1. Team Members

a. Name: Lian Zeng, AndrewID: lianz
b. Name: Kathy Wang, AndrewID: peijiaw
c. Name: Xinyuan Yang, AndrewID: xy3
d. Name: Yuting Liu, AndrewID: yutingl2

2. Libraries:

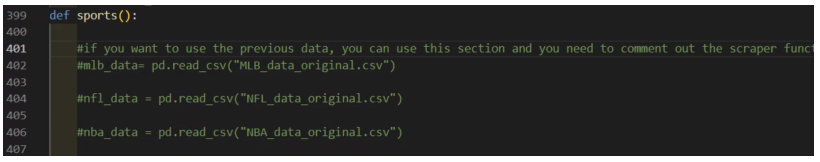
a. pip install beautifulsoup4
b. pip install selenium
c. pip install ipywidgets
d. python -m pip install -U matplotlib
e. pip install lxml
f. pip install webdriver-manager
g. python -m pip install -U matplotlib

3. Scraper Function

a. MLB Scraper Function & NBA Scraper Function &NFL Scraper Function

i. Introduction: MLB, NBA, and NFL Scraper Functions are used to scrape the data from the Stubhub website. The scraped data includes team names, locations, dates, times, and prices.
ii. Tools: selenium and BeautifulSoup
iii. Notes: The Stubhub website sometimes has problems displaying prices, and it may take a few restarts to display them. So, for convenience, if there is no price, I will just read the csv files we previously cleaned up that include the prices and run the following code.

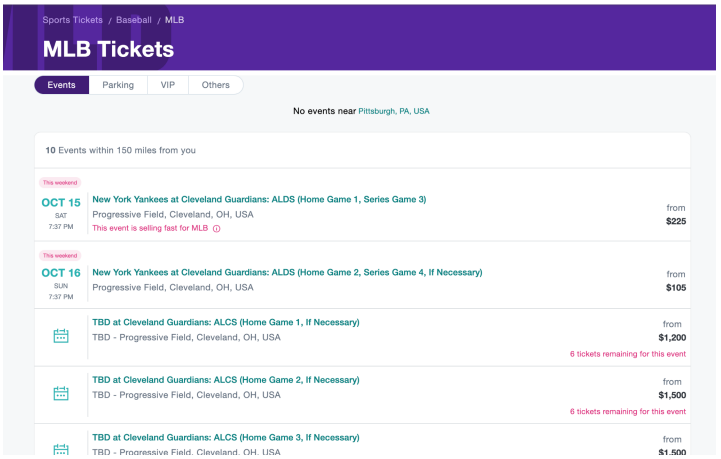
1. If you want to scrape the most recent data, use the live scraper code and run the whole file without changing any code.
2. If you want to directly use previously cleaned data, uncomment them in the sports function and run the whole file



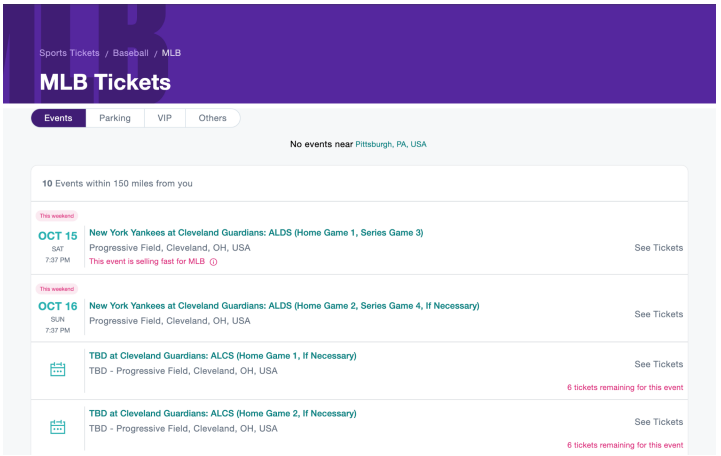
and comments on these lines

```
399  def sports():
400      mlb_data= MLB_scrapper_func()
401      try:
402          price = int(mlb_data['Price'][0])
403      except:
404          mlb_data = pd.read_csv("MLB_data_original.csv")
405      nfl_data = NFL_scrapper_func()
406      try:
407          price = int(nfl_data['Price'][0])
408      except:
409          nfl_data = pd.read_csv("NFL_data_original.csv")
410      nba_data = NBA_scrapper_func()
411      try:
412          price = int(nba_data['Price'][0])
413      except:
414          nba_data = pd.read_csv("NBA_data_original.csv")
415
```

- **Right demonstration:**



- **Wrong demonstration:**



iv. Sample outputs:

1. MLB scraper function

	Home_Team	Away_Team	Arena	City	State	Country	Date	Day	Time	Price
0	Los Angeles Dodgers	San Diego Padres	Dodger Stadium	Los Angeles	CA	USA	2022-10-11	Tue	6:37 PM	$65
1	New York Yankees	Cleveland Guardians	Yankee Stadium	Bronx	NY	USA	2022-10-11	Tue	7:37 PM	$120
2	Atlanta Braves	Philadelphia Phillies	Truist Park	Atlanta	GA	USA	2022-10-12	Wed	4:35 PM	$28
3	Los Angeles Dodgers	San Diego Padres	Dodger Stadium	Los Angeles	CA	USA	2022-10-12	Wed	5:37 PM	$57
4	Houston Astros	Seattle Mariners	Minute Maid Park	Houston	TX	USA	2022-10-13	Thu	2:37 PM	$35
5	New York Yankees	Cleveland Guardians	Yankee Stadium	Bronx	NY	USA	2022-10-13	Thu	7:37 PM	$108

2. NBA scraper function

	Home_Team	Away_Team	Arena	City	State	Country	Date	Day	Time	Price
0	Orlando Magic Preseason	Memphis Grizzlies Preseason	Amway Center	Orlando	FL	USA	2022-10-11	Tue	7:00 PM	$12
1	Detroit Pistons Preseason	Oklahoma City Thunder Preseason	Little Caesars Arena	Detroit	MI	USA	2022-10-11	Tue	7:00 PM	
2	Chicago Bulls Preseason	Milwaukee Bucks Preseason	United Center	Chicago	IL	USA	2022-10-11	Tue	7:00 PM	$22
3	Utah Jazz Preseason	San Antonio Spurs Preseason	Vivint Smart Home Arena	Salt Lake City	UT	USA	2022-10-11	Tue	7:00 PM	$2
4	Golden State Warriors Preseason	Portland Trail Blazers Preseason	Chase Center	San Francisco	CA	USA	2022-10-11	Tue	7:00 PM	$32

3. NFL scraper function

	Home_Team	Away_Team	Arena	City	State	Country	Date	Day	Time	Price
0	Chicago Bears	Washington Commanders	Soldier Field	Chicago	IL	USA	2022-10-13	Thu	7:15 PM	$80
1	New Orleans Saints	Cincinnati Bengals	Caesars Superdome	New Orleans	LA	USA	2022-10-16	Sun	12:00 PM	$109
2	Green Bay Packers	New York Jets	Lambeau Field	Green Bay	WI	USA	2022-10-16	Sun	12:00 PM	$169
3	New York Giants	Baltimore Ravens	MetLife Stadium	East Rutherford	NJ	USA	2022-10-16	Sun	1:00 PM	$115
4	Atlanta Falcons	San Francisco 49ers	Mercedes-Benz Stadium	Atlanta	GA	USA	2022-10-16	Sun	1:00 PM	$90

4. Sports Tickets

a. **Introduction**: we have three kinds of sports tickets: NBA, MLB, and NFL. First, users can select the sport they are interested in. Then, the program will filter the tickets based on the team name, city, and the time frame they choose. Finally, the user will select the one with the most desirable sports tickets.

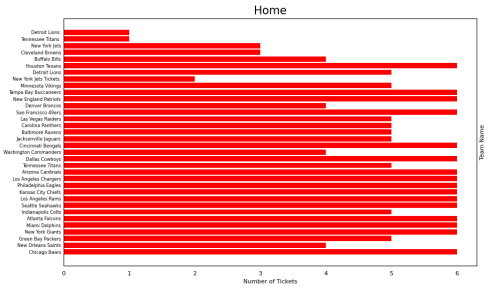
b. Choose Sports

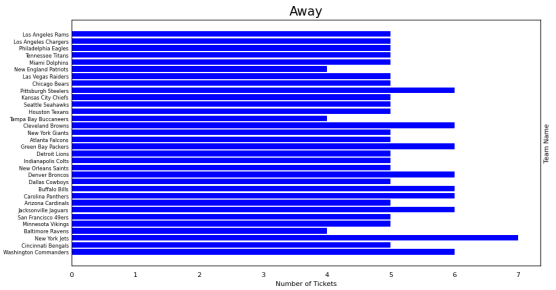
i. Choose the sports from the list (Case, and front and back whitespace Insensitive)

c. Event Overview

i. After selecting the sport, two plots will be created, with x-axis standing for the amount of the ticket, y-axis standing for the team name for each sport. The first plot stands for the home team and the second stands for the away team, this will grant user an overview of the ticket distribution for different teams:

(Taking NFL as an example)





d. Filter Results

i. Choose the team name

1. Select the team name from the list above, which includes all teams for the selected sport. You can choose not to select any teams,
2. Sample output:
    a. select Austin Spurs

```
The team you choose is: Austin Spurs
Your search result is:
```

	Home_Team	Away_Team	Arena	City	State	Country	Date	Day	Time	Price
0	Austin Spurs	Texas Legends	H-E-B Center at Cedar Park	Cedar Park	TX	USA	2022-11-04	Fri	7:30 PM	$29
1	Austin Spurs	Texas Legends	H-E-B Center at Cedar Park	Cedar Park	TX	USA	2022-11-05	Sat	7:00 PM	$29

b. Select nothing, then display all teams

```
The team you choose is:
Your search result is:
```

	Home_Team	Away_Team	Arena	City	State	Country	Date	Day	Time	Price
0	Orlando Magic Preseason	Memphis Grizzlies Preseason	Amway Center	Orlando	FL	USA	2022-10-11	Tue	7:00 PM	$12
1	Detroit Pistons Preseason	Oklahoma City Thunder Preseason	Little Caesars Arena	Detroit	MI	USA	2022-10-11	Tue	7:00 PM	
2	Chicago Bulls Preseason	Milwaukee Bucks Preseason	United Center	Chicago	IL	USA	2022-10-11	Tue	7:00 PM	$22

ii. Choose the city:

1. Input the city from the list, which includes all the cities that will be hosting games for the selected team. You can choose not to select any city.

2. Step2: After selection, the output will be:

```
The city you choose is: Boston
Your search result is:
```

	Home_Team	Away_Team	Arena	City	State	Country	Date	Day	Time	Price
0	Boston Celtics	Philadelphia 76ers	TD Garden	Boston	MA	USA	2022-10-18	Tue	7:30 PM	$80
1	Boston Celtics	Cleveland Cavaliers	TD Garden	Boston	MA	USA	2022-10-28	Fri	7:30 PM	$57

iii. Choose the date:

1. Input the start date and end date from the listed ranges(Follow the prompt date format). You can choose not to select any time period.

```
The date is from: 2022-10-11 to 2022-10-16
Your search result is:
```

	Home_Team	Away_Team	Arena	City	State	Country	Date	Day	Time	Price
0	Los Angeles Dodgers	San Diego Padres	Dodger Stadium	Los Angeles	CA	USA	2022-10-11	Tue	6:37 PM	$72
1	Los Angeles Dodgers	San Diego Padres	Dodger Stadium	Los Angeles	CA	USA	2022-10-12	Wed	5:37 PM	$60
2	San Diego Padres	Los Angeles Dodgers	Petco Park	San Diego	CA	USA	2022-10-14	Fri	1:00 PM	$175
3	San Diego Padres	Los Angeles Dodgers	Petco Park	San Diego	CA	USA	2022-10-15	Sat	1:00 PM	$160
4	Los Angeles Dodgers	San Diego Padres	Dodger Stadium	Los Angeles	CA	USA	2022-10-16	Sun	1:00 PM	$71

e. Sort Results

i. Introduction: In this function, the filtered results will be sorted by different cases. price from low to high, price from high to low, most recent, least recently

ii. Input your “sort by” choice. You can choose not to select any cases.

	Home_Team	Away_Team	Arena	City	State	Country	Date	Day	Time	Price
0	Los Angeles Dodgers	San Diego Padres	Dodger Stadium	Los Angeles	CA	USA	2022-10-11	Tue	6:37 PM	$72
1	Los Angeles Dodgers	San Diego Padres	Dodger Stadium	Los Angeles	CA	USA	2022-10-12	Wed	5:37 PM	$60
2	San Diego Padres	Los Angeles Dodgers	Petco Park	San Diego	CA	USA	2022-10-14	Fri	1:00 PM	$175
3	San Diego Padres	Los Angeles Dodgers	Petco Park	San Diego	CA	USA	2022-10-15	Sat	1:00 PM	$160
4	Los Angeles Dodgers	San Diego Padres	Dodger Stadium	Los Angeles	CA	USA	2022-10-16	Sun	1:00 PM	$71

f. Final Choice

i. Introduction: Users choose the one ticket they want most from the above results.

ii. Choose the ticket according to the index of the final results above

```
The sports teams are: Delaware Blue Coats vs.Greensboro Swarm
The game time is on 2023-01-06 at 7:00 PM
The ticket price is: $19
```

5. Flight Input and Results

a. Fight Scraper Function

i. First, install webdriver-manager before running.

ii. Second, we need to get the **chrome version** you are using. For the latest version, the stable channel has been updated to 104.0.5112.101 for Mac and Linux and 104.0.5112.102/101 for Windows. Replace your chrome version in the code:

```
# chrome version used on your device
agents = ["Chrome/104.0.5112.101"]
```

iii. After running the Flight Scraper Function, run the IATA code block to store the airports’ three-letter code and city name. Later there will be a code to change the city name and airports to an airport iata code

which is used to scrape flight info. User could simply pass in the city name and choose the correct airports they want.

b. flight input

i. The user will first input the origin city name with the first letter capitalized. For example, Los Angeles.
ii. The options of airports in the origin city will be presented. The user should copy one airport name and pass it into the answer box.
iii. The options of airports in the destination city will then be shown. The user should copy one airport name and pass it into the answer box.
iv. Then, users will be asked to pass their departure date and return date in the YYYY-MM-DD format. The passed-in date must be a future date. For example, if today is 2022-10-15, then the passed-in date has to be a date after but include 2022-10-15.
v. If the input is in wrong format or the there is no flight correspond to the given airport and city, user need to rerun all the code from the start.

c. flight results

i. This code block will present the flight information with the lowest price, there can be many flights that have the lowest price.
ii. It will also present all possible flight information. The user needs to choose the flight that works for them and passes in the row number/index of that flight info to the input box when asked to enter the flight row number.
iii. The flight price chosen by the user is stored and will be used to calculate the total budget.

6. Yelp Data and Recommendation

a. Data Overview

Yelp Data is downloaded from the yelp open data source web, and in our application, we use the business and review json files to serve as recommendation references.

b. Basic idea and Selection process

i. data combination

We combine the business and review data set into one csv file, dropping the information we do not need here (because of the file size, it may take few minutes to finish the combination). The file includes some basic information like event names, addresses, review texts, stars, and each user ID.

ii. data selection

Our recommendation is based on the reviews and stars users have rated for, so we sort the data with the index of stars and remove all the duplicates. Finally only keep the address and review details for our users, with 5 recommendations for each input.

iii. input & results

The user will first put their expected state to visit in the shell, and then they will see the results with reviews, stars, and detailed address info for their reference.

```
state_input = input('Please enter the state code: ')
recommendation = df_event.loc[state_input].head(5)
print(recommendation)
```

```
Please enter the state code: CA
```

And recommendation will just show below:

```
Here is the recommended event for you:
点击滚动输出;双击隐藏                                    name             address  \
state
CA                                Living Vehicle                 NaN
CA     Carquest Auto Parts - Larrys Auto Parts  5855 Hollister Ave
CA                                Dustie Wagens  115 W De La Guerra
CA                            Sage Hill Campground    5050 Paradise Rd
CA        Santa Barbara Auto Truck Accessories  5737 Hollister Ave

                 city  stars                                                text
state
CA     Santa Barbara    5.0  I saw this tiny home and the Santa Barbara Ear...
CA            Goleta    5.0  Very helpfull and people that know what they a...
CA     Santa Barbara    5.0  I surprised my husband with a California Road ...
CA     Santa Barbara    4.5  This place has the potential for 5 stars. A fe...
CA            Goleta    4.5  Love this place and the owners. Steve and Lynn...
```