

שיעור ראשון EXEL For Data Analysis

מבוא לאקסל מונחים בסיסיים

אקסל עובד עם טבלה ובתוכה יש תאים המזוהים בשורות ועמודות.

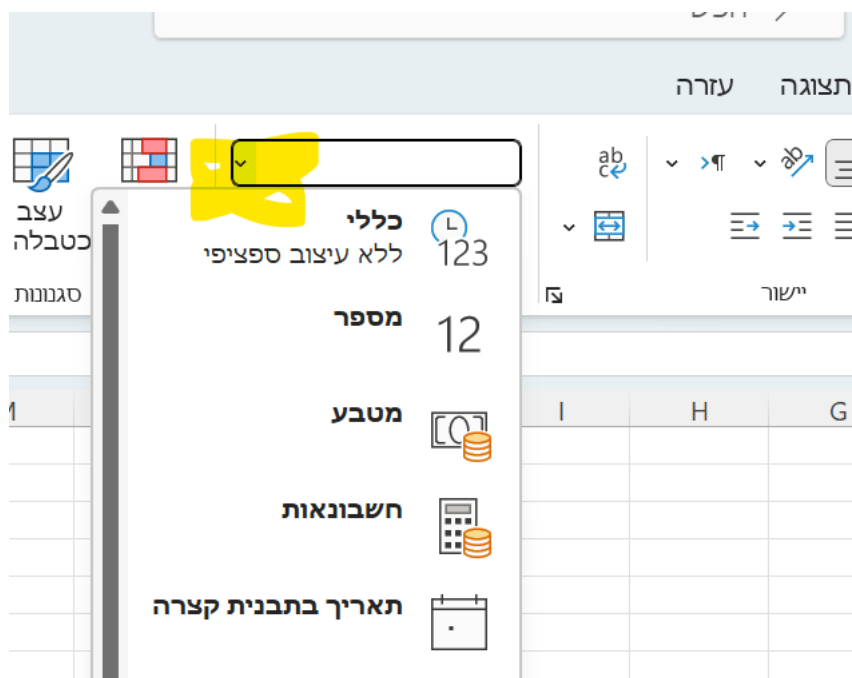
שורה - שורות מתחילות מהספרה 1 והן מסמלות את הקווים האופקיים בטבלה.

עמודה - מתחילות מ A והן מסמלות את הקווים האנכיים בטבלה.

תא - תא הוא נקודת המפגש בין השורה לעמודה, כל תא מאות באנגלית + מספר.

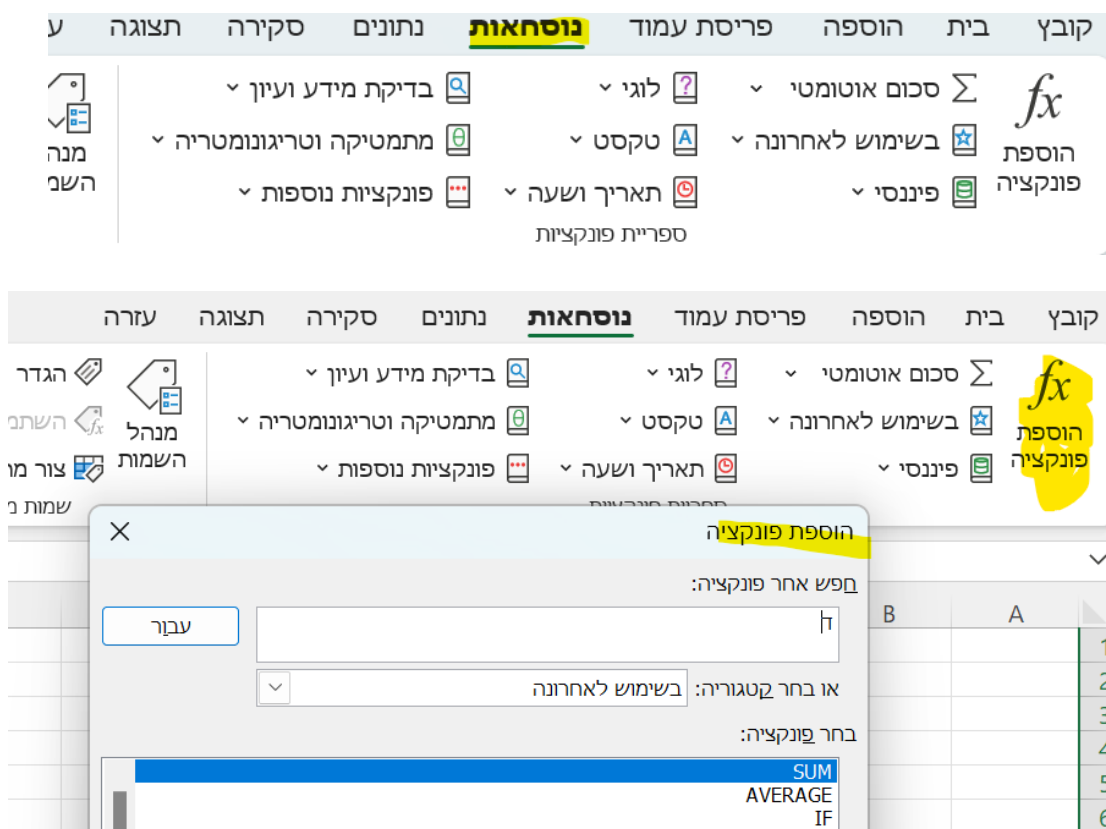
טווח תאים - אוסף של תאים נבחרים.

****לכל ערך שאנחנו מכניסים אנחנו יכולים לבחור איזה סוג הוא, יכול להיות תאריך, מספר, טקסט וכו'.. נוכל לבצע את השינוי של הסוג כך :**

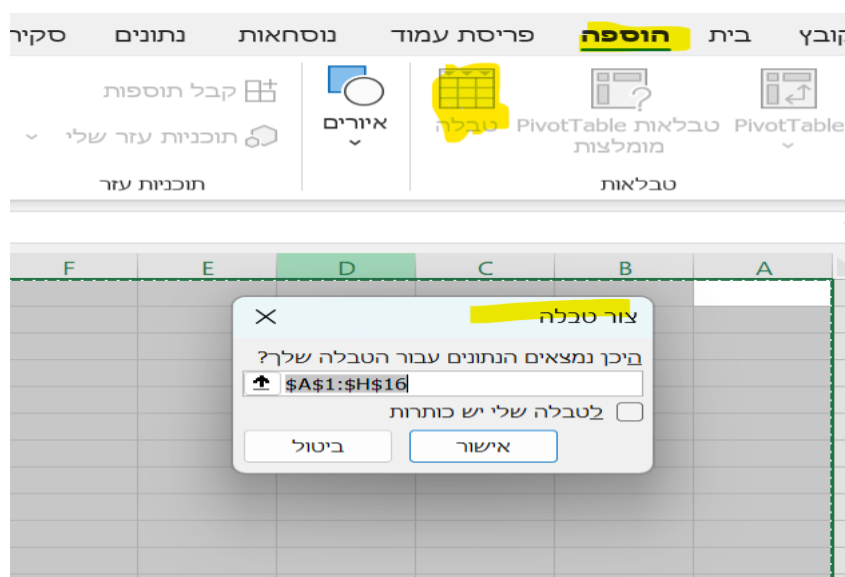


****על מנת לסמן את כל העמודה נלחץ על קונטרול+שיפט+חץ למטה.**

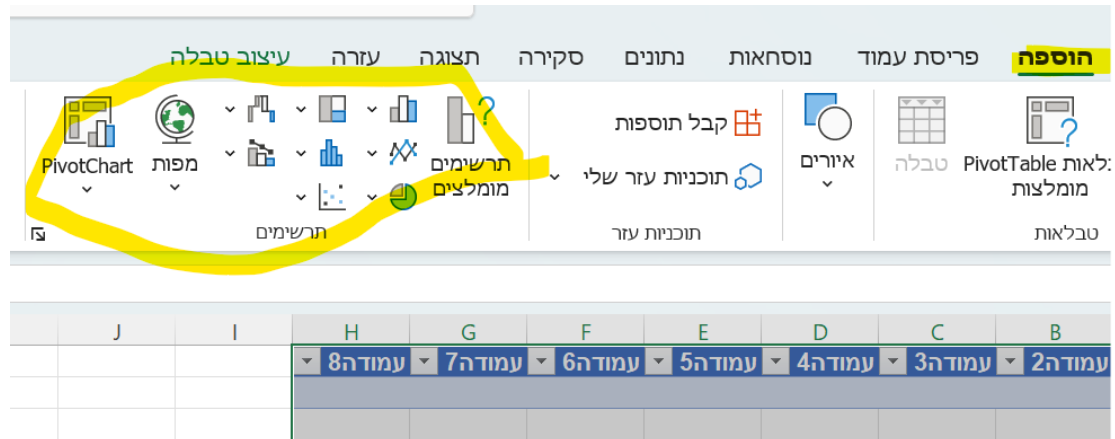
פונקציות – פונקציות הן נוסחאות מוגדרות מראש המבצעות חישובים באמצעות ערכים ספציפיים, הנקראים **ארגומנטים**, בסדר או במבנה. ישנם הרבה פונקציות אנחנו ננסה למיין אותם לפי סוג הנתון, למשל טקסט ואז נראה את כל הפונקציות הקשורות בטקסט.



****הגדרה כטבלה** – נעשה זאת על ידי סימון כל הנתונים או שניתן לרשום ידנית את הטווחים, ואז נוכל לקבץ את הנתונים כטבלה, וזה יאפשר לנו לבצע פונקציות ביותר יעילות, יעזור לנו להבין ויזואלית, וכן רוב העבודה בSQL זה עם טבלאות. ניתן ללחוץ גם על F3



גרף – תרשים, ייצוג חזותי של נתונים מגיליון העבודה בדרך ויזואלית, כך שיהיה נוח יותר לקרוא את הנתונים. אם נרצה לדעת למשל כמות מינמלית ביום מסויים בטבלה זה יהיה מאתגר כיוון שיש מספר רב של שורות ונתונים, לכן נעדיף להשתמש בגרף שנוכל לראות זאת באופן נוח יותר. יש כמות גדולה של גרפים ובתוך כל אחד יש תתי גרפים, זה מאפשר לנו לשנות את הרזולוציה של האופן בו נסתכל על המידע. בתור דאטה אנליסט נצטרך לדעת לעומק את הויזואליזציות.



טבלת ציר – pivot table נותן לנו אפשרות אדרגטיבית, כאשר נלחץ על הPIVOT נבחר לפתוח את זה בדף חדש ושם נוכל להתעסק עם הנתונים, שם נוכל לבחור לפי איזה קטגוריה אנחנו נרצה לפלטר את הטבלה, איזה קטגוריה נרצה לבחון ואיזה פונקציה נרצה לבצע עליה. הוא מאפשר לנו לבצע עיבוד של הנתונים על אותו הדאטה סט בקלות. טבלה אינטרקטיבית לרוב, המסכמת כמות גדולה של נתונים. ניתן לקבוע איזה נתון אנו רוצים שיהיה בציר ה X ואיזה בציר ה Y, וכן איזה אינטרגציה נדרשת.

פונקציות

פונקציה היא פעולה חשבונית המוגדרת מראש לביצוע חישובים וניתוח נתונים, יש באקסל כ-450 פונקציות, שכל אחת מבצעת פעולה אחרת. חוסך לנו זמן, ורשימת נוסחאות מורכבות

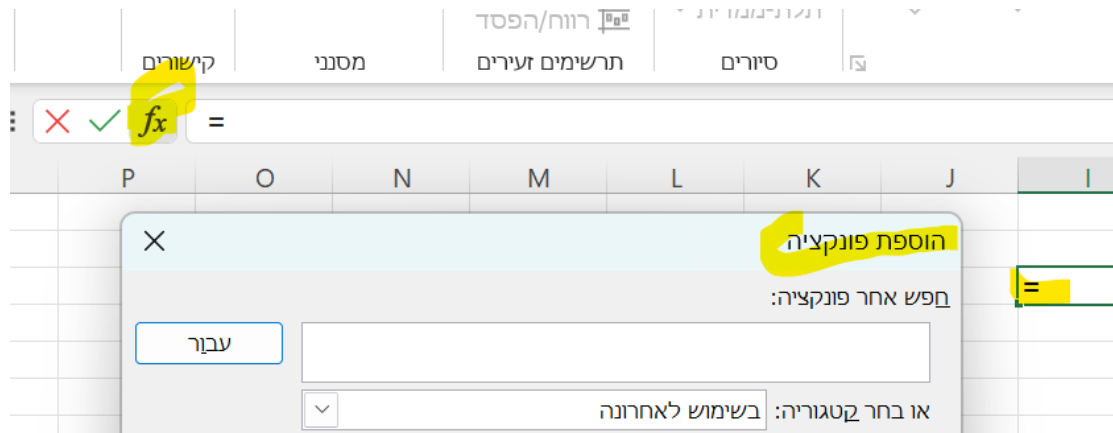
1. **ההבדל בין נוסחה לפונקציה** – נוסחה יכולה להכיל מספר פונקציות, כך שתבצע מספר פעולה, בנוסחאות אנחנו מגדירים את, נוסחה גמישה ופונקציה מוגדרת וספציפית.

**** על מנת להתחיל נוסחה נרשום בתא נטרלי = ואז נסמן אילו תאים אנחנו רוצים לכלול בנוסחה, ונשים בין התנאים את הפעולה שאנחנו רוצים לבצע, לדוגמה +, -, *, / ... לאחר מכן ENTER** (אם הנתונים בתוך טבלה הנוסחה תפעל לאורך כל הטבלה, אם לא מסומן בטבלה אז הנוסחה היתה פועלת רק על הנתונים שסימנתי לקליק כפול על הנקודה או גרירה של תא הנוסחה תאפשר לנו להכילה גם על תאים נוספים במצב שהם לא בטבלה)).

****פונקציה חייבת להיות סגורה, כלומר בין שני ערכים או יותר.**

****לשים לב!** כאשר עושים נוסחה על עמודות שלמות להקפיד שהסימון יהיה עד סוף הטבלה ולא עד סוף הגיליון. נוכל לבצע את זה על ידי קונטרול + שיפט + חץ למטה.

****גישה מהירה לחלונית הפונקציות.**



פונקציות סיכום

SUM – מצבע פעולת חיבור בין הערכים הנבחרים. חייב שיהיה משהו לסכום, אם זה יהיו שמות של ערים ושל ארצות, או תאריכים.

SUMIF – חיבור כמו SUM רק שהיא מוסיפה את כל הערכים (מספרים) העונים על קריטריון שאנחנו הגדרנו. נזין את ערכי הטווח של עמודה אחת, הקריטריון (התנאי שלי שמם) וזו את התנאי, לדוגמה "<6" (ואת הטווח ערכים של העמודה השניה. נצטרך למלא שלושה שדות. אז נקבל את הסכום של כל מי שעונה עליו.

**** תמיד** שמם את התנאי בקריטריון בתוך גרשיים ""

**** אם** נרצה יותר מקריטריון אחד נרשום SUMIFS

**** כשאני רוצה לבצע סכימה עם קריטריון צריך שיהיה אותו Range בשני צדדיו.**

COUNT – השימושית ביותר, תמיד מחפשים כמויות של רשומות. הוא סופר רק לפי מספרים.

COUNTIF – סופר רק את הערכים העונים על קריטריון מסויים. נצטרך למלא שני שדות, הראשון הוא הטווח (RANGE) והשני הוא התנאי (CRITERIA) כלומר הקריטריון, לדוגמה גדול מ, שונה מ, שווה ל, וכו.. ואז נקבל את מספר הרשומות שעונים על הקריטריון. יכול לענות על שאלה של כמה אנשים קונים מעל 10 פירות בשבוע.

COUNTA – תמיד מחפשים כמויות של רשומות. הוא לא רגיש לסוג הערך, הוא יודע לספור כמה רשומות יש לי בתוך עמודה מסויימת. יספור כמה רשומות יש לנו.

****ניתן לבצע את פעולת COUNTA על כמה עמודות, והוא יחבר אותם למספר אחד, אם יש לנו 230 רשומות ואנחנו מבצעים ספירה על שני עמודות נקבל תוצאה של 460.**

****לשים לב** שאם אנחנו רוצים לבצע ספירה על עמודה מסומיית אז שהכותרת לא מסומנת גם, כיוון שהיא עלולה להיכנס לספירה ולהטעות לנו את הנתונים.

COUNTBLANK – סופר את כמות התאים הריקים בטווח. הסיבה לשימוש בה, אם יש דאטה סט עם שדות ריקים שאין בהם נתונים(נפוץ) , חלק מהעבודה כאנליסטים עלינו לבצע QA על הדאטה. **לשים לב** שאנחנו מסמינים את הטבלה ספציפית כי אחרת אנחנו נקבל את כל התאים הריקים לאורך כל העמודה.

AVERAGE – ממוצע של מספרים בטווח של תאים (תאים ריקים ישפיעו על הממוצע)

AVERAGEIF – מבצע ממוצע רק על הערכים העונים על קריטריון מסויים. לבצע ממוצע על תאריך יהיה בעייתי מבחינת התוצאה אבל אפשרי כיוון שהפונקציה הזאת מחשיבה אותו כמספר.

** אם אנחנו רוצים לדעת איזה ערכים מכילה עמודה מסויימת ובמידה ויש ערכים שחוזרים מספר פעמים לאורך העמודה, נבצע את הפעולה "הסרת כפילויות", "Remove Duplicate" ואז נקבל את הערכים הייחודיים בעמודה.