

# News Labs – Junior Software Engineer – Coding Exercise

## Background

BBC News publishes hundreds articles on average each day. We like to make data-informed decisions about publishing and optimising our content, so we can increase the reach and quality of our journalism.

To help us better understand how our readers are consuming our content, we collect a bunch of data on each article that is published to our website. For measuring reach, we look at **Page Views**, which tells us how many people a particular story reached and indicates how well the story was promoted. This data is collected approximately every 15 minutes and includes information on the number of page views broken down by different countries/regions, different types of social media used, different device types etc.

## Dataset

In this exercise, we provide you with a subset of this data for a single article that was published on 15th October 2018 at 13:05 GMT, and covering the remainder of that same day.

Files:

- `article-Regions.csv`
- `article-Device.csv`

## Summary

For this exercise, we'd like to ask you to write some code to read in the data and generate outputs for *two* out of the five listed below.

You should generate a separate text file for each of the *two* parts that you select (name the file corresponding to the selected part, e.g., `a.txt`, `b.txt` etc...).

Use a language and an approach with which you feel most comfortable (e.g., Python/Javascript/other languages, and language tools, constructs and libraries etc).

We suggest that you should aim to spend no more than 2 or 3 hours on the exercise.

Please submit, within 1 week of receiving the exercise, a link to a github repository containing:

- Your code
- The text files containing your outputs
- A README file with instructions on how to install/run your code

Please feel free to include a text file containing a brief outline (no more than a paragraph or two) of any choices you have made.

## Exercises

- a) Generate a text file containing a table listing the mean number of **Total Page Views** during each 1-hour period (i.e., 13:00-13:59, 14:00-14:59... etc)
- b) Generate a text file containing a table listing the mean number of **Page Views** for each **Region** during each 1-hour period (i.e., 13:00-13:59, 14:00-14:59... etc)
- c) For each **Region**, output the 1-hour time period where the the mean number of **Page Views** is highest
- d) Output the percentage of people visiting the page via a **mobile phone or a tablet device**
- e) Generate a text file containing a table listing all the 1-hour time periods where the mean number of **Page Views** via **mobile phones** is higher than the overall mean (throughout the day) number of Page Views via mobile phone

During the interview, we'd like to spend a few minutes discussing your solution and the approach you've taken.

If you have any questions or need any clarifications relating to this exercise, please email [coding-exercises@bbcnewslabs.co.uk](mailto:coding-exercises@bbcnewslabs.co.uk)