In [1]:
```
import requests
import random
import pandas as pd
import numpy as np
from bs4 import BeautifulSoup
import re
import ast
```

In [ ]:
```
questions = []
choices = []
def getWonderPolls():
    for m in range(0,1000):
        URL = 'http://www.wonderpolls.com/new_polls.php?page=' + str(m)
        page = requests.get(URL)
        soup = BeautifulSoup(page.content, 'html.parser')
        if (soup.find_all(class_="poll_question") == None):
            continue
        queries = soup.find_all(class_="poll_question")
        for q in queries:
            title = q.find('font')
            question = list(filter(lambda option: option != '' and option != '
\n', title.text.split("  ")))[0]
            questions.append(question)
            answers = q.find_all(class_="poll_option")
            ans = []
            for answer in answers:
                asdf = list(filter(lambda option: option != '' and option != '
\n', answer.text.split("\n")))
                a = asdf[0][1:]
                ans.append(a)
            choices.append(ans)

getWonderPolls()
wonderPollsdf = pd.DataFrame(list(zip(questions, choices)),
               columns =['Question', 'Choices'])
```

In [ ]:
```
df_elements = wonderPollsdf.sample(n=10)
df_elements
```

In [ ]:
```
wonderPollsdf.to_csv('pollsData.csv', sep='\t', encoding='utf-8')
```

In [ ]:
```
rawPollsdf = pd.read_csv('rawPollsData.csv')

# Creating pollData.csv

df = rawPollsdf
sLength = len(df['Question'])
dateAdded = pd.Series(['2020-11-12']*sLength)
creator = pd.Series([1]*sLength)
df = df.assign(dateAdded=dateAdded.values)
df = df.assign(creator=creator.values)
df = df.drop('Choices', 1)
df.to_csv('pollData.csv', encoding='utf-8')


# Creating choiceData.csv
df = rawPollsdf
choiceList = []
pollIDList = []
for i in range(0, len(df['Question'])):
    l = ast.literal_eval(df['Choices'][i])
    for j in range(0, len(l)):
        pollIDList.append(i)
        choiceList.append(l[j])

df = pd.DataFrame(list(zip(pollIDList, choiceList)),
                              columns =['PollID', 'ChoiceTitle'])
sLength = len(df['PollID'])
dateAdded = pd.Series(['2020-11-12']*sLength)
creator = pd.Series([1]*sLength)
rating = pd.Series([0]*sLength)
df = df.assign(dateAdded=dateAdded.values)
df = df.assign(creator=creator.values)
df = df.assign(rating=rating.values)
df
df.to_csv('choiceData.csv', encoding='utf-8', index=False)
```