Day 7: *Descriptive statistics in Pandas*

With your help, Dot expertly solved the airport crisis and saved the day. The crowd cheered and raised Dot up over their heads, carrying them onto the plane. The flight crew even gave Dot a first-class seat for their short flight, providing them with complimentary snacks and drinks throughout the trip. Dot arrived at their third European destination, Barcelona, feeling rested and exhilarated. Naturally, they wanted to prolong their feelings of elation by going to a high-energy environment — a club? A festival? At the hotel, Dot asked the concierge if he had any ideas. "Of course, esteemed patron and guest, you must attend a football game," the concierge insisted.

Of course, Dot thought. They were in the home of the popular football club FC Barcelona, and the team was playing a game that very afternoon! Dot made their way over to the Camp Nou stadium, grabbing some delicious tapas to go on the way. They settled into their seat at the game among hundreds of cheering fans, some holding flags and signs, some with their faces and bodies painted to show their team loyalty. Dot didn't know very much about FC Barcelona, but they were happy to be there among so much excitement. They pulled out their smartphone to do some quick research on the team's statistics before the start of the game.

Tutorial

Part of the job of data scientists and data analysts is to *understand* the data and *extrapolate* certain findings. In most cases, the first step of understanding a numerical data set is to look at the **descriptive statistics**. Descriptive statistics are *brief descriptions that summarize datasets*. There are many great Python libraries we can use to get descriptive statistics from numerical datasets. One such library is **Pandas**. We won't be going too in-depth with pandas today, just the basics to understand how to use pandas to get some general descriptive statistics. We will be covering pandas extensively and going over some of the advanced functions starting on Day 8.

Pandas is one of the most widely used Python packages. Pandas can be used when working with large data sets or performing data cleaning, manipulation, and analysis.

Since the Pandas library is external to base Python, we will need to import it. When importing, we give an alias to pandas to shorten our code when calling on functions from pandas. Instead of writing ***pandas.name_of_function()*** we'll be able to write ***pd.name_of_function()***. The alias ***pd*** is standard within the Python community.

```
import pandas as pd # we must import our external python plugin

list_of_num = [1,2,3,4,5]

series = pd.Series(list_of_num) #converting our list_of_num to a pandas series
variable
                                #we need to do this to use some of pandas'
useful descriptive statistics functions

series.max()    #outputs maximum value in Pandas Series
series.min()    #outputs minimum value in Pandas Series
```

```
series.mean()   #outputs average value in Pandas Series
series.median() #outputs median value in Pandas Series
series.mode()   #outputs mode value in Pandas Series
```

To read more on descriptive statistics, read this article.

To learn more about the various pandas functions, check out the user guide in the pandas documentation.

Why should I use the documentation?

On the job as a data scientist or data analyst, more often than not, you may find yourself looking up the documentation of a particular function or plugin you use. Don't worry if there are a few functions you don't know by heart. There are just too many to know! An essential skill is to learn how to navigate documentation and understand how to apply the examples to your work.

Challenge

To help Dot learn more about FC Barcelona's statistics, we have the historical La Liga results of the 1988/1989 season.

Dot needs you to figure out the answers to the following questions:

1. What is the maximum amount of games Barcelona playes in 1 season?
2. What is the average attendance across the seasons?
3. What is the difference between median value of wins and losses?
4. What is the minimum number of games Barcelona managed to win in 1 season?
5. What is the difference between max and min amount of points Barcelona was able to get in all seasons?

In [4]:
```
import pandas as pd
df = pd.read_csv('fc_barcelona.csv')
df.head()
```

Out[4]:

	Season	Squad	Country	Comp	LgRank	MP	W	D	L	GF	GA	GD	Pts	Attendance	Top Team Scorer
0	2020-2021	Barcelona	es ESP	1. La Liga	3rd	38	24	7	7	85	38	47	79	NaN	Lionel Messi - 30
1	2019-2020	Barcelona	es ESP	1. La Liga	2nd	38	25	7	6	86	38	48	82	54223.0	Lionel Messi - 25

	Season	Squad	Country	Comp	LgRank	MP	W	D	L	GF	GA	GD	Pts	Attendance	Top Team Scorer
2	2018-2019	Barcelona	es ESP	1. La Liga	1st	38	26	9	3	90	36	54	87	76104.0	Lionel Messi - 36
3	2017-2018	Barcelona	es ESP	1. La Liga	1st	38	28	9	1	99	29	70	93	67142.0	Lionel Messi - 34
4	2016-2017	Barcelona	es ESP	1. La Liga	2nd	38	28	6	4	116	37	79	90	78678.0	Lionel Messi - 37

In [6]:
```
points = df.Pts
games_played = df.MP
wins = df.W
losses = df.L
attendance = df.Attendance.dropna() # skipping missing values (NaN) because there were
```

In [8]:
```
# SOLUTION

print(games_played.max())
print(attendance.mean())
print(wins.median() - losses.median())
print(wins.min())
print(points.max() - points.min())
```

```
42
72579.85714285714
19.0
15
54
```