Amadeus
Data Mining Exercise – solved in Python

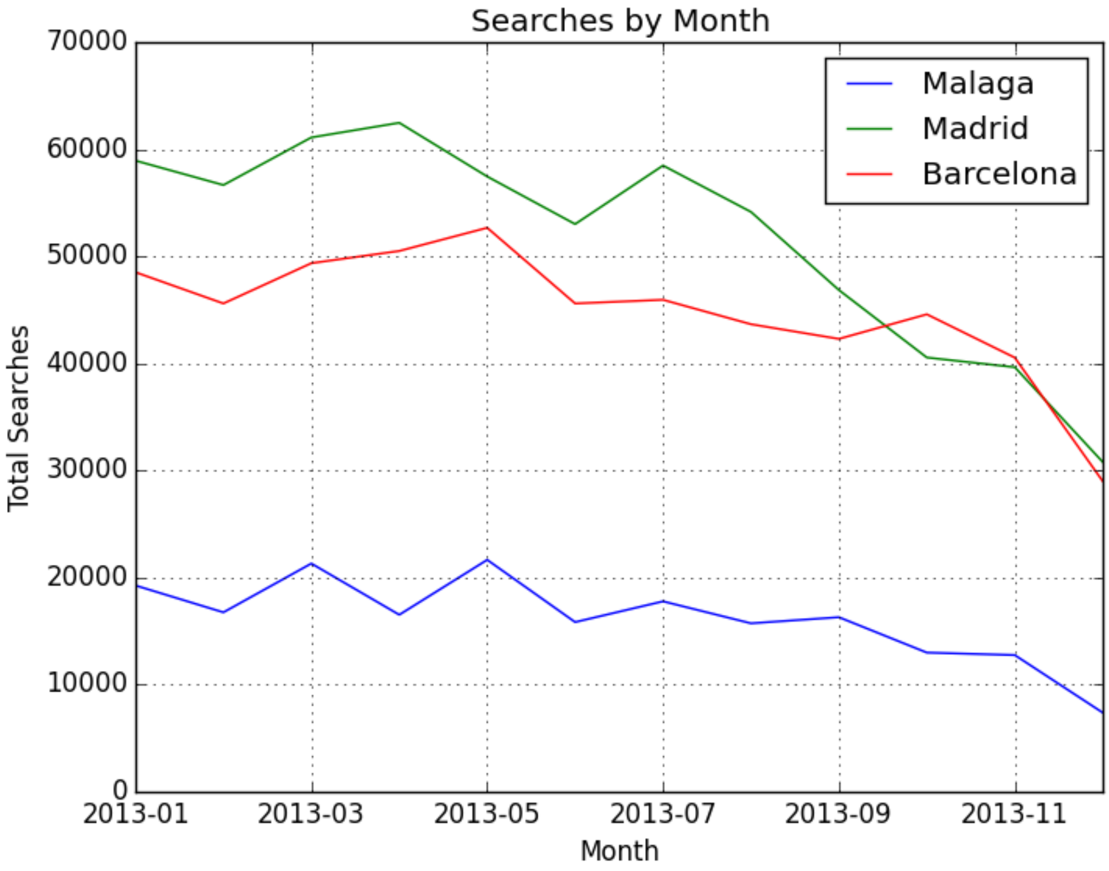
First exercise: Give number of lines in bookings & searches table

```
The number of lines in bookings is 10000010
The number of lines in searches is 20390198
```

Second exercise: Top 10 arrival airports in 2013 by number of passengers

```
#   Results:
#
#   arr_port  pax
#   LHR       88809
#   MCO       70930
#   LAX       70530
#   LAS       69630
#   JFK       66270
#   CDG       64490
#   BKK       59460
#   MIA       58150
#   SFO       58000
#   DXB       55590
```

Third exercise: Plot the monthly number of searches for flights arriving at Málaga, Madrid or Barcelona.



Bonus exercise 1

Please see table searches_booked.csv, it's only run on a small sample of the data since I didn't have the time to run it through the whole data set.

Bonus exercise 2

Please refer to the bonus2web folder.

Matching criteria:
A search is booked if:
1. All its segments has a corresponding booking record with matched Dep & Arr port, boarding date
2. The search date is on the same day of the booking date
3. If a booking is matched, it will not be reused for further matching

Assumptions:
1. Identical rows in searches are from same end-user
2. Identical rows in bookings are from different end-users