## Scraping Instructions & Suggestions

- Use selenium in python for scraping so that you can scrap all websites with maximum data.
- Data needs to be extracted using headless mode when using selenium.
- Try to use "try & exception" at all possible points where the execution of the script may stop due to run time error.
- The first preference when you extract data of any element from the website is by using element id then class name and then using the path of the element.
- Make sure that you are extracting all upcoming events only from the website.
- Ideally, the script will be designed to get the data from the event detail page, but in case, the detail pages do not have any specific design/pattern then we can take details from the event listing page only.

## Requirement

- Scrap the event details (event name, event website, date, etc.) from the website.
- Data should be stored in a TSV file in the same sequence and format as shared in the file.
- Provide final script along with the TSV file having data extracted from the website using the script.

## Sample Websites:

- https://hopin.com/explore/search
- https://digimarcon.com/events/

## INSTRUCTIONS FOR DATA FIELDS

| | Field Name | Format | Example | Remarks |
|---|---|---|---|---|
| Mandatory | scrappedUrl | URL | https://hopin.com/events/syconnect | The URL of the page from where you are scrapping the details. |
| Mandatory | eventname | string | StreamYard Connect Summit | |
| Mandatory | startdate | YYYY-MM-DD | 2022-01-20 | If only single date is mentioned then both startdate & enddate will be same. |
| Mandatory | enddate | YYYY-MM-DD | 2022-01-21 | If only single date is mentioned then both startdate & enddate will be same. |
| Optional | timing | JSON | [{"type": "general", "Start_time": "01:00PM", "end_time": "02:00PM", "timezone": "CET", "days": "all"}] | # No spaces between Time & AM/PM<br># The keys are case sensitive |
| Optional | eventinfo | string | Join us on Thursday, January 20th for the StreamYard Connect Summit! | |
| Optional | ticketlist | JSON | [{"type":"paid","price":"15.00","currency":"$"}] | |
| Optional | orgProfile | string | Better live shows start here. | |
| Mandatory | orgName | string | StreamYard | |
| Mandatory | orgWeb | URL | https://streamyard.com/ | |
| Keep Blank | logo | - | - | - |
| Optional | sponsor | JSON | ['Gordon Research Conferences', 'National Institutes of Health'] | |
| Optional | agendalist | JSON | [{"start_time":" 2:00AM","end_time":" 4:00AM","day":"2020-10-28","desc":" ","title":" Way Forward: Lessons Learned from Maison & Objet Digital Fair"}] | |
| Keep Blank | type | - | - | - |
| Keep Blank | category | - | - | - |
| Mandatory | city | string | Kuala Lumpur | Can be left blank if **online_event** is 1 |
| Mandatory | country | string | Malaysia | Can be left blank if **online_event** is 1 |
| Optional | venue | string | Kuala Lumpur Convention Centre | |
| Mandatory | event_website | URL | https://hopin.com/events/syconnect | If official website of the event is not found, then use **scrappedUrl** in **event_website** |
| Optional | googlePlaceUrl | URL | https://www.google.com/maps/place/Kuala+Lumpur+Convention+Centre/@3.1532671,101.7130429,15z/data=!4m5!3m4!1s0x0:0x7706359b57f73cb7!8m2!3d3.1532671!4d101.7130429 | URL of Google Maps obtained from searching Venue string on Google. Below are the steps:<br>Step 1: Google Search: <venue>, <city>, <country><br>Step 2: Click on Google Map Image<br>Step 3: Wait for 4 seconds<br>Step 4: Copy the URL of page opened |
| Mandatory | ContactMail | JSON | ['marketing@streamyard.com', 'contact2@domain.com', 'contact3@domain.com'] | If contact is not available on the page, then a default static contact can be taken from the website's other pages like: Contact Us, Abount Us, etc. |
| Optional | Speakerlist | JSON | [{"name": "Cornelia Hoicke", "title": "Freie Auditorin", "link": ""},{"name": "XYZ Hoicke", "title": "Freie ABC", "link": ""}] | |
| Mandatory | online_event | 0 or 1 | 1 | 0: Not Happening Online<br>1: Happening Online |

***Note:***

- *If any of the Mandatory field is not present on the website, confirm with us first before proceeding.*
- *For all the fields mentioned as Optional, if the data is present on the website, then it must be processed (unless it is too difficult).*