

Dear Raj,

Hope you had a lovely weekend! It is my pleasure to help with decision making in our company's non-profit investment. Here are my answers for your questions:

Part 1 - Revenue

1. 2014 was a key year for generating revenue. How many Ann Arbor-based companies (rows) are listed in the database that year?

There are 247 Ann Arbor-based companies are listed in the database in 2014.

```
Database changed
mysql> select count(1) from taxdata where year = 2014 and city = 'Ann Arbor';
+-----+
| count(1) |
+-----+
|      247 |
+-----+
1 row in set (4.65 sec)
```

taxdata

Field Name	Data Type	Values	Note
id	int	id of the company	primary key
ein	int	ein assigned to the company	
name	varchar	name of the company	
year	int	taxyear	
revenue	int	revenue of the company	
expenses	int	expense of the company	
purpose	text	purpose of the company	
ptid	varchar	ptid of the company	
ptname	varchar	ptname of the company	
city	varchar	city where company located	
state	varchar	state where company in	
url	varchar	url of company	

2. To get a broader picture, we also need a list of the names of ALL of the companies that generated more than \$10,000,000,000 (ten billion dollars) in 2014.

```
mysql> select name from taxdata where year = 2014 and revenue>10000000000;
+-----+
| name                                     |
+-----+
| IHC HEALTH SERVICES INC                 |
| CENTRAL STATES SE & SW AREAS HEALTH & WELFARE F |
| Banner Health                           |
| KAISER FOUNDATION HEALTH PLAN INC       |
| Thrivent Financial for Lutherans        |
| UAW RETIREE MEDICAL BENEFITS TRUST      |
| KAISER FOUNDATION HOSPITALS             |
| Massachusetts Institute of Technology  |
| UPMC GROUP                             |
| Cornell University                     |
| Trustees of the University of Pennsylvania |
| President and Fellows of Harvard College |
| YALE UNIVERSITY                        |
| STATE EMPLOYEES' CREDIT UNION          |
| DUKE UNIVERSITY                       |
| TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA |
| Howard Hughes Medical Institute         |
| Partners HealthCare System Inc & AffiliatesGroup Return |
| THE BOARD OF TRUSTEES OF THE LELAND STANFORDJUNIOR UNIVERSITY |
| DIGNITY HEALTH                         |
+-----+
20 rows in set (2.18 sec)
```

3. How does this compare to how many unique companies generated a revenue of more than \$1,000,000,000 (one billion dollars) in any year?

There are 404 unique companies made revenue more than 1, 000, 000, 000 in any year.

```
mysql> select count(distinct name) from taxdata where revenue>1000000000;
+-----+
| count(distinct name) |
+-----+
| 404                  |
+-----+
1 row in set (0.43 sec)
```

Part 2 - Expenses

1. We are looking for the top 20 unique companies by expenses (as in: who had the most expenses?). Your query should provide the name of the company and their expenses in 2013.

```
mysql> select distinct name, expenses
-> from(select name, expenses , year from taxdata where year = 2013 order by expenses desc limit 20) as subquery;
```

name	expenses
KAISER FOUNDATION HEALTH PLAN INC	41982288055
UNIVERSITY OF ROCHESTER	2776536619
ST FRANCIS HOSPITAL AND MEDICAL CENTER	713077460
WELLSTAR HEALTH SYSTEM INC	660400156
SKY LAKES MEDICAL CENTER	488333644
NAPLES COMMUNITY HOSPITAL INC	426168844
YAKIMA VALLEY MEMORIAL HOSPITAL	401985865
EMBRY-RIDDLE AERONAUTICAL UNIVERSITY INC	374154000
ST MARY'S HOSPITAL & MEDICAL CENTER INC	372041256
St Barnabas Hospital	349321637
HOSPITAL COMMITTEE FOR THE LIVERMORE/PLEASANTON AREA	341257233
VALLEY PRESBYTERIAN HOSPITAL	273789797
WINCHESTER HOSPITAL	264605951
TEAMSTERS WESTERN REGION AND NEW JERSEY HEALTH CARE FUND	231059337
OROVILLE HOSPITAL	200242937
ST JOSEPH HOSPITAL OF EUREKA	196999313
SWARTHMORE COLLEGE	168742729
ST FRANCIS HOSPITAL - POUGHKEEPSIE	164506124
Vail Clinic Inc	159617575
TUSKEGEE UNIVERSITY	152572824

20 rows in set (0.58 sec)

2. We also need to know the EINs and cities for each company that made between \$1-100,000 in revenue and between \$10000-200,000 in expenses. (If this list is long, just give us a random selection of 20 and your query)

```
mysql> select ein,city from(select ein,city,revenue, expenses from taxdata where 1<revenue<100000 and 10000 < expenses< 200000) as subquery limit 20;
```

ein	city
742661023	SAN ANTONIO
562629114	BROOKLYN
270678774	GLENDORA
464114252	NEWPORT
510311790	DOVER
261460932	WISE
270609504	DETROIT
205710892	ELK GROVE
251910030	SAN FRANCISCO
521548962	ROCKVILLE
731653383	FERNDALE
10706544	Ocala
274563334	Tucson
262511957	ADA
272260017	Cincinnati
271198702	LEESBURG
841650210	AUSTIN
760294275	WASHINGTON
541250706	WARRENTON
270526903	North Fort Myers

20 rows in set (0.04 sec)

Part 3 - Specific Words

1. First, we are interested in the number of companies with the word "toy" anywhere in the 'purpose' field.

There are 378 companies with the word 'toy' anywhere in the 'purpose' field.

```
mysql> select count(1) from taxdata where purpose like '%toy%';
```

count(1)
378

1 row in set (2.39 sec)

2. Additionally, we need data on how many rows have both the word 'smith' in the 'ptname' field and have reported revenue (e.g. revenue is not empty or 0).

There are 2796 rows meet the above requirements.

```
mysql> select count(1) from taxdata where ptname like '%smith%' and revenue is not null
and revenue <> 0;
+-----+
| count(1) |
+-----+
|      2796 |
+-----+
1 row in set (0.92 sec)
```

Part 4

1. First, we need to know the company name and length of the name for 50 random companies with a ptid of P01345770.

The following is my query and part of the output table;

```
mysql> select name, length(name) from taxdata where ptid= 'P01345770' order by rand() limit 50;
```

name	length(name)
Singhal Family Charitable Foundation	36
Xin Cheng Foundation	20
McLamore Family Foundation Inc	30
Cultures of Resistance Network Foundation	41
The EvansKabush Family Foundation	33
The Esperance Family Foundation Inc	35
The McLean Family Foundation	28
The Loftus Family Foundation	28
The Obrzut – Ling Foundation Inc	32
William T Foley Foundation Inc	30
Twomey Family Foundation	24
Lighten Family Foundation	25
The John Murray and Eleanor M StritterFoundation Inc	52
Joy and Hank Kuchta Foundation	30
The Joseph and Lucille Madri FamilyFoundation	45
The Hartings Family Charitable Foundation	41
Dilmaghani Dream Foundation	27
Turkanis Family Foundation	26
Oak Tree Foundation of Colorado AKA OakTree Foundation	54
The Hart Family Foundation	26
The Leah and Alain Lebec Foundation Inc	39
The Cosse-Finney Foundation	27
Sampson Family Foundation	25
JM McDonald Foundation Inc	26
The Sturtevant Family Foundation	32
Blue Earth Foundation Inc	25
Kanshi Ram Memorial Arora (KARMA) Foundation	44
Bhai Jaita Foundation Inc	25
Gary W Dietrich Family Foundation	33
Mary Elizabeth Conover Foundation	33
William H Wood Family Foundation	32
Robert and Ardis James Foundation Inc	37

2. Second, we need the number of companies that have a 'purpose' field containing less than 10 characters. Keep in mind that the value of the 'purpose' field should never be an empty string.

There are 917 companies meet the requirement.

```
mysql> select count(1) from taxdata where length(purpose) < 10 and purpose <> '';
+-----+
| count(1) |
+-----+
|      917 |
+-----+
1 row in set (1.01 sec)
```

Part 5 - Employee's

1. What shortest WHERE clause can you write to query the number of folks that got hired in 1994, 1995 and 1990. (You can do this using the count(1) syntax as discussed in class, or a query that returns all the records and you just count). I want the shortest WHERE query possible.

The following is my query:

```
mysql> select count(1) from employees where year(hire_date) in (1990, 1994, 1995);
+-----+
| count(1) |
+-----+
|    52560 |
+-----+
1 row in set (0.27 sec)
```

2. What query will provide a count of all 'Senior Engineer' that were at the company on 1986-06-26. (Hint, this is not a simple where date=1)

The following is my query:

```
mysql> select count(1) from titles where title = 'Senior Engineer'
-> and from_date <= '1986-06-26'
-> and (to_date > '1986-06-26' or to_date is null);
+-----+
| count(1) |
+-----+
|     2795 |
+-----+
1 row in set (0.79 sec)
```

employees

Field Name	Data Type	Values	Note
emp_no	int	employee number	primary key
birth_date	date	birthdate of employee	
first_name	varchar	first name of employee	
last_name	varchar	last name of employee	
gender	enum('M', 'F')	gender of employee	
hire_date	date	employee hired date	

3. Can you give me a list of unique names of folks that have had a title of "Engineer". We have not covered joins yet, so please make sure to do this without a join and in a single query.

There are 11930 names meet the requirements showed in the following picture, to give a short preview I added 'limit 10' command at the last:

```
SachinTendulkar
+-----+
111930 rows in set (1.59 sec)

mysql> select distinct concat(first_name, ' ', last_name) as full_name from employees where emp_no in (select distinct emp_no from titles where title = 'Engineer') limit 10;
+-----+
| full_name |
+-----+
| Chirstian Koblick |
| Sumant Peac |
| Duangkaew Piveteau |
| Patricio Bridgland |
| Berni Genin |
| Kazuhide Peha |
| Mayuko Warwick |
| Shahaf Famili |
| Bojan Montemayor |
| Yongqiao Berztiss |
+-----+
10 rows in set (0.04 sec)
```

Title

Field Name	Data Type	Values	Note
emp_no	int	employee number	primary key
title	varchar	title of employee	
from_date	date	employee hired from date	
to_date	date	employee left company date	

Hopefully my analytics will be helpful! Thank you!

Best regards,

Jun