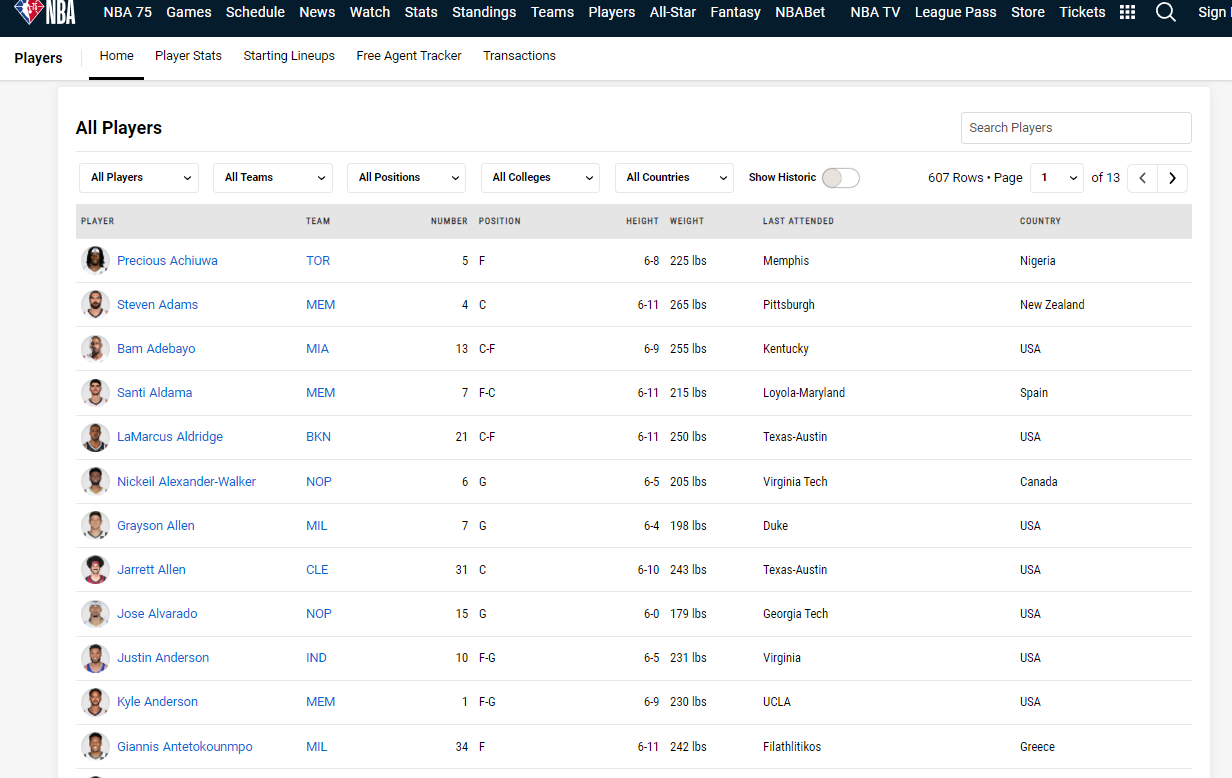
**תרגיל הגשה ב-python : web scraping – עידן חן**

בחרתי לעשות את תרגיל ה-web scraping על האתר הרשמי של ה-NBA.

המטרה: לחלץ את כל המידע על השחקנים ולשמור ב-CSV.

כשהגעתי לעמוד הראשי שמציג את כל השחקנים, הבנתי שחסר לי מלא נתונים.

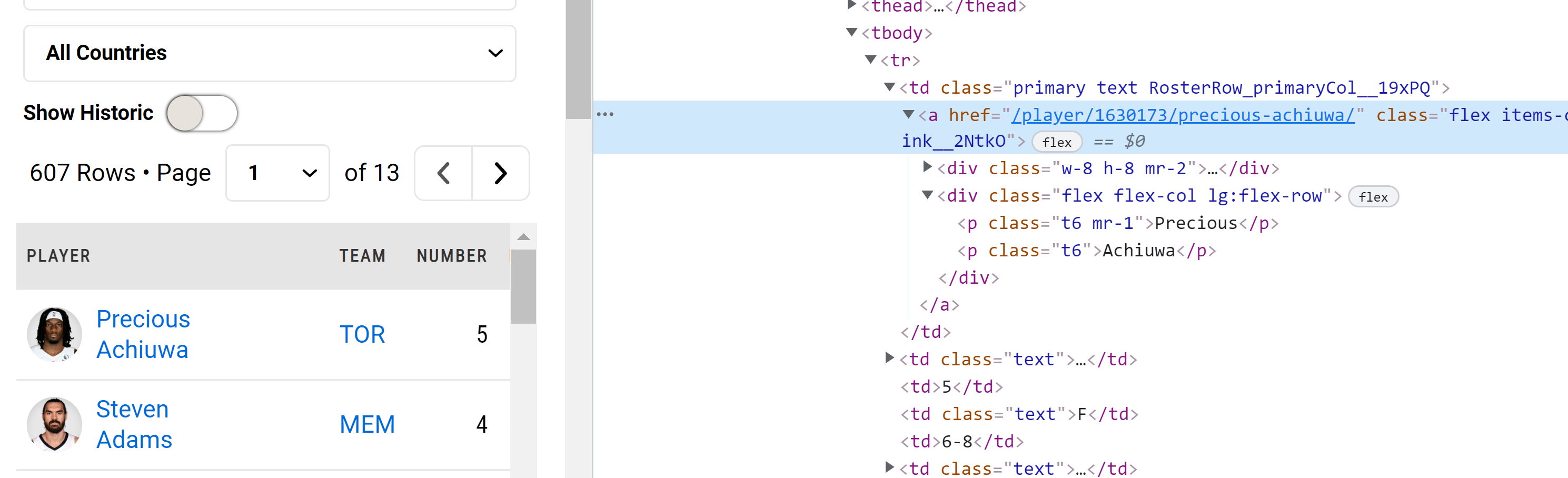
ראה עמוד מתחת:

חסרים לי נתונים כמו: כמה נקודות למשחק, ריבאונדים, ועוד נתונים שלא הכי רלוונטים כמו תאריך יום הולדת.

גיליתי שאם אני לוחץ על השם של השחקן (שמודגש בכחול) – אני נשלח לעמוד הבית של השחקן שמכיל את כל הנתונים שאני רוצה

אז הבנתי שאני חייב להגיע לעמוד הבית של כל שחקן,

נכנסתי אל דף ה-HTML של הטבלה של השחקנים ושמתי לב לדבר הבא:

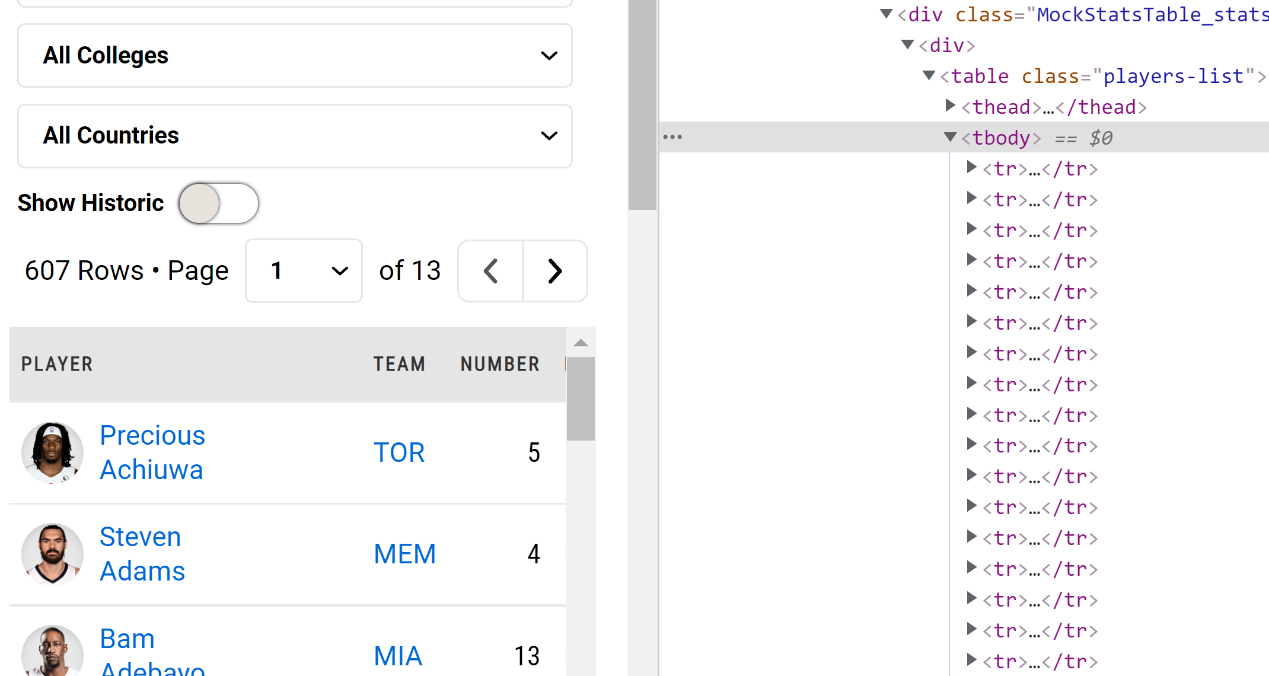


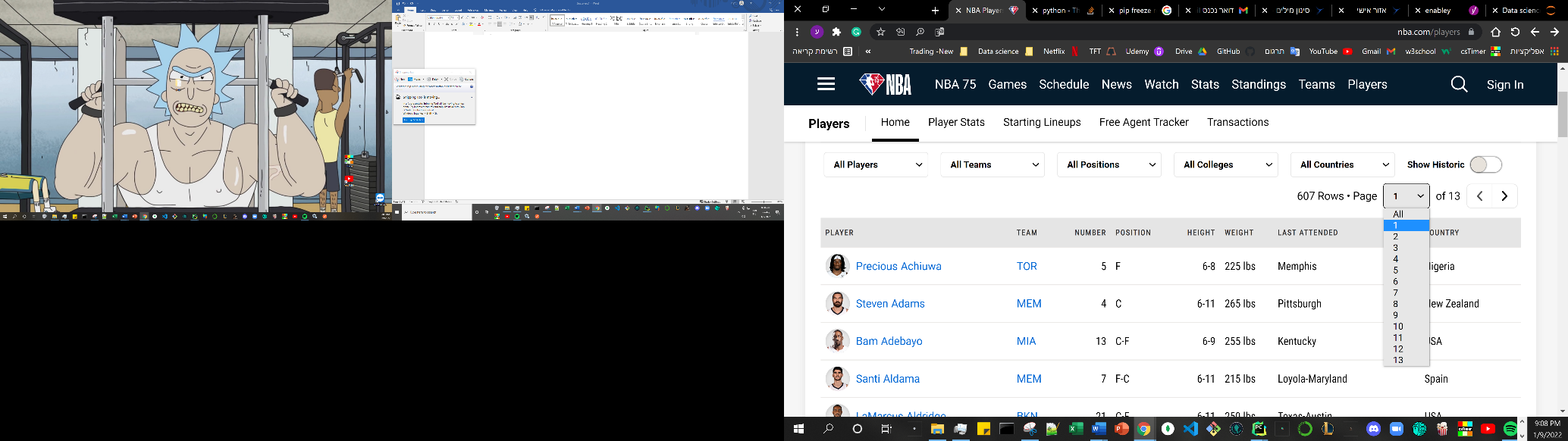
לכל שחקן יש href שמכיל את הקישור לעמוד הבית שלו,

אז הרעיון הכללי הוא ליצור מערך של קישורים, כל תא במערך הוא קישור

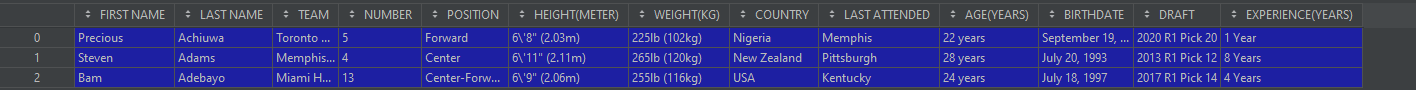
לעמוד הבית של שחקן.

חילצתי את הקישורים דרך ה-class\_name של ה-table (אין ל-tr שם שיכולתי להסתמך עליו) ובאמצעות הבנת הנתיב הוספתי xpath מתאים

וחילצתי את כל הקישורים לעמוד הבית של כל השחקנים

שמתי לב לדבר חשוב- לפני שאני מתחיל לקחת נתונים ממסך הבית, חשוב שאבחר בכפתור ה-Page בשדה ה-All על מנת שאקבל את כל השחקנים בעמוד אחד ולא אצטרך לעבור עמוד כל פעם – השתמשתי ב-Select ב-selenium ובחרתי את הערך הזה בתחילת ההרצה

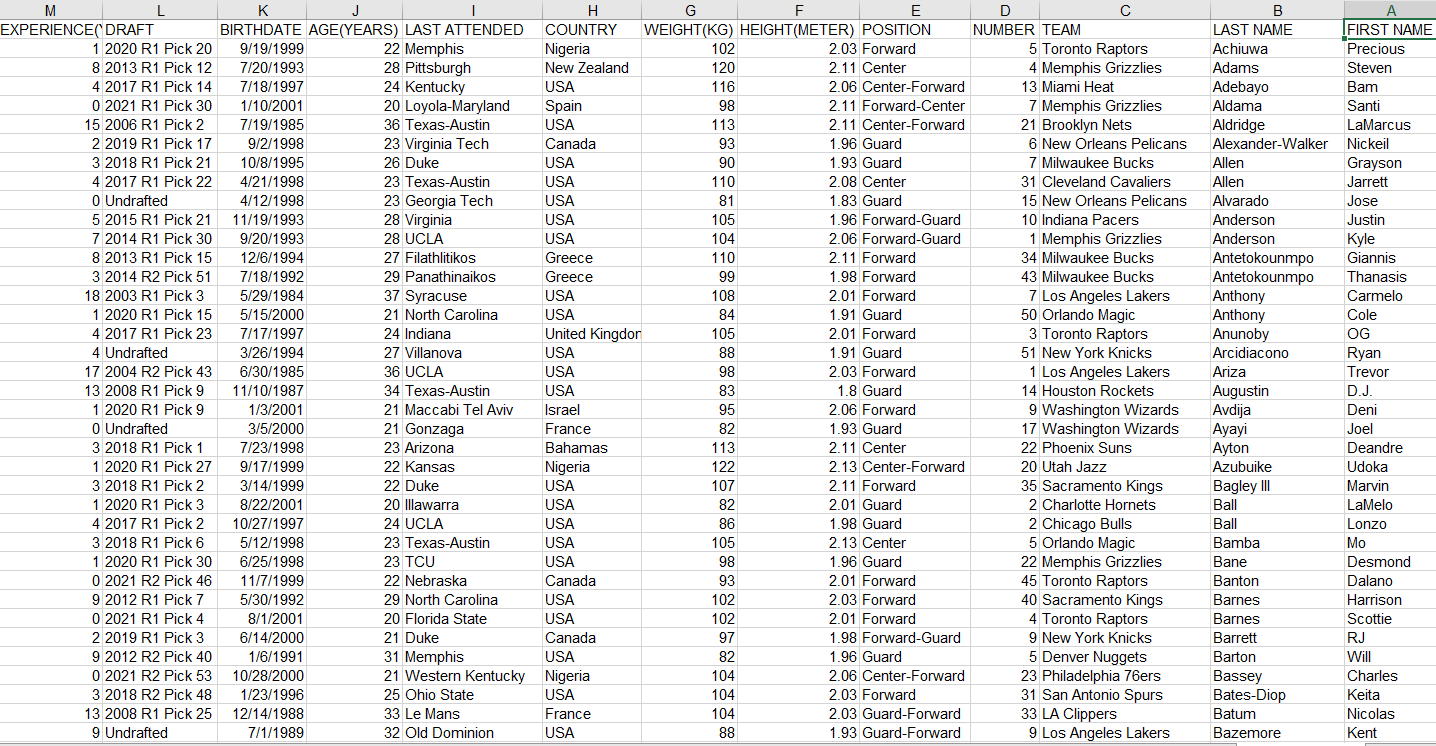
לאחר מכן התחלתי לרוץ על כל שחקן ולהכניס נתונים – אני מנוסה מאד ב-pandas ולכן השתמשתי ב-pandas.DataFrame על מנת לאחסן את הנתונים.

אחרי שסיימתי להכניס נתונים קיבלתי את התוצאה הבאה:

בסופו של דבר, המטרה הסופית היא שננתח את הנתונים ונבצע עליהם גרפים ונגיע למסקנות, כדי להגיע למטרה הזאת – צריך נתונים "נקיים" ושיהיה אפשר לבצע עליהם חישובים. ולכן אפשר לשים לב לכמה בעיות:

* HEIGHT : הוא טקסט שמכיל גם אינצ'ים וגם מטרים
* WEIGHT : הוא טקסט שמכיל גם ליברות וגם קילוגרמים
* AGE: טקסט שמכיל גם את הגיל וגם את המילה years
* EXPERIENCE: טקסט שמכיל גם את הגיל וגם את המילה years
* BIRTHDAY : טקסט שאמור לייצג תאריך אבל בפועל צריך להמיר אותו (הוא לא בפורמט תקין)

לשם כך באמצעות pandas ופונקציית ה-apply שהיא מכילה ביצעתי שינויים בכל עמודה בנפרד והגעתי לתוצאה הרצויה:

בסופו של דבר שמרתי את המסמך לקובץ CSV :

לסיכום, נהניתי מאד בתרגיל, אני שוחה ב-python אבל selenium אף פעם לא ניסיתי,

היה תרגיל כיפי ואני מאד מתחבר לכל נושא ה-data analyst and scientist ומתכנן

לעסוק בזה בעתיד.