

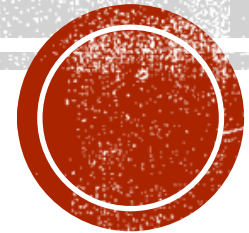
Bay Area Bike Share



Data Analysis for the Period of August 2013 to August 2015

CAPSTONE 2 PROJECT

03 June 2022





PROBLEM STATEMENT

The **Bay Area Bike Share** provides quick, easy, and affordable bike trips in 5 main cities in California Its network of cities San Jose, Redwood City, Mountain View Palo Alto, and San Francisco.

Bay Area Bike Share wants to launch a Summer campaign for San Francisco and wishes to find out the following:

- What is the Customer base & performance like?
- Does weather affect sales?
- Most & Least Popular Dock Point
 - (For marketing to boost up the dock station attendance)
- Turnover Rates Most & Least Popular Dock during Summer
 - (For Performance Monitoring)





DATA COLLECTION SOURCE & CREDITS

<https://www.kaggle.com/datasets/benhamner/sf-bay-area-bike-share?select=trip.csv>

Consist of 5 csv with a Size of Data Set estimated at 660000 or more rows covering 5 main cities in California

station.csv - Contains data that represents a station where users can pickup or return bikes.

status.csv - data about the number of bikes and docks available for given station and minute.

trips.csv - Data about individual bike trips

weather.csv - Data about the weather on a specific day for certain zip codes

Other Credits to:

<https://weatherworksinc.com/news/humidity-vs-dewpoint>

<https://uk.trotec.com/products-services/machinery-homecomfort/dehumidification/practical-knowledge-concerning-dehumidifiers/basic-knowledge-regarding-humidity/>





DATA TRANSFORMATION

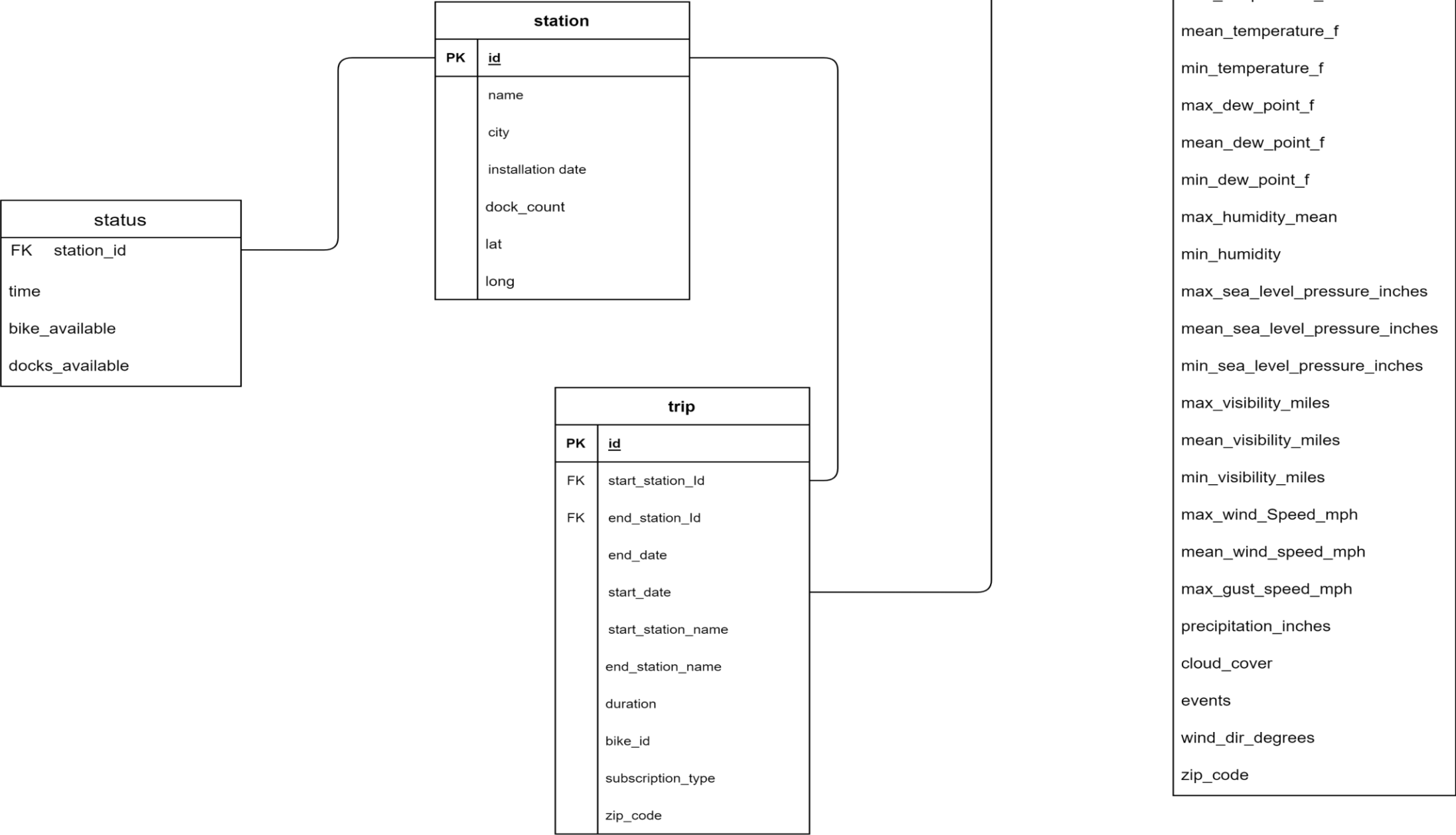
All csv files are combined into SQLite format

No more additional cleaning is required as will make use of SQL commands to adjust during queries

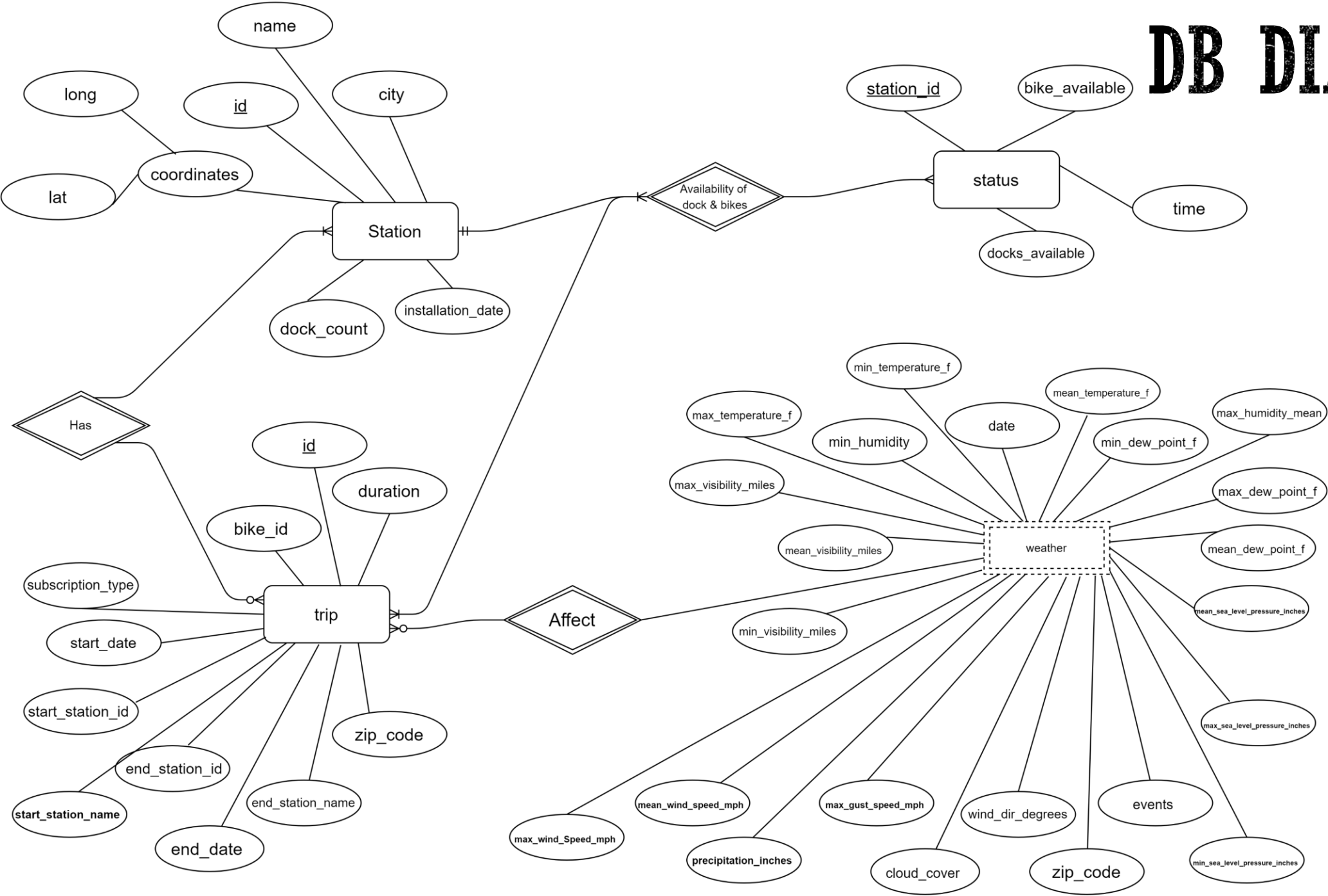
Exportation of SQL query to CSV file



ER DIAGRAM



DB DIAGRAM

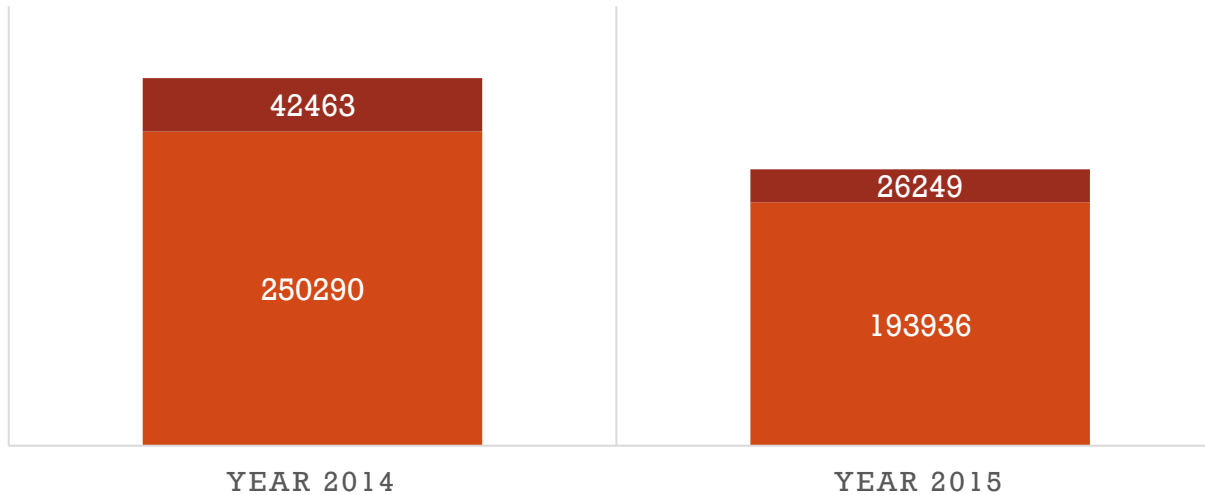


CUSTOMER BASE & PERFORMANCE



TOTAL TRIP VS CUSTOMER TYPE FOR SAN FRANCISCO

Subscriber Customer



Results

Execution Plan

Customer Type	Year 2014	Year 2015
Subscriber	250290	193936
Customer	42463	26249

```
WITH sfs AS (SELECT * FROM station
WHERE city = 'San Francisco'),
```

```
sft2014 AS (SELECT * FROM sfs
INNER JOIN trip ON sfs.id = start_station_id
WHERE start_date LIKE '%/2014%'),
```

```
sft2015 AS (SELECT * FROM sfs
INNER JOIN trip ON sfs.id = start_station_id
WHERE start_date LIKE '%/2015%'),
```

```
sfsub2014 AS (SELECT
subscription_type,COUNT(subscription_type) AS 'Year 2014' FROM
sft2014
GROUP BY subscription_type
ORDER BY subscription_type DESC),
```

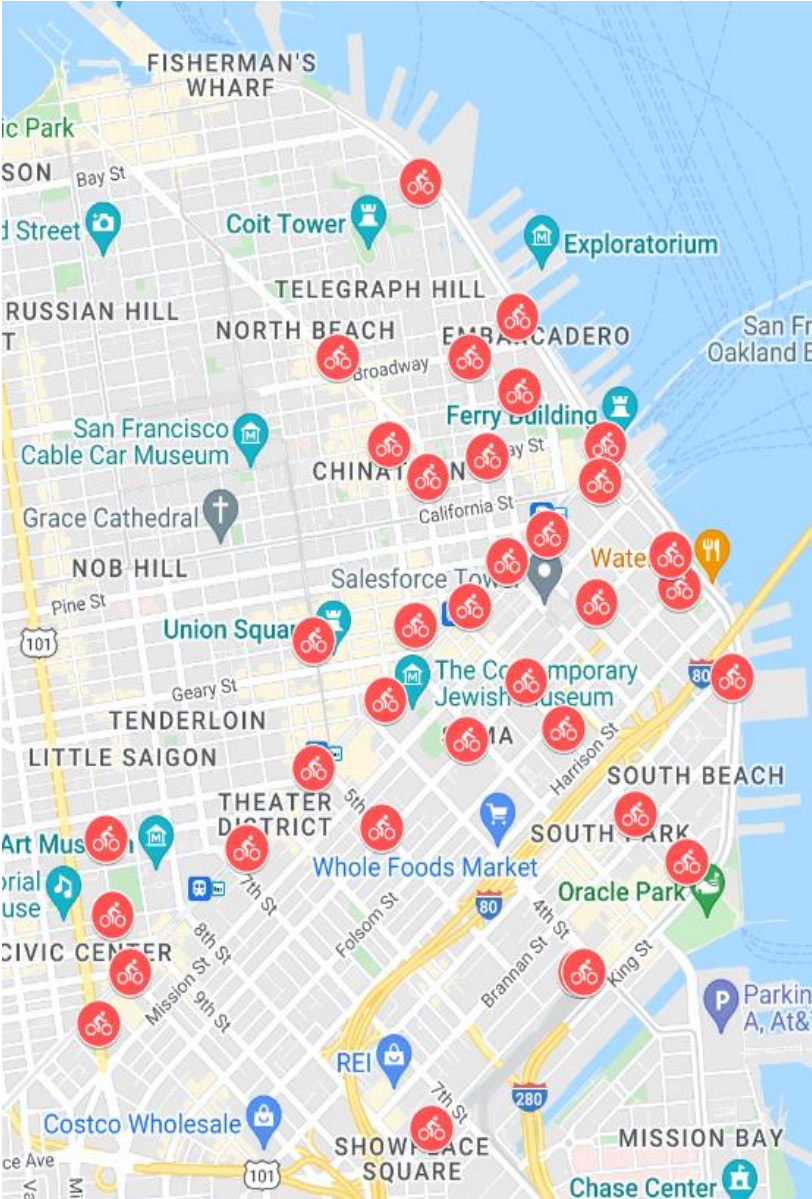
```
sfsub2015 AS (SELECT
subscription_type,COUNT(subscription_type) AS 'Year 2015' FROM
sft2015
GROUP BY subscription_type
ORDER BY subscription_type DESC)
```

```
SELECT sfsub2014.subscription_type AS 'Customer
Type',sfsub2014.'Year 2014',sfsub2015.'Year 2015' FROM sfsub2014
INNER JOIN sfsub2015
ON sfsub2014.subscription_type=sfsub2015.subscription_type;
```



BAY AREA BIKE SHARE LIST OF STATION@SAN FRANCISCO

id	Name of Station
39	Powell Street BART
41	Clay at Battery
42	Davis at Jackson
45	Commercial at Montgomery
46	Washington at Kearney
47	Post at Kearney
48	Embarcadero at Vallejo
49	Spear at Folsom
50	Harry Bridges Plaza (Ferry Building)
51	Embarcadero at Folsom
54	Embarcadero at Bryant
55	Temporary Transbay Terminal (Howard at Beale)
56	Beale at Market
57	5th at Howard
58	San Francisco City Hall
59	Golden Gate at Polk
60	Embarcadero at Sansome
61	2nd at Townsend
62	2nd at Folsom
63	Howard at 2nd
64	2nd at South Park
65	Townsend at 7th
66	South Van Ness at Market
67	Market at 10th

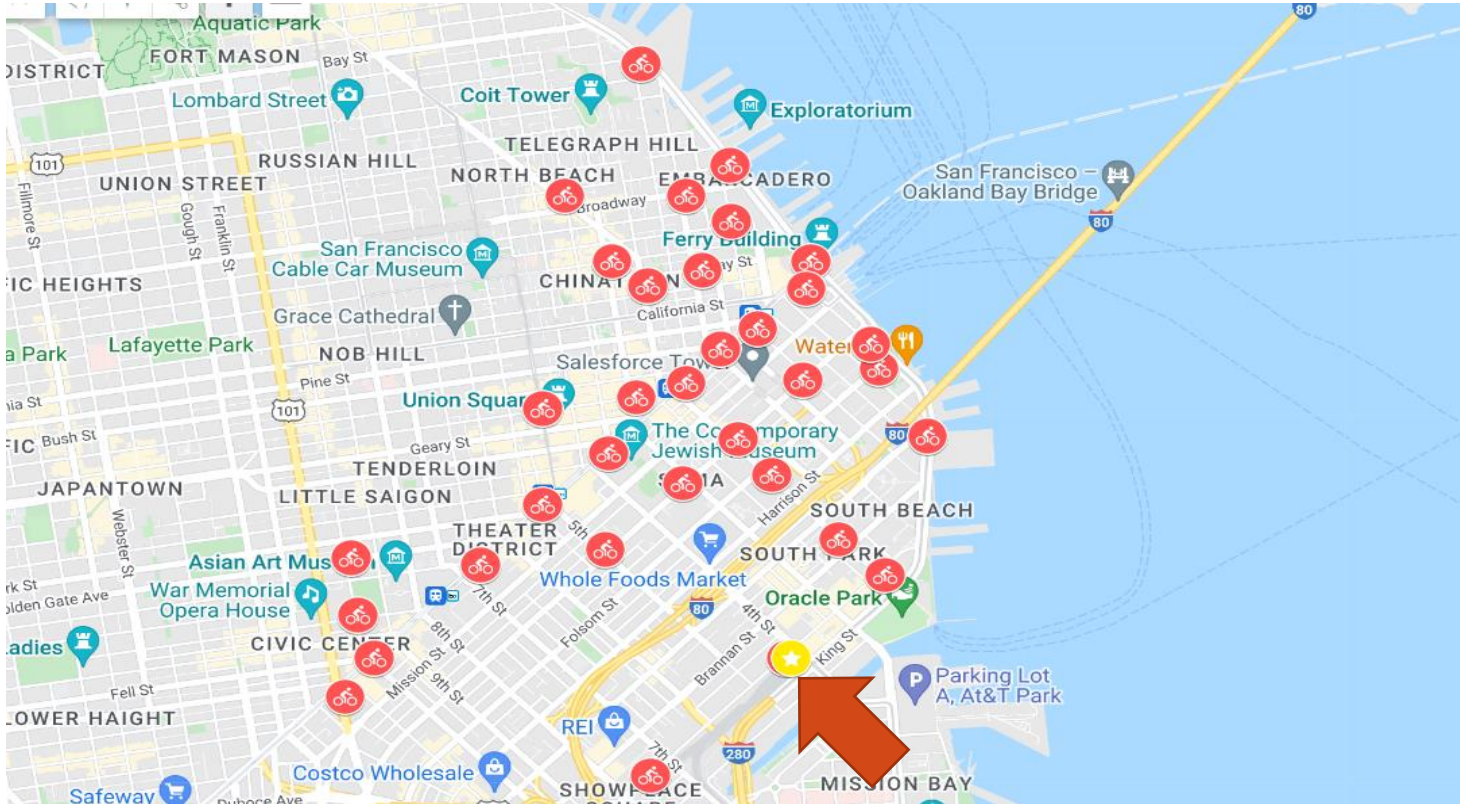


id	Name of Station
68	Yerba Buena Center of the Arts (3rd @ Howard)
69	San Francisco Caltrain 2 (330 Townsend)
70	San Francisco Caltrain (Townsend at 4th)
71	Powell at Post (Union Square)
72	Civic Center BART (7th at Market)
73	Grant Avenue at Columbus Avenue
74	Steuart at Market
75	Mechanics Plaza (Market at Battery)
76	Market at 4th
77	Market at Sansome
82	Broadway St at Battery St





MOST POPULAR STATION



```
WITH sfs AS (SELECT * FROM station
WHERE city = 'San Francisco'),

sft2014 AS (SELECT * FROM sfs
INNER JOIN trip ON sfs.id = start_station_id
WHERE start_date LIKE '%/2014%'),

sft2015 AS (SELECT * FROM sfs
INNER JOIN trip ON sfs.id = start_station_id
WHERE start_date LIKE '%/2015%'),

sfp14 AS (SELECT city,start_station_id AS
s14,COUNT(start_station_id)AS tv14 FROM sft2014
GROUP BY start_station_id
ORDER BY COUNT(start_station_id)DESC
LIMIT 5),

sfp15 AS (SELECT city,start_station_id AS
s15,COUNT(start_station_id) AS tv15 FROM sft2015
GROUP BY start_station_id
ORDER BY COUNT(start_station_id)DESC
LIMIT 5)

SELECT s14 AS 'Station for 2014',tv14 AS 'Total Visits
for 2014',s15 AS 'Station for 2015',tv15 AS 'Total Visits
for 2015' FROM sfp14
INNER JOIN sfp15
ON sfp14.city = sfp15.city
LIMIT 1;
```

Results

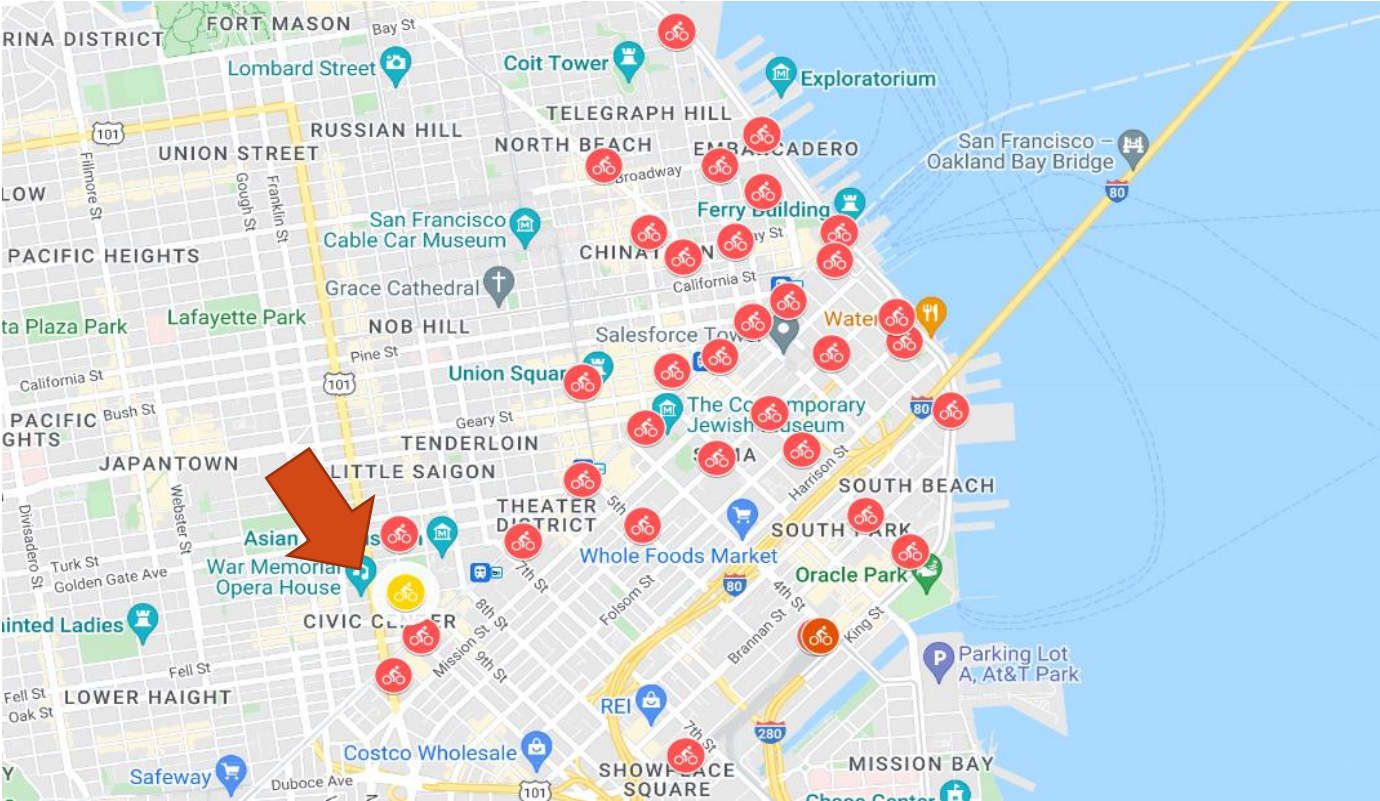
Execution Plan

Station for 2014	Total Visits for 2014	Station for 2015	Total Visits for 2015
70	25144	70	17496

70 San Francisco Caltrain (Townsend at 4th)



LEAST POPULAR STATION



```
WITH sfs AS (SELECT * FROM station
WHERE city = 'San Francisco'),
```

```
sft2014 AS (SELECT * FROM sfs
INNER JOIN trip ON sfs.id = start_station_id
WHERE start_date LIKE '%/2014%'),
```

```
sft2015 AS (SELECT * FROM sfs
INNER JOIN trip ON sfs.id = start_station_id
WHERE start_date LIKE '%/2015%'),
```

```
sfp14 AS (SELECT city,start_station_id AS
s14,COUNT(start_station_id)AS tv14 FROM sft2014
GROUP BY start_station_id
ORDER BY COUNT(start_station_id)
LIMIT 5),
```

```
sfp15 AS (SELECT city,start_station_id AS
s15,COUNT(start_station_id) AS tv15 FROM sft2015
GROUP BY start_station_id
ORDER BY COUNT(start_station_id)
LIMIT 5)
```

```
SELECT s14 AS 'Station for 2014',tv14 AS 'Total Visits for 2014',s15 AS
'Station for 2015',tv15 AS 'Total Visits for 2015' FROM sfp14
INNER JOIN sfp15
ON sfp14.city = sfp15.city
LIMIT 1;
```

Results		Execution Plan	
Station for 2014	Total Visits for 2014	Station for 2015	Total Visits for 2015
58	2145	58	1466

San Francisco City Hall





MOST POPULAR STATION TURNOVER RATES

Using the data in Most Popular Station, find out the turnover rate for both the summer period of 2014 and 2015.

Sample Size is taken 2 months (July & August)
San Francisco Caltrain (Townsend at 4th)

Results Execution Plan

Most popular station	Turnover Rate(Summer)
70	35270
70	37062

```
WITH mpaug14 AS (SELECT start_station_id,COUNT(id) AS
Total FROM trip
WHERE start_date LIKE '%/8/2014%' OR start_date LIKE
'%/7/2014%' AND start_station_id=70 OR end_station_id=70),

tmpaug14 AS (SELECT start_station_id AS 'Most popular
station',Total/2 AS 'Turnover Rate(Summer)' from mpaug14),

mpaug15 AS (SELECT start_station_id ,COUNT(id) AS Total
FROM trip
WHERE start_date LIKE '%/8/2015%' OR start_date LIKE
'%/7/2015%' AND start_station_id=70 OR end_station_id=70),

tmpaug15 AS (SELECT start_station_id AS 'Most popular
station',Total/2 AS 'Turnover Rate(Summer)' from mpaug15)

SELECT * FROM tmpaug14
UNION
SELECT * FROM tmpaug15;
```



LEAST POPULAR STATION TURNOVER RATES



Using the data in the least Popular Station, find out the turnover rate for both the summer period of 2014 and 2015.

Sample Size is taken 2 months (July & August)

San Francisco City Hall

Results

Execution Plan

Least popular station Turnover Rate(Summer)

58 5465

58 7337

```
WITH lpaug14 AS (SELECT COUNT(id) AS Total,end_station_id
FROM trip
WHERE start_date LIKE '%/8/2014%' OR start_date LIKE
'%/7/2014%' AND start_station_id=58 OR end_station_id=58),
```

```
tlpaug14 AS (SELECT end_station_id AS 'Least popular station',Total/2
AS 'Turnover Rate(Summer)' from lpaug14),
```

```
lpaug15 AS (SELECT COUNT(id) AS Total,end_station_id FROM trip
WHERE start_date LIKE '%/8/2015%' OR start_date LIKE
'%/7/2015%' AND start_station_id=58 OR end_station_id=58),
```

```
tlpaug15 AS (SELECT end_station_id AS 'Least popular station',Total/2
AS 'Turnover Rate(Summer)' from lpaug15)
```

```
SELECT * FROM tlpaug14
UNION
SELECT * FROM tlpaug15;
```

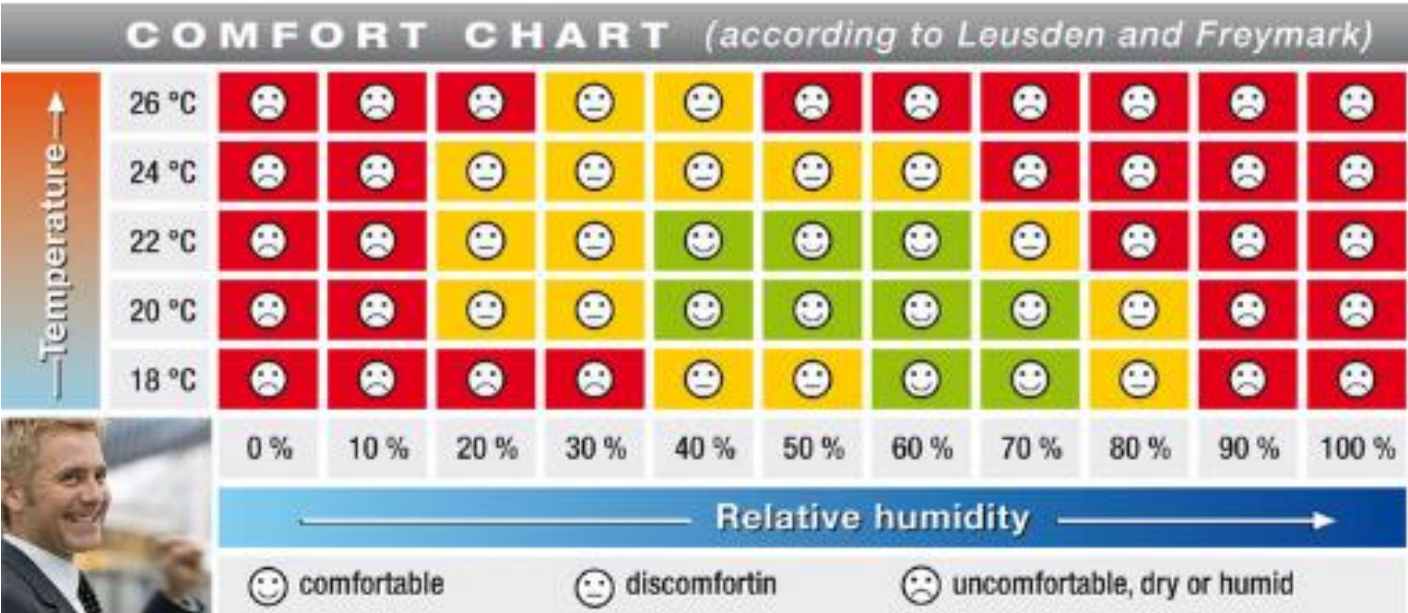


 +  =   

DOES WEATHER AFFECT SALES?

The following criteria will be taken as parameters for comfort level evaluation:

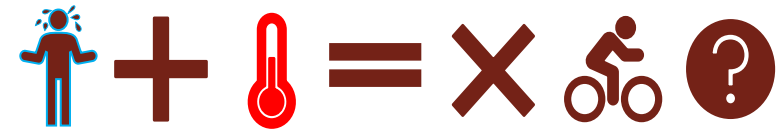
- Average temperature
- Average humidity
- Average dew point
- Sample data for July & August of each year



Use Dewpoint, Not Relative Humidity for Summer Comfort		
Dewpoint	How it Feels	Emojification
50 – 60°F	Comfortable	😊
60 – 65°F	Getting Sticky	😐
65 – 70°F	Unpleasant	😓
70°F or more	Downright Gross	🤢



DOES WEATHER AFFECT SALES?



Uncomfortable Days Vs Total Trip

Dew point above 60

Humidity above 70%*

Celsius (°C)	Fahrenheit (°F)
26 °C	78.8 °F
24 °C	75.2 °F
22 °C	71.6 °F
20 °C	68.0 °F
18 °C	64.4 °F

Results Execution Plan

date	Fahrenheit (°F)		humidity	Dew point	Total Trip
8/29/2015	72	70	62	310	
7/19/2015	77	56	57	350	
8/16/2015	82	54	56	356	
8/15/2015	76	52	54	407	
8/30/2013	76	65	60	606	
8/29/2013	72	69	61	642	
7/25/2014	78	48	52	1082	
8/28/2015	80	60	56	1113	
7/20/2015	76	61	59	1252	
7/29/2015	77	51	53	1259	
8/17/2015	77	57	58	1267	
7/28/2015	77	49	50	1310	
8/27/2015	78	48	51	1331	

```
WITH sfs AS (SELECT * FROM station
WHERE city = 'San Francisco'),

sftsa AS (SELECT * FROM sfs
INNER JOIN trip ON sfs.id = start_station_id),

tpda AS(SELECT SUBSTRING(start_date, 1, 9)
d,COUNT(SUBSTRING(start_date, 1, 9)) ucd
FROM sftsa
WHERE start_date LIKE '7%' OR start_date LIKE
'8%'
GROUP BY d),

uncomfy_dates AS (SELECT
date,mean_temperature_f,mean_humidity,mean_dew
_point_f FROM weather
WHERE zip_code= '95113' AND (date LIKE '7/%'
OR date LIKE '8%')
AND ((mean_temperature_f>75 AND
mean_humidity>40)OR mean_dew_point_f>60))

SELECT
date,mean_temperature_f,mean_humidity,mean_dew
_point_f,ucd FROM tpda
INNER JOIN uncomfy_dates
ON d=date;
```



DOES WEATHER AFFECT SALES?



Comfortable Days Vs Total Trip

Dew point below 60

Humidity 60% ~ 70%*

Celsius (°C)	Fahrenheit (°F)
26 °C	78.8 °F
24 °C	75.2 °F
22 °C	71.6 °F
20 °C	68.0 °F
18 °C	64.4 °F

Results Execution Plan

date	Fahrenheit (°F)		humidity	Dew point	Total Trip
8/24/2014	68	67	55	307	
7/12/2015	69	60	54	316	
7/13/2014	68	69	56	343	
7/12/2014	66	69	54	366	
7/11/2015	67	66	53	439	
7/26/2014	75	59	55	440	
7/11/2014	68	64	54	968	
7/24/2015	67	60	51	1115	
7/10/2015	67	68	55	1119	
7/23/2015	65	65	53	1140	
7/22/2015	68	69	55	1171	
7/14/2015	67	69	56	1191	
8/20/2014	71	60	56	1195	
7/27/2015	72	56	53	1208	
7/15/2015	72	59	56	1236	
8/31/2015	72	58	55	1238	
8/25/2014	68	66	54	1262	
8/26/2015	73	55	54	1306	
8/27/2014	72	60	56	1338	
8/26/2014	67	63	54	1351	

```
WITH sfs AS (SELECT * FROM station
WHERE city = 'San Francisco'),

sftsa AS (SELECT * FROM sfs
INNER JOIN trip ON sfs.id = start_station_id),

tpda AS(SELECT SUBSTRING(start_date, 1, 9)
d,COUNT(SUBSTRING(start_date, 1, 9)) ucd
FROM sftsa
WHERE start_date LIKE '7%' OR start_date LIKE '8%'
GROUP BY d),




comfy_dates AS (SELECT
date,mean_temperature_f,mean_humidity,mean_dew_point_f FROM
weather
WHERE zip_code= '95113' AND (date LIKE '7%' OR date LIKE
'8%')
AND ((mean_temperature_f BETWEEN 68 AND 75 AND
mean_humidity BETWEEN 40 AND 60)
OR (mean_temperature_f BETWEEN 64.4 AND 68 AND
mean_humidity BETWEEN 60 AND 70)
AND mean_dew_point_f<60))




SELECT date,mean_temperature_f AS 'Fahrenheit
(°F)',mean_humidity AS 'humidity',mean_dew_point_f AS 'Dew
point',ucd AS 'Total Trip' FROM tpda
INNER JOIN comfy_dates
ON d=date
ORDER BY ucd;
```



Does weather affect sales?

Comparing the data computed from both tables... 

Results		Execution Plan			
date	Fahrenheit (°F)	humidity	Dew point	Total Trip	
8/24/2014	68	67	55	307	
7/12/2015	69	60	54	316	
7/13/2014	68	69	56	343	
7/12/2014	66	69	54	366	
7/11/2015	67	66	53	439	
7/26/2014	75	59	55	440	
7/11/2014	68	64	54	968	
7/24/2015	67	60	51	1115	
7/10/2015	67	68	55	1119	
7/23/2015	65	65	53	1140	
7/22/2015	68	69	55	1171	
7/14/2015	67	69	56	1191	
8/20/2014	71	60	56	1195	
7/27/2015	72	56	53	1208	
7/15/2015	72	59	56	1236	
8/31/2015	72	58	55	1238	
8/25/2014	68	66	54	1262	
8/26/2015	73	55	54	1306	
8/27/2014	72	60	56	1338	
8/26/2014	67	63	54	1351	

Results		Execution Plan			
date	Fahrenheit (°F)	humidity	Dew point	Total Trip	
8/29/2015	72	70	62	310	
7/19/2015	77	56	57	350	
8/16/2015	82	54	56	356	
8/15/2015	76	52	54	407	
8/30/2013	76	65	60	606	
8/29/2013	72	69	61	642	
7/25/2014	78	48	52	1082	
8/28/2015	80	60	56	1113	
7/20/2015	76	61	59	1252	
7/29/2015	77	51	53	1259	
8/17/2015	77	57	58	1267	
7/28/2015	77	49	50	1310	
8/27/2015	78	48	51	1331	





SUMMARY OF INSIGHTS

- ❧ The Sales are lesser in the Year 2015 compared to the Year 2014
- ❧ Most Popular dock station is at San Francisco Caltrain (Townsend at 4th)
- ❧ Least Popular dock station is at San Francisco City Hall
- ❧ Highest Turnover rate during summer is 37062 & Lowest is 7337
- ❧ There is a Higher Turnover rate in the Year 2015 compared to the Year 2014 the most and least popular station
- ❧ No significant trend shows that Humid weather will affect the user demands, especially on weekdays



THANK YOU...    

