or tennis activities. Based on the frequency table, there is a high average of 10.22 and 8.7 visits for golf and tennis for promotion 1, and we see a huge spike in dining visits as well. Therefore, we believe that there was an offer of discounted dining if the consumer participated in golf or tennis activities. Looking at revenue, we see that it was quite an effective promotion. $1909 and $798 were spent on golf and tennis on average, and $4706 was spent on dining from that promotional period as well. Unfortunately, there may also be some cannibalization for the pool activity as more people went towards tennis and golf for the promotion. We can see this as the visits for the pool was only 6 despite having a revenue of $1089 for promotion 1.

A potential explanation for promotion 2 is that there was a promotion for a special event such as their Christmas party event. We believe that there was a promotion to attend the party which attracted consumers to also participate in activities such as ski and bridge tournaments. From frequency, we see that 226 people attended the special events for promotion 2, so it is possible that promotion 2 promoted a special event. And when people attended the event, they partook in skiing and bridge tournament events during the season. This is a plausible explanation as 226 different consumers partook in the special events category for promotion 2, and we also see an uptick in attended events of skiing (105) and bridge tournament (140). We can see that a large portion of the consumers spent their money for this as well since the "other" category was a whopping $3084 on average for their revenue and this category included the special events and activities.