Programming Problems
DataWeave Software Pvt. Ltd.

Ground Rules

1. Please use Python only for coding.
2. Please make reasonable assumptions when in doubt.
3. Avoid using databases. If you need to persist something, you may do so on files.
4. If you are using any readily available code, do mention the source for it.
5. We expect complete solutions to any problem that you solve.
6. Also mention in each of the problems, if you know of any alternate ways of solving this problem and could not code it due to time constraints.
7. You are free to use any publicly available data or open source models for your testing.
8. And most importantly, have fun!

Problem 1 : VoterId Reader

Mr. Khatri works for Indian government in election department. His department maintains records for all the Indian citizens' Voter Ids in pdf format. Each pdf holds thousands of Voters' details in a tabular format based on their state and the region. A major part of Mr. Khatri's job is to find out the VoterId for specific citizen given his area code and serial number.

Mr. Khatri's workflow for finding voterID is :

- First, from area code he searches pdf document of that region.
- Then, in that document, he looks for the specified serial number and find out a specific card and read Voter ID from it. (Refer Figure-1.1)

To reduce his workload Mr. Khatri wants to automate this process. He met Mr. Bazinga who has helped him to convert these pdf into images. Now, He is looking for a solution where given a serial number one can extract voterId in **machine-readable forma**t.

For example, in the image shown in Figure-1.1, if input serial number is 8 then programme should return "UOG0733866".

Please help Mr. Khatri in automating his work. **Do remember while framing your solution, He needs to deal with datasets for all Indian citizens**.

(Enclosed is required images dataset)

Photo Electoral Roll - 2015
71 - Gannavaram Assembly Constituency Part No. 1

1.Nunna,Narasimhaswamy Temple

1 UOG0238535 Elector's Name: MASTHAN MOHAMMED Father's Name: GHOUSE House No: 1-1/2 Age: 22 Sex: Male	2 UOG0476607 Elector's Name: YATHISHREDDY ARUMALLA Father's Name: KOTIREDDY ARUMALLA House No: 1-3 Age: 24 Sex: Male	3 UOG0476623 Elector's Name: SRIKANTHREDDY BHIMAVARAPU Father's Name: SUBBAREDDY BHIMAVARAPU House No: 1-7 Age: 30 Sex: Male
4 UOG0673997 Elector's Name: RAJANI BHIMAVARAPU Husband's Name: SIVA RAMA KRISHNA REDDY BHIMAVARAPU House No: 1-7 Age: 30 Sex: Female	5 UOG0869018 Elector's Name: YASHWANTH REDDY YARKAREDDY Father's Name: SRINIVASA REDDY YARKAREDDY House No: 1-8 Age: 21 Sex: Male	6 UOG0476631 Elector's Name: MANOHARREDDY KURREY Father's Name: VENKATAREDDY KURREY House No: 1-9 Age: 20 Sex: Male
7 UOG0816373 Elector's Name: SIVA KUMAR REDDY KURRE Father's Name: VENKATA REDDY KURRE House No: 1-9 Age: 18 Sex: Male	8 UOG0733866 Elector's Name: PRAVEENA BONTHU Husband's Name: DEEPENDRA REDDY BONTHU House No: 1-11 Age: 28 Sex: Female	9 UOG0525586 Elector's Name: JANSI LAKSHMI SILAM Husband's Name: PRAKASH REDDY SILAM House No: 1-13 Age: 27 Sex: Female
10 UOG0880766 Elector's Name: VANI ALLA Husband's Name: VENKATA RAMI REDDY ALLA House No: 1-15 Age: 46 Sex: Female	11 UOG0880774 Elector's Name: VENKATA RAMI REDDY ALLA Father's Name: KRISHNA REDDY ALLA House No: 1-15 Age: 51 Sex: Male	12 UOG0324863 Elector's Name: VISHNU VARDHAN REDDY KALAKOTA Father's Name: VENKATA RAMI REDDY House No: 1-25 Age: 50 Sex: Male

Figure - 1.1 (PDF document page)

Problem 2 : Fashion Matcher

For this problem, you have been given 5 sets of images. Each set contains two similar images. Your task is to determine a score depicting the percentage of similarity in those two images. The dataset is given for example purpose only. The score generated should be between 0 and 1, where a score of 1 defines identical images having 100 percent similarity and a score of 0 defines complete dissimilarity. In this task, we would like to know your approach to transfer learning.

For example:

Similarity score between these two images is 0.9

Similarity score between these two images is 0.56

For testing your solution, use the dataset given or choose any 10 similar images from internet (preferably from any e-commerce website) and store the output in a CSV file in the following format :
Image_Source,Image1_URL,Image2_URL,Similarity_Score

Problem 3 : Color Tagger

Mr. Khatri has a teenage daughter, who is suffering from color blindness. She can't recognize RED, BLUE and GREEN colors explicitly. She is interested in buying a colorful cloth online for her birthday party. Please help her in identifying the component of these three major colors in given fashion cloth.

You need to find out the proportions of Red, Green and Blue colors only for the clothing shown in given image.

For example:

Output for this image should be:

- Red : 70%
- Green : 0%
- Blue : 0%

For testing your solution use the given dataset. Write output in a csv file with ImageName, Red,Green,Blue as column header.