**מטלה 3 – בינה עסקית:**

אביב רביד 316018175

נועה דלאל 316045335

עדן בן דוד 206957797

**חלק 1 – STTM:**

1. קובץ STTM מצורף בתיקיית Analysis
2. סכימה ויזואלית המציגה את הSTTM:

Diagram

Description automatically generated

**חלק 2 – Data Mining Techniques:**

1. תהליך
2. טכניקות
3. מדד הדמיון שנציע לפרויקט הינו מדד דמיון בין בתים. התכונות אותן נשווה בין כל שני בתים יהיו התכונות המרכזיות שעונות על שאלת המחקר שבחנו, אשר הן נמצאות בטבלת Fact. בחרנו להשתמש בכל סוגי המדדים שראינו בכיתה, וליצור מדד משוקלל.

* נשתמש במדד Sorensen-Dice שהינו מדד דמיון בין משתנים נומריים בדידים עבור תכונת מספר חדרים (עמודת bedrooms), ותכונת מספר שירותים (עמודת bathrooms).
* נשתמש במדד Jaccard שהינו מדד דמיון בין משתנים קטגוריאליים עבור העיר שבה נמצא כל בית (עמודת city).
* נשתמש במדד Euclidean Distance שהינו מדד דמיון בין משתנים נומריים רציפים עבור גודל הבית במטרים רבועים (עמודת sqm\_living), ועבור המחיר של הבית (עמודת price).

נאחד את המדדים לכדי מדד אחד, נמצא את השקלול של כל המדדים, וכך נמצא את הבתים שלהם יש מדד דמיון הכי גדול ונסיק כי הם הדומים ביותר.

1. עבור שאלה עסקית 1: **Supervised**

האם דירות בעלות 3 חדרים, אשר שטח המגורים שלהן הינו מעל הממוצע של שטחי המגורים, יעלו מעל הממוצע של מחירי הדירות?

השערת H0:

השערת H1:

עבור שאלה עסקית 2: **Unsupervised**

מה מאפיין דירות בעלות 3 חדרים ובעלות שטח מגורים מעל הממוצע?

השערת H0:

השערת H1:

נחליט לקבל לפי חישובים שביצענו, ממוצע מחירי דירות בעלות 3 חדרים. ממוצע מחירי דירות כללי. ממוצע שטח מגורים.

**חלק 3 – שאילתות:**

שאילתות מצורפות בקובץ הנקרא SQL Q בתיקיית Analysis.

**חלק 4 – ניהול גרסאות:**

בוצע, קישור לפרויקט: <https://github.com/EdenBenDavid1/BI_Project>