**FoodFinder**

**פרוייקט SQL - מסמך תיעוד**

**מגישים:**

**חן אילון - 201617032**

**גל קשי - 204572861**

**יוגב ניומן - 305541021**

**בן בלייר - 302678446**

תוכן

[מבוא 2](#_Toc535757407)

[מאגר הנתונים 2](#_Toc535757408)

[תרשים 2](#_Toc535757409)

[טבלאות ואינדקסים 2](#_Toc535757410)

[מבנה הDB 4](#_Toc535757411)

[הData שבמאגר הנתונים 6](#_Toc535757412)

[אופטימיזציות? 6](#_Toc535757413)

[ספריות וחבילות חיצוניות בשימוש 6](#_Toc535757414)

[אפלקציה ושליפות SQL 7](#_Toc535757415)

[השליפות שנעשות בDB 7](#_Toc535757416)

[אפלקציה - מבנה הקוד 7](#_Toc535757417)

[ספריות וחבילות חיצוניות בשימוש 7](#_Toc535757418)

[הסבר כללי על היישום 7](#_Toc535757419)

# מבוא

"זמן הוא מה שאנחנו רוצים יותר מכל, אך גם הדבר שבוא אנו משתמשים באופן הכי גרוע" - וויליאם פן.

במציאות של היום לפעמים אין לנו מספיק זמן ללכת לקניות, או למצוא את המתכון הנכון לדיאטה שלנו.

לפעמים אנחנו צריכים מתכון מהיר עם מרכיבים שיש לנו במקרר, ולפעמים אנחנו צריכים מתכון שלא מכיל מרכיבים מסויימים, או ששייך לדיאטה כלשהי.

בשביל להקל על החיפוש יצרנו את FoodFinder, אפלקציה שמאפשרת שליפה חכמה על מתכונים.

FoodFinder מאפשרת שימוש מיטבי בזמן המשתמש, במרכיבים שנמצאים במקרר שלו ובדיאטות שלו.

האפלקציה מאפשרת שליפה על מתכון לפי זמן ההכנה שלו, מרכיבים שנמצאים או לא נמצאים בו ולפי דיאטות.

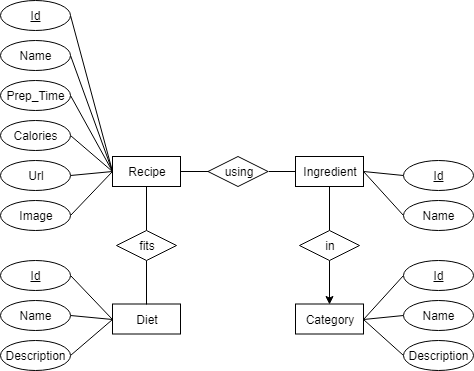
האפלקציה מאפשרת גם שליפות חכמות יותר המותאמות למצבים יומיומיים כגון:

שליפה על 3 מתכונים שניתן להכין יחד בשעתיים - למקרה שיש זמן מוגבל להכין ארוחה בת כמה מנות.

שליפה שמאפשרת לבחור מתכונים ולגזור מהן רשימת קניות של מצרכים.

# מאגר הנתונים

## תרשים



## טבלאות ואינדקסים

להלן רשימת הטבלאות בDB שלנו.

סימונים:

* מפתח ראשי של טבלה יסומן ע"י קו תחתון
* אינדקס מסומן ע"י צבע ירוק
* אינדקס מסוג FULLTEXT מסומן בצבע כחול

**Tables:**

Category(Id, Name, Description)

Diet(Id, Name, Description)

Ingredient(Id, Name, Category\_Id)

Recipe(Id, Name, Prep\_Time, Calories, Url, Image)

Recipe\_Diet( Recipe\_Id, Diet\_Id)

Recipe\_Ingredient(Recipe\_Id, Ingredient\_Id, Weight, Description)

רוב השדות בטבלאות מובנים.

נסביר על השדות בטבלה Recipe:

* Id - מזהה יחודיי של המתכון
* Name - שם המתכון
* Prep\_Time - זמן ההכנה של המתכון
* Calories - מספר הקלוריות עבור המתכון וכמויות הרכיבים המוכלות בו.
* Url - קישור לאתר שמכיל את המתכון המלא
* Image - קישור לתמונה של המנה

**Foreign Keys:**

* Ingredient.Category\_Id -> Category.Id
* Recipe\_Diet.Recipe\_Id -> Recipe.Id
* Recipe\_Diet.Diet\_Id -> Diet.Id
* Recipe\_Ingredient.Recipe\_Id ->Recipe.Id
* Recipe\_Ingredient.Ingredient\_Id -> Ingredient.Id

נשים לב כי לכל מתכון, רכיב, דיאטה וקטגורית אוכל יש Id, על מנת שההצמדות יהיו קלות יותר.

כמו כן, מפורטים לעיל המפתחות הזרים שמטרותיהם להצמיד בין טבלאות.

אינדקסים נוספים שאינם מפתחות ראשיים או זרים:

* Ingredient.Name - מאפשר חיפוש קל על רכיבים במתכונים.
* Recipe.Name - מאפשר שליפת טקסט על שם מתכון. לדוגמה - מתכונים של עוגות.
* Recipe.Prep\_Time - מאפשר שליפה קלה על זמן הכנה של מתכונים.
* Diet.Name - מאפשר שליפה נוחה על טבלת הדיאטות, אם רוצים לחפש דיאטה מסוימת.
* Category.Name - מאפשר שליפה נוחה על סוגי קטגוריות של רכיבי מזון.

## מבנה הDB

נשים לב כי התלויות בDB שלנו הן:

Recipe.Id -> Recipe.Name, Recipe.Perp\_Time, Recipe.Calories, Recipe.Url, Recipe.Image

Recipe.Id -> Diet.Id

Recipe.Id -> Ingredient.Id

Recipe.Id, Ingredient.Id -> Recipe\_Ingredient .Weight, Recipe\_Ingredient.Description

Diet.Id -> Diet.Name, Diet.Description

Category.Id - > Category.Name, Category.Description

Ingredient.Id - > Ingredient.Name, Category.Id

נסתכל על אוסף השדות:

R = { Recipe.Id ,Recipe.Name, Recipe.Perp\_Time, Recipe.Calories, Recipe.Url, Recipe.Image, Diet.Id, Ingredient.Id, Recipe\_Ingredient.Weight, Recipe\_Ingredient.Description, Diet.Name, Diet.Description, Category.Id , Category.Name, Category.Description, Ingredient.Name}

קל לראות שיש הפרה של BCNF.

נפתור את הBCNF ונראה כי חלוקת הטבלאות שלנו תואמת לBCNF.

סתירה: Category.Id - > Category.Name, Category.Description

פירוק:

R1 = {Category.Id, Category.Name, Category.Description}

R2 = { Recipe.Id ,Recipe.Name, Recipe.Perp\_Time, Recipe.Calories, Recipe.Url, Recipe.Image, Diet.Id, Ingredient.Id, Recipe\_Ingredient.Weight, Recipe\_Ingredient.Description, Diet.Name, Diet.Description, Category.Id , Ingredient.Name}

R1 היא BCNF.

R2 לא, סתירה: Diet.Id - > Diet.Name, Diet.Description

פירוק:

R1 = {Category.Id, Category.Name, Category.Description}

R2 = {Diet.Id, Diet.Name, Diet.Description}

R3 = { Recipe.Id ,Recipe.Name, Recipe.Perp\_Time, Recipe.Calories, Recipe.Url, Recipe.Image, Diet.Id, Ingredient.Id, Recipe\_Ingredient.Weight, Recipe\_Ingredient.Description, Category.Id , Ingredient.Name}

R1, R2 הן BCNF.

R3 לא, סתירה: Ingredient.Id - > Ingredient.Name, Category.Id

פירוק:

R1 = {Category.Id, Category.Name, Category.Description}

R2 = {Diet.Id, Diet.Name, Diet.Description}

R3 = { Ingredient.Id , Ingredient.Name, Category.Id }

R4 = { Recipe.Id ,Recipe.Name, Recipe.Perp\_Time, Recipe.Calories, Recipe.Url, Recipe.Image, Diet.Id, Ingredient.Id, Recipe\_Ingredient.Weight, Recipe\_Ingredient.Description}

R1, R2,R3 הן BCNF.

R4 לא, סתירה:

Recipe.Id, Ingredient.Id -> Recipe\_Ingredient .Weight, Recipe\_Ingredient.Description

פירוק:

R1 = {Category.Id, Category.Name, Category.Description}

R2 = {Diet.Id, Diet.Name, Diet.Description}

R3 = {Ingredient.Id , Ingredient.Name, Category.Id }

R4 = {Recipe.Id, Ingredient.Id, Recipe\_Ingredient .Weight, Recipe\_Ingredient.Description }

R5 = {Recipe.Id ,Recipe.Name, Recipe.Perp\_Time, Recipe.Calories, Recipe.Url, Recipe.Image, Diet.Id, Ingredient.Id}

כעת כל הטבלאות הן BCNF.

נרצה לבצע מספר טיובים על מבנה הטבלאות, כדי שלא יכילו הרבה שורות מיותרות, וגם יתאימו יותר למבנה השליפות שלנו:

1. נוריד מR5 את השדה Ingredient.Id, כיוון שההצמדה שקיימת בטבלה R4 מספקת לנו מספיק מידע.
2. נחלק את R5 ל2 הטבלאות הבאות:

R5 = { Recipe.Id ,Recipe.Name, Recipe.Perp\_Time, Recipe.Calories, Recipe.Url, Recipe.Image}

R6 = { Recipe.Id , Diet.Id}

מכיוון שלכל מתכון יכולות להיות מספק דיאטות (Many to Many), אך שאר השדות הם בעלי משמעות אחת (Many to One) - עדיף לפרק את הtuples ל2 סוגים.

העיצוב הסופי מקיים BCNF.

בנוסף, העיצוב מותאם לשליפות שהאפלקציה מאפשרת - שליפות מתכונים לפי דיאטות, רכיבים וקטגוריות מזון.

על השליפות נפרט בהמשך המסמך.

## הData שבמאגר הנתונים

את המידע שהכנסנו למאגר הנתונים השגנו מ3 מקורות:

* Edamam API -

Edamam מציעים שליפות חינמיות על מאגר המתכונים שברשותם באמצעות "Get Requests" שמחזיר מידע בJSONים. מהמידע שהAPI מחזיר סיננו את המידע שמעניין אותנו וחילקנו אותו לטבלאות.

בנוסף למתכונים, שלפנו מהמאגר לעיל סוגי דיאטות.

הקשיים העיקריים שנאלצנו לעבוד איתם בAPI הזה:

* הגבלה על כמות השליפות בפרק זמן מסויים.
* פרסור של רכיבים במתכונים - הAPI מחזיר רכיבים בצורה של משפט שכולל מספרים, מידות, רכיבים ופעולות, בפורמטים משתנים ולא מוגדרים.
* TheMealDB API -

מאגר זה כולל פחות מתכונים באופן משמעותי, לכן לא השתמשנו בו לצורך זה.

מהמאגר הזה שלפנו כ-500 רכיבים של מתכונים. במאגר רכיבים זה השתמשנו על מנת לבנות אוסף שאילתות לשליפה מ-Edamam. כמו כן המרכיבים נשמרו בטבלה Ingredient.

גם השליפות מהמאגר הזה נעשות באמצעות "Get Requests" שמחזיר מידע בJSONים.

* טיובים -
* 500 המרכיבים ששלפנו לא היו מספיק מכלילים בשביל לפרסר את כל רכיבי המתכונים שהורדנו. לכן, נאלצנו "לכרות" מהרכיבים הלא מפורסרים שקיבלנו, רכיבים מפורסרים. באופן זה הגדלנו את רשימת הרכיבים לרשימה שמונה 1308 רכיבים.
* בנוסף, על מנת לאפשר שליפות מתכונים שמכילים קטגוריות של רכיבים (לדוגמה - מתכונים עם ירקות), מיינו את הרכיבים ל14 קטגוריות שונות.

סה"כ הורדנו כ-30,000 מתכונים.

(הקוד לשליפת המידע מצורף במקום הנדרש)

## ספריות וחבילות חיצוניות בשימוש

בשימוש ביצירת הDB וכריית המידע -

השתמנו בספריות פייתון הבאות:

* requests
* json
* pandas
* datetime
* time
* os
* re
* math
* sqlalchemy (create\_engine)

# אפלקציה ושליפות SQL

## השליפות שנעשות בDB

TODO - להוסיף הסבר של 7 השאילתות המסובכות (או כמה שעשינו, לפחות 7).

הסבר נוסף על בחירות אינדקסים?

## אפלקציה - מבנה הקוד

TODO

## ספריות וחבילות חיצוניות בשימוש

TODO

## הסבר כללי על היישום

TODO - תרשים זרימה של general flow והסבר שימוש כללי.