



Yelp Gym Stars Predictions

מבוא למדעי הנתונים פרויקט גמר
316440221 אלון נאור

הקדמה

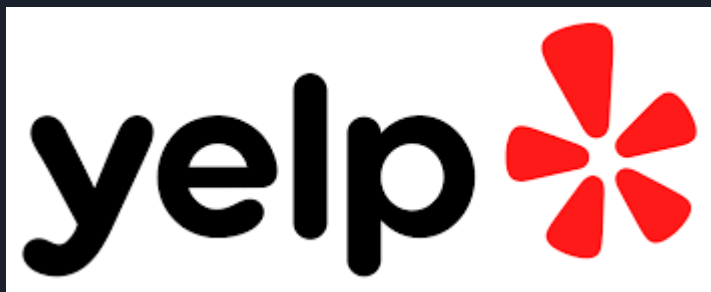


- בעידן המודרני 56% מבני האדם חיים בעיר מספר זה צפוי לגדול ל 70% עד סוף 2050
- העיר נותנת לאנשים הזדמנויות רבות יותר מאשר כפרים מגורים במקומות נחותים בנושאי עבודה, חינוך, רפואה וע
- החיסרון בערים הוא הצפיפות העצומה וזיהום אוויר ברח
- מה שמקשה על ביצוע פעילות ספורטיבית
- הפתרון ברוב המקרים הוא קניית מנוי לחדר כושר ובפרויקט זה ננסה לחזות את ציון חדר כושר

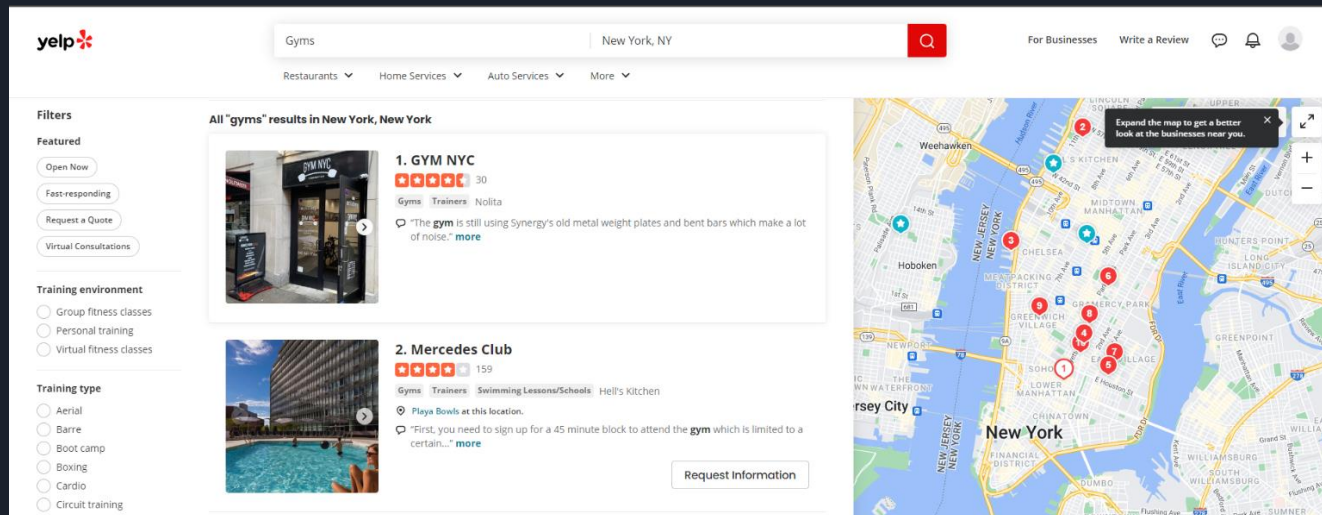


מקורות נתונים והרכשה

- נשתמש באתר www.yelp.com
- נחלץ מהאתר את העמוד של כל כושר
- נחפש חדרי כושר מהערים: ניו יורק, שיקגו, יוסטון, לוס אנגלס, פינקס, פילדלפיה וסאן אנטוניו
- לכל חדר כושר נחלץ את הפרטים שאחנו רוצים:
שם, דירוג, מספר תגובות, שעות פעילות ושירותים בחדר כושר



כאן אנחנו רואים דוגמא לחיפוש אתר YELP שמראה לנו תוצאות חיפוש בעיר ניו יורק. מהעמוד הזה נקלוט את הקישורים הרלוונטים לחדרי הכושר שלנו



The screenshot shows the Yelp website interface for searching gyms in New York, NY. The search bar at the top contains 'Gyms' and the location 'New York, NY'. The results are displayed in a list format, showing the top two gyms: '1. GYM NYC' and '2. Mercedes Club'. Each listing includes a photo, a star rating, the number of reviews, and a brief description. A 'Request Information' button is visible at the bottom right of the listings. On the left side, there are filters for 'Featured', 'Training environment', and 'Training type'. On the right side, there is a map showing the locations of the gyms in New York City.

Filters

Featured

- Open Now
- Fast-responding
- Request a Quote
- Virtual Consultations

Training environment

- ☐ Group fitness classes
- ☐ Personal training
- ☐ Virtual fitness classes

Training type

- ☐ Aerial
- ☐ Barre
- ☐ Boot camp
- ☐ Boxing
- ☐ Cardio
- ☐ Circuit training

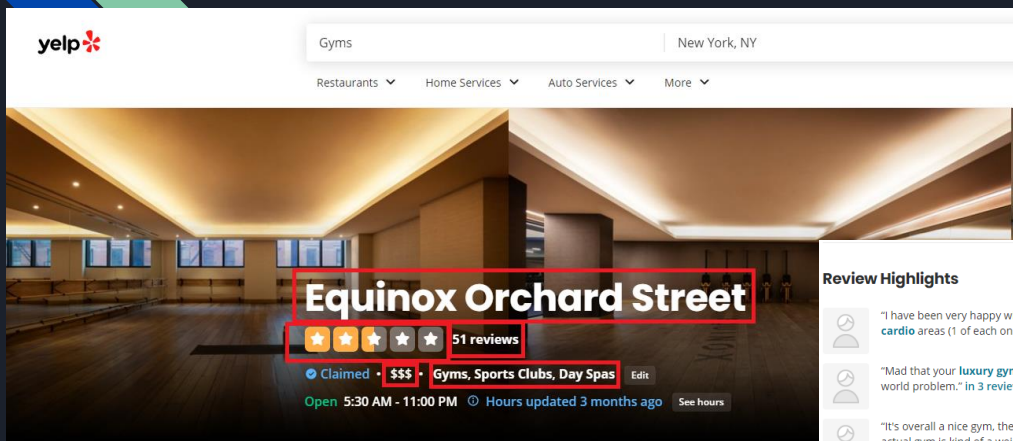
All "gyms" results in New York, New York

1. GYM NYC
★★★★★ 30
Gyms Trainers Nolita
"The gym is still using Synergy's old metal weight plates and bent bars which make a lot of noise." [more](#)

2. Mercedes Club
★★★★☆ 159
Gyms Trainers Swimming Lessons/Schools Hell's Kitchen
"Plays Bowls at this location."
"First, you need to sign up for a 45 minute block to attend the gym which is limited to a certain..." [more](#)

[Request Information](#)

אתר של חדר כושר נראה כך וניקח ממנו את הנתונים המסומנים



Review Highlights

"I have been very happy with my experience so far, everything is clean, and there are 2 weight and cardio areas (1 of each on the 2nd and 3rd floors)." [in 3 reviews](#)

"Mad that your luxury gym doesn't have wide open space for you to do your lunges, talk about a first world problem." [in 3 reviews](#)

"It's overall a nice gym, the steam room is amazing, really nice locker rooms, good classes, but the actual gym is kind of a weird set up and usually packed." [in 7 reviews](#)

(212) 501-1450

[Get Directions](#)

200 Orchard St New York, NY 10002

You Might Also Consider

Sponsored

Eye Level Learning Center

★★★★★ 5

0.6 miles

"I can't give a full review yet but when I took my son for the assessment the girl..." [read more](#)

CKO Kickboxing

★★★★☆ 70

3.2 miles

"Great atmosphere! Amazing community! Front desk workers are very friendly. The..." [read more](#)

Location & Hours

200 Orchard St
New York, NY 10002
Houston St & Stanton St
Lower East Side

[Get directions](#)

Mon	5:30 AM - 11:00 PM
Tue	5:30 AM - 11:00 PM
Wed	5:30 AM - 11:00 PM
Thu	5:30 AM - 11:00 PM
Fri	5:30 AM - 10:00 PM
Sat	8:00 AM - 8:00 PM
Sun	8:00 AM - 8:00 PM

[Open now](#)

[Edit business info](#)

יצירת ה DataFrame

```
{'name': 'GYM NYC', 'stars': 4.5, 'review_num': 30, 'expense': 0, 'activities': ['Gyms', 'Trainers'], 'Mon': 16.0, 'Tue': 16.0, 'Wed': 16.0, 'Thu': 16.0, 'Fri': 15.0, 'Sat': 8.0, 'Sun': 8.0, 'Gyms': 1, 'Trainers': 1, 'Swimming Lessons/Schools': 0, 'Coffee & Tea': 0, 'Acai Bowls': 0, 'Juice Bars & Smoothies': 0, 'Swimming Pools': 0, 'Yoga': 0, 'Cardio Classes': 0, 'Boxing': 0, 'Weight Loss Centers': 0, 'Circuit Training Gyms': 0, 'Physical Therapy': 0, 'Day Spas': 0, 'Sports Clubs': 0, 'Preschools': 0, 'Summer Camps': 0, 'Performing Arts': 0, 'Interval Training Gyms': 0, 'Boot Camps': 0, 'Massage': 0, 'Saunas': 0, 'Golf Lessons': 0, 'Climbing': 0, 'Cultural Center': 0, 'Gymnastics': 0, 'Nutritionists': 0, 'Child Care & Day Care': 0, 'Pilates': 0, 'Fitness/Exercise Equipment': 0, 'Barre Classes': 0, 'Tennis': 0, 'ROTD 9/28/2015': 0, 'Cycling Classes': 0, 'Martial Arts': 0, 'Day Camps': 0, 'Cryotherapy': 0, 'Cafes': 0, 'Pickleball': 0, 'ROTD 6/13/2019': 0, 'Tanning': 0, 'Health Coach': 0, 'ROTD 10/4/2016': 0, 'Kids Activities': 0, 'Rock Climbing': 0, 'Recreation Centers': 0, 'Bikes': 0, 'Spray Tanning': 0, 'ROTD 1/6/2019': 0, 'ROTD 4/18/2015': 0, 'Beer, Wine & Spirits': 0, 'Bars': 0, 'American (Traditional)': 0, 'Kickboxing': 0, 'Meditation Centers': 0, 'ROTD 11/19/2022': 0, 'Community Service/Non-Profit': 0, 'Hiking': 0, 'ROTD 2/24/2019': 0, 'Venues & Event Spaces': 0, 'Massage Therapy': 0, 'Community Centers': 0, 'Dietitians': 0, 'Sports Medicine': 0, 'Orthopedists': 0, 'Basketball Courts': 0, 'Brazilian Jiu-jitsu': 0, 'Tanning Beds': 0, 'Aerial Fitness': 0, 'Self-defense Classes': 0, 'ROTD 9/14/2021': 0}
```

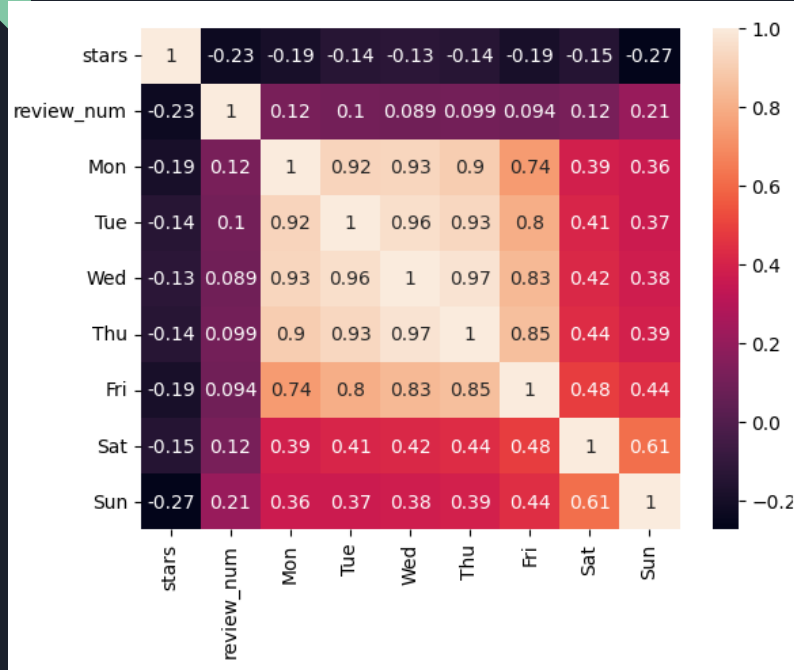
דוגמא למילון של חדר כושר אחד ב DataFrame
לא מסונן

	name	stars	review_num	expense	Mon	Tue	Wed	Thu	Fri	Sat	...	Meditation Centers	Community Service/Non-Profit	Hiking	Venues & Event Spaces	Massage Therapy	Community Centers	Dietitians
0	GYM NYC	4.5	30	0	16.0	16.0	16.0	16.0	15.0	8.0	...	0	0	0	0	0	0	0
1	Mercedes Club	4.0	159	0	16.5	16.5	16.5	16.5	15.5	12.0	...	0	0	0	0	0	0	0
2	Chelsea Piers Fitness	4.0	198	0	16.5	16.5	16.5	16.5	16.5	12.0	...	0	0	0	0	0	0	0
3	TMPL - Avenue A	4.5	26	0	17.5	17.5	17.5	17.5	16.5	12.0	...	0	0	0	0	0	0	0
4	Encore Fitness	5.0	12	0	16.0	16.0	16.0	16.0	16.0	10.0	...	0	0	0	0	0	0	0
...
607	Hays Communities YMCA	3.5	39	0	14.0	14.0	14.0	14.0	12.0	10.0	...	0	0	0	0	0	0	0
608	Gold's Gym - San Antonio	2.5	47	0	18.0	18.0	18.0	18.0	18.0	12.0	...	0	0	0	0	0	0	0
609	Gold's Gym - San Antonio Live Oak	2.0	62	0	18.0	18.0	18.0	18.0	18.0	12.0	...	0	0	0	0	0	0	0
610	Tru Fit Athletic Clubs - 410 & Bandera	2.5	62	0	7.0	7.0	7.0	7.0	17.0	11.0	...	0	0	0	0	0	0	0
611	STW Krav Maga	2.5	93	0	11.0	11.0	11.0	11.0	10.0	6.0	...	0	0	0	0	0	0	0

612 rows x 60 columns

אחרי סינון נקבל את ה DataFrame הזה

ויזואליזציה וEDA



מפת חום המציגה קשר בין כוכבים מס ביקורות

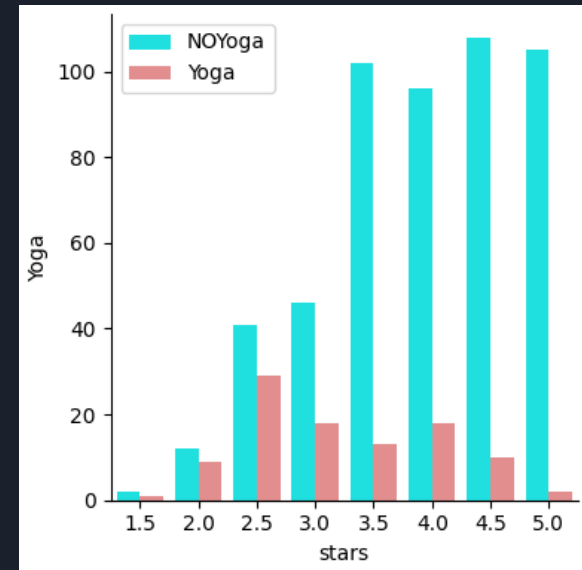
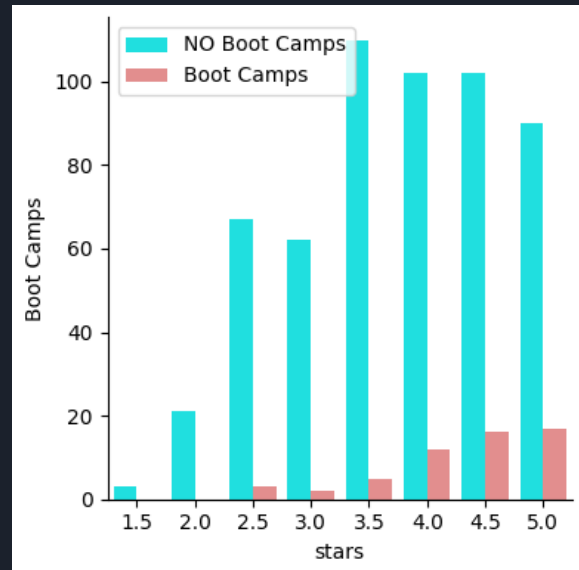
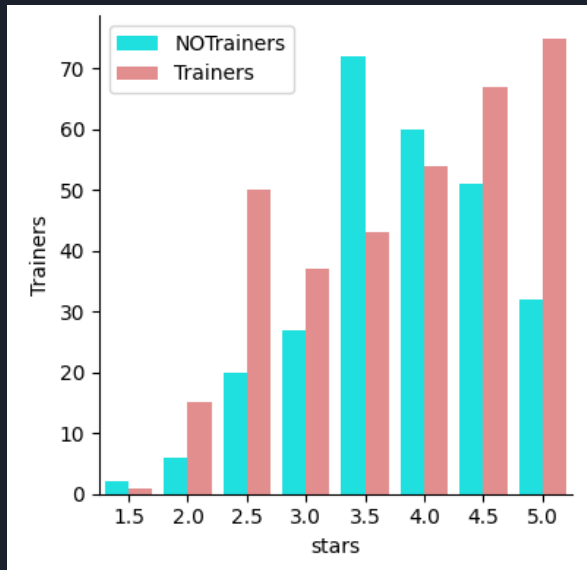
ושעות פעילות שבועיות

לצערנו לא מצאנו פה קישורים לעמודת הכוכבים



ויזואליזציה וEDA

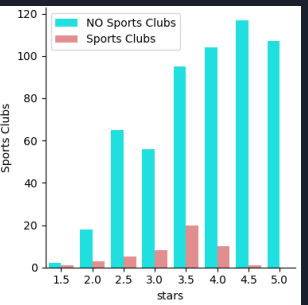
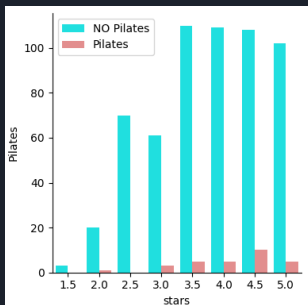
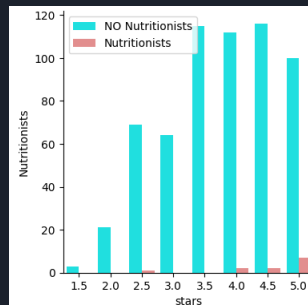
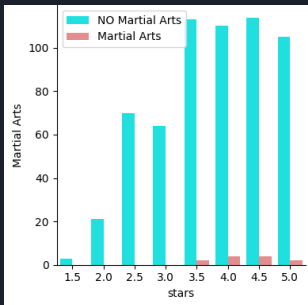
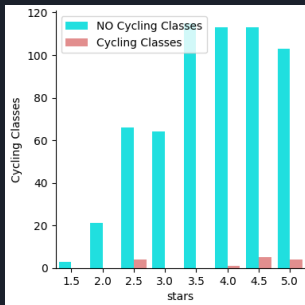
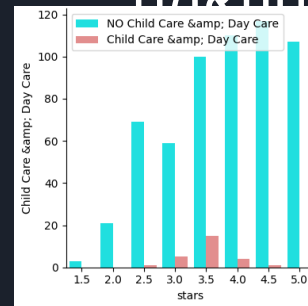
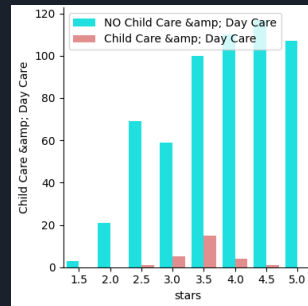
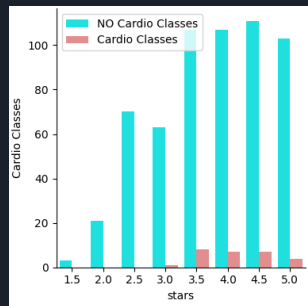
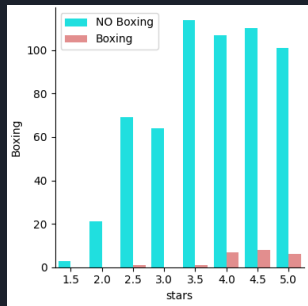
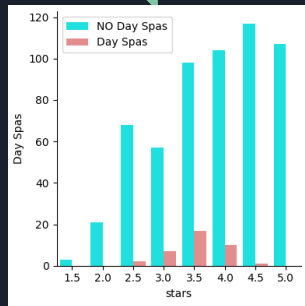
נסתכל על שירותים מיוחדים בחדרי כושר ודירוג חדר הכושר בהתאמה



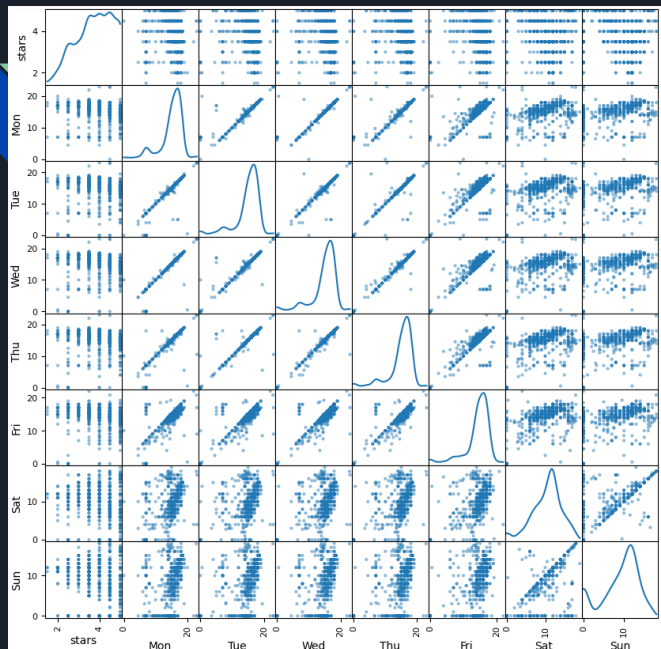
ויזואליזציה וEDA

נסתכל על שירותים מיוחדים בחדרי כושר ודירוג חדר הכושר

בהתאמה



למידת מכונה



הבעיה שלנו היא בעיית רגרסיה עם המון משתנים בינאריים (השירותים בחדרי ה

נבחר לעבוד בשיטת רגרסיה לינארית

קיבלנו $R^2: 0.30487$:

אחרי ניסיונות fitting קיבלנו $R^2: 0.2976$:

זהו ציון לא מספק

למידת מכונה

Decision Tree

0.24390



KNN

	k	train_accuracy	test_accuracy
0	1	1.000000	0.235772
1	2	0.611452	0.203252
2	3	0.533742	0.186992
3	4	0.546012	0.195122
4	5	0.552147	0.211382
5	6	0.496933	0.219512
6	7	0.472393	0.195122
7	8	0.451943	0.170732
8	9	0.431493	0.178862
9	10	0.449898	0.170732
10	11	0.437628	0.162602
11	12	0.433538	0.203252
12	13	0.415133	0.211382
13	14	0.421268	0.203252
14	15	0.408998	0.211382
15	16	0.384458	0.235772
16	17	0.384458	0.235772
17	18	0.394683	0.276423
18	19	0.392638	0.284553
19	20	0.396728	0.252033



סיכום ומסקנות

- מטרת הפרויקט הייתה לבדוק האם ניתן לחזות את דירוג חדר הכושר בהתאם לפרמטרים שלו
 - ניסינו להשתמש במספר הצגות ושיטות כדי להסתכל על הבעיה מנקודות מבט שונות
 - הגענו למסקנה ששילוב הנתונים עם שיטות למידת המכונה שלמדנו לא מספיקות
- על מנת להסיק מסקנה מהימנה עבור דירוג חדר הכושר