

# Cramstack

## Take-Home Tasks

Download all the contents of the [folder](#) to your own drive.

### 1. Optical Character Recognition

There are 2 document images in the images folder. Both images have one thing in common - each has a due date field with a value in the upper portion of it.

Open a notebook. Now the tasks are:

1. Read all the images in the folder.
2. Find all the **date** text (eg. due date:) and corresponding **values** text (eg. 1/1/1) and draw their **bounding boxes** in the image(just show the images in the notebook).
3. Print image name as id and date-value pairs in json or dictionary format eg.,

```
[{
  "id": "image_name",
  "date": "1/1/1"
},
{
  "id": "image_name_2",
  "date": "12/12/12"
}]
```

Save your notebook in a folder named 1. OCR

Hints:

- Use *pytesseract* for ocr; *cv2*,*pil* for image manipulation; *matplotlib* for visualization
- Crop the necessary upper portion of the image.
- Run the processes in a loop

### 2. Exploratory Data Analysis

The **game\_sales\_data.csv** dataset contains game sales and corresponding user and critic scores over the years. Now the focus of this task is to explore the dataset.

Open another notebook. Try to implement the following using pandas, matplotlib/seaborn.

1. Remove rows where the platform is not PC, PS4, PS3, Xone, X360. (do the following 2-4 tasks based on this). Count the number of distinct games.
2. How many games with unknown publishers are there ?
3. Visualize average critic score per developer in a scatter plot. (Include marker size depending on the value)
4. Show the top 5 games with biggest differences in critic vs user score.

**Save your notebook in a folder named 2. EDA**

### 3. Dashboard Creation

Power Bi desktop is needed for this task. You can download it from [here](#).

You are going to be using the same dataset from task 2. Load and transform the dataset as necessary. Then try to create the following charts in a single dashboard,

1. Pie chart of total shipped game copies per publisher.
2. A bar chart showing the top 10 average of critic scores by developer where the score is less than 10 but greater than or equal to 8.5 .
3. A column chart showing total shipped by platforms PC, PS4 and XOne since year 2015.
4. A line chart containing total shipped by the platforms PS, PS2, PS3, PS4 over the years.

**Save the .pbix and pdf file generated from the dashboard in a folder named 3. Dashboard**