



תוכן עניינים

4	חלק א' - ניתוח הסביבה העסקית ואפיון הנתונים
4	1. ניתוח הארגון, התהליכים העסקיים וקבלת ההחלטות
5	2. תהליך עסקי:
6	3. תרשים BPMN
7	4. תרחישי קבלת החלטות
8	5. הגדרת מדדי ביצוע (KPI's – Key Performance Indicators)
10	6. ניתוח מקורות הנתונים התפעוליים
10	7. תרשים ERD
11	8. מודל טבלאי - בסיס נתונים ראשי
13	9. אפיון של מחסן נתונים טבלאי
16	חלק ב' - אפיון תהליכי ETL
16	1. השלבים בתהליך ה-ETL
19	2. מימוש תהליכי ה-ETL – יישומי הבסיס
19	שלב 1 - MRR
20	שלב 2 - Dimension Staging
21	שלב 4 - Fact Staging
23	שלב 5 - Referential Integrity
23	שלב 6 - Fact Warehousing
25	3. מימוש תהליכי ה-ETL – יישומים מתקדמים
25	מימד משתנה לאט מסוג 2
25	ביצוע פרוצדורה שמורה
26	הוספת פילטר בשלב ה- Staging של טבלת הזמנות:
27	Referential Integrity
29	חלק ג' - ניתוח נתונים וויזואליזציה
29	ניתוח נתונים באמצעות R
32	ויזואליזציה (Tableau Dashboards)
32	Report 1
33	Report 2
34	Dashboard
34	OLAP
35	Sunburst Flavor



35.....	Spider
36	נספחים:
36.....	ERD
38.....	MRR
40.....	STG
41.....	DW

חלק א' - ניתוח הסביבה העסקית ואפיון הנתונים

1. ניתוח הארגון, התהליכים העסקיים וקבלת ההחלטות

על הארגון:

eCreamery – היא חברת אונליין לרכישת גלידות ומוצרים נלווים. החברה מציעה עיצוב אישי של טעמי הגלידה, אריזות הטעמים. יש אינספור אפשרויות שונות לעיצוב האריזה אשר כוללות טקסטרים שונים, לכל אירוע ולכל מחווה, אפשרות להוסיף תמונה וכמובן לבחור את הטעמים השונים ולשדרגם עם תוספות. כמו כן, החברה גם משלחת את ההזמנות במגוון מדינות בארה"ב. אתר החברה <https://ecreamery.com/>.

החברה הוקמה ב-2019 בארה"ב, ממיזוג של שתי חברות קטנות אשר איחדו כוחות מתוך החזון להגשים חלום קינוחי ("dessert dream-making"). חברה אחת שמתמחה באפיית עוגיות תוצרת בית, טעימות בהתאמה אישית לאירועים מיוחדים, והחברה השנייה מתמחה ביצירת טעמי גלידה מקוריים וייחודיים בשילוב של אריזה בעיצוב אישי. האיחוד נועד ליצור מנעד רחב ומודולרי של טעמים ושל קינוח הכולל עוגיות וגלידות, והכל בעיצוב אישי, אפילו האריזה.

המטרה העסקית של החברה היא להרוויח כסף. מטרת האתר הינה הנגשה וריכוז של מגוון האפשרויות השונות, וליצור ממשק אופטימלי להרכבת הקינוח הרצוי ועיצובו בכדי למקסם את כמות ההזמנות, ובאותה העת לאגד את כל ההזמנות, תהליך ההזמנה ומשלוח.

אתגר עיקרי העומד בפני החברה הינו תחזוקה ועדכון שוטפים של בסיס הנתונים בהתאם למלאי המשתנה ובהתאם להזמנות, על מנת לשמור על העדכניות בפני הלקוחות ועל מנת שלא ייווצרו הזמנות של טעם חסר למשל. אתגר נוסף העומד בפני החברה הינו פער הציפיות של הלקוחות בבחירת טעמי הגלידה והזמנתם באופן מקוון מבלי לטעום לפני כמו בגלידריות.



2. תהליך עסקי:

האתר "eCreamery" עוסק במכירת גלידות ברשת, לכן התהליך העסקי והמרכזי בו הוא בחירת המוצר- אם לרכוש מוצר קיים או להרכיב טעם גלידה ייחודי, כולל תוספות ולעצב את האריזה.

בעת כניסה לאתר למשתמש יש סרגל כלים בחלקו העליון של המסך, שם ניתן לראות את המוצרים המוכנים הנמכרים ביותר. למשתמש יש שתי אפשרויות: לרכוש מוצר קיים ומוגמר שעוצב לאירוע מיוחד כמו חגים, ימי הולדת, אירועים ונושאים, או להרכיב גלידה לפי בחירת טעם ותוספות, ועיצוב האריזה לפי בחירת תמונה, טקסט או צבע. ניתן לבחור את מספר היחידות שזמין מכל מוצר. לאחר הוספת כל המוצרים לעגלת הקניות, במעבר לתשלום, מוצג למשתמש סיכום ההזמנה הכולל את המוצרים שבחר ועלותם, כולל המס. מכאן על המשתמש להיכנס למערכת כלקוח רשום או להירשם כלקוח חדש. בעת הרישום כלקוח המשתמש צריך להזין את הפרטים הבאים: אימייל, שם מלא, כתובת מלאה, עיר, מדינה בארה"ב, מיקוד ומספר טלפון. בשלב זה תתחשב עלות המשלוח. הזנת פרטים אלו הכרחית לביצוע התהליך העסקי בשלמותו. לאחר הזנת הפרטים, המערכת תעביר את הפרטים למאגר בסיס הנתונים של החברה, וכשהלקוח ירצה לבצע הזמנה נוספת פרטיו כבר יהיו קיימים ומעודכנים לקראת התהליך העסקי. בשלב הבא יפתח חלון עם פרטי המשלוח, הלקוח יבחר את סוג המשלוח- מהיר או רגיל. חלון התשלום יפתח ובו ימלאו את פרטי התשלום. ההזמנה תושלם כאשר ישלח מייל פרטי ההזמנה.

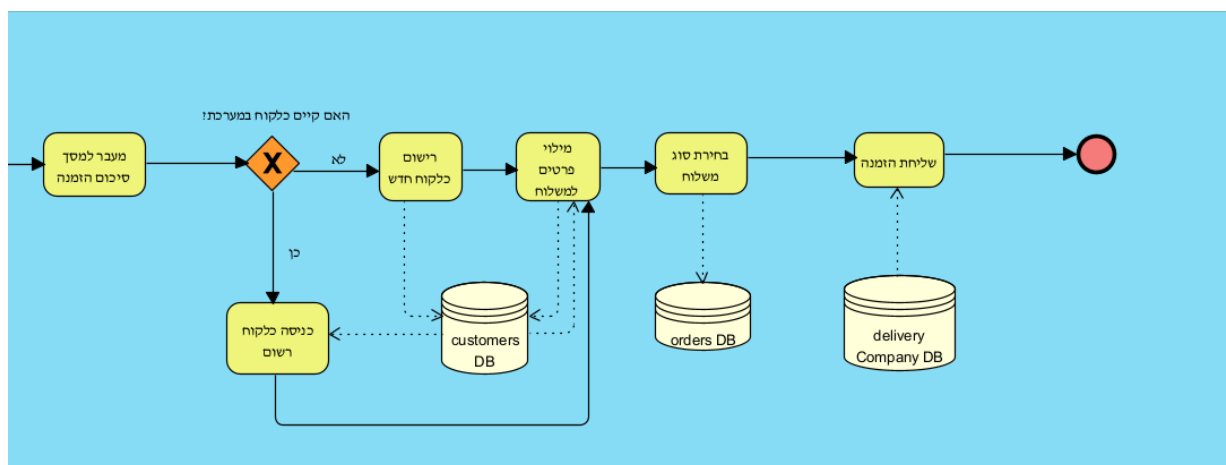
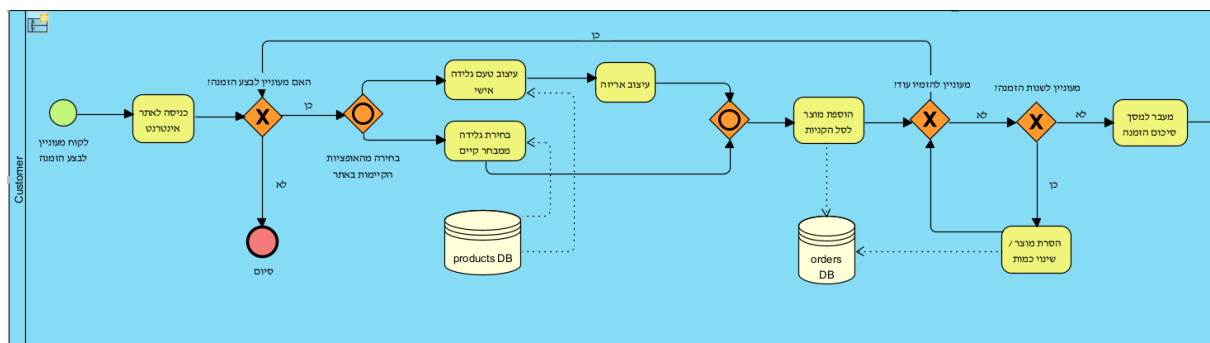
- **בעלי העניין** בתהליך הרכישה הם לקוחות האתר שמעוניינים בגלידה לפי העדפותיהם האישיות. בעלי עניין נוספים הם חברות השליחויות השונות שעובדות בשיתוף פעולה עם החברה.
- **הקלט** בתהליך הרכישה הן שלוש האפשרויות שנתונות לבחירת הלקוח (גלידה ממבחר קיים/גלידה בעיצוב אישי/גלידה מעוצבת לפי אירוע) וכן המוצרים עצמם. קלט נוסף הוא כתובת המשלוח. **הפלט** בתהליך הוא דף סיכום ההזמנה המציג ללקוח את הפריטים אותם בחר, כמויות ועלויות, בדף זה הלקוח צריך לבחור סוג משלוח ואמצעי תשלום.
- **המפעיל** בתהליך זה הוא האתר עצמו, שלפי בסיס הנתונים שלו ניתן לבחור את המוצרים שבהם הלקוח מעוניין.
- **אמצעי הבקרה** בתהליך זה הם וידוא רישום לקוח לאתר, סיווג הפרמטרים למילוי כחובה/רשות, וידוא תקינות אמצעי תשלום.
- **התהליך העסקי** שבחרנו משויך בחלקו לשלב ה"קלט" – במהלך בחירת המוצרים, ובחלקו השני הוא משויך לשלב ה"פלט" – במהלך סיכום ההזמנה והתשלום.

השלבים המרכזיים בתהליך העסקי:

- בחירת מוצרים לרכישה מתוך האפשרויות הקיימות באתר
- כניסה לאתר כלקוח או הרשמה כלקוח חדש
- מילוי פרטים למשלוח
- בחירת סוג משלוח
- ביצוע תשלום ואישור הזמנה

ניתן לראות כי יש צורך בגישה לבסיסי הנתונים של המוצרים באתר, של הלקוחות, עדכון בסיס הנתונים של הלקוחות במידת הצורך, עדכון בסיס הנתונים של ההזמנות. התהליך מסתיים בתוצר של הזמנה חדשה בבסיס הנתונים של ההזמנות.

3. תרשים BPMN:



4. תרחישי קבלת החלטות:

החלטה תפעולית	החלטה ניהולית-טקטית	החלטה ניהולית-אסטרטגית	
בקצרה – באיזו החלטה מדובר? מדוע החלטה חיונית להפעלת התהליך והצלחתו?	מה גובה הקופון שיש להזין ללקוח לא מרוצה מההזמנה? ההחלטה חיונית לצורך שימור לקוחות.	האם להמשיך לעבוד עם חברת שליוחיות מסוימת. ההחלטה חיונית לניהול נכון של הוצאות החברה.	לפתוח מועדון לקוחות VIP של החברה. חיוני לחברה על מנת לתחזק את הקשר עם הלקוחות הרווחים לחברה.
הגדרת מאפייני ההחלטה: <ul style="list-style-type: none"> פונקציית מטרה מרחב אפשרויות בחירה (בינארי? בדיד? רציף?) אילוצים? תלות בהחלטות אחרות? 	מטרה- מקסום רווחי ע"י שימור לקוחות ושמירה על שביעות רצונם. מרחב אפשרויות בדיד- הקופון ניתן באחוזים בין 0 ל-100, מספרים שלמים בלבד. אילוצים- הימנעות מהנחות גבוהות מידי שפוגעות ברווחיות החברה. תלות- בהינתן לסיבת אי שביעות הרצון ומידת החומרה.	מטרה- מקסום רווחי החברה ע"י שיתופי פעולה שמשלמים לחברה מבחינה כלכלית. מרחב אפשרויות בדיד- האם להמשיך לעבוד עם החברה או לא. אילוצים- התנאים שנחתמו בחוזה מול חברת השליוחיות. תלות- מציאת אלטרנטיבות רווחיות יותר באותה המדינה.	מטרה- מקסום רווחי החברה ע"י שדרוג האטרקטיביות והגדלת היקף המכירות. מרחב אפשרויות בדיד- לפתוח מועדון לקוחות או לא. ובמידה ותקבל ההחלטה לפתוח, יש לבנות מודל המאפיין לקוח VIP ואת תנאי המועדון. אילוצים- התקציב שיש לחברה להשקיע בקמפיין, בהקצאת כוח אדם ובמתן הנחות. תלות- הגדרת תנאי הצטרפות, עריכת חוזית, צפי הרווחים, הקצאת כוח אדם ממונה.
אילו נתונים נדרשים לקבלת ההחלטה?	צפייה בחוות הדעת (Reviews) שהלקוח השאיר. בדיקה כמה ימים לקח למשלוח להגיע (ההפרש בין יום ביצוע ההזמנה ליום קבלת המשלוח). מידת הרווחיות של הלקוח (סכום ההזמנות).	תאריך הזמנה ותאריך קבלת המשלוח מטבלת הזמנה. גובה הקנס על ימי איחור, והתשלום החודשי לחברה מטבלת חברת השליוחיות.	שקלול סך הרווחים של החברה (עלויות ייצור, זיכוי לקוחות, תשלום לחברת השליוחיות, הכנסות מההזמנות, קנסות מחברת השליוחיות) מעקב אחר הלקוחות החוזרים והרווחיים.
מי מעורב בתהליך ההחלטה ורשאי לצפות במידע המופק מהנתונים?	נציג שירות לקוחות אשר מוסמך להזין גובה קופון, וכל הדרגים מעליו.	אחראי שיפור תהליכים בחברה, המנהלת שממונה על מערך המשלוחים, מנהל קשרי לקוחות, מנהלת כספים.	הדרג הניהולי הבכיר.
מה היקף השימוש במידע בראייה חודשית?	תדירות גבוהה. יתכן שמספר פעמים בשבוע.	תדירות נמוכה, כאשר עולה בעיה ניכרת, כאשר חוזה עם חברה נגמר, או אחת לשנה.	תדירות נמוכה, החלטה חד פעמית עם השפעה לטווח הרחוק.

5. הגדרת מדדי ביצוע (KPI's – Key Performance Indicators)

1. ממוצע ההכנסה מהזמנה בודדת לתקופת זמן

מדד זה הינו מדד תועלת עבור החלטה תפעולית. עבור מקסום הרווחים בחברה אנחנו נרצה שסך הכנסות מהזמנות יהיו גבוהות ככל האפשר, כלומר אנחנו נשאף לך שהלקוחות יזמינו כמה שיותר מוצרים / מוצרים בעלי ערך יותר בהזמנה בודדת וכך עלות ההזמנה תגדל ותעלה מתקופה לתקופה.

מדד BSC- מדד זה הינו מדד בהיבט הלקוחות, נכונות לשלם.

חישוב ממוצע ההכנסה להזמנה בתקופה X:

$$\frac{\text{סך ההכנסות מכל ההזמנות בתקופה } X_t}{\text{סך ההזמנות בתקופה } X_t}$$

חישוב המדד:

$$\frac{\text{ממוצע ההכנסה להזמנה בתקופה } X_t}{\text{ממוצע ההכנסה להזמנה בתקופה } X_{t-1}}$$

1. מידת שביעות רצון הלקוחות מהמוצרים

מדד זה הינו מדד תועלת עבור החלטה תפעולית. המדד נועד לשקף את אחוז הדירוגים הנמוכים למוצר, כלומר מוצר שקיבל את הציון 1 או 2. מדד זה מייצג תועלת כי הוא בוחן מענה על צורך ושביעות רצון הלקוח. ככל שהמדד ישאף ל0 כך נדע שאנחנו עומדים ביעד המדד.

מודל BSC- מדד זה הינו מדד בהיבט לקוחות, שביעות רצון.

חישוב המדד:

$$\frac{\text{מספר הדירוגים 1 או 2 בתקופה } x \text{ למוצר}}{\text{סך הביקורות בתקופה } x \text{ למוצר}}$$

2. אחוז ניצול קיבולת המשלוחים המקסימלית

מדד זה הינו מדד יעילות עבור החלטה ניהולית טקטית. המדד נועד לייצר תמונה ברורה עבור מידת ניצול הקיבולת המקסימלית במדינה. החברה משלמת מראש פר חודש סכום קבוע עבור מספר מוצרים לשילוח בחודש לחברת השליחויות בכל מדינה. נשאף לכך שהחברה תנצל בצורה המיטבית את קיבולת המוצרים לשליחה. ככל שהשבר יהיה יותר גבוה, המצב יהיה יותר משתלם.

מודל BSC- מדד זה הינו מדד בהיבט התנהלות פנימית, ניצולת משאבים.

חישוב המדד:

$$\frac{\text{מספר המוצרים הנשלחים בחודש במדינה}}{\text{קיבולת מקסימלית למדינה}}$$

3. אחוז הזמנות שלא הגיעו בזמן

מדד זה הינו מדד יעילות עבור החלטה ניהולית טקטית. על מנת למזער את מספר הפיצויים שהחברה נותנת ללקוחותיה נרצה למדוד את אחוז ההזמנות שהגיעו בזמן עבור כל רבעון ונרצה שמדד זה ישאף ל0.

מודל BSC - מדד זה הינו מדד בהיבט התנהלות פנימית, התנהלות של תהליכים פנים ארגוניים.

חישוב המדד:

מספר הזמנות שלא הגיעו בזמן

מספר הזמנות הכולל

4. אחוז הלקוחות החוזרים

מדד זה הינו ממד תועלת עבור החלטה ניהולית אסטרטגית. על מנת למקסם את רווחי החברה ולשמור על יציבות נרצה לקוחות קבועים. את מדד זה נחשב עבור תקופת זמן קבועה ונשאף שערך המדד יהיה שווה ל1.

מודל BSC - מדד זה הינו מדד בהיבט הלקוחות.

חישוב המדד:

מספר הלקוחות שקנו גם בתקופה X_t וגם בתקופה X_{t-1}

מספר הלקוחות שקנו בתקופה X_{t-1}

5. הכנסה כוללת שנתי

מדד זה הינו מדד תועלת עבור החלטה ניהולית אסטרטגית. מקסום רווחים הינה המטרה העיקרית של החברה, ולכן יש חשיבות רבה במדד זה. אנו נרצה לראות צמיחה בהכנסות מההזמנות מידי שנה.

מודל BSC - מדד זה הינו מדד בהיבט הכספים והרגולציה.

חישוב המדד:

סך ההכנסות השנתיות בשנה X_t

סך ההכנסות השנתיות בשנה X_{t-1}

6. ניתוח מקורות הנתונים התפעוליים

מקורות הנתונים:

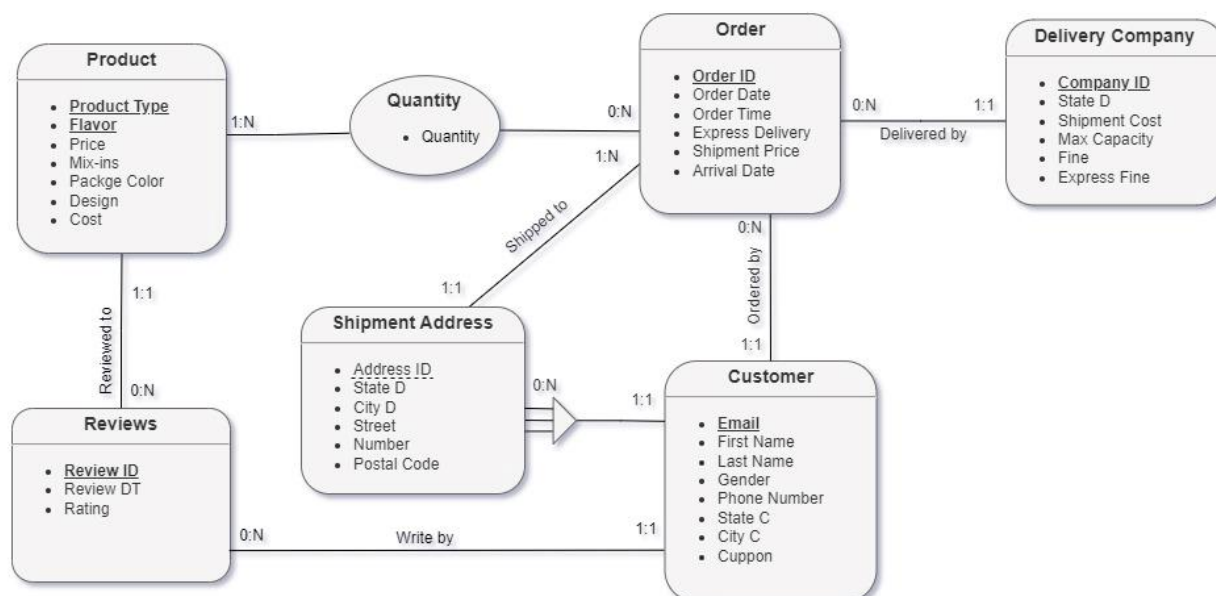
מקור נתונים ראשי

מקור הנתונים הראשי שלנו מכיל בתוכו מידע רב על הלקוחות, מוצרי החברה, ההזמנות, הביקורות ועוד. בסיס הנתונים מיושם בשרת הנתונים הטבלאי SQL server מתוך הנחה שזו הטכנולוגיה הטובה והנוחה ביותר לעבוד ולנתח נתונים עם הכלים שנלמד בקורס זה. חוללנו את הנתונים באמצעות מחולל הנתונים – **Red Gate**, הנתונים בבסיס הנתונים "נאספו" לאורך 3 שנים מ-2019-2022, מכיוון שמדובר בבסיס נתונים שאנחנו יצרנו, בסיס הנתונים אינו מתעדכן בהתאם להזמנות המתבצעות באתר ואינו מתוחזק, רק אנחנו שולטים בו.

מקור נתונים משני

הנתונים של בסיס הנתונים המשני התקבלו מאתר GLOBAL DATA LAB, והם נתוני דמוגרפיה אמיתיים של כל המדינות בארה"ב בשנים 2019-2022, והם מכילים נתונים כמו גודל האוכלוסייה, מדד הפיתוח האנושי, מדדי בריאות, השכלה, הכנסות ותוחלת חיים. הנתונים מתוחזקים בקובץ Excel, ובסיס נתונים זה מיושם גם כן דרך שרת הנתונים הטבלאי SQL server. מטרת מקור הנתונים המשני היא לעזור לארגון לקבל החלטות ולהעשיר את התובנות הניהוליות מעבר לאלו שנוכל להפיק ממקור הנתונים הראשי בלבד.

7. תרשים ERD:



8. מודל טבלאי - בסיס נתונים ראשי

הטבלה	שם השדה	סוג הנתונים	סולם מדידה	תיאור השדה
CUSTOMERS רשימת לקוחות	Email	varchar(60)	שמי	מזהה לקוח
	First Name	varchar(30)	שמי	שם פרטי של הלקוח
	Last Name	varchar(30)	שמי	שם משפחה של הלקוח
	Gender	char	שמי	מין הלקוח
	Phone Number	varchar(30)	שמי	מספר טלפון של הלקוח
	State C	varchar(30)	שמי	המדינה שבה גר הלקוח
	City C	varchar(30)	שמי	העיר שבה גר הלקוח
	Cuppon	float	סדר	אחוז ההנחה שיש ללקוח (יכול להיות 0)
DELIVERY COMPANY חברות המשלוחים שעובדות עם החברה	Company ID	int	סדר	מזהה חברת משלוח
	State D	varchar(30)	שמי	המדינה שבה חברת המשלוח עושה משלוחים
	Shipment cost	money	יחס	עלות חודשית (משולם על ידי החברה)
	Capacity	int	יחס	מספר מוצרים מקסימלי שחברת השליחויות התחייבה לשלוח בחודש
	Fine	money	יחס	עלות קנס יומית לאיחור במשלוח (משולם לחברה)
	Express Fine	money	יחס	עלות קנס יומית לאיחור במשלוח מהיר (משולם לחברה)
PRODUCTS רשימת מוצרים	Product Type	varchar(50)	שמי	מזהה מוצר - סוג המוצר
	Flavor	varchar(30)	שמי	מזהה מוצר - טעם גלידה
	Price	money	יחס	מחיר המוצר
	Mix-ins	varchar(10)	שמי	תוספות אפשריות לגלידה
	Package Color	varchar(10)	שמי	צבע אריזה
	Design	varchar(7)	שמי	עיצוב - טקסט/תמונה
	Cost	money	יחס	עלות ייצור המוצר (משולם על ידי החברה)
SHIPMENT ADDRESSES רשימת כתובות למשלוח	Email (CUSTOMERS)	varchar(60)	שמי	מזהה לקוח
	AddressID	int	סדר	מזהה כתובת
	State D	varchar(30)	שמי	המדינה שאליה הלקוח רוצה שהמשלוח יגיע
	City D	varchar(30)	שמי	העיר שאליה הלקוח רוצה שהמשלוח יגיע
	Street	varchar(30)	שמי	הרחוב שאליו הלקוח רוצה שהמשלוח יגיע
	Number	varchar(5)	סדר	המספר של הרחוב שאליו הלקוח רוצה שהמשלוח יגיע
	Postal Code	varchar(30)	סדר	המיקוד של הכתובת
ORDERS וזמנות שבוצעו על ידי לקוחות	OrderID	int	סדר	מזהה הזמנה
	Order Date	date	רווח	תאריך שבו בוצעה ההזמנה
	Order Time	datetime	רווח	שעה שבה בוצעה ההזמנה
	Express Delivery	bit	שמי	האם הלקוח בחר במשלוח מהיר או רגיל
	Shipment Price	money	יחס	מחיר המשלוח בהתאם למשלוח מהיר או רגיל
	arrival date	date	רווח	תאריך הגעת המשלוח ליעד
	Company ID (DELIVERY COMPANY)	int	סדר	מזהה חברת משלוח
	Email (CUSTOMERS)	varchar(60)	שמי	מזהה לקוח
	AddressID (ADDRESSES)	int	סדר	מזהה כתובת
	ReviewID	int	סדר	מזהה ביקורת
REVIEWS ביקורות שנכתבו על ידי לקוחות על מוצר מסוים	Review Date	datetime	רווח	תאריך כתיבת הביקורת
	Review Time	datetime	רווח	שעה שבה בוצעה כתיבת הביקורת
	Rating	int	רווח	דירוג המוצר בסקלה 1-5 כאשר 5 - מצוין, 1 - גרוע
	Review Description	varchar(400)	שמי	מלל לביקורת
	Email (CUSTOMERS)	varchar(60)	שמי	מזהה לקוח
	Product Type (PRODUCTS)	varchar(50)	שמי	מזהה מוצר - סוג המוצר
MIX-INS סוג התוספות הקיימות	Flavor (PRODUCTS)	varchar(30)	שמי	מזהה מוצר - טעם גלידה
	Product Type (PRODUCTS)	varchar(50)	שמי	מזהה מוצר - סוג המוצר
	Flavor (PRODUCTS)	varchar(30)	שמי	מזהה מוצר - טעם גלידה
	Mix-ins	varchar(30)	שמי	תוספת לגלידה
QUANTITIES כמות המוצרים שנבחרה על ידי הלקוח	OrderID (ORDERS)	int	סדר	מזהה הזמנה
	Product Type (PRODUCTS)	varchar(50)	שמי	מזהה מוצר - סוג המוצר
	Flavor (PRODUCTS)	varchar(30)	שמי	מזהה מוצר - טעם גלידה
	Quantity	int	יחס	כמות יחידות שנבחרה על ידי הלקוח

מודל טבלאי – בסיס נתונים משני:

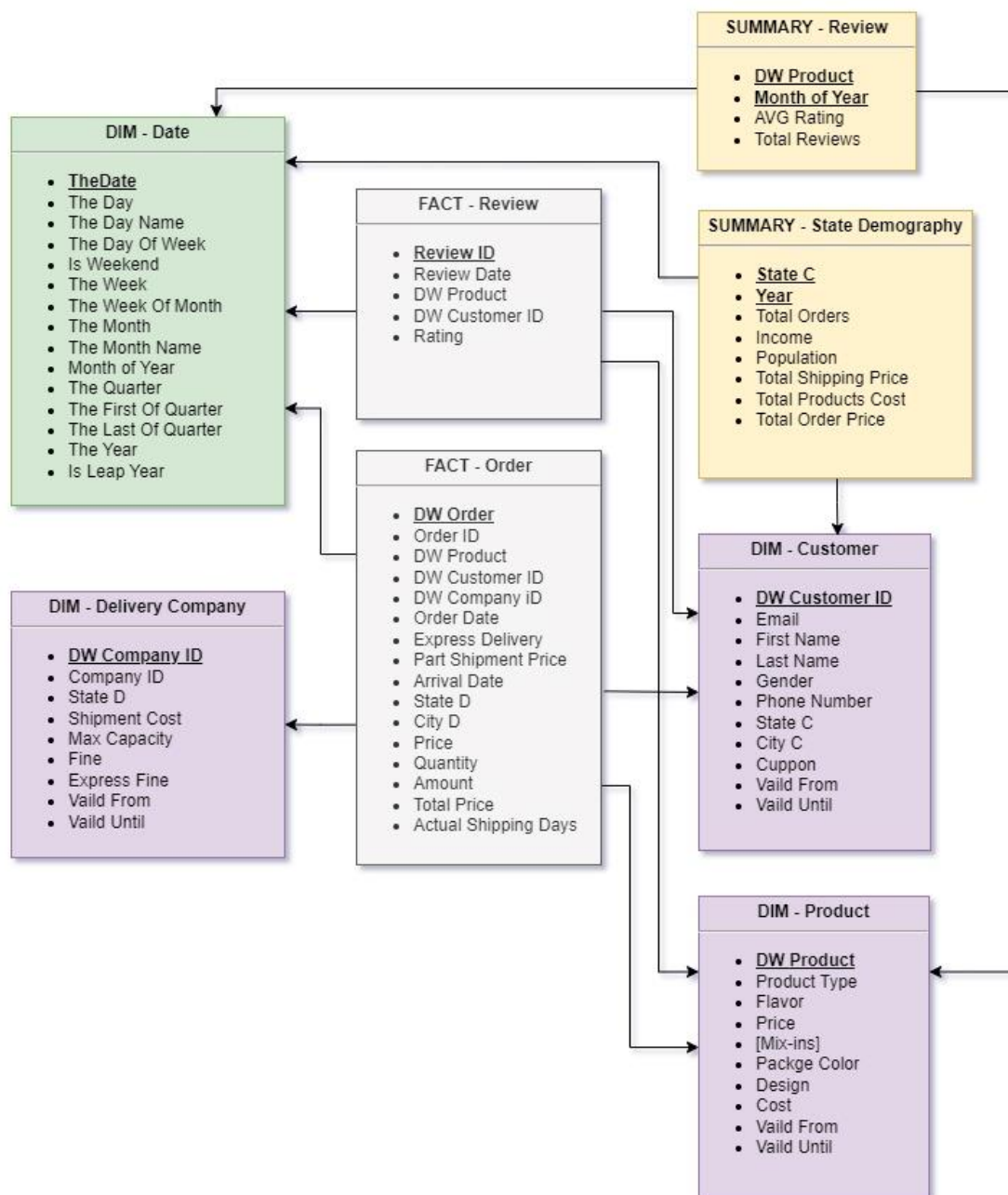
הטבלה	שם השדה	סוג הנתונים	סולם מדידה	תיאור השדה
State Demography מצב דמוגרפי של המדינות לפי שנים	State	varchar(30)	שמי	מדינה
	Year	int	רווח	שנה
	HDI	float	יחס	מדד הפיתוח האנושי
	Health index	float	יחס	מדד הבריאות
	Educational index	float	יחס	מדד ההשכלה
	Income	money	יחס	מדד ההכנסה
	Life expectancy	float	יחס	תוחלת חיים
	Population	float	יחס	גודל האוכלוסייה

שילוב מקורות הנתונים:

בסיס הנתונים המשני (State Demography) מספק לנו פילוח נתונים לפי שנים ולפי מדינות של נתונים דמוגרפים שהחברה שלנו מספקת משלוחים. ניתן לבצע מיזוג בין שדה City C שנמצא בטבלת CUSTOMERS לבין שדה City בטבלת State Demography וכך ניתן לדעת מה הפילוח הדמוגרפי של האזורים שבהם הלקוחות שלנו גרים. ניתן לבצע מיזוג בין שדות Year, City בטבלת State Demography לבין שדות Year, City C בטבלת SUMMARY- State Demography וכך נוכל לדעת מה טווח המחירים שהלקוחות מבזבזים בהתאם לנתונים הדמוגרפיים שלהם, ונראה אם יש קשר בין הנתונים (כלומר נוכל לבדוק אם ככל שהמדינה במצב כלכלי טוב – האם הלקוחות שלנו מאותה מדינה עושים יותר משלוחי גלידה ועלויות ההזמנות שלהם יותר גבוהות).

9. אפיון של מחסן נתונים טבלאי:

תרשים מחסן הנתונים:



תיאור מחסן הנתונים:

הטבלה	שם השדה	סוג הנתונים	תיאור השדה
DIM_CUSTOMERS טבלת מימד משתנה לאט מסוג 2, מכילה את רשימת הלקוחות, תוך שמירת גרסאות	DW_Customer ID	int	מזהה לקוח מפתח עזר
	Email	varchar(60)	מזהה לקוח מקורי
	First Name	varchar(30)	שם פרטי של הלקוח
	Last Name	varchar(30)	שם משפחה של הלקוח
	Gender	char	מין הלקוח
	Phone Number	varchar(30)	מספר טלפון של הלקוח
	State C	varchar(30)	המדינה שבה גר הלקוח שדה זה משקף הירארכיה על פיה ניתן לפלח את הלקוחות עשויה להשתנות לפעמים ולכן יש צורך בשמירת גרסאות
	City C	varchar(30)	העיר שבה גר הלקוח שדה זה משקף הירארכיה על פיה ניתן לפלח את הלקוחות עשויה להשתנות לפעמים ולכן יש צורך בשמירת גרסאות
	Coupon	float	אחוז ההנחה שיש ללקוח (יכול להיות 0) עשוי להשתנות לפעמים ולכן יש צורך בשמירת גרסאות
	Valid From	date	תאריך תחילת תוקף גרסה
	Valid Until	date	תאריך סיום תוקף גרסה
DIM_DELIVERY COMPANY טבלת מימד משתנה לאט מסוג 2, מכילה את רשימת חברות המשלוחים שעובדות עם החברה, תוך שמירת גרסאות	DW_Company ID	int	מזהה חברת משלוח מפתח עזר
	Company ID	int	כל ערך משקף גרסת מוצר בעלת טווח תאריכים מוגדר
	State D	varchar(30)	מזהה חברת משלוח מקורי
	Shipment cost	money	העיר שבה חברת המשלוח עושה משלוחים עלות חודשית (משולם על ידי החברה) עשויה להשתנות לפעמים ולכן יש צורך בשמירת גרסאות
	Max Capacity	int	מספר מוצרים מקסימלי שחברת השליחויות התחייבה לשלוח בחודש עשוי להשתנות לפעמים ולכן יש צורך בשמירת גרסאות
	Fine	money	עלות קנס יומית לאיחור במשלוח (משולם לחברה) עשויה להשתנות לפעמים ולכן יש צורך בשמירת גרסאות
	Express Fine	money	עלות קנס יומית לאיחור במשלוח מהיר (משולם לחברה) עשויה להשתנות לפעמים ולכן יש צורך בשמירת גרסאות
	Valid From	date	תאריך תחילת תוקף גרסה
	Valid Until	date	תאריך סיום תוקף גרסה
DIM_PRODUCTS טבלת מימד משתנה לאט מסוג 2, מכילה את רשימת המוצרים, תוך שמירת גרסאות	DW_Product	int	מזהה מוצר מפתח עזר
	Product Type	varchar(50)	משקף צירוף ייחודי של 2 המזהים של המוצר כל ערך משקף גרסת מוצר בעלת טווח תאריכים מוגדר
	Flavor	varchar(30)	מזהה מוצר מקורי - סוג המוצר שדה זה משקף הירארכיה על פיו ניתן לפלח את המוצרים
	Price	money	מזהה מוצר מקורי - טעם גלידה שדה זה משקף הירארכיה על פיו ניתן לפלח את המוצרים
	Mix-ins	varchar(30)	מחיר המוצר עשוי להשתנות לפעמים ולכן יש צורך בשמירת גרסאות
	Package Color	varchar(10)	תוספת אפשרית לגלידה
	Design	varchar(7)	צבע אריזה
	Cost	money	עיצוב - טקסט/תמונה עלות ייצור המוצר (משולם על ידי החברה) עשויה להשתנות לפעמים ולכן יש צורך בשמירת גרסאות
	Valid From	date	תאריך תחילת תוקף גרסה
	Valid Until	date	תאריך סיום תוקף גרסה
	Date	date	מזהה התאריך
	The Day	int	שנה
	The Day name	varchar(20)	שם היום
	The Day of week	int	מספר היום בשבוע
	is weekend	int	האם סוף שבוע
	the week	int	מספר השבוע בשנה
	the week of month	int	מספר השבוע בחודש
DIM_DATE טבלת מימד זמן, המכילה מאפייני תאריך משתנה לאט מסוג 0	the month of year	varchar(20)	חודש ושנה
	the month	int	מספר החודש
	the month name	varchar(20)	שם החודש
	the quarter	int	מספר הרבעון
	the first of quarter	date	תאריך תחילת הרבעון
	the last of quarter	date	תאריך סוף הרבעון
	the year	bit	השנה
	is leap year	bit	האם השנה מעוברת

מזהה הזמנה ייחודי למוצר מההזמנה המקורית	int	DW Order	FACT_ORDERS טבלת עובדה פרטנית, המפרטת את פרטי ההזמנות לרמת גרעיניות של הזמנה בודדת
מזהה הזמנה מקורי	int	OrderID	
מזהה מוצר (מפתח עזר)	int	DW PRODUCT (DIM_PRODUCTS)	
מזהה לקוח (מפתח עזר)	varchar(60)	DW Customer ID (DIM_CUSTOMERS)	
מזהה חברת משלוח (מפתח עזר)	int	DW Company ID (DIM_DELIVERY COMPANY)	
תאריך שבו בוצעה ההזמנה שדה גולמי מטבלת ORDERS	datetime	Order Date (DIM_DATE)	
האם הלקוח בחר במשלוח מהיר או רגיל (שדה גולמי מטבלת ORDERS)	bit	Express Delivery	
מחיר המשלוח בהתאם למשלוח מהיר או רגיל עם העמסה שדה מחושב - ביצוע אגרזציה לפי הזמנה ואז חישוב ממוצע של shipment price מ-ORDERS, ולאחר מכן לחלק את הממוצע במספר השורות של הזמנה אחת	money	Part Shipment Price	
תאריך הגעת המשלוח ליעד שדה גולמי מטבלת ORDERS	date	Arrival date	
המדינה שאליה הלקוח רוצה שהמשלוח יגיע שדה גולמי נלקח מטבלת SHIPMENT ADDRESSES	varchar(30)	State D	
העיר שאליה הלקוח רוצה שהמשלוח יגיע שדה גולמי נלקח מטבלת SHIPMENT ADDRESSES	varchar(30)	City D	
מחיר המוצר המסוים שדה גולמי נלקח מטבלת PRODUCTS	money	Price	
כמות יחידות ממוצר מסוים שדה גולמי נלקח מטבלת QUANTITIES	int	Quantity	
מחיר כולל של כמות היחידות מאותו מוצר שדה מחושב: Price*Quantity	money	Amount	
שדה מחושב: Part Shipment Price+Amount*(1-Cuppon)	money	Total Price	
שדה מחושב: day(arrival date-order date)	int	Actual Shipping Days	
מזהה ביקורת	int	ReviewID	FACT_REVIEWS טבלת עובדה פרטנית, הכוללת את נתוני הביקורות על מוצר מסוים
תאריך שבו בוצעה הביקורת שדה גולמי נלקח מטבלת REVIEWS	datetime	Review Date (DIM_DATE)	
מזהה מוצר (מפתח עזר)	int	DW PRODUCT (DIM_PRODUCTS)	
מזהה לקוח (מפתח עזר)	int	DW Customer ID (DIM_CUSTOMERS)	
דירוג המוצר בסקלה 1-5 כאשר 5 - מצוין, 1 - גרוע שדה גולמי נלקח מטבלת REVIEWS	int	Rating	SUMMARY - State Demography טבלת עובדה סיכומית, המציגה את נתוני ההזמנות לפי מדינה ושנה שממנה בוצעו ההזמנות
מזהה הטבלה	varchar(30)	State C (DIM_CUSTOMERS)	
מזהה הטבלה	int	Year (DIM_DATE)	
סך מספר ההזמנות שדה מחושב - סכימת מספר ההזמנות לפי מדינה ושנה	int	Total Orders	
ההכנסה הממוצעת במדינה לפי שנה פר נפש, נלקח מטבלת בסיס הנתונים המשני State Demograph	money	Income	
כמות האוכלוסיה במדינה לפי שנה, נלקח מטבלת בסיס הנתונים המשני State Demography	int	Population	
סך עלויות המשלוחים שדה מחושב - סכימת עלויות המשלוחים בהזמנות	money	Total Shipping price	
סך עלויות המוצרים שדה מחושב - סכימת עלויות המוצרים בהזמנות	money	Total Products Cost	
סך ההזמנות שדה מחושב - סכימת עלות הזמנות	money	Total Order Price	SUMMARY - Review טבלת עובדה סיכומית, המציגה את נתוני הביקורות לפי מוצר וחודש
מזהה הטבלה (מפתח עזר)	int	DW PRODUCT (DIM_PRODUCTS)	
מזהה הטבלה	varchar(20)	month of year (DIM_DATE)	
ממוצע דירוג למוצר שדה מחושב - סכימת הדירוגים וחילוק בכמות הדירוגים	float	AVG Rating	
סך ביקורות למוצר שדה מחושב - סכימת מספר הביקורות	int	Total Reviews	

חלק ב' - אפיון תהליכי ETL

1. השלבים בתהליך ה ETL



• Data Mirroring

שלב "שיקוף המידע". שלב העתקת/שכפול בסיס הנתונים התפעולי. לאחר שמירת הנתונים במבנה הטבלאי המקורי, נרוקן את כלל הטבלאות כדי להימנע ממצב של כפילות נתונים (פקודת Truncate table). נעתיק את בסיס הנתונים המקורי לטבלאות MRR. עבור טבלאות העובדה, נטען את הנתונים טעינה אינקרמנטלית.

• Dimensions Staging

שלב ביצוע הטרנספורמציות הנדרשות לטבלאות המימד על מנת להעביר את הנתונים למבנה כוכב. גם בשלב זה נרוקן את הטבלאות ולאחר מכן נעביר את הטבלאות לפי סדר היררכי, על פי שטית Top-Down.

• Dimensions Warehousing

שלב העברת טבלאות המימד מה Staging Area אל ה DW. כלומר נעדכן את הנתונים הקיימים ונוסיף נתונים חדשים במידת הצורך (פקודת UPSERT).

• Fact Staging

שלב ביצוע הטרנספורמציות לטבלאות העובדה, על מנת להתאים את הנתונים למבנה של סכמת כוכב. נרוקן את הטבלאות בטרם העברתן, ונטען אותן בצורה אינקרמנטלית. על מנת להביא את הטבלאות לרמת הגרעיניות המתאימה נבצע פעולות JOIN בין הטבלאות הרלוונטיות וכן פעולות אגריגציה והעמסה לערכי השדות מטבלאות העובדה.

• Referential Integrity

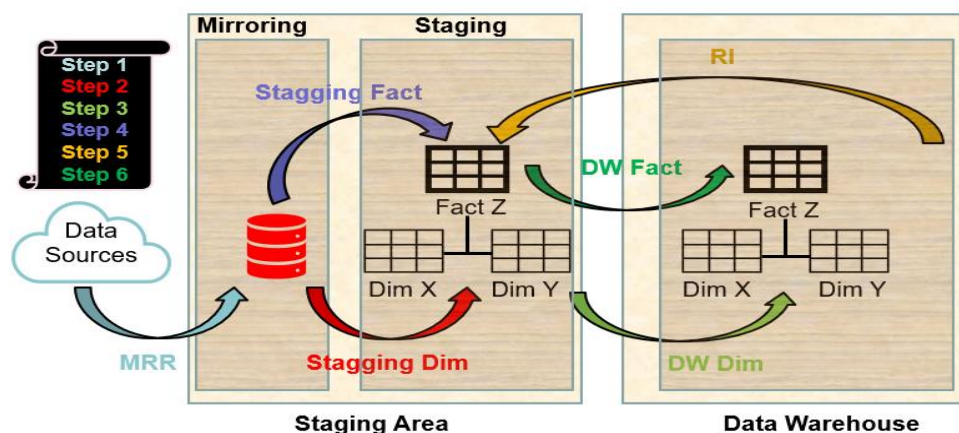
שלב Validation "אימות יושרה" של הנתונים. בשלב זה נוודא תאימות מידע בין נתוני טבלאות המימד לבין נתוני טבלאות העובדה. כמו כן, נוודא שהמפתחות הזרים קיימים ותקינים. במידה ונבחין באי התאמה, נפעל לפי אחת משתי הדרכים הבאות:

- הוספת טבלת דחויים, שתכלול את הטרנזקציות שנדחו.
- הוספת הרשומה החסרה לטבלת המימד עם השדה Unknown ומתן ערכי null לחלק מהשדות.

• Fact Warehousing

שלב עדכון הנתונים. נעביר את טבלאות העובדה מה Staging Area שיצרנו בשלב Fact Staging אל ה DW. נוסיף את הנתונים החדשים באמצעות פקודת INSERT בטעינה אינקרמנטלית.

תרשים Master:



תלויות בין שלבי התהליך:

על מנת שתהליך ETL יתבצע בהצלחה, יש לבצע את שלבי התהליך בסדר הנ"ל מכיוון שישנה תלות בין שלבי התהליך השונים. על מנת לא להעמיס על השרתים, נעתיק את הנתונים תחילה. לאחר מכן נעדכן את טבלאות המימד לפני טבלאות העובדה, ובבצע את השלבים Staging, Mirroring. טבלאות העובדה תלויות בטבלאות המימד - הרשומות בטבלאות העובדה מקושרות במפתחות זרים לטבלאות המימד. לפני טעינת הטבלאות ל-DW נבצע אימות נתונים, כלומר תקינות המפתחות הזרים.

שלבי ETL שצפויים לקחת זמן רב במיוחד במערכת האמת:

- **Fact Staging** - בשלב זה ישנן פעולות אגרגציה ופעולות JOIN רבות, על כמות גדולה של נתונים, שצפויות להימשך זמן רב. שלב זה נוגע לטבלאות העובדה, אשר חוזרות מספר רב של פעמים, לכן נוצרות רשומות חדשות. על כן, סביר ששלב זה ייקח זמן רב.
- **Fact Warehousing** - בדומה לשלב Staging, שלב זה גם נוגע לטבלאות העובדה. כלומר טעינת כמות גדולה של נתונים מתחדשים שצפויים להשתנות ולהתעדכן. לכן פעולה זו צפויה לקחת זמן רב במיוחד.
- **Referential Integrity** - שלב זה גם הוא צפוי לקחת זמן רב, משום שבמסגרתו שיש לבצע בדיקת ולידציה עבור כל רשומה בטבלת העובדה אל מול טבלאות המימד, כלומר עוברים על כמות גדולה מאוד של רשומות.

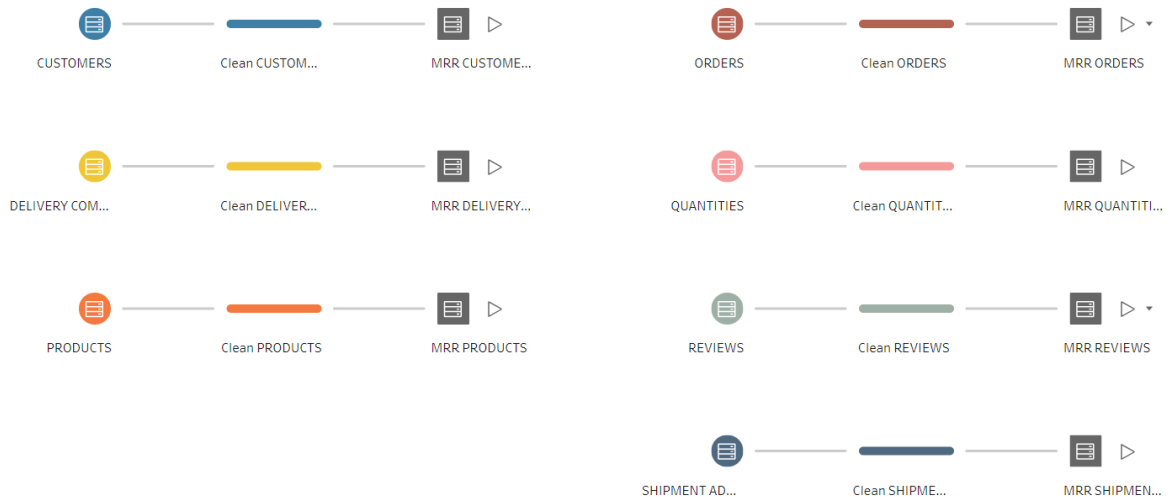


השלבים בתהליך ה-ELT אשר עלולים להיכשל במערכת האמת ודרכי התמודדות:

- **כשל בתהליכי עדכון הנתונים** – עלול לבוע מעומס נתונים ופעולות על המערכת. התמודדות- באמצעות ה-MRR לפיזור העומס, וגם ניתן לבצע שלבים עמוסים בפעולות ונתונים בזמנים בהם המערכת פחות עמוסה, למשל בשעות הלילה.
- **כשל בשלב ה-RI** – חוסר תאימות במפתחות הזרים בין טבלאות העובדה לטבלאות המימד. אשר נגרם עקב אכיפה לא קפדנית דיה של אילוצי המפתחות הזרים במקור הנתונים. התמודדות
 - הוספת טבלת דחויים, ובמקום להעביר לטבלת העובדה, נעביר אליה את הרשומות שהתגלה בהן כשל, לצורך דיווח שגיאות ותחקור נוסף.
 - הוספת הרשומה לטבלת המימד עם השדה Unknown, הרשומה שבה זוהה הערך תקושר לרשומה המלאכותית ותוזן לטבלת העובדה.
- **כשל בשלב Dimensions Staging** - פגיעה בתלות בין הנתונים תקרה במידה והעתקת הנתונים לא תתבצע בהתאם להיררכיה. התמודדות- הקפדה על ביצוע השלבים לפי הסדר הדרוש.
- **כשל בטעינה אינקרמנטלית** – הזנת הנתונים החלקית השוטפת עלולה ליצור טעויות במעקב אחר הנתונים. התמודדות- ניהול חותמת זמן ומפתח ראשי על מנת לעקוב אחר העדכונים.

2. מימוש תהליכי ה-ETL – יישומי הבסיס

שלב 1 - MRR:



Truncate Table:

MRR CUSTOMERS 0 fields

Save output to
Database table

Settings Custom SQL

Before After

Optional: Execute this SQL before writing flow data to database table.

```
TRUNCATE TABLE
eCreameryMRR.dbo.MRR_CUSTOMERS
```

טעינה אינקרמנטלית:

Input

Settings Tables Data Sample

eCreameryORG

Table
ORDERS

Incremental Refresh
Specify the input and output fields to use to identify new rows.

☒ Enable incremental refresh

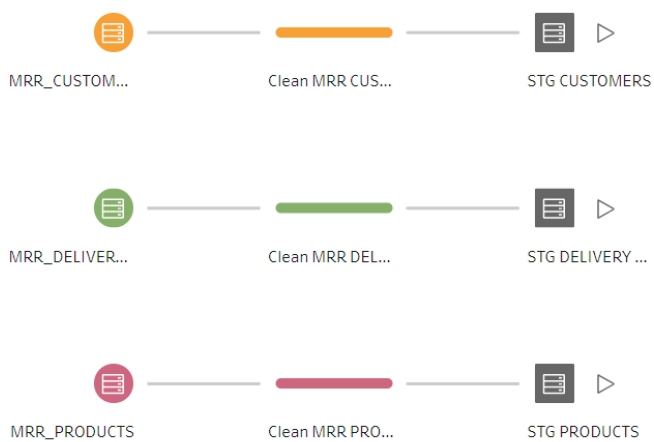
Input field
OrderID

Output
MRR ORDERS

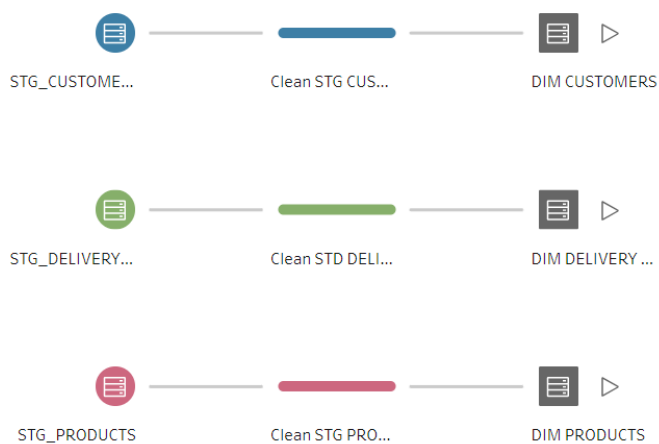
Select the output and field with the last processed value for field "OrderID".

Output field

שלב 2 - Dimension Staging:

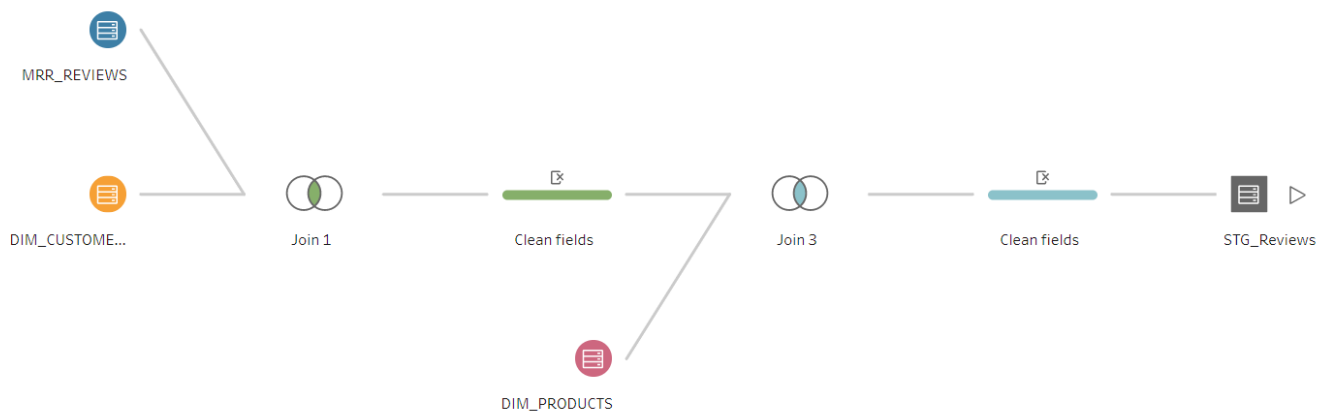


שלב 3 - Dimensions Warehousing:

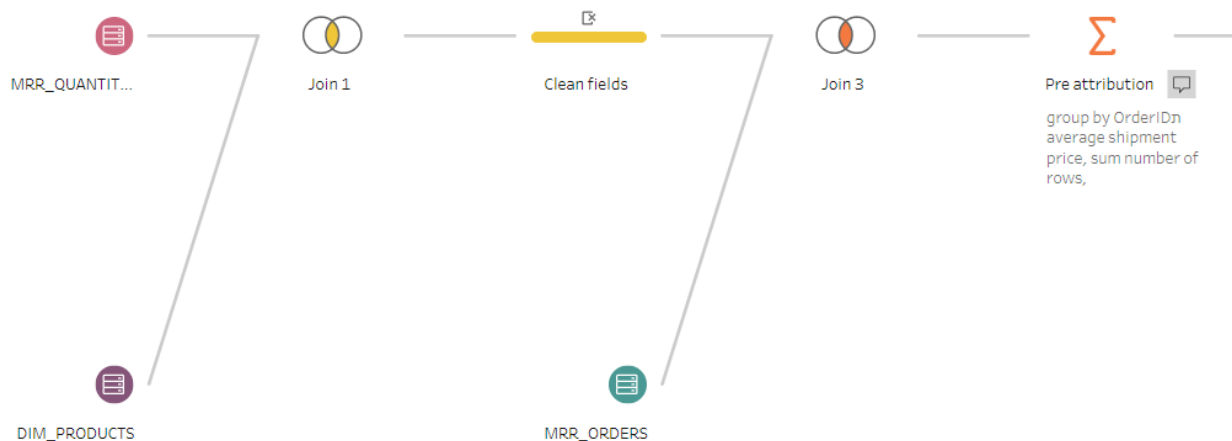


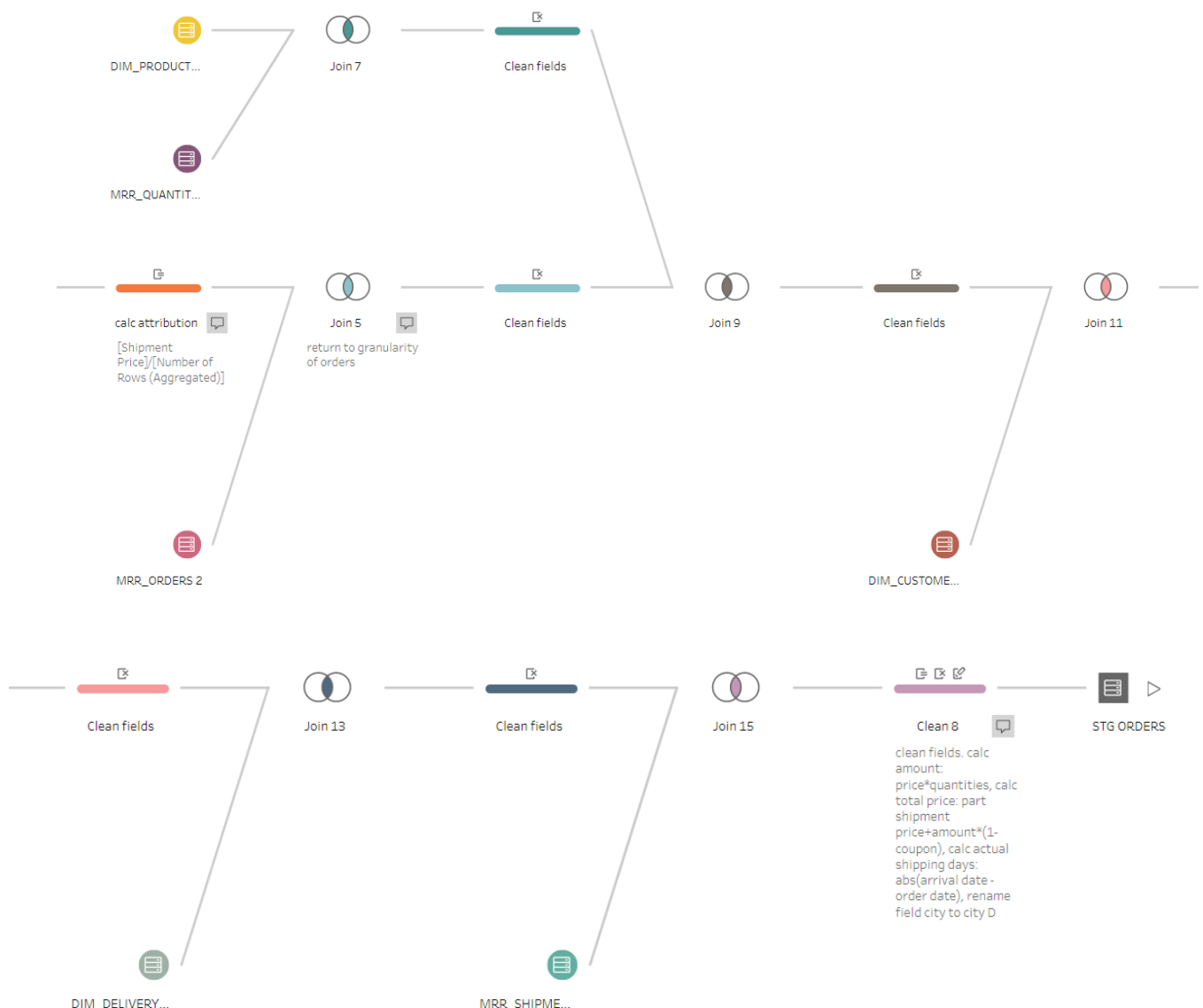
שלב 4 - Fact Staging:

STG_Reviews



STG_Orders





Grouped Fields				Aggregated Fields			
Type	Group Type	Field Name	Changes	Type	Aggregation	Field Name	Changes
#	GROUP	OrderID		#	AVG	Shipment Price	
				#	SUM	Number of Rows (Aggregated)	

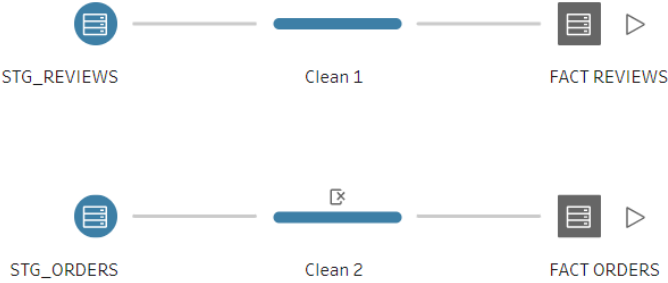
שדות מחושבים:

Calculated Field	Amount	ROUND([Price]*[Quantity],2)
Calculated Field	Total Price	ROUND([Part Shipment Price]+[Amount]*(1-[Coupon]),2)
Calculated Field	Actual Shipping Days	ABS(DATEDIFF('day',[Arrival Date],[OrderDate],'monday'))

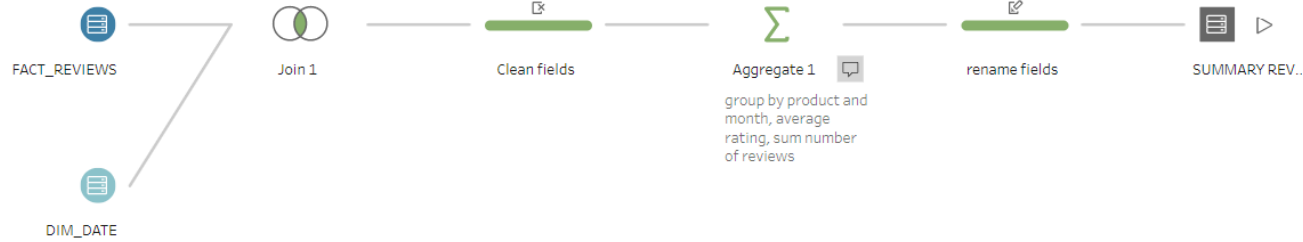
שלב 5 - Referential Integrity

מפורט בסעיף 4 – מימוש תהליכי ה ETL – יישומים מתקדמים.

שלב 6 - Fact Warehousing



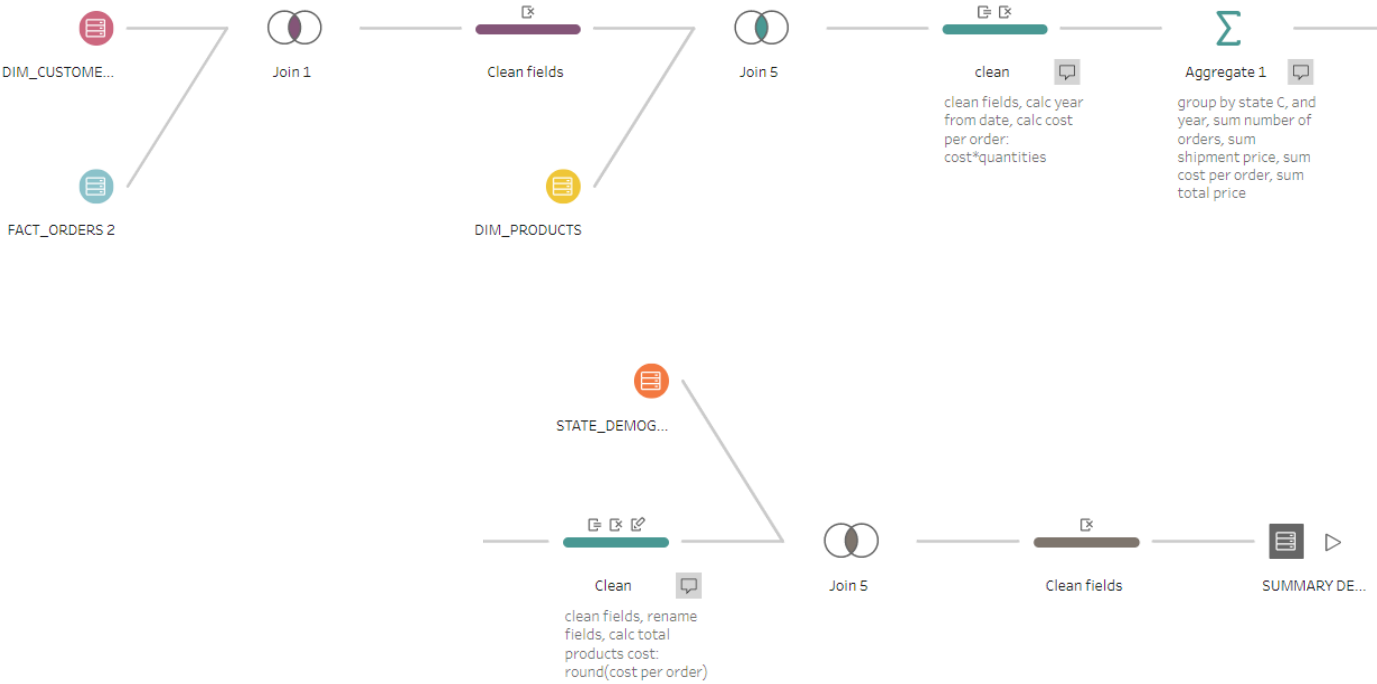
Summary Reviews



אגרגציה:

Grouped Fields			Aggregated Fields		
Type	Group Type	Field Name	Type	Aggregation	Field Name
T F	GROUP	TheMonthOfYear	#	SUM	Number of Rows (Aggregated)
T F	GROUP	DWPProduct	#	AVG	Rating

:Summary State Demography



אגרגציה:

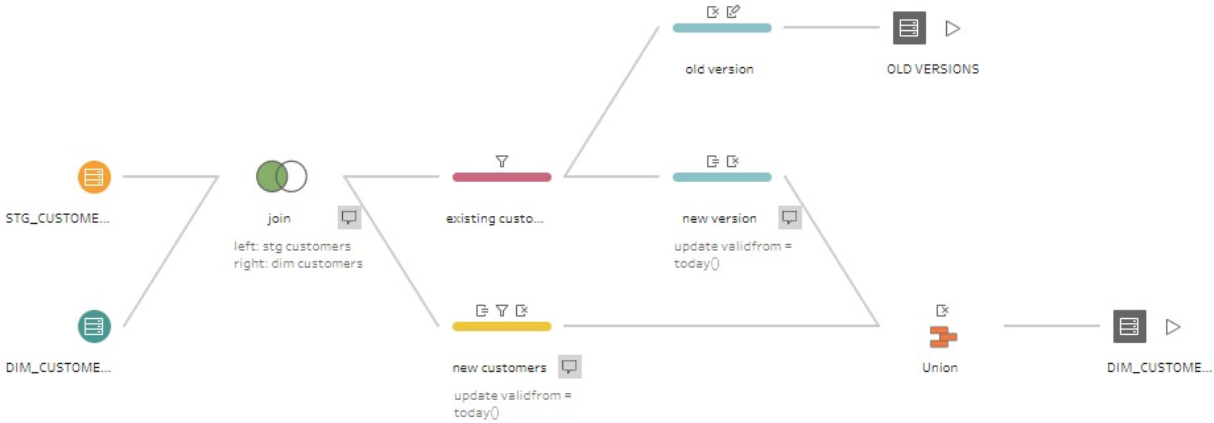
Grouped Fields			Aggregated Fields		
Type	Group Type	Field Name	ype	Aggregation	Field Name
#	GROUP	Year	#	SUM	cost per order
∑	GROUP	State C	#	SUM	Number of Rows (Aggregated)
			#	SUM	Total Price
			#	SUM	Part Shipment Price

שדות מחושבים:

Calculated Field
Year
YEAR([OrderDate])
Calculated Field
cost per order
[Cost]*[Quantity]

3. מימוש תהליכי ה-ETL – יישומים מתקדמים

מימד משתנה לאט מסוג 2



ביצוע פרוצדורה שמורה

Settings

Custom SQL

Before

After

Optional: Execute this SQL after writing flow data to database table.

```
EXEC updateOldVersions
```

New Customers

Filter

DWCustomerID

ISNULL([DWCustomerID])

Existing Customers

Changes (2)

<

Filter

DWCustomerID

ValidUntil

NOT (ISNULL([DWCustomerID]))AND ISNULL([ValidUntil])

Filter

City C

City C-1

Coupon

...

[City C]!=[City C-1]OR[Coupon]!=[Coupon-1]OR[State C]!=[State C-1]

הוספת פילטר בשלב Staging של טבלת הזמנות:



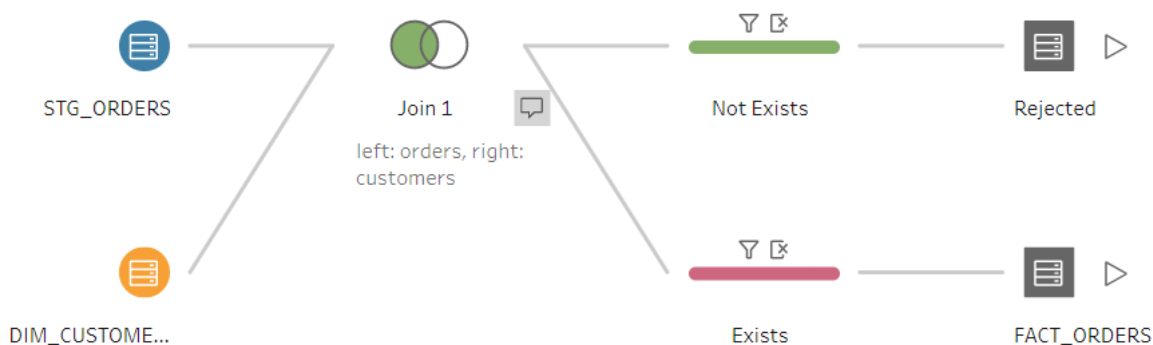
Edit Filter

Calculation Must return a boolean value

```
[OrderDate]>=[ValidFrom] AND ([OrderDate]<[ValidUntil] OR ISNULL([ValidUntil]))
```

Referential Integrity

גישה ראשונה:



-

איתור לקוחות קיימים-

Edit Filter

Calculation

```
NOT (ISNULL ([Email]))
```

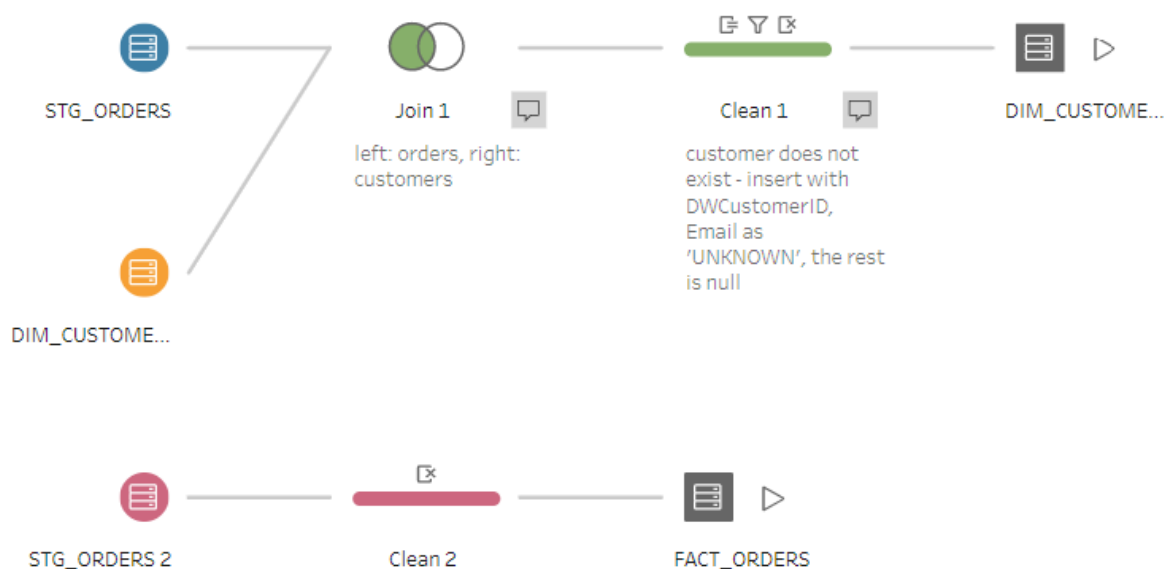
איתור לקוחות שאינם קיימים-

Edit Filter

Calculation

```
ISNULL ([Email])
```

ג'שה שנייה:



סימון אותם לקוחות ב UNKNOWN -

איתור הלקוחות הלא קיימים שביצעו הזמנה -

Edit Field

Field Name

Email

'UNKNOWN'

Edit Filter

Calculation

ISNULL([Email])

חלק ג' – ניתוח נתונים וויזואליזציה

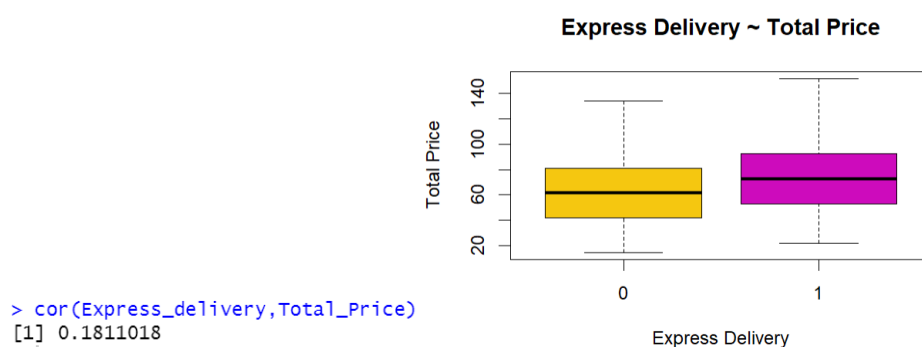
ניתוח נתונים באמצעות R

שאלה עסקית – האם קיים קשר בין בחירה במשלוח מהיר לבין סכום ההזמנה?
מימשנו בRStudio מבחן חי בריבוע לבדיקת ההשערה כאשר $\alpha = 0.05$

ניסוח מבחן ההשערה:

H_0 : אין קשר בין בחירה במשלוח מהיר לבין סכום ההזמנה.

H_1 : קיים קשר בין בחירה במשלוח מהיר לבין סכום ההזמנה.



מתרשים ה-boxplot, ניתן לראות כי כמעט ואין שוני בין התפלגות סכום ההזמנה כאשר הלקוח בחר במשלוח מהיר וכאשר הלקוח בחר במשלוח רגיל, כמו כן ניתן לראות כי הקורלציה בין שני המשתנים (סכום ההזמנה ומשלוח מהיר) נמוכה מאוד אשר מרמזת על כך שלא קיים קשר.

הפלט הסטטיסטי:

```
> chisq.test(dataset$Express.Delivery, Total_Price, correct=FALSE)
```

Pearson's Chi-squared test

data: dataset\$Express.Delivery and Total_Price
X-squared = 2999.9, df = 2896, p-value = 0.08731

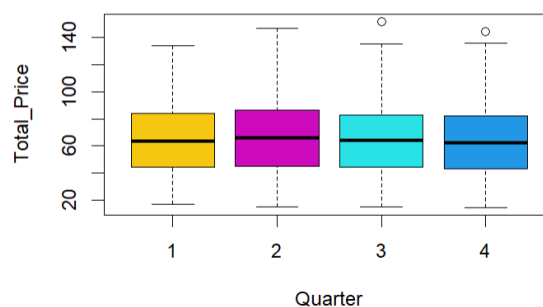
מסקנה: קבלנו כי $P\text{ value} > \alpha$, ולכן לא נדחה את השערת האפס בר"מ 5%, ונאמר כי אין קשר בין בחירה במשלוח מהיר לבין סכום ההזמנה, בנוסף נבחין כי תוצאות מבחן זה מחזקות את תוצאות תרשים ה-boxplot ואת הקורלציה בין המשתנים שקבלנו.

השאלה העסקית – האם יש הבדל בהתפלגות ההכנסות בין הרבעונים לאורך השנים 2019-2022?

ניסוח מבחן השערה:

H_0 : אין הבדל בהתפלגויות.

H_1 : יש הבדל בהתפלגויות.



מתרשים boxplot ניתן לראות כי אין כמעט שוני בהתפלגות ההכנסות בין הרבעונים השונים, לצורך ביצוע מבחן השערה זה נבצע מבחן ANOVA בRStudio, כאשר $\alpha = 0.05$

הפלט הסטטיסטי:

```
> anova(lm(Total_Price ~ dataset$Quarter, dataset))
Analysis of Variance Table

Response: Total_Price
          Df Sum Sq Mean Sq F value Pr(>F)
dataset$Quarter  3    5048   1682.61    2.441 0.06247 .
Residuals    3496 2409838    689.31
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

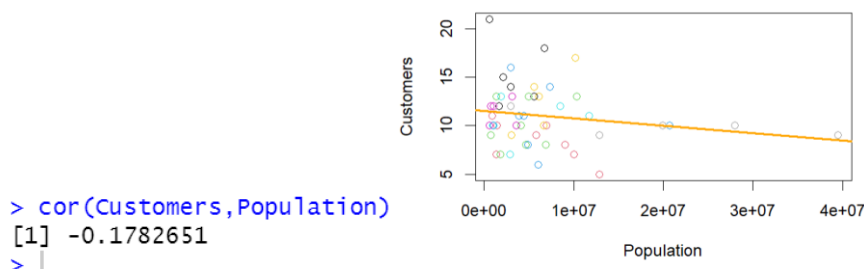
מסקנה: קבלנו כי $P\text{ value} > \alpha$, ולכן לא נדחה את השערת האפס בר"מ 5%, ונאמר כי אין הבדל בהתפלגות ההכנסות בין הרבעונים לאורך השנים (2019-2022), בנוסף נראה כי מבחן זה מחזק את תוצאות תרשים ה-boxplot שקבלנו.

השאלה העסקית - האם יש קשר בין גודל האוכלוסייה לבין מספר הלקוחות מאותה מדינה?

ניסוח מבחן ההשערה:

H_0 : אין קשר בין גודל האוכלוסייה לבין מספר הלקוחות.

H_1 : קיים קשר בין גודל האוכלוסייה לבין מספר הלקוחות.



מתרשים ה-plot ניתן לראות כי לא קיים קשר מהותי בין גודל האוכלוסייה לבין מספר הלקוחות בכל מדינה, כמו כן ניתן לראות כי הקורלציה בין שני משתנים אלו היא נמוכה יחסית וגם היא מרמזת על אי קיום קשר בין המשתנים. מימשנו בRStudio מבחן t סטטיסטי לבדיקת ההשערה כאשר $\alpha = 0.05$

הפלט הסטטיסטי:

```
> t.test(x=Population,
+        y=Customers, alternative = "two.side", conf.level = 0.95)

welch Two Sample t-test

data: Population and Customers
t = 6.2722, df = 49, p-value = 8.895e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 4407930 8564088
sample estimates:
mean of x mean of y
 6486020      11
```

מסקנה: קיבלנו כי $P\text{ value} < \alpha$, ולכן נדחה את השערת האפס בר"מ 5%, ונאמר כי קיים קשר בין גודל האוכלוסייה ובין מספר הלקוחות מאותה המדינה. קבלנו סתירה להנחה שהנחנו מתרשים ה-plot ומערך הקורלציה.

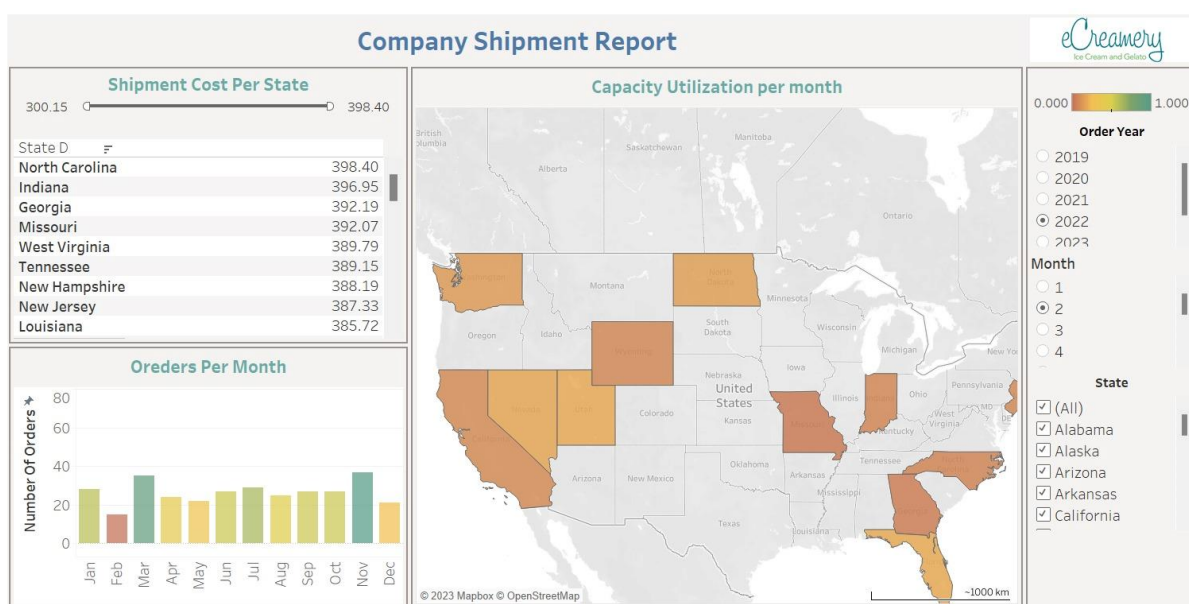
ויזואליזציה (Tableau Dashboards)

Report 1

שאלה עסקית - האם כדאי לעבוד עם חברת המשלוחים שקיימת בכל מדינה ?
ניהול טקטי - עונה על KPI של חברת המשלוחים של ניצול הקיבולת של כל חברת משלוחים בכל מדינה וכמה עולה לנו חברת המשלוחים הזו.

פילטרים – ניתן לפלטר לפי שנה, חודש, מדינה ותעדיף חודשי עבור חברת משלוחים (Shipment Cost Per State)
(הנחה – לכל מדינה יש חברת משלוחים אחת)

בדוח זה ניתן לראות את התעריף החודשי של חברת המשלוחים בכל מדינה, כמות המכירות מהזמנות לפי חודש ושנה, בנוסף ניתן לראות את מפת ארה"ב והמדינות שבה לפני צבע שמתאר את מדד היעילות, הניצולת של כל חברת המשלוחים [1,0].



Report 2

שאלה עסקית - האם לפתוח מועדון לקוחות VIP ?

ניהולי אסטרטגי - נבדוק את הרווחים מהלקוחות לקוחות חוזרים לפי חודשים נציג את הלקוחות הכי רווחיים שלנו.

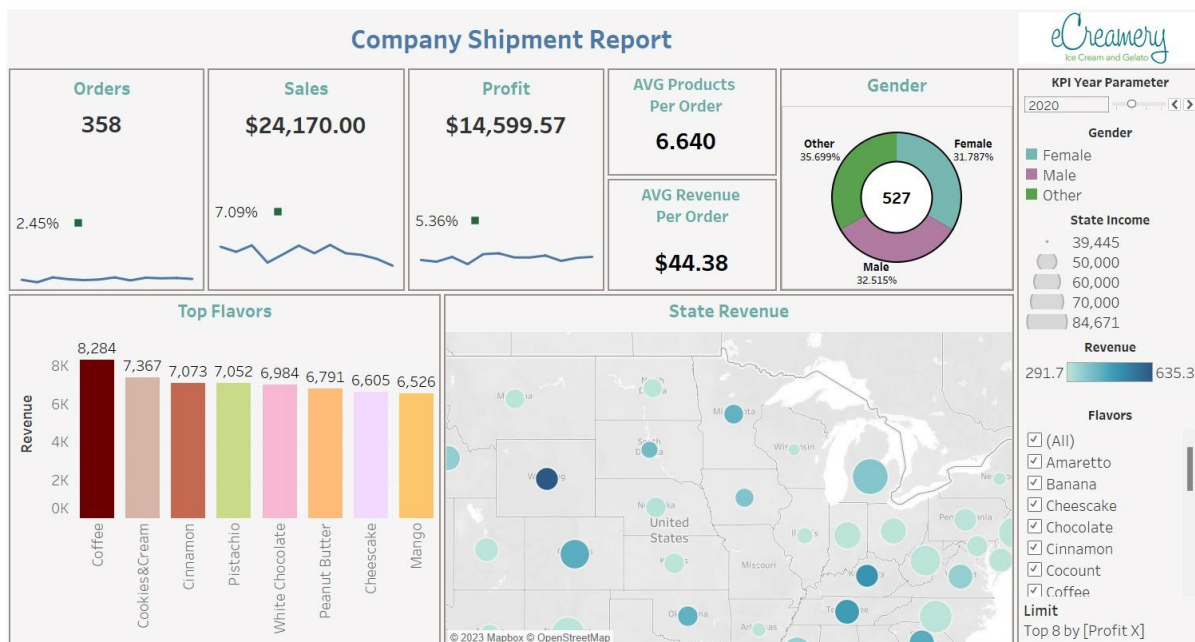
פילטרים - ניתן לפלטר לפי שנה ומדינה (הרזולוציה בדוח היא שנתי)

בדוח זה ניתן לראות גרף שנתי של כמות הלקוחות בחודש אל מול כמות הלקוחות החוזרים (לקוחות שבצעו רכישה נוספת בחודש העוקב) ויחס הלקוחות החוזרים. Top Customers הינה רשימה של הלקוחות הכי רווחיים שלנו. Revenue Per Year – רווח שנתי של מכירות והזמנות כשאר הצבע מעיד על גודל הרווח והמספר על הבועה מעיד על כמות הלקוחות באותה שנה.



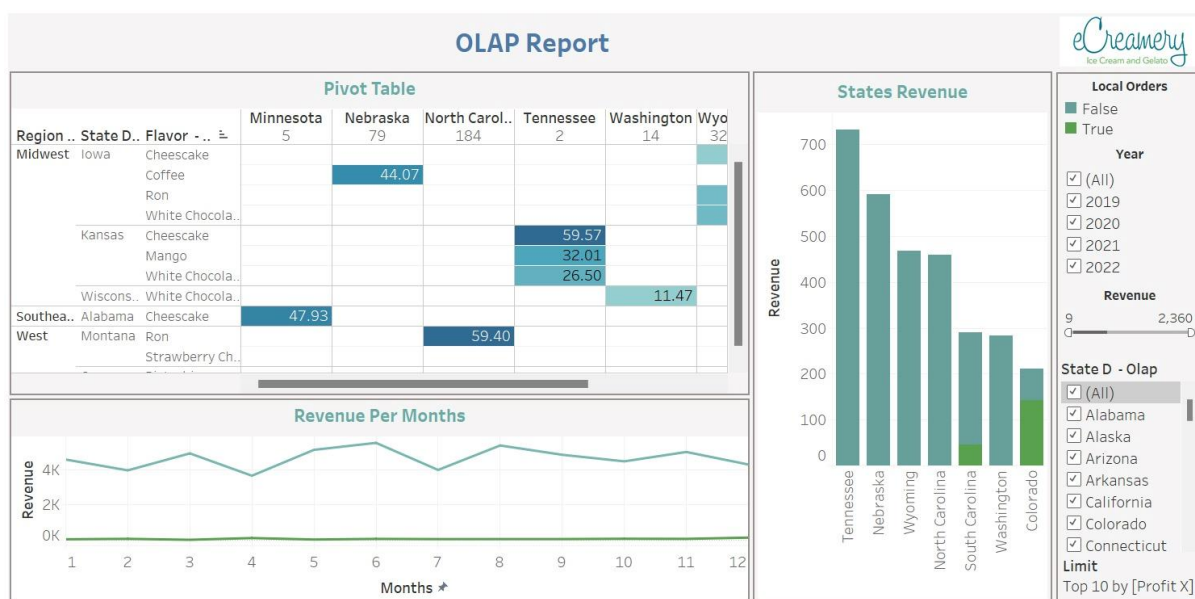
Dashboard

מציג בצורה סיכומית את מספר ההזמנות, המכירות, הרווחים של החברה, מציג את הטעמים הנבחרים של החברה, מיועד לדרג הבכיר של החברה יש חלוקה של מספר הזמנות לפי מגדר.
מפת מדינות שמציגה את ההכנסה השנתית הממוצעת של התושבים במדינה ואת ההכנסות מאותן מדינות, ניתן לבחון האם קיים קשר ביניהם.

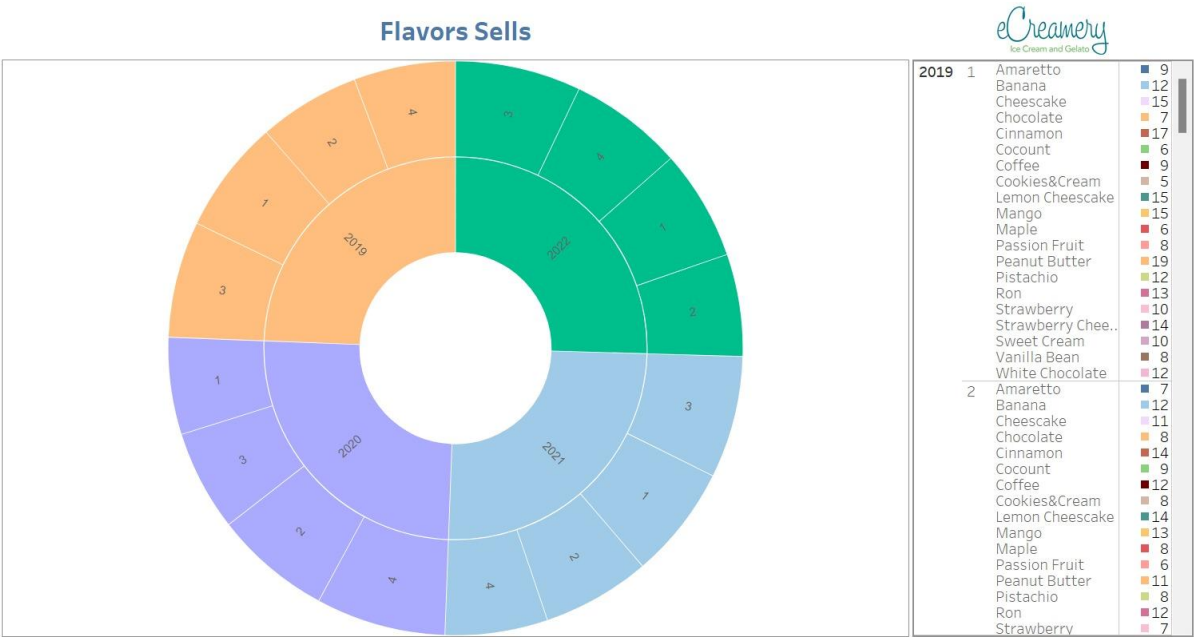


OLAP

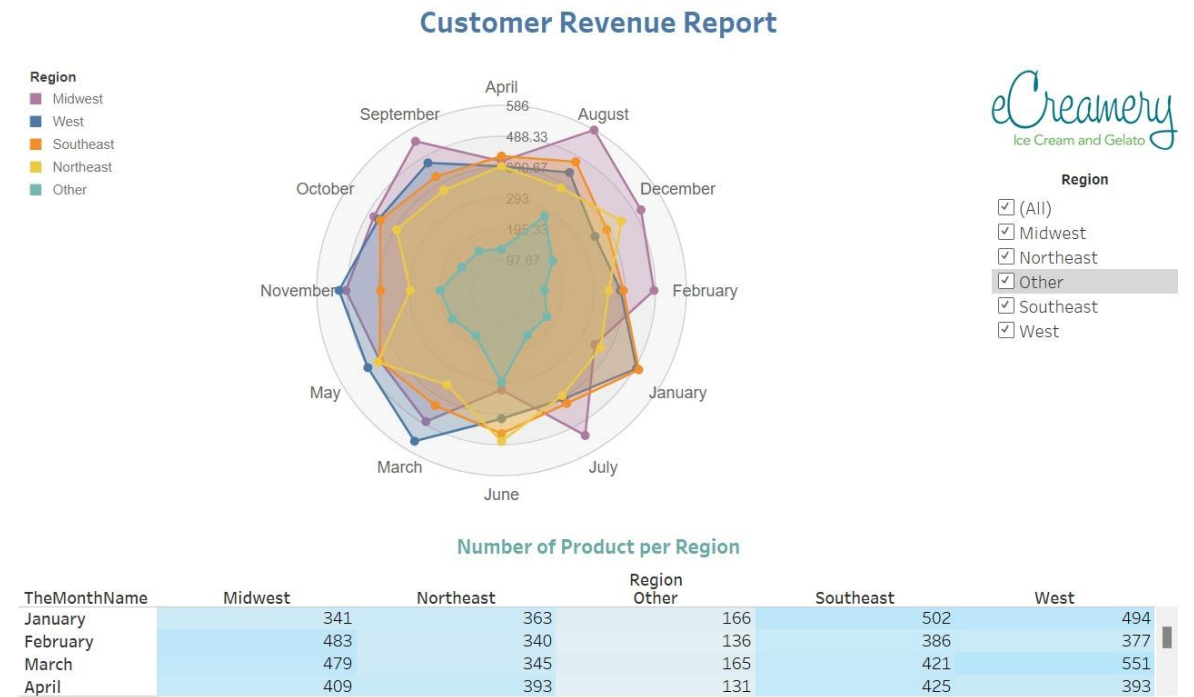
היררכיה מהמדינות שנוצרה בהן ההזמנה למדינות שאליהן נשלחה ההזמנה לפי הטעמים הכי רווחיים
רווחים פר חודש רווחים פר מדינה.



Sunburst Flavor



Spider



נספחים:

ERD

```

CREATE TABLE CUSTOMERS(
Email          varchar(60)          NOT NULL,
[First Name]   varchar(30)          NOT NULL,
[Last Name]    varchar(30)          NOT NULL,
Gender         char                  NULL,
[Phone Number] varchar(30)          NOT NULL,
[State C]      varchar(30)          NOT NULL,
[City C]       varchar(30)          NOT NULL,
Coupon        decimal              NOT NULL,
CONSTRAINT PK_CUSTOMERS PRIMARY KEY(Email)
)

CREATE TABLE [DELIVERY COMPANY](
[Company ID]    int                  NOT NULL,
[State D]       varchar(30)          NOT NULL,
[Shipment Cost] money              NOT NULL,
Capacity        int                  NOT NULL,
Fine            money              NOT NULL,
[Express Fine]  money              NOT NULL,
CONSTRAINT PK_DC PRIMARY KEY([Company ID])
)

CREATE TABLE PRODUCTS(
[Product Type] varchar(50)          NOT NULL,
Flavor          varchar(30)          NOT NULL,
Price           money              NOT NULL,
[Package Color] varchar(10)          NOT NULL,
[Mix-ins]       varchar(10)          NOT NULL,
Design          varchar(7)          NOT NULL,
Cost            money              NOT NULL,
CONSTRAINT PK_PRODUCTS PRIMARY KEY([Product Type],Flavor),
CONSTRAINT C_Price CHECK (Price >= 0),
CONSTRAINT C_Cost  CHECK (Price >= 0)
)

CREATE TABLE [SHIPMENT ADDRESSES] (
Email          varchar(60)          NOT NULL,
AddressID      int                  NOT NULL,
State          varchar(30)          NOT NULL,
City           varchar(30)          NOT NULL,
Street         varchar(30)          NOT NULL,
Number         varchar(5)           NOT NULL,
[Postal Code]  varchar(30)          NOT NULL,
CONSTRAINT PK_SA PRIMARY KEY(Email, AddressID),
CONSTRAINT FK_SA_Email FOREIGN KEY (Email)
REFERENCES CUSTOMERS(Email)
)

```




```
CREATE TABLE ORDERS(
  OrderID          int          NOT Null,
  OrderDate        date         NOT Null,
  OrderTime        time         NOT Null,
  [Express Delivery] bit        NOT Null,
  [Shipment Price] money        NOT NULL,
  [Arrival Date]   date         NOT NULL,
  [Company ID]     int          NOT NULL,
  Email            varchar(60)  NOT NULL,
  AddressID        int          NOT NULL,
  CONSTRAINT PK_ORDERS PRIMARY KEY(OrderID),
  CONSTRAINT FK_ORDERS_Company FOREIGN KEY([Company ID])
REFERENCES [DELIVERY COMPANY]([Company ID]),
  CONSTRAINT FK_ORDERS_Email FOREIGN KEY (Email)
REFERENCES CUSTOMERS(Email),
  CONSTRAINT FK_ORDERS_AddressID FOREIGN KEY (Email, AddressID)
REFERENCES [SHIPMENT ADDRESSES] (Email, AddressID),
)
```

```
CREATE TABLE REVIEWS(
  ReviewID          int          NOT Null,
  ReviewDate        date         NOT Null,
  ReviewTime        time         NOT Null,
  Rating            int          NOT Null,
  [Review Description] varchar(400) NOT Null,
  Email             varchar(60)  NOT Null,
  [Product Type]    varchar(50)  NOT Null,
  Flavor            varchar(30)  NOT Null,
  CONSTRAINT PK_REVIEWS PRIMARY KEY(ReviewID),
  CONSTRAINT FK_REVIEWS_Product FOREIGN KEY ([Product Type],Flavor)
REFERENCES PRODUCTS([Product Type],Flavor),
  CONSTRAINT FK_REVIEWS_Email FOREIGN KEY (Email)
REFERENCES CUSTOMERS(Email),
  CONSTRAINT C_Rating CHECK(Rating BETWEEN 1 AND 5)
)
```

```
CREATE TABLE [MIX-INS](
  [Product Type]    varchar(50)  NOT NULL,
  Flavor            varchar(30)  NOT NULL,
  [Mix-ins]         varchar(30)  NOT NULL,
  CONSTRAINT PK_MIX PRIMARY KEY([Product Type], Flavor, [Mix-ins]),
  CONSTRAINT FK_MIX_Product FOREIGN KEY ([Product Type],Flavor)
REFERENCES PRODUCTS([Product Type],Flavor)
)
```

```
CREATE TABLE QUANTITIES (
  OrderID          int          NOT NULL,
  [Product Type]    varchar(50) NOT NULL,
  Flavor            varchar(30) NOT NULL,
  Quantity          int          NOT NULL,
  CONSTRAINT PK_QUANTITIES PRIMARY KEY (OrderID, [Product Type], Flavor),
  CONSTRAINT FK_QUANTITIES_OrderID FOREIGN KEY (OrderID)
REFERENCES ORDERS(OrderID),
  CONSTRAINT FK_QUANTITIES_Product FOREIGN KEY ([Product Type], Flavor)
REFERENCES PRODUCTS ([Product Type], Flavor),
)
```



MRR

```
CREATE TABLE MRR_CUSTOMERS(
Email          varchar(60) NOT NULL,
[First Name]   varchar(30)      ,
[Last Name]    varchar(30)      ,
Gender         char              ,
[Phone Number] varchar(30)      ,
[State C]      varchar(30)      ,
[City C]       varchar(30)      ,
Coupon         float             ,
CONSTRAINT PK_CUSTOMERS PRIMARY KEY(Email)
)
```

```
CREATE TABLE MRR_DELIVERY_COMPANY(
[Company ID]   int      NOT NULL,
[State D]      varchar(30)    ,
[Shipment Cost] money       ,
Capacity       int           ,
Fine           money         ,
[Express Fine] money         ,
CONSTRAINT PK_DC PRIMARY KEY([Company ID])
)
```

```
CREATE TABLE MRR_PRODUCTS(
[Product Type] varchar(50)      NOT NULL,
Flavor         varchar(30)      NOT NULL,
Price          money            ,
[Package Color] varchar(30)      ,
[Mix-ins]      varchar(50)      ,
Design         varchar(7)       ,
Cost           money            ,
CONSTRAINT PK_PRODUCTS PRIMARY KEY([Product Type],Flavor)
)
```

```
CREATE TABLE MRR_SHIPMENT_ADDRESSES (
Email          varchar(60)      NOT NULL,
AddressID      int              NOT NULL,
State          varchar(30)      ,
City           varchar(30)      ,
Street         varchar(30)      ,
Number         varchar(5)       ,
[Postal Code]  varchar(30)      ,
CONSTRAINT PK_SA PRIMARY KEY(Email, AddressID)
)
```

```
CREATE TABLE MRR_ORDERS(
OrderID        int      NOT Null,
OrderDate      date      ,
OrderTime      time      ,
[Express Delivery] bit      ,
[Shipment Price] money    ,
[Arrival Date] date      ,
[Company ID]   int        ,
Email          varchar(60) ,
AddressID      int        ,
CONSTRAINT PK_ORDERS PRIMARY KEY(OrderID)
)
```



```
CREATE TABLE MRR_REVIEWS(  
ReviewID          int          NOT NULL,  
ReviewDate        date          ,  
ReviewTime        time          ,  
Rating            int          ,  
Email             varchar(60)   ,  
[Product Type]    varchar(50)   ,  
Flavor            varchar(30)   ,  
CONSTRAINT PK_REVIEWS PRIMARY KEY (ReviewID)  
)
```

```
CREATE TABLE MRR_QUANTITIES (  
OrderID           int          NOT NULL,  
[Product Type]    varchar(50) NOT NULL,  
Flavor            varchar(30) NOT NULL,  
Quantity          int          ,  
CONSTRAINT PK_QUANTITIES PRIMARY KEY (OrderID, [Product Type], Flavor)  
)
```

STG

```
CREATE TABLE STG_CUSTOMERS(
Email          varchar(60)      ,
[First Name]   varchar(30)      ,
[Last Name]    varchar(30)      ,
Gender         char              ,
[Phone Number] varchar(30)      ,
[State C]      varchar(30)      ,
[City C]       varchar(30)      ,
Coupon        float             ,
)
```

```
CREATE TABLE STG_DELIVERY_COMPANY(
[Company ID]    int              ,
[State D]       varchar(30)      ,
[Shipment Cost] money           ,
Capacity        int              ,
Fine            money           ,
[Express Fine]  money           ,
)
```

```
CREATE TABLE STG_PRODUCTS(
[Product Type]  varchar(50)      ,
Flavor          varchar(30)      ,
Price           money           ,
[Package Color] varchar(30)      ,
[Mix-ins]       varchar(50)      ,
Design          varchar(7)       ,
Cost            money           ,
)
```

```
CREATE TABLE STG_ORDERS(
OrderID         int              ,
DWProduct       int              ,
DWCustomerID    int              ,
DWCompanyID     int              ,
OrderDate       date             ,
[Express Delivery] bit           ,
[Part Shipment Price] money      ,
[Arrival Date]  date             ,
[State D]       varchar(30)      ,
[City D]       varchar(30)      ,
Price           money           ,
Quantity        int              ,
Amount          money           ,
[Total Price]   money           ,
[Actual Shipping Days] int       ,
)
```

```
CREATE TABLE STG_REVIEWS(
ReviewID        int              ,
ReviewDate      date             ,
DWProduct       int              ,
DWCustomerID    int              ,
Rating          int              ,
CONSTRAINT PK_REVIEWS PRIMARY KEY(ReviewID),
CONSTRAINT C_Rating CHECK(Rating BETWEEN 1 AND 5)
)
```

DW

```
CREATE TABLE DIM_CUSTOMERS(
DWCustomerID      int identity(1,1),
Email              varchar(60) ,
[First Name]      varchar(30) ,
[Last Name]       varchar(30) ,
Gender            char ,
[Phone Number]    varchar(30) ,
[State C]         varchar(30) ,
[City C]          varchar(30) ,
Coupon            float ,
ValidFrom         Date ,
ValidUntil        Date ,
)
```

```
CREATE TABLE DIM_DELIVERY_COMPANY(
DWCompanyID        int identity(1,1) ,
[Company ID]       int ,
[State D]          varchar(30) ,
[Shipment Cost]    money ,
Capacity           int ,
Fine              money ,
[Express Fine]     money ,
ValidFrom          Date ,
ValidUntil         Date ,
)
```

```
CREATE TABLE DIM_PRODUCTS(
DWProduct          int identity(1,1) ,
[Product Type]     varchar(50) ,
Flavor             varchar(30) ,
Price             money ,
[Package Color]    varchar(30) ,
[Mix-ins]          varchar(50) ,
Design            varchar(7) ,
Cost              money ,
ValidFrom          Date ,
ValidUntil         Date ,
)
```

```
CREATE TABLE FACT_ORDERS(
DWOrder            int identity(1,1),
OrderID            int ,
DWProduct          int ,
DWCustomerID       int ,
DWCompanyID        int ,
OrderDate          date ,
[Express Delivery]  bit ,
[Part Shipment Price] money ,
[Arrival Date]     date ,
[State D]          varchar(30) ,
[City D]           varchar(30) ,
Price              money ,
Quantity           int ,
Amount             money ,
[Total Price]      money ,
[Actual Shipping Days] int ,
)
```

```
CREATE TABLE FACT_REVIEWS(
```



```

ReviewID          int          ,
ReviewDate        date         ,
DWProduct         int          ,
DWCustomerID     int          ,
Rating            int          ,
)

```

```

create table DIM_DATE
(
TheDate           date          ,
TheDay            int           ,
TheDayName        varchar(20)  ,
TheDayOfWeek      int           ,
IsWeekend         int           ,
TheWeek           int           ,
TheWeekOfMonth    int           ,
TheMonthOfYear    varchar(20)  ,
TheMonth          int           ,
TheMonthName      varchar(20)  ,
TheQuarter        int           ,
TheFirstOfQuarter date         ,
TheLastOfQuarter  date         ,
TheYear           int           ,
IsLeapYear        int           ,
)

```

```

CREATE TABLE SUMMARY_REVIEW(
DWProduct          int          ,
MonthOfYear        varchar(30)  ,
[AVG Rating]       float        ,
[Total Reviews]    int          ,
)

```

```

CREATE TABLE SUMMARY_STATE_DEMOGRAPHY(
[State C]          varchar(30)  ,
Year              int           ,
[Total Orders]     int           ,
Income            money         ,
Population         int           ,
[Total Shipping Price] money    ,
[Total Products Cost] money    ,
[Total Orders Price] money    ,
)

```

```

CREATE TABLE FACT_REJECTED_ORDERS(
DWOrder           int identity(1,1) ,
OrderID           int           ,
DWProduct         int           ,
DWCustomerID     int           ,
DWCompanyID       int           ,
OrderDate         date          ,
[Express Delivery] bit          ,
[Part Shipment Price] money      ,
[Arrival Date]    date          ,
[State D]         varchar(30)    ,
[City D]          varchar(30)    ,
Price             money         ,
Quantity          int           ,
Amount            money         ,
)

```

```
[Total Price]          money          ,  
[Actual Shipping Days] int            ,  
)
```

```
CREATE TABLE OLD_VERSION_CUSTOMERS(  
DWCustomerID    int identity(1,1),  
Email           varchar(60)      ,  
[First Name]    varchar(30)      ,  
[Last Name]     varchar(30)      ,  
Gender          char             ,  
[Phone Number]  varchar(30)      ,  
[State C]       varchar(30)      ,  
[City C]        varchar(30)      ,  
Coupon          float            ,  
ValidFrom       Date             ,  
ValidUntil      Date             ,  
)
```

```
create procedure [dbo].[updateOldVersions]  
as BEGIN  
update DIM_CUSTOMERS set ValidUntil = GETDATE()-1 where DWCustomerID in (Select DWCustomerID from  
OLD_VERSION_CUSTOMERS)  
truncate table OLD_VERSION_CUSTOMERS  
END
```