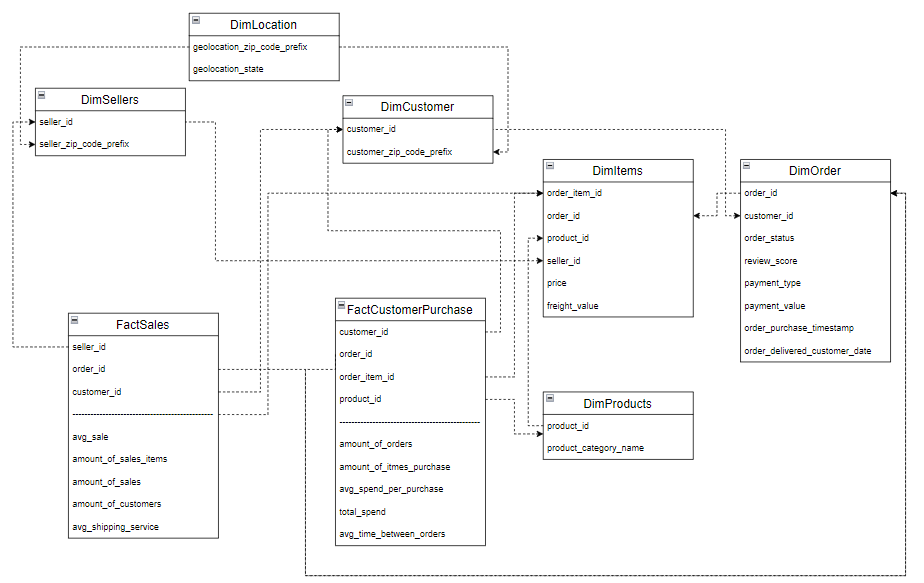
**Homework 2 – Business Intelligence**

|  |  |
| --- | --- |
| **שם** | **ת.ז** |
| תומר בן יהודה | 307854422 |
| אלעד חזיזה | 312326788 |
| עידן כהן | 313522534 |

**חלק 1 - הגדרת DW**

לאחר הגדרת ה-KPIs בש.ב. 1, נתכנן DW שמכיל את כל העמודות שחשבנו שהם רלוונטיות כדי לענות על השאלות:

* מה הוא פוטנציאל ההכנסות של חנות ברשת?
* לאילו קבוצות ניתן לחלק את הלקוחות של החנות? (לטובת קמפיין פרסומי)

1. מידלנו את הDW לפי סכימת GALAXY מכיוון ששאלות המחקר שלנו והKPIs הנגזרים מהם מייצרים שני DOMAINs מרכזיים, מידע אודות מכירות של חנויות ומידע אודות רכישות של לקוחות.
2. 
3. בוצע.
4. החלטנו לחלק את הDW לשתי טבאלות Fact וכתוצאה מכך יצור סכימת GALAXY וזאת בכדי להפריד בין המכירות של חנויות לרכישות של לקוחות. כתוצאה מכך אנליסט אשר ירצה לנתח את הנתונים יתחבר ברמה אינטואיטיבית יותר למחסן הנתונים והשליפות שלו יחסכו בכמות הjoins בעת השליפה עצמה.
5. מצורף.

**חלק 2 - הגדרת ETL**

1. תהליך הETL יבצוע כך:
   1. **EXTRACT**- בשלב זה אנחנו נעתיק את הנתונים מהמקורות שלנו לבסיס נתוני הביניים שלנו. בפועל ניצור טבלאות עם העמודות הרלוונטיות לנו
   2. **Transformation**-ניצור את טבלאות הFACT ובשלב זה הנתונים עוברים את השינויים
   3. **Load**-בשלב זה הנתונים שעברו טרנספורמציה יעלו למחסן הנתונים שלנו.
2. בסעיף זה אנחנו מגדירים את הפעולות שנעשה בETL. שלב זה חשוב שבשביל להגדיר לפני תחילת העבודה מה נעשה ובנוסף לתעד את כל השינויים שעשינו בגרסה הנוכחית על מנת שישאר תיעוד לעבודה שלנו.

תחילה נתאר מאיזה טבלה ועמודה מגיע המידע שעליו נעבוד. נתאר את הסוג (TYPPE) של העמודה. במידה הצורך, נתאר את השינויים שצריך לעשות על המידע הזה. אם לא נצטרך לעשות שינויים, נשאיר ריק. לאחר מכן, נתעד באיזה טבלה המידע יישמר, באיזה שם עמודה ובאיזה סוג. אם ישנם הערות על העמודה, נכתוב.

בקישור המצורף מתוארים כל השלבים הנדרשים לביצוע הETL:

[כאן](https://docs.google.com/spreadsheets/d/1AkkSBprZ2Jtf1A5blGhV_JDl5kFl_9Ok/edit?usp=sharing&ouid=111786400909731518215&rtpof=true&sd=true)

1. מימוש הETL מומש בקישור המתואר בפייתון:

<https://colab.research.google.com/drive/1Ex9DVell7rBqcixmDU-LTigqK2TMSbjS?usp=sharing>