Case Study
Data Analysis
Report

Programming Languages used:

Python

Other Tool:

Jupyter Notebook

API Used:

Guardian Media Group API

Packages Used:

- requests
- datetime
- pandas
- time
- matplotlib
- numpy
- schedule
- os
- import csv

SUMMARY:

I used python in Jupyter-notebook to do the assigned task. The API endpoint from Guardian API was called with python requests package with the following parameters:

```
params = {
    "from-date": Starting Date,
    "to-date": Ending Date,
    "api-key": "test",
    "q": "Justin Trudeau",
    "order-by": "oldest",
    "page": str(page_no)
}
```

The API was called twice per date (Ranging from 2018-01-01 until today). First it was called to collect articles which contain "Justin Trudeau" in its body or title, secondly, it was called to filter the articles which have "Justin Trudeau" in its Title. This was done to check the relevance of the article with respect to the query. The API returns the article without the relevance of the query to it. Since we want to focus on the articles which are focused on Justin Trudeau and not the articles which have him mentioned somewhere in the body only.

The following tasks are also implemented in Jupyter-notebook with name "`Finalised.ipynb`" and corresponding report in "`Report.pdf`". The **daily automated job** part is implemented in "`scheduled_guardian.py`"

Counting total articles about Justin Trudeau since 01.01.2018 until today:

Again, by relevance, I collected both datasets (i. Ones containing Justin Trudeau in body OR title, ii. Ones containing Justin Trudeau in title i.e., more relevant ones). Total articles collected in both case are given as follows:

```
Articles which Contain Justin Trudeau in their Title OR in their Body
Total posted articles until today:  3455

Articles which Contain Justin Trudeau in their Title Only
Total posted articles until today:  164
```

The output stored for the above mentioned two cases is presented in a dataframe as follows:

```
1 output_mentions
```

	Date	No. of Articles
0	2018-01-01	2
1	2018-01-02	1
2	2018-01-03	2
3	2018-01-04	0
4	2018-01-05	1
...	...	...
1778	2022-11-14	5
1779	2022-11-15	1
1780	2022-11-16	6
1781	2022-11-17	4
1782	2022-11-18	2

1783 rows × 2 columns

```
1 output_titles
```

	Date	No. of Articles
0	2018-01-01	0
1	2018-01-02	0
2	2018-01-03	0
3	2018-01-04	0
4	2018-01-05	0
...	...	...
1778	2022-11-14	0
1779	2022-11-15	1
1780	2022-11-16	0
1781	2022-11-17	1
1782	2022-11-18	0

1783 rows × 2 columns

Average of all days for the above mentioned period from “No. of Articles”:

For 1783 days, following come out to be the averages of the both datasets:

```
Articles which Contain Justin Trudeau in their Title OR in their Body
Average articles per day:  1.9377453729669096

Articles which Contain Justin Trudeau in their Title Only
Average articles per day:  0.09197980931015143
```

In which section are the most articles written:

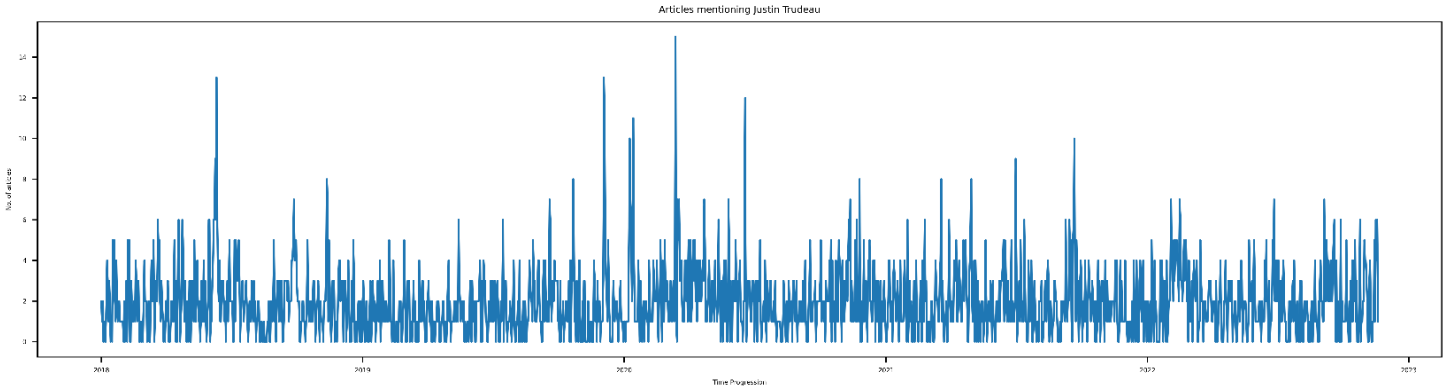
While collecting the information, I also collected the section names of the article and that output was stored in a different dataset. Later on I counted the sections and sorted them to calculate the sections in which the most articles were written. The output comes out as follows:

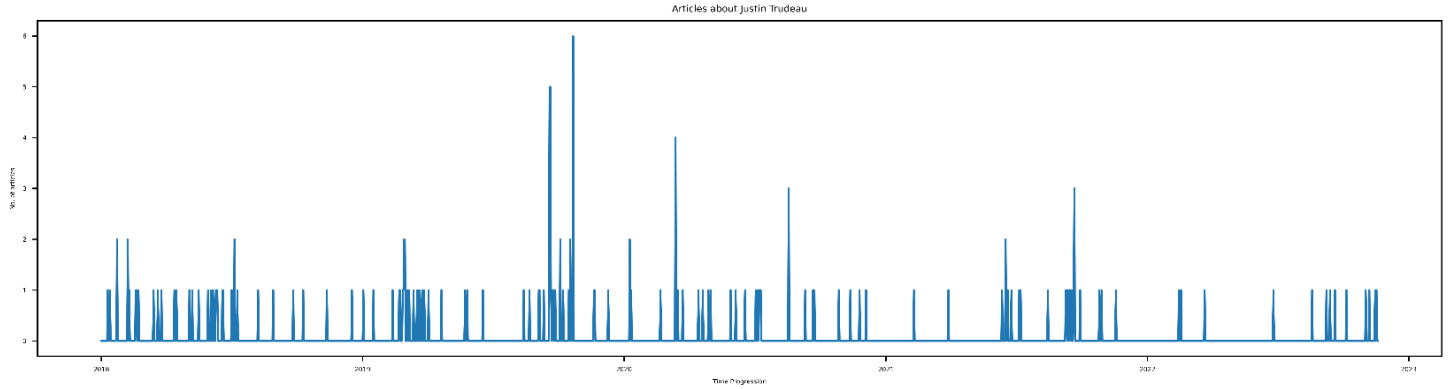
```
For the DataFrame - Articles mention Justin Trudeau in Title OR Body
Section with Most Articles:  ('World news', 960)

For the DataFrame - Articles mention Justin Trudeau in Title
Section with Most Articles:  ('World news', 127)
```

Evolution of the "No. of Articles" over time for the above period:

For that I just plotted "No. of Articles" versus Dates for both datasets to see the time evolution process as done in the Time series analysis. For the both datasets for same query, the output is presented in the

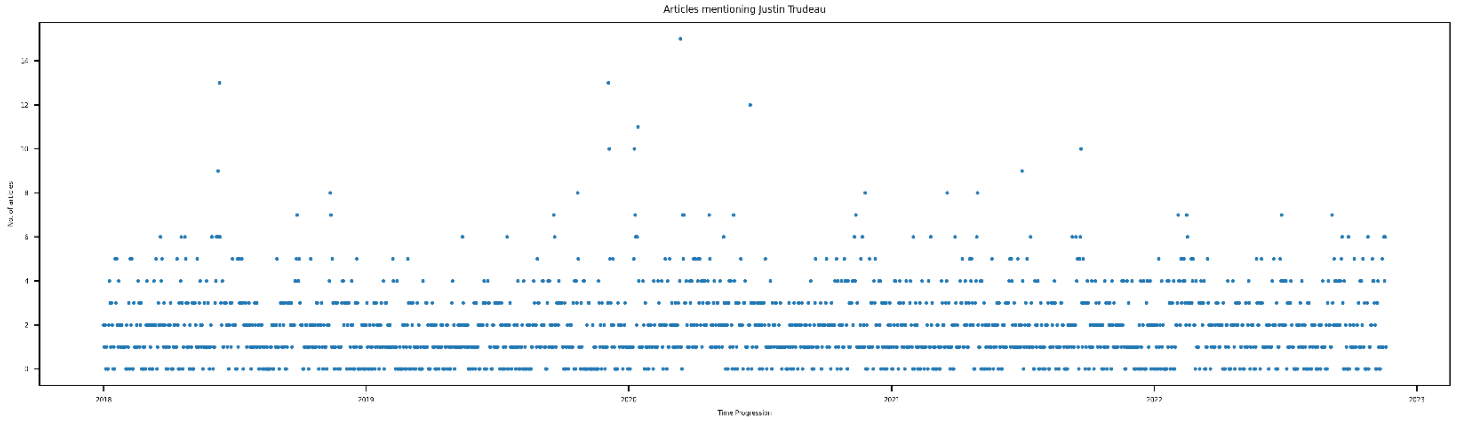


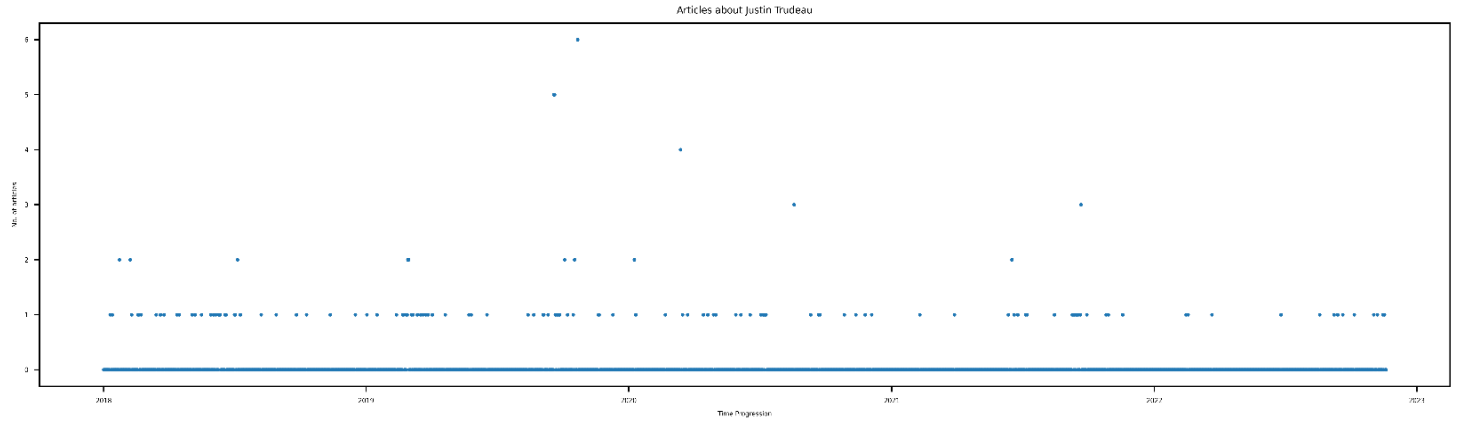


following picture:
(Please pardon the small font. I am attaching these pictures in the email as well.)
x-Axis: Time Progression
y-Axis: No. of Articles

Are there any unusual events in the time series under investigation?:

In this simple case of one dependent and one independent variable, an unusual event most likely corresponds to the obvious outliers. As we can see in the top graphs, there are larger spikes in the time series and they seem less frequent. I am therefore considering them to be the odd ones out and thence as unusual. A more visually appealing way of separating them is by using a scatter plot as it is less messy. Scatter plot versions of the previously shown graphs are as follows:



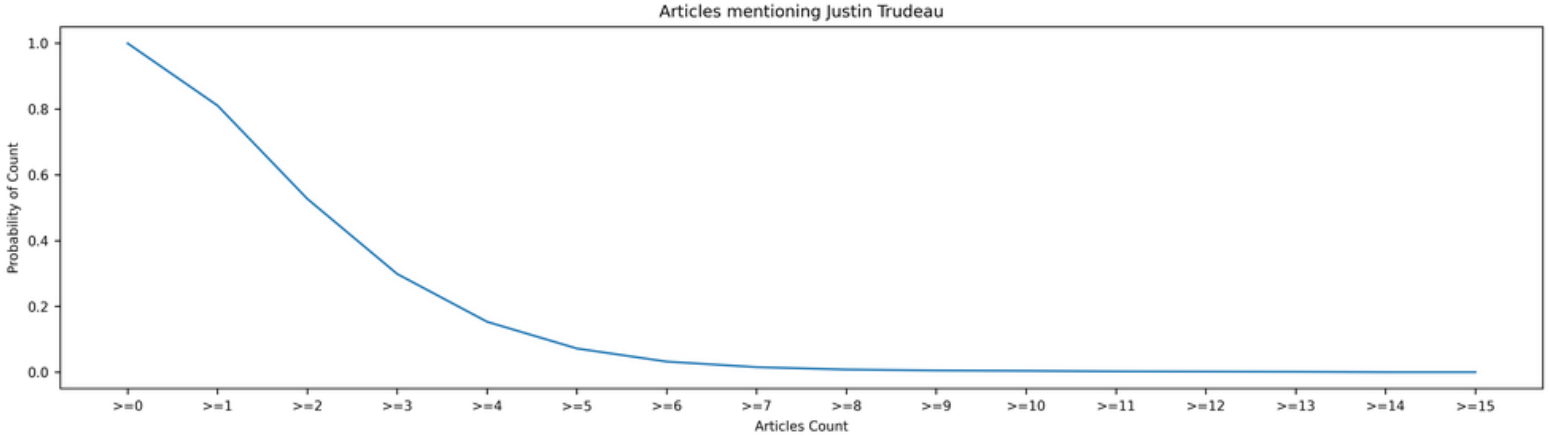


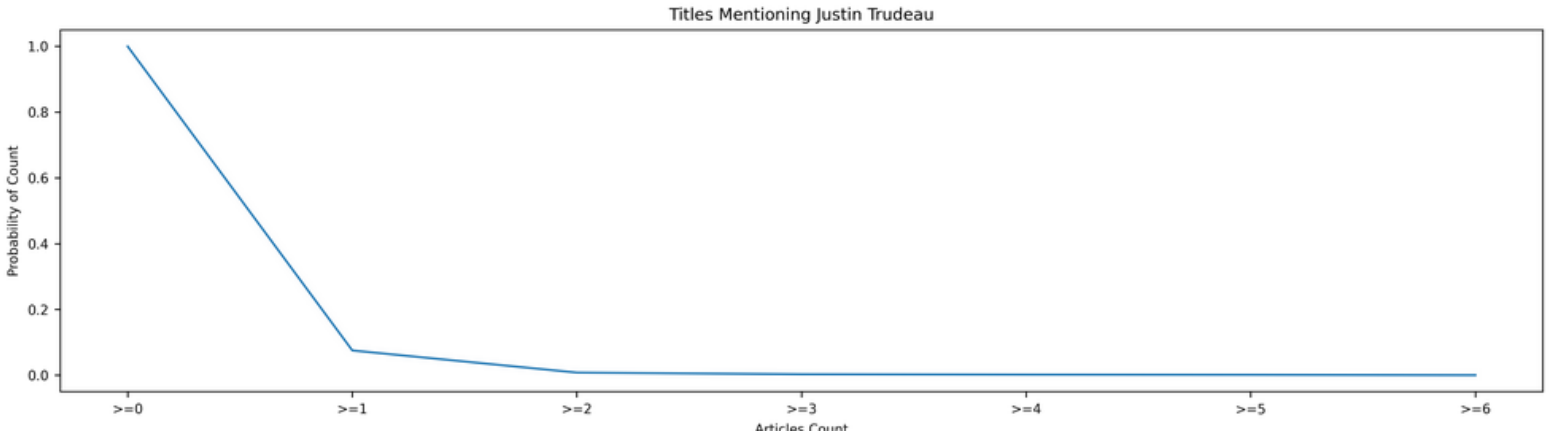
(Once again, my apologies for the small font. I am attaching this too in the email.)

x-Axis: Time Progression
y-Axis: No. of Articles

Why are these unusual? (Define for yourself what you want to show by ordinary or unusual):

These events are unusual as they correspond to the outliers. Here by an ordinary event, I refer to the values of No. of Articles which are close to the average number of articles per day. By unusual events, I refer to the values of No. of Articles which are far far away from the mean value. In terms of probability. The values which are in higher occurrence are usual and vice versa. To visualise the probabilities in a graph, I counted the articles greater than 0, 1, 2, until the largest single value per day of the total articles posted for the both datasets and plotted the probabilities of their occurrence versus the counts. The graphs come out as follows:





In the top graph, we can see that most of the days, the number of articles being posted is zero - which is the event with the highest probability of occurrence over the time period. As the number of articles being posted each day starts increasing, that event becomes more and more rare. Around the ***Elbow of the Graph*** and afterward on the **x-Axis**, is where we can see the unusual events with the most unusual event towards the rightmost of the x-axis with the least probability of occurrence.

- that point for the top graph is approximately between 6 and 7 (we will use 7 for simplicity)
- that point for the bottom graph is 2

It is rare that Articles containing Justin Trudeau (title and body) more than 7 a day
It is rare that Articles containing Justin Trudeau (title only) more than 2 a day

Now I will extract the dates where this has happened and will have a look at the Titles to infer what might have been an unusual thing during that day. This was also done using a piece of code and the output is as follows for all dates in both datasets where “No. of Articles” is greater than 2:

	Date	No. of Articles	Article_Names	Section
8	2018-01-09	4	[Cabinet reshuffle: Justine Greening quits the...	[Politics, Politics, Opinion, UK news]
9	2018-01-10	3	[The media should not settle for ‘truthiness’ ...	[Opinion, World news, Technology]
11	2018-01-12	3	[Debunked: Trump reasons for cancelling London...	[US news, Opinion, Music]
16	2018-01-17	5	[Justin Rose hits back at ‘boring golf’ tag – ...	[Sport, Environment, Music, Politics, Politics]
17	2018-01-18	3	[Tide Pod challenge: YouTube clamps down on 'd...	[Technology, US news, Culture]
...	...	...	...	...
1771	2022-11-07	3	[How can we cut soaring demand for meat? Try a...	[Opinion, Music, World news]
1772	2022-11-08	4	[Afternoon Update: Liberal senator’s ‘offensiv...	[Australia news, Politics, Environment, Football]
1778	2022-11-14	5	[G20 explainer: everything you need to know ab...	[World news, World news, Politics, World news,...
1780	2022-11-16	6	[SES assesses flood damage in NSW’s central we...	[Australia news, Politics, Film, World news, W...
1781	2022-11-17	4	[Xi angrily rebukes Trudeau over ‘leaks’ to me...	[World news, Australia news, US news, World news]

	Date	No. of Articles	Article_Names	Section
626	2019-09-19	5	[Justin Trudeau brownface: Canada PM apologise...	[World news, World news, US news, Australia ne...
627	2019-09-20	5	[Trudeau says he can't recall how many times h...	[World news, US news, Opinion, World news, Wor...
659	2019-10-22	6	[Justin Trudeau's victory is a death knell for...	[World news, World news, World news, US news, ...
802	2020-03-13	4	[Justin Trudeau in self-isolation after wife S...	[World news, US news, World news, World news]
960	2020-08-18	3	[Canada finance minister resigns amid charity ...	[World news, World news, World news]
1359	2021-09-21	3	[Canada election result: Trudeau wins third te...	[World news, World news, World news]

Now here I just took the intersection of the dates between the two dataframes and the intersection of the corresponding article names to see the date and the articles posted on those days which are focusing on Justin Trudeau. The output is as follows:

```
Date :  2019-09-19
COMMON ARTICLES:
- Justin Trudeau brownface: Canada PM apologises after image emerges
- US briefing: Greta Thunberg, Justin Trudeau and a Trump whistleblower
- How will Justin Trudeau's blackface photos affect Canada's election?
- Morning mail: climate strike, Trudeau blackface, bird extinctions
- Thursday briefing: Trudeau apologises for 'brownface' picture
Date :  2019-09-20
COMMON ARTICLES:
- US briefing: climate strike, Trump whistleblower and Justin Trudeau
- Trudeau tries to shift focus from brownface images to gun control
- Justin Trudeau’s blackface can’t be wiped away | Letters
- Justin Trudeau's brownface scandal is bad. But voting him out isn't the solution | Moustafa Bayoumi
- Trudeau says he can't recall how many times he wore blackface makeup
Date :  2019-10-22
COMMON ARTICLES:
- The Guardian view on the Canadian election: a win for Trudeau, but not a triumph | Editorial
- US briefing: Trudeau's narrow win, GOP disunity and ocean acidification
- Trudeau faces rough road as Canada's minority parties lay out their conditions
- Canada election 2019: 'We'll govern for everyone' says Trudeau, after narrow win – as it happened
```

- Justin Trudeau's victory is a death knell for Canada's fledgling far-right
- Canada elections: Trudeau wins narrow victory to form minority government

Date : 2020-03-13
COMMON ARTICLES:
- Justin Trudeau in self-isolation after wife Sophie tests positive for coronavirus
- Fears grow Sophie Grégoire Trudeau picked up coronavirus on London trip
- Justin Trudeau announces sweeping steps to tackle coronavirus in Canada
- US briefing: Trudeau quarantined, Iran mass graves and Chelsea Manning

Date : 2020-08-18
COMMON ARTICLES:
- Canada: departure of finance minister suggests Trudeau will pursue 'green' recovery plan
- Trudeau accused of attempting to cover up scandal by proroguing parliament
- Canada finance minister resigns amid charity scandal and reports of tensions with Trudeau

Date : 2021-09-21
COMMON ARTICLES:
- Trudeau didn't win the majority but still has chance to pass sweeping legislation
- Justin Trudeau secures a third victory in an election 'nobody wanted'
- Canada election result: Trudeau wins third term after early vote gamble

Show the causes of the unusual events:

By looking at the article names and dates and with a little bit of look up on the internet, following seems to be the causes behind the unusual events with respect to the dates:

1) 2019, September 19th (CAUSE: *Photo Scandal*)

An old photo of Justin Trudeau emerged on social media showing him wearing black/dark brown make-up on his face as a part of the costume during a school party where he was a teacher. This caused an outrage which eventually made him apologise for it. The timing of the picture was odd given the upcoming Federal Elections for the year 2019 were to be held on October 21st and Justin Trudeau was the serving Prime Minister at the time.

Following 4 news articles are about Justin Trudeau and related to the Brown Face Incident:

- Justin Trudeau brownface: Canada PM apologises after image emerges
- How will Justin Trudeau's blackface photos affect Canada's election?
- Morning mail: climate strike, Trudeau blackface, bird extinctions
- Thursday briefing: Trudeau apologises for 'brownface' picture

While the following article mentions him and that event along with other news:

- US briefing: Greta Thunberg, Justin Trudeau and a Trump whistleblower

2) 2019, September 20th (CAUSE: *Photo Scandal*)

In reference to the previous explanation, the talk related to his old photo in the light of upcoming elections was still going on and was pretty fresh in the press.

Following 4 news articles are about Justin Trudeau and related to the Brown Face Incident:

- Trudeau tries to shift focus from brownface images to gun control
- Justin Trudeau’s blackface can’t be wiped away | Letters
- Justin Trudeau's brownface scandal is bad. But voting him out isn't the solution | Moustafa Bayoumi
- Trudeau says he can't recall how many times he wore blackface makeup

While the following article mentions him and that event along with other news:

- US briefing: climate strike, Trump whistleblower and Justin Trudeau

3) 2019, October 22th (CAUSE: *Canadian Federal Election 2019 and Trudeau's victory*)

The day after Canadian Federal Elections of 2019 which were won by Justin Trudeau by a narrow margin yet not gaining a majority which is to have more than 35% of the seats. This led him to form a minority government in coalition.

The following 5 articles focus on Justin Trudeau's election victory and the supposed situations to follow:

- The Guardian view on the Canadian election: a win for Trudeau, but not a triumph | Editorial
- Trudeau faces rough road as Canada's minority parties lay out their conditions
- Canada election 2019: 'We'll govern for everyone' says Trudeau, after narrow win – as it happened
- Justin Trudeau's victory is a death knell for Canada's fledgling far-right
- Canada elections: Trudeau wins narrow victory to form minority government

The following article mentions briefly his electoral victory along with other topics:

- US briefing: Trudeau's narrow win, GOP disunity and ocean acidification

4) 2020, March 13th (CAUSE: *Justin Trudeau Testing Positive for COVID*)

Justin Trudeau tested positive for coronavirus along with his wife and went into self quarantine. The other article in the same relation to the news mentions Trudeau's approach for tackling Coronavirus situation in Canada

The following 2 articles focus on Justin Trudeau's covid infection and election victory and the supposed situations to follow:

- Justin Trudeau in self-isolation after wife Sophie tests positive for coronavirus
- Justin Trudeau announces sweeping steps to tackle coronavirus in Canada

The following 2 articles mentions Trudeau's wife's possible contact point for getting infected in relation to Justin Trudeau's covid positive news and the about Trudeau's quarantine being mentioned along with two other news in the same article:

- Fears grow Sophie Grégoire Trudeau picked up coronavirus on London trip

- US briefing: Trudeau quarantined, Iran mass graves and Chelsea Manning

5) 2020, August 18th (CAUSE: *WE Charity Scandal, Parliament Prorogation, Morneau's Resignation*)

Justin Trudeau's Finance Minister Bill Morneau's name came up along with Prime Minister's related to WE Charity scandal. Trudeau prorogued the parliament, as done in the past, to avoid political scrutiny. The following three articles revolve around this issue:

- Canada: departure of finance minister suggests Trudeau will pursue 'green' recovery plan
- Trudeau accused of attempting to cover up scandal by proroguing parliament
- Canada finance minister resigns amid charity scandal and reports of tensions with Trudeau

6) 2021, September 21st (CAUSE: *Federal Elections Canada 2021 and Trudeau's victory*)

Justin Trudeau called for early elections amidst the Corona crisis seeking public mandate. He secured a third victory in the office and the following three articles are related to this event:

- Trudeau didn't win the majority but still has chance to pass sweeping legislation
- Justin Trudeau secures a third victory in an election 'nobody wanted'
- Canada election result: Trudeau wins third term after early vote gamble

Create a daily automated job that updates question 5 daily and creates an output that could be sent to recipients who have not seen the data before:

For this part, please see the file "`scheduled_guardian.py`". It runs the task and saves the time evolution graphs of the data which could be sent to recipients who have not seen the data before. The Task executes by itself everyday at 9:00am. Saves the time-evolution Graphs in the same folder which could be sent to recipients who have not seen the data before.

The Time evolution graphs are stored as the following two files:

```
EVOLUTION_PLOT_NORMAL.png
EVOLUTION_PLOT_SCATTER.png
```

Jupyter Notebook report is in

```
Report.pdf
```

```
———————————————————— The End ————————————————————
```