

סיכום שיעור 4 28/12 (מתן הלל)

שאלות משיעורי הבית:

** מאחורי הקלעים הסדר של קריאת השאילתא:

FROM >> WHERE>>GROUP BY>>SELECT>>ORDER BY

** שאלה 16- שני הרכיות לORDER BY, תחילה מהמחיר הגבוהה לנמוך, ואז בכל המחירים

שהם אותו דבר תכנס ההרכיה השניה במקרה של השאלה לפי ytd_sales

ורוצים מהקטן לגדול אז נמיין את זה עם ASC.

CASE WHEN RIGHT(email, len('Hogwarts.ac')) = 'Hogwarts.ac' **

Then 1

ELSE 0

END AS is_hogwarts_email

אם נגמר בכתובת של הוגוורץ תרשום לי 1 אם לא אז 0

חיבור בין טבלאות וגרפים:

JOIN - לפעמים נרצה לחבר בין כמה **טבלאות** בDB, שיש עמודות מטבלאות שונות שאנחנו רוצים

שיוצגו לנו אחד ליד השני.

על מנת לחברא בין טבלאות אנחנו צריכים נקודה שמשותפת לשני הטבלאות, ודרך השוואת העמודה

הזו נוכל להצמיד בין שני הטבלאות לפי המזהה המשותף.

יש מספר סוגים של JOIN כרגע אנחנו מדברים על **INNER JOIN**, יסומן גם כJOIN.

ON - קובע את העמודה בה אנחנו רוצים לצור את ההצמדה, מביא ארק את העמודות שיש בהם

התאמה בין שני העמודות.

```
SELECT *  
FROM students  
JOIN DimGeography  
ON students.GeographyKey = DimGeography.GeographyKey
```

ID	StudentName	LastName	Email	GeographyKey	GeographyKey	City	Country
1	Jon	Yang	jon24@adventure-works.com	26	26	Rockhampton	Australia
2	Eugene	Huang	eugene10@adventure-works.com	37	31	Hobart	Australia
3	Ruben	Torres	ruben35@adventure-works.com	31	37	Seaford	Australia

**קורה לאחר הSELECT.

** ה JOIN קורה לאחר הFROM, כלומר מצטרף לטבלה שהיא רשומה בFROM

** על מנת לזמן את העמודה מכל אחת מהטבלאות, נרשום (שם עמודה.שם טבלה) table

. name.coulom name

**אם יש מזהה בעמודה בטבלה אחת שלא נמצא בטבלה השניה אז לא נקבל את הרשומות הללו.

ID	StudentName	LastName	Email	GeographyKey	GeographyKey	City	Country
1	Jon	Yang	jon24@adventure-works.com	26	26	Rockhampton	Australia
2	Eugene	Huang	eugene10@adventure-works.com	37	31	Hobart	Australia
3	Ruben	Torres	ruben35@adventure-works.com	31	37	Seaford	Australia
900	Joe	Zhu	joe12@adventure-works.com	300	500	Beijing	China

** יש סוגים אחרים של JOIN שיאפשר לנו לקבל את העמודות שמסוננות בINNER_JOIN

דוגמה

```
SELECT *  
FROM students  
INNER JOIN houses  
ON students.house.id = houses.ID
```

פלט

אנחנו רוצים לשלב את טבלת הביתים לטבלת הסטודנטים בעמודת ה houses.id שהוא נמצא בשניה הטבלאות. ID - PK של הסטודנט, FK ה ID של הבית.

```
SELECT *  
FROM users  
JOIN DimGeography  
ON users.GeographyKey = DimGeography.GeographyKey  
JOIN DimSalesTerritory  
ON DimSalesTerritory.SalesTerritoryKey = DimGeography.SalesTerritoryKey
```

פלט

שילוב של ששלושה טבלאות, יוזרים, גאוגרפי, סטיל.

הגדרות קשורות ל JOIN

PK- Primary Key - עמודה בטבלה שכל השדות בה חייבים להיות ייחודיים, מזהה רץ חד ערכי וייחודי, כל מזהה יופיע פעם אחת וזה יכול לשמש אותנו כPK של הטבלה. יש סימן קטן של מפתח שהיא עמודה חד ערכית ואין ערך שחוזר פעמיים.

****** קו בין הטבלאות מעיד על הקשר בין שני הטבלאות.

FK - Foreign Key - עמודה שהיא השניה בהררכיה, כלומר היא משולבת בתוך טבלה שיש לה PK אחר. יכולים להיות בעמודה זו כפילויות.

ID	StudentName	LastName	Email	GeographyKey
1	Jon	Yang	jon24@adventure-works.com	26
2	Eugene	Huang	eugene10@adventure-works.com	37
3	Ruben	Torres	ruben35@adventure-works.com	31
900	Joe	Zhu	joe12@adventure-works.com	300

PK

ID	StudentName	LastName	Email	GeographyKey
1	Jon	Yang	jon24@adventure-works.com	26
2	Eugene	Huang	eugene10@adventure-works.com	37
3	Ruben	Torres	ruben35@adventure-works.com	31
900	Joe	Zhu	joe12@adventure-works.com	300

PK

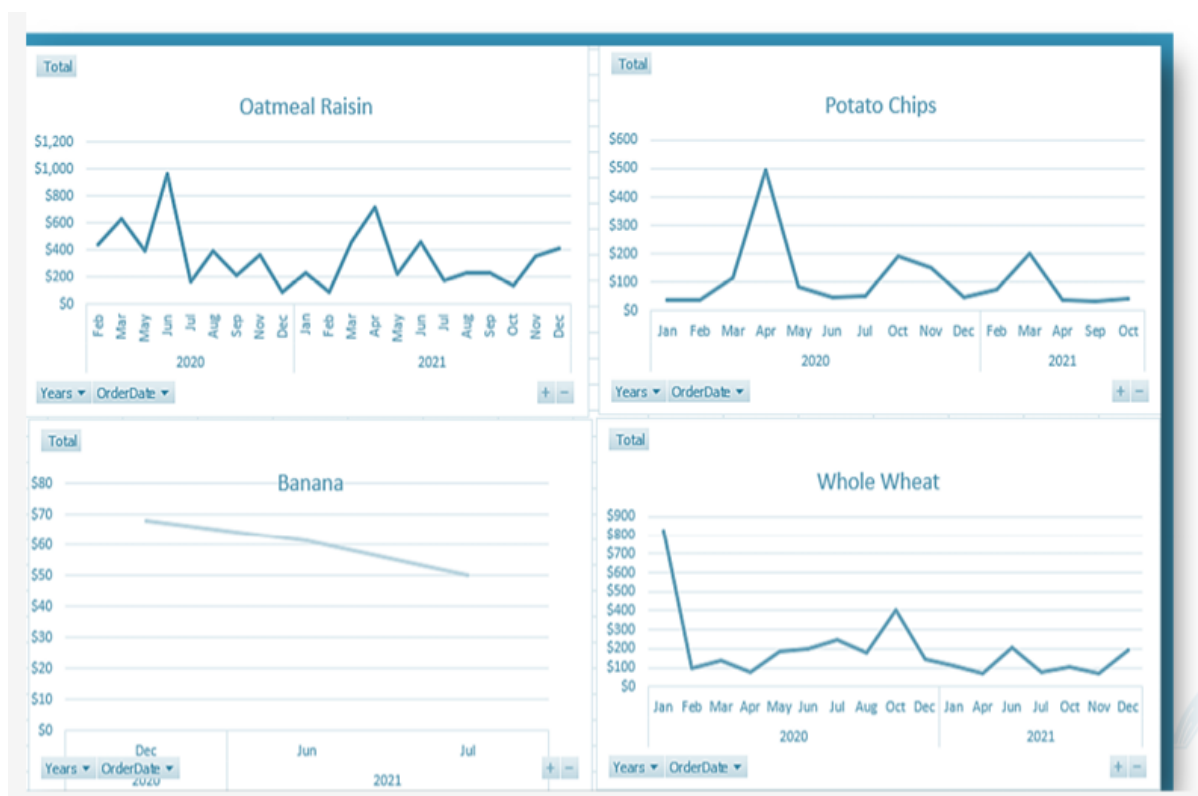
FK

הדמיית נתונים:

לאחר העיבוד של הנתונים נרצה להראות את הנתונים והמסקנות בצורה ברורה וטובה ושיבינו את זה גם אם לא מבינים בדאטה. צורת הצגה ויזואלית.

**חשוב לשים לב לא להעמיס על הגרף יותר מידי כדי שיהיה קריא, לדוגמה מי מכר הכי הרבה.

Total	Product			
Quarter	Banana	Oatmeal Raisin	Potato Chips	Whole Wheat
2020				
Jan			\$38	\$824
Feb		\$443	\$36	\$98
Mar		\$633	\$114	\$140
Apr			\$494	\$80
May		\$392	\$82	\$188
Jun		\$963	\$47	\$202
Jul		\$159	\$52	\$244
Aug		\$389		\$178
Sep		\$210		
Oct			\$192	\$405
Nov		\$366	\$153	
Dec	\$68	\$82	\$49	\$143
2021				
Jan		\$227		\$108
Feb		\$82	\$76	
Mar		\$460	\$203	
Apr		\$716	\$40	\$73
May		\$219		
Jun	\$61	\$454		\$209
Jul	\$50	\$170		\$80
Aug		\$224		
Sep		\$224	\$34	
Oct		\$133	\$42	\$105
Nov		\$352		\$70
Dec		\$409		\$192



תרשים או גרף הוא הצגה של מידע בצורה ויזואלית.

איך נדע איזה מידע נרצה להשיג?

נשאל כמה שאלות כלליות:

1. איזה סוג מידע מדובר? קטגוריות, רצף, גאוגרפיים, תאריכיים.

- נתונים עם מאפיינים גאוגרפיים, כגון מדינה, כתובת, מיקוד.
- נתונים עם מאפיינים תאריכיים, הפרוסים על פי זמן/תאריך/רצף.
- נתונים של קבוצות או תתי קבוצות. או קטגוריות

2. מה נרצה לעשות? השפעה, סך החלקים אל מול שלם, לאורך זמן, קיום יחס או כיוון היחס.

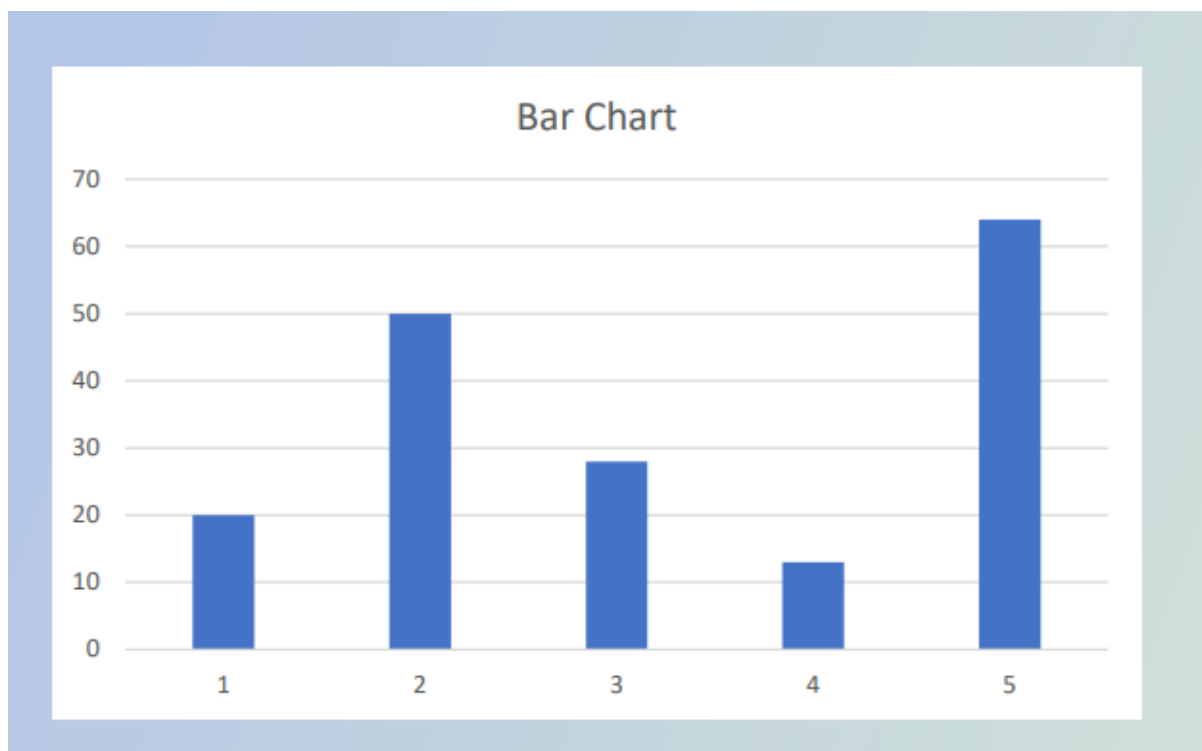
- השוואות של נתונים לאורך זמן או בין קטגוריות.
- לפרק את חלקי השלם- להוציא מידע ספציפי מדאטה בייס מורחב
- תדירות ערכים בתוך סדרה, כמה ילדים הוציאו ציון מעל 80 לדוגמה.
- יחסים בין משתנים מרובים

3. למי נציג את זה? מנהל, צוות עובדים, אדם נטרלי, ציבור.

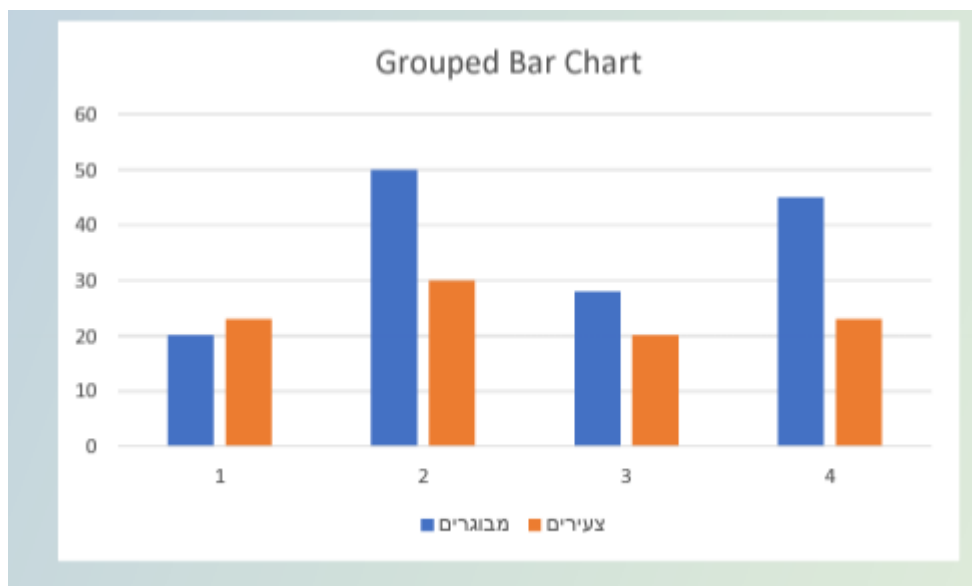
- מנתל- יעדיף לראות ולהבין מה קורה ברמה הפרטנית.
- מנהל- רוצה מידע מסכם עם תובנות ברורות הניתנות לפעולה.
- צוות בכיר- צריך KPI ברמה גבוהה, ומעקב אחר ביצועי העסק.
- ציבור - דורש ויזואליזציה ספור ברור לעקוב.

סוגי גרפים

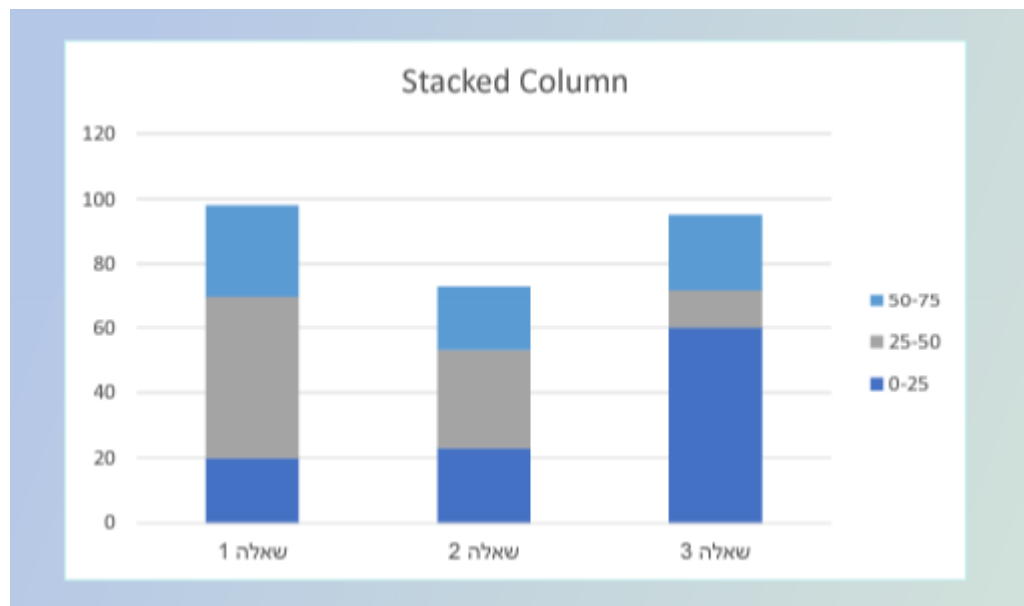
Bar Chars - אפשר להשוות בין שני קבוצות שונות. קל להראות השוואות ולהעריך גדלים..



Group bar Chart - ניתן להשוות גם אם יש קבוצות שונות, כל צבע מעיד על קבוצה אחרת.

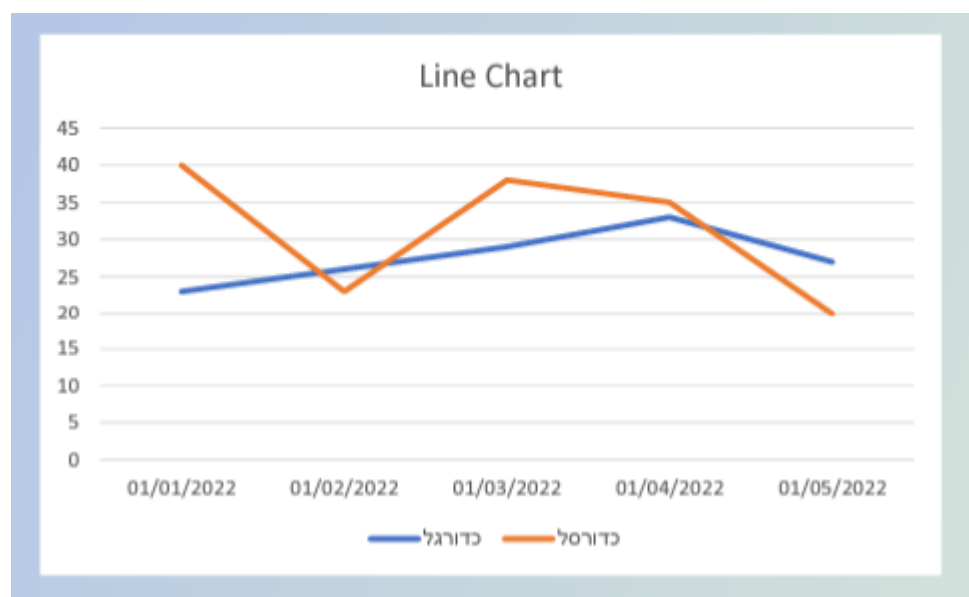


Stack Column - אנחנו יכולים להשוות בתוך השם את החלקים שלו, בצבעים השונים שמכילים את הBAR הכללי, וגם נוכל לבחון בהבדל לאחרים.

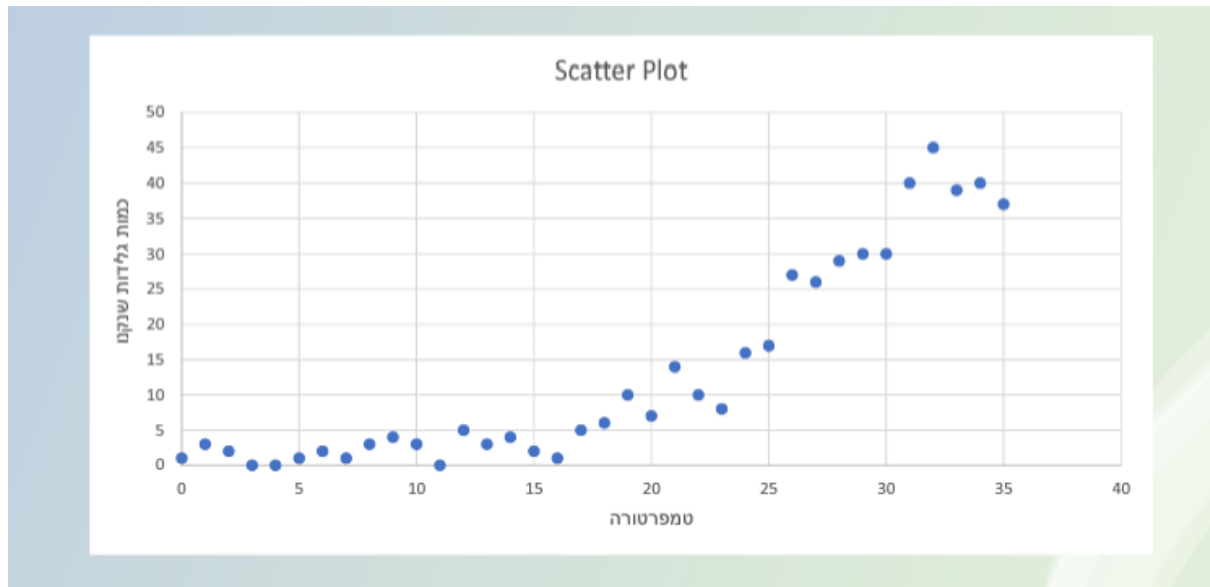


****על מנת שיהיה ברור עדיף שהפרופורציות יהיו רשומות.**

Line Chart - טוב למעט קטגוריות מראה לנו התקדמות עם תלות בשני משתנים



Scatter Plot - מאפשר לנו לראות קורלציה, מתאים למחקר, חישוב של פונקציות ליניאריות, להראות מתאמים של מקרי בוחן.



Gauge Chart - הגעה ליעד מסויים, ערך מספרי של אחוז הצלחה וכו'.



גרף עוגה- גרף מבלבל שמקשה להראות הבדלים.

