BIKE SHARING ASSIGNMENT

In [1]:
```
import sqlite3
```

In [2]:
```
import pandas as pd
```

In [3]:
```
conn = sqlite3.connect("bikesharing.db")
```

In [4]:
```
cur = conn.cursor()
```

In [6]:
```
station = pd.read_csv("station_updated.csv")
station.head()
```

Out[6]:

	id	name	lat	long	dock_count	city	installation_date
0	2	San Jose Diridon Caltrain Station	37.329732	-121.901782	27	San Jose	2013-08-06
1	3	San Jose Civic Center	37.330698	-121.888979	15	San Jose	2013-08-05
2	4	Santa Clara at Almaden	37.333988	-121.894902	11	San Jose	2013-08-06
3	5	Adobe on Almaden	37.331415	-121.893200	19	San Jose	2013-08-05
4	6	San Pedro Square	37.336721	-121.894074	15	San Jose	2013-08-07

In [7]:
```
status = pd.read_csv("status.csv")
status.head()
```

Out[7]:

	station_id	bikes_available	docks_available	time
0	2	2	25	2013/08/29 12:06:01
1	2	2	25	2013/08/29 12:07:01
2	2	2	25	2013/08/29 12:08:01
3	2	2	25	2013/08/29 12:09:01
4	2	2	25	2013/08/29 12:10:01

In [8]:
```
weather = pd.read_csv("weather_updated.csv")
weather.head()
```

Out[8]:

	Unnamed: 0	date	mean_temperature_f	mean_humidity	mean_dew_point_f	mean_wind_speed_mph	z
0	0	2013-08-29	68.0	75.0	58.0	11.0	

	Unnamed: 0	date	mean_temperature_f	mean_humidity	mean_dew_point_f	mean_wind_speed_mph	z
1	1	2013-08-30	69.0	70.0	58.0	13.0	
2	2	2013-08-31	64.0	75.0	56.0	15.0	
3	3	2013-09-01	66.0	68.0	56.0	13.0	
4	4	2013-09-02	69.0	77.0	60.0	12.0	

In [9]:
```
trip = pd.read_csv("trip_updated.csv")
trip.head()
```

Out[9]:

	id	duration	start_date	start_station_name	start_station_id	end_date	end_station_name	end_stat
0	4576	63	2013-08-29 14:13:00	South Van Ness at Market	66	2013-08-29 14:14:00	South Van Ness at Market	
1	4607	70	2013-08-29 14:42:00	San Jose City Hall	10	2013-08-29 14:43:00	San Jose City Hall	
2	4130	71	2013-08-29 10:16:00	Mountain View City Hall	27	2013-08-29 10:17:00	Mountain View City Hall	
3	4251	77	2013-08-29 11:29:00	San Jose City Hall	10	2013-08-29 11:30:00	San Jose City Hall	
4	4299	83	2013-08-29 12:02:00	South Van Ness at Market	66	2013-08-29 12:04:00	Market at 10th	

In [11]:
```
#Table creation
station.to_sql("station", conn, if_exists="replace", index=False)
```

In [13]:
```
status.to_sql("status", conn, if_exists="replace", index=False)
trip.to_sql("trip", conn, if_exists="replace", index=False)
weather.to_sql("weather", conn, if_exists="replace", index=False)
```

```
C:\Users\User\anaconda4\lib\site-packages\pandas\core\generic.py:2872: UserWarning: The
spaces in these column names will not be changed. In pandas versions < 0.14, spaces were
converted to underscores.
  sql.to_sql(
```

In [12]:
```
pd.read_sql("SELECT * FROM station", conn)
```

Out[12]:

	id	name	lat	long	dock_count	city	installation_date
0	2	San Jose Diridon Caltrain Station	37.329732	-121.901782	27	San Jose	2013-08-06
1	3	San Jose Civic Center	37.330698	-121.888979	15	San Jose	2013-08-05
2	4	Santa Clara at Almaden	37.333988	-121.894902	11	San Jose	2013-08-06
3	5	Adobe on Almaden	37.331415	-121.893200	19	San Jose	2013-08-05
4	6	San Pedro Square	37.336721	-121.894074	15	San Jose	2013-08-07
...	...	...	...	...	...	...	...
65	77	Market at Sansome	37.789625	-122.400811	27	San Francisco	2013-08-25
66	80	Santa Clara County Civic Center	37.352601	-121.905733	15	San Jose	2013-12-31
67	82	Broadway St at Battery St	37.798541	-122.400862	15	San Francisco	2014-01-22
68	83	Mezes Park	37.491269	-122.236234	15	Redwood City	2014-02-20
69	84	Ryland Park	37.342725	-121.895617	15	San Jose	2014-04-09

70 rows × 7 columns

Task - 1

Question 1

In [14]:
```
pd.read_sql("SELECT COUNT(*) from station", conn)
```

Out[14]:

	COUNT(*)
0	70

In [15]:
```
pd.read_sql("SELECT SUM(bikes_available) FROM status", conn)
```

Out[15]:

	SUM(bikes_available)
0	604295824

In [17]:
```
pd.read_sql("SELECT COUNT(*) from trip", conn)
```

Out[17]:

	COUNT(*)
0	669959

Question - 2

In [ ]:

Question - 3

Part 1 # Many to Many Part 2 # Many to One Part 3 # One to Many

Question 4

In [21]:
```
pd.read_sql("SELECT * from trip where start_date+duration = (SELECT MIN(start_date+dura
```

Out[21]:

	id	duration	start_date	start_station_name	start_station_id	end_date	end_station_name	end_
0	8576	60	2013-09-02 09:40:00	Harry Bridges Plaza (Ferry Building)	50	2013-09-02 09:41:00	Harry Bridges Plaza (Ferry Building)	
1	8651	60	2013-09-02 10:50:00	San Francisco Caltrain 2 (330 Townsend)	69	2013-09-02 10:51:00	San Francisco Caltrain 2 (330 Townsend)	
2	9444	60	2013-09-03 08:37:00	Redwood City Public Library	24	2013-09-03 08:38:00	Redwood City Public Library	
3	14644	60	2013-09-08 13:55:00	San Francisco Caltrain (Townsend at 4th)	70	2013-09-08 13:56:00	San Francisco Caltrain (Townsend at 4th)	
4	18792	60	2013-09-12 10:09:00	Civic Center BART (7th at Market)	72	2013-09-12 10:10:00	Civic Center BART (7th at Market)	
5	20271	60	2013-09-13 12:43:00	Market at 4th	76	2013-09-13 12:44:00	Market at 4th	
6	22689	60	2013-09-15 21:15:00	Embarcadero at Sansome	60	2013-09-15 21:16:00	Embarcadero at Sansome	
7	32159	60	2013-09-23 18:53:00	Townsend at 7th	65	2013-09-23 18:54:00	Townsend at 7th	
8	57581	60	2013-10-14 14:47:00	Clay at Battery	41	2013-10-14 14:48:00	Clay at Battery	
9	77650	60	2013-10-30 18:22:00	Harry Bridges Plaza (Ferry Building)	50	2013-10-30 18:23:00	Harry Bridges Plaza (Ferry Building)	
10	100171	60	2013-11-18 15:01:00	Temporary Transbay Terminal (Howard at Beale)	55	2013-11-18 15:02:00	Temporary Transbay Terminal (Howard at Beale)	

	id	duration	start_date	start_station_name	start_station_id	end_date	end_station_name	end_
11	110247	60	2013-11-28 09:52:00	2nd at Townsend	61	2013-11-28 09:53:00	2nd at Townsend	
12	122928	60	2013-12-12 09:18:00	Powell at Post (Union Square)	71	2013-12-12 09:19:00	Powell at Post (Union Square)	
13	128828	60	2013-12-17 17:28:00	Steuart at Market	74	2013-12-17 17:29:00	Steuart at Market	
14	136099	60	2013-12-26 15:55:00	2nd at South Park	64	2013-12-26 15:56:00	2nd at South Park	

In [ ]:

```
```

Question 5

In [22]:

```
pd.read_sql("SELECT AVG(duration) AS AverageDuration FROM trip", conn)
```

Out[22]:

	AverageDuration
0	1107.949846

In [23]:

```
pd.read_sql("SELECT AVG(duration) AS AverageDuration FROM trip WHERE start_station_id =
```

Out[23]:

	AverageDuration
0	6357.401109

Question- 6

In [25]:

```
query = '''
        SELECT bike_id,MAX(duration)
    FROM
    (SELECT bike_id, SUM(duration) as dur
     FROM trip
     GROUP BY bike_id
   )

        '''
pd.read_sql(query, conn)
```

Out[25]:

	bike_id	MAX(dur)
0	535	18611693

Question -7

In [1]:
```
Its on word sheet.
```

```
  File "C:\Users\User\AppData\Local\Temp/ipykernel_8552/1700198911.py", line 1
    Its on word sheet.
        ^
SyntaxError: invalid syntax
```

Task 2

Question - 1

In [9]:
```
import sqlite3
import pandas as pd
conn = sqlite3.connect("bikesharing.db")
cur = conn.cursor()
```

In [10]:
```
pd.read_sql_query("SELECT count(start_station_name) as count_station, start_station_id,
```

Out[10]:

	count_station	start_station_id	most_popular_station
0	49092	70	San Francisco Caltrain (Townsend at 4th)
1	33742	69	San Francisco Caltrain 2 (330 Townsend)
2	32934	50	Harry Bridges Plaza (Ferry Building)
3	27713	60	Embarcadero at Sansome
4	26089	55	Temporary Transbay Terminal (Howard at Beale)
5	25837	61	2nd at Townsend
6	24838	74	Steuart at Market
7	24172	77	Market at Sansome
8	23724	65	Townsend at 7th
9	20272	67	Market at 10th
10	20165	76	Market at 4th
11	18496	64	2nd at South Park
12	18378	39	Powell Street BART
13	16306	73	Grant Avenue at Columbus Avenue
14	15940	62	2nd at Folsom
15	15709	56	Beale at Market
16	14811	54	Embarcadero at Bryant

	count_station	start_station_id	most_popular_station
17	14102	72	Civic Center BART (7th at Market)
18	14054	51	Embarcadero at Folsom
19	13526	57	5th at Howard

In [11]:
```
pd.read_sql_query("SELECT count(*) as idle_time, station_id FROM status where station_i
```

Out[11]:

	idle_time	station_id
0	1031180	2

Task - 3

In [14]:
```
pd.read_sql_query("SELECT station_id, AVG(bikes_available) as bikes_avail_avg, AVG(dock
```

Out[14]:

	station_id	bikes_avail_avg	docks_avail_avg
0	2	13.172570	13.761535
1	3	8.461138	6.527884

Task 4

In [13]:
```
pd.read_sql_query("SELECT count(start_station_name) as count_station, start_station_id,
```

Out[13]:

	count_station	start_station_id	least_popular_station
0	23	80	San Jose Government Center
1	67	25	Broadway at Main
2	213	24	Redwood City Public Library
3	224	21	Franklin at Maple
4	287	23	San Mateo County Center
5	311	26	Redwood City Medical Center
6	341	83	Mezes Park
7	436	25	Stanford in Redwood City
8	750	38	Park at Olive
9	840	80	Santa Clara County Civic Center

In [ ]:
```

```