פרוייקט WALT  
מיכל זינגר ויובל שוורץ

**מבוא:** אוכל הוא אחד מצרכי חיינו, אנחנו לא יכולים לשרוד בלעדיו, ואנחנו צריכים אותו כל יום. אבל כמו שבני אדם תמיד עשו, הפכנו את הדרישה הזו שלנו לתעשייה מלאה. מזון גדל למגזר של מיליארדי דולרים בכל העולם, המעסיק מיליוני אנשים. עם עלייתן של אפליקציות משלוחי המזון, תעשיית המזון עוברת תקופת זהב וחווה עלייה אקספוננציאלית.

​עם התפתחות האפשרות להזמין באופן מקוון מהמסעדה האהובה עליהם, הלקוחות מתפנקים. ככל שיותר ויותר אנשים משתמשים בפלטפורמת הזמנת אוכל באינטרנט, תעשיית משלוחי המזון צומחת בצורה פנומנלית. הרבה נתונים נוצרים כאשר לקוח מזמין אוכל באינטרנט. סוג האוכל שהמשתמש מזמין, זמן ההזמנה, הטעם שהמשתמש מעדיף וכו'. אנשים בדרך כלל בררנים לגבי האוכל שלהם, ועל ידי ניתוח הנתונים שהוזכרו לעיל, סטארט-אפים למשלוחי מזון יכולים לספק חוויה מענגת עבור המשתמשים שלהם.

בעקבות זאת, אנו יודעים את ההשפעה שיכולה להיות לניתוח נתונים על תעשיית האוכל. מערכת Big Data Analytics חזקה יכולה להשפיע לטובה על השורה התחתונה של העסק. במסמך זה נדגים כיצד תחום ניתוח הנתונים יכול לסייע לתעשיית המזון למניעת הפסדים ולשיפור חווית המשתמש. על ידי ניתוח הבחירות, נוכל להתאים את ההצעה כך שתתאים לטעם הלקוחות. זה יעזור בשיפור ההכנסות ובשדרוג חווית המשתמש.

כדי להתאים לציפיות המשתמשים, Data Science ו-Data Analytics נכנסים לתמונה וממלאים בה תפקיד חיוני. הארגונים ממנפים את Data Analytics כדי לגלות ניתוח מדויק של תעשיית המזון בכדי להגדיל את העסק שלהם, להיות מודעים למגמות ולהקטין הוצאות. Data Analytics תומך בזיהוי צורכי הלקוח. הוא מסייע לארגונים באיסוף וניתוח נתונים המסייעים בזיהוי מגמות ודפוסים.

החידושים מביאים פתרונות יצירתיים לבעיות המטרידות את התעשייה תוך הפקת התפתחויות חיוביות בתעשיית המזון. יתרה מכך, השימוש בטכנולוגיה נחוץ יותר עבור אפליקציות של משלוחי מזון. הצורך לאיסוף נתונים הכרחי כדי לדעת על ביקוש והעדפות הלקוחות. לפי דיווח, שוק משלוחי המזון עד 2023 צפוי להגיע ל-15 מיליארד דולר.

להלן הסיבות לשימוש בData Science ו-Data Analytics בקרב תעשיית משלוחי המזון:

1. **שיפור זמן האספקה ​​וחסכון**

אחת הסיבות לכך שתעשיית המזון עולה מדרגה היא בגלל משלוחי המזון. משלוחי המזון של המסעדות לבית הלקוח הפכה את חווית הלקוח לניתנת לניהול ויעילה יותר. עם זאת, למשלוח מזון יש בעיה מסובכת לפתור מכיוון שהיא כוללת קנה מידה בין הבטחת חווית לקוח טובה תוך שמירה על פרודוקטיביות גבוהה במשלוח ההזמנות. Data Science וניתוח עוזרים להם לייעל את הזמן והעלות. לכן, אם מתרחשות נסיבות כלשהן, הם יודעים מה לעשות ומעודדים אותם לספק בזמן וכצפוי.

1. **הערכת התנהגות לקוחות**

Big Data Analytics ו-Data Science ידועים בנתונים שיכולים לעזור להם לחזות את התנהגות הלקוחות. לכל סטארטאפ של משלוחי מזון יש פלטפורמת מדיה חברתית שבה הם מעלים, מפרסמים על הצעות חדשות וכו'. עם זאת, אם משהו מתרחש, הראשון להיות יעד הפרסום הוא המדיה החברתית. לא ניתן פשוט להתעלם מהרגשות של הלקוחות ברשתות החברתיות. זה המקום הראשון שבו יעבירו ביקורת.

בעזרת יישום של Big Data Analytics, אפשר למדוד את כיוון הלקוחות לעבר המותג שלך ואת ביצועי משלוחי המזון ותגובת הלקוח; Big Data Analytics תאסוף ותנתח את כל ההערות של המותג על פני פלטפורמות שונות של מדיה חברתית כמו טוויטר, אינסטגרם, פייסבוק ואחרות. צעד זה יעזור בביצוע החלטות עסקיות על סמך הנתונים שסופקו.

1. **קידום מבוסס מיקום**

כל משלוח מזון מספק מזון למרחק מסוים. השליח לא יכול להיות בכל מקום ואם האדם המשיך, הוא לא יוכל לספק לו את המזון בזמן. היתרון של Big Data Analytics הוא בכך שהוא מאפשר לחברת המזון למקד לקוחות על ידי יצירת שימוש במיקום בזמן אמת ובזמן הנכון. חברות משלוחי המזון יישמו מודיעין מיקום כדי לקבוע את שיעורי הביטולים לפי אזור, כדי לגלות את שיעור הביקוש וההיצע באזור ולקבל מושג לגבי גדלי השלבים של המשלוחים והמסעדות וההבדלים ביניהם.

1. **מעקב אחר אלגוריתמים חכמים לביקוש**

על ידי יישום של אלגוריתמי Big Data חכמים, אפליקציית משלוחי המזון יכולה לחזות את ההזמנה הבאה של הלקוח. על ידי ניתוח הגלישה המוקדמת של משתמש והתבוננות בנתוני ההזמנה שלו בעבר, האלגוריתם יכול לחזות מתי הלקוח צפוי להזמין שוב או לא.

באמצעות אימוץ אנליטיקה חזויה, אפליקציית משלוחי מזון יכולה לחזות במדויק כמה לקוחות יזמינו, במיוחד בזמן מסוים ביום או בשבוע, ומאיפה וממה. אלגוריתמי הניתוח יכולים לדעת מה הלקוח צפוי לרצות ובאיזה אזור בעיר או במדינה, ומהו המזון המוזמן הכי הרבה.

**שאלת המחקר שלנו:**

המטרה העיקרית שלנו בפרוייקט היא חקירת דאטה לשם מערכת המלצה עבור לקוחות המזמינים משלוחי מזון. עיקר הדגש במערכת שלנו היא זיהוי דימיון בין לקוחות ומציאת דפוסי הזמנה דומים. כך נוכל לייעל את תהליך ההזמנה מצד הלקוח, מה שיגרום לריבוי הזמנות מצד העסק ולשיפור חווית קניה מצד הלקוחות. מצב זה יועיל לשני הצדדים. אז למעשה שאלת המחקר שלנו היא: Classification of similar customers

**התהליך:**

השלב הראשון והעיקרי בעבודה קשור לנתונים, שהרי – כל העבודה מסתמכת ובנויה על תוכן הנתונים, מקורם, אמינותם, סינונם, בחירת פיצ'רים רלוונטים וכו'. לאורך התהליך היו לנו לבטים רבים לגביי בחירת מקור הנתונים. בהתחלה ניסינו להשתמש בנתונים מוכנים מאתר Kaggle. לאחר ניתוח ועבודה על נתונים אלו הגענו למסקנה שהנתונים לא היו טובים מספיק, סמלוץ הנתונים היה ברמה נמוכה ולא הייתה לנו אפשרות להגיע לתובנות קונקרטיות בעזרת הניתוחים שביצענו. לכן, הוחלט על סמלוץ נתונים באופן עצמאי, בעזרת קבוצה נוספת, בנוסף החלטנו שנעבוד על תוכנה בשם Neo4j.

בהתחלה הדאטה שקיבלנו מהקבוצה הנוספת היה בנוי מ4 קבצי JSON:

* מסעדות – כל מסעדה יש לה רשימת קטגוריות שהיא משתייכת אליה ותפריט.
* הזמנות – כל הזמנה הייתה קשורה למסעדה וללקוח, ולכל הזמנה הייתה רשימה של פריטים שהוזמנו ושדות משלהם
* לקוחות
* קטגוריות

Diagram

Description automatically generatedסידרנו את הקבצים לטבלאות כך שהטבלאות שלנו היו : מסעדות, הזמנות, לקוחות וקטגוריות.  
סידור זה לא היה מוצלח מכיוון שהוא יצר הרבה כפילויות, למשל בתמונה הבאה ניתן לראות כי לכל מסעדה נוצר node לפי כמות הקטגוריות להן היא שייכת. כמו כן גם לכל הזמנה נוצר node לפי כמות המוצרים שהוזמנו באותה הזמנה.

Diagram

Description automatically generatedבתמונה הבאה ניתן לראות כי היה ניתן לראות הזמנות של לקוחות רק לפי מוצרים שהוזמנו, ולא לפי הזמנה בלבד, מה שיצר הרבה רעש וחוסר יעילות.

סידור זה יצר המון כפילויות וגם הזיק להרצת האלגוריתמים שלנו. לכן ניסו לחשוב איך ניתן לסדר את הנתונים שקיבלנו לטבלאות כך שיהיו פחות כפילויות ושהמודל שלנו יהיה יותר יעיל ויותר מובן למתבונן.

לאחר עבודה על הדאטה הכוללת ניקוי נתונים והמרה למשתנים רלוונטים למודל שלנו, יצרנו מבנה נתונים של קבצי CSV של 5 טבלאות:

Icon

Description automatically generated **Restaurants**- מסעדות

Icon

Description automatically generated **Customers**- לקוחות

Icon

Description automatically generated **Categories**- קטגוריות

Icon

Description automatically generated **Orders**- הזמנות

**OrderDescription**Icon

Description automatically generated - תיאור הזמנות

נעבור על כל טבלה ונפרט:

**Restaurants**- מסעדות  
מכיל את הרשומות:

1. Rest\_id – ערך מספרי המכיל מס' ייחודי לכל מסעדה
2. Name – מחרוזת של שמות המסעדות
3. Address – מחרוזת של כתובת המסעדה
4. Rate – ערך מספרי של דירוג המסעדה
5. Latitude & longitude – ערך מספרי המציג את הנקודות מיקום של כל מסעדה
6. קטגוריות המסעדה – מוצג כערך בוליאני, לכל קטגוריה עמודה

ערך זה כולל את הקטגוריות הבאות:

street food, mediterranean, homemade, italian, pasta, burger, mexican, salad,

vegan, vegetarian, wraps, kosher, breakfast, sandwich, soup, french, hummus,

dessert, bakery, sweets, steak, pizza, asian, meat, ice\_cream, smoothie, café,

healthy, american, indian, fish, thai, kebab, sushi, japanese, bowl, kids.

**Customers**- לקוחות  
מכיל את הרשומות:

1. customer\_id – ערך מספרי המכיל מס' ייחודי לכל לקוח
2. fullName – מחרוזת של שם הלקוח
3. Age – ערך מספרי המייצג את גיל הלקוח
4. Latitude & longitude – ערך מספרי המציג את הנקודות מיקום של כל לקוח
5. Female, male – ערך בוליאני המייצג את מין הלקוח

**Categories**- קטגוריות   
מכיל את הרשומה היחידה name – מפרטת את רשימת השמות של כל קטגוריות האוכל.

**Orders**- הזמנות  
מכיל את הרשומות:

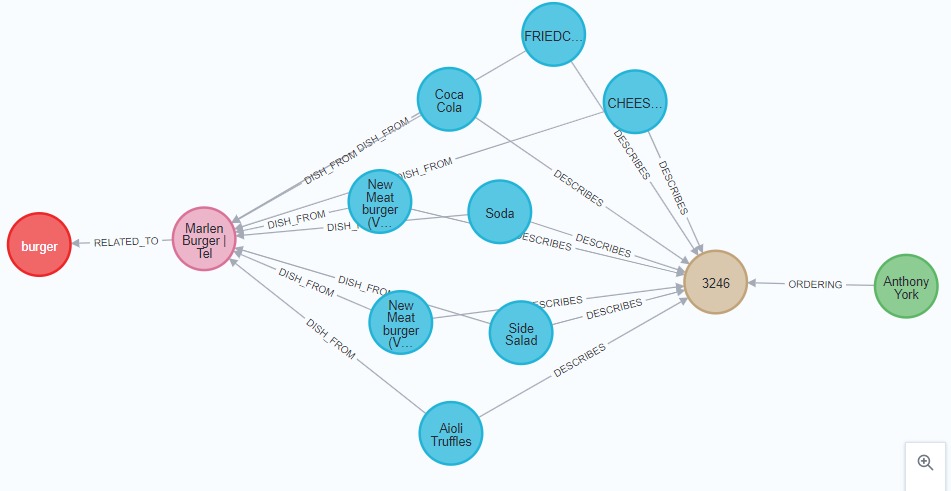
1. order\_id - ערך מספרי המכיל מס' ייחודי לכל הזמנה
2. Rest\_id – ערך מספרי המכיל מס' ייחודי לכל מסעדה
3. customer\_id – ערך מספרי המכיל מס' ייחודי לכל לקוח
4. datetime\_placed – ערך מסוג תאריך המכיל את תאריך ביצוע ההזמנה
5. datetime\_received - ערך מסוג תאריך המכיל את תאריך קבלת ההזמנה
6. total\_price – ערך מספרי המכיל את סך מחיר ההזמנה

**Order** **Description** – תיאור הזמנות  
מכיל את הרשומות:

1. order\_id - ערך מספרי המכיל מס' ייחודי לכל הזמנה
2. Rest\_id – ערך מספרי המכיל מס' ייחודי לכל מסעדה
3. Name – מחרוזת של שם המנה
4. Description – מחרוזת של תיאור שם המנה
5. Price – ערך מספרי המכיל את מחיר המנה
6. items\_id – ערך מספרי המכיל מס' ייחודי לכל מנה

**הסבר קצר על תוכנת neo4j**:

תוכנת neo4j מאפשרת העלאת נתונים בצורת nodes, כך שלכל ערך בטבלה נוצר node עם תכונות. בנוסף ניתן ליצור קשרים בין nodes ע"י קשתות וכך ליצור גרפים מעניינים. אנחנו יצרנו node לכל הזמנה, לקוח, תיאור הזמנה, מסעדה וקטגוריה. בתמונות שאנחנו מראות מהתוכנה, חסרים nodes מכיוון שיש כמות גדולה והתוכנה אינה יכולה להריץ את כולם. לכן בתמונות רואים מדגם מייצג ולא את כל הדאטה.

בתמונה שלפניכם ניתן לראות גרף מייצג הכולל את כל הטבלאות. הnode הירוק מסמן לקוח – שאליו מקושרת הזמנה מס' 3246 בצבע חום. להזמנה זו מקושרים כל הפריטים שהוזמנו בהזמנה, הם מסומנים בכחול. אליהם מקושרת המסעדה שממנה הוזמנה ההזמנה - מסומן בורוד. ולבסוף, למסעדה מקושרת הקטגוריה בצבע אדום.

**טבלת ה-nodes של Restaurants**

Chart, bubble chart

Description automatically generated

**טבלת ה-nodes של Customers**

Chart, bubble chart

Description automatically generated

**טבלת ה-nodes של Categories**

**Chart, bubble chart

Description automatically generated**

**טבלת ה-nodes של Orders**

**Chart, bubble chart

Description automatically generated**

**טבלת ה-nodes של OrderDescription**

**Chart, bubble chart

Description automatically generated**

**הקשרים בין ה nodes בגרף**

* קשר בין לקוחות להזמנות: כל לקוח מחובר להזמנות שלו ע"י קשת. קשר זה נקרא "ORDERING".

Diagram

Description automatically generated

* קשר בין הזמנות לתיאור ההזמנה: כל תיאור הזמנה המייצג פריט ממסעדה מחובר ע"י קשת "DESCRIBES" להזמנה אליה הוא שייך.

Chart

Description automatically generated

3. קשר בין תיאור הזמנה למסעדה:

כל תיאור הזמנה המייצג פריט ממסעדה מחובר ע"י קשת ”DISH\_FROM” למסעדה ממנה הוא הוז

* קשר בין תיאור הזמנה למסעדה: כל תיאור הזמנה המייצג פריט ממסעדה מחובר ע"י קשת ”DISH\_FROM” למסעדה ממנה הוא הוזמן.

Chart, bubble chart

Description automatically generated

* קשר בין הזמנה לקטגוריה: כל מסעדה השייכת לקטגוריה יש בינה ובין אותה קטגוריה קשת “RELATED\_TO” המצביעה על הקטגוריה. יש מסעדות השייכות לכמה קטגוריות. סה"כ יש 37 קטגוריות.

Chart

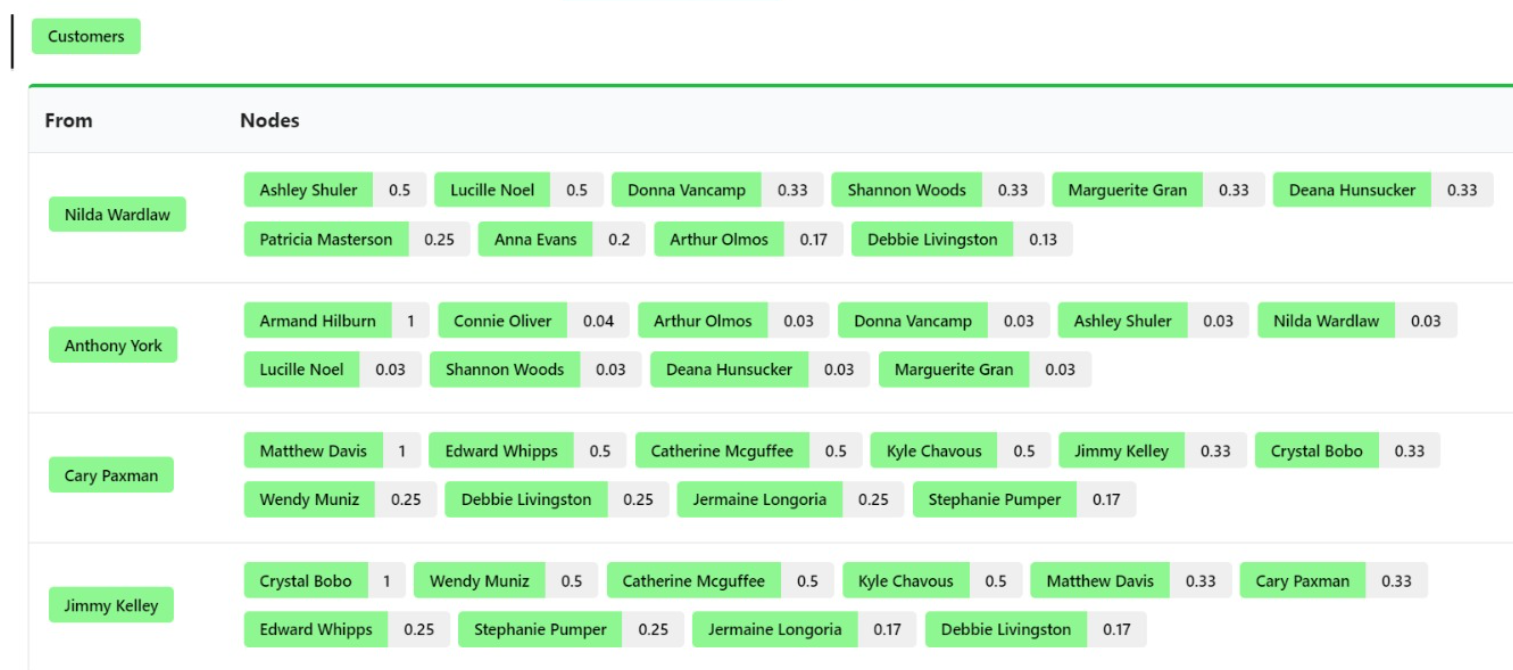
Description automatically generated with medium confidence

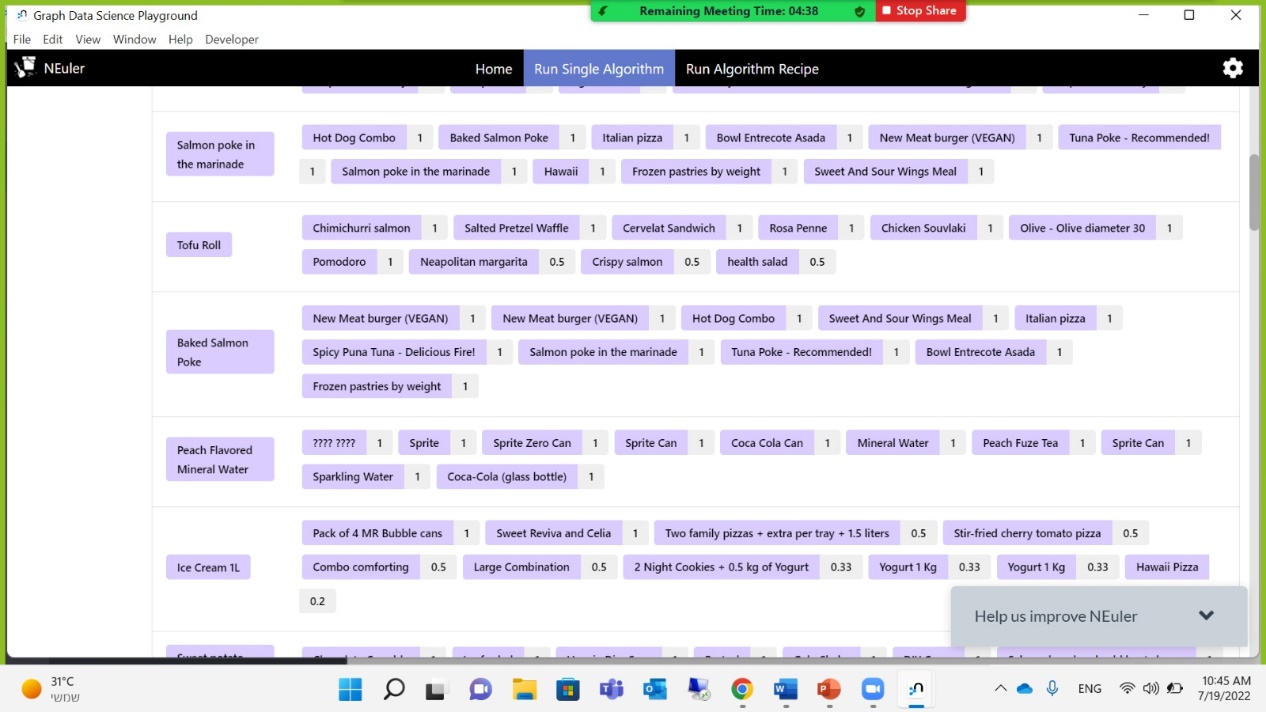
**בחירת האלגוריתמים ופרמטרים:**

בחרנו להשתמש באלגוריתמים הבאים:

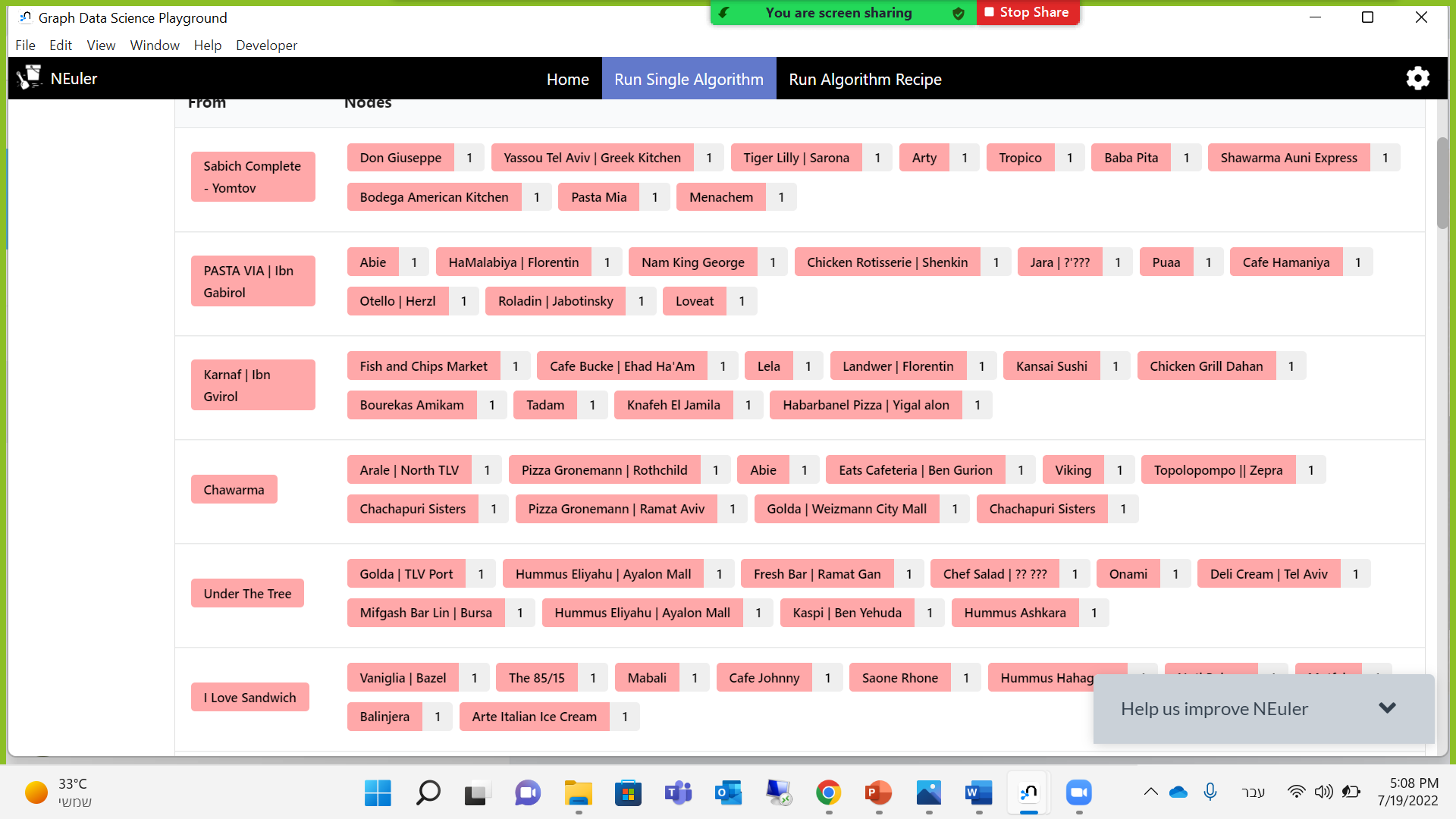
**KNN** - k-Nearest Neighbors algorithm הוא אלגוריתם חסר פרמטרים לסיווג ולרגרסיה מקומית. בהינתן קלט של דוגמה חדשה, האלגוריתם משייך אותו לקבוצה. הדוגמה משויכת למחלקה הנפוצה ביותר בקרב k השכנים הקרובים.

השתמשנו באלגוריתם זה בכדי לשייך את הלקוחות שלנו לקבוצות של לקוחות שונים על פי גיל לקוח כך שכל קבוצה מונה 10 לקוחות הכי דומים. כך ניתן לייחס לכל לקוח את הלקוח בסביבת הגיל שלו, זו יכולה להיות דרך להצעת מנות ללקוח על פי קבוצת גיל.



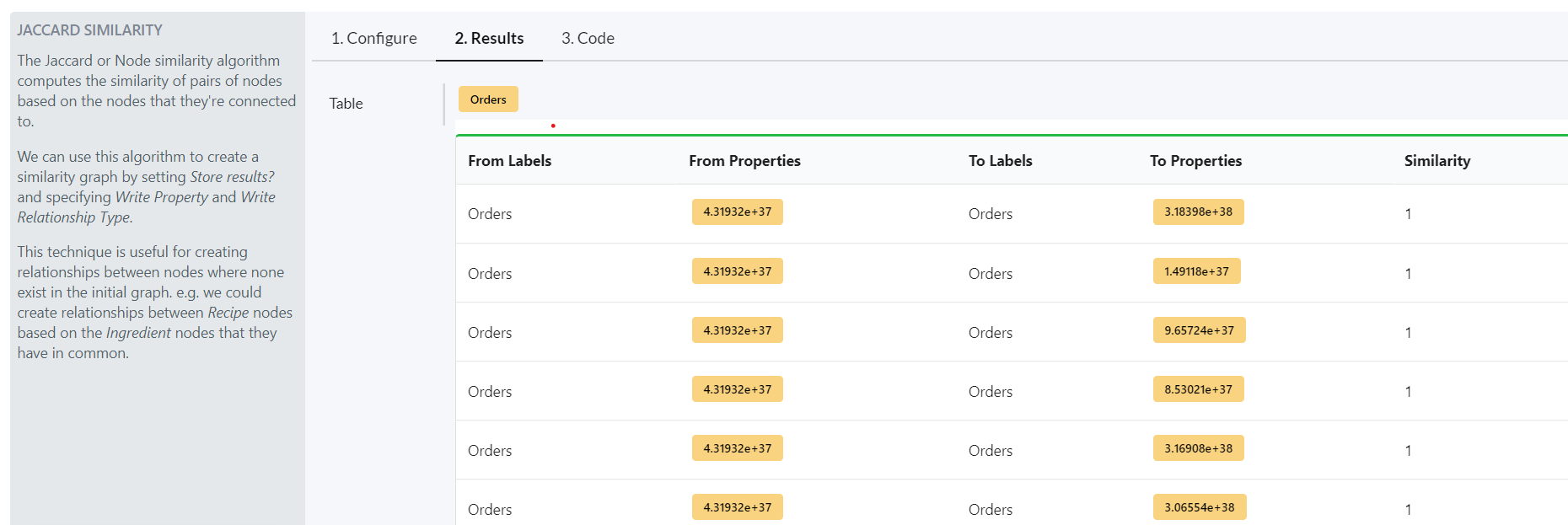
עוד שימוש של האלגוריתם – הפעם בחרנו בשימוש על פי מחיר, ניתן לראות שלכל קבוצת מנות משויכת מנה המשתייכת אליהן על פי סכום מחיר המנה ובנוסף ניתן לראות את דירוג הדימיון. שיוך מנות על פי מחיר נותן לנו אינדיקציה לקבוצת מנות לפי מחיר, שכן לקוח שמזמין בדרך כלל באיזור סכום כסף מסוים, אפשר להציע לו את המנות המשויכויות למנה שאותה הזמין.

שימוש נוסף של האלגוריתם – הוא מציאת מסעדות דומות על פי דירוג. האלגוריתם אוסף את עשר המסעדות הכי קרובות אליו בדירוג, כך שניתן להשתמש באלגוריתם זה בכדי להמליץ ללקוח איזו מסעדות בעלות דירוג טוב ובעלות דימיון בדירוג למסעדות שהוא אוהב.

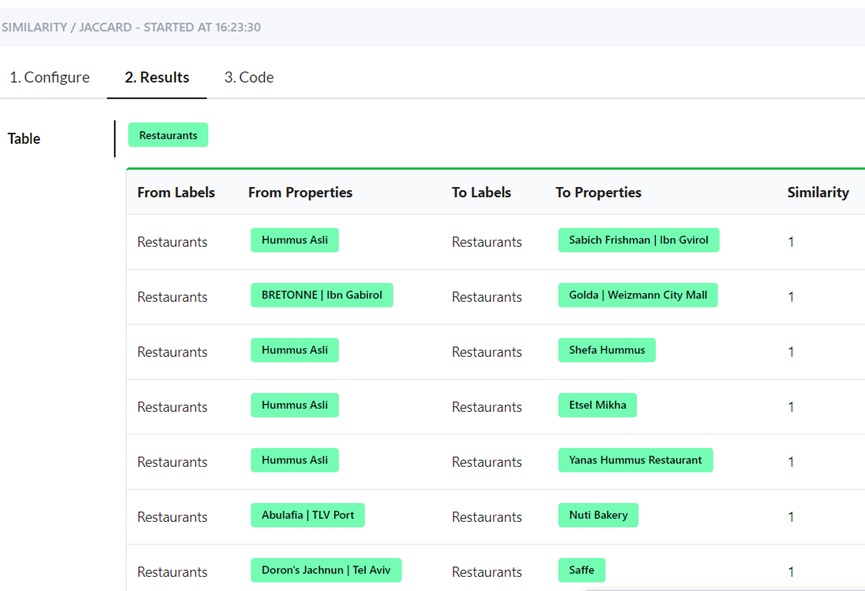


**Jaccard Similarity** - מודד קווי דמיון בין קבוצות. מוגדר כגודל הצומת חלקי גודל האיחוד של שתי קבוצות.

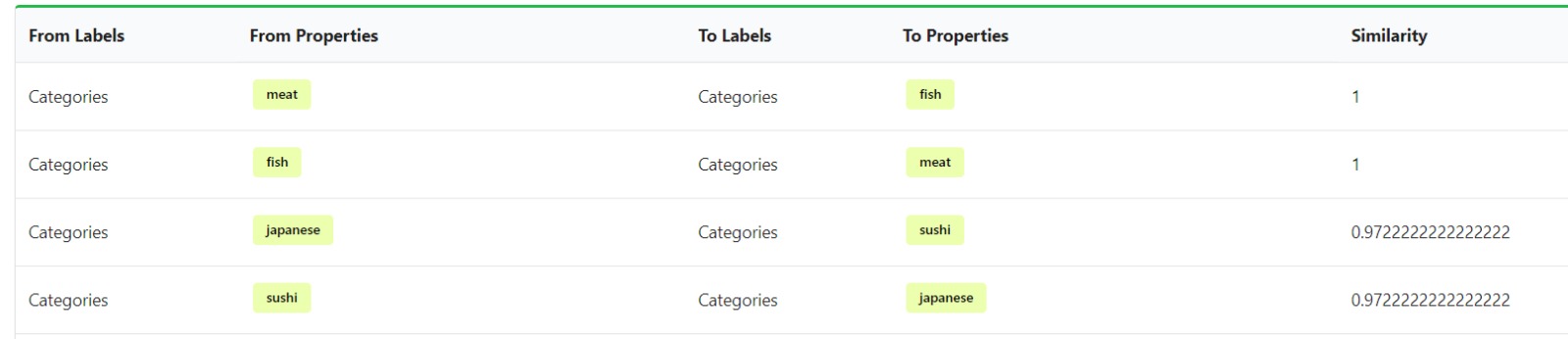
בעזרת אלגוריתם זה ניתן לראות דמיון בין הזמנות. בנוסף ישנו פרמטר של ציון לדימיון.  
לדוגמא, ככל שציון הדימיון יותר גבוה, כך הפריטים מההזמנה יותר דומים. ניתן לנתח את כמות ההזמנות הדומות בין לקוח ללקוח וכך אפשר להשתמש בזה כדי להמליץ ללקוח מנות חדשות שלא ניסה על פי הדימיון של ההזמנות שלו להזמנות אחרות.



עוד שימוש באלגוריתם זה, הוא בבחירת מסעדות דומות. ניתן להציע ללקוח מסעדה דומה או לייצר רשימת מסעדות דומות למסעדה שהוא אוהב על פי ציון הדימיון.



שימוש נוסף, המלצת קטגוריות דומות, אם לקוח אוהב להזמין הרבה מקטגוריה מסויימת, נוכל להציע לו קטגוריה דומה.

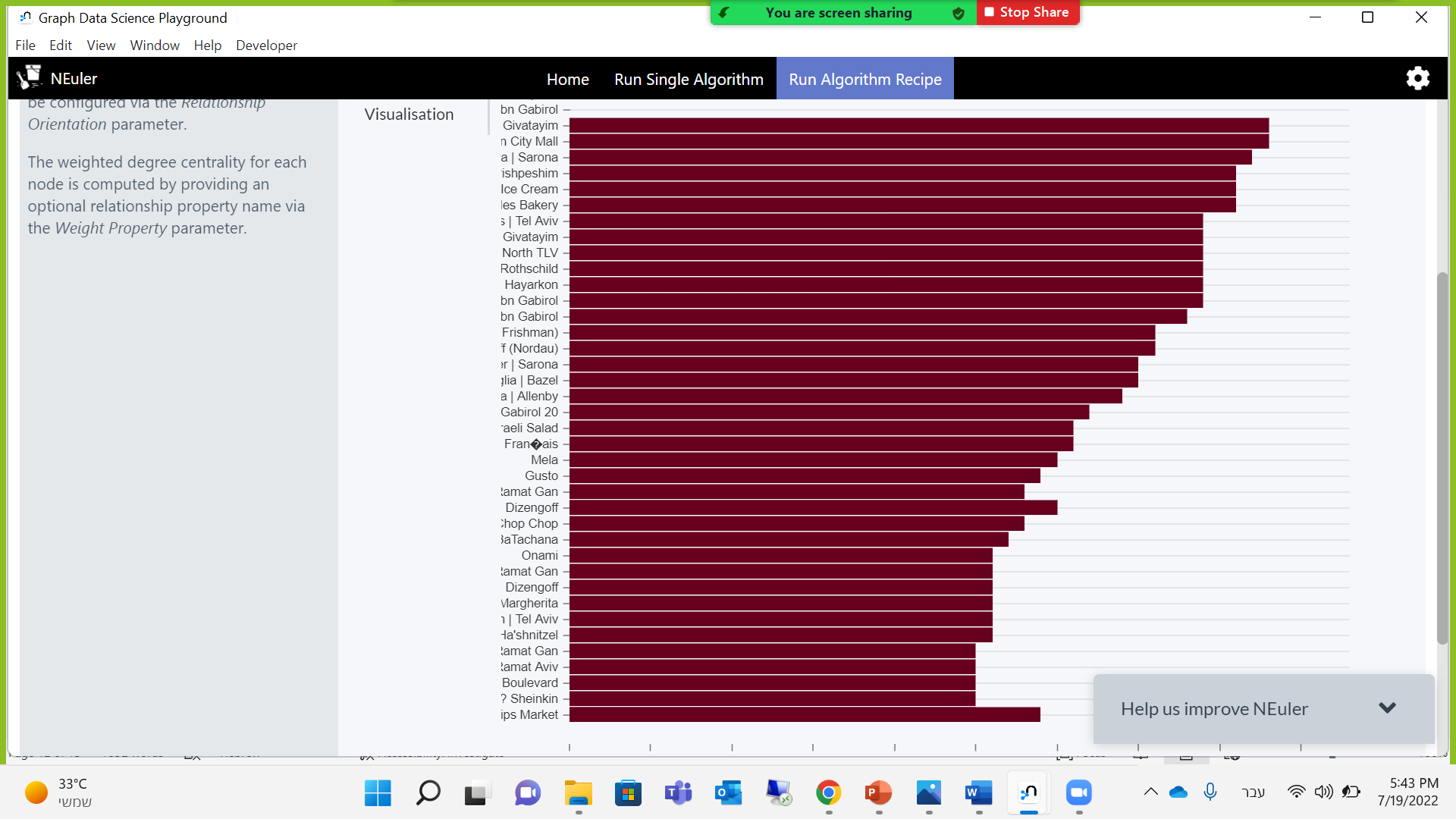


**Degree Centrality** - מוצא את הצמתים המשפיעים ביותר או המרכזיים ביותר בגרף בהתבסס על מספר הקשרים שיש לצומת

Chart

Description automatically generatedבטבלה זו ניתן לראות את הקטגוריות הכי משפיעות. במילים אחרות, הקטגוריות שהוזמנו מהן הכי הרבה הזמנות. נרצה לקחת אותן בחשבון כאשר נרצה להמליץ ללקוח מאיזה קטגוריות כדאי לו להזמין. נעדיף להמליץ לו על קטגוריה שהיא יותר משפיענית מאשר לא, כי כך סיכויי ההתאמה אליו יגדלו.

באופן דומה, הרצנו את אותו אלגוריתם על המסעדות, כדי לראות איזו מסעדות הן הכי משפיעניות, כמובן שניתן להשתמש בזה ולהציע ללקוח מסעדות פופולריות.



**Overlap** – מודד חפיפה בין שני משתנים ומראה את ציון הדימיון ביניהם.

בעזרת שימוש באלגוריתם זה אפשר לראות חפיפה בין סוגי מנות, ניתן להמליץ ללקוח מנה דומה למנה שהזמין על פי דירוג הדימיון.

**Graphical user interface, text, application

Description automatically generated**

**סיכום ומסקנות:**

עם הביקוש הגובר למשלוחי מזון, חברות סטארט-אפ וחברות משלוחי מזון יצטרכו להשתמש בBig Data Analytics וב .Data Science כמו כן, אלו יכולים לעזור לכסות מגוון נושאים. זה כולל תעריפי משלוח, זמני הכנת מזון ומסלולי משלוח.

מסעדות וחברות משלוחי מזון שאינן משתמשות בניתוח ביג דאטה מפספסים הזדמנות משתלמת להגביר את החזר ההשקעה שלהן ולשפר את שביעות רצון הלקוחות. יתרה מכך, יהיה קשה לחזור אחורה בזמן, לשחזר את הנתונים ולנתח אותם אם תעשיית המזון לא תוכל לאגור את הנתונים הגדולים שלה.

להלן כמה מסקנות שעלו מן הפרוייקט:

1. עבודה על התוכנה של Neo4j הייתה מאתגרת וחדשה עבורנו, לקח לנו זמן להבין כיצד להזין את הנתונים ליצירת הNodes וכמו כל שפה חדשה לוקח זמן להבין כיצד לכתוב כמו שצריך. אך השימוש בכלי ויזואלי בהחלט התנה יתרון עם שימוש הדאטה, בעזרת הדרך הויזואלית היה ניתן להבין יותר טוב את הנתונים ואיך כל טבלה מתקשרת לאחרת בנוסף ניתן היה לעצב ויזאולית את הגרפים בכדי שיווצר משהו יותר מובן עבור כאלו שלא היו חלק מהפרויקט.
2. השימוש בתוכנה של האלגוריתמים בneo4j לעומת זאת הייתה יותר מאתגרת ונראה כי בכדי להשתמש בכל האלגוריתמים שמוצעים שם יש לעשות התאמות לנתונים, מה שלא היה רלוונטי עבורנו, לכן הרגשנו שמבחינת הפן הזה היה אולי יותר כדאי להשתמש בכלי אחר. העבודה עם האלגוריתמים הייתה נחמדה מבחינת גיוון האלגוריתמים, שלל הכלים לכל אלגוריתם והעובדה שניתן לראות את התוצאות באופן ויזואלי וברור אך לא היה ניתן להשתמש בתוצאות הנתונים להרצת ניתוח הנתונים והסקת מסקנות. לסיכום היינו ממליצות בneo4j להרצת אלגוריתמים ולהצגת נתונים באופן ויזואלי אך התוכנה פחות מתאימה לניתוח נתונים.
3. ניתן להסיק מהרצת האלגוריתמים כי יש הרבה דמיון ומדדים אחרים בין לקוחות, הזמנות, מסעדות וקטגוריות ומאפיינים אלו אכן מדידים. מסקנותינו מכאן היא שניתן לשפר מערכות המלצה של אוכל ע"י שימת דגש על הלקוחות – חלוקת הלקוחות לקבוצות שונות לפי דמיון תכונותיהם ומאפיינים של ההזמנות שלהם, ובנוסף להתחשב במאפיינים אחרים של שאר הדאטה כמו שהצגנו כאן במסמך ולייצר מערכת המלצה טובה יותר למשתמש ואף לחברה.

**דברים לשיפור:**

מאחר והייתה בעיתיות עם התוכנה של neo4j והדאטה המוגבל שקיבלנו נאלצנו לוותר על ניתוח של שדות חשובים. מבחינת ניתוח מיקום, היה ניתן לבצע סיווג של לקוחות לפי איזורים ולהתאים להם מסעדות רלוונטיות באיזור, כמו לדוגמא אם לקוח הזמין ממסעדה מסוימת שקרובה אליו, היה ניתן להציע ללקוח אחר שגר קרוב – להזמין ממסעדה זו.

בנוסף לא הייתה לנו אפשרות לעבוד על התאריכים והשעות בתוכנה מכיוון שהיא לא תמכה בפורמט. במידה והייתה תומכת היה ניתן לבצע ניתוחים של השעות כדי לתת ללקוח אינדיקציה לזמן הגעה משוער של ההזמנה על פי ניתוח ימי עומס. כך שתהיה שקיפות בין המסעדה ללקוח לגביי זמן ההגעה.