**Task**

At HackerRank, we love helping our users find interesting challenges so we're always impoving our challenge recommendation system. In this challenge, you must predict a set of challenges that a HackerRank user (or *hacker*) is likely to solve. We'll provide you with a list of challenges along with information on where they are located on HackerRank. You'll also have a list of submissions made by a subset of hackers on these challenges. You must attempt to predict challenge recommendations for each hacker to solve as their next challenge.

Your recommendation must satisfy the following conditions:

* The recommended *challenge\_id* should be present in the dataset provided to you.
* The *hacker\_id* should be present in the dataset provided to you.
* The hacker should not have already successfully solved the recommended challenge.
* The challenge should be available in the contest c8ff662c97d345d2.

**Data Set**

Download the zip file (*MD5* checksum is d4305bffee2b3dd0350a8df94e2f38b4) [here](https://www.hackerrank.com/external_redirect?to=https://s3.amazonaws.com/hr-testcases-us-east-1/24058/assets/hackerrank-challenge-recommendation-dataset.zip). The dataset contains two files named *challenges.csv* and *submissions.csv*. The files are organized as follows:

* **challenges.csv**: This dataset contains information about the challenges. The file is organized into the following fields:
  + *challenge\_id* is the challenge's unique ID.
  + *contest\_id* is the unique ID for the contest where the challenge was featured.
  + *domain* is the name of the domain where the challenge is located.
  + *subdomain* is the name of the chapter within the domain where this challenge lies.
  + *difficulty* is a decimal value between and (if this value is greater than , it cannot be trusted) denoting the difficulty of the challenge. The higher the value, the higher the challenge's difficulty.
  + *solved\_submission\_count* is the number of unique hackers who have successfully solved this challenge.
  + *total\_submission\_count* is the number of the unique hackers who have made at least one submission for this challenge.
* **submissions.csv**: This dataset contains information about the submissions made by hackers. The file is organized into the following fields:
  + *hacker\_id* is the unique ID of the hacker who made the submission.
  + *contest\_id* is the unique ID of the contest where the hacker made the submission.
  + *challenge\_id* is the unique ID of the challenge where the hacker made the submission.
  + *language* is the name of programming language used for the submission.
  + *solved* is a boolean value (i.e., or ) denoting whether or not the hacker completely solved the challenge by passing all the test cases. If the value is , it means the hacker successfully solved the challenge; otherwise, the value is (indicating that the submitted solution did not satisfy the challenge's requirements).
  + *created\_at* is the time at which the hacker made the submission.

Note that all the IDs in the dataset are -digit hexadecimal numbers.

**Submission Details**

You are required to upload the following three files:

* The output file, *recommendation.csv* (maximum allowed size is *10MB*), will contain your predicted challenges for the hackers. Each row should contain your challenge recommendations for one hacker. The first value of every row should be the *hacker\_id* followed by at most recommended *challenge\_ids* for that hacker.

A valid output file has the following format:

7abc5cd136d75672,9580a525da7618e9,659f2ceb67eca1ef,feb666a0b4f1d685,6ed2a6b18d1ac4e4, 8d9ec6e3f0500705,286001f675d5dc00,c0af13494532539f,ad1cc7894c991bbd,e70a3d1b641f8e2e,0251520ace81bc57

478761f505a4f75b,1e5b724075cd3d1e,8d9ec6e3f0500705,a8fa4ef7a8bc43dc,5e85f62bcbb42ea2, 6ed2a6b18d1ac4e4,970eca328814a522,425f241df843cd89,9580a525da7618e9,76fba8ff25f765b6,286001f675d5dc00

002d8a73bc9e7734,0ea21bb9feaf46c8,55f8a66f057a7e76,c9713aed1d68c5b5,34386eca594b7bc2,b49bb8ec9c2b8481,26c95b75d8237f39,76fba8ff25f765b6,e05e63bcccbafedb,feb666a0b4f1d685,659f2ceb67eca1ef

98d363d73a97efdd,6289494731c566da,a177f57b92fd38d6,59e33283b96c6407,1d7afa1b244e70c9, feaaa68f9218a16a,08620fa095caf8f5,ff883052c3a1d18b,72ccb69fd2fd9480,d72597ae8480885a,e6f5e45bb2f4f804

Note that the output file *should not* have any header row.

* A *PDF* file (maximum allowed size is *2MB*) providing the findings and justification on the following topics:
  + Write a few lines about training dataset quality and any errors found in the training dataset.
  + Explain the data preprocessing steps.
  + Explain and justify the model you've chosen for the prediction.
* The source code of your approach for this task. Upload a *zip* file (maximum allowed size is *5MB*) with all relevant files to reproduce your results. The submitted file must have a *README* file with a detailed description about how to run the model to predict the challenges recommendations and generate the *recommendation.csv* file. Do not forget to include links to any external libraries or packages you use for the generation of your model.

There is no limit on execution time, but the code should generate the output file: *recommendation.csv*.