

```
In [1]: #Install reqired Libraries  
!pip install beautifulsoup4 requests  
!pip install pandas  
!pip install matplotlib  
!pip install lxml  
!pip install nltk  
!pip install wordcloud
```

Requirement already satisfied: beautifulsoup4 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (4.12.2)

Requirement already satisfied: requests in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (2.29.0)

Requirement already satisfied: soupsieve>1.2 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from beautifulsoup4) (2.4.1)

Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from requests) (2.0.4)

Requirement already satisfied: idna<4,>=2.5 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from requests) (3.4)

Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from requests) (1.26.16)

Requirement already satisfied: certifi>=2017.4.17 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from requests) (2023.5.7)

Requirement already satisfied: pandas in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (1.4.1)

Requirement already satisfied: python-dateutil>=2.8.1 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from pandas) (2.8.2)

Requirement already satisfied: pytz>=2020.1 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from pandas) (2022.7)

Requirement already satisfied: numpy>=1.18.5 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from pandas) (1.21.5)

Requirement already satisfied: six>=1.5 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from python-dateutil>=2.8.1->pandas) (1.16.0)

Requirement already satisfied: matplotlib in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (3.7.1)

Requirement already satisfied: contourpy>=1.0.1 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib) (1.0.5)

Requirement already satisfied: cyclor>=0.10 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib) (0.11.0)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib) (4.25.0)

Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib) (1.4.4)

Requirement already satisfied: numpy>=1.20 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib) (1.21.5)

Requirement already satisfied: packaging>=20.0 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib) (23.0)

Requirement already satisfied: pillow>=6.2.0 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib) (9.4.0)

Requirement already satisfied: pyparsing>=2.3.1 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib) (3.0.9)

Requirement already satisfied: python-dateutil>=2.7 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib) (2.8.2)

Requirement already satisfied: importlib-resources>=3.2.0 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib) (5.2.0)

Requirement already satisfied: zipp>=3.1.0 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from importlib-resources>=3.2.0->matplotlib) (3.11.0)

Requirement already satisfied: six>=1.5 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from python-dateutil>=2.7->matplotlib) (1.16.0)

Requirement already satisfied: lxml in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (4.9.2)

Requirement already satisfied: nltk in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (3.8.1)

Requirement already satisfied: click in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from nltk) (8.0.4)

Requirement already satisfied: joblib in c:\users\grigor\appdata\roaming\python\python38\site-packages (from nltk) (1.2.0)

Requirement already satisfied: regex>=2021.8.3 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from nltk) (2022.7.9)

Requirement already satisfied: tqdm in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from nltk) (4.65.0)
Requirement already satisfied: colorama in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from click->nltk) (0.4.6)
Requirement already satisfied: wordcloud in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (1.9.2)
Requirement already satisfied: numpy>=1.6.1 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from wordcloud) (1.21.5)
Requirement already satisfied: pillow in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from wordcloud) (9.4.0)
Requirement already satisfied: matplotlib in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from wordcloud) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib->wordcloud) (1.0.5)
Requirement already satisfied: cyclor>=0.10 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib->wordcloud) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib->wordcloud) (23.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: importlib-resources>=3.2.0 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from matplotlib->wordcloud) (5.2.0)
Requirement already satisfied: zipp>=3.1.0 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from importlib-resources>=3.2.0->matplotlib->wordcloud) (3.11.0)
Requirement already satisfied: six>=1.5 in c:\users\grigor\.conda\envs\ads_502\lib\site-packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)

```
In [2]: #Import required libraries
import requests
from bs4 import BeautifulSoup
import pandas as pd
import matplotlib.pyplot as plt
import lxml

import os
import re
import numpy as np
import string
import glob
import nltk
from wordcloud import WordCloud
from collections import Counter, defaultdict
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from string import punctuation

sw = stopwords.words("english")
```

```
C:\Users\Grigor\.conda\envs\ADS_502\lib\site-packages\numpy\_distributor_init.py:30:
UserWarning: loaded more than 1 DLL from .libs:
C:\Users\Grigor\.conda\envs\ADS_502\lib\site-packages\numpy\.libs\libopenblas.XWYDX2I
KJW2NMTWSFYNGFUWKQU3LYTCZ.gfortran-win_amd64.dll
C:\Users\Grigor\.conda\envs\ADS_502\lib\site-packages\numpy\.libs\libopenblas64__v0.
3.21-gcc_10_3_0.dll
warnings.warn("loaded more than 1 DLL from .libs:")
```

```
In [3]: def scrape_website(url, file_path):
# Send a GET request to the website
response = requests.get(url)

# Parse the HTML content using BeautifulSoup
soup = BeautifulSoup(response.content, "html.parser")

# Find all the news article elements
articles = soup.find_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)  
return df
```

```
Angels_url = "https://www.mlbtraderumors.com/los-angeles-angels"
response = requests.get(Angels_url)
# Parse the HTML content using BeautifulSoup
Beaut_soup = BeautifulSoup(response.content, "html.parser")
article = Beaut_soup.find_all("article")
# Create a file to save the scraped data
file_path = "C:/Users/Grigor/OneDrive/Desktop/Master_Program/ADS509/Project/angels.txt"
```

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

| | Title | Author | Date | Content |
|---|---|------------------|---------------|---|
| 0 | Angels To Select Kevin Padlo | Anthony Franco | June 19, 2023 | The Angels are selecting the contract of corne... |
| 1 | Angels Place Anthony Rendon On Injured List | Darragh McDonald | June 19, 2023 | The Angels announced that infielder Anthony Re... |
| 2 | Angels Place Gio Urshela On Injured List With ... | Anthony Franco | June 16, 2023 | The Angels announced they've placed infielder ... |
| 3 | Angels Place Zach Neto On Injured List, Recall... | Anthony Franco | June 15, 2023 | The Angels put rookie shortstop Zach Neto on t... |
| 4 | Angels Sign Daniel Murphy To Minor League Deal | Darragh McDonald | June 12, 2023 | Infielder Daniel Murphy, who had been playing ... |

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

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

def impurity(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
```

```

Out[9]: <bound method Series.apply of 0      The Angels are selecting the contract of corne...
1      The Angels announced that infielder Anthony Re...
2      The Angels announced they've placed infielder ...
3      The Angels put rookie shortstop Zach Neto on t...
4      Infielder Daniel Murphy, who had been playing ...
5      The Angels announced that they have selected t...
6      The Angels have been without starter José Suar...
7      Last week, MLBTR took an early look at offseas...
8      The calendar has flipped to June, and more tha...
9      Recently retired slugger Albert Pujols has bee...
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
Name: Content, dtype: object>

```

```

In [10]: pd.options.display.max_colwidth = 100 ###
# add new column to data frame
df['impurity'] = df['Content'].apply(impurity, min_len=10)

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

```

```

Out[10]:

```

| | Content | impurity |
|----|---|----------|
| 6 | The Angels have been without starter José Suarez for a month on account of a strain in his throw... | 0.001748 |
| 0 | The Angels are selecting the contract of corner infielder Kevin Padlo, reports Robert Murray of ... | 0.000000 |
| 30 | | 0.000000 |

```
In [11]: # Example function to normalize the text in the "Contents" column
def normalize_text(text):
    # Lowercase the text
    text = text.lower()

    # Remove punctuation
    text = text.translate(str.maketrans("", "", string.punctuation))

    # Remove special characters and digits
    text = re.sub(r"^[^a-zA-Z]", " ", 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["Content"].apply(normalize_text)

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

```
Out[11]:
```

| | Content | Contents_Normalized |
|---|---|---|
| 0 | The Angels are selecting the contract of corner infielder Kevin Padlo, reports Robert Murray of ... | [angel, select, contract, corner, infield, kevin, padlo, report, robert, murray, fansid, twitter... |
| 1 | The Angels announced that infielder Anthony Rendon has been placed on the 10-day injured list du... | [angel, announc, infield, anthoni, rendon, place, day, injur, list, due, left, wrist, contus, co... |
| 2 | The Angels announced they've placed infielder Gio Urshela on the 10-day injured list with a frac... | [angel, announc, place, infield, gio, urshela, day, injur, list, fractur, left, pelvi, michael, ... |
| 3 | The Angels put rookie shortstop Zach Neto on the 10-day injured list due to a left oblique strai... | [angel, put, rooki, shortstop, zach, neto, day, injur, list, due, left, obliqu, strain, even, in... |
| 4 | Infielder Daniel Murphy, who had been playing for the Long Island Ducks of the independent Atlan... | [infield, daniel, murphi, play, long, island, duck, independ, atlant, leagu, contract, purchas, ... |

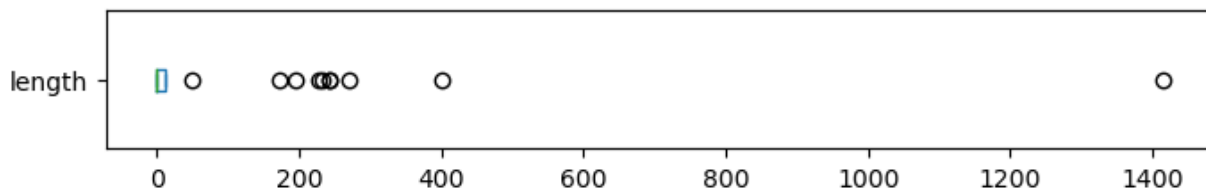
```
In [12]: df['length'] = df['Contents_Normalized'].str.len()
df.head()
```

Out[12]:

| | Title | Author | Date | Content | impurity | Contents_Normalized | length |
|---|---|------------------|------------|---|----------|---|--------|
| 0 | Angels To Select Kevin Padlo | Anthony Franco | 2023-06-19 | The Angels are selecting the contract of corner infielder Kevin Padlo, reports Robert Murray of ... | 0.0 | [angel, select, contract, corner, infield, kevin, padlo, report, robert, murray, fansid, twitter... | 196 |
| 1 | Angels Place Anthony Rendon On Injured List | Darragh McDonald | 2023-06-19 | The Angels announced that infielder Anthony Rendon has been placed on the 10-day injured list du... | 0.0 | [angel, announc, infield, anthoni, rendon, place, day, injur, list, due, left, wrist, contus, co... | 271 |
| 2 | Angels Place Gio Urshela On Injured List With Pelvic Fracture | Anthony Franco | 2023-06-16 | The Angels announced they've placed infielder Gio Urshela on the 10-day injured list with a frac... | 0.0 | [angel, announc, place, infield, gio, urshela, day, injur, list, fractur, left, pelvi, michael, ... | 244 |
| 3 | Angels Place Zach Neto On Injured List, Recall Kolton Ingram | Anthony Franco | 2023-06-15 | The Angels put rookie shortstop Zach Neto on the 10-day injured list due to a left oblique strai... | 0.0 | [angel, put, rooki, shortstop, zach, neto, day, injur, list, due, left, obliqu, strain, even, in... | 228 |
| 4 | Angels Sign Daniel Murphy To Minor League Deal | Darragh McDonald | 2023-06-12 | Infielder Daniel Murphy, who had been playing for the Long Island Ducks of the independent Atlan... | 0.0 | [infield, daniel, murphi, play, long, island, duck, independ, atlant, leagu, contract, purchas, ... | 173 |

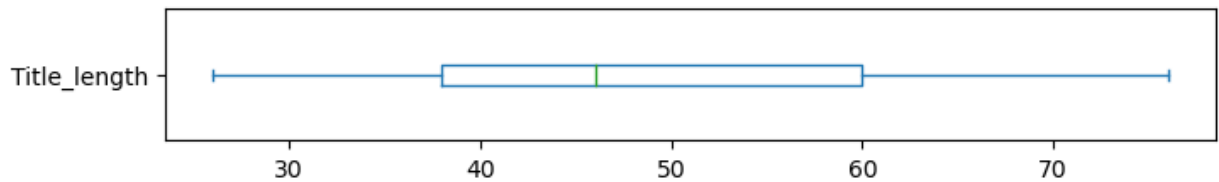
In [13]: `df['length'].plot(kind='box', vert=False, figsize=(8, 1))`

Out[13]: <Axes: >



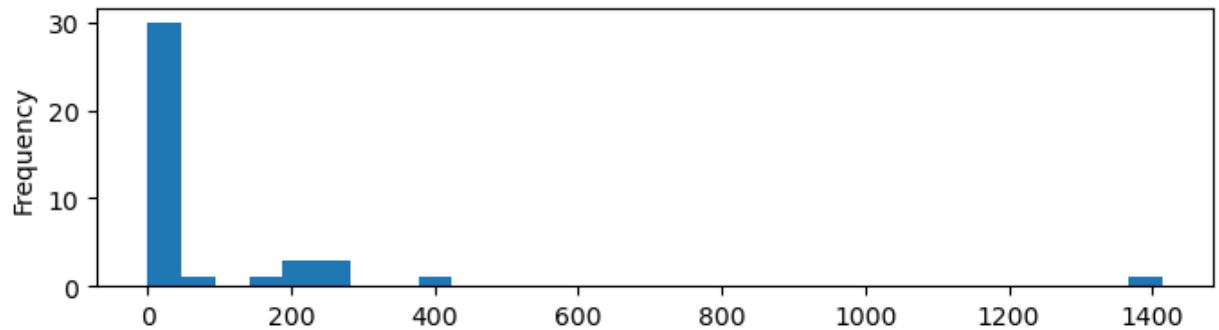
In [14]: `df['Title_length'] = df['Title'].str.len()
df['Title_length'].plot(kind='box', vert=False, figsize=(8, 1))`

Out[14]: <Axes: >



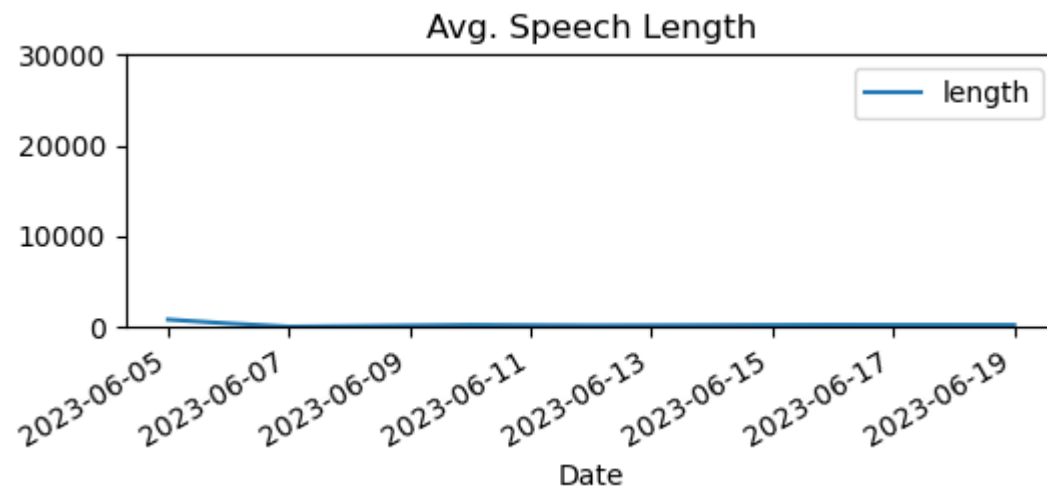
```
In [15]: df['length'].plot(kind='hist', bins=30, figsize=(8,2))
```

```
Out[15]: <Axes: ylabel='Frequency'>
```



```
In [16]: df.groupby('Date').agg({'length': 'mean'}) \
        .plot(title="Avg. Speech Length", ylim=(0,30000), figsize=(6,2))
```

```
Out[16]: <Axes: title={'center': 'Avg. Speech Length'}, xlabel='Date'>
```



```
In [17]: # Create a file to save the scraped data
file_path = "C:/Users/Grigor/OneDrive/Desktop/Master_Program/ADS509/Project/angels.txt"
with open(file_path, "w", encoding="utf-8") as file:
    # Initialize empty lists to store the extracted values
    titles = []
    authors = []
    dates = []
    contents = []

    # Iterate over the articles and extract the text
    for article in article:
        # 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")

print("Data saved to", file_path)

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

df = pd.DataFrame(data)
df.head()

```

Data saved to C:/Users/Grigor/OneDrive/Desktop/Master_Program/ADS509/Project/angels.txt

Out[17]:

| | Title | Author | Date | Content |
|---|---|------------------|---------------|---|
| 0 | Angels To Select Kevin Padlo | Anthony Franco | June 19, 2023 | The Angels are selecting the contract of corner infielder Kevin Padlo, reports Robert Murray of ... |
| 1 | Angels Place Anthony Rendon On Injured List | Darragh McDonald | June 19, 2023 | The Angels announced that infielder Anthony Rendon has been placed on the 10-day injured list du... |
| 2 | Angels Place Gio Urshela On Injured List With Pelvic Fracture | Anthony Franco | June 16, 2023 | The Angels announced they've placed infielder Gio Urshela on the 10-day injured list with a frac... |
| 3 | Angels Place Zach Neto On Injured List, Recall Kolton Ingram | Anthony Franco | June 15, 2023 | The Angels put rookie shortstop Zach Neto on the 10-day injured list due to a left oblique strai... |
| 4 | Angels Sign Daniel Murphy To Minor League Deal | Darragh McDonald | June 12, 2023 | Infielder Daniel Murphy, who had been playing for the Long Island Ducks of the independent Atlan... |

```
In [18]: Attendance_url = "https://www.baseball-reference.com/teams/ANA/attend.shtml"
```

```
# Read the HTML table into a list of DataFrames  
tables = pd.read_html(Attendance_url)  
  
# Select the table of interest (index 0 in this case)  
table = tables[0]
```

```
In [19]: Angels_Attendance = "C:/Users/Grigor/OneDrive/Desktop/Master_Program/ADS509/Project/Ar  
Angels_Attendance
```

```
Out[19]: 'C:/Users/Grigor/OneDrive/Desktop/Master_Program/ADS509/Project/Angels_Attendance.cs  
v'
```

```
In [20]: path = "C:/Users/Grigor/OneDrive/Desktop/Master_Program/ADS509/Project/"  
  
Angels_Attendance = pd.DataFrame(pd.read_csv(path + 'Angels_Attendance.csv'))  
Angels_Attendance
```

Out[20]:

| | Year | Tm | Lg | W | L | Finish | Playoffs | Attendance | Attend/G | Rank | Est. Payroll |
|-----|--------|--------------------|---------|------|------|--------|----------|------------|----------|-----------|---------------|
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 1 | 2023.0 | Los Angeles Angels | AL West | 35.0 | 31.0 | 3.0 | NaN | 1,053,117 | 31,913 | 4th of 15 | \$224,228,095 |
| 2 | 2022.0 | Los Angeles Angels | AL West | 73.0 | 89.0 | 3.0 | NaN | 2,457,461 | 30,339 | 5th of 15 | \$177,063,095 |
| 3 | 2021.0 | Los Angeles Angels | AL West | 77.0 | 85.0 | 4.0 | NaN | 1,515,689 | 18,484 | 6th of 15 | \$188,408,595 |
| 4 | 2020.0 | Los Angeles Angels | AL West | 26.0 | 34.0 | 4.0 | NaN | NaN | NaN | 8th of 15 | \$181,254,762 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 59 | 1965.0 | California Angels | AL | 75.0 | 87.0 | 7.0 | NaN | 566,727 | 7,084 | 8th of 10 | NaN |
| 60 | 1964.0 | Los Angeles Angels | AL | 82.0 | 80.0 | 5.0 | NaN | 760,439 | 9,388 | 7th of 10 | NaN |
| 61 | 1963.0 | Los Angeles Angels | AL | 70.0 | 91.0 | 9.0 | NaN | 821,015 | 10,136 | 5th of 10 | NaN |
| 62 | 1962.0 | Los Angeles Angels | AL | 86.0 | 76.0 | 3.0 | NaN | 1,144,063 | 14,124 | 4th of 10 | NaN |
| 63 | 1961.0 | Los Angeles Angels | AL | 70.0 | 91.0 | 8.0 | NaN | 603,510 | 7,360 | 9th of 10 | NaN |

64 rows × 14 columns

```

In [21]: Angels_05_df = pd.DataFrame(pd.read_csv(path + 'Angels_Attendance.csv'))
Angels_05_df = Angels_05_df[Angels_05_df["Year"] > 2004]
Angels_05_df

```

Out[21]:

| | Year | Tm | Lg | W | L | Finish | Playoffs | Attendance | Attend/G | Rank | Est. Payroll |
|-----------|--------|-------------------------------|---------|------|------|--------|-----------------|------------|----------|-----------|---------------|
| 1 | 2023.0 | Los Angeles Angels | AL West | 35.0 | 31.0 | 3.0 | NaN | 1,053,117 | 31,913 | 4th of 15 | \$224,228,095 |
| 2 | 2022.0 | Los Angeles Angels | AL West | 73.0 | 89.0 | 3.0 | NaN | 2,457,461 | 30,339 | 5th of 15 | \$177,063,095 |
| 3 | 2021.0 | Los Angeles Angels | AL West | 77.0 | 85.0 | 4.0 | NaN | 1,515,689 | 18,484 | 6th of 15 | \$188,408,595 |
| 4 | 2020.0 | Los Angeles Angels | AL West | 26.0 | 34.0 | 4.0 | NaN | NaN | NaN | 8th of 15 | \$181,254,762 |
| 5 | 2019.0 | Los Angeles Angels | AL West | 72.0 | 90.0 | 4.0 | NaN | 3,023,012 | 37,321 | 2nd of 15 | \$158,078,584 |
| 6 | 2018.0 | Los Angeles Angels | AL West | 80.0 | 82.0 | 4.0 | NaN | 3,020,216 | 37,287 | 2nd of 15 | \$166,849,666 |
| 7 | 2017.0 | Los Angeles Angels | AL West | 80.0 | 82.0 | 2.0 | NaN | 3,019,585 | 37,279 | 3rd of 15 | \$181,125,500 |
| 8 | 2016.0 | Los Angeles Angels | AL West | 74.0 | 88.0 | 4.0 | NaN | 3,016,142 | 37,236 | 3rd of 15 | \$139,712,000 |
| 9 | 2015.0 | Los Angeles Angels of Anaheim | AL West | 85.0 | 77.0 | 3.0 | NaN | 3,012,765 | 37,195 | 2nd of 15 | \$131,522,500 |
| 10 | 2014.0 | Los Angeles Angels of Anaheim | AL West | 98.0 | 64.0 | 1.0 | Lost ALDS (3-0) | 3,095,935 | 38,221 | 2nd of 15 | \$128,667,000 |
| 11 | 2013.0 | Los Angeles Angels of Anaheim | AL West | 78.0 | 84.0 | 3.0 | NaN | 3,019,505 | 37,278 | 4th of 15 | \$116,532,500 |
| 12 | 2012.0 | Los Angeles Angels | AL West | 89.0 | 73.0 | 3.0 | NaN | 3,061,770 | 37,800 | 3rd of 14 | \$141,073,500 |

| | Year | Tm | Lg | W | L | Finish | Playoffs | Attendance | Attend/G | Rank | Est. Payroll |
|----|--------|---|------------|-------|------|--------|------------------------|------------|----------|--------------|---------------|
| | | of Anaheim | | | | | | | | | |
| 13 | 2011.0 | Los Angeles Angels of Anaheim | AL West | 86.0 | 76.0 | 2.0 | NaN | 3,166,321 | 39,090 | 3rd of 14 | \$138,543,166 |
| 14 | 2010.0 | Los Angeles Angels of Anaheim | AL West | 80.0 | 82.0 | 3.0 | NaN | 3,250,814 | 40,134 | 2nd of 14 | \$104,963,866 |
| 15 | 2009.0 | Los Angeles Angels of Anaheim | AL West | 97.0 | 65.0 | 1.0 | Lost ALCS (4- 2) | 3,240,386 | 40,005 | 2nd of 14 | \$118,169,000 |
| 16 | 2008.0 | Los Angeles Angels of Anaheim | AL West | 100.0 | 62.0 | 1.0 | Lost ALDS (3-1) | 3,336,747 | 41,194 | 2nd of 14 | \$119,216,333 |
| 17 | 2007.0 | Los Angeles Angels of Anaheim | AL West | 94.0 | 68.0 | 1.0 | Lost ALDS (3-0) | 3,365,632 | 41,551 | 2nd of 14 | \$109,251,333 |
| 18 | 2006.0 | Los Angeles Angels of Anaheim | AL West | 89.0 | 73.0 | 2.0 | NaN | 3,406,790 | 42,059 | 2nd of 14 | \$103,472,000 |
| 19 | 2005.0 | Los Angeles Angels of Anaheim | AL West | 95.0 | 67.0 | 1.0 | Lost ALCS (4- 1) | 3,404,686 | 42,033 | 2nd of 14 | \$94,867,822 |

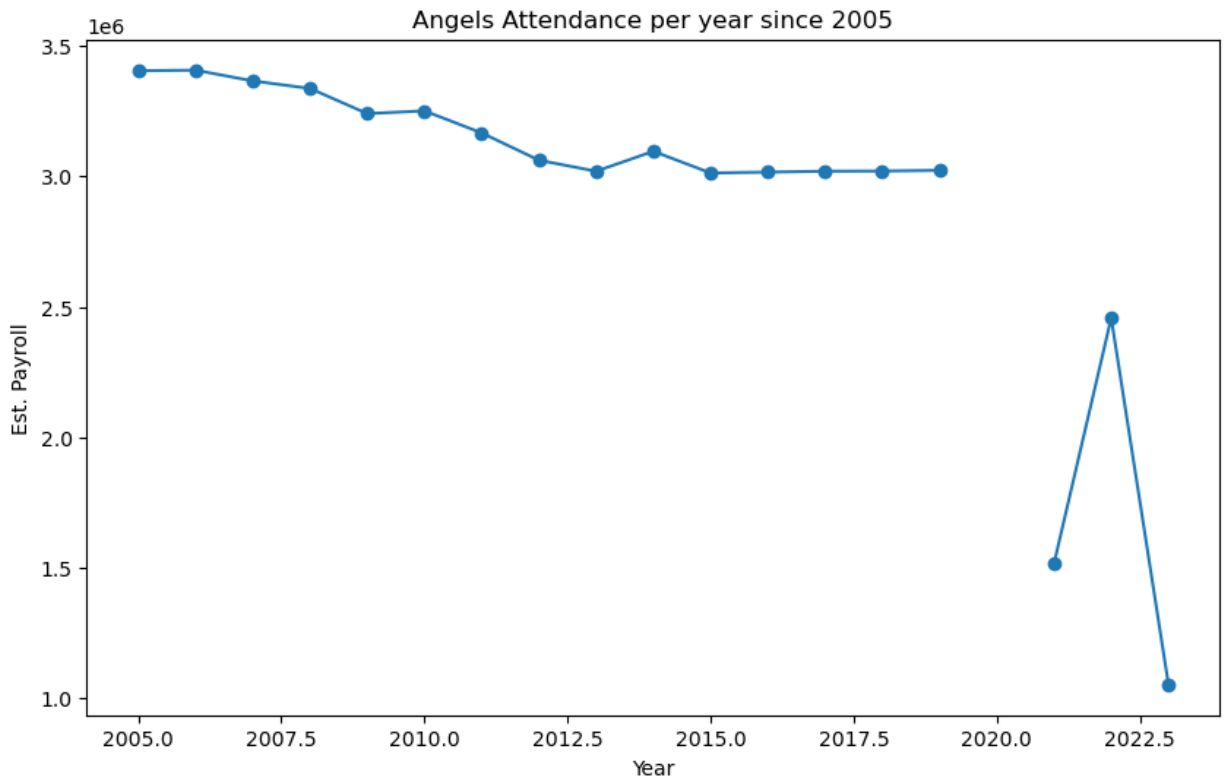
```
In [22]: Angels_05_df['Est. Payroll'] = Angels_05_df['Est. Payroll'].replace({'\$':''}, regex =
Angels_05_df['Est. Payroll'] = Angels_05_df['Est. Payroll'].replace({'\\','':''}, regex =
Angels_05_df['Attendance'] = Angels_05_df['Attendance'].replace({'\\','':''}, regex = True
```

```
In [23]: #plot data
Angels_05_df['Attendance'] = Angels_05_df['Attendance'].astype(float)
Angels_05_df.dtypes
# Plotting
plt.figure(figsize=(10, 6))

# Plotting Attendance over the years
plt.plot(Angels_05_df['Year'], Angels_05_df['Attendance'], marker='o')
plt.xlabel('Year')
```

```
plt.ylabel('Est. Payroll')
plt.title('Angels Attendance per year since 2005')
```

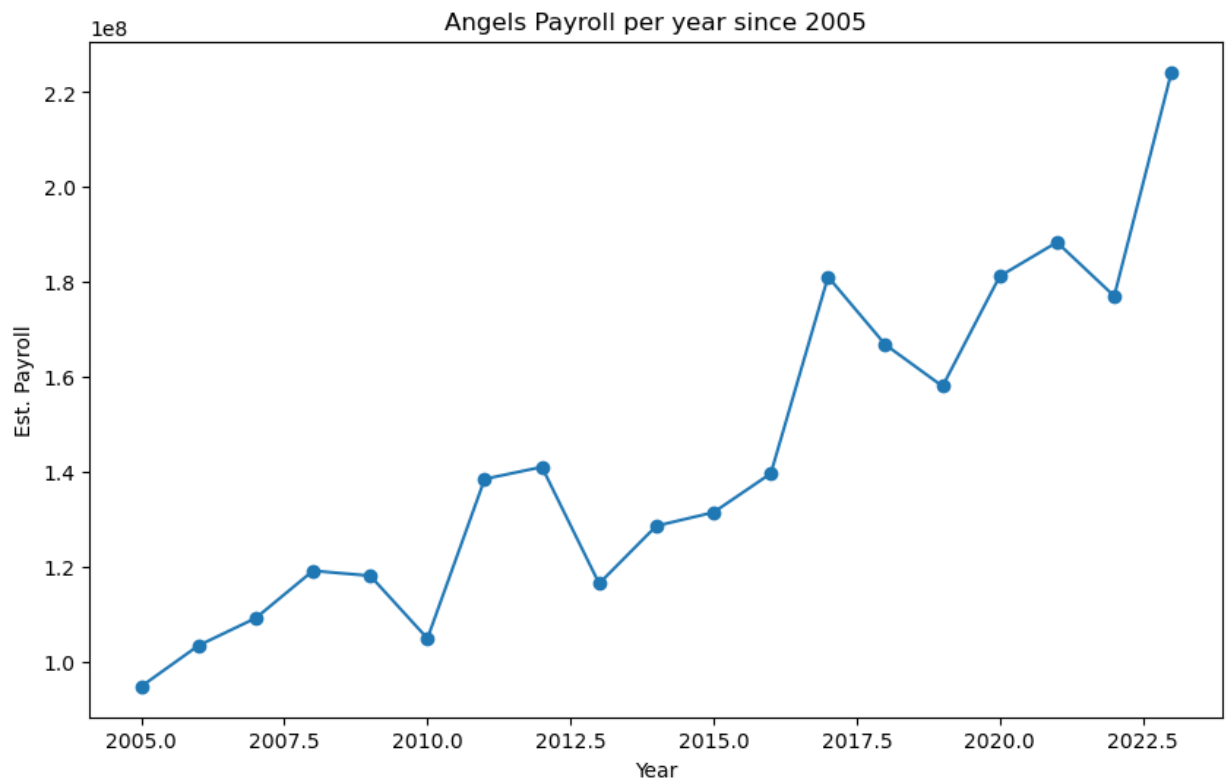
Out[23]: Text(0.5, 1.0, 'Angels Attendance per year since 2005')



```
In [24]: #plot data
Angels_05_df['Est. Payroll'] = Angels_05_df['Est. Payroll'].astype(float)
Angels_05_df.dtypes
# Plotting
plt.figure(figsize=(10, 6))

# Plotting Attendance over the years
plt.plot(Angels_05_df['Year'], Angels_05_df['Est. Payroll'], marker='o')
plt.xlabel('Year')
plt.ylabel('Est. Payroll')
plt.title('Angels Payroll per year since 2005')
```

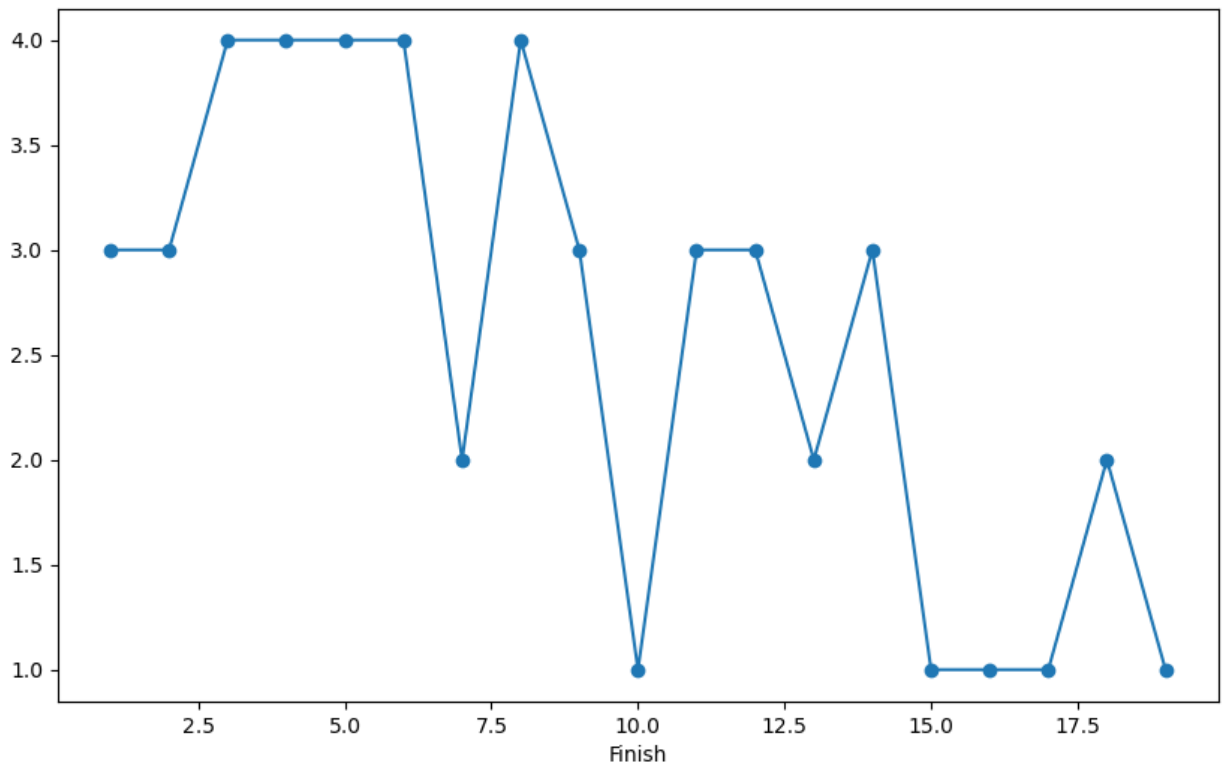
Out[24]: Text(0.5, 1.0, 'Angels Payroll per year since 2005')



```
In [26]: plt.figure(figsize=(10, 6))

# Plotting Attendance over the years
plt.plot(Angels_05_df['Finish'], marker='o')
plt.xlabel('Finish')
```

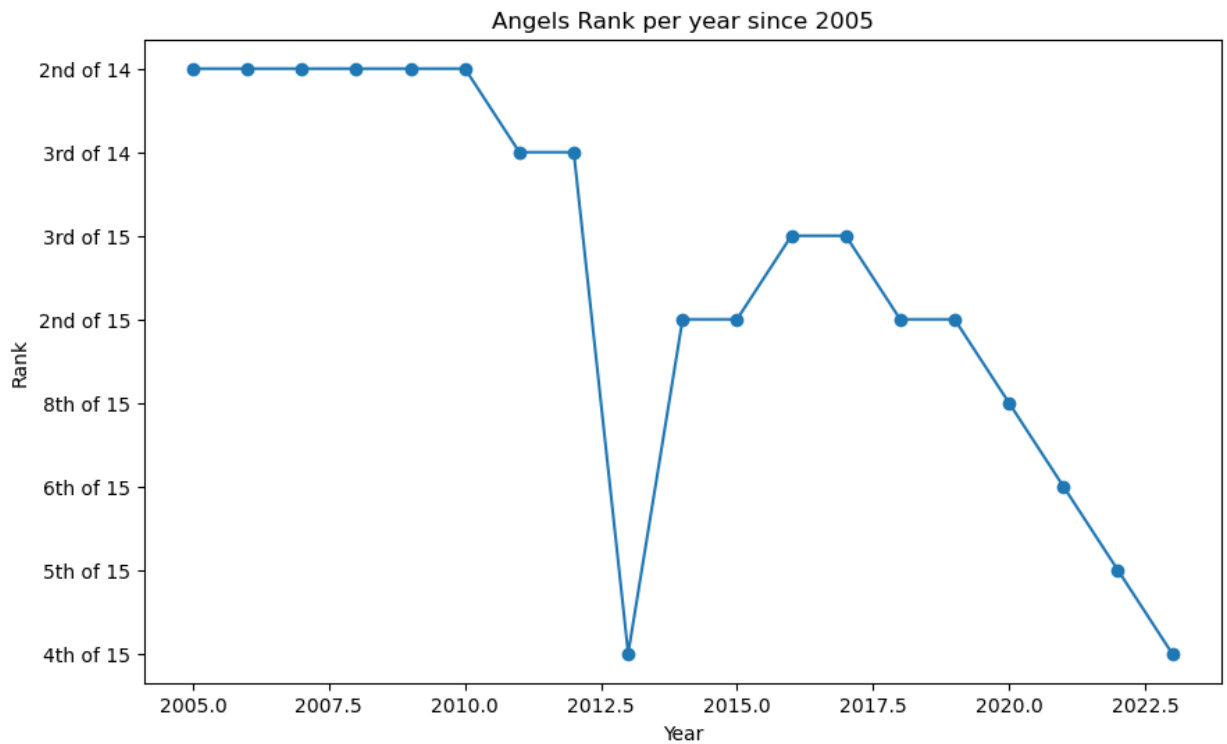
```
Out[26]: Text(0.5, 0, 'Finish')
```




```
In [27]: plt.figure(figsize=(10, 6))

# Plotting Attendance over the years
plt.plot(Angels_05_df['Year'], Angels_05_df['Rank'], marker='o')
plt.xlabel('Year')
plt.ylabel('Rank')
plt.title('Angels Rank per year since 2005')
```

```
Out[27]: Text(0.5, 1.0, 'Angels Rank per year since 2005')
```



```
In [28]: corr = Angels_05_df["Est. Payroll"].corr(Angels_05_df['Finish'])
print(corr)

0.5637391640975834
```