Homework 2

Group 3
TA: 溫郁婷

E-maill: ytwen.cs03g@nctu.edu.tw

Introduction

In this homework, you have to load a data set into your database and use SQL to answer the following questions. Here is the link for downloading the data set.

Due Day:

4/10 23:59:59

Dataset:

Airline on-time performance dataset

http://stat-computing.org/dataexpo/2009/

Descriptions:

The data consists of flight arrival and departure details for all commercial flights within the USA, from October 1987 to April 2008.

1. Download the data from year **2006** to **2008** (totally three years)

http://stat-computing.org/dataexpo/2009/the-data.html

2. Download **airports.csv** and **plane-data.csv**

http://stat-computing.org/dataexpo/2009/supplemental-data.html

Tables:

on-time performance data

Year	2006-2008
Month	1-12
DayofMonth	1-31
DayOfWeek	1 (Monday) - 7 (Sunday)
DepTime	actual departure time (local, hhmm)
CRSDepTime	scheduled departure time (local, hhmm)
ArrTime	actual arrival time (local, hhmm)
CRSArrTime	scheduled arrival time (local, hhmm)
UniqueCarrier	Unique Carrier Code. When the same code has been used by multiple carriers, a numeric suffix is used for earlier users, for example, PA, PA(1), PA(2). Use this field for analysis across a range of years.

FlightNum	flight number
TailNum	plane tail number
ActualElapsedTime	in minutes
CRSElapsedTime	in minutes
AirTime	in minutes
ArrDelay	arrival delay, in minutes
DepDelay	departure delay, in minutes
Origin	origin **IATA airport code***
Dest	destination **IATA airport code***
Distance	in miles
TaxiIn	taxi in time, in minutes
TaxiOut	taxi out time in minutes
Cancelled	was the flight cancelled?
CancellationCode	reason for cancellation (A = carrier, B = weather, C = NAS, D = security)
Diverted	1 = yes, 0 = no
CarrierDelay	in minutes
WeatherDelay	in minutes
NASDelay	in minutes
SecurityDelay	in minutes
LateAircraftDelay	in minutes

*check airports data

airports data

iata	the international airport abbreviation code
airport	name of the airport
city and country	in which airport is located
lat and long	the latitude and longitude of the airport

plane data

tailnum	plane tail number
type, ...	plane information

Questions:

【Every task needs to contain (1) **the SQL**, (2) **time** and (3) **the answer**】

1. Please list the number of different routes (origin → destination) and contain "**DISTINCT**" syntax in usage of SQL
   請列出所有不同飛行航線 (起點→終點) 的數量，請包含 DISTINCT 語法

2. Please list the number of flights which destination is TPA and the actual flight time is between 1 to 3 hours and contain "**BETWEEN**" syntax in usage of SQL
   請列出目的地為 TPA 且實際飛行時間介於 1~5 小時的航班數量，請包含 BETWEEN 語法

3. Please list both the aircraft type name and its total amount which the aircraft type name beginning with "767" and contain "**LIKE**" syntax in usage of SQL
   請列出機型名稱開頭為 767 的機型與其數量，請包含 LIKE 語法

4. Please list the average speed (use the actual arrival time) of all aircraft type and contain "**AVG**" in usage of SQL
   請列出所有機型的平均速率(用實際時間算)，請包含 AVG 語法

5. Please list the top 50 time zone count and route (origin → destination) by the count of routes which across time zone(assume one time zone is formed by 15 degree longitude and not related to country) in decrease order and contain "**ORDER BY**" in usage of SQL
   請根據跨越的時區(假設每 15 度為 1 個時區，不看國家)數量由大到小列出前五十名航線(起點→終點)與數量，請包含 ORDER BY 語法 (若數量一樣根據航線(起點→終點)名稱以字典排序)

6. Please list the aircraft type which never appear after 2009 (appeared before) and contain "**NOT IN**" syntax in usage of SQL
   請列出在 2009 年再也沒出現的機型 (在之前出現過) 請包含 NOT IN 語法

7. Please list the manufacturer which airplanes' average delay time is more than half an hour
   找出飛機平均延誤時間大於半小時的製造廠商

8. Please list the average delay time of week (7 day) and time (24 hour), and find out the day of week / time slot that has minimum delay

請列出一週 7 天以及一天 24 小時中的飛機平均延誤時間，什麼時間搭飛機遇到的 Delay 會最少?

9. Do older planes suffer more delays? (Answer by your own view)
比較老的飛機 Delay 時間會比較久嗎? (此題為開放式答案，請利用 SQL 找出的結果闡述你的觀點)

10. How well does weather or season predict plane delays? (Answer by your own view)
天氣或是季節會影響 Delay 時間嗎? (此題為開放式答案，請利用 SQL 找出的結果闡述你的觀點)

Bonus.
Find something that is interesting, meaningful or useful using SQL query, and explain your idea.
請自由發想 SQL，找出有趣的資訊並解釋它的含意

If you have any questions, feel free to contact me.