

פרויקט "צרכנים רציונאליים"

מגישים: שגית אפרת, גל דנינו ועדי רגב

תוכן עניינים

1.....	הקדמה
2.....	תיאור הפרוייקט
2.....	שלב ראשון – Web crawling - שליפת המידע מאתרי הרשתות
3.....	שלב שני- סידור הנתונים מ XML ל CSV
3.....	שלב שלישי- חלוקת מוצרים לקטגוריות
4.....	שלב רביעי- Clustering
8.....	שלב חמישי- תוצאות וניתוח הנתונים
9.....	רציונליות הצרכן
10.....	הוראות הפעלה
12.....	נספחים
12.....	קטגוריות של ירקות:
12.....	מילון הקטגוריות:
15.....	תוצאות רציונליות - פלט:
17.....	references

הקדמה

פרויקט זה בוצע במסגרת קורס "בחירה חברתית" בטכניון בפקולטה להנדסת תעשייה וניהול.

בפרויקט בחרנו להתמקד בתחום הרלוונטי לחיי היום-יום, והוא קניות בסופרים. האם אנחנו קונים בסופרים שמשלמים לנו? האם יש סופר אחר אשר המחירים עבור סלי המוצרים שלנו נמוכים בו יותר? האם בסופר אחד משתלם יותר לקנות מוצרי פרימיום ובאחר מוצרים בסיסיים?

בחרנו 20 סופרים באיזור חיפה, והטכניון בפרט, מרשתות שונות, שעליהם נבצע את ניתוחי המידע.

מטרת הפרוייקט הינה לבחון את רמת הרציונליות של הצרכן בבואו לבחור סופר בו יבצע את רכישותיו. נעשה זאת באמצעות דגימת סלי קניות שנרכשו בסופרים והסל המדגמי של המועצה הישראלית לצרכנות, עליהם נשפוט האם בחירת הסופר בו קנה היא המשתלמת ביותר עבורו. על מנת שבחירה תהיה מחושבת, ראשית צרכן נדרש להשתמש בכל המידע הרלוונטי, שנית, חייבת להיות לו היכולת לעבד את המידע על מנת להגיע להחלטה מושכלת. מרבית הצרכנים לא מבצעים את ה"מחקר" הדרוש. מעבר לדרישות אלה

לכל צרכן צפויות להיות העדפות שאינן כלולות במחיר הסל כגון קרבה לסניף, עיצוב ונוחות, מגוון, אורך התורים, זמינות, כו'. הנחת העבודה שלנו היא שצרכן עשוי להתקבע על סופר מסוים על ידי קירוב אינטואיטיבי למדדים הנ"ל.

אנו מעוניינים לחפש "רציונל חלקי" לבחירה של סופר הנובע מחיפוש של סופרים המתמחים בסוגי המוצרים אותם הצרכן אוהב, למשל "קפה איכותי", "ירקות" או "חלב פשוט".

על מנת לבצע ניתוחים אלה, אנו נחקור את הסופרים, נעשה בחינה של כל סופר שנבחר לפי הקטגוריות שיוגדרו. נבצע clustering של המידע, וממנו נסיק עבור כל קטגוריה מה הסופר שהמוצר הכי זול בו, ועבור כל סופר מה הקטגוריה שהוא הכי זול בה ביחס לאחרים.

תיאור הפרוייקט

שלב ראשון – Web crawling - שלילת המידע מאתרי הרשתות

על פי תקנות קידום התחרות בענף המזון (שקיפות מחירים), התשע"ד, רשתות השיווק מפרסמות את רשימות המחירים ברשת, אך עושות זאת בצורה שאינה נגישה לצרכן הממוצע. רב הקבצים מפורסמים בפורמט XML הקשה לפענוח וקשה להשוואה בין הסופרים, הנתונים של אותו הסופר מחולקים בין קבצים שונים (קובץ מחירים מלאים, קובץ למבצעים וכדומה). לכן, צרכן החפץ לעשות קניה חכמה ולחקור מהו המחיר הנמוך ביותר עבורו על ידי השוואה בין הסופרים צריך לכתוב תכנה יעודית לטובת הנושא.

מעבר לכך, ברשתות הבאות: טיב טעם, דור אלון, אושר עד, כל בו חצי חינום, קשת טעמים רשת סופרמרקטים, רמי לוי שיווק השקמה, סופר דוש, סופרשוק יוחננוף, פרשמרקט, וסטופמרקט המידע המוצג אינו מחולק לפי סניפים, אלא, המידע על מחירי המוצרים מפוזר בין קבצים שונים עבור סניפים שונים. מחיר מוצר עלול להשתנות בין סניף לסניף, ולכן, בכדי שהצרכן יוכל לבחון את שווי קנייתו בסניף מסוים, עליו ראשית לאחד בין כל הקבצים, אחר כך לעשות סינון לפי מספר סניף ולאחר מכן לחפש את המוצרים אותם הוא מבקש לרכוש.

כך שלמעשה, הצרכן הממוצע אינו יכול לבצע השוואת מחירים ברשתות אלה ע"י כלים בסיסיים הנמצאים במחשבו או במכשיר הטלפון הנייד שלו.

בנוסף, ברשתות אלה הכניסה מוגנת על ידי מנגנון התחברות עם שם וסיסמה וכן על ידי אמצעי אבטחה נוספים כך שבכל כניסה מתקבל csrf-token שונה- מעין מזהה יחודי להתחברות, ונדרשת שליחת נתונים ספציפיים עם בקשת המידע לשרת, מה שמסבך משמעותית את אופן פעולת התוכנה, אותה יש לבנות על מנת להוריד את הקבצים באופן אוטומטי מהאתרים השונים ואיחודם לכדי קובץ קריא המחולק ראשית לפי סניפים ושנית לפי מוצרים.

בשל כך, נדרשנו לכתוב תוכנה שסורקת את כל תכנם של האתרים השונים של רשתות השיווק, שולפת ומורידה קבצים בפורמט gz שקשורים לסניפים הרלוונטיים, ולאחר מכן מתקנת את שמות הקבצים ומחלצת את קבצי ה XML. בשלב הבא התוכנה סורקת את תוכן קבצי ה XML וקובעת שם קובץ המזהה את מועד העדכון ואת הסניף.

שלב שני- סידור הנתונים מ XML ל CSV

קבצי ה XML אינם קריאים לצרכן בקלות וקשה להוסיף להם קטגוריות שישמשו אותנו בהמשך, לשם כך נרצה להמירם לקבצי CSV. קבצי ה XML השונים משתמשים בשמות שונים של תגיות, המקשים על אוטומציה בתרגום הקבצים לפורמט אחר. בנוסף אין אחידות בין הסופרים בקבצי ה XML, העמודות אינן בסדר זהה בין הקבצים וכן בחלקם ישנן עמודות שאין בסופרים אחרים.

- כל קובץ CSV מקבל שם אינפורמטיבי בפורמט:

cat-prices-ChainID-SubChainID-Branch-Date

שלב שלישי- חלוקת מוצרים לקטגוריות

שלב זה הינו שלב הכרחי על מנת שנוכל לסווג את המוצרים לפי קטגוריות מחירים באמצעות Clustering. בשלב זה אנו מחלקים את המוצרים לקבוצות בעלות מכנה משותף, כדי שנוכל לחקור את המוצרים השונים. הקטגוריות השונות נבחרו על סמך סל המוצרים של המועצה הישראלית לצרכנות. בסה"כ בחרנו 33 קטגוריות מוצרים, מתוכן 20 קטגוריות מוצרים שאינם ירקות, ועוד 13 קטגוריות של ירקות (מפורטות בנספחים).

על מנת ליצור סט נתונים נוח לקריאה המחולק לפי קטגוריות שונות, היה עלינו להתמודד עם שמות קבצים בעייתיים ונאלצנו לעבור באופן ידני על הקבצים בשילוב עם עיבוד אוטומטי. למשל, בכדי שלקוח יוכל למצוא את כל סוגי החלב הקיימים, הוא אינו יכול לחפש "חלב" משום שיקבל הרבה קטגוריות שאינן רלוונטיות, כדוגמת שוקולד חלב, סוגי חלב צמחי, וכו'. בכדי להתמודד עם בעיה זו עברנו על המידע ידנית ויצרנו "מילון" אשר באמצעותו ניתן לחפש את המוצרים בקטגוריות השונות בעזרת תיאורים מקוצרים או תיאורים ספציפיים המוכלים בשמות או תיאורי המוצרים הרלוונטיים אך לא מופיעים בתיאורי מוצרים אחרים.

פתרון זה הולם בחלק מן המקרים, אך לא נותן מענה לכל המוצרים בהם התעניינו. למשל, בתחום הירקות הבעיה רחבה יותר משום שלכל סופר מספר ברקוד שונה למוצר, ועל כן, הבעיה של החיפוש לפי שם מחריפה. לדוגמא, בחיפוש של "מלפפונים" מוצאים גם מלפפונים חמוצים ושימורים לסוגיהם, בחיפוש של "עגבניה" נמצאים סוגים רבים של רטבים לפסטה ורסקים שונים. לכן, עברנו על כלל הסופרים אותם בחנו,

ציינו את הברקודים הרלוונטיים לכל קטגוריה מהקטגוריות שנבחרו (מופיעות בנספחים), וסיווגנו את הנתונים בהתאם.

תהליך המיון לקטגוריות כולל סיווג בהתאם למילות המפתח על ידי סריקת כל המוצרים בכל הסניפים ובניית מילון למוצרים, בניית מילון למוצרים מיוחדים על ידי סיווג ידני, ולבסוף, ריכוז כל המוצרים לסניף בציון הקטגוריה הרלוונטית, כאשר כל קובץ מתייחס לסניף מסוים. בכל פעם שנעבד עדכון מחירים חדש ניצור מחדש את המילונים מכל הקבצים ונייצר קבצים חדשים עם קטגוריות, מכיוון שתיאורי מוצרים מסופר מסוים מאפשר לנו לזהות מוצרים חדשים בסופר אחר, גם כאשר באחרון תיאור המוצר אינו מזוהה עם הקטגוריה.

שלב רביעי - Clustering

בשלב זה הנתונים כבר מחולקים לפי קטגוריות, כלומר, לכל סופר מתוך ה-20 ישנו קובץ מוצרים, וכל מוצר נמצא תחת קטגוריה מסוימת.

השלב התחלק לכמה חלקים:

(1) קיבוץ כל המוצרים מקטגוריה מסוימת ל-3, כל מוצר קיבל דירוג 0-2 כאשר 0 מסמל את מיקום המוצרים הזולים ו-2 את מיקום המוצרים היקרים.

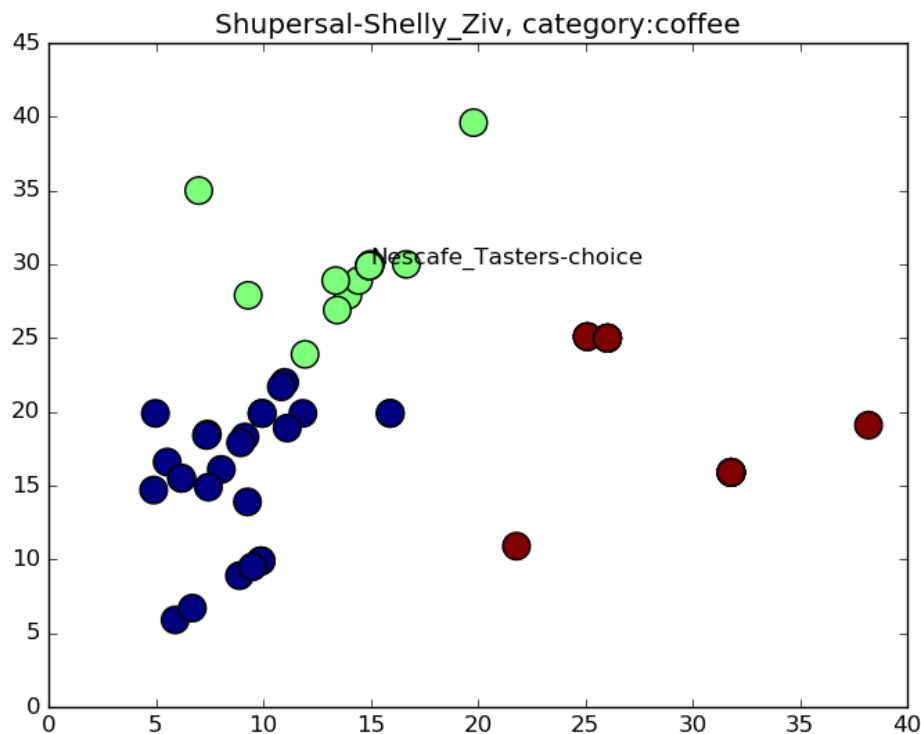
הסיבה שבחרנו לקבץ לפי קטגוריות היא משום שלא כל הסופרים מוכרים את כל המוצרים (לדוג' לשופרסל ישנו מותג בית ולרשתות האחרות ישנם מותגים כמו המותג המשותפים להם).

הקיבוץ נעשה ע"י שימוש באלגוריתם: K-Means חלוקה ל-3 קבוצות שאינן בהכרח שוות גודל ע"מ למזער את סה"כ מרחקי מחירי המוצרים ממרכז הקבוצה.

מחירי המוצרים - לכל מוצר בחרנו להסתכל על שני מדדים: מחיר ליחידה ומחיר למוצר. בחירה זו נעשתה משום שכשאר צרכנים מגיעים לסופר הם מסתכלים על מחיר המוצר (ללא תלות בכמות אותה הם מקבלים) וכן על מחיר ליחידת מידה.

קטגוריות בהן ישנם פחות מ-3 מוצרים בסופר מסוים: ישנם סופרים שהיצע המלפפונים בהם אינו גדול ולכן לא יכולנו לדרג את המוצרים הללו.

לדוג', אם נרצה למצוא את דירוג מחיר המכירה של קפה מגורען Nescafe_Tasters-choice (7290000072753) בחנות שופרסל שלי זיו, נראה כי המוצר מדורג בקיבוץ מספר 1 שהוא האמצעי. כלומר סוג זה של מוצר הינו בקטגוריה עם שאר המוצרים המתומחרים בינונית.



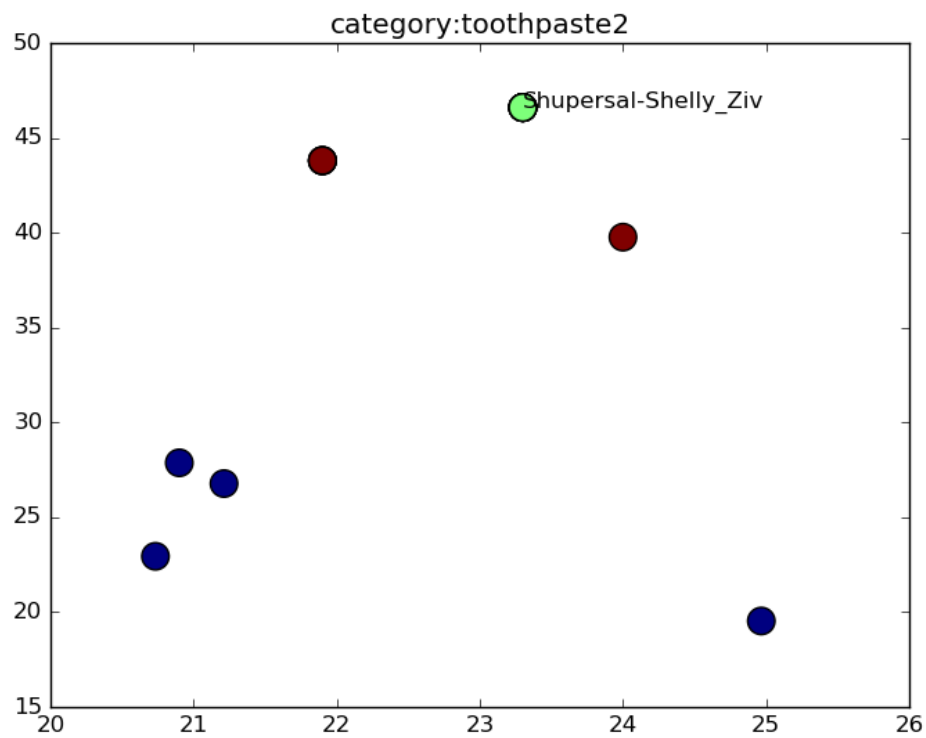
לאחר הקיבוץ הראשוני יכולנו לנתח דברים נוספים:

(2) לכל דירוג מוצר (לדוגמה משחת שיניים בדירוג הנמוך 0): קיבוץ כלל הסופרים ל-3 קבוצות שונות. מקיבוץ זה נוכל להסיק עבור כל דירוג מוצר מיהם הסופרים הזולים ביותר לקנייתו בהם ומיהם היקרים.

הקיבוץ נעשה ע"י שימוש באלגוריתם K-Means: חלוקה ל-3 קבוצות שאינן בהכרח שוות גודל ע"מ למזער את סה"כ מרחקי המחיר הממוצע למרכז כל הקבוצה.

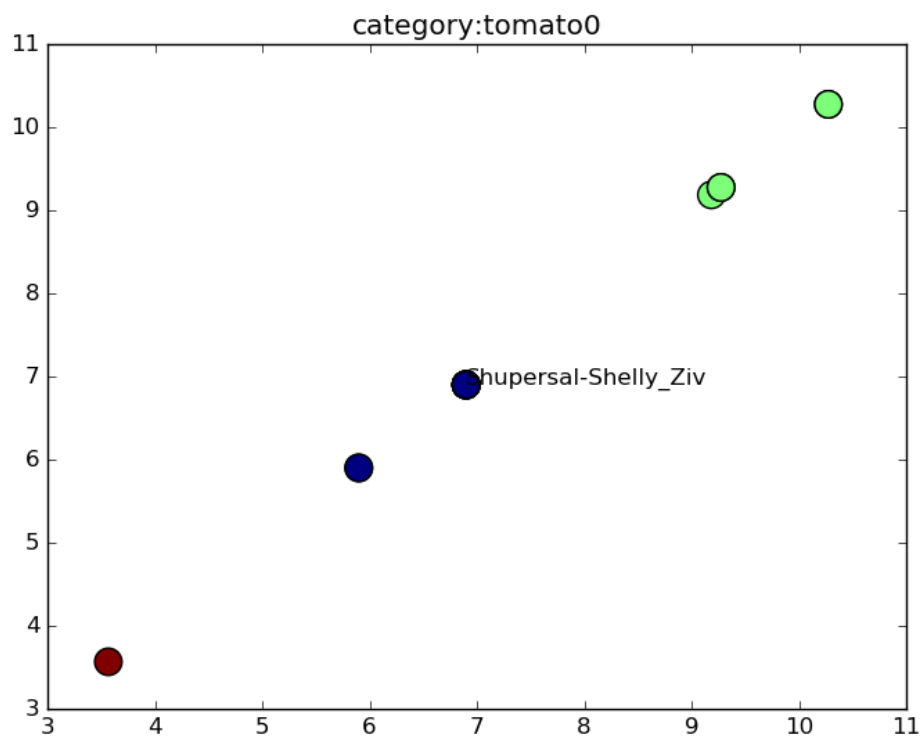
המחיר הממוצע: המרכז של כל דירוג קטגוריה בסופר מסוים, שקיבלנו ע"י הקיבוץ הקודם.

לדוגמה אם נרצה למצוא את דירוגו של הסופר שופרסל שלי זיו בקטגוריית משחות שיניים בדרג פרימיום, נגלה שסופר זה מתמחר את מוצרם אלו יקר יחסית לשאר הסופרים בחיפה.



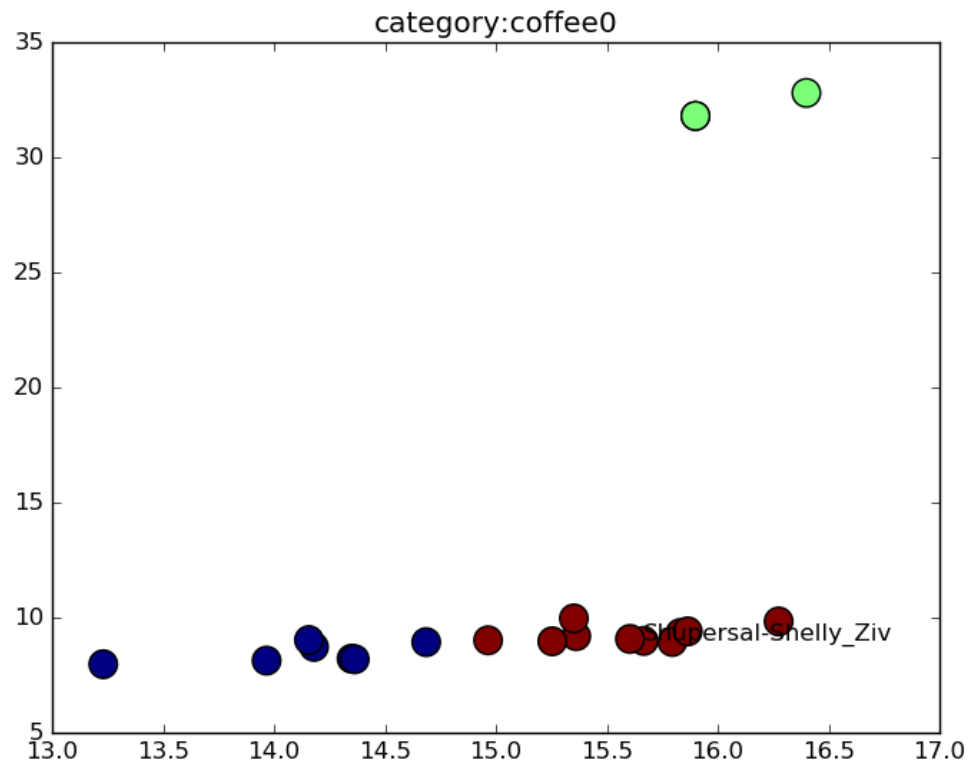
בעגבניות במחיר הנמוך, שופרסל שלי זיו מדורג כמוכר במחיר ביניים.

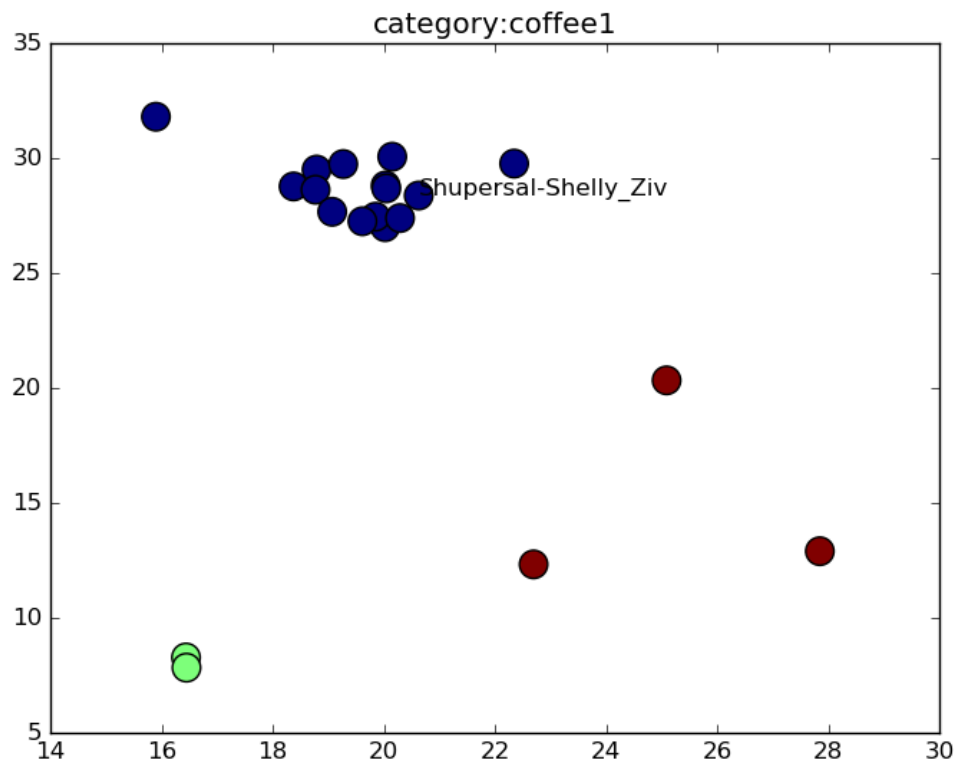
בנוסף, ניתן לראות כי עבור ירקות, מחיר המכירה הינו זהה למחיר ליחידת מידה.



עבור קפה, נתבונן בשני הדירוגים הבאים, 0 ו-1:

ניתן לראות כי שופרסל שלי זיו מתמחרת את הקפה הזול בתמחור בינוני ואת הבינוני במחיר יקר יחסית לשאר הסופרים בחיפה.





3) לכל קבלה (שאספנו) : לכל מוצר בקבלה ישנו מחיר קניה, כל מוצר קיבל דירוג חשיבות לפי סה"כ המחיר שלו מכלל הקנייה, מוצרים בעלי נתח גדול מהמחיר יקבלו חשיבות גבוהה וההפך. לכן, לכל קבלה אמיתית מהסופר הוצאנו המלצה מהיכן כדאי לקנות כל מוצר וכן מהיכן כדאי לרכוש את כלל המוצרים בקבלה

שלב חמישי- תוצאות וניתוח הנתונים

אספנו 25 קבלות, על מנת לבחון האם הסלים שהם קנו אכן נקנו במיקום המשתלם עבורם.

לאחר בחינה של המוצרים והתוצאות שהתקבלו- חלקן לא היו הגיוניות, נתגלו מספר לא מבוטל של שגיאות בקבצים המקוריים (למשל על פי הקבצים, בויקטורי ניתן לקנות סבון כלים של "המותג" בנפח 1.5 ליטר, במחיר 726 ש"ח).

רציונליות הצרכן

ישנם מגוון גורמים שמשפיעים על צרכן בבחירת הסופר אליו הוא נוהג להגיע, ונצפה שאם נתייחס לבחירת הסופר בו הסל המתוכנן הוא הזול ביותר כבחירה רציונלית, אז אדם אקראי שנבחן אינו נתפס כרציונלי. מעבר לגורמים הסובייקטיביים שעשויים להיות לאותו צרכן, אנחנו מניחים שיש לו גם ידע על מחירי הסופר הצפויים למוצרים אותם הוא נוהג לרכוש, אף אם הם חלקיים ומוערכים בצורה גסה. אנו משערים שצרכן מעריך מוצר מסוים כזול-פשוט (בין אם בהוצאה הנדרשת על מנת לרכשו ובין אם ביחס עלות תועלת שבו), בינוני, או יקר-איכותי. מעבר לכך, הסופרים השונים עצמם עשויים להתמחות בקומבינציות קטגורית מוצר-איכות שונות, כך שסניף מסוים עשוי למשל להיתפש כזול עבור אבקת כביסה איכותית (נסתכל על ממוצעי המחירים והמחירים ליחידת מוצר של הקומבינציות) אך יקר עבור קפה פשוט.

ראשית בחנו קבלות כדי לראות האם בחירת הסופר היא רציונלית (ראה דוגמה בנספח תוצאות רציונליות - פלט). חישבנו את העלות הצפויה מרשימת קניות של מוצרים דומים באותה איכות במגוון סניפים ברחבי חיפה.

בשלב הבא בדקנו את טיב הבחירה של סופר על פי הידע הגס שברשות הצרכן. סיווגנו את הסופרים לפי רמת ההתמחות שלהם בכל קומבינציות קטגוריה-איכות כ0, 1, או 2, כאשר 0 מציין שמדובר בסניף משתלם למוצר או למוצר חלופי דומה. לכל רשימת קניות, נתנו ניקוד כתלות בטיב הבחירה, על ידי ממוצע משוקלל (לפי ההוצאה לכל פריט כמשקלות) של רמת ההתמחות של הסניף בקטגורית המוצר באותה האיכות. הציונים האפשריים הם בין 0 ל2, כאשר 0 מציין רשימת קניות שכל אחד ממוצריה (או מוצר חלופי דומה באותה איכות) נחשב כמשתלם בסניף. לא צפוי שסניף מסוים יציע את המחירים הזולים ביותר לכל המוצרים ברשימה טיפוסית, אלא יציע מוצרים 'מושכי תשומת לב' מסוימים במחיר משתלם וירוויח יותר על מוצרים אחרים.

בדוגמה בנספח ניתן לראות קבלה שנקנתה בשופרסל שלי סניף כרמל. ישנם אמנם מספר סניפים זולים יותר עבור אותה הרשימה, כך שצרכן רציונלי היה עשוי לחסוך כסף, אך הציון שקיבל הסניף הינו השני הכי נמוך ברשימה. תוצאה זו עשויה להעיד על כך שאופן קבלת ההחלטה של הצרכן דומה לאופן שתיארנו, כך שהוא מבצע הערכות גסות (חלוקות לשלושה קיבוצים) ובעזרתם שואף לבצע בחירה המובילה לתועלת גבוהה.

הוראות הפעלה

במידה ואין למשתמש קבצי CSV עם קטגוריות-

ראשית יש לפתוח תיקיה בשם scrapedfiles ובתוכה תיקיות:

full, renamed, xml, xml_renamed, csv, cat

את הקוד לcrawl אפשר למצוא ב: <https://github.com/gald33/prices>

יש לשים את הקבצים בתיקיה בשם prices, ובנוסף יש לשנות בקובץ config.py את המשתנה home_dir כך שיצביע על התיקיה המכילה את התיקיות scrapedfiles וprices. על מנת להוריד את הנתונים של הסופרים, יש לפתוח סביבה וירטואלית ובה יש להשתמש ב'scrappy' לאחר התקנת החבילות הדרושות (הסביבה הוירטואלית נועדה למנוע התנגשויות בחבילות) בעזרת הספרים spiders וקבצי ההגדרות שכתבנו.

יש להכנס לסביבה הוירטואלית ובתיקיה prices להריץ:

```
scrapy crawl supermarkets
```

ישנו תהליך מורכב יותר לסופרים בעלי זיהוי משתמש משתנה, בו יש לשלוח את הטוקן מבעוד מועד ולנסח בקשת פוסט לשרת. ניתן לחלופין להשתמש ישירות בפקודה:

```
python update_prices.py
```

הסקריפט יוריד את הנתונים הדרושים מאתרי הסופרים, ובהתאם לפרטים ולתוכן, יתקן שם, יחליף מהדחיסה, יוסיף פרטים על העדכון. את המשך התהליך יש לבצע מחוץ לסביבה הוירטואלית. לאחר מכן יש להמיר את קבצי הXML לCSV על ידי:

```
python xml_to_csv.py
```

בשלב הבא יש להריץ את הפקודות:

```
python create_categories.py
```

```
python add_categories.py
```

סקריפטים אלה משתמשים בקבצי categories.json, items.csv, special_items.csv ומוסיפים לכל הקבצים קטגוריות למוצרים מקבצי הCSV עצמם, מקבצי הCSV השכנים ומקבצי הJSON. הקבצים הסופיים עם כל הפרטים הדרושים עבור כל סופר, ימצאו בתיקית cat.

ניתן להוריד את הקובץ scrape.sh, לשים אותו בתיקית home_dir ולבצע את כל האמור לעיל על ידי הפקודה:

./scrape.sh

הקמנו שרת ב-AWS שמבצע את כל התהליך הנ"ל באופן יומי.

במידה ויש למשתמש קובץ CSV עם קטגוריות ניתן להתחיל מנקודה זו:

את הקוד לניתוח הנתונים אפשר למצוא ב: <https://github.com/sagittefrat/consumer-rationality>

1. יש להריץ דרך command line:

`Interface.py -c [folder of the categorized data]`

הפלט הינו מספר קבצים:

'category_position_centers.csv':

מחזיק קטגוריה ותת קטגוריה - מיקום בקבוצה, מבין 3 הקבוצות שיצר מודול הקלאסטר (למשל 0,milk), ואת המחיר הממוצע ליחידת מידה והמחיר הממוצע של המוצרים באותה תת קטגוריה.

'barcode_super_category_position.csv':

קובץ זה מכיל את כל הברקודים בכל הסופרים ולכל ברקוד משייך את הקטגוריה אליה הוא שייך (milk) ואת תת הקטגוריה שלו (0,1,2). בנוסף, תצוגה גרפית של הקלאסטרים שנבחרו.

2. יש להריץ:

`cluster.py [category_position_centers.csv path]`

הפלט הינו הקובץ: super_scores.csv

מחזיק עבור כל קטגוריה ודירוג, לכל סופר הנמצא בקטגוריה זו, מה מיקומו בה בנוסף מהם מרכזי המחירים של הקטגוריה והדירוג.

למשל: "יש צרפת" הוא בקטגוריה חלב 0 אז הוא נמצא בקלאסטר 'discount' עבור הקטגוריה הזו. או שהסופר "דיל קניון-חיפה" ממקודם במקום 2 בקטגוריה של חלב אומר שהוא נמצא בקלאסטר 'premium' עבור הקטגוריה הזו.

3. יש להריץ:

`reciepts.py -p [barcode_super_category_position.csv path] -e [receipts files folder path] -d [CSV files with categories folder path] -c [super_scores.csv path]`

הפלט הינו דו"ח של סיכומי מחירים וסיכומי ניקוד של רשימות הקניות, עבור כל הסופרים הרלוונטיים.

נספחים

קטגוריות של ירקות:

קטגוריה	המוצרים הרלוונטיים
Cucumber	מלפפון, מלפפון בייבי
tomato	עגבניות שרי, עגבניות תמר, עגבניות...
carrot	גזר
lemon	לימון
banana	בננה
orange	תפוז
potato	תפו"א אדום, תפו"א לבן
apple	תפוח עץ מוזהב, תפוח עץ יונתן, ת"ע סמיט...
cabbage	כרוב אדום, כרוב לבן, כרוב למילוי
pepper	פלפל אדום, פלפל ירוק, פלפל חריף, פלפל שישקה, פלפל כתום, פלפל צהוב, פלפל צ'ילי
onion	בצל סגול, בצל לבן, בצל יבש, בצל ירוק, בצלצלי שאלוט
lettuce	חסה רומית, חסה ערבית, חסה עגולה, חסה שטופה...
frozen	אפונה קפואה, שעועית קפואה, ברוקולי מוקפא, כרובית קפואה, תירס קפואה, גזר גמדי קפואה

מילון הקטגוריות:

קטגוריה	ביטויים לחיפוש	שם הקטגוריה
משחת שיניים	'משחת שיניים'	toothpaste
חטיף בוטנים	'חטיף בוטנים', 'במבה'	peanut snack

tampons	'טמפונים', 'טמפון', 'טמפוני'	טמפונים
bleach	'אקונומיקה'	אקונומיקה
stain remover	'מסיר כתמים', 'כתמים'	מסיר כתמים
fats remover	'מסיר שומנים'	מסיר שומנים
deodorant	'דאו', 'דאודורנט', 'בודי ספריי'	דאודורנט
coconut cream	'קרם קוקוס', 'חמאת קוקוס', 'מי קוקוס'	קרם קוקוס
bag	'שקיות מזון', 'שקיות רב"ש', 'שקיות אוכל'	שקיות מזון
garbage bag	'שקיות אשפה', 'אשפתון'	שקיות אשפה
white cheese	'גבינה לבנה', 'גבינת שמנת'	גבינה לבנה
yellow cheese	'גבינה צהובה', 'אמנטל', 'גבינת עמק', 'טל העמק', 'הצהובה הטובה', 'הצהובה טובה', 'גבינת נועם', 'גלבע', 'גוש חלב'	גבינה צהובה
milk	'חלב עמיד', 'חלב מועשר', 'חלב 3%', 'חלב 1%', 'חלב תנובה', 'חלב מפוסטר', 'חלב טרי', 'חלב הרים', 'חלב שתיה', 'חלב על הבוקר', 'חלב יטבתה', 'חלב נטול לקטוז'	חלב
toilet paper	'נייר טואלט', 'נ.טואלט'	נייר טואלט
bread	'לחם שיפון', 'לחם אחיד', 'לחם עינן', 'לחם ברמן', 'לחם דגני', 'לחם פרוס', 'לחם עם', 'לחם כפרי', 'לחם פומפ', 'בגט', 'באגט', 'לחמני', 'חלה רגילה', 'חלה מתוקה', 'חלה מקמ'	לחם
fabric softener	'מרכך כביסה'	מרכך כביסה
Dish soap	'נוזל כלים', 'נוזל כלים'	סבון כלים
pasta	'פסטה ספגטי', 'פסטה מקרוני', 'פסטה תלתל', 'פרפקטו', 'פסטה מסולסלת', 'רביולי', 'פתיתים א', 'פתיתים קוס', 'פתיתים פלוס', 'פתיתים	פסטה

	טבע', 'ספגטי מס', 'ספגטיני'	
oil	'תרסיס שמן ק', 'שמן קנולה מ'	שמן קנולה
coffee	'קפה נמס', 'קפה שחור', 'קפה טורקי', 'קפה אל', 'קפה אספרסו', 'קפסולות קפ'	קפה

תוצאות רציונאליות - פלט:

תוצאות של קבלה משופרסל שלי סניף כרמל:

Prices Analysis

```
[('Shupersal-Shelly_Nativ-Hen', 262.37513764900916),  
 ('Shupersal-Shelly_Keler', 285.4197521167557),  
 ('Shupersal-Deal-Extra_Tel-Hanan', 230.0231430378603),  
 ('Shupersal-Deal_Grand-mall', 218.45966235533976),  
 ('Shupersal-Deal_Admirality', 210.99601776018105),  
 ('Shupersal-Shelly_Stela', 291.6384792020073),  
 ('Shupersal-Deal_Eliezer', 197.81093881448902),  
 ('Shupersal-Shelly_Sapir', 307.81166436423047),  
 ('Shupersal-Shelly_Carmel', 255.27709736909122),  
 ('Shupersal-Shelly_Horev', 321.98688202973494),  
 ('Shupersal-Deal_Haifa-mall', 252.40266999922505),  
 ('Shupersal-Shelly_Ziv', 286.2011262838468)]
```

Scores Analysis

```
[('Shupersal-Shelly_Nativ-Hen', 0.9811900191938578),  
 ('Shupersal-Shelly_Keler', 1.092514395393474),
```

('Shupersal-Deal-Extra_Tel-Hanan', 1.1500959692898274),
('Shupersal-Deal_Grand-mall', 1.0245681381957774),
('Shupersal-Deal_Admirality', 1.2664107485604608),
('Shupersal-Shelly_Stela', 1.092514395393474),
('Shupersal-Deal_Eliezer', 1.1642994241842608),
('Shupersal-Shelly_Sapir', 1.325143953934741),
('Shupersal-Shelly_Carmel', 0.9811900191938578),
('Shupersal-Shelly_Horev', 1.2030710172744723),
('Shupersal-Deal_Haifa-mall', 0.7535508637236086),
('Shupersal-Shelly_Ziv', 1.092514395393474)]

references

1. Vanhoof, T. B. G. S. K., & Wets, G. Using Shopping Baskets to Cluster Supermarket Shoppers.
2. Brijs, T., Swinnen, G., Vanhoof, K., & Wets, G. (2000). A data mining framework for optimal product selection in convenience stores.
3. http://www.consumers.org.il/category/semel_supermarkets_list - סל קניות המועצה הישראלית לצרכנות
4. <http://www.antitrust.gov.il/subject/188.aspx> - שקיפות מחירים