In [2]:
```
##Connect to SQL server, create table and import data

import psycopg2
#from sqlalchemy import create_engine
#engine=create_engine('postgresql://postgres:postgres@localhost:5432/mydb')
%load_ext sql
%sql postgresql://localhost/mydb
```

```
The sql extension is already loaded. To reload it, use:
  %reload_ext sql
(psycopg2.OperationalError) could not connect to server: Connection
refused
        Is the server running on host "localhost" (::1) and acceptin
g
        TCP/IP connections on port 5432?
could not connect to server: Connection refused
        Is the server running on host "localhost" (127.0.0.1) and ac
cepting
        TCP/IP connections on port 5432?

(Background on this error at: http://sqlalche.me/e/13/e3q8)
Connection info needed in SQLAlchemy format, example:
               postgresql://username:password@hostname/dbname
               or an existing connection: dict_keys([])
```

In [13]:
```
%%sql
create table airport_code(City varchar(255),State varchar(255),Country
varchar(255),IATA varchar(255));
create table airlines(Year int,Month int,DayofMonth int,DayofWeek int,
DepTime varchar(255),CRSDepTime int,ArrTime varchar(255),CRSArrTime in
t,UniqueCarrier varchar(255),FlightNum int,TailNum varchar(255),Actual
ElapsedTime varchar(255),CRSElapsedTime int,AirTime varchar(255),ArrDe
lay varchar(255),DepDelay varchar(255),Origin varchar(255),Dest varcha
r(255),Distance varchar(255),Taxiln varchar(255),TaxiOut varchar(255),
Cancelled int, CancellationCode varchar(255),Diverted int,CarrierDelay
varchar(255),WeatherDelay varchar(255),NASDelay varchar(255),SecurityD
elay varchar(255),LateAircraftDelay varchar(255),IsArrDelayed varchar(
255),IsDepDelayed varchar(255));
copy airport_code(City,State,Country,IATA)
from '/Users/vyshnavigovindankutty/Desktop/Airport_data/airportcode.cs
v'
delimiter ','
csv header;

copy airlines(Year,Month,DayofMonth,DayofWeek,DepTime,CRSDepTime,ArrTi
me,CRSArrTime,UniqueCarrier ,FlightNum,TailNum,ActualElapsedTime,CRSEl
apsedTime,AirTime,ArrDelay,DepDelay,Origin,Dest,Distance,Taxiln,TaxiOu
t,Cancelled, CancellationCode,Diverted,CarrierDelay,WeatherDelay,NASDe
lay,SecurityDelay ,LateAircraftDelay,IsArrDelayed ,IsDepDelayed)
from '/Users/vyshnavigovindankutty/Desktop/Airport_data/airlines.csv'
delimiter ','
csv header;

select * from airlines limit 15;
```

```
 * postgresql://localhost/mydb
15 rows affected.
```

Out[13]:

year	month	dayofmonth	dayofweek	deptime	crsdeptime	arrtime	crsarrtime	uniquecarrier
2008	1	30	3	1917	1905	2021	2019	NW
2008	1	6	7	1610	1617	1752	1752	NW
2008	1	25	5	1326	1326	1611	1616	NW
2008	1	9	3	1126	1130	1251	1256	NW
2008	1	29	2	1130	1130	1257	1256	NW
2008	1	27	7	1006	925	1235	1049	NW
2008	1	21	1	1726	1720	2055	2130	NW
2008	1	20	7	1213	1140	1628	1623	NW
2008	1	8	2	1137	1140	1415	1445	NW
2008	1	29	2	1149	1140	1437	1445	NW
2008	1	29	2	1853	1139	2013	1230	NW
2008	1	20	7	1338	1335	1625	1626	NW
2008	1	22	2	1348	1307	1700	1608	NW
2008	1	31	4	1300	1307	1548	1608	NW
2008	1	27	7	926	930	1041	1051	NW

In [7]:

```
##1 Count of flights that departed late at origin and reached their de
stination early or on time

%%sql
select count(*) from airlines where isdepdelayed='YES' and isarrdelaye
d='NO';
```

```
 * postgresql://localhost/mydb
1 rows affected.
```

Out[7]:

count
54233

In [16]:
```
##2 Count of flights which departed late from origin by more than 60 minutes

%%sql
select count(*) from
airlines
where (deptime!='NA' and isdepdelayed='YES'and
cast(deptime as int)>crsdeptime+100) or
(cast(deptime as int)<crsdeptime
and ((2400+cast(right(concat('000',ltrim(cast(deptime as char(4)))),4)
as int))-crsdeptime)>100
 and isdepdelayed='YES'
 and deptime!='NA') ;
```

```
 * postgresql://localhost/mydb
1 rows affected.
```

Out[16]:

count
40104

In [ ]:
```

```

In [18]:
```
##3 Count of flights which departed early or on time but arrived late by at least 15 minutes

%%sql
select count(*) from airlines where (isdepdelayed='YES' and deptime!='
NA' and cast(deptime as int)>crsdeptime+15) or (cast(deptime as int)<c
rsdeptime and ((2400+cast(right(concat('000',ltrim(cast(deptime as cha
r(4)))),4)as int))-crsdeptime)>15 and isdepdelayed='YES' and deptime!=
'NA') ;
```

```
 * postgresql://localhost/mydb
1 rows affected.
```

Out[18]:

count
132792

In [13]:
```
##4 Count of flights departed from following major airports - ORD, DFW
, ATL, LAX, SFO

%%sql
select count(*) from airlines where origin in ('ORD','DFW','ATL','LAX'
,'SFO');
```

```
 * postgresql://localhost/mydb
1 rows affected.
```

Out[13]:

count
118212

In [17]:
```
##5 Add a column FlightDate by using Year, Month and DayOfMonth. Forma
t should be yyyyMMdd

%%sql
update airlines set Flightdate=concat(cast(year as char(4)),right(conc
at('00',ltrim(cast(month as char(2)))),2),right(concat('00',ltrim(cast
(dayofmonth as char(2)))),2));
```

```
 * postgresql://localhost/mydb
605659 rows affected.
```

Out[17]:
```
[]
```

In [18]:
```
##6 Count of flights that departed late between January 1 2008 to Janu
ary 9 2008 using FlightDate

%%sql
select count(*) from airlines where substring(Flightdate,1,4)='2008' a
nd substring(Flightdate,5,2)='01'and cast(substring(Flightdate,7,2) as
int)>1 and cast(substring(Flightdate,7,2) as int)<9;
```

```
 * postgresql://localhost/mydb
1 rows affected.
```

Out[18]:

count
140722

In [19]:
```
##7 Count of flights that departed late on Sundays

%%sql
select count(*) from airlines where dayofweek=7 and cast(deptime as in
t)>crsdeptime and deptime!='NA';
```

```
 * postgresql://localhost/mydb
1 rows affected.
```

Out[19]:

count
32561

In [10]:
```
%sql postgresql://localhost/mydb
```

In [10]:
```
##8 Get number of flights that had delayed departure and number of fli
ghts delayed in arrival for each day along with number of flights depa
rted for each day for January 2009
       #i. Output should contain 4 columns - FlightDate, FlightCount,
DepDelayedCount, ArrDelayedCount
       #ii.FlightDate should be of YYYY-MM-dd format.
       #iii. Data should be sorted in ascending order by flightDate

%%sql
select a1.Flightdate,count(*) as FlightCount,
(select count(*) from airlines a2 where a1.Flightdate=a2.Flightdate
     and cast(deptime as int)>crsdeptime
     and deptime!='NA')as DepDelayedCount,
(select count(*) from airlines a2 where a1.Flightdate=a2.Flightdate
    and cast(arrtime as int)>crsdeptime and arrtime!='NA') as ArrDelay
edCount
from airlines a1 where substring(Flightdate,1,4)='2008'
and substring(Flightdate,5,2)='01'
group by Flightdate
order by cast(substring(FlightDate,7,2)as int) desc;
```

```
 * postgresql://localhost/mydb
31 rows affected.
```

Out[10]:

flightdate	flightcount	depdelayedcount	arrdelayedcount
20080131	20260	9038	18292
20080130	19766	6644	18604
20080129	19485	6277	18010
20080128	20147	7551	18920
20080127	18903	8802	17389
20080126	16276	5732	15460
20080125	20313	8781	19089
20080124	20257	8087	19330
20080123	19769	7330	18621
20080122	19504	9111	18147
20080121	20133	10001	18850
20080120	18653	6678	17861
20080119	16249	6381	14910
20080118	20347	10010	19462
20080117	20273	9574	18579
20080116	19764	5039	18603
20080115	19503	5299	18728
20080114	20176	5913	18724
20080113	18946	6615	18023
20080112	16572	3886	15902
20080111	20349	7224	19195
20080110	20297	7005	19339
20080109	19820	5948	18930
20080108	19603	7461	18592
20080107	20341	8078	19022
20080106	19893	10466	18404
20080105	18066	8956	16791
20080104	20929	9304	19235
20080103	20937	11725	19643
20080102	20953	13203	19515
20080101	19175	10417	17717

In [11]:
```
##9 Get number of airports (IATA Codes) for each state in the US. Sort
the data in descending order by count

%%sql
select State,count(*) as c from airport_code where country='USA' and S
tate!='None' group by State order by c desc;
```

```
 * postgresql://localhost/mydb
51 rows affected.
```

Out[11]:

state	c
CA	29
TX	26
AK	25
FL	18
NY	18
MI	18
MT	14
PA	13
IL	12
CO	12
WY	10
NC	10
NM	9
HI	9
GA	9
NE	9
WI	9
WA	9
KS	9
AZ	8
AR	8
IA	8
WV	8
ND	8
MN	8

MA	8
MO	8
LA	7
MS	7
ME	7
SD	7
VA	7
OR	7
AL	6
OH	6
IN	6
SC	6
ID	6
TN	6
OK	5
KY	4
NH	3
MD	3
NV	3
VT	3
NJ	3
Hawaii	2
CT	2
UT	2
RI	1
DE	1

In [5]:
```
##10 Get number of flights departed from each US airport

%%sql
select airport_code.IATA,count(*) from airlines inner join airport_cod
e on airlines.origin=airport_code.IATA where country='USA' group by IA
TA,country ;
```

```
 * postgresql://localhost/mydb
270 rows affected.
```

Out[5]:

iata	count
ABE	413
ABI	240
ABQ	3447
ABY	102
ACT	209
ACV	301
ACY	31
ADQ	62
AEX	205
AGS	202
ALB	1224
ALO	27
AMA	628
ANC	1338
ASE	736
ATL	33897
ATW	423
AUS	4359
AVL	260
AVP	211
AZO	359
BDL	2729
BET	88
BFL	403
BGM	62
BGR	208
BHM	2047
BIL	359
BIS	221
BLI	30
BMI	445

BNA	4935
BOI	1520
BOS	9717
BPT	39
BQK	61
BRO	106
BRW	62
BTM	62
BTR	755
BTV	580
BUF	2051
BUR	2797
BWI	8883
BZN	449
CAE	919
CAK	744
CDV	62
CHA	376
CHO	74
CHS	1129
CIC	114
CID	781
CLD	229
CLE	5264
CLL	151
CLT	10752
CMH	3162
CMI	243
CMX	31
COD	93
COS	1445
CPR	301
CRP	361

CRW	258
CSG	116
CVG	8659
CWA	120
DAB	270
DAL	4717
DAY	1228
DBQ	120
DCA	7304
DEN	19477
DFW	23861
DHN	117
DLH	201
DRO	318
DSM	1349
DTW	14357
EGE	516
EKO	149
ELM	113
ELP	1818
ERI	109
EUG	552
EVV	498
EWN	61
EWR	12467
EYW	92
FAI	403
FAR	402
FAT	1266
FAY	171
FCA	216
FLG	148
FLL	6100

FLO	55
FNT	791
FSD	506
FSM	240
FWA	527
GEG	1373
GFK	120
GGG	93
GJT	372
GNV	171
GPT	729
GRB	675
GRK	361
GRR	1273
GSO	1083
GSP	995
GTF	217
GTR	81
GUC	93
HHH	67
HLN	133
HNL	5660
HOU	4810
HPN	879
HRL	409
HSV	901
IAD	6786
IAH	15531
ICT	1272
IDA	300
ILM	221
IND	3579

IPL	117
ISP	912
IYK	82
JAC	292
JAN	1132
JAX	2929
JFK	10023
JNU	309
KTN	186
LAN	369
LAS	15292
LAW	178
LAX	18945
LBB	717
LCH	96
LEX	749
LFT	452
LGA	10300
LGB	1244
LIH	1371
LIT	1365
LNK	270
LRD	169
LSE	162
LWS	52
LYH	56
MAF	589
MBS	229
MCI	5577
MCN	70
MCO	11070
MDT	662
MDW	7702

MEI	53
MEM	7046
MFE	353
MFR	474
MGM	323
MHT	1622
MIA	5545
MKE	2843
MLB	175
MLI	778
MLU	236
MOB	543
MOD	259
MOT	93
MQT	85
MRY	705
MSN	1102
MSO	298
MSP	11800
MSY	3453
MTJ	277
MYR	360
OAJ	82
OAK	5932
OGG	2079
OKC	2270
OMA	2277
OME	92
ONT	3558
ORD	29936
ORF	1503
OXR	113
PBI	2767

PDX	4898
PFN	272
PHF	481
PHL	8191
PHX	17695
PIA	500
PIH	154
PIT	3843
PLN	24
PNS	777
PSC	231
PSP	1189
PVD	1983
PWM	594
RAP	338
RDD	152
RDM	297
RDU	5312
RFD	58
RHI	31
RIC	1687
RNO	2322
ROA	292
ROC	1352
ROW	62
RST	356
RSW	2608
SAN	8043
SAT	3920
SAV	997
SBA	1109
SBN	436

SBP	550
SCE	62
SDF	2028
SEA	8543
SFO	11573
SGF	902
SGU	308
SHV	687
SIT	92
SJC	4976
SJT	125
SLC	12401
SMF	4774
SMX	149
SNA	4273
SPI	150
SPS	209
SRQ	677
STL	5329
SUN	222
SUX	38
SWF	450
SYR	1048
TEX	54
TLH	270
TOL	152
TPA	6748
TRI	173
TUL	2040
TUP	10
TUS	2545
TVC	306
TWF	249

TXK	124
TYR	155
TYS	1019
VLD	82
VPS	571
WRG	62
XNA	1199
YAK	62
YKM	33
YUM	380

In [13]:
```
##11 Get number of flights departed from each US state

%%sql
select State,count(*) from airlines
inner join airport_code on airlines.origin=airport_code.IATA
where country='USA'and State!='None'  group by State;
```

```
 * postgresql://localhost/mydb
49 rows affected.
```

Out[13]:

state	count
AK	2818
AL	3931
AR	2928
AZ	20768
CA	72853
CO	23288
CT	2729
FL	41042
GA	35527
HI	9110
IA	2315
ID	2497
IL	39812
IN	5040
KS	1272

KY	2777
LA	5884
MA	9717
MD	8883
ME	802
MI	17824
MN	12357
MO	11808
MS	2005
MT	1734
NC	17942
ND	836
NE	2547
NH	1622
NJ	12498
NM	3509
NV	17763
NY	28414
OH	19209
OK	4488
OR	6221
PA	13491
RI	1983
SC	3525
SD	844
TN	13549
TX	63930
UT	12709
VA	4093
VT	580
WA	10210
WI	5356

WV	258
WY	686

In [2]:
```
##12 Get the list of airports in the US from which flights have not departed

%%sql
select IATA from airport_code where country='USA' and IATA not in (select IATA from airport_code inner join airlines on airlines.origin=airport_code.IATA) order by IATA limit 10;
```

```
UsageError: Cell magic `%%sql` not found.
```

In [8]:
```
##13 Check if there are any origins in airlines data which do not have record in airport-codes

%%sql
select distinct origin from airlines where origin not in (select distinct origin from airlines inner join airport_code on airlines.origin=airport_code.IATA)order by origin limit 10;
```

```
 * postgresql://localhost/mydb
10 rows affected.
```

Out[8]:

origin
ADK
BQN
CDC
CEC
HDN
ITO
KOA
OTZ
PMD
PSE

In [3]:
```
##14 Get the total number of flights from the airports that do not con
tain entries in airport-codes

%%sql
select count(*) from airlines
left join airport_code on airlines.origin=airport_code.IATA
where airport_code.IATA is NULL;
```

```
 * postgresql://localhost/mydb
1 rows affected.
```

Out[3]:

count
5585

In [6]:
```
##15 Get the total number of flights per airport that do not contain e
ntries in airport-codes

%%sql
select airlines.origin, count(*) from airlines
left join airport_code on airlines.origin=airport_code.IATA
where airport_code.IATA is NULL group by airlines.origin;
```

```
 * postgresql://localhost/mydb
16 rows affected.
```

Out[6]:

origin	count
ADK	9
BQN	124
CDC	48
CEC	88
HDN	429
ITO	786
KOA	1316
OTZ	92
PMD	57
PSE	110
PSG	62
SCC	62
SJU	1997
SLE	54
STT	311
STX	40

In [ ]:

```
##16 Based on your analysis, give us 3 interesting facts you could gat
her from the datasets.

Atlanta is the busiest airport with maximum number of flights departin
g from it-33897
select origin,count(*) from airlines group by origin order by count(*)
desc limit 1;

O'Hare has maximum number of flights that are delayed- 13099
select origin,count(*) from airlines where cast(deptime as int)>crsdep
time and cast(arrtime as int)>crsarrtime and arrtime!='NA' and deptime
!='NA' group by origin order by count(*) desc limit 1;

O'Hare also has maximum number of flights that reach before time even
after it starts late- 3283
select origin,count(*) from airlines where cast(deptime as int)>crsdep
time and cast(arrtime as int)<crsarrtime and arrtime!='NA' and deptime
!='NA' group by origin order by count(*) desc limit 1;

Nearly 45% flights have a delayed departure
select count(*) as total_count,(select count(*) from airlines where is
depdelayed='YES') as del_count from airlines where deptime!='NA' and a
rrtime!='NA';

Only 9% of flights arrive on time inspite of departing late
select count(*) as total_count,(select count(*) from airlines where is
depdelayed='YES' and isarrdelayed='NO') as del_count from airlines whe
re deptime!='NA' and arrtime!='NA';

Thurday is the busiest day of the week with maximum number of flights
departing
select month,dayofweek,count(*) from airlines group by dayofweek,month
order by count(*) desc;
```

In [6]:
```
##17 Combine date,month and year as flight_date

%%sql
select concat(cast(year as char(4)),right(concat('00',ltrim(cast(month
as char(2)))),2),right(concat('00',ltrim(cast(dayofmonth as char(2))))
,2)) as Flight_date from airlines limit 10;
```

```
 * postgresql://localhost/mydb
10 rows affected.
```

Out[6]:

flight_date
20080121
20080121
20080122
20080122
20080123
20080123
20080123
20080123
20080124
20080124

In [15]:
```
##18 Find the lowest value from flightdate

%%sql
select flightdate from airlines order by cast(flightdate as int) asc l
imit 1;
```

```
 * postgresql://localhost/mydb
1 rows affected.
```

Out[15]:

flightdate
20080101

In [16]:
```
##18 Find the smallest and highest value from flightdate

%%sql
select min(flightdate),max(flightdate) from airlines;
```

```
 * postgresql://localhost/mydb
1 rows affected.
```

Out[16]:

min	max
20080101	20080131

In [ ]:
```

```