**I - Description and data population**

March 20th 2018

**Analysis & Recommendation**

**GROUP 4**

Mai Nguyen

Yuan Fang

Eddy Zhu

1. Introduction about Meetup

The application we use is Meetup.com, a website and mobile app that allow people to organize groups of interest, hold and promote events within the groups and the members.

Business model: Meetup works like a double-sided platform where organizers and members connect with each others via groups and events. Organizer will pay for Meetup to create groups and events and to promote it among Meetup's members. As other platform, Meetup will be affected by network effect. This means that the more number of members and member engagement, the more likely that organizer will pay to organize a meetup. On the other hand, the more relevant groups and events that are organized, the more likely members will engage with Meetup.



Groups

Events

1. Description of the application & Entity Relationship Diagram

Here is the description of the application around groups, members, event, topics, venue and category.

1, Each group is identified by its Group ID, Group name, description of the group, location (including city, state, country time), number of members, join mode and ratings.

2, Each member is identified by ID, Name, Status, Hometown, Location (including City, State, Country, Latitude and Longitude)

3, Each event is identified by its event ID, description, created time, duration, fee (accepts, including amount, currency, label, require), visibility and count of yes RSVP.

4, Each topic has topic ID as its identification, topic name, description, number of members interested in the topic and URL key.

5, Each venue is identified by Venue ID, address, city, state and country.

6, Each group is organized by one member and each member can organize multiple groups. Also, created time is associated with it.

7, Each group can have multiple members and each member can join in multiple groups. There is a joined time and last visited time associating with each joining.

8, Each member can select more than one topic as their interests and each topic can be selected by multiple members.

9, Each group can have multiple topics and each topic can be included in multiple groups.

10, Each event is held by only one group, but one group can host multiple events. For each event created by a group, created time is attributed.

11, Each event takes place at only one venue, but one venue can be the location of multiple events. For each event taking place at the venue, there is event time associated with it.

12, Each group can have only one category, but one category can have multiple groups.

13, Each city is identified by its City\_ID, Distance, state, country, latitude, longitude, zipcode and ranking. Each group can register at only one city but a city can have multiple groups.

In the scope of this analysis, we ignore the relationship between member and event.

**ER Diagrams and Tables**

**Members**

**Member\_ID**

Name

Status

Hometown

Place (city, state, country, lat, lon)

**Categories**

**Category ID**

Category Name

Short name

Sort name

**Events**

**Event ID**

Description

Duration  
Fee (accepts, amounts, currency, label, required)

**Joined time,**

**Visited time**

**M**

**1**

**1**

**Hold**

**Join**

**Belong**

**M**

**M**

**M**

**M**

**Organize**

**M**

**Created time**

**Select**

**Created time**

**1**

**M**

**Groups**

**Group ID**

Group Name

City

Country

Created time

Description

Join Mode

Members

Ratings

**Topics**

**Topic\_ID**

Name

Description

Number of members

URL key

**Take place**

**M**

**Event\_Time**

**Have**

**M**

**M**

**Melong**

**1**

**M**

**Venues**

**Venue ID**

Address

City

State

Country

**Cities**

**City ID**

Distance

Place (country, lat, lon, state, zip)

Ranking

**1**

**Register**

**Tables**

(Group\_ID, Group name, City, Country, Description, Join mode, Members, Ratings, Organizer\_ID, Created\_Time, City\_ID, Category\_ID)

**Topics**

(Topic\_ID, Topic name, Description, Number of member, URL Key)

**Members**

(Member\_ID, Name, Status, Hometown, City, State, Country, Lat, Lon)

**Events**

(Event\_ID, Description, Duration, Fee\_accepts, Fee amounts, Fee currency, Fee label, Fee required, Group\_ID, Created\_time, Venue\_ID, Event\_Time)

**Venues**

(Venue\_ID, Address, City, State, Country)

**Categories**

(Category\_ID, Category name, Short name, Sort name)

**Cities**

(City\_ID, Distance, Country, Lat, Lon, State, Zip, Ranking)

**Join**

(Member\_ID, Group\_ID, Joined time, Visited time)

**Have**

(Topic\_ID, Group\_ID)

**Select**

(Topic\_ID, Member\_ID)

Note: Attributes in blue are primary keys, and attributes in yellow are foreign keys.

1. Data population & Limitation

**Population of the data**

Our dataset contains 16330 rows of group with Group ID as unique key, 178218 rows of member with member ID as unique key, 5807 rows of event with event ID as unique key, 2509 rows of topic with topic ID as unique key, 107093 rows of venue with venue as unique key and 33 rows of categories with categories as unique key.

**Limitation:**

This dataset have some notable limitation:

* Data about events are the upcoming events at a certain point in time that will happen in the future. We don't have data about the events that already happened in the past.
* Data are collected from only 3 main cities New York, Chicago and San Francisco.
* This application ignore the relationship between Member and Event.

**II - Analyzing the data**

**Users’ Interests Analysis**

In this part, we want to investigate in what are Meetup users most interested in. To answer the question, we break it down into three sub-questions and will answer from three perspectives. And we want to use this analysis to help get some insights regarding what kind of groups to create and what kind of events should be promoted in the future.

**Question 1: Which categories are more popular in general? (in terms of number of groups created/number of members)**

**Q1-1**

--Measured by number of groups created

--Top 3 Category: Tech (3644), Career & Business (2544), Socializing (1278)

select g.category\_id, c.category\_name, count(g.group\_id) as NumGroup

from [Groups Table] g, [Categories Table] c

where g.category\_id=c.category\_id

group by g.category\_id, c.category\_name

order by NumGroup desc

Go

**Q1-2**

--Measured by number of members

--Top 3 Category: Tech, Career & Business, Socializing,

DROP VIEW IF EXISTS dbo.PopCatMem

go

create view PopCatMem as

select g.category\_id, c.category\_name, count(distinct j.member\_id) as NumofMem

from [Groups Table] g, [Categories Table] c, [Join Table] j

where g.category\_id=c.category\_id and g.group\_id=j.group\_id

group by g.category\_id, c.category\_name

go

select \*

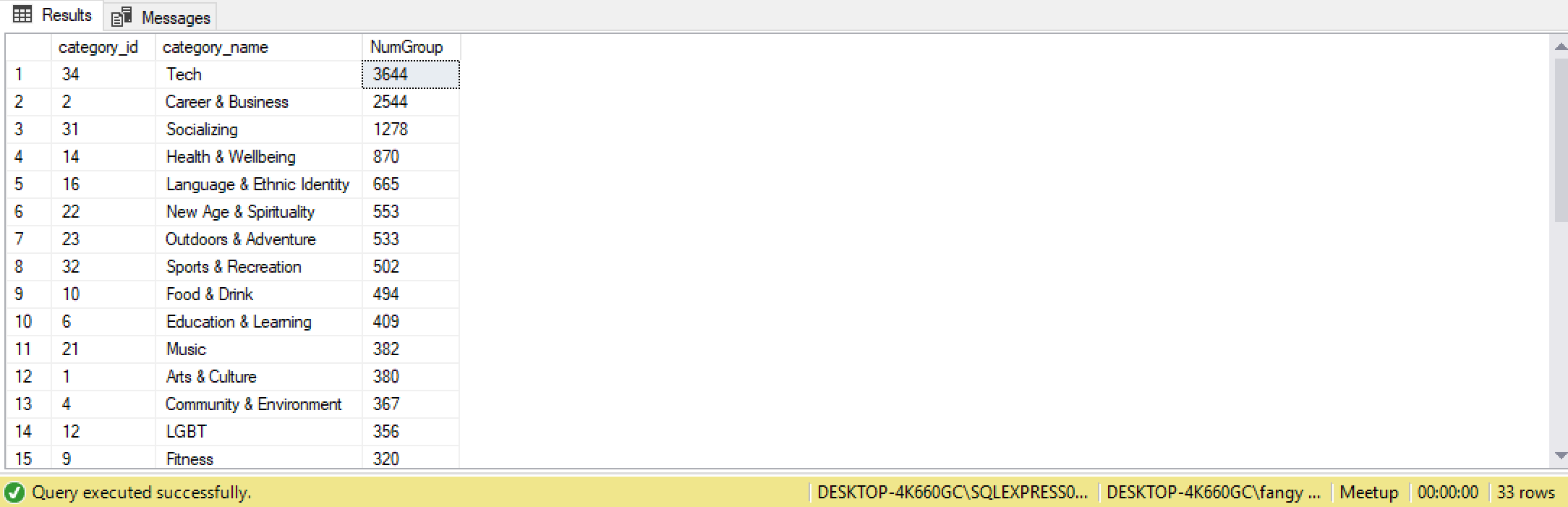
from PopCatMem

order by NumofMem desc

go

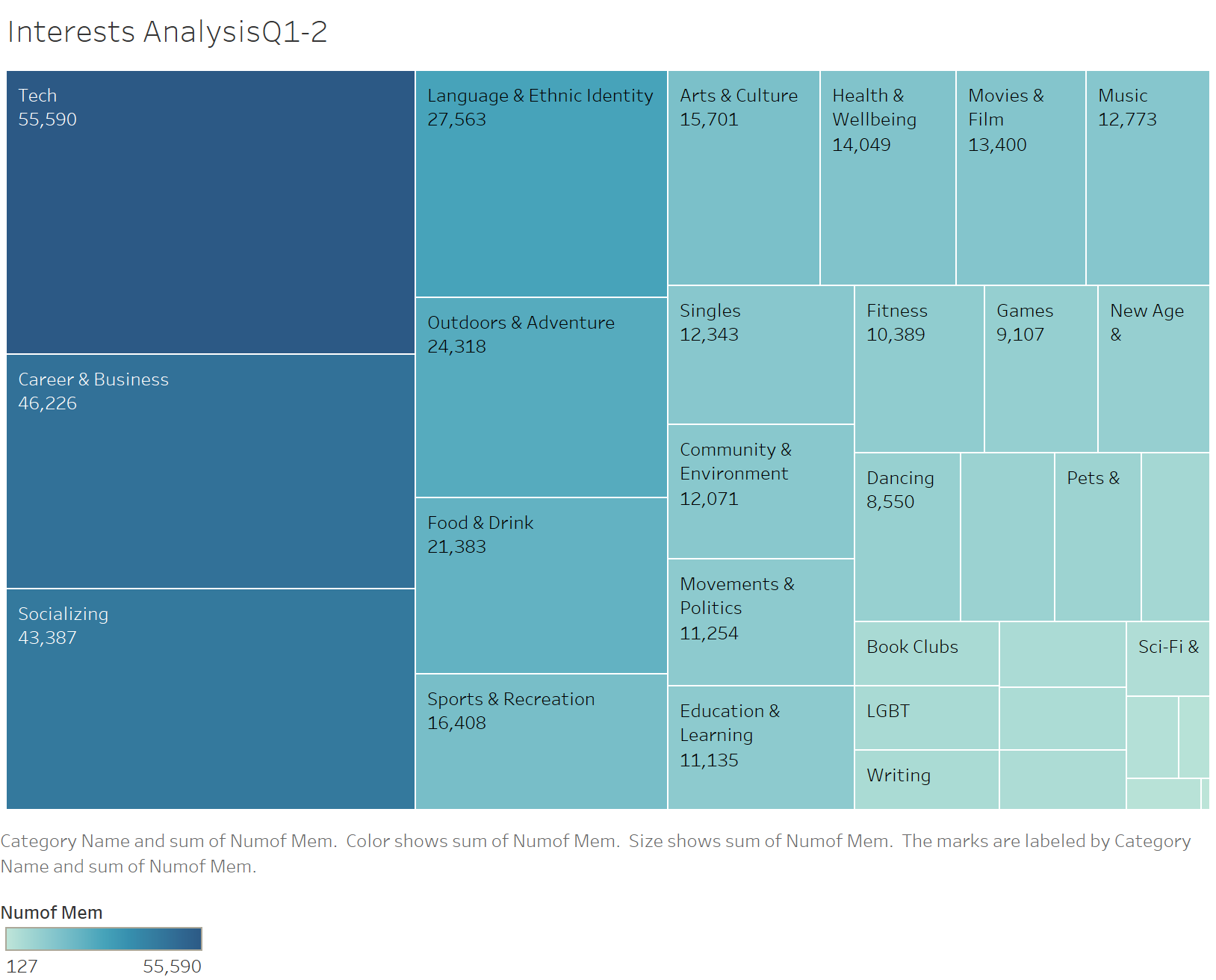
**Result:**

**Q1-1**

****

**Q1-2**

****

****

**Insight and application:**

1) Measured by number of groups created, the top 5 categories are: Tech, Career & Business, Socializing, Health & Wellbeing, and Language & Ethnic Identity.

Measured by number of members joined, the top 5 categories are: Tech, Career & Business, Socializing, Language & Ethnic Identity, and Outdoors & Adventure.

2) We can see the top 3 categories are the same for both measurements, and those will be the main categories if the meetup company want to create large events to attract most people.

3) To allocate company resource for new trendy categories, we think the measurement of number of members makes more sense because it takes the activity level into consideration. We would recommend the company to create more events for the categories of Language & Ethnic Identity and Outdoors & Adventure given the large user base.

**Question 2: What are the difference among cities in terms of interests in categories?**

Q2-1

--Count the Number of Categories for each city

--We notice only 3 cities(New York, Chicago, San Francisco) include all categories

--We will use these 3 cities for the following analysis

SELECT dbo.[Cities Table].city\_id, dbo.[Cities Table].city\_name, COUNT(DISTINCT dbo.[Categories Table].category\_id) AS NumCat

FROM dbo.[Categories Table] INNER JOIN

dbo.[Groups Table] ON dbo.[Categories Table].category\_id = dbo.[Groups Table].category\_id INNER JOIN

dbo.[Cities Table] ON dbo.[Groups Table].city\_id = dbo.[Cities Table].city\_id

GROUP BY dbo.[Cities Table].city\_id, dbo.[Cities Table].city\_name

ORDER BY NumCat desc

go

Q2-2

--See the popular categories for the 3 big cities(New York, Chicago, San Francisco)

--People in those 3 big cities have similar interests distribution based on the 33 categories

DROP VIEW IF EXISTS dbo.PopCat

go

create view PopCat as

select ct.city\_name,c.category\_id,c.category\_name,count(\*) as NumofGroup

from [Categories Table] c, [Groups Table] g, [Cities Table] ct

where c.category\_id=g.category\_id and g.city\_id=ct.city\_id

group by ct.city\_name,c.category\_id,c.category\_name

go

select \*

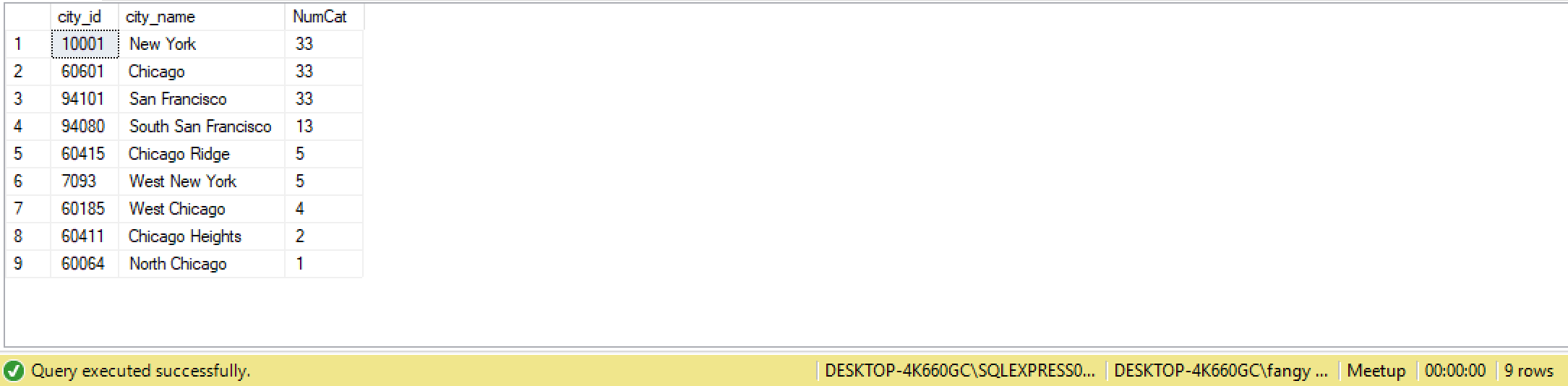
from PopCat

order by PopCat.city\_name, NumofGroup desc

go

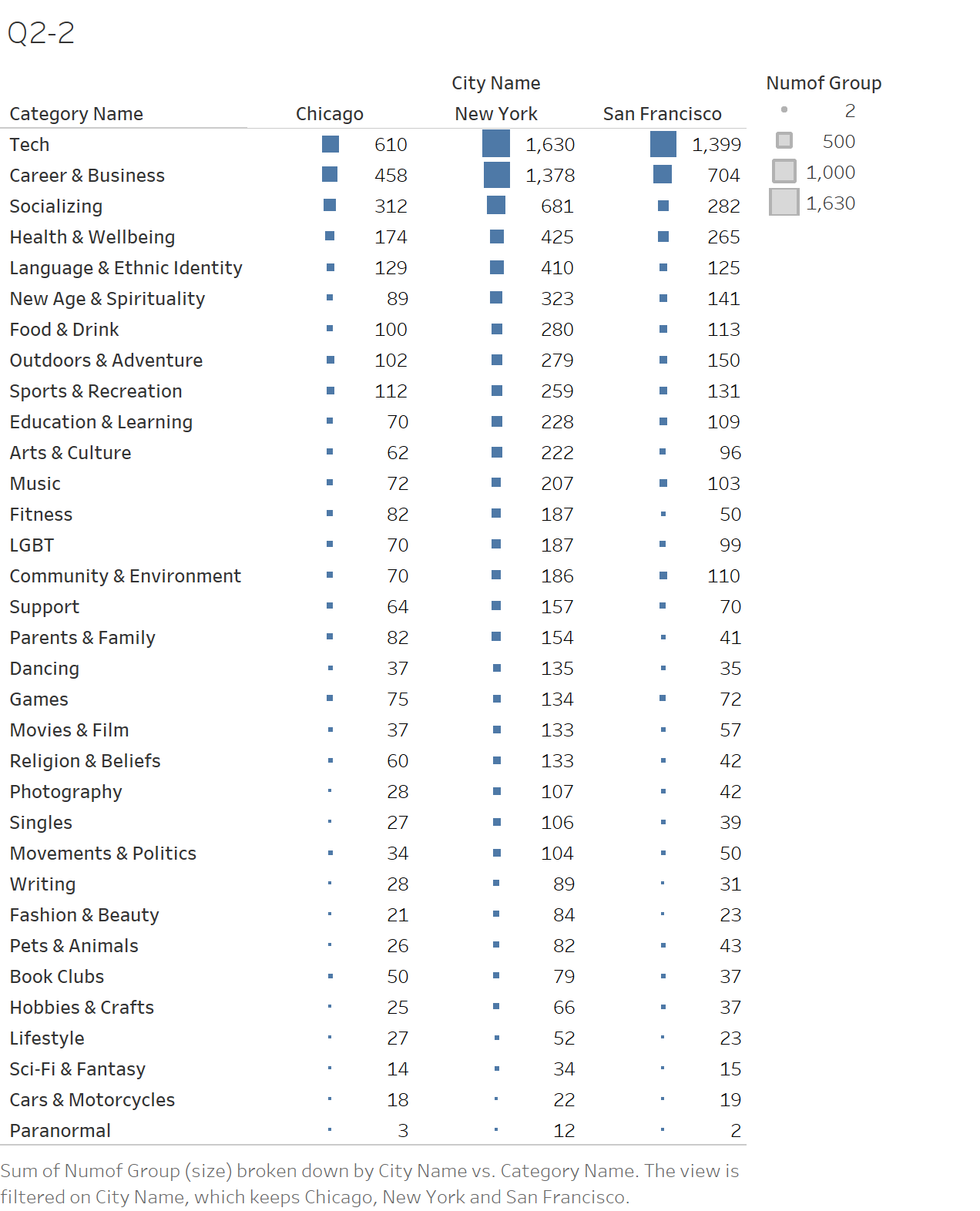
**Result:**

**Q2-1**



**Q2-2**





**Insight and application:**

1. People in those 3 big cities have similar interests’ distribution based on the 33 categories The most popular categories are Tech, Career & Business, Socializing, and Health & Wellbeing. So we suggest the company to focus on these four categories as its core “business”.
2. For new trendy categories, we may have different strategies for different cities. For example, the company can create more events regarding Outdoors & Adventure and Sports & Recreation in San Francisco, create more events regarding Food & Drink in Chicago. For New York, the company can use this city as the test stone for most new events because the group numbers are least centralized in New York and interests are most diversified in New York City.

**Question 3: The popularity of different categories (in term of number of group created) across different years (2002-2017).**

--The popularity of different categories (in term of number of group created) across different years.

--Which is the most popular categories in these years?

--For the top 3 popular categories, how many groups are there in each categories in chronological(year) order? (get a sense of times series trend)

--Step 1: Check the year range

--(2002-2017)

select distinct datepart(year,created) as years

from [Groups Table]

order by years

go

--Step 2:

DROP VIEW IF EXISTS dbo.TimePopCat

go

create view TimePopCat as

select g.category\_id, c.category\_name, datepart(year,g.created) as Years, count(g.group\_id) as NumGroup

from [Categories Table] c, [Groups Table] g

where c.category\_id=g.category\_id

group by g.category\_id, c.category\_name, datepart(year,g.created)

go

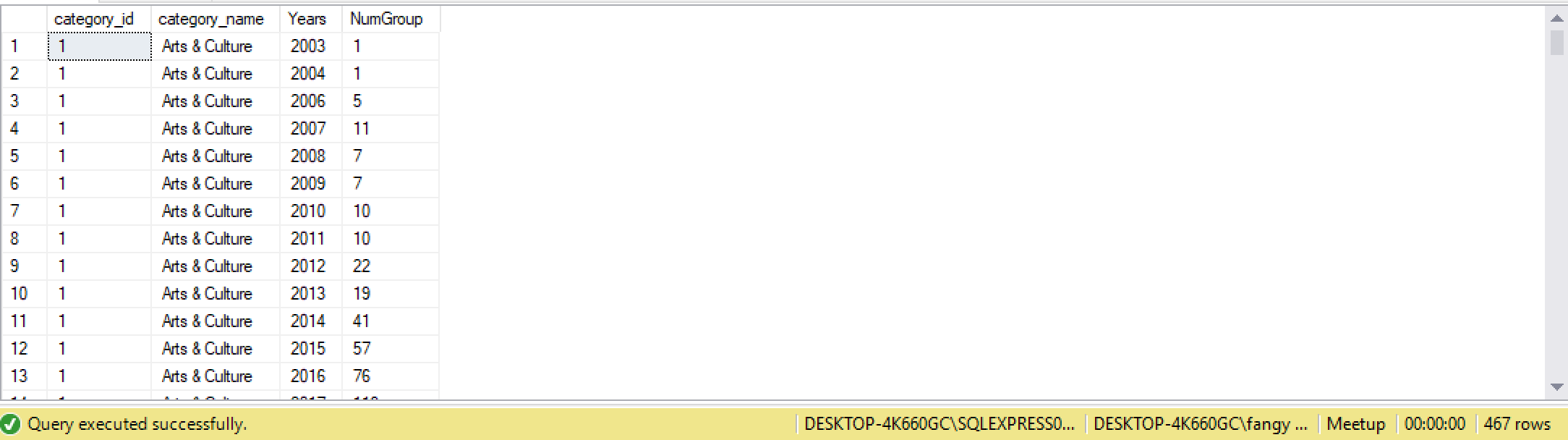
select \*

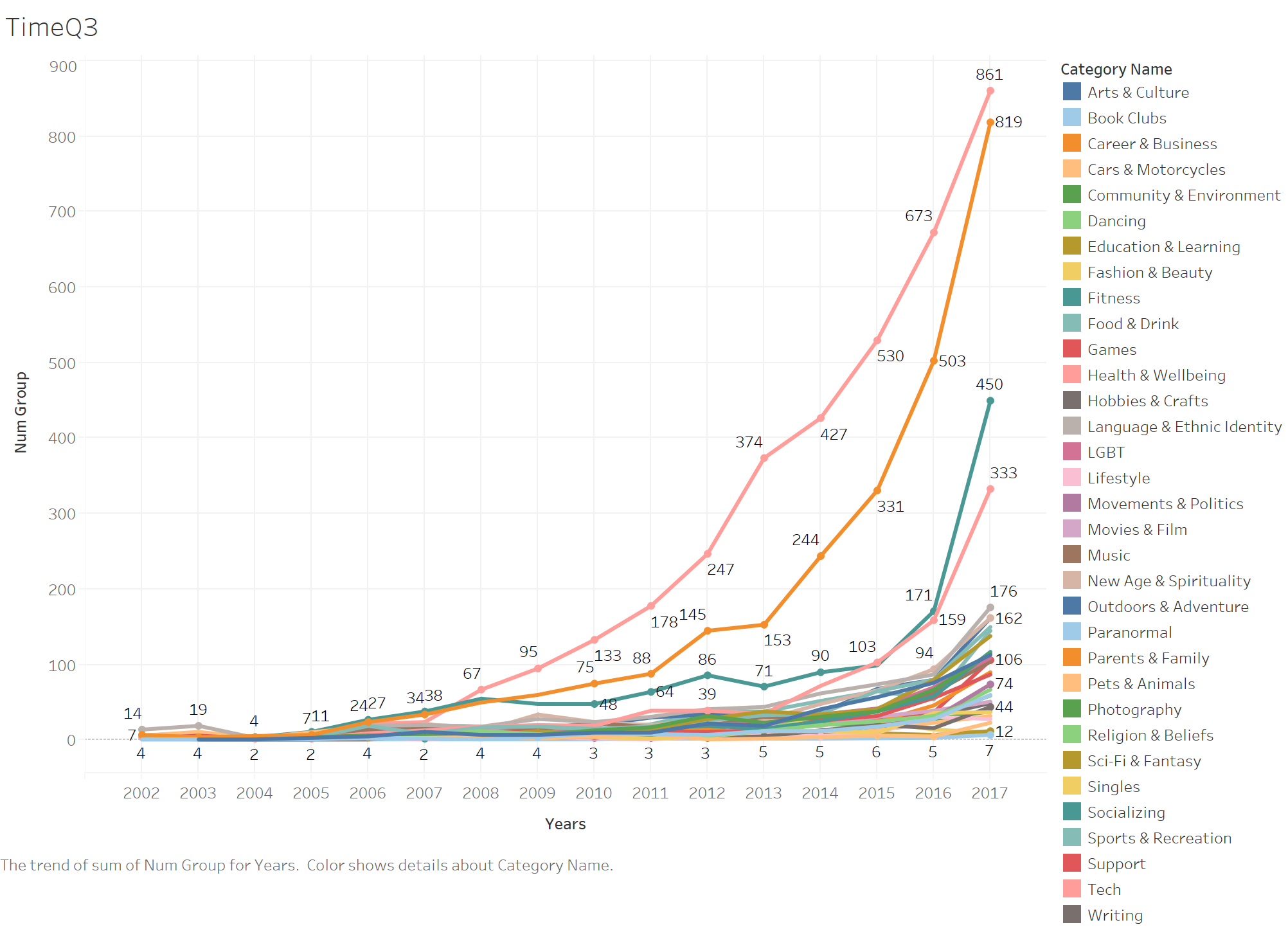
from TimePopCat

order by category\_id,Years

go

**Result:**

****

****

**Insight & Application:**

Look from the perspective of time series, the results are quite consistent with what we have got in previous part. The four most popular categories are also the fastest growing categories: Tech, Career & Business, Socializing, and Health & Wellbeing. So, we suggest the company to focus on these four categories as its core “business”.

**Frequently interested together**

**Question 4: What are the topics that are frequently interested together by members.**

Meetup want to answer this correlation question to build a recommendation system in which it can automatically suggest the topics that are frequently interested together with a certain topic.

**Queries:**

--Question: what are the topics that are frequently interested together by the members?

--Duplicate the member\_topic table

DROP VIEW IF EXISTS dbo.Dup\_member\_topic

go

create view Dup\_Member\_topic as

select mt.member\_id,mt.topic\_id from Member\_topic MT

go

--Join 2 similar tables (member\_topic and dup\_member\_topic) to pair the topics that are selected by one member

DROP VIEW IF EXISTS dbo.membertopic

go

create view memberTopic as

select mt.topic\_id as Topic1, dmt.topic\_id as Topic2, count(distinct mt.member\_id) as NumMem

from member\_topic MT, dup\_member\_topic DMT

where Mt.member\_id = dmt.member\_id and mt.topic\_id <> dmt.topic\_id

group by mt.topic\_id, dmt.topic\_id

go

--Get topic name for each topic ID

DROP VIEW IF EXISTS dbo.membertopic\_name01

go

create view memberTopic\_name01 as

select t.topic\_name as TopicName1, mt.Topic2 as Topic2, mt.NumMem as NumMem

from memberTopic MT, Topic T

where mt.topic1 = t.topic\_id

go

DROP VIEW IF EXISTS dbo.membertopic\_name

go

create view memberTopic\_name as

select mtn.TopicName1 as TopicName1, t.topic\_name as TopicName2, mtn.NumMem as NumMem

from memberTopic\_name01 MTN, Topic T

where mtn.topic2 = t.topic\_id

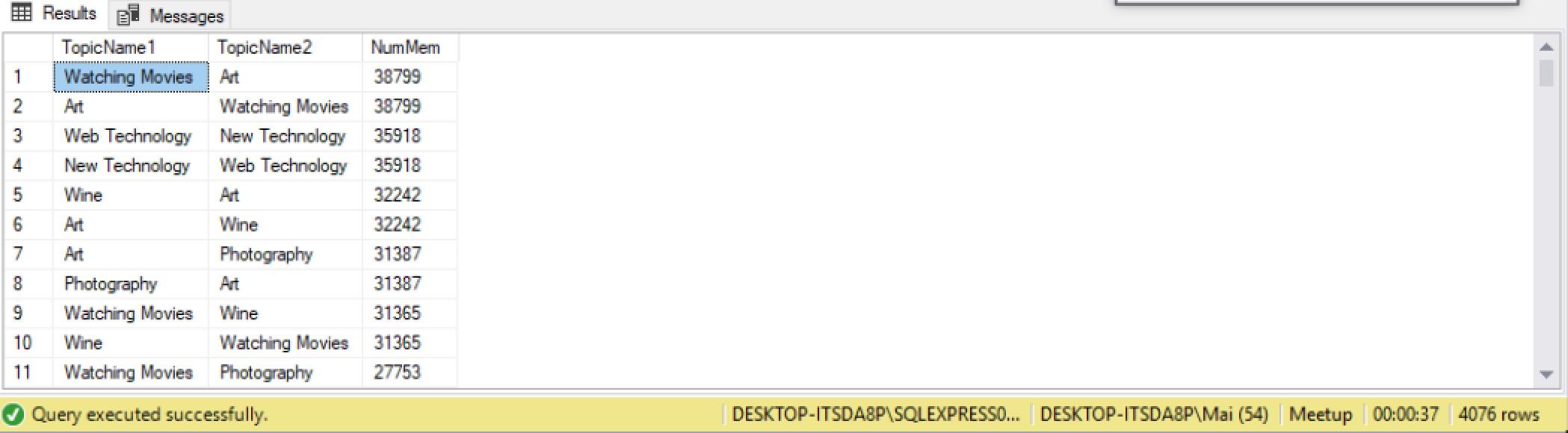
go

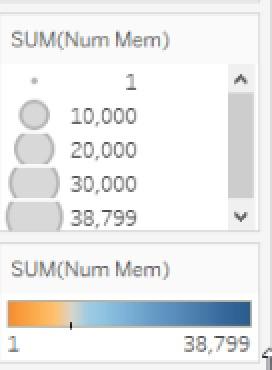
select \* from membertopic\_name

order by NumMem desc

go

**Result:**





**Insight and application:**

As we can see from the result, the pairs of topics that are frequently selected by members concentrate on a certain numbers of topics such as Art, Watching Movie, Wine, Photography, New technology, Education, etc.

Using this result, Meetup can suggest a member who already selected a topic as his interest other topics that are usually selected together with this topic.

For example: A new member selected Art as his interest. Based on this information and the result above, Meetup can suggest other topics that are frequently selected together with “Art” by other members such as Watching Movie, Wine, Photography, Craft,...to this new member. Therefore, Meetup can widen the range of interest among members, and automatically recommends the groups that are relevant with their interests or other members that have similar interests.

**Question 5: What are the categories that are frequently joined by members?**

Meetup want to build a recommendation system in which it can automatically suggest new groups belonging to categories that are frequently interested together to its members.

**Queries:**

---- What are the 2 categories that people are frequently interested in together.

---- Create a table to collect information about members joining category

drop view if exists dbo.mem\_category

go

create view mem\_category as

select gm.member\_id as member\_id, c.category\_id as category\_id, max(c.category\_name) as category\_name

from Group\_member GM, Categories C, Groups G

where Gm.group\_id = G.group\_id and g.category\_id = c.category\_id

group by gm.member\_id, c.category\_id

go

---Duplicate the member - category table

DROP VIEW IF EXISTS dbo.dup\_mem\_category

GO

create view Dup\_mem\_category as

select member\_id, category\_id, category\_name

from mem\_category

go

--Join 2 similar tables to pair the categories that members usually take togethers.

Create view Together\_categories as

select mc.category\_name as Category1, dmc.category\_name as Category2, count(distinct mc.member\_id) as NumMem

from mem\_category MC, Dup\_mem\_catagory DMC

where mc.member\_id = dmc.member\_id and mc.category\_id <> dmc.category\_id

group by mc.category\_name, dmc.category\_name

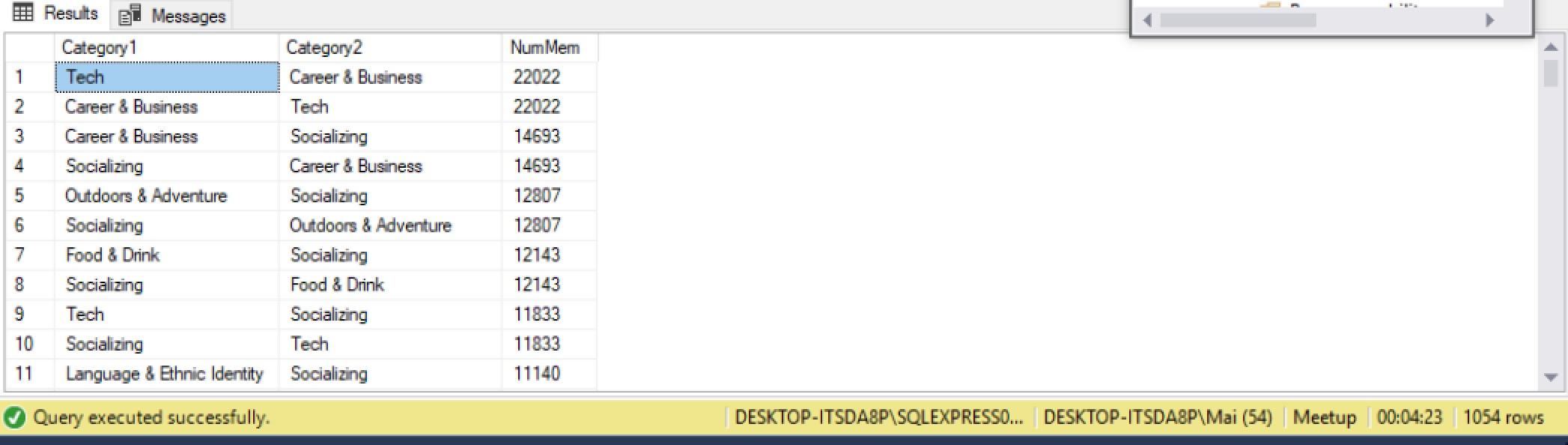
go

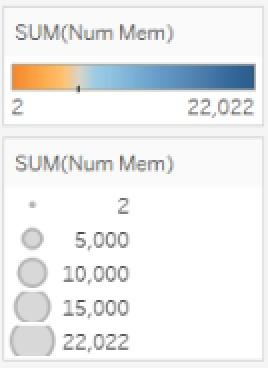
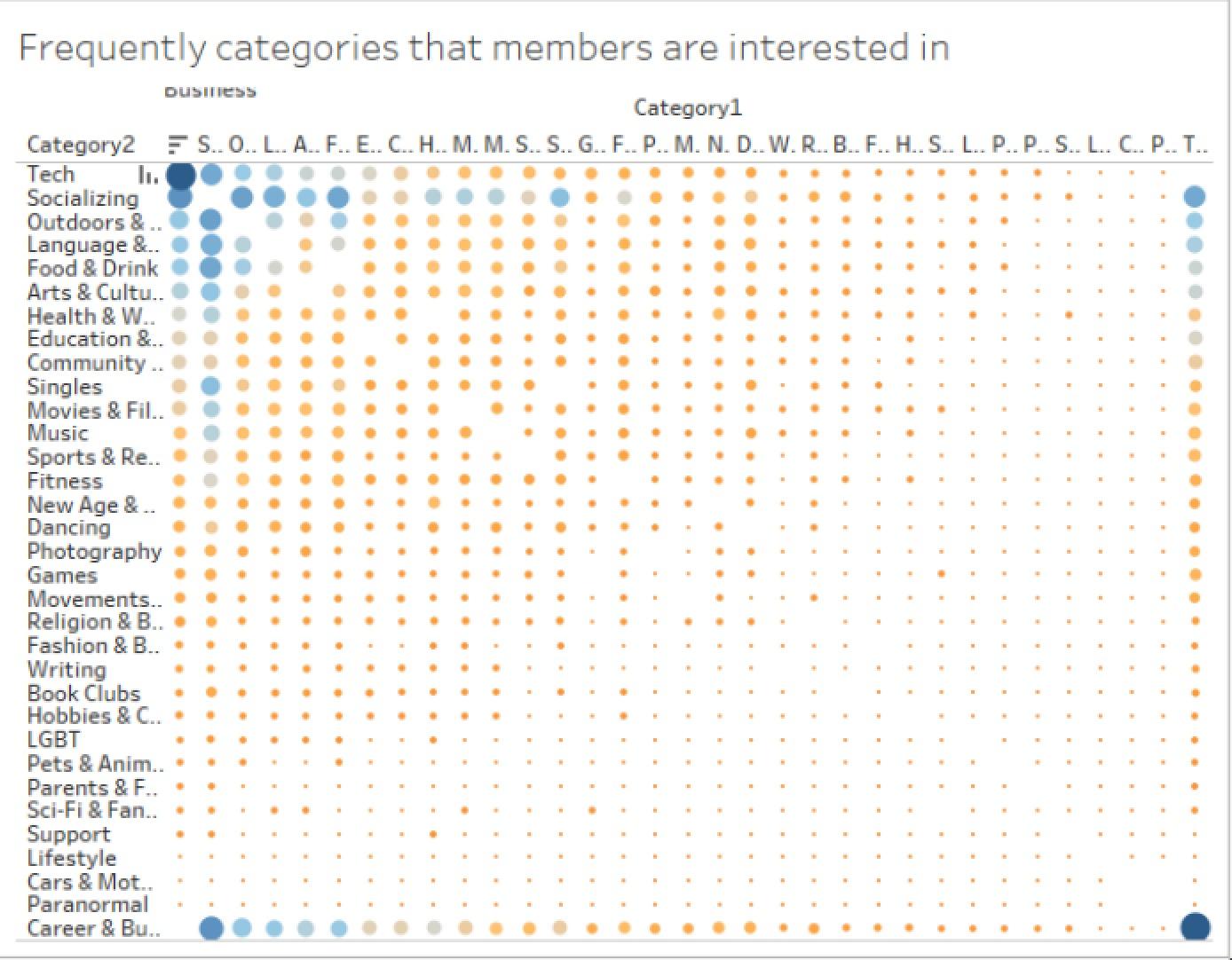
Select \* from Together\_categories

order by NumMem desc

go

**Result:**





**Insight and Application:**

The categories that are frequently selected together concentrate on a certain number of categories such as Tech, Career & Business, Socializing, Outdoors, Language & Ethnicity, etc.

This result illustrates that the groups that members usually join together are related to Tech, Career & Business, Socializing, Outdoors, etc. Again, based on this information, Meetup can suggest a certain member other groups that usually joined together by other members,

For example, a new member decided to joined a group belonging to category Tech. Meetup has the data about the other groups belonging to Career & Business, Socializing, Outdoors, etc that are frequently selected by other members. Therefore, it can suggest this member to join these types of groups.

From question 3.1 and question 3.2, we also see that there is a misalignment between the topics that members selected and the groups that they join. While many people are interested in Art, Watching Movie, Photography, the number of people join categories like Art & Culture, Photography and Movies & Films are not aligned. This can be an area for Meetup to develop more groups and more events related to these interests.

**Question 6: Comparison between 3 cities**

**Queries:**

select max(event\_time) as lasttime

from [Events Table] e

go

drop view if exists dbo.recency\_event

go

create view recency\_event as

select group\_id, datediff(day, max(e.event\_time), '2018-10-25') as recency

from [Events Table] e

group by group\_id

go

drop view if exists dbo.frequency\_event

go

create view frequency\_event as

select group\_id, count(event\_id) as fequency

from [Events Table] e

group by group\_id

go

drop view if exists dbo.group\_view

go

create view group\_view as

select g.group\_id, g.city\_id, r.recency as Recency\_days, f.fequency as Frequency\_times

from [Groups Table] g, recency\_event r, frequency\_event f

where g.group\_id=r.group\_id and g.group\_id=f.group\_id

group by g.group\_id, g.city\_id, r.recency, f.fequency

go

drop view if exists dbo.mapping

go

create view mapping as

select v.group\_id, v.Frequency\_times as Fequency, v.Recency\_days as Recency, v.city\_id as City\_id, c.city\_name as City\_Name, c.zip as Zip

from [Cities Table] c, group\_view v

where c.city\_id=v.city\_id

group by v.group\_id, v.Frequency\_times,v.Recency\_days, v.city\_id, c.city\_name, c.zip

go

drop view if exists dbo.mapping2

go

create view mapping2 as

select g.city\_id, COUNT(group\_id) as Num\_group, c.zip as Zip

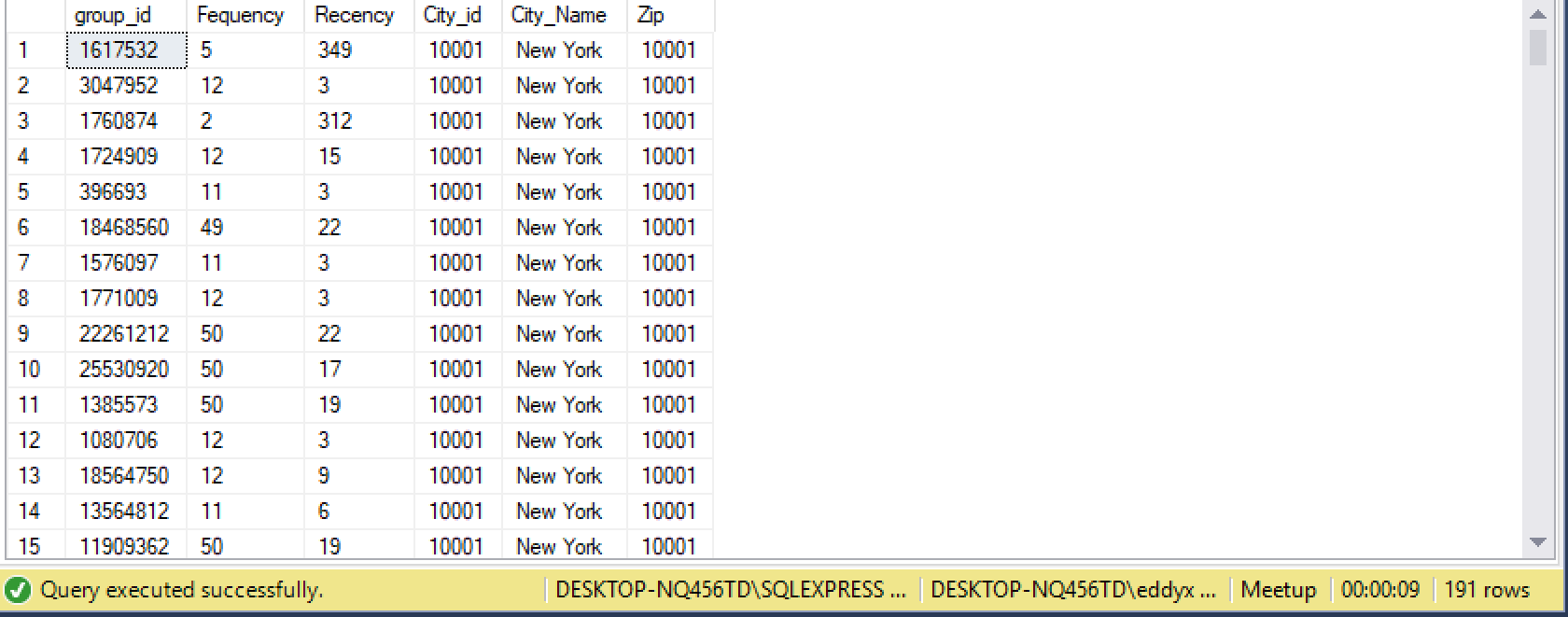
from [Groups Table] g, [Cities Table] c

where g.city\_id=c.city\_id

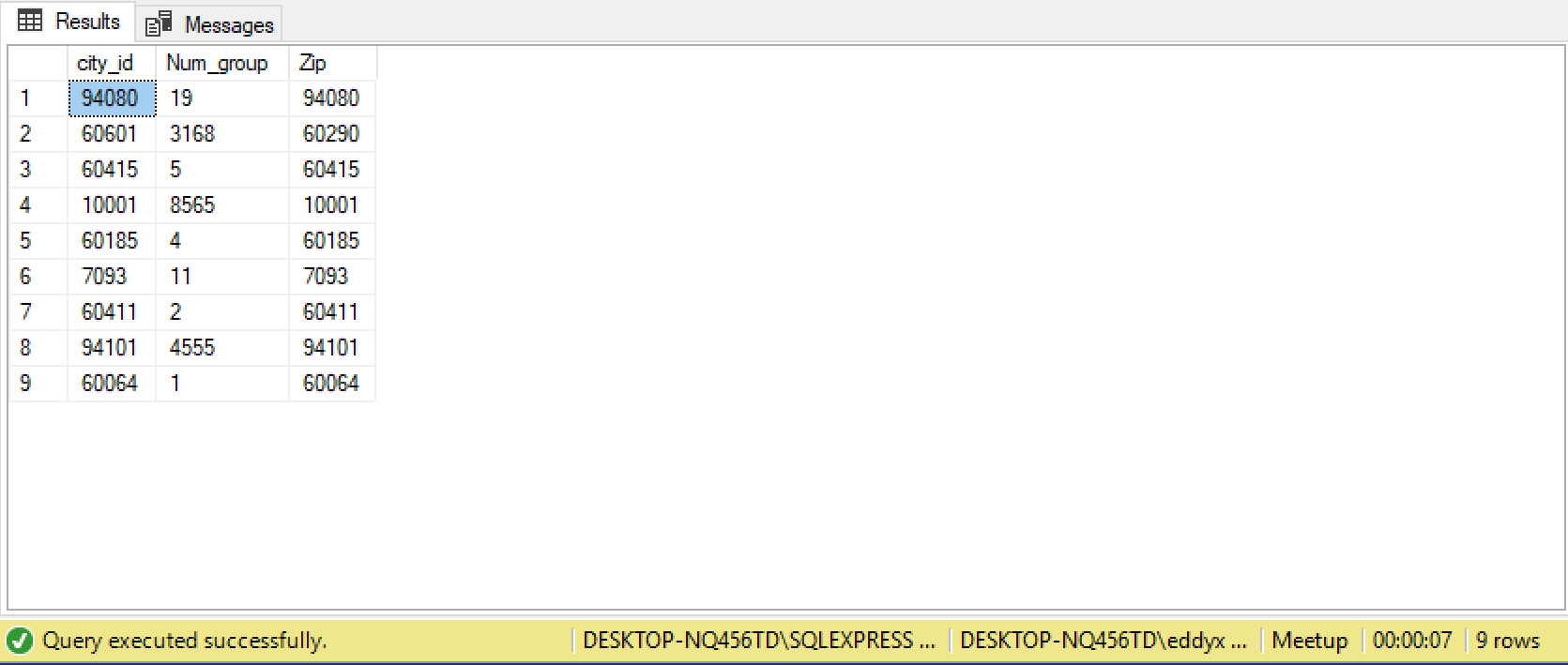
group by g.city\_id, c.zip

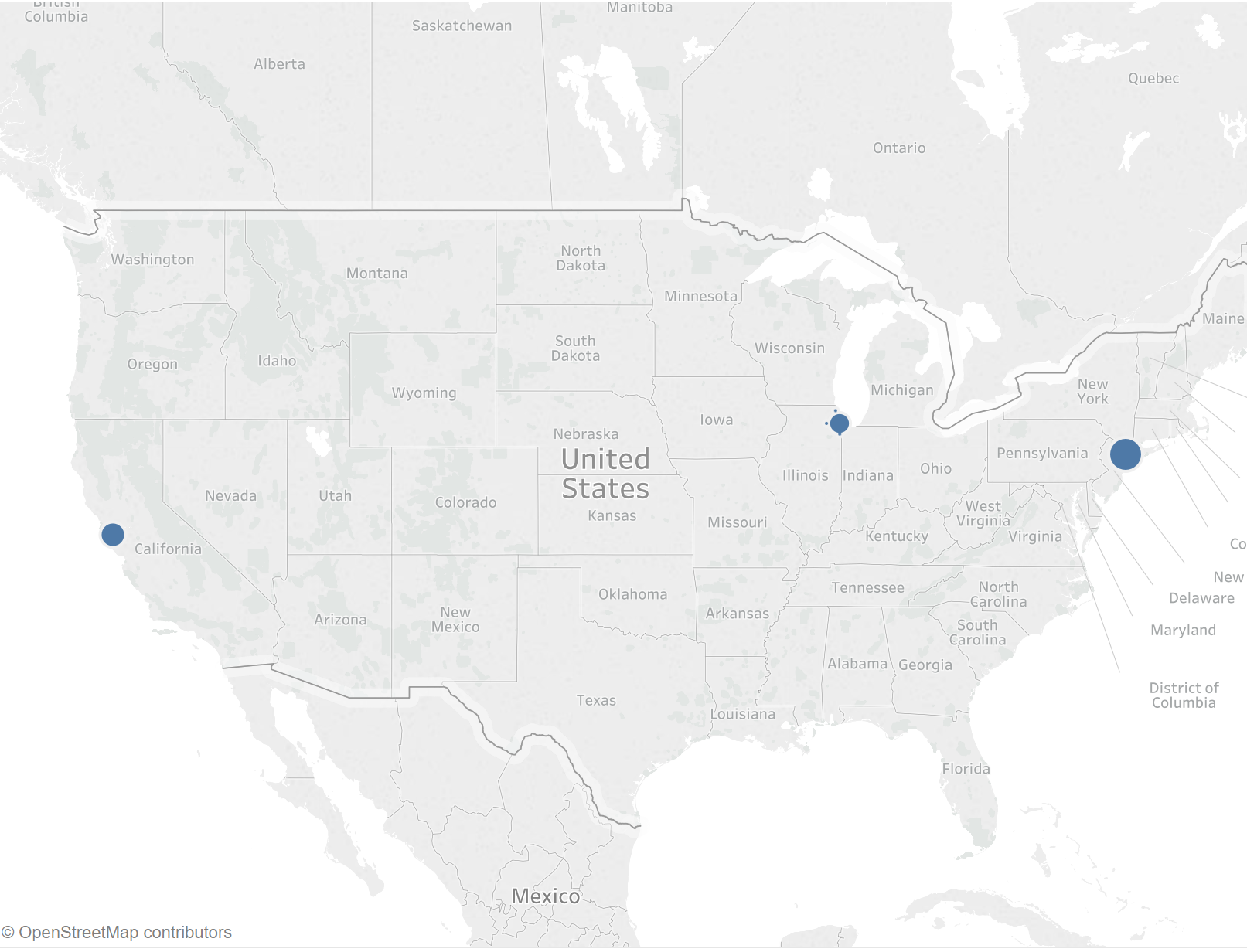
go

**Result:**

****

****

****

****

**Insight and application:**

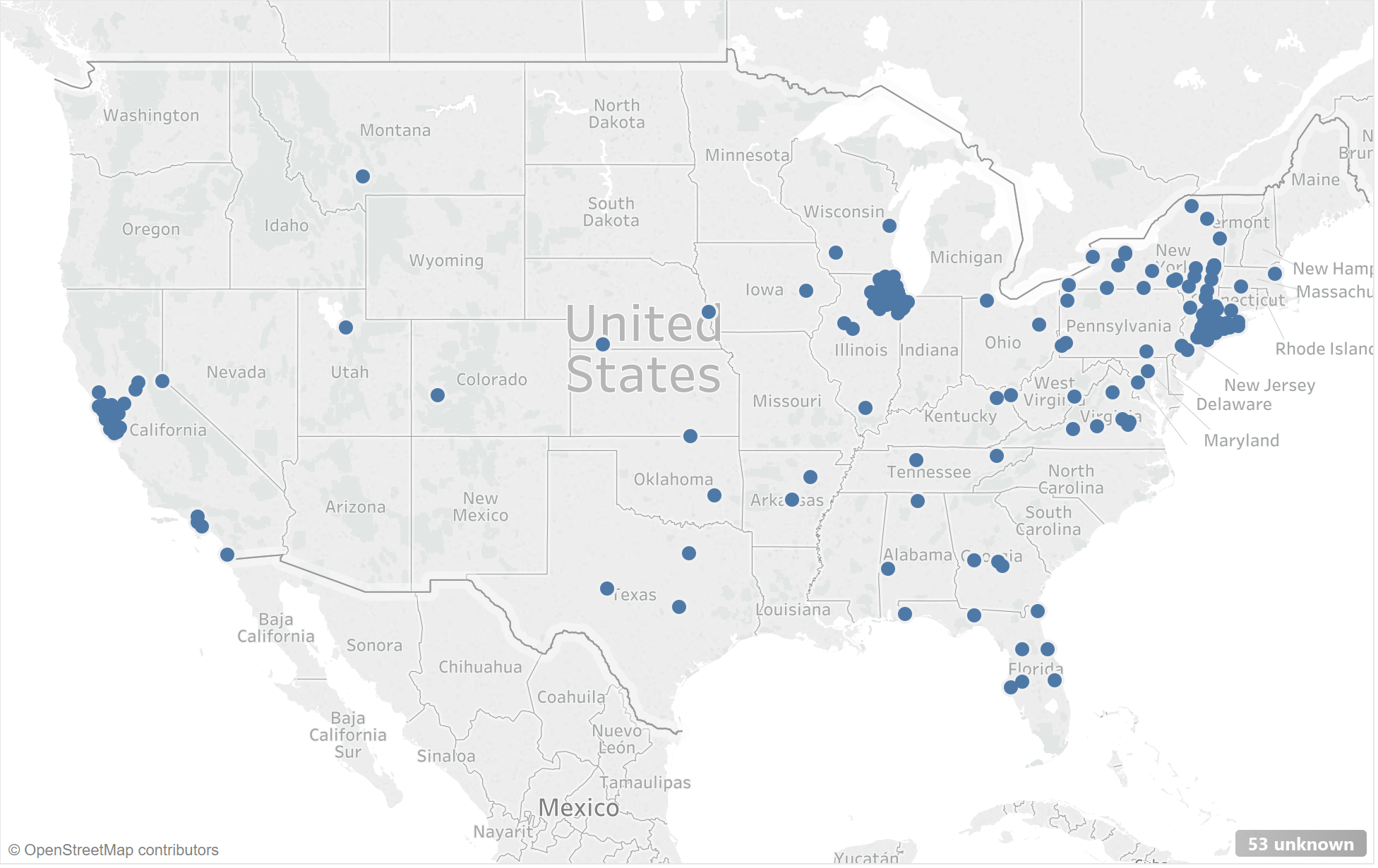
|  |  |  |  |
| --- | --- | --- | --- |
|  | New York | Chicago | San Francisco |
| Number of events | 1,267 | 1,841 | 1,320 |
| Recency (days) | 89.4 | 195.6 | 125.0 |
| Number of groups | 8,565 | 3,168 | 4,555 |

Chicago has the most events held, but with the least active level among these three cities. New York is on the other side. Although its number of events is the lowest, its events are most active. San Francisco is in the middle in both event numbers and active level.

To activate the ‘sleeping’ groups in the Chicago and to develop more groups in the New York City.

**Question 7: What is the venue distribution**

**Result:**



**Insight and application:**

Most of the events are held around the three cities. However, there are also some events held in the southeast.

**Recommendations:**

Meetup can consider opening another branch in the Florida.

**Question 8: The number of group created month-by-month, weekday-by-weekday for every year (2002-2017)**

**Queries:**

drop view if exists dbo.Group\_CreatedTime

Go

create view Group\_CreatedTime as

select month(g.created) as MOY, c.dow as DOW,

sum(case when year(g.created) = 2002 then 1 else 0 end) as n2002,

sum(case when year(g.created) = 2003 then 1 else 0 end) as n2003,

sum(case when year(g.created) = 2004 then 1 else 0 end) as n2004,

sum(case when year(g.created) = 2005 then 1 else 0 end) as n2005,

sum(case when year(g.created) = 2006 then 1 else 0 end) as n2006,

sum(case when year(g.created) = 2007 then 1 else 0 end) as n2007,

sum(case when year(g.created) = 2008 then 1 else 0 end) as n2008,

sum(case when year(g.created) = 2009 then 1 else 0 end) as n2009,

sum(case when year(g.created) = 2010 then 1 else 0 end) as n2010,

sum(case when year(g.created) = 2011 then 1 else 0 end) as n2011,

sum(case when year(g.created) = 2012 then 1 else 0 end) as n2012,

sum(case when year(g.created) = 2013 then 1 else 0 end) as n2013,

sum(case when year(g.created) = 2014 then 1 else 0 end) as n2014,

sum(case when year(g.created) = 2015 then 1 else 0 end) as n2015,

sum(case when year(g.created) = 2016 then 1 else 0 end) as n2016,

sum(case when year(g.created) = 2017 then 1 else 0 end) as n2017

from Groups G, Calendar C

where day(g.created) = day(c.date) and month(g.created) = month(c.date) and year(g.created)=year(c.date)

group by month(g.created), c.dow

Go

select \* from Group\_createdTime

order by moy, case

when dow='Mon' then 1

when dow='Tue' then 2

When dow='Wed' then 3

When dow='Thu' then 4

When dow='Fri' then 5

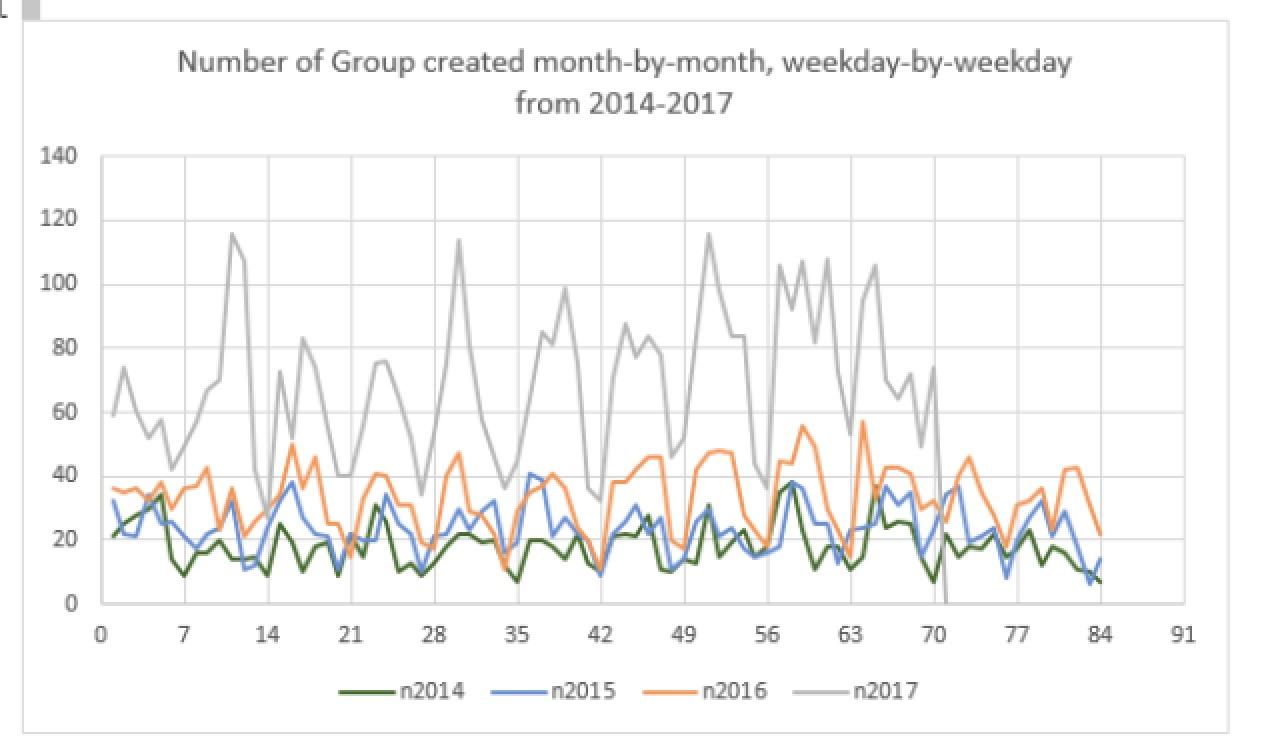
When dow='Sat' then 6

When dow='Sun' then 7 end

go

**Result:**





**Insight and application:**

Basically, the number of group created has increased over the year. However, number of group created have significantly increased in 2017 in compared with that in 2016, 2015 and 2014.

Usually groups will be created in the middle of the week such as Tuesday, Wednesday and Thursday. The number of groups created are significantly dropped in weekend and this insight holds true for all the years.

**Question 9: What are the number of time a member joined a group month-by-month, weekday-by-weekday for every year (2002-2017)?**

**Queries:**

drop view if exists dbo.member\_joinedtime

go

create view Member\_joinedtime as

select month(gm.joined) as MOY, c.dow as DOW,

sum(case when year(gm.joined) = 2002 then 1 else 0 end) as n2002,

sum(case when year(gm.joined) = 2003 then 1 else 0 end) as n2003,

sum(case when year(gm.joined) = 2004 then 1 else 0 end) as n2004,

sum(case when year(gm.joined) = 2005 then 1 else 0 end) as n2005,

sum(case when year(gm.joined) = 2006 then 1 else 0 end) as n2006,

sum(case when year(gm.joined) = 2007 then 1 else 0 end) as n2007,

sum(case when year(gm.joined) = 2008 then 1 else 0 end) as n2008,

sum(case when year(gm.joined) = 2009 then 1 else 0 end) as n2009,

sum(case when year(gm.joined) = 2010 then 1 else 0 end) as n2010,

sum(case when year(gm.joined) = 2011 then 1 else 0 end) as n2011,

sum(case when year(gm.joined) = 2012 then 1 else 0 end) as n2012,

sum(case when year(gm.joined) = 2013 then 1 else 0 end) as n2013,

sum(case when year(gm.joined) = 2014 then 1 else 0 end) as n2014,

sum(case when year(gm.joined) = 2015 then 1 else 0 end) as n2015,

sum(case when year(gm.joined) = 2016 then 1 else 0 end) as n2016,

sum(case when year(gm.joined) = 2017 then 1 else 0 end) as n2017

from Group\_member GM, Calendar C

where day(gm.joined) = day(c.date) and month(gm.joined) = month(c.date) and year(gm.joined)=year(c.date)

group by month(gm.joined), c.dow

go

select \* from member\_joinedtime

order by moy, case

when dow='Mon' then 1

when dow='Tue' then 2

When dow='Wed' then 3

When dow='Thu' then 4

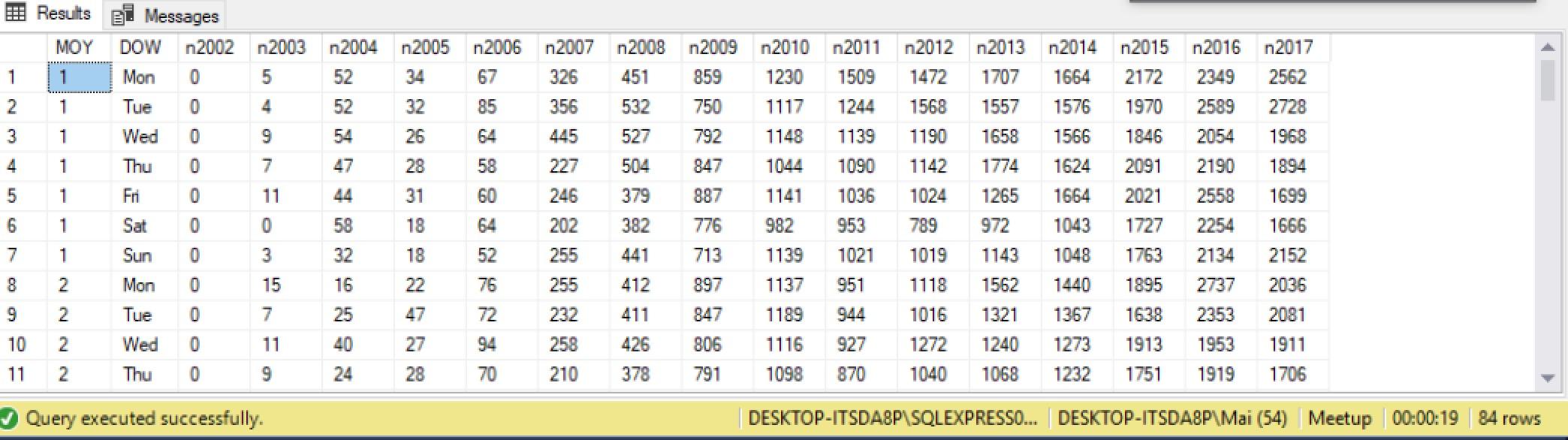
When dow='Fri' then 5

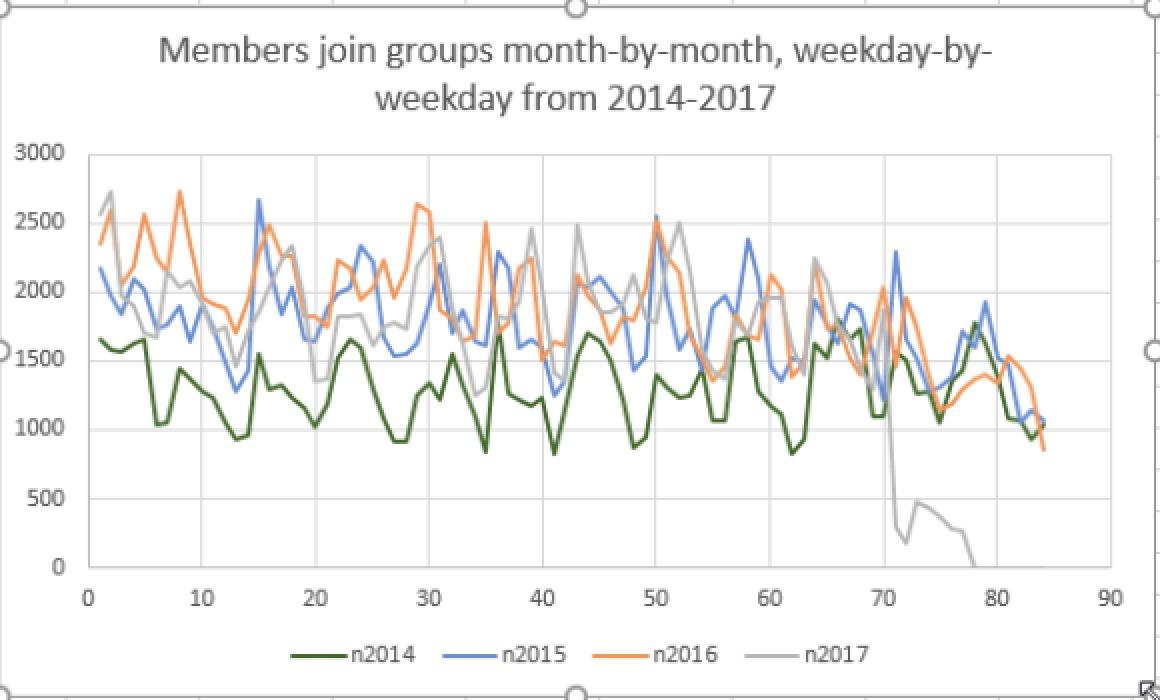
When dow='Sat' then 6

When dow='Sun' then 7 end

go

Result:



****

**Insight and Application:**

Compared this graph with the graph in question 8: Although the number of group created in 2017 have increased significantly, but the number of time a member join a group in 2017 do not change much in comparison with that in 2016 and 2015. Therefore, though there are a great number of new group, the quality of these groups may not be enough to attract new members. As we mentioned in the introduction, Meetup is associated with network effect, in which member engagement affects group and event creation by organizers and vice versa. Therefore, it is important for the platform to have a balance between these two.

Meetup may need to take action to raise the time a member joining a group to align with the a bigger number of group created. Based on what we learned in Question 5 and 6, Meetup can develop more groups and events that are more relevant to members' interests.

Again, members usually join group in the middle of the week and the number drops considerably in weekend. This phenomenon is aligned with the graph in 8 and demonstrates that Meetup members are very inactive in weekend.

**III – Conclusion**

Meetup is a platform for groups of interest and event organization. From the analysis result, it is clear that the majority of events on Meetup are about Technology, Career & Business and Socializing. However, individual interest shown via topics are more diverse and not really aligned with the individual interests shown via group categories. Our recommendation for Meetup is to maintain their “core business” in Technology, Career & Business and Socializing, keeping what they’re doing well. At the same time, they can develop the platform toward more diversified interests, catering for each big cities since different cities have certain differences in interests.