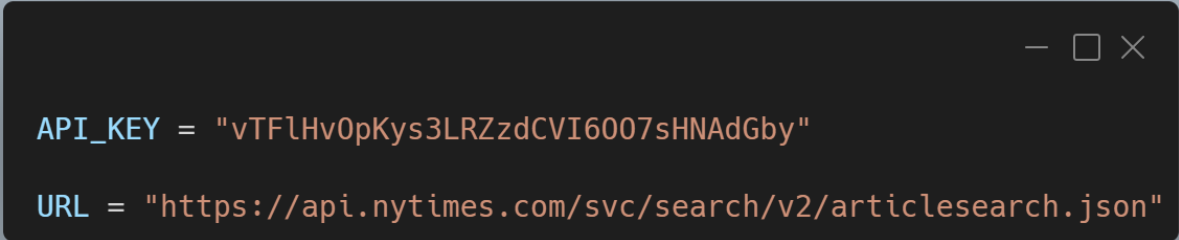# FC Barcelona and Travel

Section 1: Introduction

As a soccer fan, I was interested in how soccer can affect a city in tangible ways. As such, I wanted to see if, looking at data from 2014-2024, the number of New York Times Travel articles that reference Barcelona rise and fall in conjunction with FC Barcelona's final placement in La Liga.

I needed to make use of the NY Times Article Search API to gather data on the amount of hits for articles during each season of soccer in this timeframe. To accomplish my task, I needed to take the raw data from the API, turn it into just the integer of hits for each year, and put that cleaned data into a csv file to analyze easier.

I believe I will find that the higher FC Barcelona places in the league, the more hits for NYT travel articles referencing Barcelona over the year. This would suggest that soccer contributes to a city in other ways than just the revenue it generates–it also makes the city a tourist destination, a travel spot for Americans, contributing more money to the area.
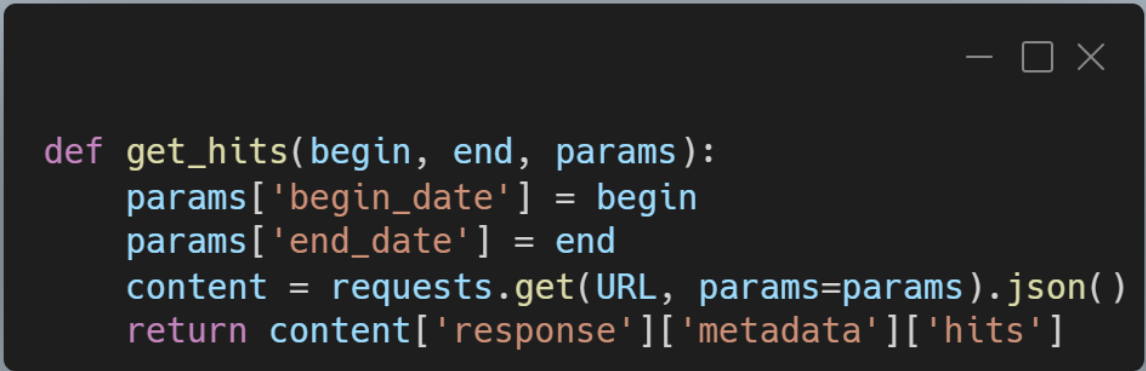
Section 2: Methods

I created a Python script to make use of the NY Times API and gather the data I needed. Here are some of the most important aspects of the script.

```
API_KEY = "vTFlHvOpKys3LRZzdCVI6OO7sHNAdGby"

URL = "https://api.nytimes.com/svc/search/v2/articlesearch.json"
```

The first part, and arguably the most important, was inputting the constants for the API key and URL. Without these, I would be unable to make use of the NY Times API and gather the information I was searching for.

```
def get_hits(begin, end, params):
    params['begin_date'] = begin
    params['end_date'] = end
    content = requests.get(URL, params=params).json()
    return content['response']['metadata']['hits']
```

After my constants and various variables were prepared, I needed to create a function to obtain the number of hits of certain parameters. I knew I needed to do 10 API calls, so a function would allow me to avoid repeating code and keep everything more succinct. The function takes in a beginning date, end date, and any other parameters that I need (like queries and tags), and returns the hits.

```
counter = 0
for begin, end in DATES.items():
    data[counter]['Hits'] = get_hits(begin, end, params)
    counter += 1
    time.sleep(12)
```

This for loop makes use of a constant DATES dictionary that I created that holds key:value pairs of the beginning and ending dates for each season of La Liga from 2014-2024. It then iterates through each season, using the get_hits function that we created earlier to input a key:value pair into our data dictionary that holds the amount of hits procured for each season. I also use a sleep function to take a break every 12 seconds, so I do not exceed our maximum API calls per minute.
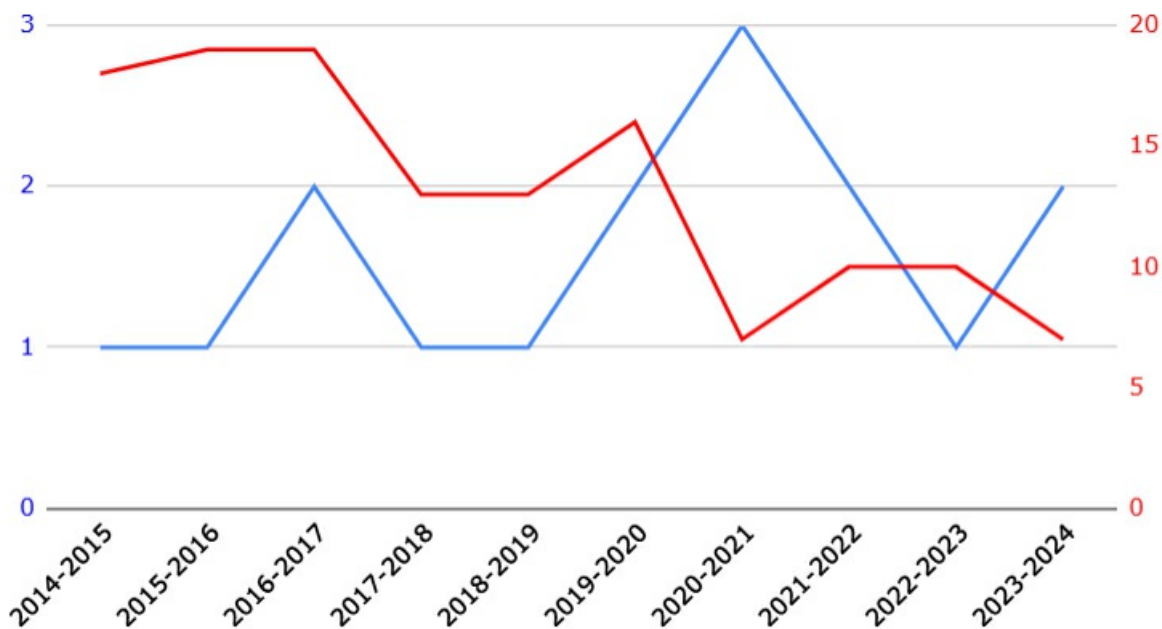
```
cwd = pathlib.Path.cwd()
file = cwd/"final.csv"
file.touch()
with file.open(mode='w', encoding='utf-8',newline='') as python_final:
    writer = csv.DictWriter(python_final, fieldnames = \
                            ["Season", "Standing", "Hits"])
    writer.writeheader()
    writer.writerows(data)
```

Finally, I must permanently save what I have done. I write the newly cleaned up data to a CSV file, alongside the season it is associated with and the final standing for FC Barcelona in each season (which is information I was able to find documented on Wikipedia).

With this, the process is complete. Data has been extracted from the NY Times API, transformed into data that is specifically useful to my question, and loaded into a CSV file for human use.

Section 3: Findings

Using the CSV file that I created, I was able to create a graph of the data.

In the red, we see the number of hits for travel articles referencing Barcelona during each season, and in the blue, we see the FC Barcelona final standing for each season. What we expect to see is an inverse relationship, where a lower number for the standing (which is really a higher placement in the league) leads to a higher amount of hits.

There are some years where this is true, like 2014-2015 or, notably, 2020-2021, and some years where it is not, like 2015-2020. It is unsurprising that the data does not show a clear picture, as COVID affected both travel and soccer around the 2019-2020 season. I was surprised, however, that there was a direct relationship between the two variables prior to COVID, when the data would be closest to normal.

It seems that there may be some correlation between the two variables, but further testing is certainly required to be sure.

Section 4: Next Steps

The most important step I should take to continue this project is broadening its horizons by searching across more years. A decade is too short of a time period to gather representative data, especially considering half of it was affected by (or dealing with the aftermath of) a global pandemic.

I can also dig into the articles themselves and see in what context they reference Barcelona. Is the city the main focus of the article, declared directly in the title, or is it simply mentioned at some point in the article with no real substance?

Finally, I can look across other teams and see what the data says about them. Realistically, sports teams do change cities and countries entirely, affecting how they are perceived by the public and how much money they bring in. America being less involved in soccer than other countries also means that the writers at the New York Times may not focus on it as much, leading to less articles inspired in any way by La Liga. My question is a small one in the grander understanding of how sports teams affect the places they represent, and if I were to develop this further, I would certainly go further than simply focusing on Barcelona.