**Problem Description:**

An advertising company sells a service of buying keywords in search engines on behalf of their customers. They’re trying to optimise their keyword and funds allocation. The first towards the optimal solution is to predict performance by keyword and fund.

**Goal:**

Predicting for any keyword (not necessarily the ones in the dataset file), CPC, and market (US/UK) the traffic a website would receive (I.e., the clicks).

**Task Evaluation:**

The evaluation of the task will use an input dataset of new keywords and CPC for each market (US/UK) at the date of 14/2/2013. The model's results will be compared to real results for that day.

**Questions:**

1. How did you manipulate the data, and why? Illustrate your answer with plots.
2. How did you perform NLP, if any?
3. How did you model the problem, and why?
4. How did you evaluate your model? What were the results of the evaluation?
5. If you had extra time, what would you do next?

**Table Structure:**

|  |  |
| --- | --- |
| ***Column*** | ***Description*** |
| Date | The date the data was collected from yyyymmdd |
| Market | The market (US/UK) |
| Keyword | The keyword |
| Average.Position | The average position a keyword had in the search engine |
| CPC | The amount of money agreed to be paid per click on a keyword |
| Clicks | The number of clicks a keyword had |
| CTR | The percentage of users who saw the ad and clicked it (clicks / impressions) |
| Impressions | The number of users who saw the ad |
| Cost | The total cost of a keyword in a day (clicks \* CPC) |