NFL_EDA

June 2, 2023

1 National Omnibus Data Analytics Platform

1.0.1 Setup

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import requests
from bs4 import BeautifulSoup
import matplotlib.pyplot as plt

from IPython.display import Image, display
```

1.1 Social Media Trends

1.1.1 NFL Social Media (2023)

[2]:
```
url = 'https://www.wsn.com/nfl/least-most-popular-nfl-teams-facebook/'
response = requests.get(url)
html_content = response.content
soup = BeautifulSoup(html_content, 'html.parser')

# Headers
head_tags = soup.find_all('strong')
test = [tag.text for tag in head_tags]
test_keep = test[:9]

# Data
table_tags = soup.find_all('td')
test_nfl = [tag.text for tag in table_tags]

# Removing header copies
test_nfl = test_nfl[6:]

# Separating into categories for headers
team = [test_nfl[i] for i in range(1, len(test_nfl), 6)]
fb_likes = [int(test_nfl[i].replace(',', '')) for i in range(2, len(test_nfl),␣
 ↪6)]
```

```
twt_follow = [int(test_nfl[i].replace(',', '')) for i in range(3,␣
 ↪len(test_nfl), 6)]
ig_follow = [int(test_nfl[i].replace(',', '')) for i in range(4, len(test_nfl),␣
 ↪6)]
```

[3]:
```
# Creating dataframe for NFL social media data
nfl_social = pd.DataFrame()
nfl_social['Tm'], nfl_social['fb_likes'], nfl_social['twt_follow'],␣
 ↪nfl_social['ig_follow'] = team, fb_likes, twt_follow, ig_follow
```

[4]:
```
# Checking data
nfl_social.head()
```

[4]:

	Tm	fb_likes	twt_follow	ig_follow
0	Dallas Cowboys	8558000	3731000	3200000
1	New England Patriots	7035000	4279000	4100000
2	Pittsburgh Steelers	6333000	3329000	2300000
3	Green Bay Packers	5303000	2211000	1800000
4	Seattle Seahawks	3917000	2323000	2300000

1.1.2 NBA Social Media (2022)

[5]:
```
url = 'https://www.wsn.com/nba/most-least-popular-nba-teams-social-media/'
response = requests.get(url)
html_content = response.content
soup = BeautifulSoup(html_content, 'html.parser')

# Headers
head_tags = soup.find_all('th')
headers_nba = [tag.text for tag in head_tags]

# Data
table_tags = soup.find_all('td')
test_nba = [tag.text for tag in table_tags]

# Separating data
team = [test_nba[i] for i in range(1, len(test_nba), 6)]
twt_likes = [int(test_nba[i].replace(',', '')) for i in range(2, len(test_nba),␣
 ↪6)]
ig_follow = [int(test_nba[i].replace(',', '')) for i in range(3, len(test_nba),␣
 ↪6)]
fb_follow = [int(test_nba[i].replace(',', '')) for i in range(4, len(test_nba),␣
 ↪6)]
```

[6]:
```
# Creating a dataframe for NBA social media data
nba_social = pd.DataFrame()
```

```
nba_social['Tm'], nba_social['twt_likes'], nba_social['ig_follow'],␣
↪nba_social['fb_follow'] = team, twt_likes, ig_follow, fb_follow
```

[7]:
```
# Checking dataframe
nba_social.head()
```

[7]:

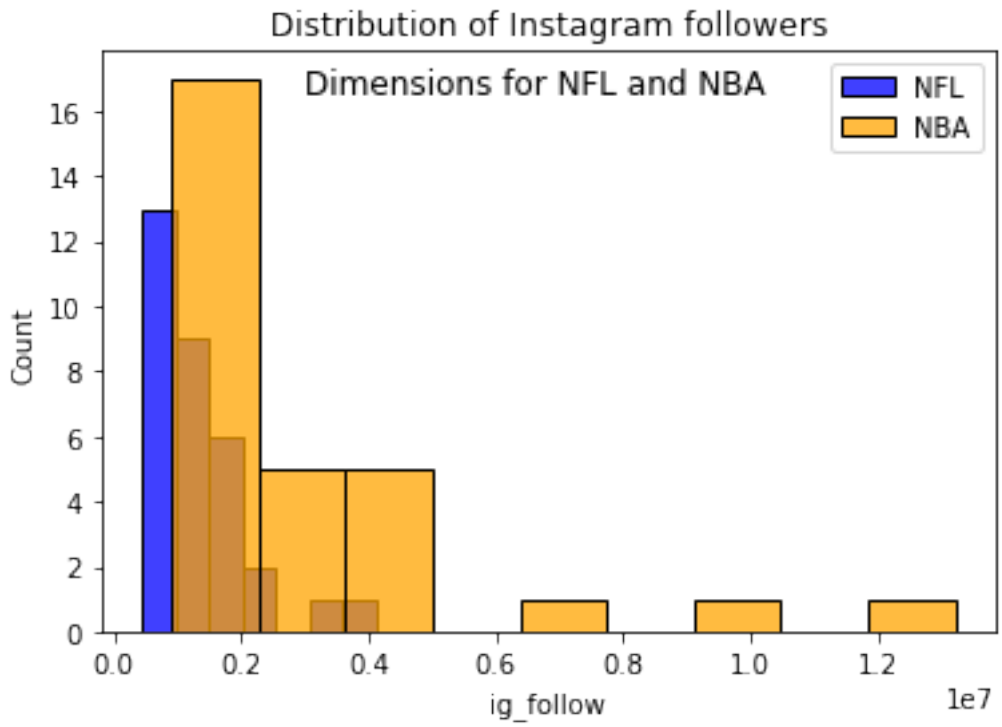
	Tm	twt_likes	ig_follow	fb_follow
0	Los Angeles Lakers	8155399	9960137	21583829
1	Golden State Warriors	6382948	13201981	11695064
2	Chicago Bulls	4113064	4776771	17651233
3	Miami Heat	4677528	3710653	15188432
4	Houston Rockets	2972379	4344989	12974848

1.1.3 NFL vs NBA Social Media

On average, NFL teams have much lower social media following as seen with instagram followers below. However, other research indicates this difference is correlated to the age groups interested in each sport. This data supports the idea that most NFL fans are in their 30s and above whereas NBA has greater social media presence due to younger fan base and greater amounts of brand deals for players and teams.

[8]:
```
# Plotting IG follower distributions
sns.histplot(nfl_social['ig_follow'], color='blue', label='NFL')
sns.histplot(nba_social['ig_follow'], color='orange', label='NBA')

plt.title('Distribution of Instagram followers')
plt.suptitle('Dimensions for NFL and NBA', y=0.86, fontsize=12)
plt.legend()
```

[8]:
```
<matplotlib.legend.Legend at 0x7f67662e8e50>
```



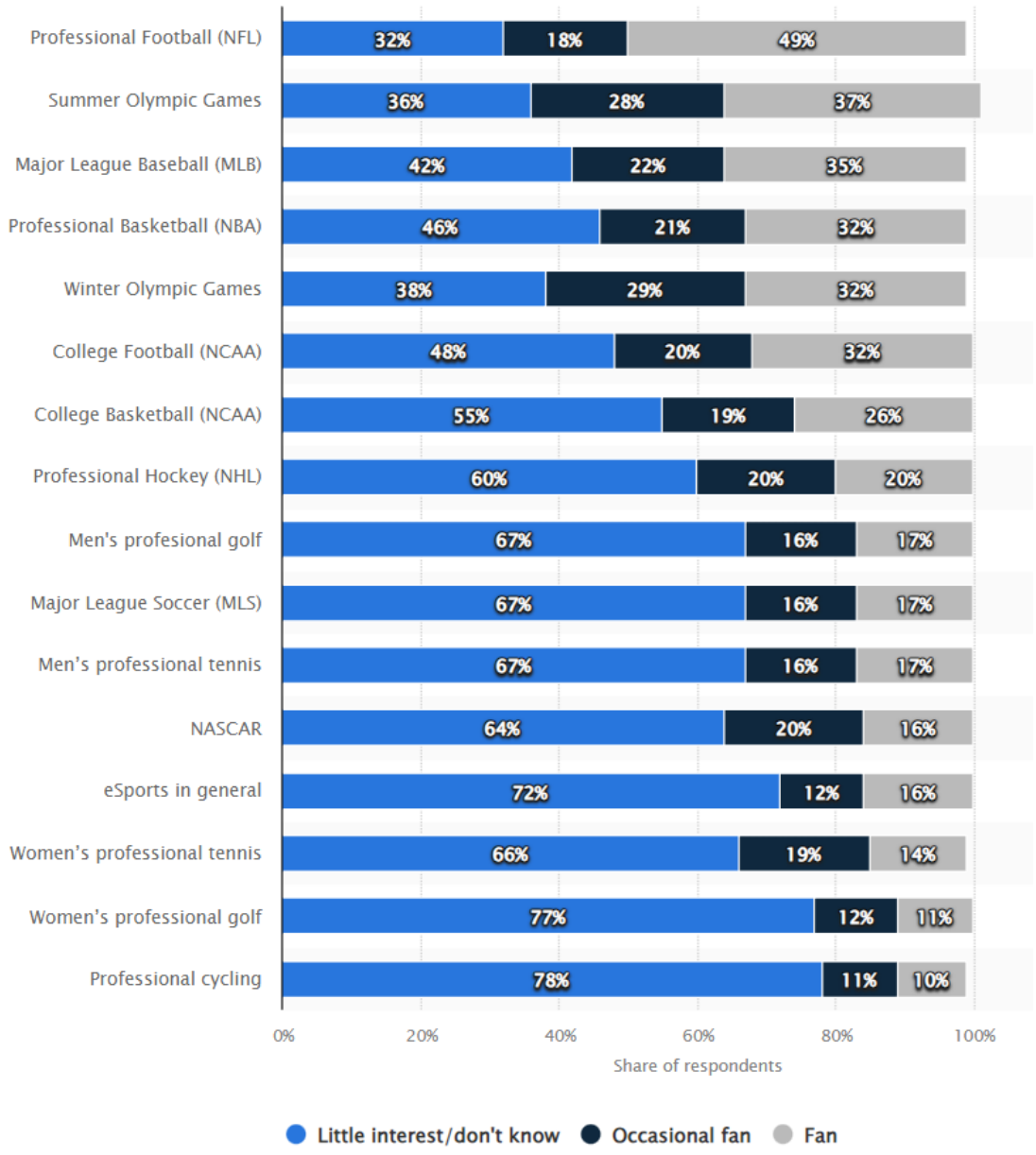
1.1.4 Sports Familiarity

Source: https://www.statista.com/statistics/198409/us-adults-favorite-sports/

This data was collected from a survey across American adults of familiarity/level of interest in different American sports.

As indicated below, while social media activity indicates favor for sports like basketball, there is a much greater amount of people who are partial or total fans of American football than any other sport.

[9]:
```
display(Image(filename='sports_popularity_graph.png'))
```



1.2 Platform Attributes

1.2.1 (Webscraping) NFL 2021 Average Price Dataset

[10]:
```
# Scraping data
url = 'https://www.statista.com/statistics/193595/
  ↪average-ticket-price-in-the-nfl-by-team/'
response = requests.get(url)
html_content = response.content
```

```
soup = BeautifulSoup(html_content, 'html.parser')

td_tags = soup.find_all('td')
numbers = [tag.text for tag in td_tags]
```

[11]:
```
# Separating team names and avg_price values
teams = numbers[::2]
avg_price = numbers[1::2]
```

[12]:
```
# Creating dataframe for average ticket price data
nfl_2021 = pd.DataFrame()
nfl_2021['Tm'], nfl_2021['avg_price'] = teams, avg_price

# Changing average price to float for plots
nfl_2021['avg_price'] = nfl_2021['avg_price'].astype(float)
```

[13]:
```
# Checking data
nfl_2021.head()
```

[13]:

	Tm	avg_price
0	Las Vegas Raiders	153.47
1	San Francisco 49ers	139.71
2	New England Patriots	131.45
3	Green Bay Packers	128.93
4	Philadelphia Eagles	127.06

1.2.2 NFL 2021 Attendance Dataset

Source: https://www.pro-football-reference.com/years/2021/attendance.htm

[14]:
```
nfl_2021_att = pd.read_csv('nfl_2021_attendance.csv')

# Removing empty/useless columns
nfl_2021_att = nfl_2021_att.drop(['Week 3-additional','Week 4-additional','Week␣
 ↪7-additional','Week 8-additional',
                  'Week 11-additional','Week 13-additional','Week␣
 ↪15-additional','Week 16-additional',
                  'Week 18-additional'], axis=1)

# Removing 'Bye' placeholder with nan
nfl_2021_att.replace('Bye', np.nan, inplace=True)
```

[15]:
```
# Checking data
nfl_2021_att.head()
```

[15]:

	Tm	Total	Home	Away	Week 1	Week 2	Week 3	\
0	Arizona Cardinals	1090961	500981.0	589980.0	67216	60115	58012	

```
1    Atlanta Falcons  1173982  540764.0  633218.0   68633   63694   75307
2   Baltimore Ravens  1141157  634840.0  506317.0   61756   70417   50788
3      Buffalo Bills  1093983  610352.0  483631.0   69787   65040   68434
4  Carolina Panthers  1202133  575249.0  626884.0   70211   70675   68320

   Week 4  Week 5 Week 6  … Week 9 Week 10 Week 11 Week 12 Week 13 Week 14  \
0   70003   62716  67431  …  59012   62400   68833     NaN   60803   63999
1   68333   60589    NaN  …  69170   93436   68286   60031   69366   72243
2   76490   70510  70704  …  70599   65948   60679   70355   59303   67431
3   68087   73389  69419  …  63110   77625   69866   69170   69694     NaN
4   93262   72545  72104  …  72203   62400   73350   62389     NaN   72243

   Week 15 Week 16  Week 17  Week 18
0  45211.0   63900    93459    63858
1  71650.0   68638    64948    68336
2  70815.0   63922    70328    70453
3  65398.0   65878    64948    64783
4  65398.0   71918    69170    65473

[5 rows x 22 columns]
```

1.2.3 Merging Datasets

[16]:
```
# Merging dataframes
nfl_2021 = pd.merge(nfl_2021_att,nfl_2021,on='Tm')
```

[17]:
```
# Checking data
nfl_2021.head()
```

[17]:
```
                 Tm    Total      Home      Away  Week 1  Week 2  Week 3  \
0  Arizona Cardinals  1090961  500981.0  589980.0   67216   60115   58012
1    Atlanta Falcons  1173982  540764.0  633218.0   68633   63694   75307
2   Baltimore Ravens  1141157  634840.0  506317.0   61756   70417   50788
3      Buffalo Bills  1093983  610352.0  483631.0   69787   65040   68434
4  Carolina Panthers  1202133  575249.0  626884.0   70211   70675   68320

   Week 4  Week 5 Week 6  … Week 10 Week 11 Week 12 Week 13 Week 14  \
0   70003   62716  67431  …   62400   68833     NaN   60803   63999
1   68333   60589    NaN  …   93436   68286   60031   69366   72243
2   76490   70510  70704  …   65948   60679   70355   59303   67431
3   68087   73389  69419  …   77625   69866   69170   69694     NaN
4   93262   72545  72104  …   62400   73350   62389     NaN   72243

   Week 15 Week 16 Week 17  Week 18  avg_price
0  45211.0   63900   93459    63858      84.83
1  71650.0   68638   64948    68336     105.57
2  70815.0   63922   70328    70453     110.53
```

```
3  65398.0   65878   64948    64783      82.49
4  65398.0   71918   69170    65473     114.67

[5 rows x 23 columns]
```

1.2.4 Average Weekly Attendance Subset

[18]:
```
# Creating dataframe for weekly attendance
week_att = pd.DataFrame()
week_att['week'], week_att['avg_att'] = nfl_2021.sum().index, nfl_2021.sum().
 ↪values

# Removing rows not related to weekly attendance
week_att = week_att.drop([0,1,2,3,13])

# Converting string to int
week_att['avg_att'] = week_att['avg_att'].astype(int)
```

```
/tmp/ipykernel_5606/487627312.py:3: FutureWarning: The default value of
numeric_only in DataFrame.sum is deprecated. In a future version, it will
default to False. In addition, specifying 'numeric_only=None' is deprecated.
Select only valid columns or specify the value of numeric_only to silence this
warning.
  week_att['week'], week_att['avg_att'] = nfl_2021.sum().index,
nfl_2021.sum().values
/tmp/ipykernel_5606/487627312.py:3: FutureWarning: The default value of
numeric_only in DataFrame.sum is deprecated. In a future version, it will
default to False. In addition, specifying 'numeric_only=None' is deprecated.
Select only valid columns or specify the value of numeric_only to silence this
warning.
  week_att['week'], week_att['avg_att'] = nfl_2021.sum().index,
nfl_2021.sum().values
```

[19]:
```
# Checking data
week_att.head()
```

[19]:
```
     week  avg_att
4  Week 1  1991107
5  Week 2  2060004
6  Week 3  2114340
7  Week 4  2180915
8  Week 5  2078557
```

1.3 EDA

1.3.1 Summary Statistics

[20]:
```
nfl_2021.describe()
```

[20]:

	Total	Home	Away	Week 1	Week 2
count	3.100000e+01	31.000000	31.000000	31.000000	31.000000
mean	1.136744e+06	572196.225806	564547.580645	64229.258065	66451.741935
std	6.714522e+04	67074.675823	43382.704534	9392.703768	6124.150357
min	1.027002e+06	412177.000000	483631.000000	35242.000000	50118.000000
25%	1.091767e+06	539582.000000	524643.500000	61756.000000	63076.000000
50%	1.118294e+06	562276.000000	559519.000000	67216.000000	67431.000000
75%	1.175798e+06	613727.000000	603758.000000	69999.000000	70328.500000
max	1.366866e+06	747368.000000	633218.000000	74119.000000	77240.000000

	Week 3	Week 4	Week 5	Week 15	Week 16
count	31.000000	31.000000	31.000000	29.000000	31.000000
mean	68204.516129	70352.096774	67050.225806	66299.931034	69004.000000
std	9609.928286	7453.706932	8950.069086	7953.942055	6074.842752
min	50788.000000	61468.000000	50137.000000	45211.000000	60972.000000
25%	62594.500000	66290.500000	60589.000000	60252.000000	65776.000000
50%	67914.000000	69170.000000	66538.000000	66391.000000	68492.000000
75%	73092.500000	70236.000000	70375.000000	71565.000000	69796.000000
max	93267.000000	93262.000000	93476.000000	77413.000000	93482.000000

	Week 17	Week 18	avg_price
count	31.000000	31.000000	31.000000
mean	69182.806452	66756.806452	106.949677
std	8255.027819	4837.705196	17.797936
min	51563.000000	56735.000000	80.380000
25%	64948.000000	64320.500000	94.090000
50%	69091.000000	66625.000000	104.500000
75%	70970.000000	69796.000000	116.585000
max	93459.000000	76012.000000	153.470000

1.3.2 Top 5 Teams with Greatest Attendance

The top 5 NFL teams in terms of attendance are as in order: - Dallas Cowboys - Denver Broncos - New York Giants - Carolina Panthers - Kansas City Chiefs

[21]:
```
nfl_2021.nlargest(5, 'Total')
```

[21]:

	Tm	Total	Home	Away	Week 1	Week 2	Week 3
8	Dallas Cowboys	1366866	747368.0	619498.0	65566	70240	93267
9	Denver Broncos	1243881	686129.0	557752.0	74119	58461	75882
23	New York Giants	1207885	591052.0	616833.0	74119	50118	75307
4	Carolina Panthers	1202133	575249.0	626884.0	70211	70675	68320

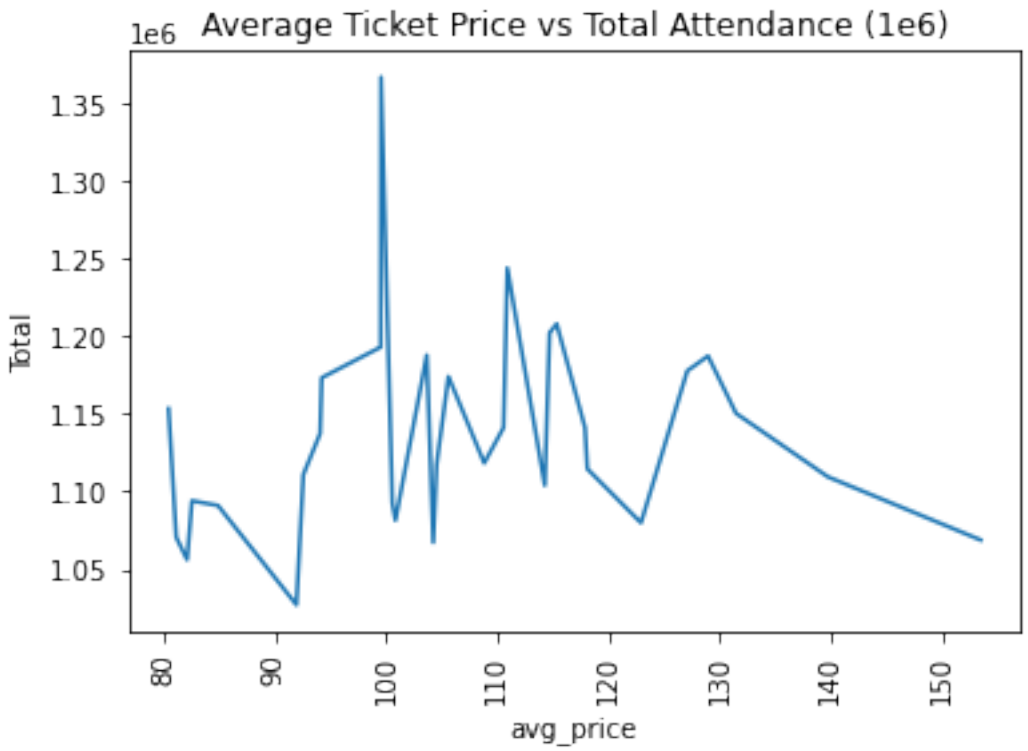
```
15  Kansas City Chiefs  1192945  659044.0  533901.0   72973   70417   72980

    Week 4  Week 5 Week 6  … Week 10 Week 11 Week 12 Week 13 Week 14  \
8    93262   93476  65878  …   93436   73494   93483   69170   61308
9    76490   59841  76854  …   76766     NaN   75807   73185   76211
23   69170   93476  73920  …     NaN   65643   73969   65898   70240
4    93262   72545  72104  …   62400   73350   62389     NaN   72243
15   69796   73389  51322  …   62125   73494     NaN   73185   73278

    Week 15 Week 16 Week 17  Week 18  avg_price
8   77413.0   93482   93459    69796      99.50
9   76134.0   60972   70240    76012     110.87
23  77413.0   69796   59594    69923     115.31
4   65398.0   71918   69170    65473     114.67
15  70240.0   73371   64505    76012      99.47

[5 rows x 23 columns]
```

1.3.3 Total Attendance vs. Average Price

Peak attendance occurs around an average ticket price of ~$100.00

[22]:
```
plt.xticks(rotation=90)
plt.title('Average Ticket Price vs Total Attendance (1e6)')
sns.lineplot(x='avg_price',y='Total',data=nfl_2021)
```

[22]:
```
<AxesSubplot:title={'center':'Average Ticket Price vs Total Attendance (1e6)'},
xlabel='avg_price', ylabel='Total'>
```

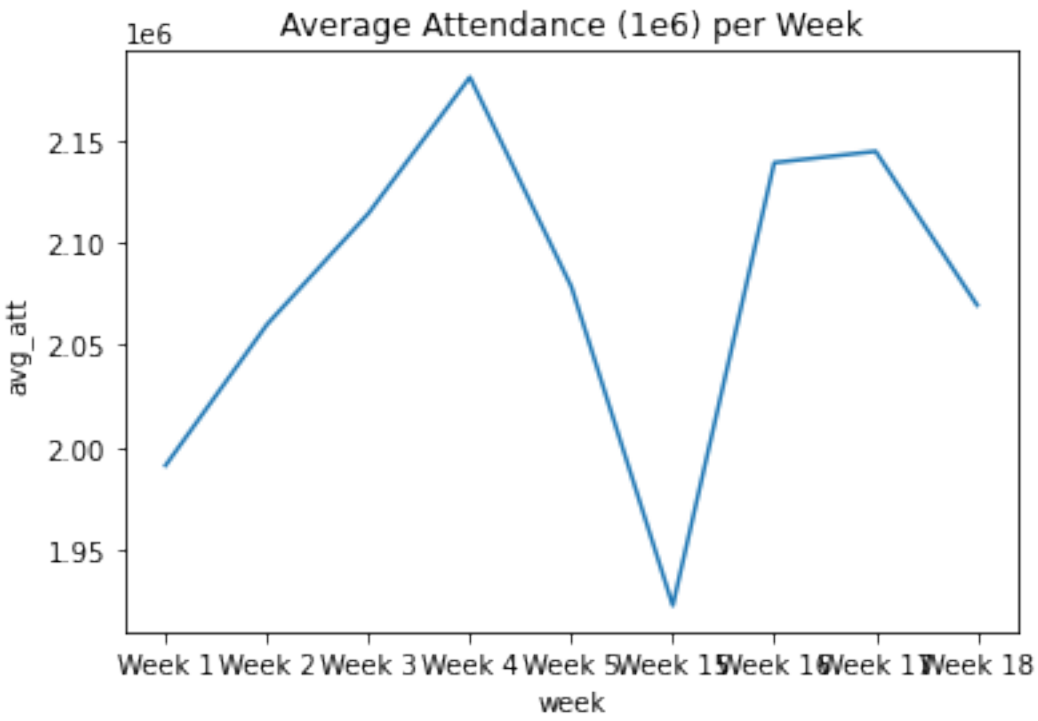


1.3.4 Average Attendance per Week

- Max Attendance: Week 3 (end of September 2021)
- Min Attendance: Week 5 (2nd week of October 2021)

[23]:
```
plt.title('Average Attendance (1e6) per Week')
sns.lineplot(x='week',y='avg_att',data=week_att)
# Week 4 is end of Sept
# Week 5 is week Oct 11, 2021
```

[23]:
```
<AxesSubplot:title={'center':'Average Attendance (1e6) per Week'},
xlabel='week', ylabel='avg_att'>
```

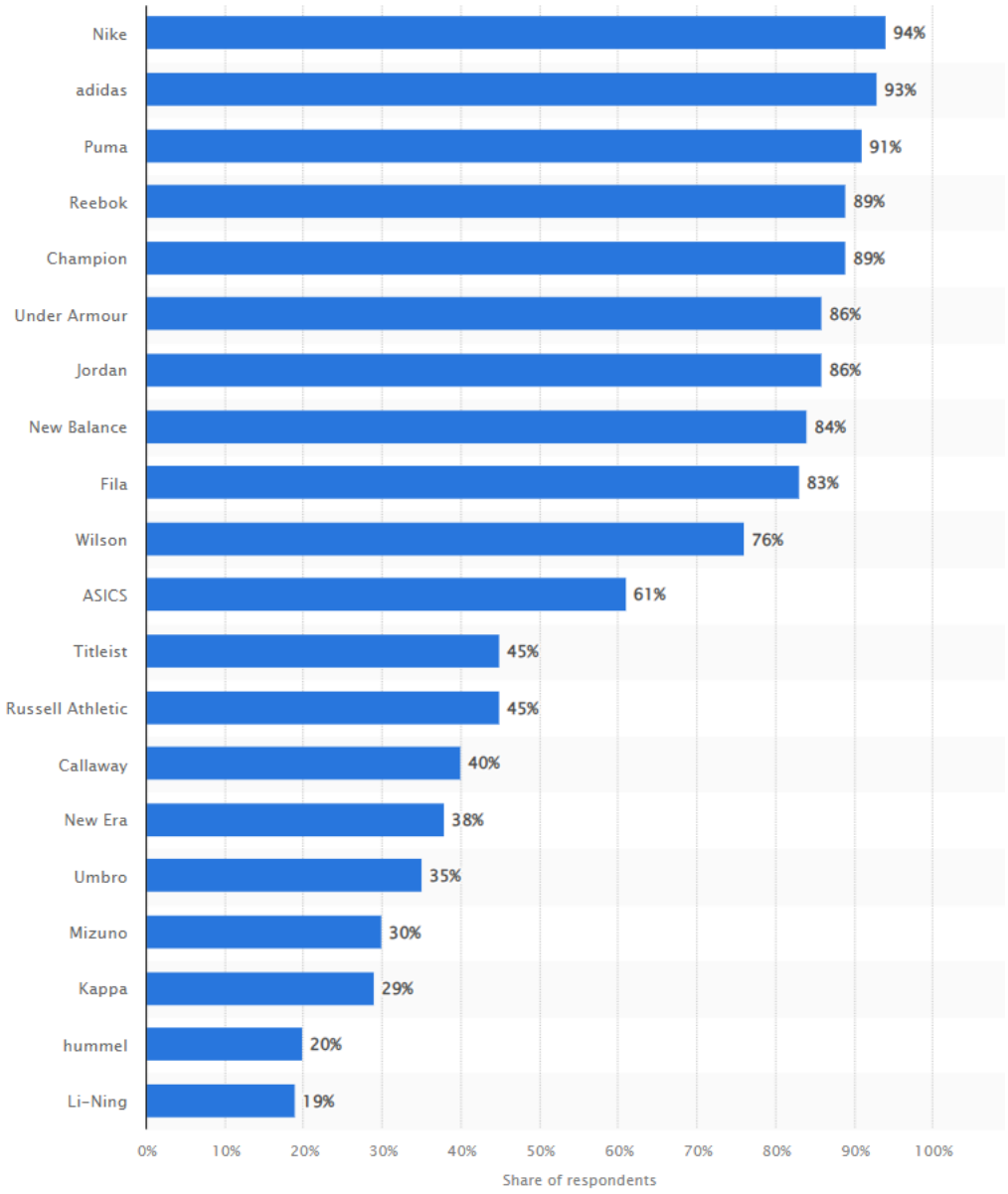


1.4 Top Sports Brands

Statista Survey: Most Well Known Brands Source: https://www.statista.com/statistics/1343534/most-well-known-sportswear-brands-in-the-united-states/

The top 5 most well known brands from this survey are: - Nike - Adidas - Puma - Reebok - Champion

[24]:
```
display(Image(filename='statista_knownbrands_graph.png'))
```

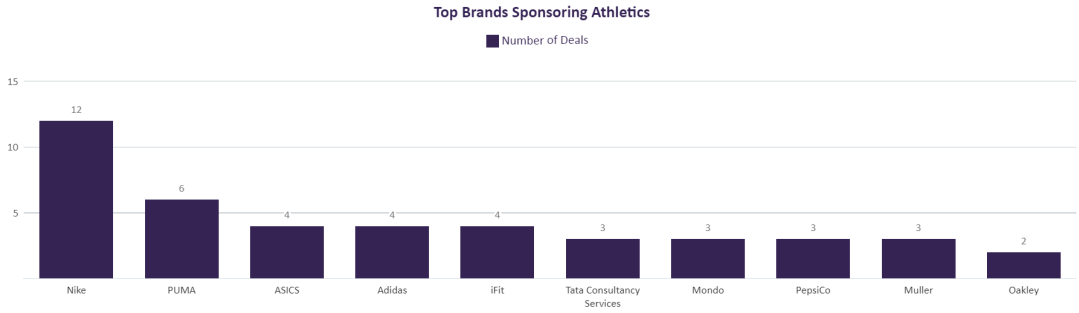


Top Sponsorship Brands

- Financial Brands: https://www.globaldata.com/data-insights/sport/most-active-financial-services-brands-sponsoring-american-football/
- General Brands in Athletics: https://www.globaldata.com/data-insights/sport/most-active-brands-sponsoring-athletics/

As of June 2022, the top 5 brands sponsoring athletics are various sports brands. In order, these top 5 brands are: - Nike - PUMA - ASICS - Adidas - iFit

[25]:
```
display(Image(filename='gd_topbrandsponsors_graph.png'))
```



Published: June 2022 • Source: GlobalData Sport Intelligence Center