

تحلیل داده‌های تنیس

بررسی عملکرد بازیکنان و
نتایج مسابقات ماه می ۲۰۲۴

نگار قیاسی فر
زمستان ۱۴۰۳



معرفی داده

این مجموعه داده از سایت [sofascore.com](https://www.sofascore.com) جمع آوری شده و شامل اطلاعات مربوط به بازیکنان تنیس، مسابقات، رده بندی و موارد دیگر می باشد.

هدف تحلیل

هدف این تحلیل بررسی تاثیر قد، راست دست یا چپ دست بودن، رده بندی و موارد دیگر در برد یا خطاهای بازیکنان می باشد. همچنین عملکرد بازیکنان نیز در این تحلیل بررسی شده است.



معرفی داده

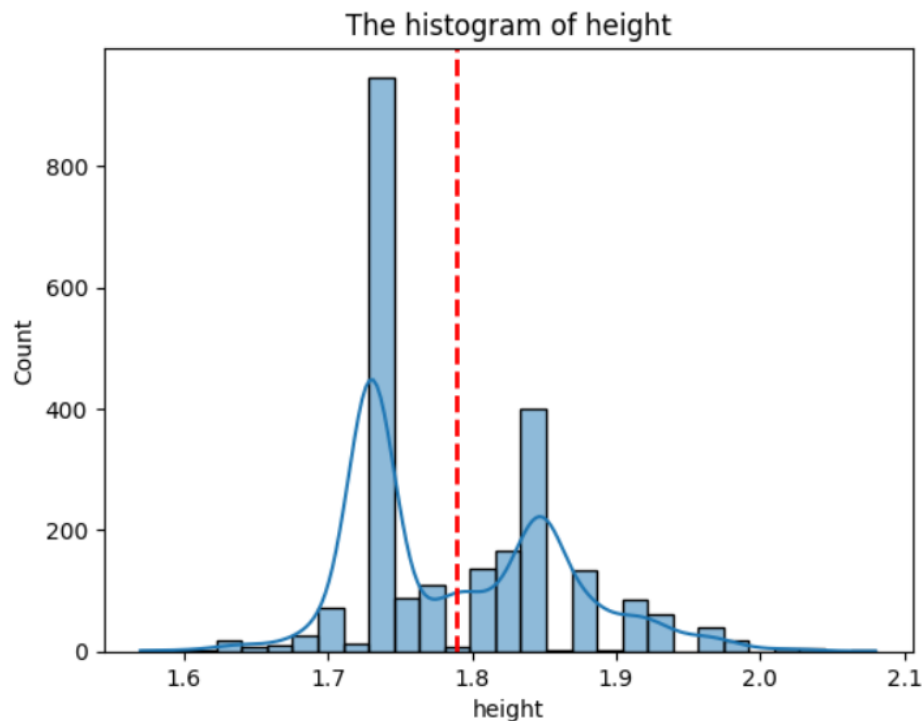
- داده از ۱۵ جدول حاوی چندین هزار سطر و ده ها ستون تشکیل شده است و هر جدول اطلاعاتی از بازیکنان، مسابقات، زمان، فصل و موارد دیگر را شامل می شود.
- ستون های مربوط به قد، جنسیت و راست دست یا چپ دست بودن و ... در این جداول شامل مقادیر از دست رفته می باشند که از چند روش برای مدیریت آنها استفاده شده است.
- ۱. پر شدن مقادیر خالی جنسیت بر اساس جنسیت بازیکن مقابل
- ۲. پر شدن مقادیر خالی قد از متوسط قد بر اساس جنسیت
- ۳. در نظر نگرفتن مقادیر خالی راست/چپ دست به دلیل حجم بالای آنها
- بسیاری از مقادیر تکراری نیز از تحلیل حذف شده اند تا موجب bias نشوند.
- در جداول مربوط به بازیکنان هر مسابقه (میزبان/مهمان)، گاهی یکی از بازیکنان یک مسابقه در جدول مربوطه موجود نبوده است. در تحلیل های مربوط به برنده بودن، تعداد ست ها و سایر موارد، تنها مسابقاتی در نظر گرفته شده اند که هر دو بازیکن آنها در جدول مذکور وجود داشته باشند.
- همچنین مسابقاتی که شروع نشدند یا به هر دلیلی کنسل شدند در تحلیلهای مربوطه قرار ندارند.



این دیتاست حاوی ۲۳۵۲ بازیکن تنیس می باشد.



میانگین قد بازیکنان تنیس

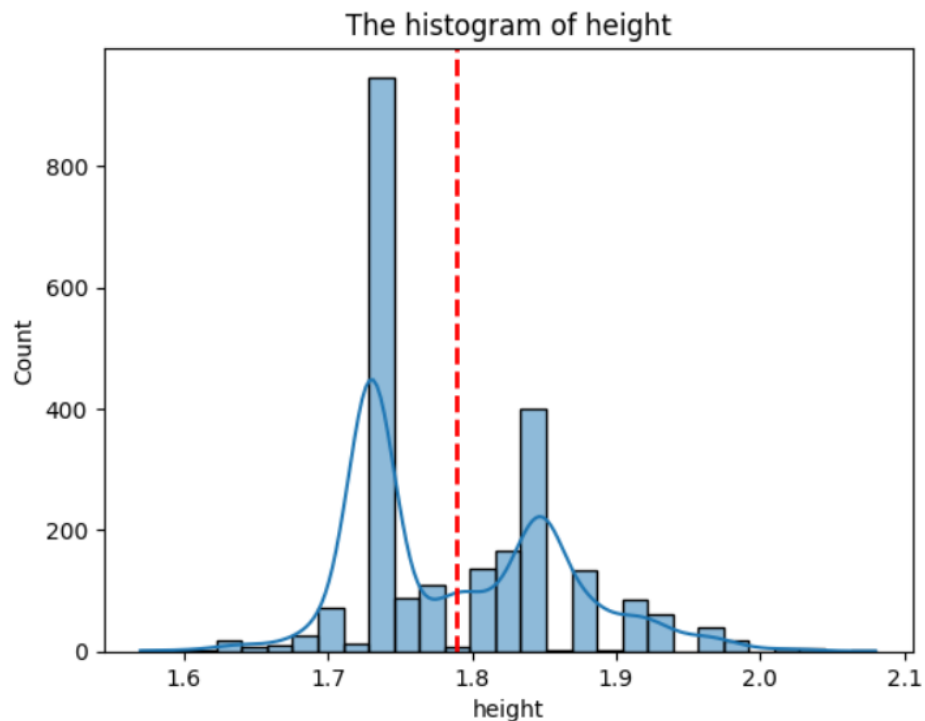


میانگین قد بازیکنان تنیس در این دیتاست ۱ متر و ۷۹ سانتی متر می باشد.

در نمودار زیر میانگین قد (قرمز رنگ) و فراوانی قد بازیکنان نمایش داده شده است.

بلندقدترین بازیکن این دیتاست Andrea Vavassori با قد ۲ متر و ۸ سانتی متر و کوتاه قدترین بازیکن Lauren Davis با قد ۱ متر و ۵۷ سانتی متر می باشد.

میانگین قد بازیکنان تنیس



چرا نمودار دارای دو قله می باشد؟

دلیل این اتفاق پر کردن مقادیر خالی قد بازیکنان از میانگین قد مردان (۱ متر و ۸۵ سانتی متر) و زنان (۱ متر و ۷۳ سانتی متر) می باشد.

برنده های ماه می



در ماه می بیشترین برد در دست ۵ بازیکن مشخص شده در جدول بود.

player_id	full_name	total_wins
253356	Uchijima, Moyuka	17
255428	Back, Dayeon	13
88310	Sherif, Mayar	12
163438	Sun, Fajing	12
137948	Chang, Hanna	12

بیشترین برد در ماه می ۲۰۲۴، در دست Moyuka Uchijima با ۱۷ برد می باشد.

بازه زمانی مسابقات



در این دیتاست طولانی ترین مسابقه به مدت یک روز و ۲۳ ساعت و ۳۲ دقیقه و ۳۳ ثانیه بوده است.

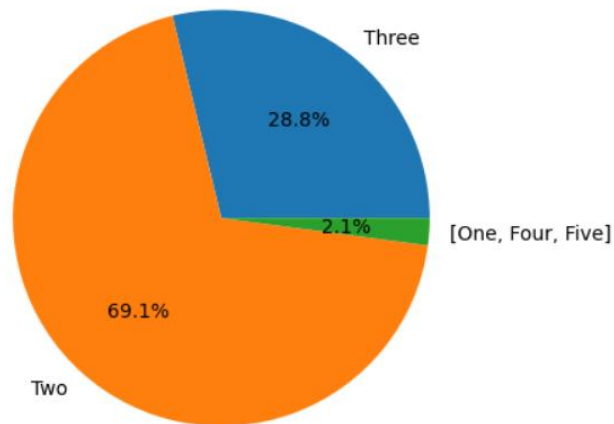
این مسابقه بین Gabriele Vulpitta و Samuele Pieri در تاریخ ۲۰ می ۲۰۲۴ برگزار شده است.

به طور میانگین مدت زمان هر مسابقه ۱ ساعت و ۵۹ دقیقه و ۲۰ ثانیه می باشد.

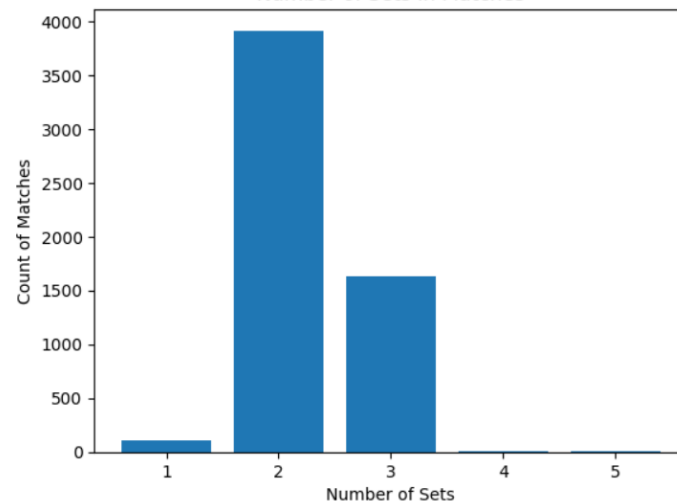
ست های مسابقات

در این دیتاست به طور میانگین بیشتر مسابقات تنها در دو ست به اتمام رسیدند.

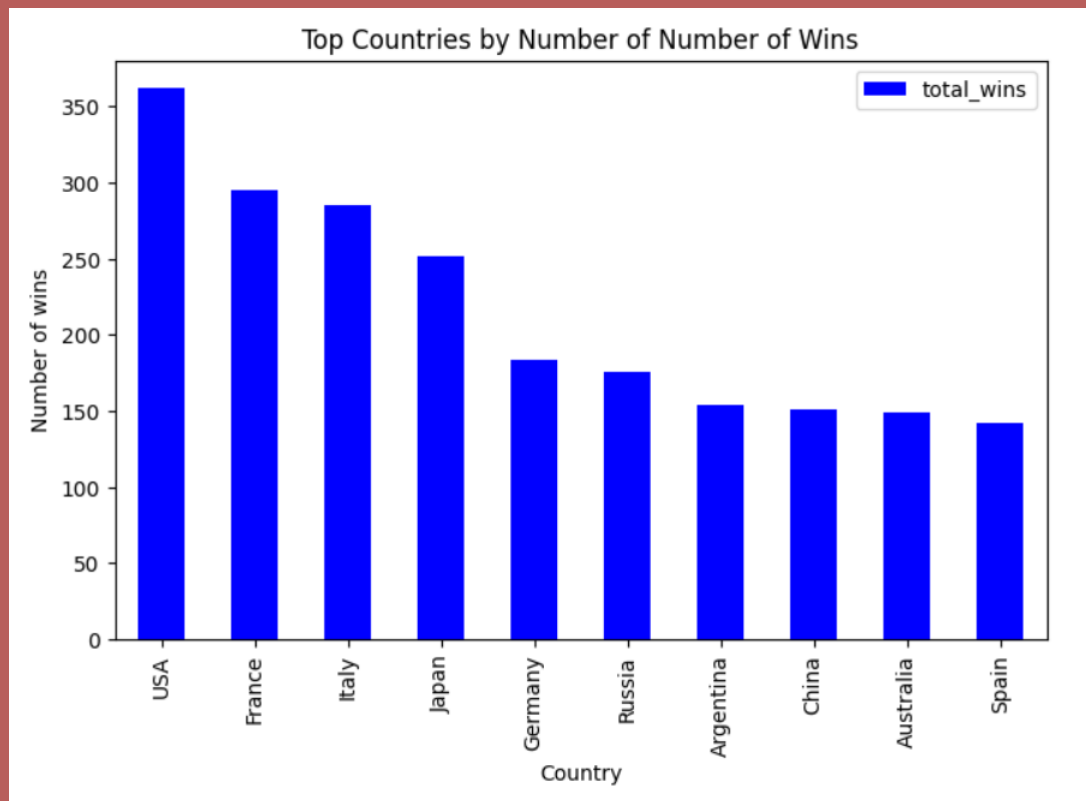
Distribution of Number of Sets in Tennis Games



Number of Sets in Matches

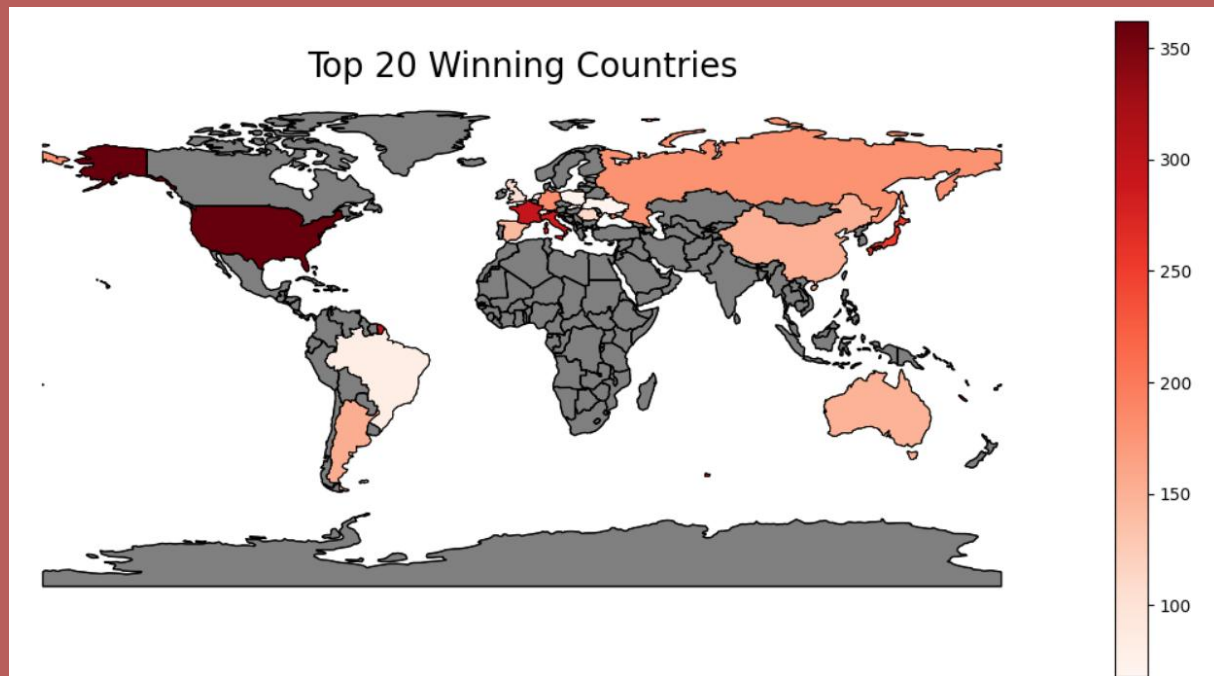


کشورها



در این دیتاست ۱۰۰ کشور
وجود دارد که کشور های آمریکا،
فرانسه و ایتالیا در سه مقام اول
بیشترین برد ها می باشند.

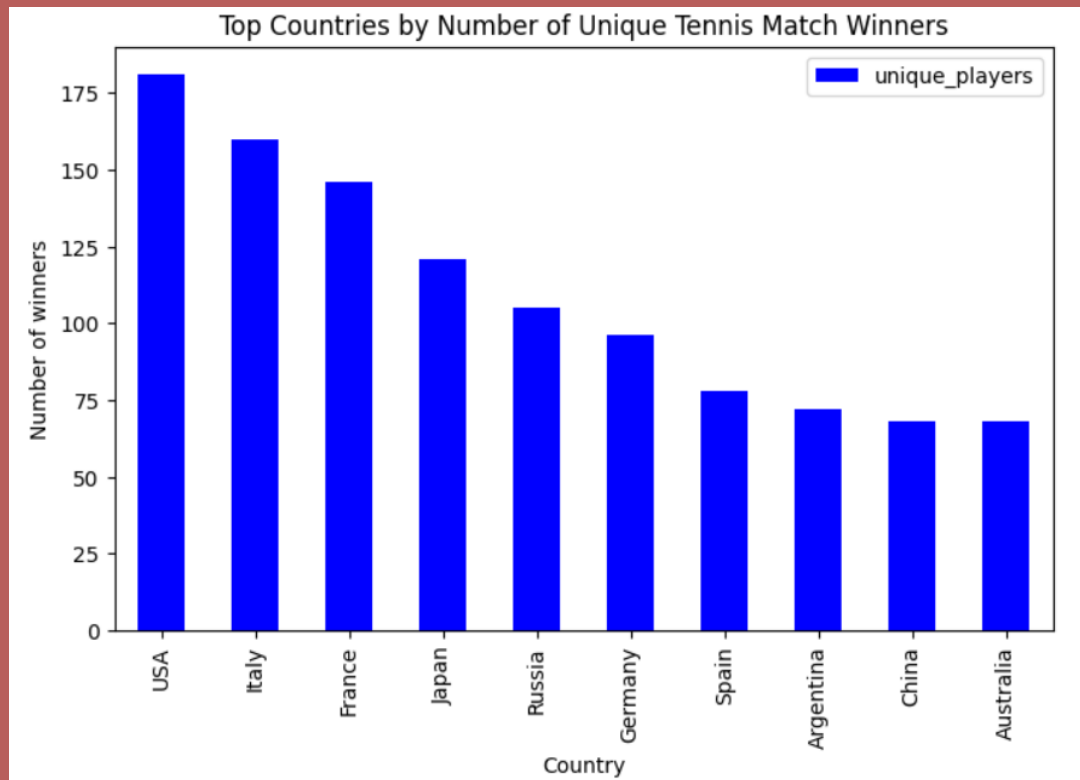
کشورها



در این نمودار ۲۰ کشور برتر از نظر تعداد بردها مشخص شده است.

تعداد برد هر کشور با رنگ مشخص شده است.

کشورها



در صورتی که تعداد برندگان هر کشور را در نظر بگیریم، مقام هر کشور مقداری جابجا خواهد شد اما به صورت کلی ۱۰ کشور اول در هر دو جدول یکسان می باشند.

این نشان می دهد که بازیکنان فرانسوی با وجود تعداد کمتر، پیروزی های بیشتری کسب کرده اند. این نشان دهنده عملکرد قوی تر یا موفقیت بیشتر بازیکنان برتر فرانسه باشد.

ایس های مسابقات

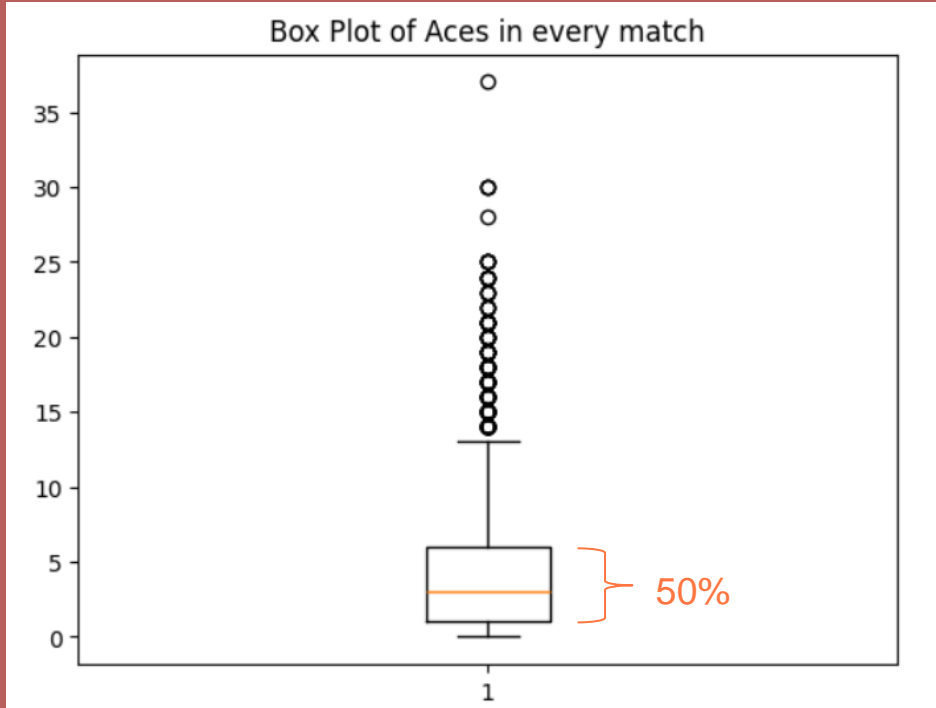


ایس چیست؟

زمانی که بازیکنی سرویس میزند و بدون دفاع توپ به زمین حریف برخورد میکند یک ایس رخ داده است.

بیشترین تعداد ایس در این مسابقات برای بازی بین Philip Ramanathan و Sekulic Ramkumar در تاریخ ۲۳ می ۲۰۲۴ با ۳۷ ایس بوده است.

ایس های مسابقات



در این نمودار توزیع داده ها و مقادیر پرت نمایش داده شده است.

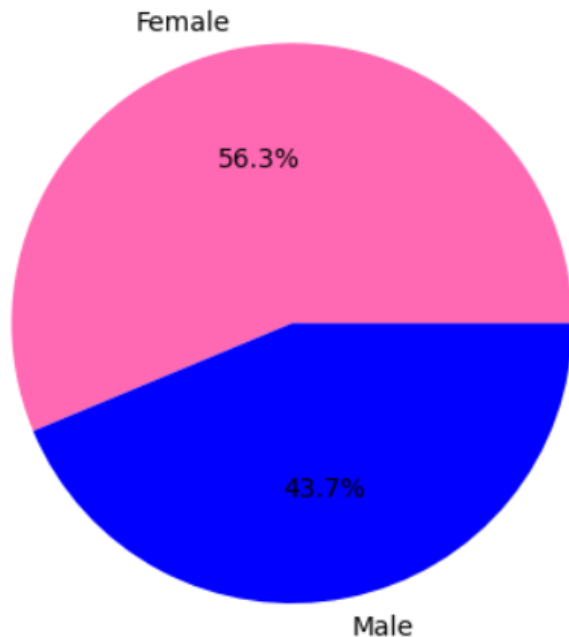
میانگین تعداد ایس ها در این دیتاست حدوداً ۴ می باشد.

۵۰٪ داده ها در باکس مشخص شده قرار دارند. یعنی ۵۰ درصد مسابقات بین ۱ و ۶ ایس داشتند.

بسیاری از مسابقات تعداد ایس های بسیار زیادی داشتند. این مقادیر بیشتر از ۱۳.۵ ایس هستند که با نقطه تو خالی مشخص شدند.

دابل فالت ها بر اساس جنسیت

Gender Distribution of double faulters in Tennis Games



دابل فالت چیست؟

اگر سرویس اول به تور برخورد کند یا در محدوده مجاز فرود نیاید، خطا شده و بازیکن مجاز به یک پرتاب دیگر است اما اگر پرتاب دوم نیز خطا شود بازیکن دچار دابل فالت (Double Fault) شده و یک امتیاز از دست خواهد داد.

در این دیتاست ۱۴۹۹۵ دابل فالت توسط خانم ها و ۱۱۶۴۵ دابل فالت توسط آقایان انجام شده است.

تورنمنت های ماه می



در این دیتاست با Miho Kuramochi
برد در ۷ تورنمنت بیشترین تعداد
تورنمنت را برده است.

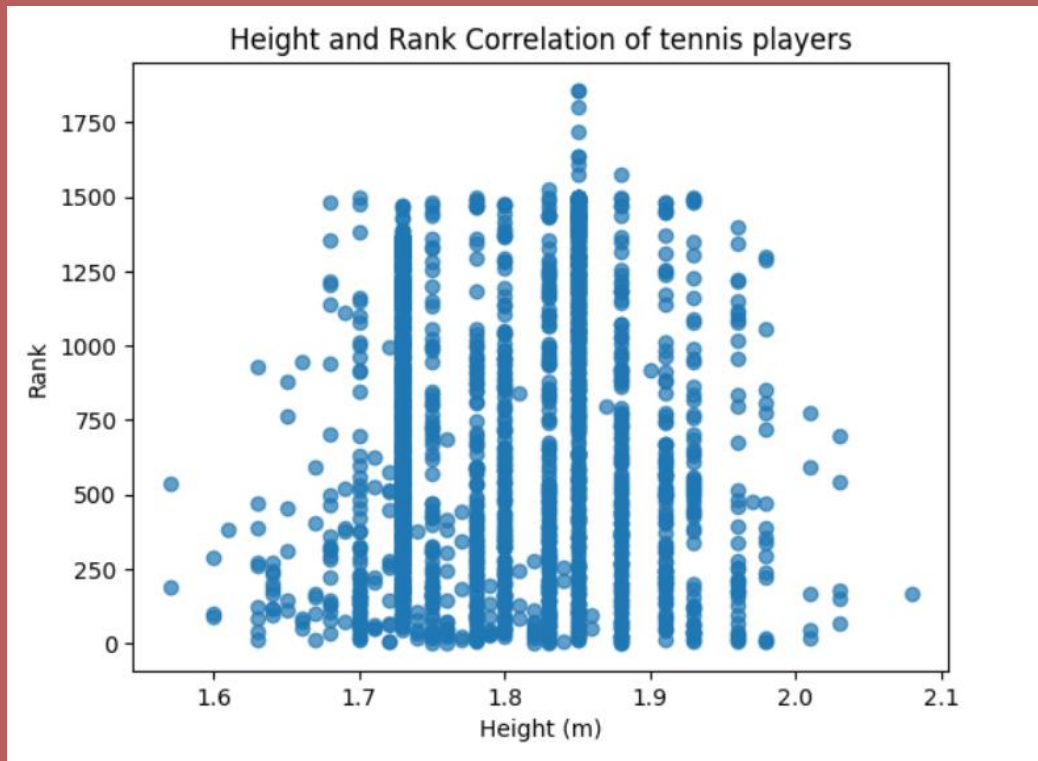
رابطه قد و رتبه بازیکنان

برای بررسی رابطه بین قد و رتبه از چند راه امکان بررسی وجود دارد.

طبق نمودار مقابل بسیاری از افراد با قد ۱.۷۵ تا ۱.۸۵ رنک بهتری دارند.

همچنین بازیکنان با قد یکسان رتبه های متعددی دارند.

اما به طور کلی رابطه خاصی بین قد و رتبه مشاهده نشد.



رابطه قد و رتبه بازیکنان

از سوی دیگر، طبق قضیه کوهن از کتاب تحلیل قدرت آماری برای علوم رفتاری (۱۹۸۸) نوشته جیکوب کوهن، در صورتی که شاخص همبستگی (correlation) در بازه‌های مشخصی باشد، میزان رابطه قد و رتبه تعیین می‌شود.

رابطه قوی : $|corr| \geq 0.5$

رابطه متوسط : $0.3 \leq |corr| \leq 0.5$

رابطه ضعیف : $0 \leq |corr| \leq 0.3$

طبق محاسبات ضریب همبستگی بین قد و رتبه بازیکنان حدوداً ۰.۰۱۵۱۷ می‌باشد که در بازه آخر جا دارد. در نتیجه رابطه ی ضعیفی بین قد و رتبه بازیکنان وجود دارد.

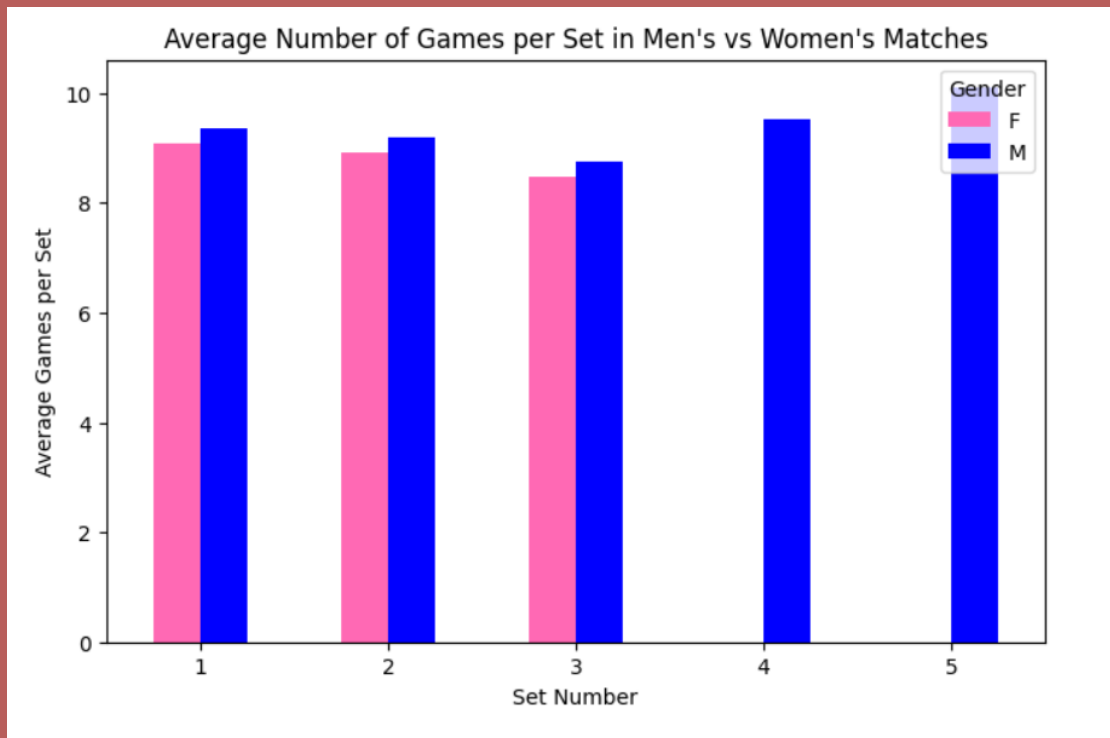
گیم ها مسابقات و جنسیت بازیکنان

ابتدا به مفهوم گیم می پردازیم:

در صورتی که بازیکنی ۴ امتیاز بدست آورد و حداقل ۲ امتیاز از حریف جلوتر باشد یک گیم برده است.
هر ۶ گیم یک ست را تشکیل می دهد و هر بازیکن برای بردن مسابقه باید ۲ ست را برنده شود.



گیم ها مسابقات و جنسیت بازیکنان



مسابقات خانم ها حداکثر ۳ ست و
مسابقات آقایان حداکثر ۵ ست
دارند.

اما میانگین گیم های آقایان و خانم ها
در سه ست اول تفاوت چندانی
ندارد.

برای ست ۴ و ۵ میزان گیم ها بیشتر
از قبل بوده است که نشاندهنده
رقابتی تر بودن مسابقات می باشد.

زمین مسابقات

به طور کلی برای هر تورنمنت ممکن است از جنس های مختلفی برای ساخت زمین تنیس استفاده کرد.

در این دیتاست زمین خاکی رسی با ۷۳.۳۶٪ رایج ترین نوع زمین تنیس می باشد.

پس از آن زمین هارد کورت در فضای باز با ۲۴.۸۱٪ و زمین های هارد کورت سرپوشیده و زمین مصنوعی فضای باز با ۱.۸۳٪ در رتبه های بعدی قرار دارند.



درصد برد در مقابل ۱۰ حریف رتبه اول

برای این سوال دو دسته از بازیکنان را می توان در نظر گرفت:

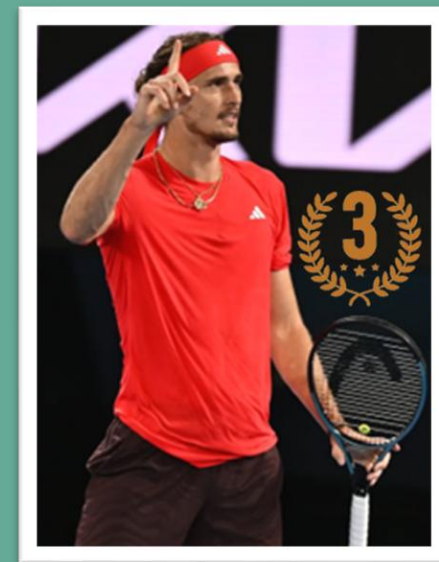
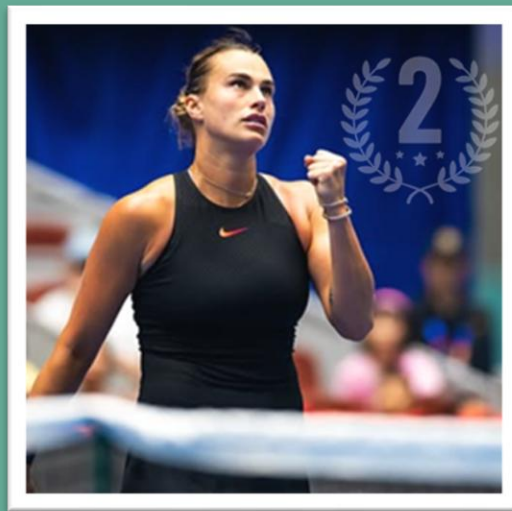
۱. بازیکنان برتر و بازیکنانی که با ۱۰ بازیکن برتر مسابقه داشتند.
۲. بازیکنانی که با ۱۰ بازیکن برتر مسابقه داشتند. (بدون در نظر گرفتن ۱۰ بازیکن برتر)

در این تحلیل از هر دو روش بررسی خواهیم کرد.



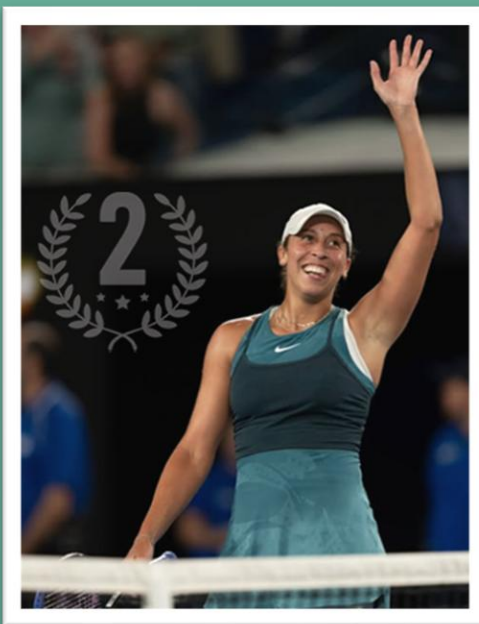
درصد برد در مقابل ۱۰ حریف رتبه اول

در صورتی که بازیکنان برتر را در تحلیل در نظر بگیریم ، Iga Swiatek با ۱۰۰٪ برد ۹ بازی،
Aryna Sabalenka با ۸۰٪ برد ۱۰ بازی،
Alexander Zverev با ۸۷.۵٪ برد ۸ بازی
در سه رتبه اول قرار دارند.



درصد برد در مقابل ۱۰ حریف رتبه اول

در صورتی که بازیکنان برتر را در تحلیل در نظر نگیریم ، Tommy Paul با ۱۰۰٪ برد ۲ بازی،
Madison Keys با ۵۰٪ برد ۴ بازی،
Sorana Cîrstea با ۱۰۰٪ برد ۱ بازی
در سه رتبه اول قرار دارند.

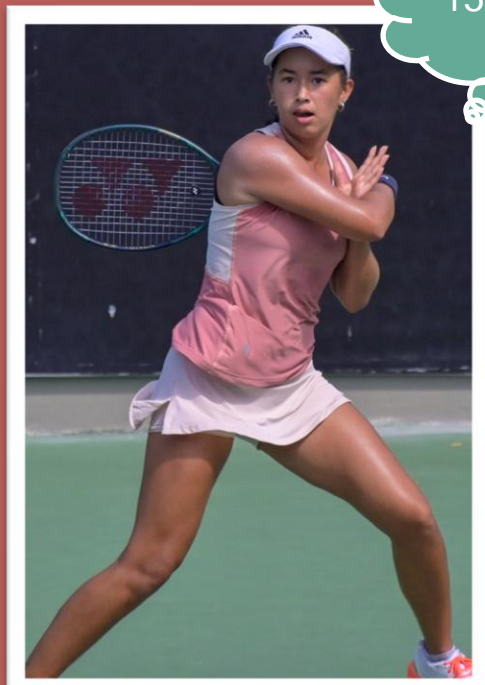


برنده سرویس ها

11 breaks!



15 breaks!

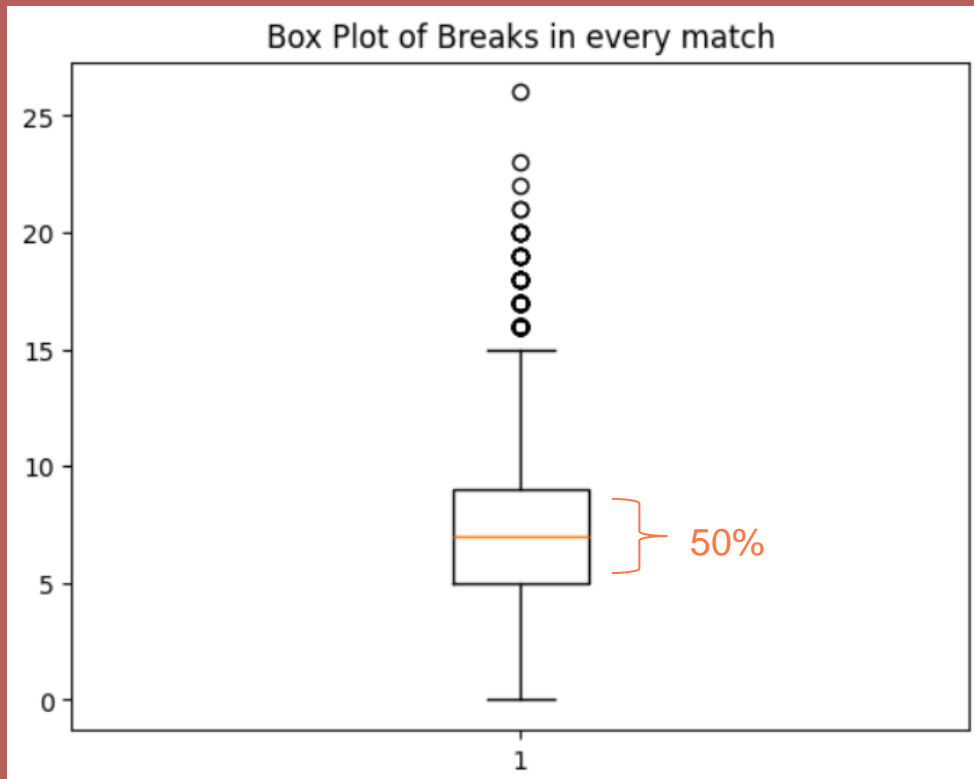


Break چیست؟

زمانی که بازیکنی در گیمی که حریف او سرویس زده است برنده شود، یک break انجام داده است.

بیشترین تعداد break ۲۶ عدد بوده که در مسابقه بین Ema Burgic و Higueta Barraza در تاریخ ۷ می رخ داده است.

برنده سرویس ها



در این دیتاست به طور میانگین ۷ break در هر مسابقه رخ داده است.

۵۰ درصد مسابقات بین ۵ و ۹ break داشتند.

بسیاری از مسابقات تعداد break های بسیار زیادی داشتند. این مقادیر بیشتر از ۱۵ break هستند که با نقطه تو خالی مشخص شدند.



پایان