Utilizing SQL to Create Different Visuals with Data that was Given

By: Ashley Krause

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import sqlite3 as sql
```

[169]:
```
#Read in a csv with weather data
weather = pd.read_csv('weather.csv')
print(weather)
conn = sql.connect('weather.db')
weather.to_sql('weather', conn)
```

```
     Year  Month  Tmax  Tmin  Rain    Sun
0    1957      1   8.7   2.7  39.5   53.0
1    1957      2   9.0   2.9  69.8   64.9
2    1957      3  13.9   5.7  25.4   96.7
3    1957      4  14.2   5.2   5.7  169.6
4    1957      5  16.2   6.5  21.3  195.0
..    ...    ...   ...   ...   ...    ...
743  2018     12  10.7   5.2  60.6   40.3
744  2019      1   7.6   2.0  33.2   56.4
745  2019      2  12.4   3.3  34.2  120.2
746  2019      3  13.1   5.8  49.6  119.0
747  2019      4  15.8   5.7  12.8  170.1

[748 rows x 6 columns]
```

[170]:
```
sql_statement = """SELECT Year, AVG(Rain)
                   FROM weather
                   GROUP BY Year
                   HAVING Year NOT LIKE 2019"""


weather_results = pd.read_sql_query(sql_statement, conn)
weather_results
```

[170]:

	Year	AVG(Rain)
0	1957	46.466667
1	1958	61.641667
2	1959	40.608333
3	1960	61.283333
4	1961	46.566667
...	...	...
57	2014	72.000000
58	2015	46.833333
59	2016	49.216667
60	2017	47.683333
61	2018	48.333333

62 rows × 2 columns

[171]:
```
sql_statement = """SELECT Year, Month, Tmin, Tmax
                   FROM weather
                   WHERE Tmin >= (Tmax-4.5)"""

weather_results = pd.read_sql_query(sql_statement, conn)
weather_results
```

[171]:

	Year	Month	Tmin	Tmax
0	1966	1	1.0	5.3
1	1968	12	1.0	5.4
2	1986	2	-2.7	1.7
3	1996	1	3.1	7.2
4	2000	12	4.6	9.0
5	2002	12	5.2	9.0
6	2013	1	2.0	6.5

```
sql_statement = """SELECT Year, AVG(Tmax), MAX(TMax), MIN(Tmax)
                        FROM weather
                        WHERE Month IN (6,7,8)
                        GROUP BY Year
                        HAVING AVG(Tmax) > MAX(Tmax)"""

weather_results = pd.read_sql_query(sql_statement, conn)
weather_results
```

Year	AVG(Tmax)	MAX(Tmax)	MIN(Tmax)
1957	22.400000	23.6	21.1
1959	23.666667	24.7	22.1
1970	22.400000	23.6	21.3
1973	22.333333	23.7	21.4
1975	23.933333	25.9	21.8
1976	25.733333	26.6	25.1
1983	24.300000	27.6	20.8
1984	23.300000	24.4	21.3
1989	24.033333	25.8	22.1
1990	23.400000	26.0	19.5
...	...	...	...
2010	23.366667	25.0	21.6
2013	23.866667	27.0	20.3
2014	23.200000	25.8	21.7
2015	22.700000	23.7	22.2
2016	23.133333	24.7	20.7
2017	23.266667	24.0	22.0
2018	25.666667	28.2	24.2

```
sql_statement = """SELECT *
                   FROM nba_15_16"""

nba_results = pd.read_sql_query(sql_statement, conn)
nba_results
```

	index	player	position	team	salary
0	0	Paul Millsap	PF	Atlanta Hawks	18.671659
1	1	Al Horford	C	Atlanta Hawks	12.000000
2	2	Tiago Splitter	C	Atlanta Hawks	9.756250
3	3	Jeff Teague	PG	Atlanta Hawks	8.000000
4	4	Kyle Korver	SG	Atlanta Hawks	5.746479
...	...	...	...	...	...
412	412	Gary Neal	PG	Washington Wizards	2.139000
413	413	DeJuan Blair	C	Washington Wizards	2.000000
414	414	Kelly Oubre Jr.	SF	Washington Wizards	1.920240
415	415	Garrett Temple	SG	Washington Wizards	1.100602
416	416	Jarell Eddie	SG	Washington Wizards	0.561716

417 rows × 5 columns

[177]:
```
sql_statement = """SELECT player, salary
                   FROM nba_15_16
                   WHERE team = 'Philadelphia 76ers' AND salary >= (SELECT salary
                                                                    FROM nba_15_16
                                                                    WHERE player = 'Nik Stauskas')"""

nba_results = pd.read_sql_query(sql_statement, conn)
nba_results
```

[177]:

	player	salary
0	Gerald Wallace	10.105855
1	Carl Landry	6.500000
2	Joel Embiid	4.626960
3	Jahlil Okafor	4.582680
4	Nerlens Noel	3.457800
5	Nik Stauskas	2.869440

```
sql_statement = """SELECT team
                   FROM nba_15_16
                   WHERE 2.0 <=  (SELECT salary
                                  FROM nba_15_16 AS x
                                  WHERE position IS 'PG')
                                  GROUP BY position"""

nba_results = pd.read_sql_query(sql_statement, conn)
nba_results.head(50)
```

	team
0	Atlanta Hawks
1	Atlanta Hawks
2	Atlanta Hawks
3	Atlanta Hawks
4	Atlanta Hawks

```
sql_statement = """SELECT *
                   FROM nba_20_21"""

nba_results = pd.read_sql_query(sql_statement, conn)
nba_results
```

	index	jersey	name	pos	height	weight	age	team	yos	predraft_team	draft_status	nationality
0	0	15	Precious Achiuwa	SF	81	210	21	Miami Heat	0	Memphis	2020 Rnd 1 Pick 20	Nigeria
1	1	12	Steven Adams	C	84	265	27	New Orleans Pelicans	7	Pittsburgh	2013 Rnd 1 Pick 12	New Zealand
2	2	13	Bam Adebayo	C	73	255	23	Miami Heat	3	Kentucky	2017 Rnd 1 Pick 14	United States
3	3	0	Ty-Shon Alexander	SG	76	195	23	Phoenix Suns	0	Creighton	2020 NBA Draft, Undrafted	United States
4	4	6	Nickeil Alexander-Walker	SG	77	205	22	New Orleans Pelicans	1	Virginia Tech	2019 Rnd 1 Pick 17	Canada
...	...	...	...	...	...	...	...	...	...	...	...	...
490	490	55	Delon Wright	G	77	185	29	Sacramento Kings	5	Utah	2015 Rnd 1 Pick 20	United States
491	491	21	Thaddeus Young	F	80	235	32	Chicago Bulls	13	Georgia Tech	2007 Rnd 1 Pick 12	United States
492	492	11	Trae Young	PG	73	180	22	Atlanta Hawks	2	Oklahoma	2018 Rnd 1 Pick 5	United States
493	493	40	Cody Zeller	F	84	240	28	Charlotte Hornets	7	Indiana	2013 Rnd 1 Pick 4	United States
494	494	40	Ivica Zubac	C	85	240	24	Los Angeles Clippers	4	KK Mega Bemax (Serbia)	2016 Rnd 2 Pick 2	Bosnia and Herzegovina Croatia

495 rows × 12 columns

```
sql_statement = """SELECT team, COUNT(player) AS num_player, AVG(salary) AS avg_salary
                   FROM nba_15_16
                   GROUP BY team
                   ORDER BY COUNT(player) DESC"""

nba_results = pd.read_sql_query(sql_statement, conn)
nba_results
```

	team	num_player	avg_salary
0	Memphis Grizzlies	21	4.466497
1	Charlotte Hornets	18	4.672355
2	Washington Wizards	17	5.296912
3	Utah Jazz	17	3.095993
4	Toronto Raptors	17	4.392507
5	Phoenix Suns	17	2.971813
6	Oklahoma City Thunder	16	6.052010
7	New Orleans Pelicans	16	5.032163
8	Boston Celtics	15	3.352367
9	Portland Trail Blazers	14	3.246206
10	Orlando Magic	14	5.544567
11	Golden State Warriors	14	6.720367
12	Denver Nuggets	14	4.459243
13	Atlanta Hawks	14	4.969507
14	San Antonio Spurs	13	6.511698
15	Philadelphia 76ers	13	3.267796

```
sql_statement = """SELECT position, COUNT(player) AS num_player, AVG(salary) AS avg_salary
                   FROM nba_15_16
                   GROUP BY position
                   ORDER BY COUNT(position) DESC"""

nba_results = pd.read_sql_query(sql_statement, conn)
nba_results
```

	position	num_player	avg_salary
0	SG	96	3.988195
1	PG	85	5.165487
2	PF	85	4.951344
3	SF	82	5.532675
4	C	69	6.082913

[184]:
```
sql_statement = """SELECT COUNT(nb.player)
                   FROM nba_15_16 nb
                   JOIN nba_20_21 n ON (nb.player = n.name)
                   """
```

[185]:
```
nba_results = pd.read_sql_query(sql_statement, conn)
nba_results
```

[185]:

	COUNT(nb.player)
0	158

```
86]: sql_statement = """SELECT nb.player, nb.team AS team15_16, n.team AS team20_21
                        FROM nba_15_16 nb
                        LEFT JOIN nba_20_21 n ON (nb.player = n.name)
                        """

nba_results = pd.read_sql_query(sql_statement, conn)
nba_results
```

86]:

	player	team15_16	team20_21
0	Paul Millsap	Atlanta Hawks	Denver Nuggets
1	Al Horford	Atlanta Hawks	Oklahoma City Thunder
2	Tiago Splitter	Atlanta Hawks	None
3	Jeff Teague	Atlanta Hawks	Milwaukee Bucks
4	Kyle Korver	Atlanta Hawks	None
...	...	...	...
412	Gary Neal	Washington Wizards	None
413	DeJuan Blair	Washington Wizards	None
414	Kelly Oubre Jr.	Washington Wizards	None
415	Garrett Temple	Washington Wizards	Chicago Bulls
416	Jarell Eddie	Washington Wizards	None

417 rows × 3 columns

187]:
```
sql_statement = """SELECT nb.player, nb.team AS team15_16, n.team AS team20_21
                   FROM nba_15_16 nb
                   LEFT JOIN nba_20_21 n ON (nb.player = n.name)
                   WHERE n.team IS NULL"""

nba_results = pd.read_sql_query(sql_statement, conn)
nba_results
```

187]:

	player	team15_16	team20_21
0	Tiago Splitter	Atlanta Hawks	None
1	Kyle Korver	Atlanta Hawks	None
2	Thabo Sefolosha	Atlanta Hawks	None
3	Walter Tavares	Atlanta Hawks	None
4	Jason Richardson	Atlanta Hawks	None
...	...	...	...
254	Ramon Sessions	Washington Wizards	None
255	Gary Neal	Washington Wizards	None
256	DeJuan Blair	Washington Wizards	None
257	Kelly Oubre Jr.	Washington Wizards	None
258	Jarell Eddie	Washington Wizards	None

259 rows × 3 columns

189]:
```
sql_statement = """SELECT nb.team, COUNT(nb.player) AS num_in_both, AVG(n.age) AS avg_age, AVG(nb.salary) AS avg_salary
                   FROM nba_15_16 nb
                   LEFT JOIN nba_20_21 n ON (nb.player = n.name)
                   WHERE nb.player IS n.name
                    GROUP BY nb.team"""

nba_results = pd.read_sql_query(sql_statement, conn)
nba_results
```

189]:

	team	num_in_both	avg_age	avg_salary
0	Atlanta Hawks	7	31.571429	6.724702
1	Boston Celtics	5	28.800000	4.389403
2	Brooklyn Nets	4	30.750000	4.374236
3	Charlotte Hornets	4	29.500000	8.090966
4	Chicago Bulls	6	29.833333	7.137244
5	Cleveland Cavaliers	4	31.750000	18.331968
6	Dallas Mavericks	1	33.000000	1.270964
7	Denver Nuggets	5	28.000000	4.451963
8	Detroit Pistons	6	29.000000	5.111883
9	Golden State Warriors	6	30.833333	9.641412
10	Houston Rockets	6	31.000000	9.173006
11	Indiana Pacers	3	28.666667	6.945582
12	Los Angeles Clippers	4	32.750000	12.643054
13	Los Angeles Lakers	2	25.500000	4.117680