

# Final Project

Data Analyst – Appelseeds organization

Sapir Cardona, Anan Sirhan, Tomer Dahan, Ishay Reiner



# Datasets Considered

## Soil Pollution and Health Impacts

נפסל כיוון שהיה נישתי מדי ולא מייצג תחום מבוקש בשוק או בעל יישומים רחבים.

## Job Descriptions Dataset

הכיל מעל מיליון רשומות עם מגוון התפלגויות ועמודות טקסט חופשי, מה שהיה מצריך ניקוי נרחב ומורכב (ספליטים, טקסטואלי).

## Customer Personality Analysis

✓ רצון להתנסות בתחום המרקטינג

✓ דאטה סט ממוצע בגודלו עם טבלה אחת – אתגר לבניית סכמת DWH באופן אסטרטגי מתוכנן

✓ מאפשר עבודה עם קלאסטרים – משהו חדש שטרם התנסינו ורלוונטי לפרוטפוליו

# Data Tools & Technologies



## Excel

חקירה ראשונית של העמודות וסוגי הערכים  
(קובץ variables Inventory בנספחים)



## SQL

ניקוי הנתונים, טיפולים בערכים חריגים, שינוי שמות עמודות ויצירת טבלאות  
staging וסכמת כוכב של DWH



## Python (pandas)

ניקוי הדאטה, ניתוח משתנים, יצירת שדות חדשים כמו גיל, סך ההוצאה, ולביצוע  
אנליזות מתקדמות



## Python (seaborn)

לצורך ביצוע ויזואליזציות כחלק מ-EDA, כולל התפלגויות, קשרים והשוואות בין  
קבוצות, שסייעו בהבנה ראשונית של הדאטה והעלו תובנות ראשוניות



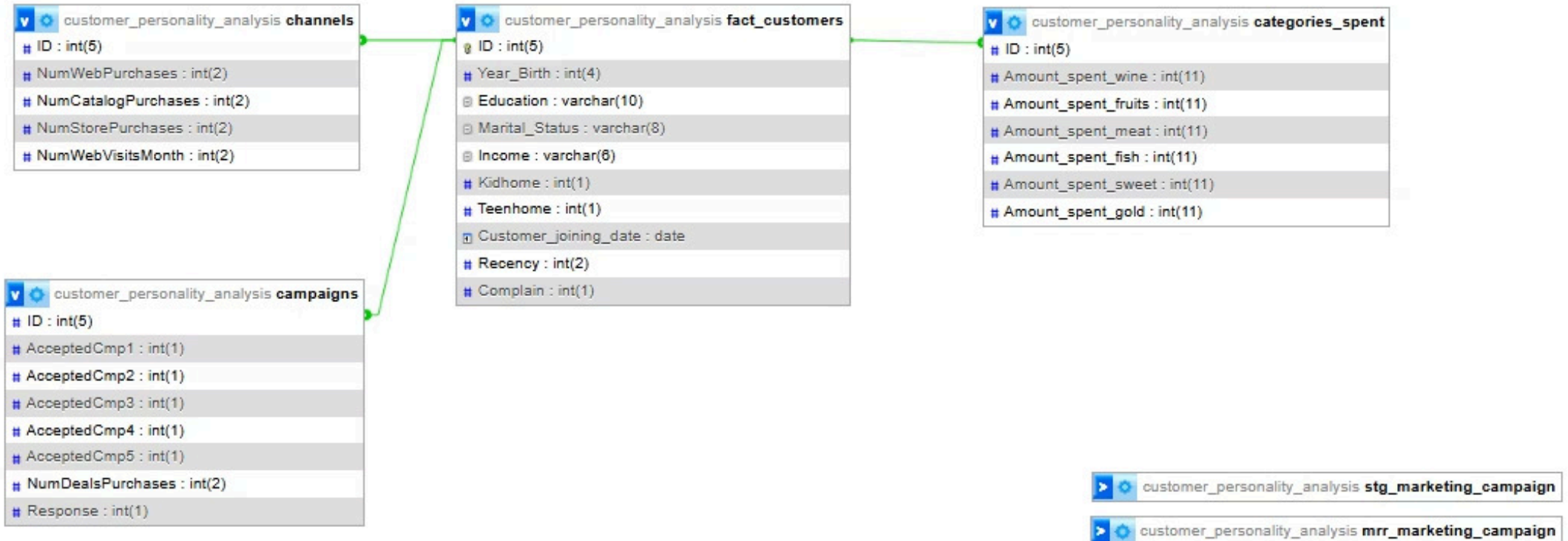
## Tableau

לבניית דשבורד אינטראקטיבי המאפשר חקר של התפלגויות, פילוחים ודפוסי  
היענות, תוך אפשרות לסינון לפי קלאסטרים וקבלת תובנות ויזואליות

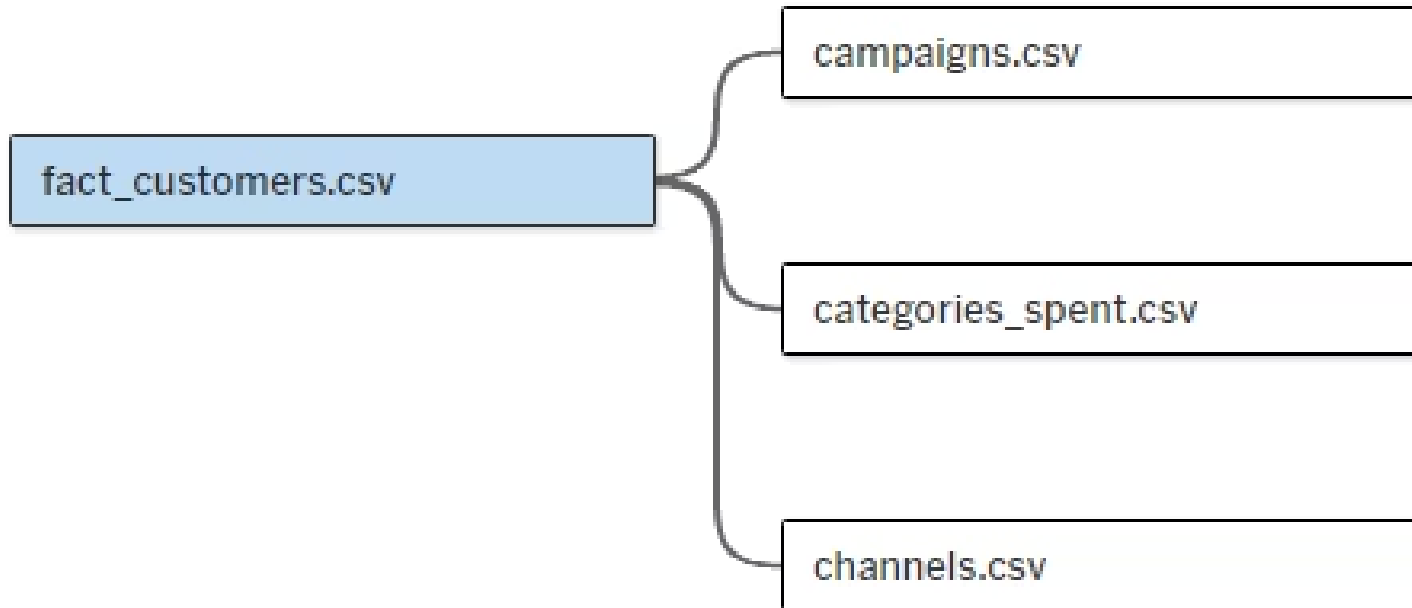
שילוב הכלים אפשר לנו לעבור תהליך מלא מקצה לקצה - מניקוי וארגון הדאטה ועד  
להצגת תובנות עסקיות רלוונטיות ואקטיביות בקונטקסט שיווקי

# ERD – SQL

## Data Warehouse - Star Schema



# Tableau – Data Relationships



# Define Your Audience

קהל היעד העיקרי לדוח ולדשבורד הוא מנהלי מחלקת השיווק של הרשת – האחראים על תכנון קמפיינים שיווקיים, יצירת מבצעים והנעת לקוחות לרכישה חוזרת – הן בחנויות הפיזיות והן באתר האונליין. בחירה זו נובעת מהאופי של הדאטה, שמכיל מידע עשיר על מאפייני לקוחות והרגלי צריכה, באופן שמתחבר ישירות לשאלות שיווקיות פרסונליות.

## מחלקת השיווק עוסקת בבניית קמפיינים מותאמים לקהלים שונים, במטרה:

- להגדיל את כמות הרכישות
- להעלות את הרווחיות הכוללת (AOV)
- להתאים מסרים, הטבות ומבצעים לקבוצות לקוחות שונות

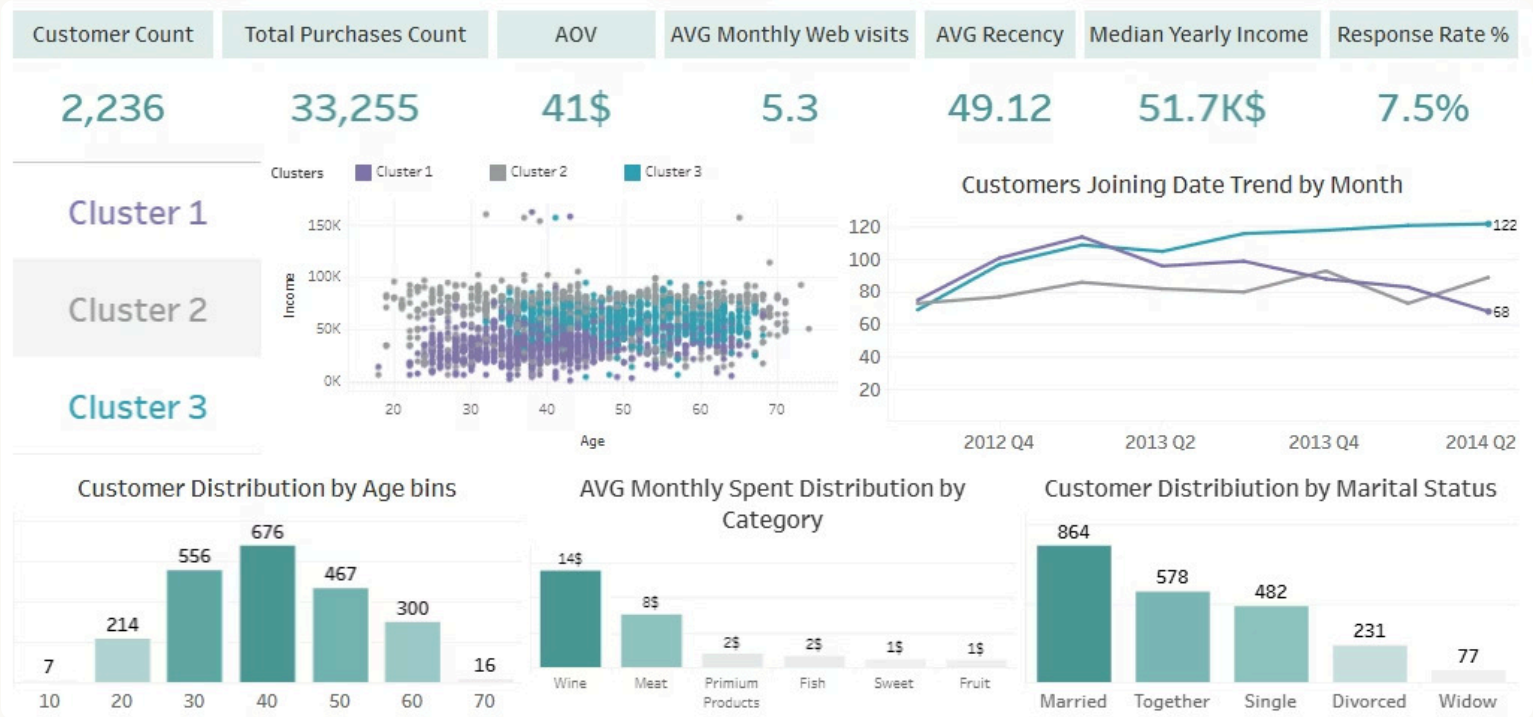
בשאיפה להבין אילו לקוחות מגיבים לקמפיינים ואילו לא, ולזהות מה מאפיין כל קבוצה – כדי לדייק את המסרים השיווקיים בעתיד. נכון לשלב זה, הדגש הוא בעיקר על שיפור הפעילות מול לקוחות קיימים, מתוך כוונה להגדיל את שווי הרכישה ולהשפיע על ההיענות.

## שאלות עסקיות מרכזיות:

1. אילו קבוצות לקוחות נוטות להגיב לקמפיינים שיווקיים, ואילו לא?
2. מהם דפוסי הרכישה לפי מאפייני לקוח (כמו השכלה, סטטוס משפחתי והכנסה)?
3. באילו ערוצי רכישה (אתר, חנות, קטלוג) יש שימוש גבוה יותר, ומה כדאי לחזק?

# Tableau Dashboard Development

הדשבורד בנוי במבנה תלת-שכבתי ברור, המאפשר מעבר הדרגתי בין סקירה כללית, ניתוח מאפייני קלאסטרים, וניתוחים השוואתיים.



## שכבה עליונה – KPI

- מספר לקוחות
- מספר רכישות
- ערך AOV
- ממוצע היענות לקמפיינים
- ממוצע Recency
- ממוצע ביקורים חודשיים באתר
- חציון הכנסה שנתית

## שכבה אמצעית – דינמיקה וזיהוי קבוצות

- גרף **Line Chart** המציג את מגמת הצטרפות הלקוחות לפי רבעונים
- גרף **Cluster Attribute Breakdown** - תצוגה השוואתית של ממוצעי מאפייני כל קלאסטר:
  - גיל (Age)
  - הכנסה שנתית ממוצעת
  - מספר ילדים (Kidhome) ומתבגרים (Teenhome) בבית
  - מספר רכישות מהחנות, מהאתר, מהקטלוג, וממבצעים
- **שורת פילטרים עליונים:** מאפשרת סינון לפי קלאסטר

## שכבה תחתונה – ניתוחים השוואתיים

- גרף **Bar Chart**: התפלגות לקוחות לפי קבוצות גיל
- גרף **Bar Chart**: הוצאה ממוצעת חודשית לפי קטגוריות מוצרים
- גרף **Bar Chart**: פילוח לקוחות לפי סטטוס משפחתי

# Key Findings

במהלך הניתוח זיהינו **שלוש קבוצות מובהקות של לקוחות (קלאסטרים)**, המאופיינות בדמיון בפרמטרים כגון:

- גיל
- הכנסה
- מספר ילדים ומתבגרים בבית
- דפוסי רכישה בחנות, באתר, במבצעים ובקטלוג

עבור כל קלאסטר ניתחנו:

- את המאפיינים הדמוגרפיים המרכזיים
- את הרגלי הצריכה והקטגוריות המובילות בהוצאה
- את רמת ההיענות לקמפיינים שיווקיים

הניתוח הזה איפשר לנו לגבש **המלצות שיווקיות ממוקדות ופרסונליות** עבור כל קבוצה, במטרה להגביר מעורבות, לשפר התאמה של מסרים שיווקיים ולהוביל לעלייה בפעילות והרווחיות



# Limitations & Challenges

## עמודות לא ברורות

עמודה לא ברורה של מוצרי פרימיום לא כללה הסבר או תיוג ברור למוצרים שנכללים בה  
שתי עמודות נוספות הכילו ערכים זהים בכל הרשומות, ולכן הוסרו (drop) בשל חוסר אינפורמטיביות



## ערכים חריגים

נמצאו גילאים קיצוניים (מעל גיל 100) – הוחרגו מהדשבורד כדי לא לעוות את הפילוחים  
אותרה רשומת הכנסה חריגה (מעל \$600,000) – הוצאה מהגרפים מטעמי תצוגה



## נתונים חסרים

היו סטטוסים משפחתיים לא מזהים (YOLO, Absurd) שהומרו ל "Unknown"  
בעמודת Income אותרו 24 ערכי NULL, שמולאו לפי ממוצע כלל הדאטה – פעולה פשוטה אך לא אופטימלית  
(הערכים החסרים נקלטו בSQL כ- " " ולא הוחשבו NULL, לעומת python שבוצע זיהוי מידי אוטומטי)



## חוסר נתוני רווחיות

לא הייתה לנו גישה לרווחיות בפועל של המוצרים - מה שמנע ניתוח עומק כלכלי  
נאלצנו להסתמך על מחירים וסך רכישות, ולא על מדדי רווח אמיתיים



# New Tool Learning

במהלך הפרויקט בחרנו לשלב כלי שלא השתמשנו בו קודם – **t-SNE** (t-distributed Stochastic Neighbor Embedding) מתוך ספריית `sklearn.manifold`.

מדובר בכלי להפחתת מימדים, אשר מאפשר להציג נתונים מרובי תכונות (כגון גיל, מספר ילדים, סוגי רכישות וכו') במישור דו־ממדי – ובכך מאפשר **הבנה ויזואלית של מבנה הקלאסטרים**.

השימוש ב־t-SNE שיפר משמעותית את תהליך העבודה שלנו בשלב ה־EDA המתקדם, בכך שאיפשר:

- לזהות קבוצות מובחנות של לקוחות בצורה אינטואיטיבית
- לבחון האם הקלאסטרים שנבנו הם בעלי משמעות ויזואלית
- היווה בסיס לקראת יצירת קלאסטרים ב- Tableau בניתוח אנליטי מבוסס Python

# Future Needs

השלמת ערכים חסרים בצורת אופטימלית

השלמת ערכים חסרים לפי ממוצע בתוך קבוצת הקלאסטר

הרחבת מספר הקלאסטרים

לבחון חלוקה ליותר קלאסטרים – ייתכן ש-3 קבוצות בלבד מפספסות תתי-קבוצות מעניינות, והגדלת מספר הקלאסטרים תאפשר שיווק ממוקד יותר.

העשרת הדאטה בנתוני רווחיות

העשרת הדאטה בנתוני רווחיות – שילוב מידע עסקי אמיתי לגבי רווח למוצר היה מאפשר להמליץ לא רק מה לקדם, אלא גם מה משתלם לקדם.

בחינת הצלחת קמפיינים עתידיים

לבחון את הצלחת הקמפיינים העתידיים בפועל, בהתאם להמלצות השיווקיות שנבנו לכל קלאסטר.



# Conclusion & Closing Remarks



הפרויקט אפשר לנו לבחון לעומק את התנהגות הלקוחות של הרשת, לזהות קבוצות מובהקות לפי מאפיינים דמוגרפיים והתנהגותיים, ולבנות עבור כל קבוצה המלצות שיווקיות מותאמות אישית. דרך תהליך מקיף של ניקוי, ניתוח, קלאסטרונג והצגה ויזואלית – הצלחנו לחבר בין הדאטה לבין צרכים עסקיים ברורים, בדגש על מיקוד שיווקי, הגדלת רווחיות והבנה טובה יותר של קהל הלקוחות.

# Links & Extra

**Kaggle Dataset Link**

**Google Colab Notebook**

**Tableau Public Dashboard**

**Managers report**

**GitHub Repo**

