1. Go to <https://www.expedia.co.uk> and look for a Flight + Hotel (Package) trip to Las Vegas (notice you can save up to 15% by booking a flight + hotel together!), whatever the dates.
2. You land on the “Start by choosing your hotel” page (the Hotel Search Results page). This page is listing all the available hotels you can book, along with the total package price per person (i.e. final package price also including the flight part). Note the strikethrough price: this corresponds to the sum of the hotel and flight price if you were booking them separately.
3. Download “vegas\_HSR\_test.csv” from <https://github.com/ndelagex/test>.   
   Open the csv using a jupyter notebook or any Python/R editor.   
     
   This dataset is logging all the search results you were seeing previously, for users that have made a Package search to Vegas on the 6th March. This includes:

* departurelocationcode and arrivallocationcode: departure and arrival airport code
* start and end dates: trip start and end date
* adultcount and childcount: # of adults and child traveling
* result\_position: position of the hotel in the listing
* pricing info: total package price, splitting between the hotel and the flight part, plus the savings (difference between actual and strikethrough price
* transaction\_id: tells you if the user eventually made a transaction for this hotel

Do not hesitate if you have any questions on the dataset!

1. At first sight, can you think of some metrics/graphs/stats we can use to monitor the destination, if we had a year of data for e.g.? (just list them somewhere)
2. Data handling questions:
   * Delete duplicates, NA values, round numerical values.
   * Compute basic statistics about numerical fields (average, median, min…)
   * Compute the distribution (boxplot) of hotel price per Room Night?   
     (Room Night = Roomcount \* Length of Stay).
   * Histogram of number of searches by Savings %
3. Regression model:
   * Which variables do you think have an impact on conversion rate?
   * Can you think of a basic model that would use the savings to predict the conversion rate?