**Challenge Goal:**

1. Extract Data:
   1. Use ***extract.py*** as a starting code to extract data from google app store. You have to modify the code as needed. For detail documentation of the library use this github link: <https://github.com/JoMingyu/google-play-scraper>

You have to submit the ***submission.py*** code with clean commented code.

* 1. Generate a csv file from the extracted data. A sample data **jobswipe.csv** is provided. You have to generate ***submission\_competiror\_name.csv and have*** to submit this file.
  2. Your submitted csv file should have the exact same structure as the sample data file. Column name should be the same.
  3. Your submitted data should have N number of rows. Where N = number of google play store reviews for a particular app. For example for this following app, <https://play.google.com/store/apps/details?id=com.adswipe.jobswipe&hl=en_US&gl=US> total number of reviews is 2.13K. So the expectation is that your extracted data set should have **2.13K** rows.
  4. Submitted data data file name has a string concatenation part that depends on the app name. For example in the app we mentioned in 1(d), the app name is jobswipe. So the expected file name is submission\_jobswipe.csv .
  5. In order to extract the app name please parse the URL to get the app id. And take the last word in the app id. For example, id=com.adswipe.**jobswipe**
  6. Now you will repeat steps a to f for these four more competitors.

i) <https://play.google.com/store/apps/details?id=com.jobget&hl=en_US&gl=US>

ii) <https://play.google.com/store/apps/details?id=com.harri.candidate&hl=en_US&gl=US>

iii) <https://play.google.com/store/apps/details?id=com.poachedjobs.mobile&hl=en_US&gl=US>

iv) <https://play.google.com/store/apps/details?id=com.snagajob.jobseeker&hl=en_US&gl=US>

1. **Generate Insights:**
2. Merge all 5 extracts into a single data frame with an added column with the app name (explained in 1.f). Submit the file as final\_submission.csv
3. Categorize the reviews with as +ve and -ve and neutral sentiment
4. For score/rating 4 & 5 extract the top 5 **keywords** to understand why they rated high.
5. Modify the sentiment analysis to classify reviews with +ve and -ve experiences.
6. Bonus point: Understand what are the downsides of each app that causes negative experience of the users.

**3. Deliverable:**

1. Six data extracts in csv format
2. One word, pdf document explaining insights and findings.
3. Python codes for extracting and generating insights.
4. Zip all these files together and reply to [info@ergool.com](mailto:info@ergool.com) with a subject line “Ergool App Store Insight Challenge\_ your last name”
5. Deadline: You have to submit by **March 5th 11:59 PM Bangladesh Time**.

Thank you for taking the challenge!

Good luck!

Team Ergool