DavidGrice_Final_AnalyzeData

December 18, 2019

0.1 Libraries For Analyzing

[46]:
```
import sqlite3
import pandas as pd
import numpy as np
import os
import matplotlib as plt
```

0.2 Connecting to Database

[2]:
```
path = './AliensExist.db'
conn = sqlite3.connect(path)
curs = conn.cursor()
curs.execute("PRAGMA foreign_keys=ON")
```

[2]:
```
<sqlite3.Cursor at 0x2733f9219d0>
```

0.3 Functions used in analyzing data

[3]:
```
def createView(name,sql):
    curs.execute("DROP VIEW IF EXISTS " + name)
    curs.execute("CREATE VIEW " + name + " AS " + sql)
    conn.commit()
```

0.4 Setup views for answering Question 2-A

[4]:
```
# Combine the flight info with the origin airport and destination airport
sql = """SELECT tFlightPlaneInfo.oAirport, tFlightPlaneInfo.year,
                tFlightPlaneInfo.dAirport, tFlightPlaneInfo.flights
                    FROM tFlightOrigAirportInfo
                        JOIN tFlightPlaneInfo ON tFlightOrigAirportInfo.
 ↪orig_airport = tFlightPlaneInfo.oAirport
```

```
                        JOIN tFlightDestAirportInfo ON tFlightPlaneInfo.
 ↪dAirport = tFlightDestAirportInfo.dest_airport"""
createView('vPlaneDatesArriveLeave',sql)
pd.read_sql(sql,conn)
```

[4]:

	oAirport	year	dAirport	flights
0	MHK	2000	MLU	1
1	MHK	2001	CLL	1
2	MHK	2001	MSY	1
3	MHK	2001	ROC	1
4	MHK	2002	ACT	1
…	…	…	…	…
3606798	BOK	2008	CEC	1
3606799	BOK	2009	CEC	1
3606800	BIH	2009	SFO	1
3606801	MQJ	2009	KY3	1
3606802	LCI	2007	CHO	1

[3606803 rows x 4 columns]

[5]:
```
# Gather the unique years to combine with UFO info
sql = """SELECT DISTINCT vPlaneDatesArriveLeave.year
            FROM vPlaneDatesArriveLeave
            ORDER BY year ASC"""
createView('vPlanesDate',sql)
pd.read_sql(sql,conn)
```

[5]:

	year
0	1990
1	1991
2	1992
3	1993
4	1994
5	1995
6	1996
7	1997
8	1998
9	1999
10	2000
11	2001
12	2002
13	2003
14	2004
15	2005
16	2006
17	2007
18	2008

```
19  2009
```

[6]:
```
# Combine the UFO dates with the sightings info
sql = """SELECT tUFOOriginalDate.sighting_id, tUFOOriginalDate.origMonth,␣
 ↪tUFOOriginalDate.origDay,
                tUFOOriginalDate.origYear, tUFOSightingInfo.city,␣
 ↪tUFOSightingInfo.shape,
                tUFOSightingInfo.time, tUFOSightingInfo.count
                FROM tUFOOriginalDate
                    JOIN tUFOSightingInfo ON tUFOOriginalDate.sighting_id =␣
 ↪tUFOSightingInfo.sighting_id"""
createView('vUFOSightingsWithDate',sql)
pd.read_sql(sql,conn)
```

[6]:

	sighting_id	origMonth	origDay	origYear	city	shape	time	\
0	0	10	10	1949	SAN MARCOS	CYLINDER	20:30	
1	1	10	10	1949	LACKLAND AFB	LIGHT	21:00	
2	2	10	10	1956	EDNA	CIRCLE	21:00	
3	3	10	10	1960	KANEOHE	LIGHT	20:00	
4	4	10	10	1961	BRISTOL	SPHERE	19:00	
…	…	…	…	…	…	…	…	…
74779	74779	9	9	2013	NASHVILLE	LIGHT	21:15	
74780	74780	9	9	2013	BOISE	CIRCLE	22:00	
74781	74781	9	9	2013	NAPA	OTHER	22:00	
74782	74782	9	9	2013	VIENNA	CIRCLE	22:20	
74783	74783	9	9	2013	EDMOND	CIGAR	23:00	

	count
0	1
1	1
2	1
3	1
4	1
…	…
74779	1
74780	1
74781	1
74782	1
74783	1

[74784 rows x 8 columns]

0.5 Question 2-A

[7]:
```
# Question 2-a
# Find Year with most sightings reported joined with flight information
sql = """SELECT vUFOSightingsWithDate.origYear AS Year,
                SUM(vUFOSightingsWithDate.count) AS TotalSightings
                    FROM vUFOSightingsWithDate
                        INNER JOIN vPlanesDate ON vUFOSightingsWithDate.
 ↪origYear = vPlanesDate.year
                    GROUP BY vUFOSightingsWithDate.origYear
                    ORDER BY TotalSightings DESC"""
createView('vTopUFOSightingsYear',sql)
pd.read_sql(sql,conn).head(1)
```

[7]:

	Year	TotalSightings
0	2008	4529

0.6 Question 2-B

[8]:
```
# Question 2-b-i
# Top 5 cities
sql = """SELECT vPlanesDate.year AS Year, vUFOSightingsWithDate.city AS City,
                SUM(vUFOSightingsWithDate.count) AS TotalSightings
                    FROM vPlanesDate
                        JOIN vUFOSightingsWithDate ON vPlanesDate.year =␣
 ↪vUFOSightingsWithDate.origYear
                    WHERE vPlanesDate.year = 2008
                    GROUP BY vUFOSightingsWithDate.city
                    ORDER BY TotalSightings DESC"""
topCities = pd.read_sql(sql,conn).head(5)
topFiveCities = pd.DataFrame(topCities,␣
 ↪columns=['Year','City','TotalSightings'],)

# Question 2-b-ii
# Bottom 5 cities
sql = """SELECT vPlanesDate.year AS Year, vUFOSightingsWithDate.city AS City,
                SUM(vUFOSightingsWithDate.count) AS TotalSightings
                    FROM vPlanesDate
                        JOIN vUFOSightingsWithDate ON vPlanesDate.year =␣
 ↪vUFOSightingsWithDate.origYear
                    WHERE vPlanesDate.year = 2008
                    GROUP BY vUFOSightingsWithDate.city
                    ORDER BY TotalSightings DESC"""
botCities = pd.read_sql(sql,conn).tail(5)
```

```
botFiveCities = pd.DataFrame(botCities,␣
↪columns=['Year','City','TotalSightings'],)
topBotCities = pd.concat([topFiveCities,botFiveCities], sort = False)
topBotCities
```

[8]:

	Year	City	TotalSightings
0	2008	PHOENIX	37
1	2008	HOUSTON	28
2	2008	SAN DIEGO	27
3	2008	LOS ANGELES	27
4	2008	SEATTLE	26
2604	2008	ADAMSVILLE	1
2605	2008	ADAMS	1
2606	2008	ACWORTH	1
2607	2008	ACTON	1
2608	2008	ACHUSNET	1

0.7 Setup Views for Question 2-C

[9]:
```
# Origin to Destination combined flights
sql = """SELECT vPlaneDatesArriveLeave.year AS Year, tFlightOrigAirportInfo.
↪orig_city AS City,
                SUM(vPlaneDatesArriveLeave.flights) AS TotalOriginalFlights
                FROM vPlaneDatesArriveLeave
                    JOIN tFlightOrigAirportInfo ON vPlaneDatesArriveLeave.
↪dAirport = tFlightOrigAirportInfo.orig_airport
                WHERE vPlaneDatesArriveLeave.year = 2008
                GROUP BY tFlightOrigAirportInfo.orig_city
                ORDER BY TotalOriginalFlights DESC"""
createView('vFlightsOrigCity',sql)
pd.read_sql(sql,conn)
```

[9]:

	Year	City	TotalOriginalFlights
0	2008	CHICAGO	440964
1	2008	ATLANTA	395197
2	2008	DALLAS	327994
3	2008	NEW YORK	296033
4	2008	HOUSTON	265651
..	…	…	…
380	2008	BIG RAPIDS	1
381	2008	AUBURN	1
382	2008	ANNISTON	1
383	2008	AMES	1
384	2008	ALTUS	1

[385 rows x 3 columns]

[10]:
```
# Destination to Origin combined flights
sql = """SELECT vPlaneDatesArriveLeave.year AS Year, tFlightDestAirportInfo.
 ↪dest_city AS City,
                SUM(vPlaneDatesArriveLeave.flights) AS TotalDestinationFlights
                FROM vPlaneDatesArriveLeave
                    JOIN tFlightDestAirportInfo ON vPlaneDatesArriveLeave.
 ↪oAirport = tFlightDestAirportInfo.dest_airport
                WHERE vPlaneDatesArriveLeave.year = 2008
                GROUP BY tFlightDestAirportInfo.dest_city
                ORDER BY TotalDestinationFlights DESC"""
createView('vFlightsDestCity',sql)
pd.read_sql(sql,conn)
```

[10]:

	Year	City	TotalDestinationFlights
0	2008	CHICAGO	441629
1	2008	ATLANTA	395732
2	2008	DALLAS	328133
3	2008	NEW YORK	295969
4	2008	HOUSTON	265042
..	…	…	…
381	2008	DYERSBURG	1
382	2008	CARBONDALE	1
383	2008	BROOKINGS	1
384	2008	BIG RAPIDS	1
385	2008	AMES	1

[386 rows x 3 columns]

0.8 Question 2-C

[11]:
```
# 2-c-i
# Top 5 cities with flights
sql = """SELECT vFlightsOrigCity.year, vFlightsOrigCity.City,
                SUM(TotalOriginalFlights+TotalDestinationFlights) AS␣
 ↪TotalFlights
                FROM vFlightsOrigCity
                JOIN vFlightsDestCity ON vFlightsOrigCity.year =␣
 ↪vFlightsDestCity.year
                                            AND vFlightsOrigCity.City =␣
 ↪vFlightsDestCity.City
                GROUP BY vFlightsOrigCity.City
                ORDER BY SUM(TotalOriginalFlights+TotalDestinationFlights)␣
 ↪DESC"""
```

```
topCityFlights = pd.read_sql(sql,conn).head(5)
topFiveCityFlights = pd.DataFrame(topCityFlights,␣
↪columns=['Year','City','TotalFlights'],)

# 2-c-ii
# Bottom 5 cities with flights
sql = """SELECT vFlightsOrigCity.year, vFlightsOrigCity.City,
                SUM(TotalOriginalFlights+TotalDestinationFlights) AS␣
↪TotalFlights
                FROM vFlightsOrigCity
                JOIN vFlightsDestCity ON vFlightsOrigCity.year =␣
↪vFlightsDestCity.year
                                          AND vFlightsOrigCity.City =␣
↪vFlightsDestCity.City
                GROUP BY vFlightsOrigCity.City
                ORDER BY SUM(TotalOriginalFlights+TotalDestinationFlights)␣
↪DESC"""
botCityFlights = pd.read_sql(sql,conn).tail(5)
botFiveCityFlights = pd.DataFrame(botCityFlights,␣
↪columns=['Year','City','TotalFlights'],)
topBotCityFlights = pd.concat([topFiveCityFlights,botFiveCityFlights], sort =␣
↪False)
topBotCityFlights
```

[11]:

	Year	City	TotalFlights
0	2008	CHICAGO	882593
1	2008	ATLANTA	790929
2	2008	DALLAS	656127
3	2008	NEW YORK	592002
4	2008	HOUSTON	530693
375	2008	HOLLAND	2
376	2008	FRANKFORT	2
377	2008	CARBONDALE	2
378	2008	BIG RAPIDS	2
379	2008	AMES	2

0.9 Setup Views for Question 2-D

[12]:
```
# States with UFO Sightings
sql = """SELECT vPlanesDate.year AS Year, tUFOCityStateCountry.state AS State,
                SUM(vUFOSightingsWithDate.count) AS TotalSightings
                    FROM vPlanesDate
                        JOIN vUFOSightingsWithDate ON vPlanesDate.year =␣
↪vUFOSightingsWithDate.origYear
```

```
                        JOIN tUFOCityStateCountry ON vUFOSightingsWithDate.
↪sighting_id = tUFOCityStateCountry.sighting_id
                    WHERE vPlanesDate.year = 2008
                    GROUP BY tUFOCityStateCountry.state
                    ORDER BY TotalSightings DESC"""
createView('vUFOTotalSightingsByState',sql)
pd.read_sql(sql,conn)
```

[12]:

	Year	State	TotalSightings
0	2008	CA	655
1	2008	TX	341
2	2008	FL	270
3	2008	NY	195
4	2008	0	188
5	2008	WA	185
6	2008	PA	179
7	2008	AZ	164
8	2008	IL	151
9	2008	MI	129
10	2008	OH	124
11	2008	NC	123
12	2008	IN	115
13	2008	MO	97
14	2008	NJ	96
15	2008	GA	94
16	2008	VA	92
17	2008	TN	92
18	2008	CO	91
19	2008	OR	88
20	2008	MA	70
21	2008	KY	67
22	2008	WI	66
23	2008	NV	60
24	2008	MD	58
25	2008	CT	57
26	2008	NM	53
27	2008	MN	53
28	2008	OK	51
29	2008	LA	43
30	2008	AL	43
31	2008	SC	38
32	2008	AR	36
33	2008	NH	32
34	2008	IA	30
35	2008	WV	28
36	2008	KS	28
37	2008	UT	26

```
38  2008    ME              24
39  2008    MS              22
40  2008    ID              21
41  2008    MT              18
42  2008    NE              17
43  2008    RI              16
44  2008    HI              16
45  2008    VT              12
46  2008    WY              11
47  2008    ON              11
48  2008    DE              11
49  2008    ND              10
50  2008    AK               9
51  2008    SD               7
52  2008    DC               7
53  2008    AB               3
54  2008    PE               2
55  2008    SK               1
56  2008    QC               1
57  2008    NS               1
58  2008    BC               1
```

0.10 Question 2-D

[13]:
```
# Question 2-d-i
# Top 5 states
sql = """SELECT vPlanesDate.year AS Year, tUFOCityStateCountry.state AS State,
                SUM(vUFOSightingsWithDate.count) AS TotalSightings
                    FROM vPlanesDate
                        JOIN vUFOSightingsWithDate ON vPlanesDate.year =␣
 ↪vUFOSightingsWithDate.origYear
                        JOIN tUFOCityStateCountry ON vUFOSightingsWithDate.
 ↪sighting_id = tUFOCityStateCountry.sighting_id
                    WHERE vPlanesDate.year = 2008
                    GROUP BY tUFOCityStateCountry.state
                    ORDER BY TotalSightings DESC"""
topStateSights = pd.read_sql(sql,conn).head(5)
topFiveStateSights = pd.DataFrame(topStateSights,␣
 ↪columns=['Year','State','TotalSightings'],)

# Question 2-d-ii
# Bottom 5 states
sql = """SELECT vPlanesDate.year As Year, tUFOCityStateCountry.state As State,
                SUM(vUFOSightingsWithDate.count) AS TotalSightings
                    FROM vPlanesDate
```

```
                        JOIN vUFOSightingsWithDate ON vPlanesDate.year =␣
↪vUFOSightingsWithDate.origYear
                        JOIN tUFOCityStateCountry ON vUFOSightingsWithDate.
↪sighting_id = tUFOCityStateCountry.sighting_id
                    WHERE vPlanesDate.year = 2008
                    GROUP BY tUFOCityStateCountry.state
                    ORDER BY TotalSightings DESC"""
botStateSights = pd.read_sql(sql,conn).tail(5)
botFiveStateSights = pd.DataFrame(botStateSights,␣
↪columns=['Year','State','TotalSightings'],)
topBotStateSights = pd.concat([topFiveStateSights,botFiveStateSights], sort =␣
↪False)
topBotStateSights
```

[13]:

	Year	State	TotalSightings
0	2008	CA	655
1	2008	TX	341
2	2008	FL	270
3	2008	NY	195
4	2008	0	188
54	2008	PE	2
55	2008	SK	1
56	2008	QC	1
57	2008	NS	1
58	2008	BC	1

0.11 Setup Views for Question 2-E

[14]:
```
# Origin to Destination combined flights
sql = """SELECT vPlaneDatesArriveLeave.year AS Year, tFlightOrigAirportInfo.
↪orig_state AS State,
                SUM(vPlaneDatesArriveLeave.flights) AS TotalOriginalFlights
                FROM vPlaneDatesArriveLeave
                    JOIN tFlightOrigAirportInfo ON vPlaneDatesArriveLeave.
↪dAirport = tFlightOrigAirportInfo.orig_airport
                WHERE vPlaneDatesArriveLeave.year = 2008
                GROUP BY tFlightOrigAirportInfo.orig_state
                ORDER BY TotalOriginalFlights DESC"""
createView('vFlightsOrigState',sql)
pd.read_sql(sql,conn)
```

[14]:

	Year	State	TotalOriginalFlights
0	2008	TX	786876
1	2008	CA	701768
2	2008	FL	489114

3	2008	IL	468576
4	2008	NY	427239
5	2008	GA	424069
6	2008	OH	340679
7	2008	NC	334166
8	2008	PA	302058
9	2008	MI	250578
10	2008	DC	247136
11	2008	TN	235664
12	2008	AZ	217029
13	2008	MO	195569
14	2008	NV	191898
15	2008	MN	191795
16	2008	WA	180919
17	2008	NJ	156999
18	2008	MA	128528
19	2008	UT	121577
20	2008	OR	115703
21	2008	WI	111985
22	2008	MD	107225
23	2008	IN	102411
24	2008	LA	86696
25	2008	HI	80247
26	2008	SC	74590
27	2008	AK	58279
28	2008	AL	56028
29	2008	VA	55928
30	2008	OK	54410
31	2008	NM	50740
32	2008	IA	42927
33	2008	CT	41515
34	2008	AR	40998
35	2008	NE	35172
36	2008	RI	31480
37	2008	MT	27783
38	2008	MS	27597
39	2008	NH	25722
40	2008	ME	24725
41	2008	KS	22956
42	2008	CO	18905
43	2008	ND	16556
44	2008	VT	16487
45	2008	SD	15864
46	2008	KY	15418
47	2008	WV	13670
48	2008	ID	9646
49	2008	WY	8694

```
50  2008    DE                    230
```

[15]:
```
# Destination to Origin combined flights
sql = """SELECT vPlaneDatesArriveLeave.year AS Year, tFlightDestAirportInfo.
 ↪dest_state AS State,
                SUM(vPlaneDatesArriveLeave.flights) AS TotalDestinationFlights
                FROM vPlaneDatesArriveLeave
                    JOIN tFlightDestAirportInfo ON vPlaneDatesArriveLeave.
 ↪oAirport = tFlightDestAirportInfo.dest_airport
                WHERE vPlaneDatesArriveLeave.year = 2008
                GROUP BY tFlightDestAirportInfo.dest_state
                ORDER BY TotalDestinationFlights DESC"""
createView('vFlightsDestState',sql)
pd.read_sql(sql,conn)
```

[15]:

	Year	State	TotalDestinationFlights
0	2008	TX	786072
1	2008	CA	701771
2	2008	FL	488571
3	2008	IL	468905
4	2008	NY	426738
5	2008	GA	424093
6	2008	OH	341755
7	2008	NC	334671
8	2008	PA	301136
9	2008	MI	250880
10	2008	DC	247261
11	2008	TN	240352
12	2008	AZ	217346
13	2008	MO	195384
14	2008	MN	192093
15	2008	NV	191162
16	2008	WA	179897
17	2008	NJ	157664
18	2008	MA	128612
19	2008	UT	121635
20	2008	OR	115674
21	2008	WI	112076
22	2008	MD	107084
23	2008	IN	102212
24	2008	LA	85888
25	2008	HI	78628
26	2008	SC	74199
27	2008	AK	61856
28	2008	AL	55691
29	2008	VA	55660
30	2008	OK	54127

```
31  2008    NM                    50470
32  2008    IA                    42523
33  2008    CT                    41237
34  2008    AR                    40958
35  2008    NE                    35143
36  2008    RI                    31137
37  2008    MS                    27601
38  2008    MT                    27538
39  2008    NH                    25437
40  2008    ME                    24384
41  2008    KS                    22636
42  2008    CO                    18597
43  2008    ND                    16525
44  2008    VT                    16377
45  2008    SD                    15879
46  2008    KY                    15597
47  2008    WV                    13400
48  2008    ID                     9442
49  2008    WY                     8694
50  2008    DE                      162
```

[16]:
```
# Top States with flights
sql = """SELECT vFlightsOrigState.year, vFlightsOrigState.State,
                SUM(TotalOriginalFlights+TotalDestinationFlights) AS␣
 ↪TotalFlights
                FROM vFlightsOrigState
                JOIN vFlightsDestState ON vFlightsOrigState.year =␣
 ↪vFlightsDestState.year
                                          AND vFlightsOrigState.State =␣
 ↪vFlightsDestState.State
                GROUP BY vFlightsOrigState.State
                ORDER BY SUM(TotalOriginalFlights+TotalDestinationFlights)␣
 ↪DESC"""
createView('vTotalFlightsByStateByYear',sql)
pd.read_sql(sql,conn)
```

[16]:
```
    Year State  TotalFlights
0   2008    TX       1572948
1   2008    CA       1403539
2   2008    FL        977685
3   2008    IL        937481
4   2008    NY        853977
5   2008    GA        848162
6   2008    OH        682434
7   2008    NC        668837
8   2008    PA        603194
9   2008    MI        501458
```

10	2008	DC	494397
11	2008	TN	476016
12	2008	AZ	434375
13	2008	MO	390953
14	2008	MN	383888
15	2008	NV	383060
16	2008	WA	360816
17	2008	NJ	314663
18	2008	MA	257140
19	2008	UT	243212
20	2008	OR	231377
21	2008	WI	224061
22	2008	MD	214309
23	2008	IN	204623
24	2008	LA	172584
25	2008	HI	158875
26	2008	SC	148789
27	2008	AK	120135
28	2008	AL	111719
29	2008	VA	111588
30	2008	OK	108537
31	2008	NM	101210
32	2008	IA	85450
33	2008	CT	82752
34	2008	AR	81956
35	2008	NE	70315
36	2008	RI	62617
37	2008	MT	55321
38	2008	MS	55198
39	2008	NH	51159
40	2008	ME	49109
41	2008	KS	45592
42	2008	CO	37502
43	2008	ND	33081
44	2008	VT	32864
45	2008	SD	31743
46	2008	KY	31015
47	2008	WV	27070
48	2008	ID	19088
49	2008	WY	17388
50	2008	DE	392

0.12 Question 2-E

[17]:
```
# 2-e-i
# Top 5 states with flights
sql = """SELECT vFlightsOrigState.year, vFlightsOrigState.State,
                SUM(TotalOriginalFlights+TotalDestinationFlights) AS␣
 ↪TotalFlights
                FROM vFlightsOrigState
                JOIN vFlightsDestState ON vFlightsOrigState.year =␣
 ↪vFlightsDestState.year
                                          AND vFlightsOrigState.State =␣
 ↪vFlightsDestState.State
                GROUP BY vFlightsOrigState.State
                ORDER BY SUM(TotalOriginalFlights+TotalDestinationFlights)␣
 ↪DESC"""
topStateFlights = pd.read_sql(sql,conn).head(5)
topFiveStateFlights = pd.DataFrame(topStateFlights,␣
 ↪columns=['Year','State','TotalFlights'],)

# 2-e-ii
# Bottom 5 states with flights
sql = """SELECT vFlightsOrigState.year, vFlightsOrigState.State,
                SUM(TotalOriginalFlights+TotalDestinationFlights) AS␣
 ↪TotalFlights
                FROM vFlightsOrigState
                JOIN vFlightsDestState ON vFlightsOrigState.year =␣
 ↪vFlightsDestState.year
                                          AND vFlightsOrigState.State =␣
 ↪vFlightsDestState.State
                GROUP BY vFlightsOrigState.State
                ORDER BY SUM(TotalOriginalFlights+TotalDestinationFlights)␣
 ↪DESC"""
botStateFlights = pd.read_sql(sql,conn).tail(5)
botFiveStateFlights = pd.DataFrame(botStateFlights,␣
 ↪columns=['Year','State','TotalFlights'],)
topBotStateFlights = pd.concat([topFiveStateFlights,botFiveStateFlights], sort␣
 ↪= False)
topBotStateFlights
```

[17]:

	Year	State	TotalFlights
0	2008	TX	1572948
1	2008	CA	1403539
2	2008	FL	977685
3	2008	IL	937481
4	2008	NY	853977
46	2008	KY	31015

```
47  2008    WV         27070
48  2008    ID         19088
49  2008    WY         17388
50  2008    DE           392
```

0.13 Question 2-F

[18]:
```
# Question 2-f
# Find three common shapes with most sightings reported joined with flight␣
 ↪information
sql = """SELECT vUFOSightingsWithDate.origYear AS Year, vUFOSightingsWithDate.
 ↪shape AS Shape,
                SUM(vUFOSightingsWithDate.count) AS TotalSightings
                    FROM vUFOSightingsWithDate
                        INNER JOIN vPlanesDate ON vUFOSightingsWithDate.
 ↪origYear = vPlanesDate.year
                    WHERE vUFOSightingsWithDate.origYear = 2008
                    GROUP BY vUFOSightingsWithDate.shape
                    ORDER BY TotalSightings DESC"""
createView('vTopUFOShapeSightingsForYear',sql)
pd.read_sql(sql,conn).head(3)
```

[18]:
```
   Year     Shape  TotalSightings
0  2008     LIGHT             990
1  2008  TRIANGLE             463
2  2008    CIRCLE             380
```

0.14 Question 2-G

[19]:
```
# Question 2-g
# Find three common shapes with most sightings reported joined with flight␣
 ↪information
sql = """SELECT vUFOSightingsWithDate.origYear AS Year, vUFOSightingsWithDate.
 ↪time AS Time,
                SUM(vUFOSightingsWithDate.count) AS TotalSightings
                    FROM vUFOSightingsWithDate
                        INNER JOIN vPlanesDate ON vUFOSightingsWithDate.
 ↪origYear = vPlanesDate.year
                    WHERE vUFOSightingsWithDate.origYear = 2008
                    GROUP BY vUFOSightingsWithDate.time
                    ORDER BY TotalSightings DESC"""
createView('vTopUFOCommonTimeSightingsForYear',sql)
pd.read_sql(sql,conn).head(1)
```

[19]:

	Year	Time	TotalSightings
0	2008	23:00	215

0.15 Question 2-H

[20]:
```
# Question 2-h
# State list by year with total sightings to total flights and ratio
sql = """SELECT vTotalFlightsByStateByYear.Year, vUFOTotalSightingsByState.
 ↪State, vUFOTotalSightingsByState.TotalSightings,
                vTotalFlightsByStateByYear.TotalFlights,
                "%" || ((vUFOTotalSightingsByState.TotalSightings/
 ↪vTotalFlightsByStateByYear.TotalFlights)*1.0) AS SightToFlightRatio
                FROM  vUFOTotalSightingsByState
                JOIN vTotalFlightsByStateByYear ON vUFOTotalSightingsByState.
 ↪State = vTotalFlightsByStateByYear.State"""
pd.read_sql(sql,conn)
```

[20]:

	Year	State	TotalSightings	TotalFlights	SightToFlightRatio
0	2008	CA	655	1403539	%0.0
1	2008	TX	341	1572948	%0.0
2	2008	FL	270	977685	%0.0
3	2008	NY	195	853977	%0.0
4	2008	WA	185	360816	%0.0
5	2008	PA	179	603194	%0.0
6	2008	AZ	164	434375	%0.0
7	2008	IL	151	937481	%0.0
8	2008	MI	129	501458	%0.0
9	2008	OH	124	682434	%0.0
10	2008	NC	123	668837	%0.0
11	2008	IN	115	204623	%0.0
12	2008	MO	97	390953	%0.0
13	2008	NJ	96	314663	%0.0
14	2008	GA	94	848162	%0.0
15	2008	VA	92	111588	%0.0
16	2008	TN	92	476016	%0.0
17	2008	CO	91	37502	%0.0
18	2008	OR	88	231377	%0.0
19	2008	MA	70	257140	%0.0
20	2008	KY	67	31015	%0.0
21	2008	WI	66	224061	%0.0
22	2008	NV	60	383060	%0.0
23	2008	MD	58	214309	%0.0
24	2008	CT	57	82752	%0.0
25	2008	NM	53	101210	%0.0
26	2008	MN	53	383888	%0.0

27	2008	OK	51	108537	%0.0
28	2008	LA	43	172584	%0.0
29	2008	AL	43	111719	%0.0
30	2008	SC	38	148789	%0.0
31	2008	AR	36	81956	%0.0
32	2008	NH	32	51159	%0.0
33	2008	IA	30	85450	%0.0
34	2008	WV	28	27070	%0.0
35	2008	KS	28	45592	%0.0
36	2008	UT	26	243212	%0.0
37	2008	ME	24	49109	%0.0
38	2008	MS	22	55198	%0.0
39	2008	ID	21	19088	%0.0
40	2008	MT	18	55321	%0.0
41	2008	NE	17	70315	%0.0
42	2008	RI	16	62617	%0.0
43	2008	HI	16	158875	%0.0
44	2008	VT	12	32864	%0.0
45	2008	WY	11	17388	%0.0
46	2008	DE	11	392	%0.0
47	2008	ND	10	33081	%0.0
48	2008	AK	9	120135	%0.0
49	2008	SD	7	31743	%0.0
50	2008	DC	7	494397	%0.0

0.16 Question 2-I

[21]:
```
# Question 2-i
# Sightings in Williamsburg, VA
sql = """SELECT vUFOSightingsWithDate.city AS City, tUFOCityStateCountry.state␣
 ↪AS State,
                SUM(vUFOSightingsWithDate.count) AS TotalSightings
                    FROM vUFOSightingsWithDate
                        JOIN tUFOCityStateCountry ON vUFOSightingsWithDate.
 ↪sighting_id = tUFOCityStateCountry.sighting_id
                    WHERE State = "VA" AND vUFOSightingsWithDate.city =␣
 ↪"WILLIAMSBURG" """
pd.read_sql(sql,conn)
```

[21]:

	City	State	TotalSightings
0	WILLIAMSBURG	VA	16

0.17 Bonus Question

[37]:
```
# Find Year with most sightings reported joined with flight information
sql = """SELECT vUFOSightingsWithDate.origYear AS Year,
                SUM(vUFOSightingsWithDate.count) AS TotalSightings
                    FROM vUFOSightingsWithDate
                        INNER JOIN vPlanesDate ON vUFOSightingsWithDate.
 ↪origYear = vPlanesDate.year
                    GROUP BY vUFOSightingsWithDate.origYear
                    ORDER BY vUFOSightingsWithDate.origYear DESC"""
ufoBonusSight = pd.read_sql(sql,conn)
ufoBonusSightDF = pd.DataFrame(ufoBonusSight,␣
 ↪columns=['Year','TotalSightings'],)
ufoBonusSightDF
```

[37]:

	Year	TotalSightings
0	2009	4185
1	2008	4529
2	2007	3948
3	2006	3374
4	2005	3733
5	2004	3834
6	2003	3481
7	2002	2889
8	2001	2830
9	2000	2567
10	1999	2655
11	1998	1663
12	1997	1173
13	1996	771
14	1995	1042
15	1994	398
16	1993	281
17	1992	230
18	1991	222
19	1990	236

[38]:
```
# Origin to Destination combined flights
sql = """SELECT vPlaneDatesArriveLeave.year AS Year,
                SUM(vPlaneDatesArriveLeave.flights) AS TotalOriginalFlights
                FROM vPlaneDatesArriveLeave
                    JOIN tFlightOrigAirportInfo ON vPlaneDatesArriveLeave.
 ↪dAirport = tFlightOrigAirportInfo.orig_airport
                GROUP BY vPlaneDatesArriveLeave.year
                ORDER BY vPlaneDatesArriveLeave.year DESC"""
createView('vFlightsOrigBonus',sql)
pd.read_sql(sql,conn)
```

[38]:

	Year	TotalOriginalFlights
0	2009	7209225
1	2008	7782824
2	2007	8165853
3	2006	8130069
4	2005	8489757
5	2004	8441579
6	2003	8041906
7	2002	6788772
8	2001	6644521
9	2000	6826495
10	1999	6546752
11	1998	6252185
12	1997	6207831
13	1996	6203237
14	1995	6072590
15	1994	5650424
16	1993	5366247
17	1992	5267990
18	1991	5091717
19	1990	5097235

[39]:
```
sql = """SELECT vPlaneDatesArriveLeave.year AS Year,
                SUM(vPlaneDatesArriveLeave.flights) AS TotalDestinationFlights
                FROM vPlaneDatesArriveLeave
                    JOIN tFlightDestAirportInfo ON vPlaneDatesArriveLeave.
 ↪oAirport = tFlightDestAirportInfo.dest_airport
                GROUP BY vPlaneDatesArriveLeave.year
                ORDER BY vPlaneDatesArriveLeave.year  DESC"""
createView('vFlightsDestBonus',sql)
pd.read_sql(sql,conn)
```

[39]:

	Year	TotalDestinationFlights
0	2009	7209238
1	2008	7782830
2	2007	8165858
3	2006	8130064
4	2005	8489770
5	2004	8441590
6	2003	8041898
7	2002	6788776
8	2001	6644524
9	2000	6826496
10	1999	6546752
11	1998	6252185
12	1997	6207832
13	1996	6203237

```
14  1995                  6072590
15  1994                  5650424
16  1993                  5366248
17  1992                  5267990
18  1991                  5091717
19  1990                  5097236
```

[41]:
```
sql = """SELECT vFlightsOrigBonus.Year,
                SUM(TotalOriginalFlights+TotalDestinationFlights) AS␣
 ↪TotalFlights
                FROM vFlightsOrigBonus
                    JOIN vFlightsDestBonus ON vFlightsOrigBonus.year =␣
 ↪vFlightsDestBonus.year
                GROUP BY vFlightsOrigBonus.Year
                ORDER BY vFlightsOrigBonus.Year DESC"""
flightBonusSight = pd.read_sql(sql,conn)
flightBonusSightDF = pd.DataFrame(flightBonusSight,␣
 ↪columns=['Year','TotalFlights'],)
flightBonusSightDF
```

[41]:

	Year	TotalFlights
0	2009	14418463
1	2008	15565654
2	2007	16331711
3	2006	16260133
4	2005	16979527
5	2004	16883169
6	2003	16083804
7	2002	13577548
8	2001	13289045
9	2000	13652991
10	1999	13093504
11	1998	12504370
12	1997	12415663
13	1996	12406474
14	1995	12145180
15	1994	11300848
16	1993	10732495
17	1992	10535980
18	1991	10183434
19	1990	10194471

[42]:
```
ALIENS = pd.merge(ufoBonusSightDF, flightBonusSight, on='Year')
ALIENS
```

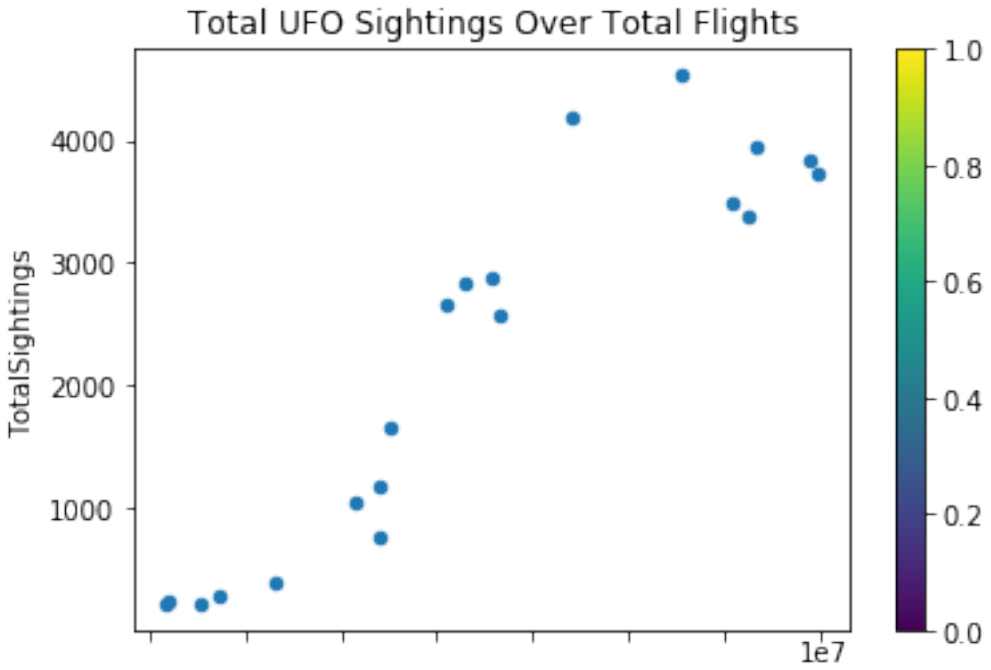
[42]:

	Year	TotalSightings	TotalFlights
0	2009	4185	14418463

1	2008	4529	15565654
2	2007	3948	16331711
3	2006	3374	16260133
4	2005	3733	16979527
5	2004	3834	16883169
6	2003	3481	16083804
7	2002	2889	13577548
8	2001	2830	13289045
9	2000	2567	13652991
10	1999	2655	13093504
11	1998	1663	12504370
12	1997	1173	12415663
13	1996	771	12406474
14	1995	1042	12145180
15	1994	398	11300848
16	1993	281	10732495
17	1992	230	10535980
18	1991	222	10183434
19	1990	236	10194471

[63]:
```
ALIENS.plot.scatter(x='TotalFlights', y = 'TotalSightings',
                    title = "Total UFO Sightings Over Total Flights",
                   colormap = 'viridis')
```

[63]: <matplotlib.axes._subplots.AxesSubplot at 0x2734fe1d748>



[ ]:
```
conn.close()
```