Databases

Luca Pernigo

Università della Svizzera italiana[1]

Course Test 4.18: Mail Marketing
May 25, 2023

Outline

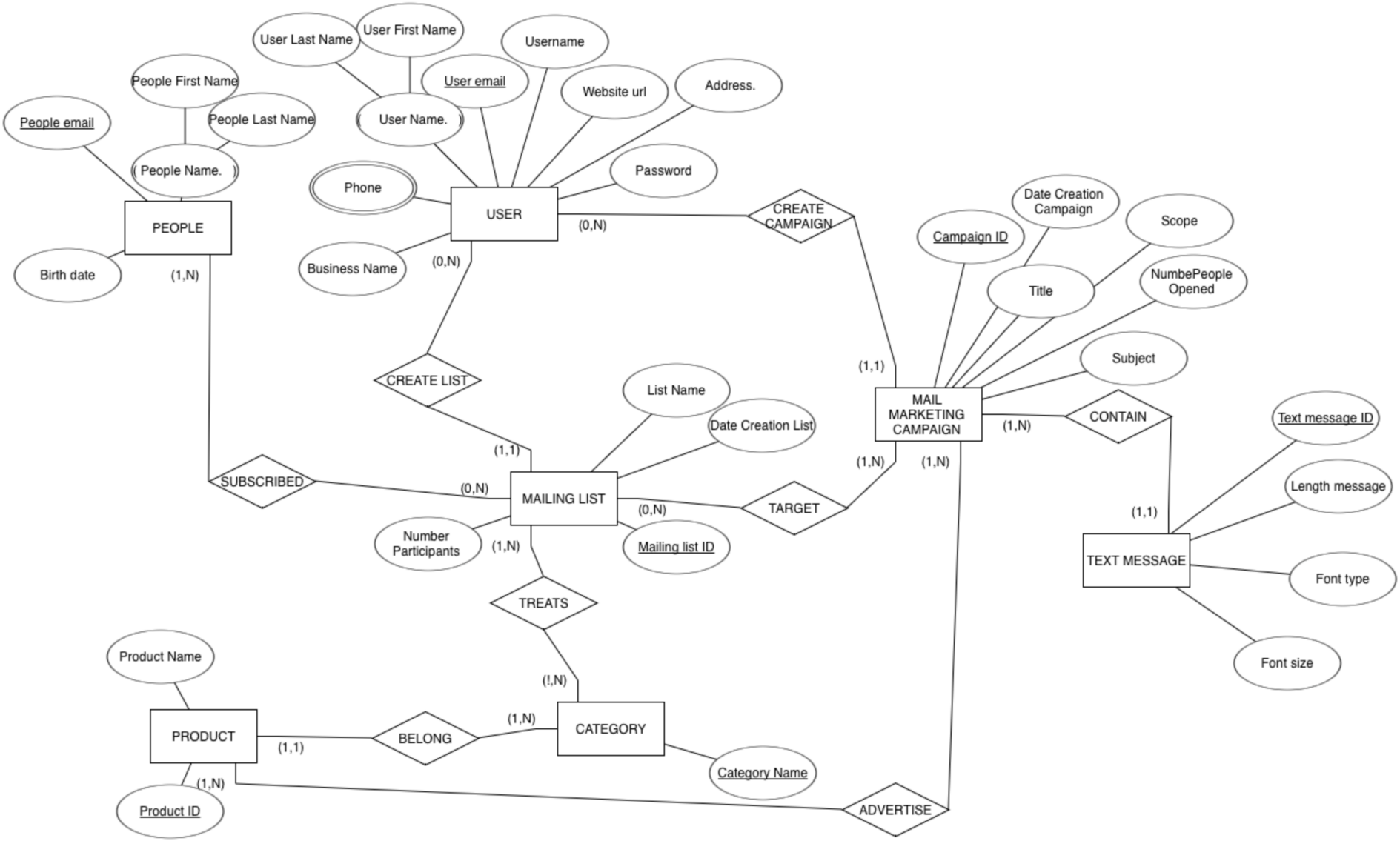
Table of Contents

Database description

Assume you need to create a database for a company offering a mail marketing tool. The database should have users of the tools who can create mail marketing campaigns. A mail marketing campaign can have a scope. A campaign is made of one or more mail text messages that are sent to the targets. The targets can be one or more mailing lists created by the users of the tool. A mailing list is simply a group of one or more mail addresses of people who agreed to receive updates about some products. The products about which they agreed to receive updates should also be stored since they are used to retrieve the possible targets of a campaign.

Table of Contents

ER/EER


User Last Name
User First Name
Username
People First Name
User email
Website url
Address.
People email
People Last Name
User Name.
People Name.
Password
Phone
USER
PEOPLE
CREATE CAMPAIGN
(0,N)
Date Creation Campaign
Scope
Campaign ID
Title
NumbePeople Opened
Birth date
(1,N)
Business Name
(0,N)
Subject
CREATE LIST
List Name
(1,1)
MAIL MARKETING CAMPAIGN
(1,1)
CONTAIN
(1,N)
Text message ID
Date Creation List
(1,1)
(1,N)
(1,N)
SUBSCRIBED
(0,N)
MAILING LIST
TARGET
(0,N)
Length message
(1,1)
Number Participants
(1,N)
Mailing list ID
TEXT MESSAGE
Font type
TREATS
Font size
Product Name
(!,N)
PRODUCT
BELONG
(1,N)
CATEGORY
(1,1)
Category Name
(1,N)
Product ID
ADVERTISE

Motivation Assumptions

- Each PEOPLE is uniquely identified by its email because people cannot have the same email address; that is all mail providers do not allow the registration of a new email if this email is already registered with their domain. Furthermore, assume that the field First Name and Last Name are mandatory fields when a PEOPLE instance wants to sign into a MAILING LIST. The birthdate is not mandatory, but the USER designing the mailing list can choose to ask for it in the sign up form of the mailing list; for example the USER might want to wish happy birthday to their subscribers.

Motivation Assumptions

- When a USER registers for the mail marketing tool, is asked for an User email, an Username, a Password, a Business Name their First Name, Last Name and for an Address. All these fields are mandatory for the USER if he/she wants to use the mail marketing tool. Note that all mail marketing tools ask for an Address, this follows from the fact that the Address field is mandatory by the International Anti Spam Laws. To give you an example consider the following link that points to the website of one of the most popular mail marketing tool (MailChimp). There they explain why they require their USERs to provide a valid Address.
https://mailchimp.com/it/help/anti-spam-requirements-for-email/

Motivation Assumptions

- USER-MAIL MARKETING CAMPAIGN Relationship(CREATE CAMPAIGN).
  On the USER side we have (0,N), because each USER can create N of MARKETING CAMPAIGNs. Additionally, consider the case of a USER that has just registered himself to the website where the mail marketing tool is hosted, in this case the new USER has not created any campaign yet, thus we have 0.
  On the MAIL MARKETING CAMPAIGN we have (1,1), because each MARKETING CAMPAIGN is created by a single USER.

Motivation Assumptions

- MAIL MARKETING CAMPAIGN-MAIL TEXT MESSAGE Relationship(CONTAIN).
  On the MAIL MARKETING CAMPAIGN we have (1,N), this can be easily inferred from the requirements that our client specified within the database description of the project: "A campaign is made of one or more mail text messages that are sent to the targets, line 3".
  On the MAIL TEXT MESSAGE, I assume that the ID of the TEXT MESSAGEs is the concatenation between the MAIL MARKETING CAMPAIGN ID to which it belongs and the text number within the MAIL MARKETING CAMPAIGN. So for example consider a CAMPAIGN with ID=NVT, composed by a signle TEXT MESSAGE. Then the ID of the TEXT MESSAGE will be NVT-1. Follwing this assumption, we have as cardinality on the MAIL TEXT MESSAGE (1,1).

Motivation Assumptions

- MAIL MARKETING CAMPAIGN-MAILING LIST Relationship(TARGET).
  On the CAMPAIGN side we have (1,N), a single CAMPAIGN may target more than one MAILING LIST and it has to target at least one MAILING LIST, this follows from the fact that when you create a CAMPAIGN, the mail marketing tool will ask you to specify at least one MAILING LIST in order to proceed with the creation of the CAMPAIGN.
  On the LIST side we have (0,N), Because first you create a MAILING LIST and then you create a CAMPAIGN targeting that specific LIST. Some time may pass before you create a CAMPAIGN for a new LIST, thus the partial participation in the participation constraint. Moreover, it can happen that a MAILING LIST is the target of multiple MARKETING CAMPAIGNs.

Motivation Assumptions

- USER-MAILING LIST Relationship(CREATE LIST).
  On the USER side we have (0,N), since USERs can create how many LISTs they want and it can be that some USER has not created any MAILING LIST yet.
  On the LIST side we have (1,1), this because each LIST is created by one single USER.

Motivation Assumptions

- PEOPLE-MAILING LIST Relationship(SUBSCRIBED). On the PEOPLE side we have (1,N), because only the data of PEOPLE subscribed to at least one MAILING LIST are tracked.
  On the MAILING LIST side we have (0,N) because one LIST can have N but also 0 subscribers, consider the case when a new MAILING LIST has just been created and no PEOPLE have been yet added to it or signed up to it.

Motivation Assumptions

- PEOPLE-MAILING LIST Relationship(SUBSCRIBED). On the PEOPLE side we have (1,N), because only the data of PEOPLE subscribed to at least one MAILING LIST are tracked.
  On the MAILING LIST side we have (0,N) because one LIST can have N but also 0 subscribers, consider the case when a new MAILING LIST has just been created and no PEOPLE have been yet added to it or signed up to it.

Motivation Assumptions

- MAILING LIST-CATEGORY Relationship(TREATS).
  It might be the case that a MAILING LIST treats various topics that are related but still different. For example consider a MAILING LIST that is about health and wellbeing, this LIST may have as topics healthy recipes and sports gear. For this reason we store information about these different topics in the CATEGORY relation.
  On the MAILING LIST side we have (1,N) because each MAILING LIST is about something and each MAILING LIST can treat more than one single topic.
  On the CATEGORY side we have (1,N) because only CATEGORIES that are treated by at least one MAILING LIST are stored in the database.
  Additionally, the same CATEGORY can be treated by different MAILING LISTs.

Motivation Assumptions

- PRODCUT-CATEGORY Relationship(BELONG). On the Product side we have (1,1) because each product can belong just to one CATEGORY. On the CATEGORY side we have (1,N) because N PRODUCTs can belong to the same CATEGORY. Furthermore, we store data about CATEGORY, only if there is at least one PRODUCT that belongs to such CATEGORY.

Motivation Assumptions

- MAIL MARKETING CAMPAIGN-PRODUCT Relationship(ADVERTISE).
  On the CAMPAIGN side we have (1,N), since each MAILING CAMPAIGN can advertise one or more products at the same time.
  On the PRODUCT side we have (1,N). On the minimum participation constraint we have total participation because our database will store only the PRODUCT that are advertised by at least one CAMPAIGN. On the other hand, on the maximum participation constraint we have N, since the same PRODUCT may be advertised by multiple CAMPAIGNs.

Motivation Assumptions

- The Phone attribute is multivalued since some USER may have more than one single phone registered in the mail marketing tool.

Motivation Assumptions

- The Scope attribute of the MAIL_MARKETING_CAMPAIGN relation is intended as the type of audience for which a particular CAMPAIGN is designed for. In case of a CAMPAIGN with no particular type of audience the scope attribute is filled with NULL.

Table of Contents

Relational


PEOPLE
Pemail
PFirstName
PLastName
Birth date
USER
Uemail
Username
Password
Address
BusinessName
Website url
UFirstName
ULastName
MAIL_MARKETING_CAMPAIGN
Campaign ID
Uemail
DateCreationCampaign
Title
Subject
Scope
NumberPeopleOpened
MAILING LIST
Mailing list ID
Uemail
DateCreationList
ListName
NumberParticipants
TEXT MESSAGE
Text Message ID
Campaign ID
Length message
Font type
Font size
CATEGORY
CategoryName
PRODUCT
Product ID
ProductName
CategoryName
TARGET
Campaign ID
Mailing list ID
SUBSCRIBED
Pemail
Mailing list ID
PHONES
Uemail
Phone
TREAT
Mailing list ID
CategoryName
ADVERTISE
Campaign ID
Product ID

Table of Contents

1NF

To achieve 1NF, it is necessary to rearrange the database design such that no composite or multivalued attributes are used in the database. To turn the mail marketing database into its 1NF form two steps have been performed.

- Since Phone in the USER relation is multivalued, each tuple of this relation has to have in its Phone column one telephone number only. USERs with N phones will have N tuples, one for each of the telephone numbers they registered in their mail marketing account.
- Since the attribute Address in USER is not atomic, it is decomposed into its component attributes: City, Zip and Country.

2NF

- PEOPLE
  Upon regristration each PEOPLE is always required to insert its Pemail along with his/her PFirstName and PLastName; in this setting we thus have that Pemail identifies PFirstName and PLastName but not Birthdate, therefore the PEOPLE relation is split in the following two relations.

Pemail	PFirstName	PLastName

Pemail	Birthdate

2NF

- USER
  In the USER relation everything is identified by the Uemail, because all its attributes have to be specified upon registration time on the mail marketing tool website. Therefore the USER relation is already in 2NF and for this reason is left as it is.

Uemail	Username	Password	City	Zip	Country	BusinessName	Website url

UFirstName	ULastName

2NF

- MAIL MARKETING CAMPAIGN
  When an USER creates a CAMPAIGN it has to specify a TITLE and a SUBJECT while its Uemail and DateCreationCampaign are automatically retrieved by the mail marketing tool. Hence, those attributes are fully functional dependent on the primary key Campaign ID. On the contrary the attribute NumberPeopleOpened and Scope do not depend on the Campaign ID. For these reasons, the CAMPAIGN relation is broken in the two relations that follow.

Campaign ID	Uemail	DateCreationCampaign	Title	Subject

Campaign ID	Scope	NumberPeopleOpened

2NF

- MAILING LIST
  In the MAILING LIST relation the dependent attributes on the primary key are Uemail, DateCreationList and ListName, while NumberParticipants is independent on the primary key Mailing list ID. Thus, in order to turn the MAILING LIST relation to its 2NF, it is split into two relations.

Mailing list ID	Uemail	DateCreationList	ListName

Mailing list ID	NumberParticipants

2NF

- TEXT
  In the TEXT MESSAGE relation every nonprime attribute is fully dependent on the primary key Text Message ID, thus this relation is already in 2NF.

Text Message ID	Campaign ID	Length message	Font type	Font size

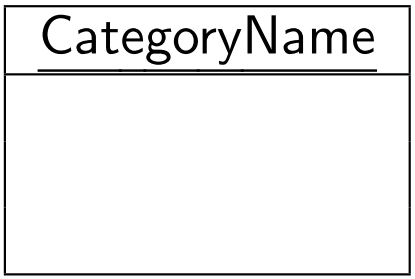
2NF

- PRODUCT
  In the PRODUCT relation, every nonprime attribute is fully functional dependent on the primary key Product ID; ProductName is directly related to Product ID, while CategoryName is indirectly related to Product ID through ProductName(assuming that given a ProductName it is trivial to retrieve to which CATEGORY this PRODUCT belongs). Thus, the PRODUCT relation is already in 2NF.

Product ID	ProductName	CategoryName

2NF

- CATEGORY
  This relation is composed by only its primary key CategoryName, therefore it is already in 2NF.



2NF

The following relations consist all of two attributes that together form composite keys. For this reason TARGET, SUBSCRIBED, PHONE, TREAT and ADVERTISE are all already in 2NF.figurefigure

- TREAT

Mailing list ID	CategoryName

- ADVERTISE

Campaign ID	Product ID

- TARGET

Campaign ID	Mailing list ID

- SUBSCRIBED

Pemail	Mailing list ID

- PHONE

Uemail	Phone

3NF

- PRODUCT
  CategoryName is related indirectly to Product ID through ProductName. Hence, there is a transitive functional dependency between Product Name and Category Name, therefore the Category ID is introduced and the table is split in two.

Product ID	Product Name	Category ID

Category ID	CategoryName

3NF

- TREAT
  After the above step we have a CATEGORY table, whit primary key Category_ID. Hence, the TREAT table is modified by switching the CatgoryName attribute with the Category_ID attribute. The new TREAT table is reported hereafter.

Mailing list ID	Category_ID

Table of Contents

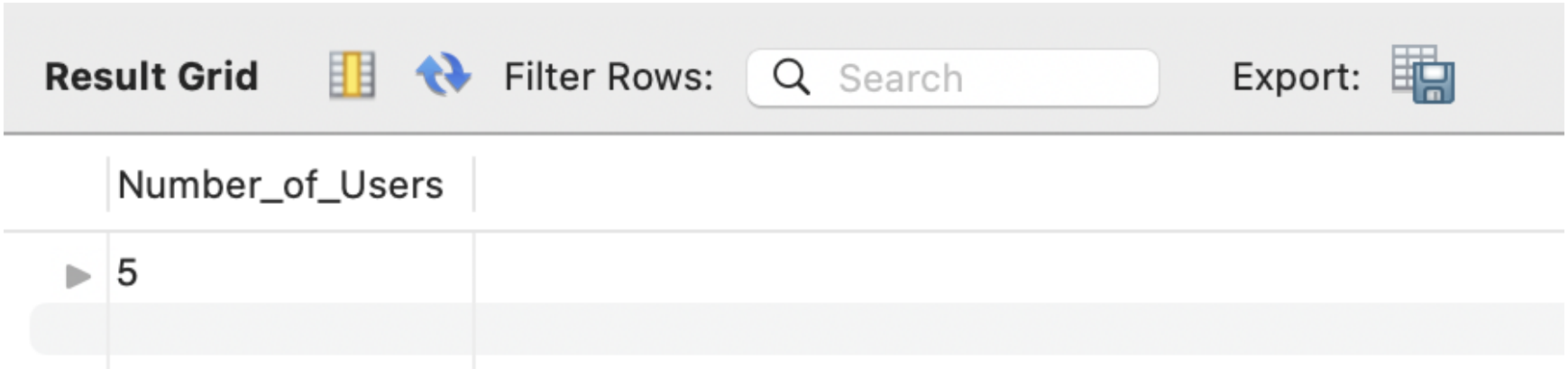
Queries

After the Relational Model design and normalization, the actual database has been implemented and data has been inserted, please refer to the attached MySQL export to see implementation details and the commands. In this section, some creative queries and their corresponding results are reported.

Query1

1 See how many users we have in our database

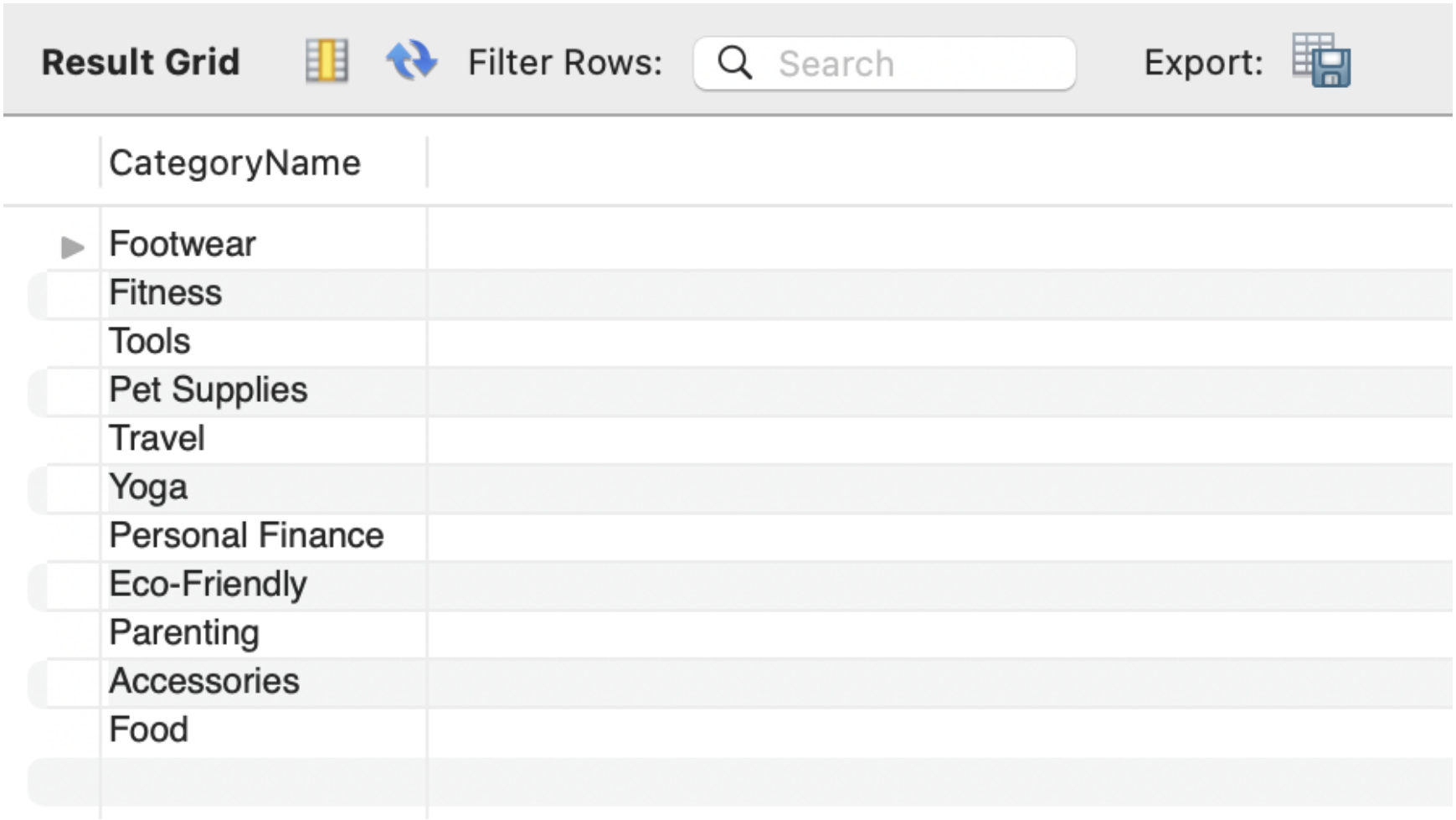
```
SELECT count(*) Number_of_Users FROM USER;
```



Query2

2 See categories of products in the database

```
SELECT DISTINCT CategoryName FROM CATEGORY;
```

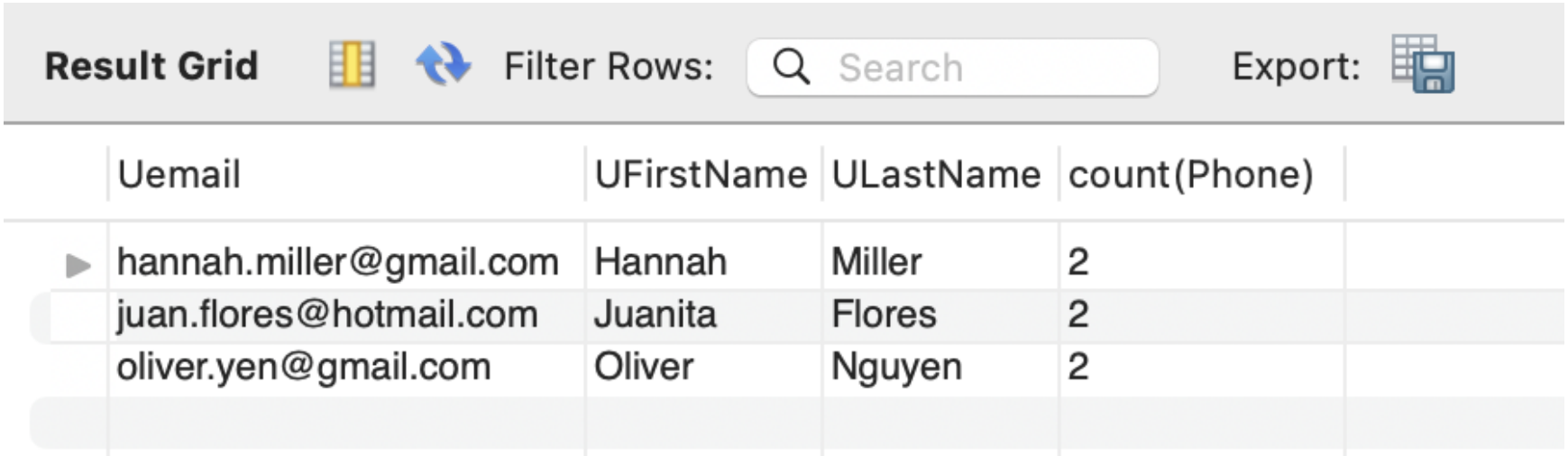


CategoryName
Footwear
Fitness
Tools
Pet Supplies
Travel
Yoga
Personal Finance
Eco-Friendly
Parenting
Accessories
Food

Query3

3 See which user has more than one phone

```
SELECT USER.Uemail, UFirstName, ULastName, count(Phone)
FROM USER JOIN PHONES ON USER.Uemail=PHONES.Uemail
GROUP BY Uemail
HAVING count(Phone)>1
;
```

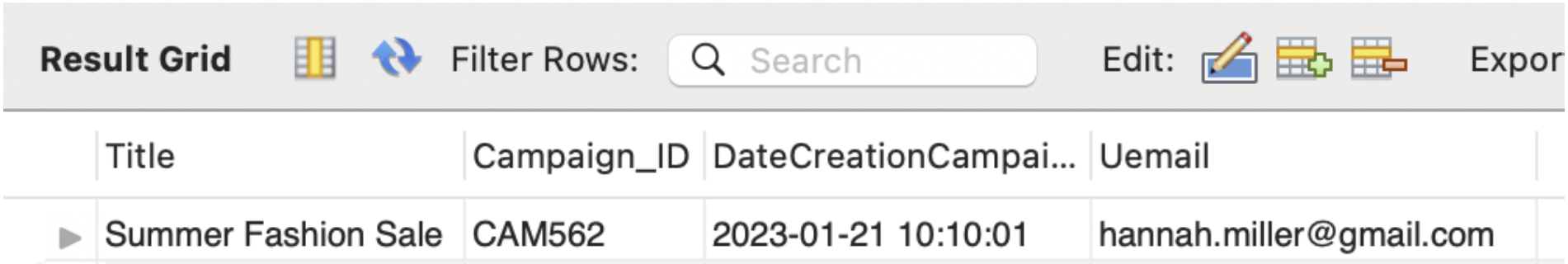


Uemail	UFirstName	ULastName	count(Phone)
hannah.miller@gmail.com	Hannah	Miller	2
juan.flores@hotmail.com	Juanita	Flores	2
oliver.yen@gmail.com	Oliver	Nguyen	2

Query4

4. Find the first campaign created on the mail marketing database

```
SELECT Title, Campaign_ID, DateCreationCampaign, Uemail FROM MAIL_MARKETING_CAMPAIGN
WHERE DateCreationCampaign=(SELECT MIN(DateCreationCampaign)  FROM MAIL_MARKETING_CAMPAIGN)
;
```



Title	Campaign_ID	DateCreationCampai...	Uemail
Summer Fashion Sale	CAM562	2023-01-21 10:10:01	hannah.miller@gmail.com

Query5

5. Find out how many subscribers each newsletter has

```
SELECT List_name, Subscribed
FROM (
SELECT List_name, count(Pemail) Subscribed
FROM SUBSCRIBED e JOIN  MAILING_LIST l ON e.List_ID=l.List_ID
GROUP BY List_name) as totsubs
ORDER BY Subscribed DESC
;
```

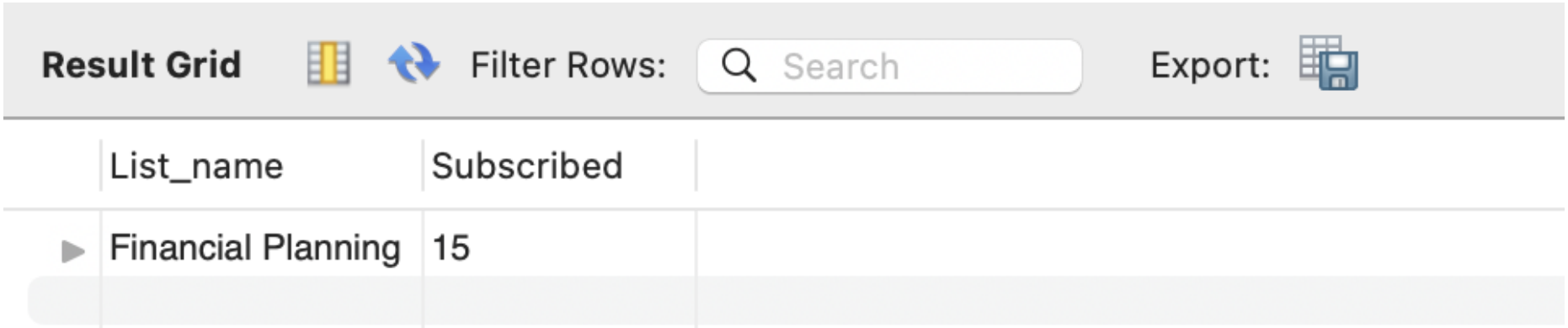
Result Grid   Filter Rows:  Search Export: 

List_name	Subscribed
Financial Planning	15
VIP Customers	14
Home Decor Ideas	14
Weekend Getaways	13
DIY Home Improvement	12
Saving for the Future	9
Healthy Living	9
Fashion Trends	8
E-Planet	8
Pet Owners	8
We Parents	6
Homeowners	5
Travel Enthusiasts	5
DIY Projects	4
Fashionistas	4
Foodies	4
Healthy Eating Tips	3
Happy Subscribers	3

Query6

6. Prompt newsletter with most subscribers

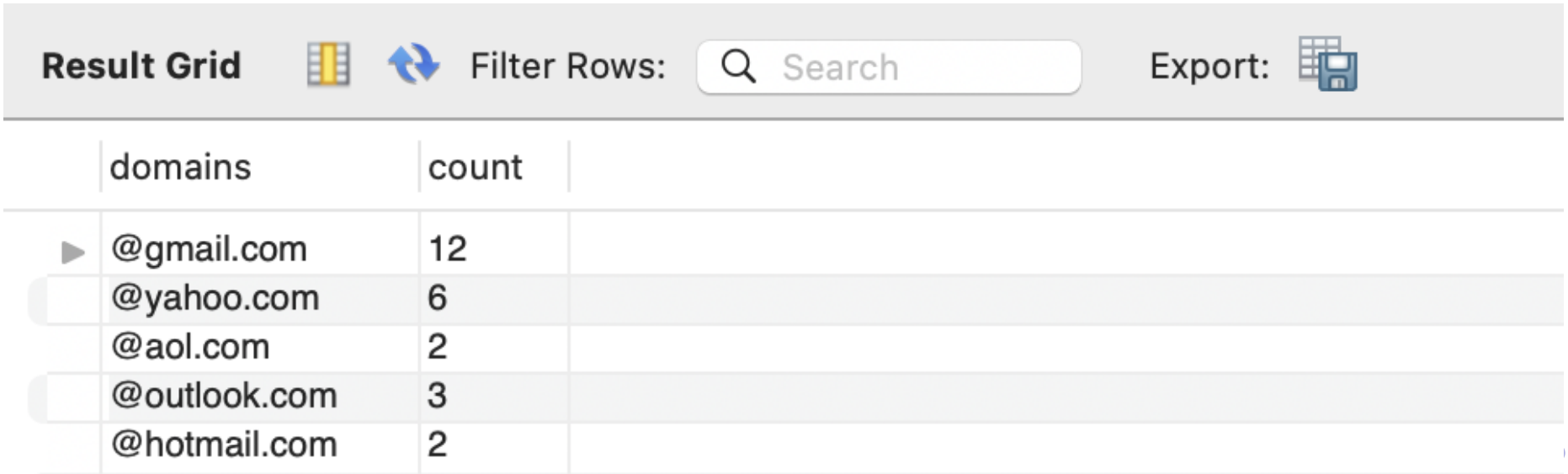
```
SELECT List_name, Subscribed
FROM (
SELECT List_name, count(Pemail) Subscribed
FROM SUBSCRIBED e JOIN  MAILING_LIST l ON e.List_ID=l.List_ID
GROUP BY List_name) as totsubs
WHERE Subscribed=(SELECT MAX(Subscribed)
FROM (SELECT List_name, count(Pemail) Subscribed
FROM SUBSCRIBED e JOIN  MAILING_LIST l ON e.List_ID=l.List_ID
GROUP BY List_name) as totsubs)
;
```



Query7

7. Find most common email domains in the database

```
SELECT
SUBSTRING(Pemail,
POSITION('@' IN Pemail)) as domains,
count(SUBSTRING(Pemail,
POSITION('@' IN Pemail))) as count
FROM PEOPLE_NAMES
GROUP BY SUBSTRING(Pemail,POSITION('@' IN Pemail))
;
```

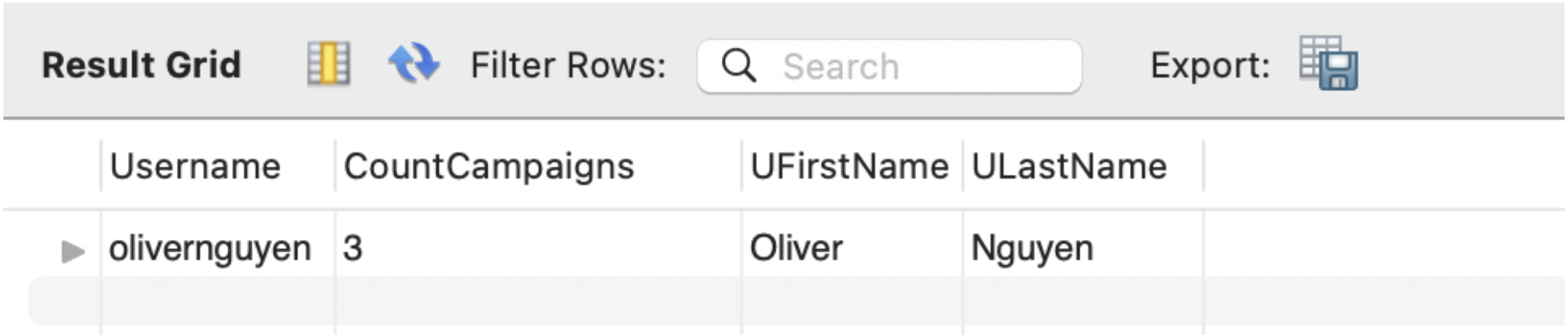


domains	count
@gmail.com	12
@yahoo.com	6
@aol.com	2
@outlook.com	3
@hotmail.com	2

Query8

8. Find user that created the most mail marketing campaigns

```
SELECT *
FROM
(SELECT Username, count(USER.Uemail) CountCampaigns, UFirstName, ULastName
FROM USER JOIN MAIL_MARKETING_CAMPAIGN ON USER.Uemail=MAIL_MARKETING_CAMPAIGN.Uemail
GROUP BY USER.Uemail) as countmail
WHERE CountCampaigns=(SELECT max(CountCampaigns)
FROM
(SELECT Username, count(USER.Uemail) CountCampaigns, UFirstName, ULastName
FROM USER JOIN MAIL_MARKETING_CAMPAIGN ON USER.Uemail=MAIL_MARKETING_CAMPAIGN.Uemail
GROUP BY USER.Uemail) as countmail2)
;
```

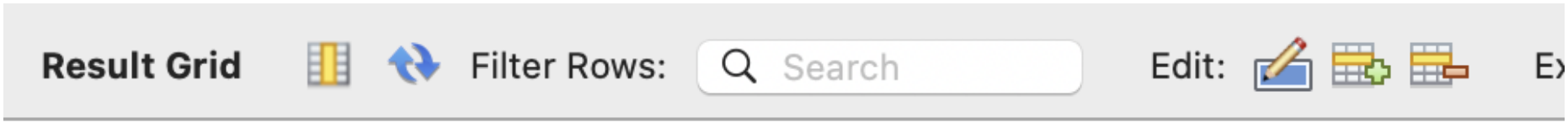


Username	CountCampaigns	UFirstName	ULastName
olivernguyen	3	Oliver	Nguyen

Query9

9. Find shortest text message

```
SELECT * FROM TEXT_MESSAGE
WHERE Lenght_message=(SELECT min(Lenght_message) FROM TEXT_MESSAGE)
;
```

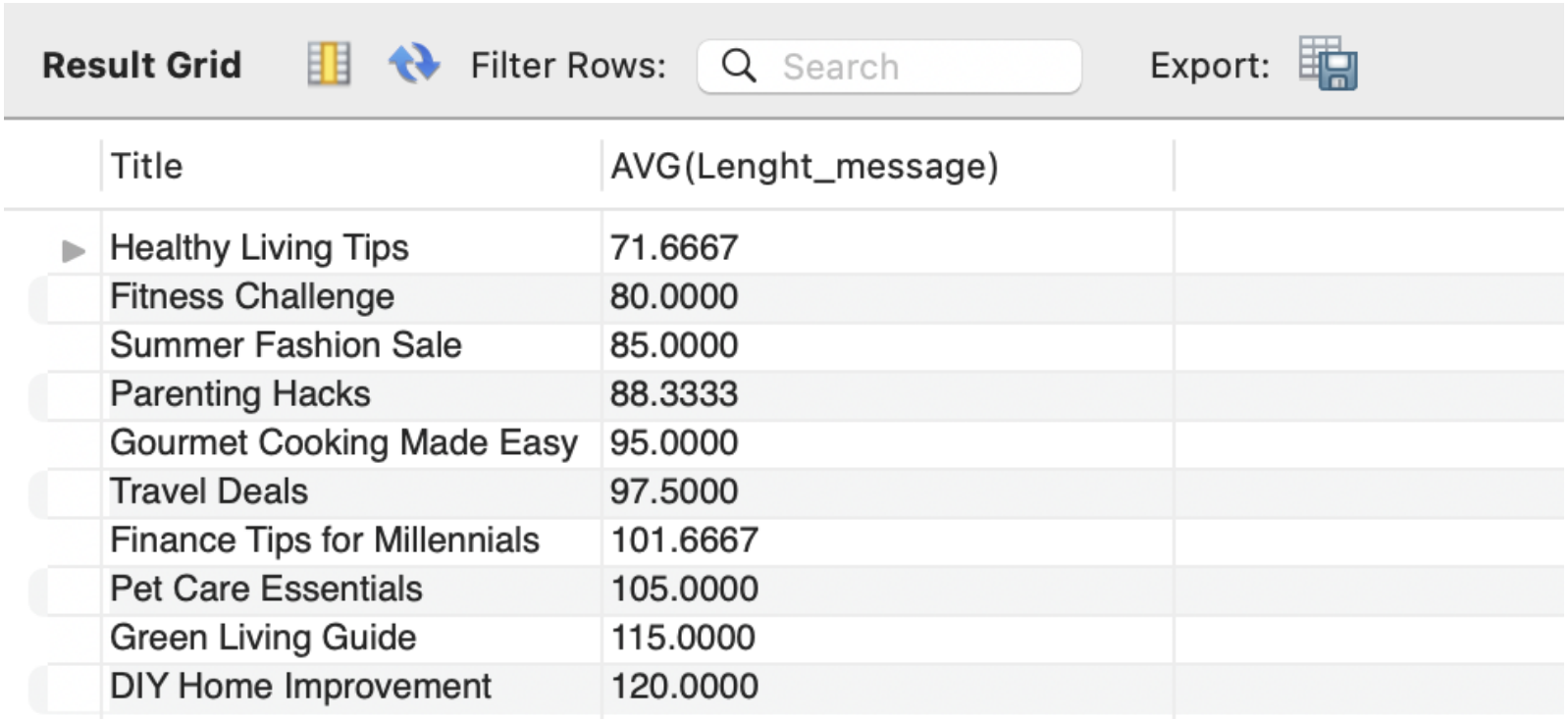


Message_ID	Campaign_ID	Lenght_message	Font_type	Font_size
MKD024-2	MKD024	70	Verdana	10
MKD024-3	MKD024	70	Verdana	10

Query10

10. Search average length of text message for campaign and order them by average length in ascending order

```
SELECT Title, AVG(Lenght_message) FROM MAIL_MARKETING_CAMPAIGN
JOIN TEXT_MESSAGE
  ON  MAIL_MARKETING_CAMPAIGN.Campaign_ID=TEXT_MESSAGE.Campaign_ID
GROUP BY Title
ORDER by AVG(Lenght_message);
```

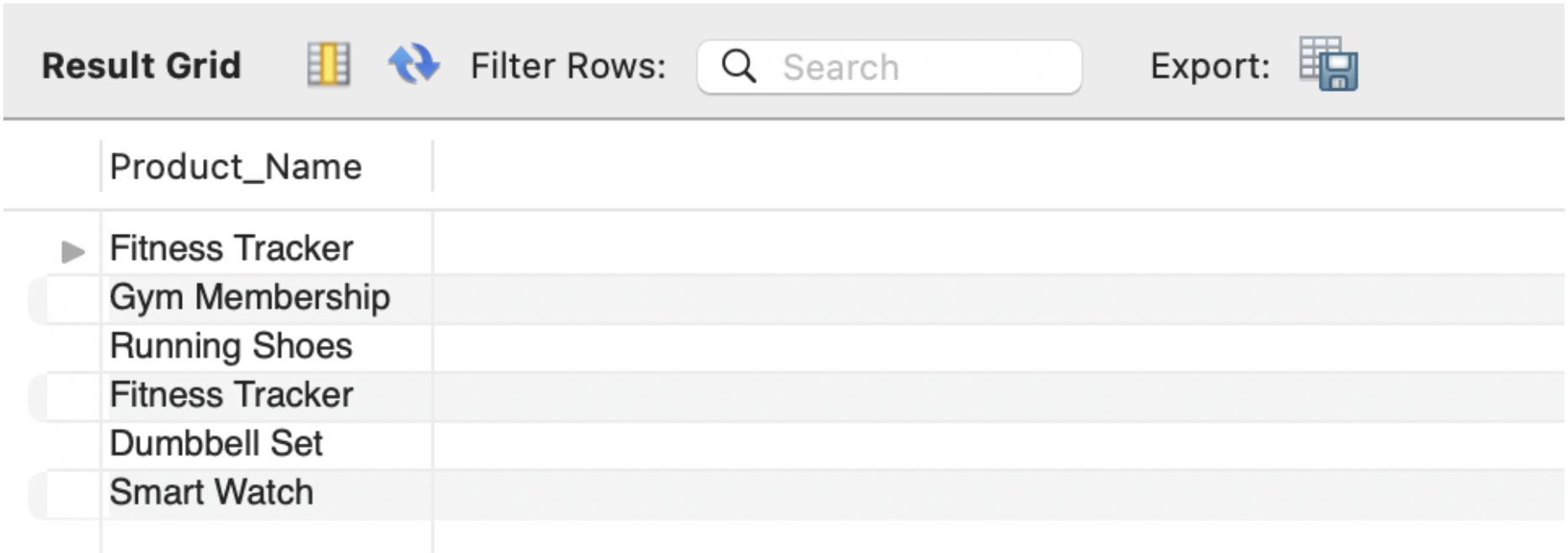


Title	AVG(Lenght_message)
Healthy Living Tips	71.6667
Fitness Challenge	80.0000
Summer Fashion Sale	85.0000
Parenting Hacks	88.3333
Gourmet Cooking Made Easy	95.0000
Travel Deals	97.5000
Finance Tips for Millennials	101.6667
Pet Care Essentials	105.0000
Green Living Guide	115.0000
DIY Home Improvement	120.0000

Query11

11. Inspect all products belonging to the category Fitness

```
SELECT Product_Name
FROM PRODUCT JOIN CATEGORY on PRODUCT.Category_ID=CATEGORY.CATEGORY_ID
WHERE CategoryName='Fitness';
```

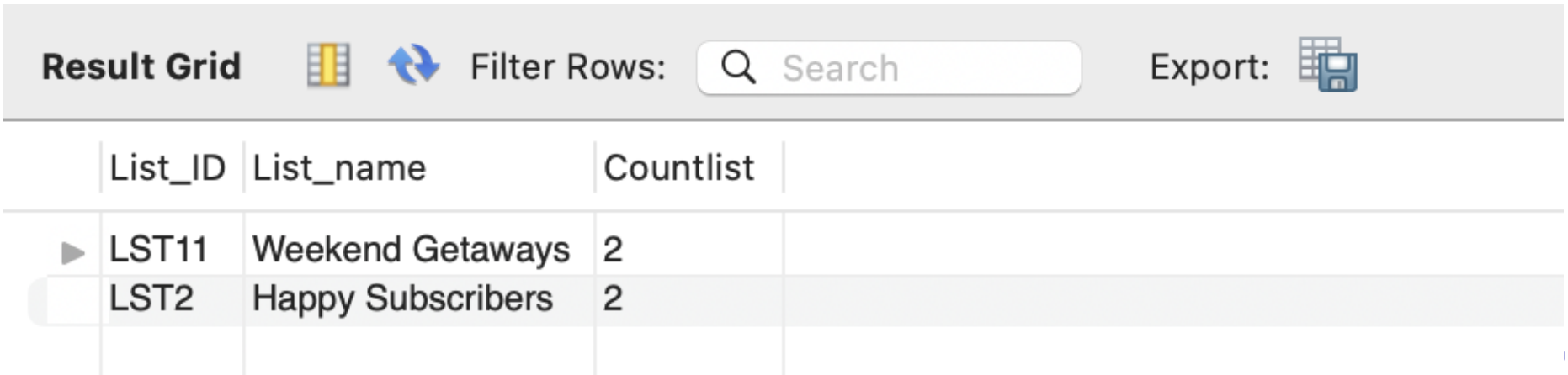


Product_Name
Fitness Tracker
Gym Membership
Running Shoes
Fitness Tracker
Dumbbell Set
Smart Watch

Query12

12. Select the most targeted mailing list, where Countlist corresponds to the numer of MAIL MARKETING CAMPAIGNs that target a LIST

```
SELECT List_ID,List_name, Countlist
FROM
(SELECT MAILING_LIST.List_ID,List_name, count(MAILING_LIST.List_ID) Countlist FROM TARGET
JOIN MAILING_LIST ON TARGET.List_ID=MAILING_LIST.List_ID
GROUP BY List_ID) as countlist
WHERE Countlist=(SELECT max(Countlist)
FROM
(SELECT List_ID, count(List_ID) Countlist FROM TARGET
GROUP BY List_ID) as countlist)
;
```

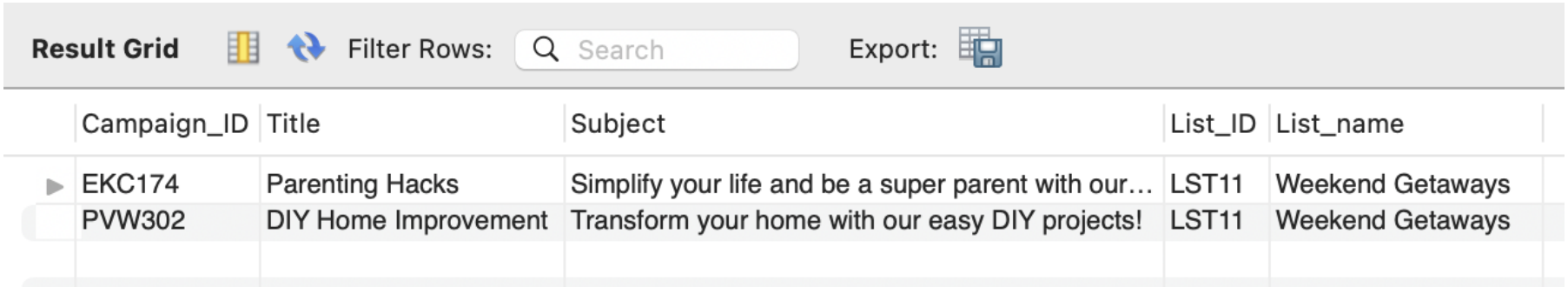


List_ID	List_name	Countlist
LST11	Weekend Getaways	2
LST2	Happy Subscribers	2

Query13

13. See the mailing campaigns that target Weekend Gateways

```
SELECT MAIL_MARKETING_CAMPAIGN.Campaign_ID, Title, Subject, MAILING_LIST.List_ID, List_name FROM TARGET
JOIN MAIL_MARKETING_CAMPAIGN ON TARGET.Campaign_ID=MAIL_MARKETING_CAMPAIGN.Campaign_ID
JOIN MAILING_LIST ON  TARGET.List_ID=MAILING_List.List_ID
WHERE List_name='Weekend Getaways'
;
```



Campaign_ID	Title	Subject	List_ID	List_name
EKC174	Parenting Hacks	Simplify your life and be a super parent with our…	LST11	Weekend Getaways
PVW302	DIY Home Improvement	Transform your home with our easy DIY projects!	LST11	Weekend Getaways

Query14

14. See the PEOPLEs that are subsribed to Weekend Gateways

```
SELECT Pemail
FROM SUBSCRIBED JOIN MAILING_LIST ON SUBSCRIBED.List_ID=MAILING_LIST.List_ID
WHERE List_name='Weekend Getaways'
;
```

Result Grid | Filter Rows: Search | Export:

Pemail
bay.klosterm@yahoo.com
careyne@yahoo.com
cornelianmalla@gmail.com
elgar_luken@aol.com
er.he@outlook.com
jutenniso@gmail.com
la_enoch@gmail.com
mark.johnson@outlook.com
penric.am@outlook.com
sarah.jackson@gmail.com
sherma.gran@gmail.com
ubu-senior@gmail.com
wyn.loes@yahoo.com

Thank You for your attention