

דו"ח פרויקט מס' 4 אלעד ישראל ואמבט בלאצ'או

השאלה הנבחנת:

יש כמות מסוימת (קטנה) של "מצביעים" וכמות (עוד יותר קטנה) של פרוייקטים שאפשר לממן ומגבלת תקציב B.

השאלה היא, אילו פרוייקטים לממן כך שההחלטה תחשב "הוגנת"?

בניין עם N דיירים רוצה להחליט איך להשקיע את הכסף שנשאר בקופה של ועד הבית. כל דייר יצביע על k הפרוייקטים הכי מועדפים עליו. לכן המטריצה תהיה של $1 \setminus 0$ – רוצה את הפרוייקט או לא, תחת המגבלה שכל אדם מצביע לבדיוק K פרוייקטים. עכשיו, צריך להחליט – מה ההשמה ההוגנת ביותר?

נסתכל על 5 אפשרויות/הנחות לתועלת שמבוססות על סכום:

U1: מקסום כמות האנשים שיקבלו לפחות פרוייקט אחד שהם רצו.

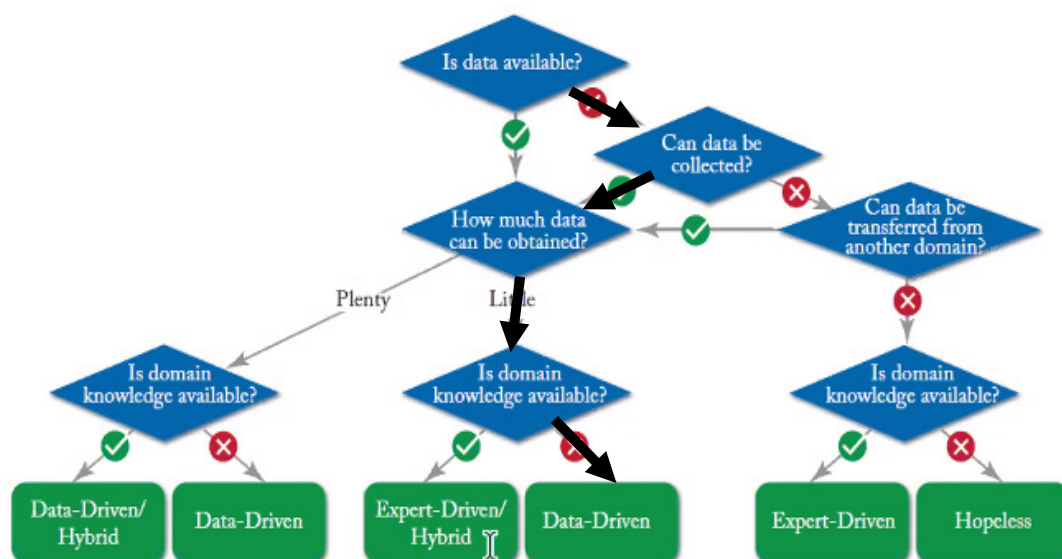
U2: מקסום סכום התועלות בהנחה שהתועלת של כל אדם היא כמות הפרוייקטים שקיבל מתוך הבחירה שלו.

U3: מקסום סכום התועלות בהנחה שהתועלת של כל אדם הוא עלות הפרוייקטים להם זכה מתוך הבחירה שלו.

U4: מקסום סכום התועלות בהנחה שהתועלת של כל אדם היא לפי SQRT של עלות הפרוייקטים להם זכה.

U5: מקסום סכום התועלות בהנחה שהתועלת של אדם היא עלות הפרוייקט הכי "טוב" שקיבל מהבחירה שלו.

תהליך הבחירה/בחינה של המודלים השונים:



אין ברשותנו מידע זמין המתאים לבעיה המתוארת, אך ניתן לאסוף מידע.

כמות המידע הניתנת לאיסוף הינה קטנה, שכן אנו מתבססים על סקר בהיקף מצומצם.

אין לנו מידע על הדומיין (וגם אם היה, כמו שלמדנו בהרצאה, transfer learning בדר"כ לא עובד טוב. בדומינים שונים הקבלת ההחלטות יכולה להיות שונה אפילו אם מדובר באותו אדם. הצבעה של אדם בועד הבית או בבחירות לממשלה למשל לא בטוח שדומות בכלל) או מידע ממומחה כלשהו ולכן בחרנו ב-Data driven נסתכל על ההחלטות של אנשים ומזה נסיק איך אנשים מקבלים החלטה באופן כללי.

	A	B	C	D	E	F
1	Pref Votes	P1	P2	P3	P4	P5
2	V1	0	1	0	0	0
3	V2	1	1	1	1	1
4	V3	1	1	1	1	1
5	V4	1	1	1	1	1
6	V5	1	1	1	1	1
7	Project Cost	90	10	45	50	99
8	Total Cost U1	90				
9	Total Cost U2	100				
10	Total Cost U3	100				
11	Total Cost U4	100				
12	Total Cost U5	90				
13	Budget	100				
14						
15	k	2				
16						
17						
18	Decision variables					
19	Assignment	P1	P2	P3	P4	P5
20	Max U1	1	0	0	0	0
21	Max U2	1	1	0	0	0
22	Max U3	1	1	0	0	0
23	Max U4	1	1	0	0	0
24	Max U5	1	0	0	0	0
25						
26						
27		U1	U2	U3	U4	U5
28	V1	0	1	10	3.162278	0
29	V2	1	2	100	12.64911	90
30	V3	1	2	100	12.64911	90
31	V4	1	2	100	12.64911	90
32	V5	1	2	100	12.64911	90
33	Max	4	9	410	53.75872	360
34						

תהליך איסוף המידע:

לשם איסוף המידע לסקר היינו צריכים לייצר דוגמאות שבהן כל בחירה אפשרית של פרוייקטים הממקסמת פונקציית תועלת כלשהי לא תעלה בקנה אחד עם הבחירות הממקסמות את הפונקציות תועלת האחרות.

בהתחלה ניסינו לייצר דוגמאות באקסל בעזרת שימוש Solver:

לאחר ניסיונות רבים ונשנים ראינו שזה קשה עד בלתי אפשרי למצוא דוגמא בצורה ידנית שבה אף בחירה של פרוייקטים הממקסמת פונ' תועלת מסויימת לא עולה בקנה אחד עם פונקציית תועלת אחרת.

לכן עברנו לכתיבת קוד:

תיאור האלגוריתם באופן כללי:

עד שלא נמצאו מספיק דוגמאות טובות:

1. לכל מספר פרוייקטים אפשרי וריאלי:

1.1. לכל מספר מצביעים אפשרי וריאלי:

1.1.1. לכל K אפשרי וריאלי:

1.1.1.1. בחר בצורה אקראית עלויות לפרוייקטים.

1.1.1.2. צור את כל החלוקות האפשריות לבחירת הפרוייקטים העומדות

בתקציב הנתון.

1.1.1.3. בצע 20 פעם:

1.1.1.3.1. צור מטריצת העדפות כאשר בכל שורה שים K אחדות

במיקום אקראי.

1.1.1.3.2. חשב את הפרוייקטים הממקסמים את התועלת לפי U1,

כנ"ל לפי U2 וכו'... הכנס את כולם לרשימה מאוחדת.

1.1.1.3.3. אם אין פרוייקטים כפולים ברשימה:

1.1.1.3.3.1. נמצאה דוגמא טובה! הדפס את הדוגמא.

הקוד עצמו מצורף בקובץ נפרד.

כך מצאנו לאחר כמה שעות הרצה מאות דוגמאות טובות. מתוכן בחרנו דוגמאות שנראו לנו קלות להבנה ושאינן גדולות מידי עד שלבסוף בחרנו 10 דוגמאות שאותם הכנסנו לסקר.

הסקר והדוגמאות מצורפים בקבצים נפרדים.

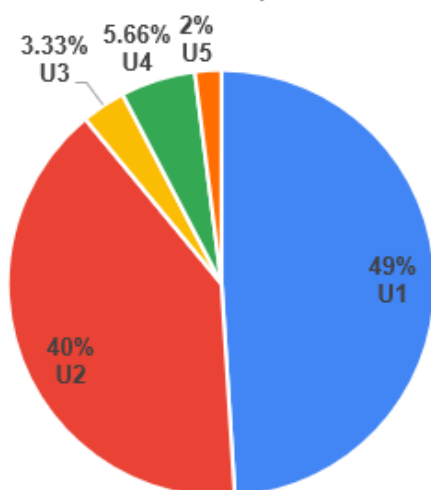
תיאור הסקר באופן כללי:

הסקר התחלק ל־3 חלקים:

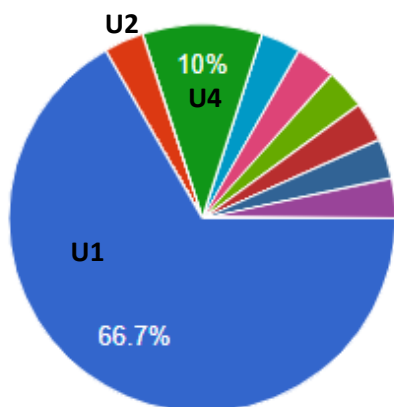
1. שאלות אישיות- לקבלת מידע על מאפייני האדם הנשאל.
 2. בהינתן מטריצת העדפות של כל דייר ועלויות של כל פרוייקט, הנשאל התבקש לבחור חלוקה שלדעתו נחשבת הכי הוגנת. מה שהנשאלים לא ידעו זה שכל חלוקה ייצגה מקסום של פונקציית תועלת אחרת. חלק זה נועד לתת לנו מושג על איזה פונקציית תועלת אנשים מנסים בפועל למקסם(בהתאם לחלוקות שראינו שנבחרו על ידם).
 3. ממלאי הסקר נשאלו: מה אתם מחשיבים כחלוקה "הוגנת"? מה הקריטריון שהנחה אתכם בבחירות שלכם? חלק זה נועד לתת לנו מושג על איזה פונקציית תועלת אנשים חושבים שהיא הוגנת, כלומר פונקציית התועלת שאמורה להנחות אותם בבחירה שלהם.
- על הסקר השיבו 30 נשאלים, ולכן (כנלמד בקורס שיטות סטטיסטיות עבור מדגם בעל 30 דגימות ומעלה) ניתן להניח שהוא מייצג התפלגות נורמלית בקרב האוכלוסיה.

עיקרי התוצאות:

העדפות בקרב כלל הנדגמים



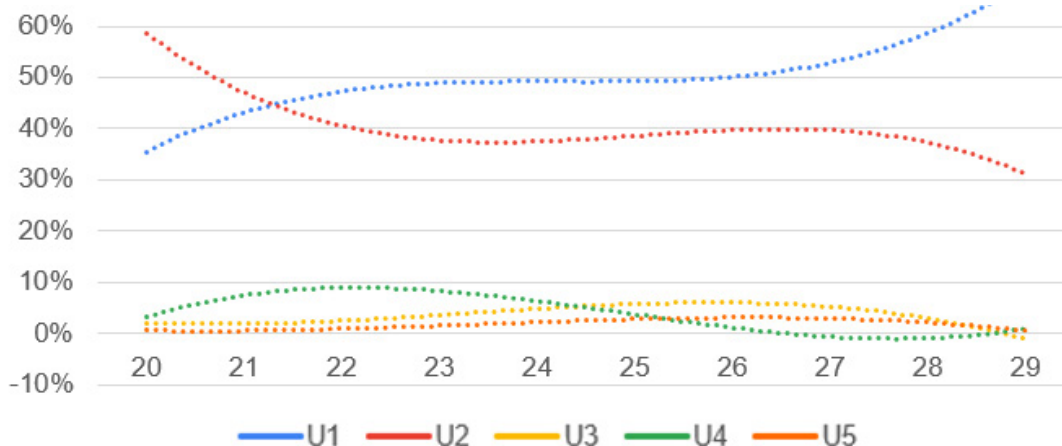
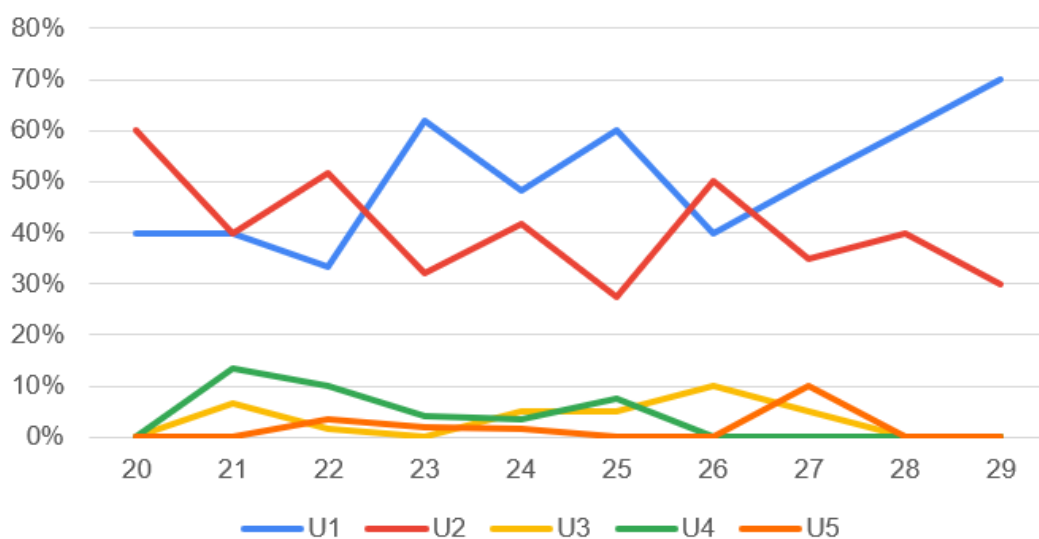
- יש העדפה ברורה ל U1 ו U2 בקרב המשיבים, עם יתרון ל U1:



- למרות ש-40% מהמשיבים בחרו בפועל לפי U2, כשנשאלו מה הקריטריון שהנחה אותם בבחירות שלהם רק 3% מהם ענו U2. כלומר למרות שהרוב המוחלט של המשיבים אמרו שהבחירה שנחשבת הכי הגיונית מבחינתם היא U1, בפועל בבחירות שלהם הם לא תמיד פעלו באופן רציונלי והעדיפו לא פעם את U2 על פני U1.

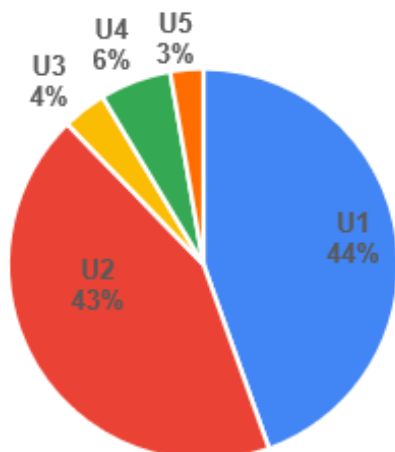
- ככל שגיל הנשאל עולה- יש לו יותר נטייה להעדפת U1, ופחות נטייה להעדפת U2. בהערכה גסה, לכל שנה נוספת- ההעדפה ל U1 גדלה ב-3.5% וההעדפה ל U2 קטנה ב-3%.
כמות המשיבים התפלגה א-סימטרית ימנית סביב 23 ולכן עבור גיל 26 ומעלה אין הרבה משיבים ולכן יש לקחת את הנתונים בעירבון מוגבל.

אחוזי בחירה בכל פונקציית תועלת כתלות בגיל

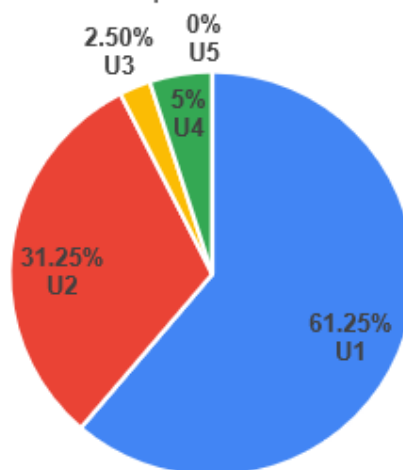


- גברים נוטים להעדיף את U1 על פני U2. גברים נוטים להעדיף את U1 ב-17% יותר מנשים, ואת U2 ב-12% פחות מנשים.

פילוח העדפות בקרב נשים

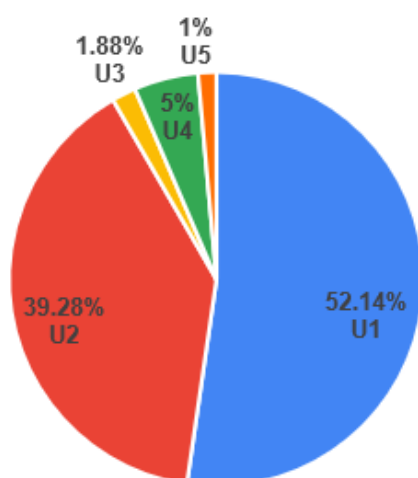


פילוח העדפות בקרב גברים

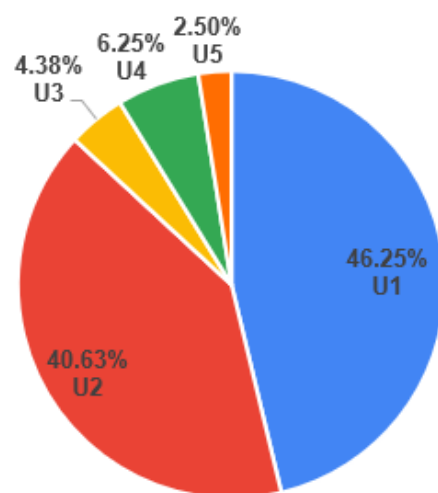


- לא ניכר הבדל בין העדפותיהם של משיבים המשלמים דמי ועד בית לעומת אלה שלא.

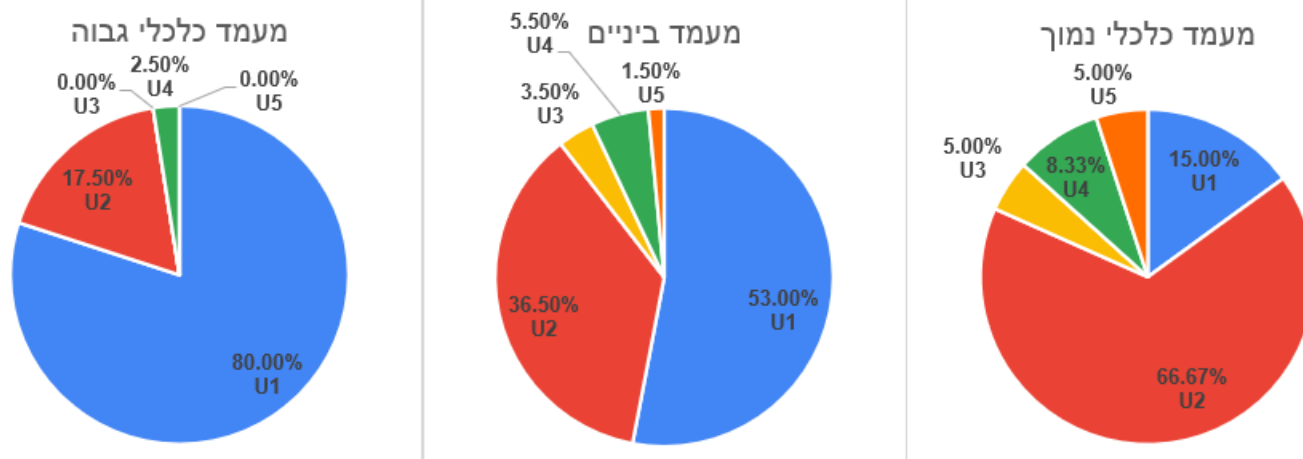
העדפות המשלמים לועד בית



העדפות לא משלמים לועד בית



- ככל שהמעמד הכלכלי של הנשאל גבוה יותר- יש לו יותר נטייה להעדפת U1, ופחות נטייה להעדפת U2.
מעמד כלכלי גבוה נוטה להעדיף את U1 ב-27% יותר ממעמד הביניים ומעמד כלכלי נמוך נוטה להעדיף את U1 ב-38% פחות ממעמד הביניים.
מעמד כלכלי גבוה נוטה להעדיף את U2 ב-19% פחות ממעמד הביניים ומעמד כלכלי נמוך נוטה להעדיף את U2 ב-30% יותר ממעמד הביניים.



לא היה ניתן להסיק על אופן השפעת רמת ההשכלה של אדם על הקריטריון הנבחר על ידו מכיוון שהנדגמים היו מאותה רמת השכלה.

המודל:

מכיוון ופונקציות התועלת U3, U4, U5 כמעט ולא נבחרו- לא ניתן להסיק מסקנות לגביהן בצורה מובהקת.

לכן התרכזנו בבניית מודל לחיזוי ההתפלגות עבור U1 ו-U2.

בחרנו לבנות מודל בעזרת נוסחא הנבנית על בסיס הנתונים שאספנו ולא לנסות לאמן מודל בעזרת למידת מכונה מכיוון שמספר הדוגמאות שברשותנו קטן ביחס למספר הדוגמאות הנדרש כדי לאמן מודל בעל שיעור הצלחה גבוה (ניסינו בכל זאת בעזרת כלי Data Mining של Frontline Excel Solvers לבדוק היתכנות למודל supervised שיהיה טוב על המידע הקיים ולא ראינו תוצאות מבטיחות כלל).

בהינתן אדם ומאפייניו:

הסיכוי שבחירתו תתואר ע"י קריטריון U1:

$$49\% + (\text{age} * 3.5\%) + (\text{gender}[\text{male}=1, \text{female}=0] * 17.25\%) + (\text{economical class}[\text{high}=1, \text{medium}=0, \text{low}=-1.4] * 27\%)$$

הסיכוי שבחירתו תתואר ע"י קריטריון U2:

$$40\% - (\text{age} * 3\%) - (\text{gender}[\text{male}=1, \text{female}=0] * 11.75\%) - (\text{economical class}[\text{high}=1, \text{medium}=0, \text{low}=-1.57] * 19\%)$$

נתייחס לגיל 20 כ age=0, גיל 21 כ age=1 וכו'. ניתן להסיק מסקנות בעזרת age אך ורק כאשר גיל הנבדק בין 20 ל 29.

דיון ומסקנות:

- ראינו שהקריטריון המנחה את רוב האנשים כאשר הם חושבים על חלוקה הוגנת הוא U1- כלומר שכמה שיותר דיירים יקבלו לפחות פרויקט 1 שהם רצו. כלומר חלוקה שבה יהיה דייר שלא קיבל אף פרויקט שהוא רצה נחשבת למאוד לא הוגנת בעיני אנשים.
- ראינו שלמרות עובדה זו, באופן לא רציונלי אנשים רבים בוחרים בחלוקות המתאימות ל-U2- כלומר שכל דייר יקבל כמה שיותר פרויקטים מתוך הבחירה שלו, גם אם יהיה דייר שלא קיבל אף פרויקט מהבחירה שלו בעקבות כך.
- ראינו שמאפייני אדם מסוימים נוטים להשפיע על קבלת ההחלטות שלו: אנשים בעלי אחד או יותר מהמאפיינים הבאים: מבוגרים יותר, גברים, ממעמד כלכלי גבוה - נוטים לבחור לפי U1. אנשים בעלי אחד או יותר מהמאפיינים הבאים: אנשים צעירים יותר, נשים, ממעמד כלכלי נמוך - נוטים לבחור לפי U2. לעומת זאת ראינו שישנם מאפיינים שאינם משפיעים בצורה ניכרת על קבלת ההחלטה, לדוגמא תשלום ועד בית.
- בנינו מודל המסוגל לחזות בקירוב את בחירת האדם בהינתן המאפיינים שלו. המודל נבנה ע"ב מדגם המייצג את האוכלוסייה ולכן הוא חוזר בצורה טובה(גם אם לא מושלמת) את ההחלטות שאותם יקבל אדם שאינו במדגם.