

How to Meet and Retrieve Your Data

Watershed's intern put together a couple of sources of information that will be useful for your project. These three types of information are contained in the capstone database:

1. the current monthly rent Watershed charges for all of their client's 244 properties, as well as the property type and geographic location of those properties.
2. some general information about examples of short-term rental properties. This information can be used to get a sense of what kind of nightly rental price Watershed's client's properties *could* be listed for, if they were converted to short-term rentals.
3. records about when those short-term rental properties were rented out, so that you can calculate their occupancy rates.

Your job is to determine how the database is organized so that you can retrieve all of the available information about Watershed's client's 244 properties, as well as the corresponding short-term rental information for comparable properties in the same location and of the same type.

1. Start by determining what tables the database contains, and what fields are included in each table.
2. Then, we recommend that you make at least a rough relational schema of how the database is organized, so that you know what fields you can use to join tables.
3. Next, make a list of the columns of data you want to retrieve in your final output.
4. Finally, write your query to retrieve the desired data from the database.

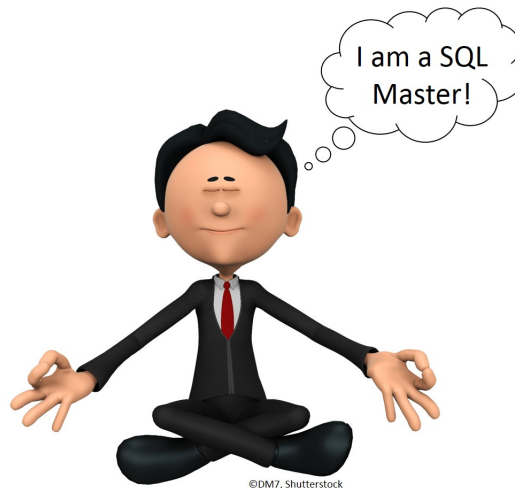
Here are some hints about how to write your query:

- Start by joining no more than two tables. After you have made sure the query works as written and that the output makes sense, add other tables one at a time, checking the new query and its results each time.
- Your final output should have 244 rows. Given the limited output, the easiest way to extract the results will be to copy and paste the output from your query into Excel, although you could also extract as a .csv file and open that with Excel. If you choose the .csv option, you might find it necessary to write your query on multiple lines when you declare it as a variable. To do this, type a space (if you forget the space the lines will run together) and a "\ " at the end of each line of your query:

```
my_data= %sql SELECT DISTINCT user_guid, state, membership_type \
FROM users \
WHERE country="US" AND state IS NOT NULL and membership_type IS NOT NULL \
ORDER BY state ASC, membership_type ASC ;
```

```
my_data.csv( 'my_data.csv' )
```

- We recommend that you calculate the occupancy rates of the example short-term rental properties within MySQL, rather than within Excel (it will be much faster!) To do this, only examine rental dates during 2015, and remember that there are 365 days in the year. The final output of your calculation should be the percentage of days in 2015 that the property was occupied. You may want to consider using a subquery for this calculation.
- Make sure that you extract information from short-term rentals ** that have the same location and property type ** as the 244 Watershed properties.
- If you run into trouble, use your workbooks and Teradata notes from “Managing Big Data with MySQL” to remind you how to implement different parts of your query.



Good luck and have fun!

To get started, connect to the capstone database and set the database as your default database using the following commands:

```
%load_ext sql
%sql mysql://studentuser:studentpw@localhost/capstone
%sql USE capstone
```

Load and connect to the database

```
In [1]: %load_ext sql
        %sql mysql://studentuser:studentpw@localhost/capstone
        %sql USE capstone

        * mysql://studentuser:***@localhost/capstone
0 rows affected.

Out[1]: []
```

Queries

You can add as many "cells" as you need in order to explore the database and extract the appropriate data. For a reminder about what "cells" are, how to add them, or how to use Jupyter in general, please refer to the "How to Use Jupyter Notebooks" video at: <https://www.coursera.org/learn/analytics-mysql/lecture/oxkUg/how-to-use-jupyter-notebooks> (<https://www.coursera.org/learn/analytics-mysql/lecture/oxkUg/how-to-use-jupyter-notebooks>).

1. Know the Database

First, we need to know how many tables in the database

```
In [4]: %sql SHOW tables

        * mysql://studentuser:***@localhost/capstone
6 rows affected.
```

```
Out[4]:      Tables_in_capstone
          location
          property_type
          st_property_info
          st_rental_dates
          st_rental_prices
          watershed_property_info
```

Second, we need to know the column details in each table

```
In [5]: %sql DESCRIBE location

        * mysql://studentuser:***@localhost/capstone
4 rows affected.
```

```
Out[5]:      Field      Type  Null  Key  Default  Extra
location_id  varchar(255)  NO    PRI   None
city         varchar(255)  YES
state        varchar(255)  YES
zipcode      int(5)       YES
```

```
In [12]: %%sql
SELECT *
FROM location
LIMIT 25

* mysql://studentuser:***@localhost/capstone
25 rows affected.
```

Out[12]:

location_id	city	state	zipcode
L1	Anchorage	AK	99501
L10	Anchorage	AK	99518
L100	Cherokee	AL	35616
L1000	Azusa	CA	91702
L10000	Winston-Sa	NC	27101
L10001	Winston-Sa	NC	27103
L10002	Winston-Sa	NC	27104
L10003	Winston-Sa	NC	27105
L10004	Winston-Sa	NC	27106
L10005	Winston-Sa	NC	27107
L10006	Winston-Sa	NC	27127
L10007	Woodleaf	NC	27054
L10008	Yadkinvill	NC	27055
L10009	Yanceyvill	NC	27379
L1001	Bakersfiel	CA	93301
L10010	Youngsvill	NC	27596
L10011	Zebulon	NC	27597
L10012	Zionville	NC	28698
L10013	Zirconia	NC	28790
L10014	Belfield	ND	58622
L10015	Carrington	ND	58421
L10016	Center	ND	58530
L10017	Cooperstow	ND	58425
L10018	Ellendale	ND	58436
L10019	Enderlin	ND	58027

```
In [6]: %sql DESCRIBE property_type

* mysql://studentuser:***@localhost/capstone
5 rows affected.
```

Out[6]:

Field	Type	Null	Key	Default	Extra
property_type_id	varchar(255)	NO	PRI	None	
apt_house	varchar(255)	YES		None	
num_bedrooms	varchar(255)	YES		None	
kitchen	varchar(255)	YES		None	
shared	varchar(255)	YES		None	

```
In [13]: %%sql
SELECT *
FROM property_type;

* mysql://studentuser:***@localhost/capstone
16 rows affected.
```

Out[13]:

property_type_id	apt_house	num_bedrooms	kitchen	shared
R1	apartment	1	Y	Y
R10	house	1	Y	N
R11	house	1	N	Y
R12	house	1	N	N
R13	house	2	Y	Y
R14	house	2	Y	N
R15	house	2	N	Y
R16	house	2	N	N
R2	apartment	1	Y	N
R3	apartment	1	N	Y
R4	apartment	1	N	N
R5	apartment	2	Y	Y
R6	apartment	2	Y	N
R7	apartment	2	N	Y
R8	apartment	2	N	N
R9	house	1	Y	Y

```
In [7]: %%sql DESCRIBE st_property_info

* mysql://studentuser:***@localhost/capstone
3 rows affected.
```

Out[7]:

Field	Type	Null	Key	Default	Extra
st_property_id	varchar(255)	NO	PRI	None	
location	varchar(255)	YES		None	
property_type	varchar(255)	YES		None	

```
In [15]: %%sql
SELECT *
FROM st_property_info
LIMIT 25;

* mysql://studentuser:***@localhost/capstone
25 rows affected.
```

Out[15]:

st_property_id	location	property_type
ST1	L9531	R6
ST10	L9533	R6
ST100	L1944	R2
ST1000	L5957	R1
ST101	L15257	R2
ST102	L15257	R6
ST103	L15257	R10
ST104	L15257	R14
ST105	L15260	R2
ST106	L15260	R6
ST107	L15260	R10
ST108	L15260	R14
ST109	L15264	R2
ST11	L9533	R10
ST110	L15264	R6
ST111	L15264	R10
ST112	L15264	R14
ST113	L15278	R2
ST114	L15278	R6
ST115	L15278	R10
ST116	L15278	R14
ST117	L15280	R2
ST118	L15280	R6
ST119	L15280	R10
ST12	L9533	R14

```
In [8]: %%sql DESCRIBE st_rental_dates

* mysql://studentuser:***@localhost/capstone
2 rows affected.
```

Out[8]:

Field	Type	Null	Key	Default	Extra
rental_date	date	NO	PRI	None	
st_property	varchar(255)	NO	PRI	None	

```
In [21]: %%sql
SELECT MAX(rental_date), MIN(rental_date)
FROM st_rental_dates;

* mysql://studentuser:***@localhost/capstone
1 rows affected.
```

```
Out[21]:  MAX(rental_date)  MIN(rental_date)

          2015-12-31      2014-01-31
```

```
In [23]: %%sql
SELECT *
FROM st_rental_dates
LIMIT 25;

* mysql://studentuser:***@localhost/capstone
25 rows affected.
```

```
Out[23]:  rental_date  st_property

2014-01-31      ST100
2014-01-31      ST106
2014-01-31      ST107
2014-01-31      ST109
2014-01-31      ST113
2014-01-31      ST114
2014-01-31      ST116
2014-01-31      ST118
2014-01-31      ST119
2014-01-31      ST123
2014-01-31      ST126
2014-01-31      ST127
2014-01-31      ST128
2014-01-31      ST13
2014-01-31      ST131
2014-01-31      ST134
2014-01-31      ST137
2014-01-31      ST139
2014-01-31      ST14
2014-01-31      ST142
2014-01-31      ST143
2014-01-31      ST147
2014-01-31      ST149
2014-01-31      ST15
2014-01-31      ST150
```

In [9]: %%sql DESCRIBE st_rental_prices

```
* mysql://studentuser:***@localhost/capstone
5 rows affected.
```

Out[9]:

	Field	Type	Null	Key	Default	Extra
	location	varchar(255)	NO	PRI	None	
	property_type	varchar(255)	NO	PRI	None	
	percentile_10th_price	int(11)	YES		None	
	percentile_90th_price	int(11)	YES		None	
	sample_nightly_rent_price	int(11)	YES		None	

In [19]:

```
%%sql
SELECT *
FROM st_rental_prices
LIMIT 25;
```

```
* mysql://studentuser:***@localhost/capstone
25 rows affected.
```

Out[19]:

location	property_type	percentile_10th_price	percentile_90th_price	sample_nightly_rent_price
L10126	R1	83	273	192
L10126	R10	155	494	257
L10126	R11	82	411	242
L10126	R12	119	366	229
L10126	R13	137	416	375
L10126	R14	151	391	286
L10126	R15	116	312	172
L10126	R16	126	343	258
L10126	R2	91	342	229
L10126	R3	41	242	117
L10126	R4	83	263	193
L10126	R5	153	340	218
L10126	R6	168	392	322
L10126	R7	140	401	350
L10126	R8	129	409	230
L10126	R9	111	371	155
L10130	R1	62	188	129
L10130	R10	190	462	221
L10130	R11	136	393	260
L10130	R12	158	415	313
L10130	R13	146	440	233
L10130	R14	205	411	316
L10130	R15	103	309	169
L10130	R16	186	402	317
L10130	R2	99	265	180


```
In [10]: %%sql DESCRIBE watershed_property_info
```

* mysql://studentuser:***@localhost/capstone
4 rows affected.

Out[10]:

	Field	Type	Null	Key	Default	Extra
	ws_property_id	varchar(255)	NO	PRI	None	
	location	varchar(255)	YES		None	
	property_type	varchar(255)	YES		None	
	current_monthly_rent	int(11)	NO		None	

```
In [2]: %%sql  
SELECT *  
FROM watershed_property_info  
LIMIT 25;
```

* mysql://studentuser:***@localhost/capstone
25 rows affected.

Out[2]:

ws_property_id	location	property_type	current_monthly_rent
W1	L9531	R6	1060
W10	L9533	R6	1200
W100	L1944	R2	3300
W101	L15257	R2	1400
W102	L15257	R6	2000
W103	L15257	R10	1600
W104	L15257	R14	2800
W105	L15260	R2	1100
W106	L15260	R6	1900
W107	L15260	R10	1800
W108	L15260	R14	3200
W109	L15264	R2	1000
W11	L9533	R10	1000
W110	L15264	R6	1300
W111	L15264	R10	1200
W112	L15264	R14	1600
W113	L15278	R2	800
W114	L15278	R6	1200
W115	L15278	R10	900
W116	L15278	R14	1100
W117	L15280	R2	1000
W118	L15280	R6	1400
W119	L15280	R10	1500
W12	L9533	R14	1300
W120	L15280	R14	1600

2. Join tables

Get occupancy rate for each st_property

```
In [3]: %%sql
SELECT d.st_property, CONCAT(FORMAT(COUNT(d.st_property)/365*100,2),'%') AS occupancy_rate
FROM st_rental_dates d
WHERE rental_date>"2014-12-31"
GROUP BY d.st_property
LIMIT 25;
```

```
* mysql://studentuser:***@localhost/capstone
25 rows affected.
```

Out[3]:

st_property	occupancy_rate
-------------	----------------

ST1	16.16%
ST10	34.79%
ST100	39.73%
ST1000	98.08%
ST101	36.44%
ST102	41.10%
ST103	41.10%
ST104	52.60%
ST105	43.29%
ST106	69.59%
ST107	10.96%
ST108	22.47%
ST109	21.92%
ST11	39.18%
ST110	53.70%
ST111	51.23%
ST112	36.16%
ST113	84.38%
ST114	91.51%
ST115	43.01%
ST116	48.22%
ST117	49.04%
ST118	52.33%
ST119	44.93%
ST12	66.03%

Get occupancy rate for each property type in each location And from the extract, except for property type 13 and location L17835 contains information for 2 properties (occupancy rate for these two are pretty close), all other property type in each location only has information one property. I will just take avg_occupancy rate into consideration.

```
In [4]: %%sql
SELECT prop.location
      ,prop.property_type
      ,COUNT(occu_temp.st_property) AS count_property
      ,CONCAT(FORMAT(avg(occu_temp.occupancy_rate),2),'%') AS avg_occupancy_rate
      ,CONCAT(FORMAT(min(occu_temp.occupancy_rate),2),'%') AS low_occupancy_rate
      ,CONCAT(FORMAT(max(occu_temp.occupancy_rate),2),'%') AS high_occupancy_rate
FROM st_property_info AS prop
LEFT JOIN
  (SELECT d.st_property, CONCAT(FORMAT(COUNT(d.st_property)/365*100,2),'%') AS occupancy_r
ate
   FROM st_rental_dates d
   WHERE rental_date>"2014-12-31"
   GROUP BY d.st_property) AS occu_temp
ON prop.st_property_id = occu_temp.st_property
GROUP BY prop.location, prop.property_type
ORDER BY count_property DESC
LIMIT 25;
```

```
* mysql://studentuser:***@localhost/capstone
25 rows affected.
```

```
Out[4]:
```

location	property_type	count_property	avg_occupancy_rate	low_occupancy_rate	high_occupancy_rate
L17835	R13	2	97.94%	97.81%	98.08%
L3368	R13	1	97.81%	97.81%	97.81%
L9573	R8	1	97.81%	97.81%	97.81%
L11495	R14	1	50.68%	50.68%	50.68%
L1882	R10	1	45.75%	45.75%	45.75%
L6800	R7	1	98.36%	98.36%	98.36%
L3515	R5	1	98.08%	98.08%	98.08%
L18071	R12	1	98.90%	98.90%	98.90%
L15260	R2	1	43.29%	43.29%	43.29%
L11738	R7	1	99.45%	99.45%	99.45%
L4804	R6	1	40.82%	40.82%	40.82%
L10386	R9	1	99.45%	99.45%	99.45%
L17473	R4	1	98.08%	98.08%	98.08%
L8120	R5	1	97.53%	97.53%	97.53%
L9506	R13	1	97.81%	97.81%	97.81%
L16890	R6	1	29.32%	29.32%	29.32%
L3244	R2	1	41.37%	41.37%	41.37%
L13649	R7	1	96.99%	96.99%	96.99%
L3810	R13	1	98.63%	98.63%	98.63%
L16976	R16	1	98.36%	98.36%	98.36%
L14484	R13	1	97.81%	97.81%	97.81%
L1902	R14	1	39.73%	39.73%	39.73%
L14418	R2	1	29.04%	29.04%	29.04%
L6075	R16	1	98.36%	98.36%	98.36%
L9533	R14	1	66.03%	66.03%	66.03%

Combine occupancy and nightly rental price together

In [5]: %%sql

```
SELECT occu.location
      ,occu.property_type
      ,occu.avg_occupancy_rate
      ,price.percentile_10th_price AS 10th_percentile
      ,price.percentile_90th_price AS 90th_percentile
FROM
  (SELECT prop.location
      ,prop.property_type
      ,COUNT(occu_temp.st_property) AS count_property
      ,CONCAT(FORMAT(avg(occu_temp.occupancy_rate),2),'%') AS avg_occupancy_rate
    FROM st_property_info AS prop
  LEFT JOIN
    (SELECT d.st_property, CONCAT(FORMAT(COUNT(d.st_property)/365*100,2),'%') AS occupan
cy_rate
    FROM st_rental_dates d
    WHERE rental_date>"2014-12-31"
    GROUP BY d.st_property) AS occu_temp
  ON prop.st_property_id = occu_temp.st_property
  GROUP BY prop.location, prop.property_type
  ORDER BY count_property DESC) AS occu
LEFT JOIN st_rental_prices AS price
ON occu.location=price.location
  AND occu.property_type=price.property_type
LIMIT 25;
```

```
* mysql://studentuser:***@localhost/capstone
25 rows affected.
```

```
Out[5]:
```

location	property_type	avg_occupancy_rate	10th_percentile	90th_percentile
L17835	R13	97.94%	None	None
L16239	R13	99.18%	None	None
L7408	R11	97.53%	None	None
L7747	R16	96.16%	None	None
L15276	R4	96.99%	None	None
L11495	R10	40.27%	102	799
L7099	R8	97.81%	None	None
L12252	R14	49.32%	175	368
L10126	R14	45.21%	151	391
L12097	R8	97.81%	None	None
L1741	R8	97.81%	None	None
L371	R12	98.36%	None	None
L11972	R15	98.08%	None	None
L12264	R2	47.95%	80	156
L2234	R12	99.18%	None	None
L1735	R6	34.25%	270	543
L3179	R1	97.26%	None	None
L4761	R10	62.19%	189	588
L1944	R6	31.78%	265	644
L15371	R5	98.36%	None	None
L11597	R1	97.26%	None	None
L6987	R13	97.53%	None	None
L2362	R11	97.26%	None	None
L11766	R1	98.08%	None	None
L16887	R10	52.33%	106	267

Combine location detail and property type detail to Watershed's property

```
In [6]: %%sql
SELECT ws.ws_property_id
      ,ws.location
      ,l.city
      ,l.state
      ,l.zipcode
      ,ws.property_type
      ,p_type.apartment
      ,p_type.num_bedrooms
      ,p_type.kitchen
      ,p_type.shared
      ,ws.current_monthly_rent
FROM watershed_property_info ws
LEFT JOIN location l
ON l.location_id=ws.location
LEFT JOIN property_type p_type
ON p_type.property_type_id=ws.property_type
LIMIT 25;

* mysql://studentuser:***@localhost/capstone
25 rows affected.
```

```
Out[6]:
```

ws_property_id	location	city	state	zipcode	property_type	apartment	num_bedrooms	kitchen	shared	current_mor
W1	L9531	Chapel Hill	NC	27514	R6	apartment	2	Y	N	
W10	L9533	Chapel Hill	NC	27517	R6	apartment	2	Y	N	
W100	L1944	San Francisco	CA	94129	R2	apartment	1	Y	N	
W101	L15257	Austin	TX	78702	R2	apartment	1	Y	N	
W102	L15257	Austin	TX	78702	R6	apartment	2	Y	N	
W103	L15257	Austin	TX	78702	R10	house	1	Y	N	
W104	L15257	Austin	TX	78702	R14	house	2	Y	N	
W105	L15260	Austin	TX	78705	R2	apartment	1	Y	N	
W106	L15260	Austin	TX	78705	R6	apartment	2	Y	N	
W107	L15260	Austin	TX	78705	R10	house	1	Y	N	
W108	L15260	Austin	TX	78705	R14	house	2	Y	N	
W109	L15264	Austin	TX	78723	R2	apartment	1	Y	N	
W11	L9533	Chapel Hill	NC	27517	R10	house	1	Y	N	
W110	L15264	Austin	TX	78723	R6	apartment	2	Y	N	
W111	L15264	Austin	TX	78723	R10	house	1	Y	N	
W112	L15264	Austin	TX	78723	R14	house	2	Y	N	
W113	L15278	Austin	TX	78744	R2	apartment	1	Y	N	
W114	L15278	Austin	TX	78744	R6	apartment	2	Y	N	
W115	L15278	Austin	TX	78744	R10	house	1	Y	N	
W116	L15278	Austin	TX	78744	R14	house	2	Y	N	
W117	L15280	Austin	TX	78746	R2	apartment	1	Y	N	
W118	L15280	Austin	TX	78746	R6	apartment	2	Y	N	
W119	L15280	Austin	TX	78746	R10	house	1	Y	N	
W12	L9533	Chapel Hill	NC	27517	R14	house	2	Y	N	
W120	L15280	Austin	TX	78746	R14	house	2	Y	N	

Combine all information together

In [29]: %%sql

```
SELECT ws.ws_property_id AS property_id
      ,ws.location
      ,l.city
      ,l.state
      ,l.zipcode
      ,ws.property_type
      ,st.avg_occupancy_rate
      ,p_type.apartment_house
      ,p_type.num_bedrooms
      ,p_type.kitchen
      ,p_type.shared
      ,ws.current_monthly_rent
      ,st.10th_percentile AS 10th_percentile_nightly
      ,st.90th_percentile AS 90th_percentile_nightly
      ,st.sample AS sample_nightly
FROM watershed_property_info ws
LEFT JOIN location l
ON l.location_id=ws.location
LEFT JOIN property_type p_type
ON p_type.property_type_id=ws.property_type
LEFT JOIN
  (SELECT occu.location
    ,occu.property_type
    ,occu.avg_occupancy_rate
    ,price.percentile_10th_price AS 10th_percentile
    ,price.percentile_90th_price AS 90th_percentile
    ,price.sample_nightly_rent_price AS sample
  FROM
    (SELECT prop.location
      ,prop.property_type
      ,CONCAT(FORMAT(avg(occu_temp.occupancy_rate),2),'%') AS avg_occupancy_rate
    FROM st_property_info AS prop
    LEFT JOIN
      (SELECT d.st_property, CONCAT(FORMAT(COUNT(d.st_property)/365*100,2),'%') AS occ
    upancy_rate
      FROM st_rental_dates d
      WHERE rental_date>"2014-12-31"
      GROUP BY d.st_property) AS occu_temp
    ON prop.st_property_id = occu_temp.st_property
    GROUP BY prop.location, prop.property_type) AS occu
  LEFT JOIN st_rental_prices AS price
  ON occu.location=price.location
  AND occu.property_type=price.property_type) AS st
ON st.location=ws.location
AND st.property_type=ws.property_type;
```