**Candidate Name:**

**Date:**

**General Overview**

You are required to complete 2 challenges out of the given 3 challenges below. Please select Challenge 1 or Challenge 2; completing Challenge 3 is required. Please Return the results to us within the next three days.

**Additional Notes for the Candidate**

* Please document your work and provide us the results along with the commented code.
* Your work will not be used for commercial purposes outside of the interview process.
* We are interested not just in the results, but also in your thought process and approach.
* Use any format you like that allows to show both the results and the process, including Jupyter Notebooks, PowerPoint, Word, Excel, etc.
* All data and included scenarios are fictitious.

**Our Expectations from each Challenge Solution**

* Even before starting answering the business questions, we expect to see how you dive deep into the data, analyze it, validate, clean if needed, and visualize when appropriate.
* We expect you not just use the right tools to solve the challenge, but also provide an explanation why these tools are appropriate (and what are other alternatives). For example, in which case it is sufficient to rely on visual interpretation, and in which case a more accurate statistical tool is required.
* We expect you not just arrive at results, but also provide an analysis of the quality and accuracy of the result, as well as provide a quantitative confidence ratio, where appropriate.
* Your code (Python / R / ...) should be well structured, easily readable, following good coding and style practices, and well documented. Same goes about the structure of your notebook or any other document you are attaching.
* You should use standard existing tools as much as possible, avoid reinventing the wheel.
* Your answers to business questions should be well formulated, clear, convincing, and backed up by evidence.
* When you can’t arrive at an answer, you should be very clear and transparent about it. Similarly, if your confidence in an answer is low, you should clearly communicate this.

# **Challenge 1:** Price Optimization

A pricing experiment was conducted by Amazon.com.au with the main aim of increasing revenue. In the experiment, users were grouped into two groups: A (66% of the user base) and B (33% of user base). Group A users were offered a cheaper price while group B users received a higher price.

## Business Questions

* What is your recommendation to the company in terms of setting an optimum price for their product?
* What are your main findings looking at the data? What is your overall view into user behavior, especially focusing on actionable insights that might increase conversion rate?
* Can you optimize the number of days that the test is run? After how many days would you have stopped the test? Why?

## Data Details

Table 1: test\_results

**user\_id**: the Id of the user. Can be joined to user\_id in user\_table

**timestamp**: the date and time when the user first hit the webpage. It is in user local time

**source**: marketing channel that led to the user coming to the site. It can be:

* ads-["google", "facebook", "bing", "yahoo", "other"]. That is, user coming from google ads, yahoo ads, etc.
* seo - ["google", "facebook", "bing", "yahoo", "other"]. That is, user coming from google search, yahoo, facebook, etc.
* friend\_referral : user coming from a referral link of another user
* direct\_traffic: user coming by directly typing the address of the site on the browser

**device**: Can be mobile or web

**operative\_system**: user operative system.

**test**: (i.e. 1 -> higher price) and (0 -> old lower price)

**price**: the price the user sees.

**converted**: 1 -> bought the product - 0 -> left the site without buying it.

Table 2: user\_table

**Columns:**

**user\_id**: the Id of the user. Can be joined to user\_id in test\_results table

**city**: the city where the user is located. Comes from the user ip address

**country**: in which country the city is located

**lat**: city latitude - should match user city

**long**: city longitude - should match user city

# **Challenge 2:** Recommendation Engine for Prime Video

Today, videos shown to new users on Prime Video’s home page are manually chosen. You need to implement a recommendation engine to increase the conversation rate.

## Business Questions

* Classify each video into the following buckets:
  + "Hot" - means trending up. These videos are candidates to be shown.
  + "Stable and Popular" - video view counts are flat, but very high. These videos are candidates to be shown too.
  + "Everything else" - these videos won't be shown.
* What are the main characteristics of the "hot videos"?
* After having identified the characteristics of the hot videos, how would you use this information from a product standpoint?

## Data Details

Table 1: video\_count

**Columns:**

**video\_id**: unique video id

**count**: total count of views for the specific video\_id on the specific date

**date**: the date that these views occurred

Table 2: video\_features

**Columns:**

**video\_id**: video id, unique by video and joinable to the video id in the other table

**video\_length**: length of the video in seconds

**video\_language**: language of the video, as selected by the user when they uploaded the video

**video\_upload\_date**: when the video was uploaded

**video\_quality**: quality of the video. It can be [ 240p, 360p, 480p, 720p, 1080p]

# **Challenge 3:** Understanding user and product interaction

Amazon’s employee travel provider has asked you to review how users interact with their online travel website.

They store their data in JSON files. Each row in these files lists all the different cities that have been searched for by a user within the same session (as well as some other info about the user).

## Business Questions

* There was a bug in the code and one country didn't get logged. Can you guess which country? How?
* For each city, find the most likely city to be also searched for within the same session.
* Travel sites are browsed by two kinds of users. Users who are actually planning a trip and users who just dream about a vacation. The first group obviously has a much higher purchasing intent. Users planning a trip often search for cities close to each other, while users who search for cities far away from each other are often just dreaming about a vacation (or a great work trip!). Based on this idea, come up with an algorithm that clusters sessions into two groups: high intent and low intent.

## Data Details

Table 1: city\_search

**Columns:**

* **session\_id**: session id.
* **unix\_timestamp**: unix timestamp of session start time
* **cities**: the unique cities which were searched within the same session
* **user**:
  + **user\_id**: the id of the user
  + **joining\_date**: when the user created the account
  + **country**: where the user is based