CHANCE TO GO PRO IN MEN'S INDOOR VOLLEYBALL IF YOU'RE FROM THE USA

Aaron Dierkes

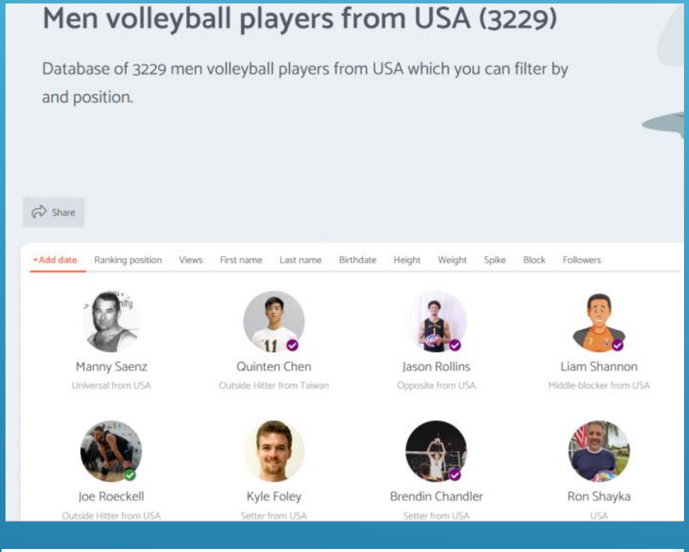
VOLLEYBALL BACKGROUND

- Men's indoor volleyball has 6 positions: **outside 1, outside 2, middle, setter, opposite, and libero.**
- The progression for volleyball players that take the sport seriously is, **play club** growing up, **play in high school**, try and be recruited for an **NCAA college**, be a top player in the NCAA, try to play **professionally overseas.**
- **4%** of high school players play in the NCAA and **2%** become professionals.
- All the professional clubs in the world **exist outside of the US**. There are "professional teams' in the US but the league started in 2020 and you don't get paid. For this project we will consider them not pro.

PROBLEM/OPPORTUNITY

- Players have no metric to check to determine their chances to go pro.
- Rely on coaches to guide them into positions.
- I see an opportunity to create something that will give you a realistic answer based on genetics and scalable to include things like awards, clubs played growing up, and where you're from.
- This will help players see if they need to train harder to overcome genetic differences that may decrease their odds.

PROPOSED VISION TO TACKLE THE PROBLEM



Men volleyball players from USA (3229)

Database of 3229 men volleyball players from USA which you can filter by and position.

Manny Saenz
Universal from USA

Quinten Chen
Outside Hitter from Taiwan

Jason Rollins
Opposite from USA

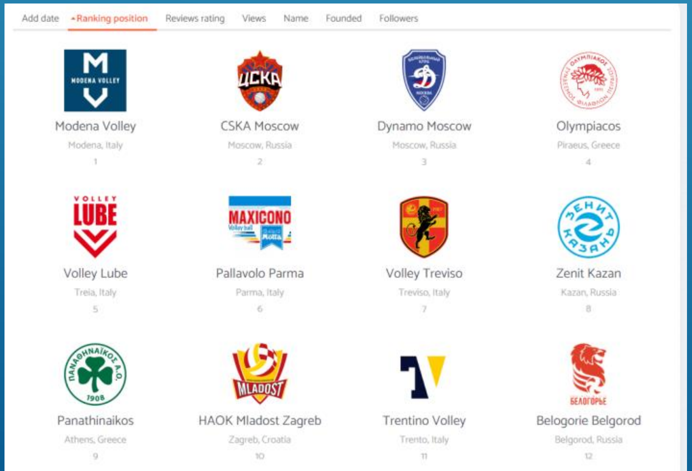
Liam Shannon
Middle-blocker from USA

Joe Roeckell
Outside Hitter from USA

Kyle Foley
Setter from USA

Brendin Chandler
Setter from USA

Ron Shayka
USA



Modena Volley
Modena, Italy
1

CSKA Moscow
Moscow, Russia
2

Dynamo Moscow
Moscow, Russia
3

Olympiacos
Piraeus, Greece
4

Volley Lube
Treia, Italy
5

Pallavolo Parma
Parma, Italy
6

Volley Treviso
Treviso, Italy
7

Zenit Kazan
Kazan, Russia
8

Panathinaikos
Athens, Greece
9

HAOK Mladost Zagreb
Zagreb, Croatia
10

Trentino Volley
Trento, Italy
11

Belogorie Belgorod
Belgorod, Russia
12

- My plan is to use all the US player data from the volleybox website.
- Compare the US players club information to the list of clubs on the website to get an accurate assessment of US players playing actual professional volleyball.
- Get the full roster data from the previous 14ish years or more if available of NCAA D1 schools and add all the data together on **names** to see who went pro out of college and who didn't.
- I'll then be able to do some EDA to see average heights by position for pro teams, dominant hands per position, vertical jump per position, etc which will help me create a model that will help predict the outcome, going pro.

POTENTIAL IMPACT

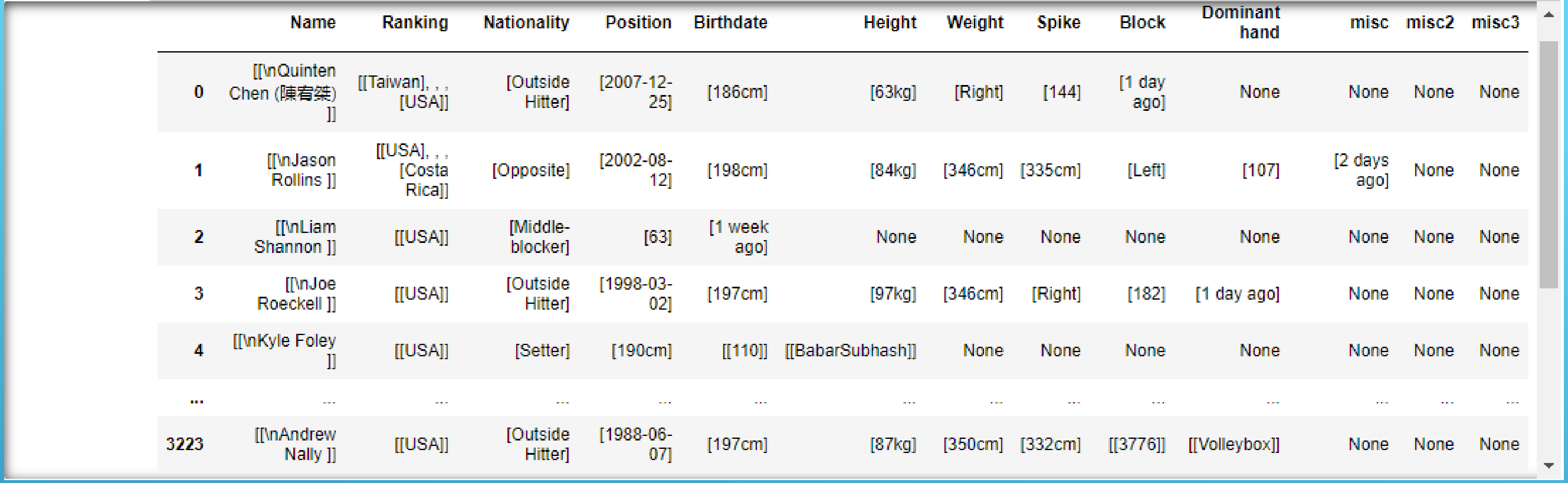
- This model would help players determine what they're lacking by position compared to the pros and help motivate them to work harder or help them make the decision to switch to another sport if they think they'll have a better shot going pro elsewhere.
- This model would also help coaches and scouts determine whether a player shows potential based on the results of the model.
- Could potentially save coaches and scouts time when looking for players.
- Could potentially save players time and money associated with playing club and training.

THE DATA SET

- The data I can get will ultimately provide me with all of the necessary features for a baseline model
- My ‘y’ is whether or not they play for a professional club which I can determine from extracting the information from volleybox.
- My ‘X’ or features are all available in the data sets I’ll be scraping.

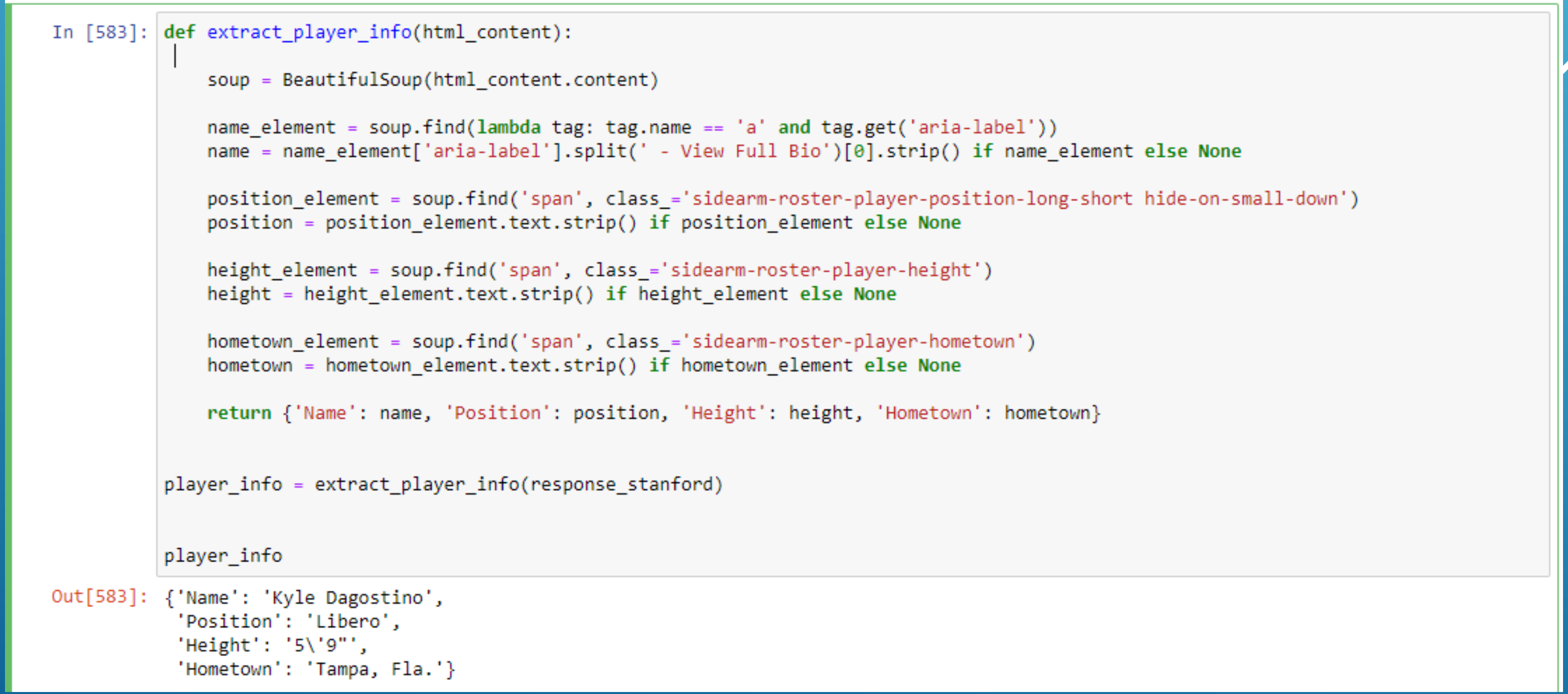
Name, Rank, Nationality, Position, Birthdate, Height, Weight, Spike, Block, Dominant hand, Hometown, Club.

- The opportunity to scale this model by adding additional features is there.



	Name	Ranking	Nationality	Position	Birthdate	Height	Weight	Spike	Block	Dominant hand	misc	misc2	misc3
0	[[\nQuinten Chen (陳宥榮)]]	[[Taiwan], , , [USA]]	[Outside Hitter]	[2007-12-25]	[186cm]	[63kg]	[Right]	[144]	[1 day ago]	None	None	None	None
1	[[\nJason Rollins]]	[[USA], , , [Costa Rica]]	[Opposite]	[2002-08-12]	[198cm]	[84kg]	[346cm]	[335cm]	[Left]	[107]	[2 days ago]	None	None
2	[[\nLiam Shannon]]	[[USA]]	[Middle-blocker]	[63]	[1 week ago]	None	None	None	None	None	None	None	None
3	[[\nJoe Roeckell]]	[[USA]]	[Outside Hitter]	[1998-03-02]	[197cm]	[97kg]	[346cm]	[Right]	[182]	[1 day ago]	None	None	None
4	[[\nKyle Foley]]	[[USA]]	[Setter]	[190cm]	[[110]]	[[BabarSubhash]]	None	None	None	None	None	None	None
...	...	...	...	...	...	...	...	...	...	...	...	...	...
3223	[[\nAndrew Nally]]	[[USA]]	[Outside Hitter]	[1988-06-07]	[197cm]	[87kg]	[350cm]	[332cm]	[[3776]]	[[Volleybox]]	None	None	None

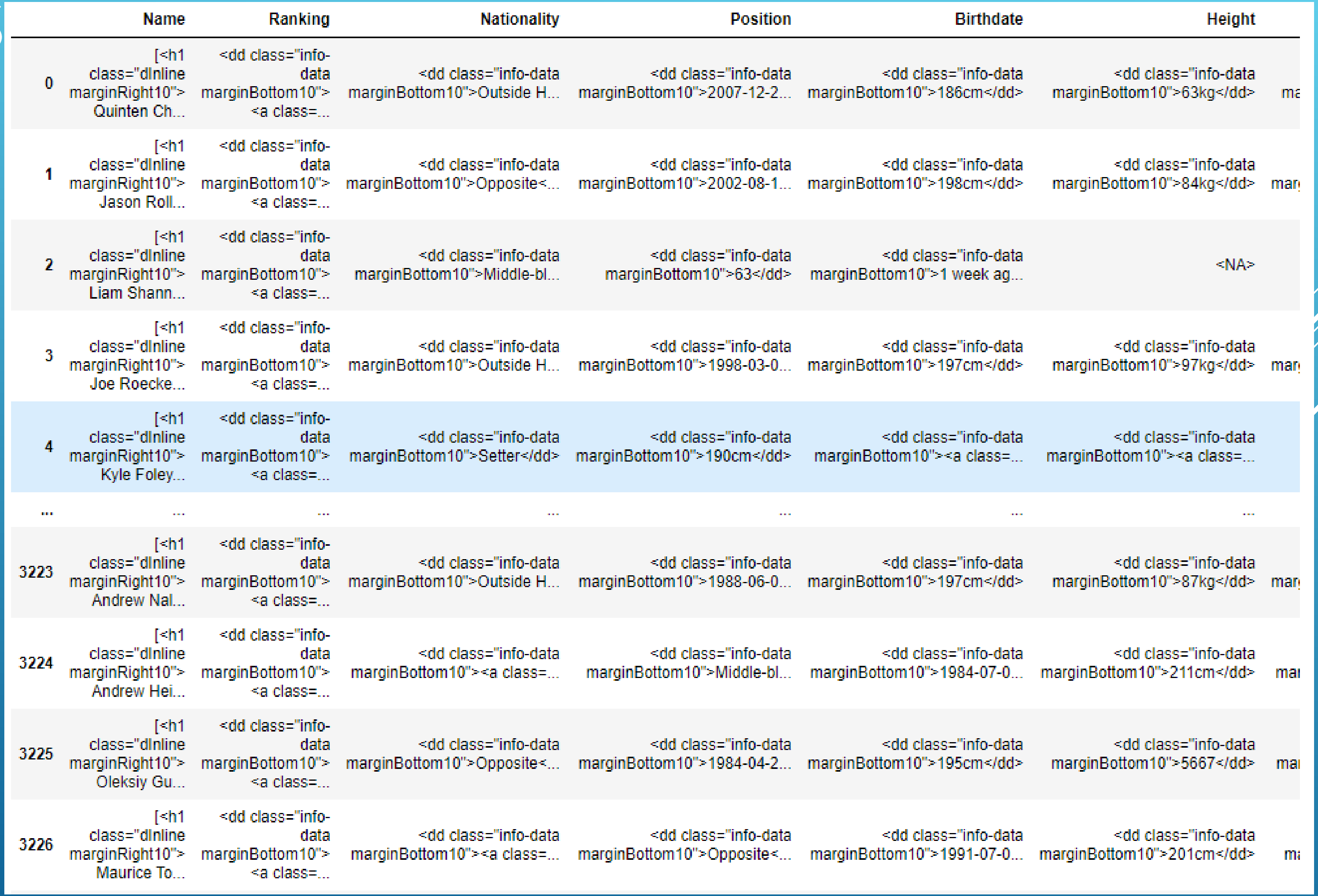
This is uncleaned data pulled from the volleybox players page



```
In [583]: def extract_player_info(html_content):
          |
              soup = BeautifulSoup(html_content.content)

              name_element = soup.find(lambda tag: tag.name == 'a' and tag.get('aria-label'))
              name = name_element['aria-label'].split(' - View Full Bio')[0].strip() if name_element else None

              position_element = soup.find('span', class_='sidearm-roster-player-position-long-short hide-on-small-down')
              position = position_element.text.strip() if position_element else None

              height_element = soup.find('span', class_='sidearm-roster-player-height')
              height = height_element.text.strip() if height_element else None

              hometown_element = soup.find('span', class_='sidearm-roster-player-hometown')
              hometown = hometown_element.text.strip() if hometown_element else None

              return {'Name': name, 'Position': position, 'Height': height, 'Hometown': hometown}


          player_info = extract_player_info(response_stanford)


          player_info

Out[583]: {'Name': 'Kyle Dagostino',
           'Position': 'Libero',
           'Height': '5\'9"',
           'Hometown': 'Tampa, Fla.'}
```

This is the work in progress for extracting data from one of the rosters I’ll be using data from

THIS IS WHAT IT LOOKS LIKE WHEN I CONVERT IT INTO A STRING AND TRY TO EXTRACT AND PUT THE VALUES INTO THE CORRECT COLUMNS



	Name	Ranking	Nationality	Position	Birthdate	Height	
0	[<h1 class="dInline marginRight10"> Quinten Ch...	<dd class="info-data marginBottom10"> <a class=...	<dd class="info-data marginBottom10">Outside H...	<dd class="info-data marginBottom10">2007-12-2...	<dd class="info-data marginBottom10">186cm</dd>	<dd class="info-data marginBottom10">63kg</dd>	ma
1	[<h1 class="dInline marginRight10"> Jason Roll...	<dd class="info-data marginBottom10"> <a class=...	<dd class="info-data marginBottom10">Opposite<...	<dd class="info-data marginBottom10">2002-08-1...	<dd class="info-data marginBottom10">198cm</dd>	<dd class="info-data marginBottom10">84kg</dd>	mar
2	[<h1 class="dInline marginRight10"> Liam Shann...	<dd class="info-data marginBottom10"> <a class=...	<dd class="info-data marginBottom10">Middle-bl...	<dd class="info-data marginBottom10">63</dd>	<dd class="info-data marginBottom10">1 week ag...	<NA>	
3	[<h1 class="dInline marginRight10"> Joe Roecke...	<dd class="info-data marginBottom10"> <a class=...	<dd class="info-data marginBottom10">Outside H...	<dd class="info-data marginBottom10">1998-03-0...	<dd class="info-data marginBottom10">197cm</dd>	<dd class="info-data marginBottom10">97kg</dd>	mar
4	[<h1 class="dInline marginRight10"> Kyle Foley...	<dd class="info-data marginBottom10"> <a class=...	<dd class="info-data marginBottom10">Setter</dd>	<dd class="info-data marginBottom10">190cm</dd>	<dd class="info-data marginBottom10"><a class=...	<dd class="info-data marginBottom10"><a class=...	
...	...	...	...	...	...	...	
3223	[<h1 class="dInline marginRight10"> Andrew Nal...	<dd class="info-data marginBottom10"> <a class=...	<dd class="info-data marginBottom10">Outside H...	<dd class="info-data marginBottom10">1988-06-0...	<dd class="info-data marginBottom10">197cm</dd>	<dd class="info-data marginBottom10">87kg</dd>	mar
3224	[<h1 class="dInline marginRight10"> Andrew Hei...	<dd class="info-data marginBottom10"> <a class=...	<dd class="info-data marginBottom10"><a class=...	<dd class="info-data marginBottom10">Middle-bl...	<dd class="info-data marginBottom10">1984-07-0...	<dd class="info-data marginBottom10">211cm</dd>	mar
3225	[<h1 class="dInline marginRight10"> Oleksiy Gu...	<dd class="info-data marginBottom10"> <a class=...	<dd class="info-data marginBottom10">Opposite<...	<dd class="info-data marginBottom10">1984-04-2...	<dd class="info-data marginBottom10">195cm</dd>	<dd class="info-data marginBottom10">5667</dd>	mar
3226	[<h1 class="dInline marginRight10"> Maurice To...	<dd class="info-data marginBottom10"> <a class=...	<dd class="info-data marginBottom10"><a class=...	<dd class="info-data marginBottom10">Opposite<...	<dd class="info-data marginBottom10">1991-07-0...	<dd class="info-data marginBottom10">201cm</dd>	ma

Essentially the data sets I'm going to collect are going to require a lot of cleaning and a lot of soup.findall() to get what I need for web scraping.

NEXT STEPS

- The most important thing for me to focus on right now is web scraping and consolidating all the roster data from the D1 schools.
- There's 59 D1 programs with men's volleyball so I need to web scrape as much as possible.
- After I need to clean it all up and combine them all together into one data set.
- I should be able to convert most of the data I've collected into numerical data so I believe that using linear regression will be my best bet.