**MSDA 607 - Week 2 Hands On Labs**

These ungraded hands-on-lab will help you build your skills working with combining information from multiple tables. Solutions to all problems are immediately available, but you’ll learn the material more durably if you attempt each problem before consulting its solution.

All of the queries use tables in the flights database. You may find the data descriptions here helpful:

<http://cran.r-project.org/web/packages/nycflights13/nycflights13.pdf>

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

select count(\*) from flights – returned 336776

select origin,dest from flights order by distance desc

Origin is JFK and Destination is NHL which is 4983 miles

1. What are the different numbers of engines in the planes table? For each number of engines, which aircraft have the most number of seats?

select distinct(engine) from planes

"Reciprocating"

"4 Cycle"

"Turbo-shaft"

"Turbo-fan"

"Turbo-jet"

"Turbo-prop"

select max(seats), engine from planes group by engine order by max(seats) desc

output is :

450;"Turbo-jet"

400;"Turbo-fan"

102;"Reciprocating"

14;"Turbo-shaft"

10;"Turbo-prop"

4;"4 Cycle"

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

As per IATA, the new York city has

New York, NY - All airports (NYC)  
New York, NY - Kennedy (JFK)  
New York, NY - La Guardia (LGA)

select \* from weather where origin='JFK' or origin='LGA' are the weather conditions

"JFK";2013;2;18;4;17.96;-0.94;42.69;290;29.92028;34.4316598184;0;1016.2;10

"JFK";2013;2;20;19;32;8.06;36.03;280;26.46794;30.4587759932;0;1011.2;10

"JFK";2013;7;2;11;71.6;69.8;94.06;180;11.5078;13.242946084;0;;0.5

"JFK";2013;7;2;13;71.6;69.8;94.06;190;10.35702;11.9186514756;0;;1

"JFK";2013;7;31;6;71.06;55.04;56.93;320;8.05546;9.2700622588;0;1020.4;10

"JFK";2013;9;2;20;75.2;73.4;94.14;200;4.60312;5.2971784336;0;;4

"JFK";2013;10;23;10;48.92;39.02;68.51;60;4.60312;5.2971784336;0;1008.1;10

"JFK";2013;10;23;11;48.92;39.02;68.51;40;4.60312;5.2971784336;0;1008.5;10

"JFK";2013;12;17;5;26.96;10.94;50.34;40;4.60312;5.2971784336;0;1023.9;10

"LGA";2013;8;22;22;75.92;66.92;73.68;210;8.05546;9.2700622588;0;1011.9;10

1. Are older planes more likely to be delayed?

Not necessarily

select \* from planes order by year asc give output as below

"N381AA";1956;"Fixed wing multi engine";"DOUGLAS";"DC-7BF";4;102;232;"Reciprocating"

"N201AA";1959;"Fixed wing single engine";"CESSNA";"150";1;2;90;"Reciprocating"

"N567AA";1959;"Fixed wing single engine";"DEHAVILLAND";"OTTER DHC-3";1;16;95;"Reciprocating"

"N378AA";1963;"Fixed wing single engine";"CESSNA";"172E";1;4;105;"Reciprocating"

"N575AA";1963;"Fixed wing single engine";"CESSNA";"210-5(205)";1;6;;"Reciprocating"

"N14629";1965;"Fixed wing multi engine";"BOEING";"737-524";2;149;;"Turbo-fan"

"N615AA";1967;"Fixed wing multi engine";"BEECH";"65-A90";2;9;202;"Turbo-prop"

"N425AA";1968;"Fixed wing single engine";"PIPER";"PA-28-180";1;4;107;"Reciprocating"

"N383AA";1972;"Fixed wing multi engine";"BEECH";"E-90";2;10;;"Turbo-prop"

"N364AA";1973;"Fixed wing multi engine";"CESSNA";"310Q";2;6;167;"Reciprocating"

"N840MQ";1974;"Fixed wing multi engine";"CANADAIR LTD";"CF-5D";4;2;;"Turbo-jet"

"N508AA";1975;"Rotorcraft";"BELL";"206B";1;5;112;"Turbo-shaft"

The oldest flight is 1956

select t1.dep\_delay,t1.tailnum, t2.year from flights t1, planes t2 where t1.tailnum=t2.tailnum order by t1.dep\_delay

gives output as

-43;"N592JB";2004

-33;"N612DL";1985

-32;"N825AS";1997

-30;"N934DL";1989

-27;"N208FR";2011

-26;"N711MQ";1976

-25;"N13994";2000

-24;"N531JB";

-24;"N929XJ";2008

-24;"N8747B";2003

-23;"N29906";2001

-23;"N800AY";2004

-23;"N190JB";2005

-23;"N279JB";2007

43 hrs of delay comes from 2004 year flight. Not the old flight.

1. 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.

This is the concept of relational database. The concept is to normalize it to the maximum possible level. So when query is written, joins are must to get appropriate values .otherwise, the output will be in some numerical or primary key format which cannot be understood by someone who does not know the query and technicalities behind it.

Some of the additional information that can be collected is Primary Key, foreign key etc.