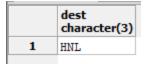# Week 2 Assignment

Week 2 assignment is due end of day on **Sunday February 8th**. You may work in a small group on the assignment. Please use the tables in the flights database. Your deliverable should include the **SQL queries that you write in support of your conclusions.** You should **submit your assignment in githu**b, and provide a link to your github "repo" in your assignment submission in Blackboard. We'll look together at some of the most interesting student solutions in our office hours on Wednesday February 11th.

One of the differences between the work of a data scientist and that of a software developer is that a data scientist explores data, asking and answering questions (where the questions asked may evolve as answers are found), whereas the endgame of the software developer is often to operationalize and improve business processes. Often work in data science leads to downstream work in software development. For each question, please provide the SQL statement or statements that you wrote to support your conclusions.

1. Which destination in the flights database is the furthest distance away?

select dest from flights
where distance = (select max(distance) as Max_Distance from flights)
group by dest;

	dest character(3)
1	HNL

2.
   a. What are the different numbers of engines in the planes table?

select engines from planes
group by engines;

	engines integer
1	2
2	1
3	4
4	3

   b. For each number of engines, which aircraft have the most number of seats?

select planes.engines, seats, tailnum

```
from planes
join(select engines, max(seats) as Max_Seats
        from planes
        group by engines) as v1 on
        planes.engines = v1.engines and
        planes.seats = v1.max_seats
group by planes.engines, seats, tailnum
order by engines, seats, tailnum
```

	engines integer	seats integer	tailnum character(6)
1	1	16	N567AA
2	2	400	N206UA
3	2	400	N228UA
4	2	400	N272AT
5	2	400	N57016
6	2	400	N77012
7	2	400	N777UA
8	2	400	N78003
9	2	400	N78013
10	2	400	N787UA
11	2	400	N862DA
12	2	400	N863DA
13	2	400	N865DA
14	3	379	N854NW
15	3	379	N856NW
16	4	450	N670US

3. What weather conditions are associated with New York City departure delays?

```
select
        flights.origin,
        dep_delay,
        temp ,
         dewp,
        humid ,
        wind_dir ,
        wind_speed ,
        wind_gust ,
        precip ,
        pressure ,
        visib
from weather  join flights on
        weather.origin = flights.origin and
```

```
        flights.year = weather.year and
        flights.month = weather.month and
        flights.day = weather.day
where flights.origin = 'JFK' and
        dep_delay > 0
```

	origin character(3)	dep_delay integer	temp double precision	dewp double precision	humid double precision	wind_dir integer	wind_speed double precision	wind_gust double precision	precip double precision	pressure double precision	visib double precision
1	JFK	181	48.92	39.02	68.51	60	4.60312	5.2971784336	0	1008.1	10
2	JFK	38	71.6	69.8	94.06	190	10.35702	11.9186514756	0		1
3	JFK	53	26.96	10.94	50.34	40	4.60312	5.2971784336	0	1023.9	10
4	JFK	146	32	8.06	36.03	280	26.46794	30.4587759932	0	1011.2	10
5	JFK	86	71.6	69.8	94.06	190	10.35702	11.9186514756	0		1
6	JFK	168	75.2	73.4	94.14	200	4.60312	5.2971784336	0		4
7	JFK	64	17.96	-0.94	42.69	290	29.92028	34.4316598184	0	1016.2	10
8	JFK	71	71.6	69.8	94.06	190	10.35702	11.9186514756	0		1
9	JFK	134	71.6	69.8	94.06	180	11.5078	13.242946084	0		0.5
10	JFK	19	17.96	-0.94	42.69	290	29.92028	34.4316598184	0	1016.2	10
11	JFK	20	32	8.06	36.03	280	26.46794	30.4587759932	0	1011.2	10
12	JFK	58	71.6	69.8	94.06	180	11.5078	13.242946084	0		0.5
13	JFK	3	48.92	39.02	68.51	60	4.60312	5.2971784336	0	1008.1	10
14	JFK	37	48.92	39.02	68.51	60	4.60312	5.2971784336	0	1008.1	10
15	JFK	28	75.2	73.4	94.14	200	4.60312	5.2971784336	0		4
16	JFK	34	71.6	69.8	94.06	190	10.35702	11.9186514756	0		1
17	JFK	46	75.2	73.4	94.14	200	4.60312	5.2971784336	0		4
18	JFK	31	48.92	39.02	68.51	40	4.60312	5.2971784336	0	1008.5	10
19	JFK	103	48.92	39.02	68.51	60	4.60312	5.2971784336	0	1008.1	10
20	JFK	25	32	8.06	36.03	280	26.46794	30.4587759932	0	1011.2	10
21	JFK	57	75.2	73.4	94.14	200	4.60312	5.2971784336	0		4
22	JFK	10	26.96	10.94	50.34	40	4.60312	5.2971784336	0	1023.9	10
23	JFK	52	75.2	73.4	94.14	200	4.60312	5.2971784336	0		4
24	JFK	25	48.92	39.02	68.51	40	4.60312	5.2971784336	0	1008.5	10
25	JFK	118	48.92	39.02	68.51	40	4.60312	5.2971784336	0	1008.5	10
26	JFK	35	26.96	10.94	50.34	40	4.60312	5.2971784336	0	1023.9	10
27	JFK	31	32	8.06	36.03	280	26.46794	30.4587759932	0	1011.2	10
28	JFK	32	71.6	69.8	94.06	180	11.5078	13.242946084	0		0.5
29	JFK	48	26.96	10.94	50.34	40	4.60312	5.2971784336	0	1023.9	10
30	JFK	79	32	8.06	36.03	280	26.46794	30.4587759932	0	1011.2	10
31	JFK	114	71.6	69.8	94.06	180	11.5078	13.242946084	0		0.5
32	JFK	43	71.6	69.8	94.06	190	10.35702	11.9186514756	0		1
33	JFK	40	17.96	-0.94	42.69	290	29.92028	34.4316598184	0	1016.2	10
34	JFK	33	32	8.06	36.03	280	26.46794	30.4587759932	0	1011.2	10
35	JFK	27	71.6	69.8	94.06	180	11.5078	13.242946084	0		0.5
36	JFK	62	75.2	73.4	94.14	200	4.60312	5.2971784336	0		4
37	JFK	24	32	8.06	36.03	280	26.46794	30.4587759932	0	1011.2	10

OK   Unix   Ln 82, Col 1, Ch 1676   463 chars   613 rows   156 ms

4. Are older planes more likely to be delayed?

-No! the avg max delay from JFK was from a more recent year (1983) plane.

```
select flights.origin, planes.year, avg(flights.dep_delay) as AvgDelay
from flights join planes on
        planes.tailnum = flights.tailnum
where   dep_delay > 0 and origin = 'JFK'
group by flights.origin,  planes.year
order by planes.year,avg(flights.dep_delay) desc
```

	origin character(3)	year integer	avgdelay numeric
1	JFK	1956	14.9090909090909091
2	JFK	1963	32.2500000000000000
3	JFK	1967	36.7500000000000000
4	JFK	1972	41.9090909090909091
5	JFK	1973	10.5000000000000000
6	JFK	1974	36.6538461538461538
7	JFK	1975	38.6363636363636364
8	JFK	1976	64.5000000000000000
9	JFK	1978	32.3846153846153846
10	JFK	1980	4.6000000000000000
11	JFK	1983	130.8000000000000000
12	JFK	1984	29.2222222222222222
13	JFK	1985	38.9090909090909091
14	JFK	1986	34.9630350194552529
15	JFK	1987	39.6270928462709285
16	JFK	1988	37.4747191011235955
17	JFK	1989	37.3787465940054496
18	JFK	1990	35.7994100294985251
19	JFK	1991	36.8676923076923077
20	JFK	1992	36.0213980028530670
21	JFK	1993	43.8315018315018315
22	JFK	1994	21.3378378378378378
23	JFK	1995	32.1470588235294118
24	JFK	1996	23.7172995780590717
25	JFK	1997	33.2426817752596789
26	JFK	1998	36.7804090419806243
27	JFK	1999	27.5262732417138238
28	JFK	2000	30.8580077050082554
29	JFK	2001	34.7797543413807709
30	JFK	2002	40.5318829707426857
31	JFK	2003	38.2396541074737492
32	JFK	2004	40.1432360742705570
33	JFK	2005	38.7506870828425599
34	JFK	2006	38.0144526445264453
35	JFK	2007	42.3947944539041596
36	JFK	2008	42.0954712362301102
37	JFK	2009	36.6512059369202226

5. Ask (and if possible answer) a question that also requires joining information from two or more tables in the flights database, and/or assumes that additional information can be collected in advance of answering your question.

What is my Departure Information?


```
select 'Departure Information' as DepInfo
     , flight
     ,airlines.carrier
     ,airlines.name
     , cast(flights.day as varchar(2))|| '/'||cast(flights.month as varchar(2)) ||'/' ||
cast(flights.year as varchar(4)) as DepartureDate
     , dep_time
     ,airports.faa
     ,airports.name
from flights
     inner join planes on flights.tailnum = planes.tailnum
     inner join airlines on flights.carrier = airlines.carrier
     inner join airports on flights.origin = airports.faa
where flight = 1318
     and flights.year = 2013 and flights.month = 12 and flights.day = 9
```

depinfo unknown	flight integer	carrier character(2)	name character varying	departuredate text	dep_time integer	faa character(3)	name character varying
Departure Information	1318	B6	JetBlue Airways	9/12/2013	1441	JFK	John F Kennedy Intl