

ADS-509 Final Project Team 11

```
#Install required libraries
!pip install beautifulsoup4 requ
!pip install tqdm
```

```
Requirement already satisfied: BeautifulSoup4 in c:\users\earne\anaconda3\lib\site-packages (4.9.3)
Requirement already satisfied: requests in c:\users\earne\anaconda3\lib\site-packages (2.31.0)
Requirement already satisfied: soupselect>1.2 in c:\users\earne\anaconda3\lib\site-packages (from BeautifulSoup4) (17.2.1)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\earne\anaconda3\lib\site-packages (from requests) (1.26.6)
Requirement already satisfied: certifi<=2017.4.17 in c:\users\earne\anaconda3\lib\site-packages (from requests) (2010.12.5)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\earne\anaconda3\lib\site-packages (from requests) (1.26.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\earne\anaconda3\lib\site-packages (from requests) (2.10)
Requirement already satisfied: tqdm in c:\users\earne\anaconda3\lib\site-packages (4.59.0)

# Import required libraries
import requests
from bs4 import BeautifulSoup
import pandas as pd
import matplotlib.pyplot as plt
from nltk.corpus import stopwords
from string import punctuation
from wordcloud import WordCloud
import re
from collections import Counter
from tqdm import tqdm

The first step in our code is to scrape blog post text from the mlbtradernews.com website. This posts include text, dates posted, and the authors name. We will scrape this information and store it in a dataframe that we can use to start our text analysis.
```

```
# Find all the news article elements
articles = soup.find_all_all("article")

# Create lists to store the extracted values
titles = []
authors = []
dates = []
contents = []

# Iterate over the articles and extract the text
with open(file_path, "w", encoding="utf-8") as file:
    for article in articles:
        # Extract the article title
        title_element = article.find("h2")
        title = title_element.text.strip() if title_element else ""
        titles.append(title)

        # Extract the article author
        author_element = article.find("span", class_="entry-author")
        author = author_element.text.strip() if author_element else ""
        authors.append(author)

        # Extract the article date
        date_element = article.find("time", class_="entry-time")
        date = date_element.text.strip() if date_element else ""
        dates.append(date)

        # Extract the article content
        content_element = article.find("div", class_="entry-content")
        content = content_element.text.strip() if content_element else ""
        contents.append(content)

        # Write the title and content to the file
        file.write("Title: " + title + "\n")
        file.write("Author: " + author + "\n")
        file.write("Date: " + date + "\n")
        file.write("Content: " + content + "\n")
        file.write("----\n")

# Create a DataFrame from the extracted values
data = {
    "Title": titles,
    "Author": authors,
    "Date": dates,
    "Content": contents
}

df = pd.DataFrame(data)

# Drop rows without content
df.dropna()

return df
```

```
# URL of the website to scrape
url = "https://www.mlbtoddermors.com/washington-nationals?show=all"
file_path = "C:/Users/easne/OneDrive - University of San Diego/MSADS/ADS-509/Project/data/mlbtoddermors.csv"

# Scrape the website, write the scraped text to a file, and get the head of the DataFrame
df = scrape_website(url, file_path)
df.head()
```

	Title	Author	Date	Content
0	Sean Doolittle Completes Rehab Assignment, Heads To Triple-A Rochester	Steve Adams	June 15, 2023	Veteran lefty Sean Doolittle was transferred from the Nationals' Double-A affiliate to their Tri-
1	Nationals Release Erasmo Ramirez	Darragh McDonald	June 9, 2023	The Nationals have released right-hander Erasmo Ramirez, according to his transactions tracker a..
2	Best Deadline Rental Returns In Recent History, #8: Nationals Get Everyday Outfielder For Jon Le..	Darragh McDonald	June 8, 2023	With the trade deadline now less than two months away, we at MLBTR are setting our sights backwa..
3	Nationals Claim Joe La Sorsa	Steve Adams	June 8, 2023	The Nationals have claimed left-hander Joe La Sorsa off waivers from the Rays, the team announce..
4	Nationals Outright Andres Machado	Anthony Franco	June 7, 2023	Nationals reliever Andrés Machado has gone unclaimed on waivers after being designated for assign..

```
# Change the Date column to a datetime object:
df['Date'] = pd.to_datetime(df['Date'], format="%B %d, %Y", errors='coerce')
```

```
df.dtypes
```

```
Title      object
Author     object
Date       datetime64[ns]
Content    object
dtype: object
```

```
# Check for missing values
df.isna().sum()
```

```
Title      0
Author     0
Date      30
Content   64
dtype: int64
```

```
len(df)
```

```
1000
```

```
df = df.dropna()
df
```

Title	Author	Date	Content
-------	--------	------	---------

		A Rochester	Adams	06-15	Double-A affiliate to their Tri...
1		Nationals Release Erasmo Ramirez	Darragh McDonald	2023-06-09	The Nationals have released right-hander Erasmo Ramirez, according to his transactions tracker a...
2	Best Deadline Returnals Retent in Recent History: #8: Nationals Get Everyday Outfielder For Jon Le...		Darragh McDonald	2023-06-08	With the trade deadline now less than two months away, we at MLBTR are setting our sights backwa...
3		Nationals Claim Joe La Sorsa	Steve Adams	2023-06-08	The Nationals have claimed left-hander Joe La Sorsa off waivers from the Rays, the team announce...
4		Nationals Outright Andres Machado	Anthony Franco	2023-06-07	Nationals reliever Andres Machado has gone unclaimed on waivers after being designated for assign...
...		...	...	...	...
995		Nationals Notes: Deadline, Scherzer, Zimmerman	Steve Adams	2019-06-28	The resurgent Nationals have won eight of their past 10 games and reached the .500 mark at 40-40...
996		Alex Meyer Retires	Connor Byrne	2019-06-25	Right-hander Alex Meyer has announced his retirement via Instagram, Mike DiGiovanna of the Los A...
997		Nationals Select Jonny Venters	Steve Adams	2019-06-25	3:25pm: To make room for Venters and Rodney, the Nationals moved injured right-hander Jeremy Hel...
998		Nationals To Select Fernando Rodney On Tuesday	Connor Byrne	2019-06-24	The Nationals will add right-handed reliever Fernando Rodney to their roster before Tuesday's ga...
999		Nationals Release Trevor Rosenthal	Connor Byrne	2019-06-23	The Nationals have released reliever Trevor Rosenthal, the team announced. The club recalled fel...

1000 rows x 4 columns

## Cleaning the Data

```

#Identify any noise in the data
RE_SUSPICIOUS = re.compile(r'{{<>|\\[\\]\\\\}}')

def impunity(text, min_len=10):
    """returns the share of suspicious characters in a text"""
    if text == None or len(text) < min_len:
        return 0
    else:
        return len(RE_SUSPICIOUS.findall(text))/len(text)

df['Content'].apply

#bound method Series.apply of 0
Veteran lefty Sean Doolittle was transferred from the Nationals' Double-A affiliate to their Tri...
2 With the trade deadline now less than two months away, we at MLBTR are setting our sights backwa...
3 The Nationals have claimed left-hander Joe La Sorsa off waivers from the Rays, the team announce...
4 Nationals reliever Andres Machado has gone unclaimed on waivers after being designated for assign...

995 The resurgent Nationals have won eight of their past 10 games and reached the .500 mark at 40-40...
996 Right-hander Alex Meyer has announced his retirement via Instagram, Mike DiGiovanna of the Los A...
997 3:25pm: To make room for Venters and Rodney, the Nationals moved injured right-hander Jeremy Hel...
998 The Nationals will add right-handed reliever Fernando Rodney to their roster before Tuesday's ga...
999 The Nationals have released reliever Trevor Rosenthal, the team announced. The club recalled fa...
Name: Content, Length: 1000, dtype: object>

pd.options.display.max_colwidth = 100 ###
# add new column to df&#223; frame
df['impunity'] = df['Content'].apply(impunity, min_len=10)

# get the top 3 records
df[['Content', 'impunity']].sort_values(by='impunity', ascending=False).head(3)

```

	Content	impunity
601	The Nationals announced they've selected the contract of right-hander Steven Fuentes. The 23-yr...	0.00333
559	Kyle Schwarber is now a member of the Washington Nationals, in no small part because of Nats nam...	0.00355
565	The Nationals accomplished a big part of their winter to-do list when they acquired Josh Bell fr...	0.00276

The above shows the highest impunity levels for this data set. Note, these are extremely low...well below 1% of all characters in each blog post are suspicious.

```
import re
```

```
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer

# Example function to normalize text
```

```
# Lowercase the text
text = text.lower()

# Remove punctuation
text = re.sub(r'[%s]' % string.punctuation, "", text)

# Remove special characters and digits
text = re.sub(r'[%s]' % string.punctuation, "", text)

# Tokenize the text
tokens = word_tokenize(text)

# Remove stopwords
stop_words = set(stopwords.words("english"))
tokens = [token for token in tokens if token not in stop_words]

# Perform stemming
stemmer = PorterStemmer()
tokens = [stemmer.stem(token) for token in tokens]

return tokens

# Apply normalization to the "Contents" column
df["Contents_Normalized"] = df["Contents"].apply(normalize_text)

# Print the head of the DataFrame with the normalized contents
df[["Content", "Contents_Normalized"]].head()
```

				setting our sights backward...			highlight, past, trad...
2	The Nationals have claimed left-hander Joe La Sorsa off waivers from the Rays, the team announce...						(nation, claim, lefthand, joe, la, sorsa, waiver, ray, team, announc, thursday, la, sorsa, design...
3	Nationalis reliever Andrés Machado has gone unclaimed on waivers after being designated for assign...						(nation, reliev, andr, machado, gone, unclaim, waiver, design, assign, week, club, announc, mach...
<pre>df['length'] = df['Contents_Normalized'].str.len() df.head()</pre>							
	Title	Author	Date	Content	impurity	Contents_Normalized	length
0	Sean Doolittle Completes Rehab Assignment, Heads To Triple-A Rochester	Steve Adams	2023-06-15	Veteran lefty Sean Doolittle was transferred from the Nationals' Double-A affiliate to their Tri...	0.000000	(jeter, left, sean, doolitt, transfer, nation, doubles, affil, triple, club, thursday, sig...	60
1	Nationals Release Erasmo Ramirez	Darragh McDonald	2023-06-09	The Nationals have released right-hander Erasmo Ramirez, according to his transactions tracker a...	0.000000	(nation, releas, righthand, erasmo, ram, res, accord, transact, track, milbcom, design, assign...	130
	Best Deadline Return Returns In Recent History; #8: National Get Everyday Outfielder For Jon Le...	Darragh McDonald	2023-06-08	With the trade deadline now less than two months away, we at MLBTR are setting our sights backward...	0.000441	(trade, deadlin, less, two, month, away, milb, set, sight, backward, sig, highlight, past, trad...	607
3	Nationals Claim Joe La Sorsa	Steve Adams	2023-06-08	The Nationals have claimed left-hander Joe La Sorsa off waivers from the Rays, the team announce...	0.000000	(nation, claim, lefthand, joe, la, sorsa, waiver, ray, team, announc, thursday, la, sorsa, design...	158
		Anthony	2023-	Nationals Reliever Andrés Machado		(nation, reliev, andr, machado, none, unclaim, waiver, design...	

```
df['length'].plot(kind='box', vert=False, figsize=(8, 1))
```


<AxesSubplot>

```
df['Title_length'] = df['Title'].str.len()
df['Title_length'].plot(kind='box', vert=False, figsize=(8, 1))
```

<AxesSubplot>

```
df['length'].plot(kind='hist', bins=30, figsize=(8,2))
```

<AxesSubplot:ylabel='Frequency'>



Length Bin (approx.)	Frequency
10-12	120
12-14	220
14-16	210
16-18	150
18-20	100

```
# Extract the month from the date and create a new column
df['Month'] = df['Date'].dt.month

# Plot the average post length
df.groupby('Month').agg({'length': 'mean'}) \
    .plot(title='Avg. Post Length', ylim=(0,500), figsize=(6,2))
```

Month	Avg. Post Length (characters)
1	300
2	310
3	305
4	300
5	295
6	290
7	280
8	220
9	200
10	230
11	230
12	240

```
def count_words(df, column='Contents_Normalized', preprocess=None, min_freq=2):  
    # process tokens and update counter  
    def update(doc):  
        tokens = doc if preprocess is None else preprocess(doc)  
        counter.update(tokens)  
  
    # create counter and run through all data  
    counter = Counter()  
    tqdm.pandas() # initialize tqdm for progress bar  
    df[column].apply(update)
```

```
# transform counter into data frame
freq_df = pd.DataFrame.from_dict(counter, orient='index', columns=['freq'])
freq_df = freq_df.query('freq >= @min_freq')
freq_df.index.name = 'token'

return freq_df.sort_values('freq', ascending=False)
```

	freq
token	
season	3102
nation	2758
year	2411
leagu	1804
nat	1609

```
#How many tokens are in the df?
len(freq_df)

6440

# top words with 10+ characters
count_words(df, column="Content",
             preprocess=lambda Content: re.findall(r"\\w{10,}", Content)).head(5)
```

	freq
token	
Washington	1259
appearances	618
organization	384
outfielder	258
exhibition	214

```
ax = freq_df.head(15).plot(kind='barh', width=0.95, figsize=(8,3))
ax.invert_yaxis()
ax.set(xlabel='Frequency', ylabel='Token', title='Top Words')

[Text(0.5, 0, 'Frequency'), Text(0, 0.5, 'Token'), Text(0.5, 1.0, 'Top Words')]
```

Token	Frequency
season	~10
nation	~8
star	~7

A horizontal bar chart titled "Words" on the y-axis and "Frequency" on the x-axis. The x-axis has major tick marks at 0, 500, 1000, 1500, 2000, 2500, and 3000. There are ten blue bars representing different words. From top to bottom, the words and their approximate frequencies are: hit (~1700), run (~1600), team (~1500), player (~1400), game (~1300), win (~1250), washington (~1200), last (~1150), deal (~1100), and injury (~1050). A legend indicates that the blue bars represent "freq".

Word	Frequency
hit	~1700
run	~1600
team	~1500
player	~1400
game	~1300
win	~1250
washington	~1200
last	~1150
deal	~1100
injury	~1050

Import data on the washington nationals from baseball-reference.com

```
1 https://www.baseball-reference.com/teams/WSN/attend.shtml

import pandas as pd

# URL of the webpage to scrape
url = "https://www.baseball-reference.com/teams/WSN/attend.shtml"

# Read the HTML table into a list of DataFrames
tables = pd.read_html(url)

# Print the first DataFrame (the table with the most data)
print(tables[0])
```

```
table = tables[0]

# Save the DataFrame as a CSV file
file_path = "C:/Users/earne/OneDrive - University of San Diego/MSADS/ADS-509/Project/data/attendance_data.csv"
table.to_csv(file_path, index=False, header=True)

print("Data saved to", file_path)

Data saved to C:/Users/earne/OneDrive - University of San Diego/MSADS/ADS-509/Project/data/attendance_data.csv

path = "C:/Users/earne/OneDrive - University of San Diego/MSADS/ADS-509/Project/data/"
```

data_df = pd.DataFrame(pd.read_csv(path + "attendance_data.csv"))														
data_df.head()														
	Year	Team	Lg	W	L	Finish	Playoffs	Attendance	Attend/G	Rank	Est. Payroll	PPF	BPF	Stadium
0	2023	Washington Nationals	NL East	25	36	5	NaN	709171.0	21490.0	11th of 15	\$98,370,676	96	94	National's Park
1	2022	Washington Nationals	NL East	25	107	5	NaN	2026401.0	25017.0	11th of 15	\$125,051,666	96	94	National's Park
		Washington Nationals	NL East							12th of 15				National's Park

	Nations	East													Park
3	2020	Washington Nationals	NL East	26	34	5	NaN	NaN	NaN	15th of 15	\$172,237,283	99	99		Nationals Park
4	2019	Washington Nationals	NL East	93	69	2	Won WS (4- 3)	2259781.0	27899.0	11th of 15	\$203,016,595	102	102		Nationals Park

```
nats_df=nats_df[nats_df['Year'] > 2004]
nats_df
```

Year	Team	LG	W	L	Finish	Playoffs	Attendance	Avg. Attend./Game	Rank	Est. Payroll	PPF	BPF	Stadium
0	Washington Nationals	NL East	25	36	5	Na/N	291,771.0	214,900	11th of 15	\$98,375,066	96	94	Nationals Park
1	Washington Nationals	NL East	55	107	5	Na/N	202,640.1	250,170	11th of 15	\$125,051,666	96	94	Nationals Park
2	Washington Nationals	NL East	65	97	5	Na/N	146,554.0	180,930	12th of 15	\$172,370,147	96	95	Nationals Park
3	Washington Nationals	NL East	26	34	5	Na/N	Na/N	Na/N	15th of 15	\$172,237,283	99	99	Nationals Park

4	2019	Washington Nationals	NL East	93	69	2	Won WS (4-3)	22,597,810	278,990	11th of 15	\$203,016,595	102	102	Nationals Park
5	2018	Washington Nationals	NL East	82	80	2	Na/N	25,296,040	312,200	8th of 15	\$188,886,699	105	106	Nationals Park
6	2017	Washington Nationals	NL East	97	65	1	Lost NLDS (3-2)	25,498,800	311,730	7th of 15	\$175,587,301	103	104	Nationals Park
7	2016	Washington Nationals	NL East	95	67	1	Lost NLDS (3-2)	24,819,380	306,410	7th of 15	\$152,967,400	102	103	Nationals Park

9	2014	Washington Nationals	NL East	96	66	1	Lost NLDS (3-1)	257,939.80	318,440	7th of 15	\$137,235,008	102	104	Nationals Park
10	2013	Washington Nationals	NL East	86	76	2	NaN	265,242.20	327,460	6th of 15	\$112,493,250	101	103	Nationals Park
11	2012	Washington Nationals	NL East	98	64	1	Lost NLDS (3-2)	237,079.40	292,690	9th of 16	\$92,386,000	101	101	Nationals Park
12	2011	Washington Nationals	NL East	80	81	3	NaN	194,047.80	242,560	14th of 16	\$68,492,928	100	100	Nationals Park

13	2010	Washington Nationals	NL East	69	93	5	Na/N	1828066.0	22569.0	14th of 16	\$67,701,000	100	99	Nationals Park
14	2009	Washington Nationals	NL East	59	103	5	Na/N	1817226.0	22435.0	13th of 16	\$64,384,000	101	99	Nationals Park
15	2008	Washington Nationals	NL East	59	102	5	Na/N	2320400.0	29005.0	13th of 16	\$54,961,000	100	98	Nationals Park
16	2007	Washington Nationals	NL East	73	89	4	Na/N	1943812.0	23998.0	14th of 16	\$36,947,500	95	94	Robert F. Kennedy Stadium

17	2006	Washington Nationals	NL East	71	91	5	Na/N	2133056.0	26580.0	11th of 16	\$63,143,000	95	94	Robert F. Kennedy Stadium
18	2005	Washington Nationals	NL East	81	81	5	Na/N	2731993.0	33728.0	8th of 16	\$48,581,500	96	95	Robert F. Kennedy Stadium

```
#remove the dollar sign from Est. Payroll column
mats_df['Est. Payroll'] = mats_df['Est. Payroll'].replace({'$':''}, regex = True)
mats_df['Est. Payroll'] = mats_df['Est. Payroll'].replace({'$':''}, regex = True)
```

data_or.head()														
Year		Tm	Lg	W	L	Finish	Playoffs	Attendance	Attend/G	Rank	Est. Payroll	PPF	BPF	Stadium
0	2023	Washington Nationals	NL East	25	36	5	Na/N	709171.0	21490.0	11th of 15	98375607	96	94	Nationals Park
1	2022	Washington Nationals	NL East	55	107	5	Na/N	2026401.0	15001.0	11th of 15	125051666	96	94	Nationals Park
2	2021	Washington Nationals	NL East	57	103	5	Na/N	1465654.0	15003.0	12th of 15	17337034.3	96	94	Nationals

3	2020	Washington Nationals	NL East	26	34	5	NaN	NaN	NaN	15th of 15	172237283	99	99	National's Park
4	2019	Washington Nationals	NL East	93	69	2	Won WS (4-3)	2259781.0	27899.0	11th of 15	203016595	102	102	National's Park

```
nats_df[['Est. Payroll']] = nats_df[['Est. Payroll']].astype(float)
nats_df.dtypes
```

Tm	object
Lg	object
W	int64
L	int64
Finish	int64
Playoffs	object
Attendance	float64
Attend/G	float64
Rank	object
Est. Payroll	float64
PFF	int64
RFF	int64

```
stats_df.describe()
```

	Year	W	L	Finish	Attendance	Attend/G	Est. Payroll	PPF	BPF
count	19.000000	19.000000	19.000000	19.000000	1.800000e+01	18.000000	1.900000e+01	19.000000	19.000000
mean	2014.000000	73.315789	77.894737	3.368421	2.164161e+06	27462.111111	1.163850e+08	99.526316	99.263158
std	5.627314	21.587169	20.215216	1.738790	5.051465e+05	4574.987970	5466072.07e+07	2.988281	3.983886

	min	2005.000000	25.000000	34.000000	1.000000	7.091710e+05	18093.000000	3.694750e+07	95.000000	94.000000
<b>25%</b>	2009.500000	62.000000	62.000000	66.500000	2.000000	1.941312e+06	24062.500000	6.604250e+07	96.000000	95.000000
<b>50%</b>	2014.000000	80.000000	80.000000	84.000000	5.000000	2.290909e+06	28452.500000	1.124932e+08	100.000000	99.000000
<b>75%</b>	2018.500000	89.500000	92.000000	92.000000	5.000000	2.528448e+06	31215.750000	1.273037e+08	101.500000	102.500000
<b>max</b>	2023.000000	98.000000	107.000000	5.000000	5.000000	2.731993e+06	33728.000000	2.030166e+08	105.000000	106.000000

```
plt.figure(figsize=(10, 6))

# Plotting Attendance over the years
plt.plot(mats_df['Year'], mats_df['Attendance'], marker='o')
plt.xlabel('Year')
plt.ylabel('Attendance')
plt.title('Attendance of Washington Nationals over the years')

Text(0.5, 1.0, 'Attendance of Washington Nationals over the years')
```

Taxon	Abundance
1	265
2	215
3	195
4	230
5	185
6	185
7	235
8	265
9	255
10	255
11	245
12	250
13	250
14	225
15	155
16	200
17	155

```


# A plotting wine over the years

```

```
plt.plot(nats_df['Year'], nats_df['W'], marker='o')
plt.xlabel('Year')
plt.ylabel('Wins')
plt.title('Wins of Washington Nationals over the years')

Text(0.5, 1.0, 'Wins of Washington Nationals over the years')
```

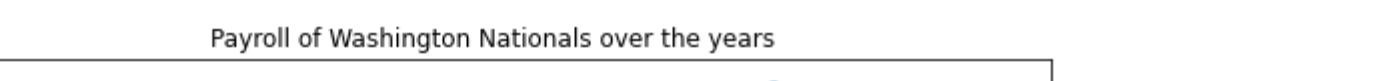
Month	Visits
Jan	82
Feb	71
Mar	73
Apr	59
May	59
Jun	69
Jul	79
Aug	85
Sep	82
Oct	85
Nov	80
Dec	52



```
# Plotting Payroll over the years
plt.figure(figsize=(10, 6))
plt.plot(nats df['Year'], nats df['Est. Payroll'], markers='o')
```

```
plt.ylabel('Payroll')
plt.title('Payroll of Washington Nationals over the years')

# Text(0.5, 1.0, 'Payroll of Washington Nationals over the years')



| Year | Payroll (Millions) |
|------|--------------------|
| 2000 | 175.0              |
| 2001 | 175.0              |
| 2002 | 175.0              |
| 2003 | 175.0              |
| 2004 | 175.0              |
| 2005 | 175.0              |
| 2006 | 175.0              |
| 2007 | 175.0              |
| 2008 | 175.0              |
| 2009 | 175.0              |
| 2010 | 175.0              |
| 2011 | 175.0              |
| 2012 | 175.0              |
| 2013 | 175.0              |
| 2014 | 175.0              |
| 2015 | 175.0              |
| 2016 | 175.0              |
| 2017 | 175.0              |
| 2018 | 175.0              |


```

Age	Percent
18-24	0.75
25-34	0.85
35-44	0.80
45-54	1.50
55-64	1.25
65+	1.00

```
# Correlation matrix
correlation = nats_df.corr()
print(correlation)

# Heatmap of correlation matrix
```

```
plt.imshow(corrrelation, cmap="coolwarm", interpolation='none')
plt.colorbar()
plt.xticks(range(len(corrrelation)), correlation.columns, rotation=90)
plt.yticks(range(len(corrrelation)), correlation.columns)
plt.title("Correlation Matrix")
```

	Est.	Payroll	BFF	BFF
Est. Payroll	0.772917	0.173634	0.374900	0.491810
PPF	0.146666	0.520898	-0.180205	-0.734371
BFF	0.133642	0.605363	-0.261083	-0.830812
				0.596436
Year	Est.	Payroll	BFF	BFF
L	0.772917	0.148668	0.153642	
W	0.173634	0.520898	0.605363	
N	-0.374900	-0.180205	-0.261083	
Finish	-0.491810	-0.734371	-0.830812	
Attendance	0.209334	0.508759	0.596436	
PPF	0.232467	0.505249	0.601454	
Est. Payroll	1.000000	0.510400	0.561694	

BPFF 0.561694 0.967702 1.000000  
: Text (0.5, 1.0, 'Correlation Matrix')

	BPFF	W	Text
BPFF	1.0	0.561694	0.967702
W	0.561694	1.0	0.967702
Text	0.967702	0.967702	1.0

Heatmap showing the correlation matrix for variables: Finish, Attendance, Attend/G, and Est. Revenue. The color scale ranges from -0.25 (blue) to 0.25 (red).

Heatmap showing the correlation between PPF and BPF across various variables. The color scale ranges from -0.75 (blue) to 0.50 (red).

Variable	PPF	BPF
Year	0.50	0.50
W	0.25	0.25
L	-0.25	-0.25
Finish	-0.50	-0.50
Attendance	0.25	0.25
Attendance	0.25	0.25
Est. Rayleigh	0.25	0.25
PPF	0.50	0.50
BPF	0.50	0.50