שיעור 2

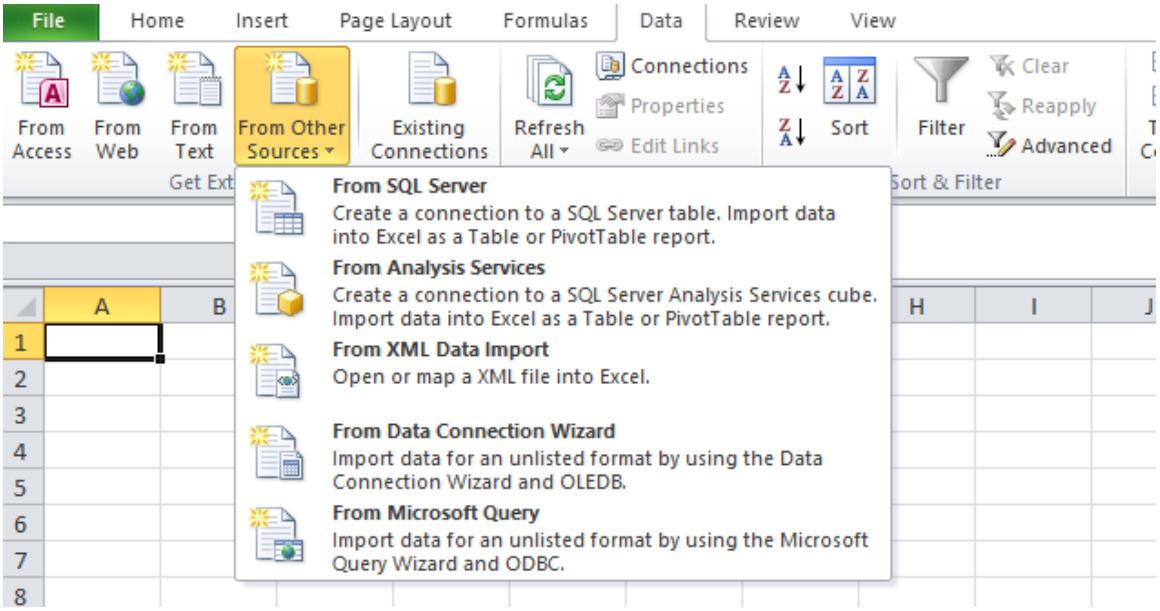
ETL – EXTRACT, TRANSFORM, LOAD

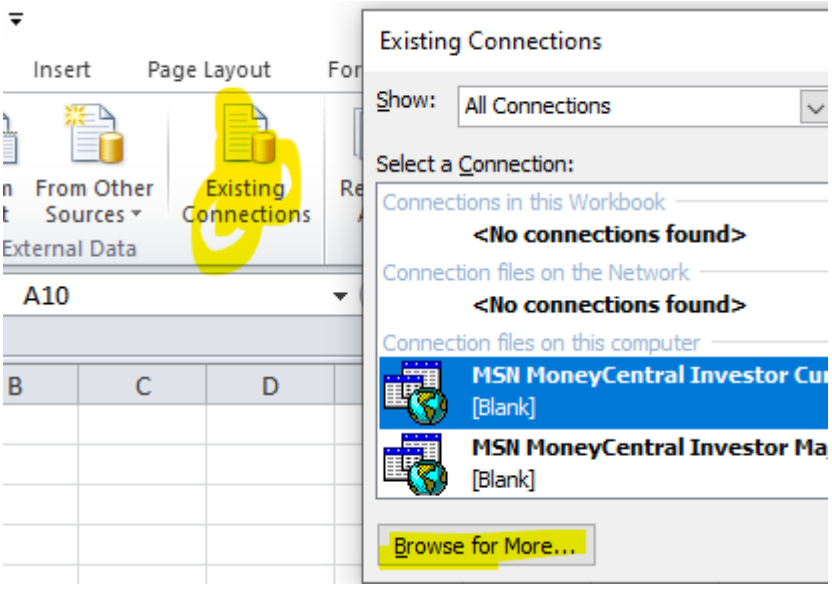
שלושה פונקציות נפרדות הנדרשות לשלב נתונים ממקורות שונים למיקום אחד, ניתן לבצע טרנספורמציות על הנתונים , וחילוץ המידע הרלוונטי. פונקציית החילוץ (Extract) מתייחסת לקריאת נתונים ממקור מוגדר ולחילוץ המידע הנדרש. פונקציית השינוי (Transform) מתייחסת לניקוי, עיצוב, צבירת נתונים והמרת מבנה הנתונים. פונקציית טעינה (Load) מתייחסת לטעינת נתונים ליעדם.

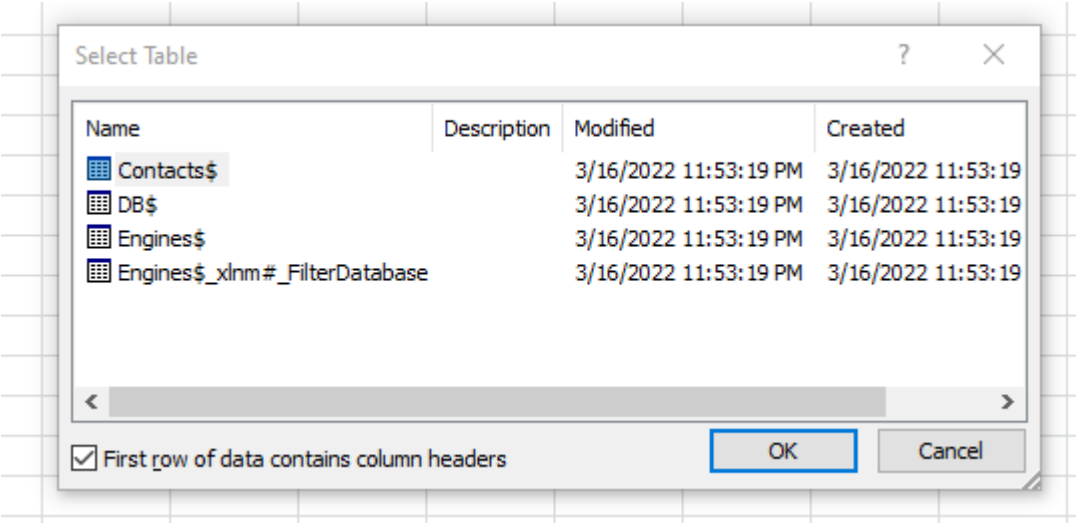
Power Query – שיפור חווית התהליך על ידי הצעה של מנגנון אינטואיטיבי לחילוץ נתונים, ביצוע טרנספורמציות מורכבות,וטעינת נתונים לחוברת עבודה.

Extract – המערכת יכולה לבצע ייבוא נתונים ממקורות מידע שונים, אם יש לנו את הנתונים מראש התהליך הוא יחסית פשוט. נבצע את הפעולה לפי השלבים הבאים:

1. **מיקום מדוייק של הקובץ:** במידה ואנו מיבאים נתונים משרת SQL אנו צריכים את כתובת/שם השרת במידה ואנו מייבאים מקובץ יש צורך במקומו המדוייק.
2. **שם וסיסמא:** חלק מהמקומות שמהם נצתרך את המידע יצריכו אותנו להשיג שם משתמש וסיסמא.







אקסל היא תוכנה שמייצרת גליונות אלקטרונים], שמאפשרות ניהול נתונים באמצעות צערכות מפותחות , כלים גרפים, ופשת תכנות בשם VBA (Visual Basic For Application)

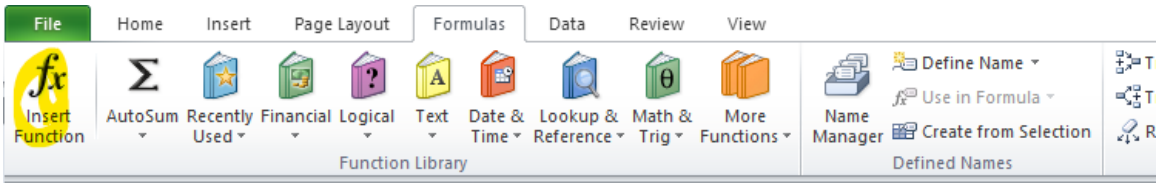
יתרונות:

1. נגישות – אחד הכלים הפופולרים ביותר, מאופיס.
2. כלים אנליטים – ניתוח כמויות מידע גדולות, ביצועי תחזיות בעזרת ידע בסיסי.
3. טיפול בנתונים – כלי מצויין לניתוח נתונים , אחסון, וביצוע חישובים, כלי סינון ומיון.

חסרונות:

1. כמות Volume גדולה, במקרים של DB מאד גדול, האקסל יתקשה בעיבוד המידע, כיוון שהוא מערכת טקסטואלית. הוא יכול להגיע עד כמה מאות אלפי נתונים למרות תוכנות SQL שיכולות לעבד מליונים.
2. שלישה ואבטחה – עקב כך שיש פעמים שהטבלאות גדולות במיוחד נצתרך לפעמים לחלק את המידע והטבלה שלנו לתת טבלאות קטנות , מה שיכול לגרום לאיבוד של מידה או הפרת סדר. יש לאקסל נטייה להתשמש בקירב במספרים גדולים (חוסר דיוק), חשופים למידע שגוי.
3. סנכרון נתונים: לאחר הטעינה הראשונה האקסל לא מסתנכנן בצורה אוטומטית.

בלשונית נוסחאות נמצא את על הנוסחאות הקיימות באקסל



- סכום – יסכם את העמודה או השורה
- ממוצע – יבצע ממוצע של כל הנתונים הנבחרים
- ספירת מספרים – ספירת תאים עם נתונים בלבד.
- מקסימום/מינימום – הערך המקסימלי או המינימאלי בנתונים.

E Transform L

המרת נתונים, כדי שיתאימו למערכת אליה אנו רוצים להטעים את הנתונים , ההמרה יכולה להיות ניקוי הנתונים מחסרי ערך, או מכפילויות. או לדוגמה שינוי ערך השדה מטקסט למספר וכדומה.

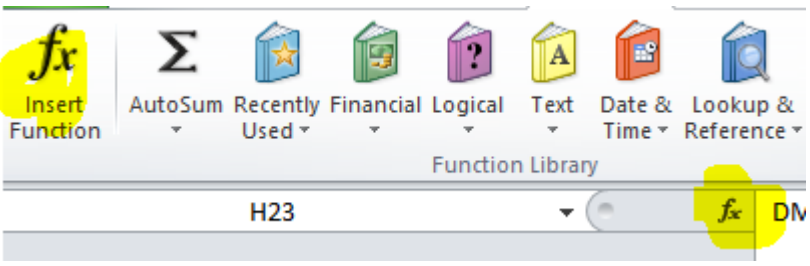
Power Query - חלונית שיש בה את כל האופציות הנדרשות לשינוי הוספה ומחיקה של הנתונים אותם אנו רוצים לקלוט, ואז נוכל גם לשנות את הסוג של הנתונים. נוכל גם בחלונית זו למחוק מידע כפול, או לא רלוונטי ולטפל בשיבושים שונים במידע. בלשונית המרת נתונים נקבל את היכולת להמיר נתונים ולהחליף ערכים. ניתן לשכפל לחשב ולהכפיל עמודות מסויימות.

ET Load

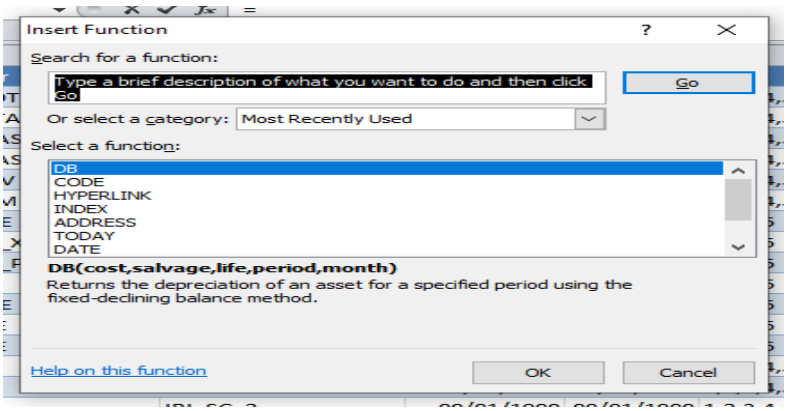
הטעינה היא לחוברת עבודה ולא לבסיס נתונים ף הטעינה ניתו לבצע כבר בשלב היבוא או לאחר איבוד הנתונים, יהיה גם בחלונית Power Query ואז ניתון להשתמש בביצוע ויזיאולזיות בניתוח הנתונים.

Standard Function

על מנת להתעסק עם פונקציות נזמן את החלונית פונקציות.



בחלונית שנפתחת נוכל לבחור את הפונקציה ולאחר מכן להשתמש במסך הערה על מת להשלימה.



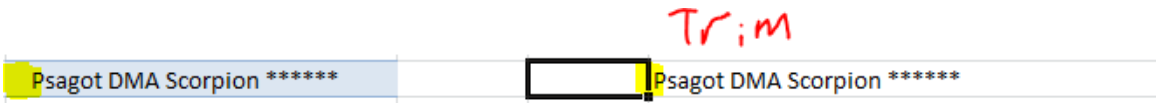
<u>סוגי פונקציות</u>

Concatenate – פונקציה זו תחבר בין מחרוזות בתאים או בטווח של תאים. פעולה זאת תצרף לנו כמה נתונים בתאים בנבחרים לתא אחד. זאת על מנת לשלב מידע שכולל חלקי מידע ממספר עמודות ותאים. תצוגתי. דוגמה:

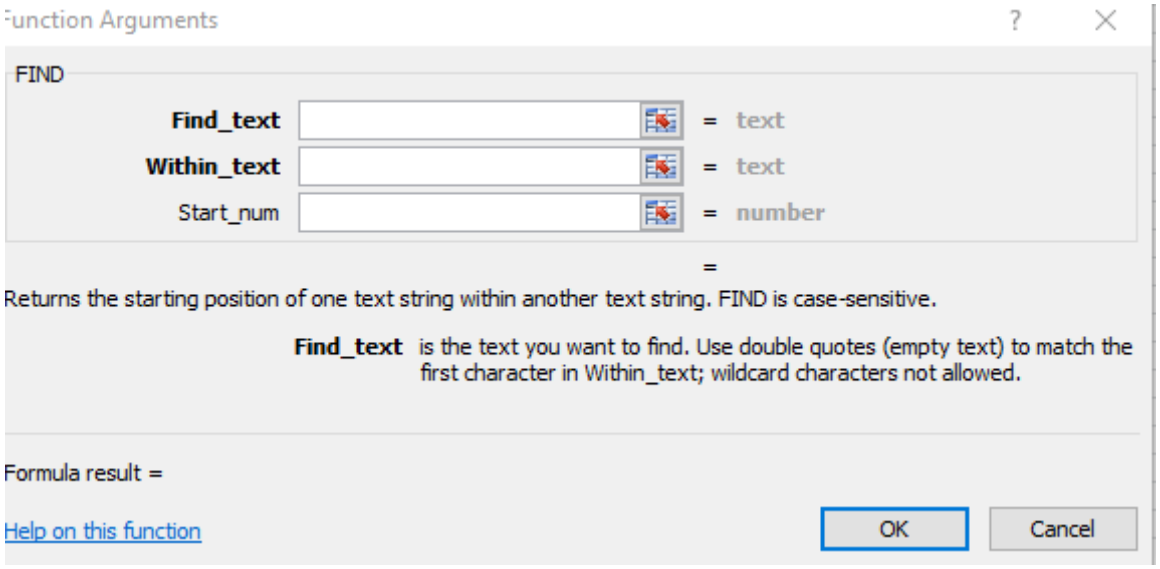
=ADRESS],[AGE],[NAME]([CONCATENATE])

נקבל עמודה שכל תא יכיל את השילוב של השם גיל וכתובת.

Trim – ניקוי רווחים לפני ואחרי הטקסט הנבחר. מסדר לנו את הויזואליזציה, מצמיד לנו את הטסקט לצד אחד באופן מסודר.



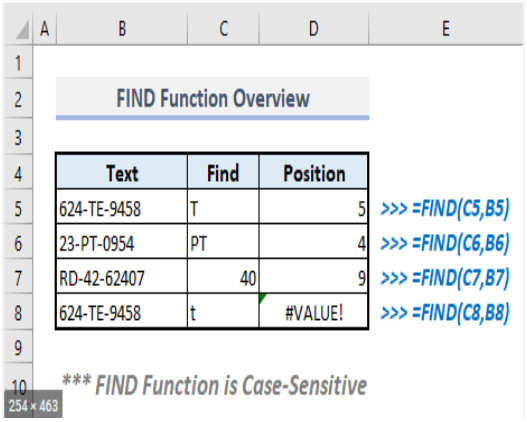
Find – פונקציה צחפש טקסט בתוך טקסט כאשר יהיה ניתן להגדיר מאיזה תו אני רוצה שיתחיל לחפש. התוצאה תיהיה המיקום של הטקסט שחיפשנו.



Find_text – איזה טקסט אני רוצה למצוא, יש להשתמש בגרשיים " " , וההתאמה בין מה שאנחנו מחפשים לבין הקיים בטבלה חייב להיות זהה לחלוטין.

Within_text – הטקסט שמכיל את הטקסט שאני רוצה למצוא.

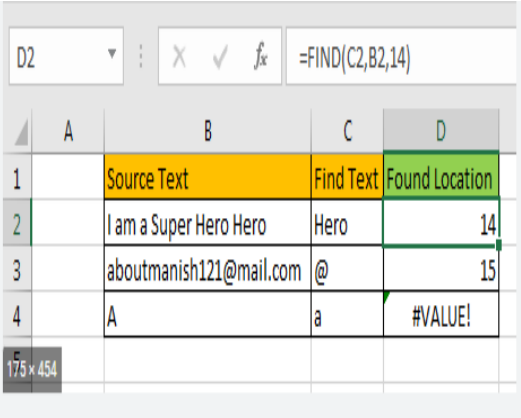
Start_Num – איפה מתחיל החיפוש , האות הראשונה ב Within_text.



FIND Function Overview

Text	Find	Position	
624-TE-9458	T	5	>>> =FIND(C5,B5)
23-PT-0954	PT	4	>>> =FIND(C6,B6)
RD-42-62407	40	9	>>> =FIND(C7,B7)
624-TE-9458	t	#VALUE!	>>> =FIND(C8,B8)

*** FIND Function is Case-Sensitive



D2 =FIND(C2,B2,14)

	Source Text	Find Text	Found Location
	I am a Super Hero Hero	Hero	14
	aboutmanish121@mail.com	@	15
	A	a	#VALUE!

Replace - מחליפה את ערך הטקסט לערך טקסט אחר.

REPLACE(old_text, start_num, num_chars, new_text)

Old text - הטקסט שתרצה להחליף חלק מהתווים שבו

Start num - מיקומו של התו ב OLD TEXT שברצוני להחליף ל NEW TEXT

Num chars - מספר התווים ב OLD TEXT שברצוני להחליף בNEW TEXT

New text - הטקסט שיחליף את התוים המסויימים.

=REPLACE(B11,C11,D11,E11)

פונקציית REPLACE

=REPLACE(OLD_TEXT,START_NUM,NUM_CHARS,NEW_TEXT)

=REPLACE(B11,C11,D11,E11)

דוגמא: החלפת קטע טקסט בטקסט אחר

תוצאה	טקסט חדש	כמה אותיות להחליף	מספר תו התחלה	טקסט ישן
מכבי חיפה	חיפה	8	6	מכבי תל אביב
Atletico Madrid	Atletico	4	1	Rial Madrid
יום נהדר	נהדר	3	5	יום טוב
השומר אחי אנוכי !	!	1	17	השומר אחי אנוכי ?

LEFT/RIGHT - מחזירה ערך של טקסט בגודל מוגדר מצד שמאל או ימין של הטקסט.

LEFT(text, [num_chars])

Text - מחרוזת הטקסט המכילה את התווים שברצונך לחלץ.

Num_chars - מציין את מספר התווים שברצונך לחלץ באמצעות LEFT.

- Num_chars חייב להיות גדול מאפס או שווה לאפס.
- אם num_chars גדול מהאורך של text, הפונקציה LEFT מחזירה את כל הטקסט.
- אם num_chars מושמט, המערכת מניחה כי הוא שווה ל- 1.

פונקציית LEFT

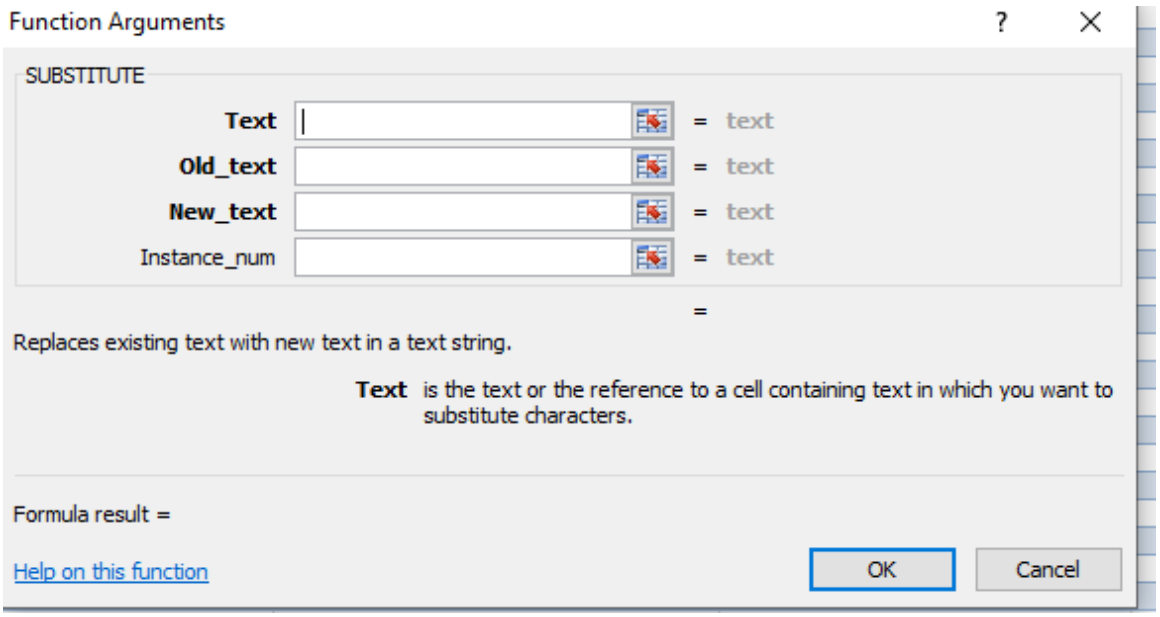
=LEFT(TEXT,[NUM_CHARS])

=LEFT(B11,C11)

דוגמא: חילוץ מספר נבחר של אותיות מתחילת הטקסט
בעברית LEFT יתחיל לספור מצד ימין- כי ההתחלה היא בימין

תוצאה	מספר אותיות	טקסט
באר	3	באר שבע
ישראל	5	ישראל ישראלי
מוצר	4	מוצר 1123
30523	5	30523-1098765
1000	4	1000.21
Today	5	Today is

Substitute – פוקציה תחפש ערך טקסטואלי בתוך טקסט ותחליף אותו בערך אחר, ניתן לבחור גם מה להחליף בטקסט.



Text – איפה נמצא הטקסט אותו ארצה לערוך

Old_text – הטסקט אותו נרצה להחליף אם זה לא תואם למה שיש בתא הוא לא יוחלף.

New_Text – איזה טסקט אנחנו רוצים שיופיע במקום הטקסט הקיים.

Instance_Num – מקומות ספציפים בטקסט בהם נקצה שתתבצע ההחלפה. זה לא הכרחי.

Example 1:

Replace "EMP" with "ID" for given dummy text

Dummy Text	Result	Formula Text
EMP-T-8895-EMP	EMP-T-8895-ID	=SUBSTITUTE(C9,"EMP","ID",2)

Example 2:

Remove all spaces from following name and return in cell

Name	Result	Formula Text
Jack Alderson	JackAlderson	=SUBSTITUTE(C14," ","")

Example 3:

Get the count of number of words in each name

Name	Result	Formula Text
James Smith	2	=LEN(TRIM(C19))-LEN(TRIM(SUBSTITUTE(C19,"
Christopher J. Murphy III	4	=LEN(TRIM(C20))-LEN(TRIM(SUBSTITUTE(C20,"
Ronald J. Carlson	3	=LEN(TRIM(C21))-LEN(TRIM(SUBSTITUTE(C21,"

MID - הפונקציה MID מחזירה מספר מסוים של תווים ממחרוזת טקסט, החל מהמיקום שהוגדר, על בסיס מספר התווים שהוגדר, משתמשת בד"כ בפונקציות עזר נוספות, כמו LEN, FIND וכד' המסייעות לדעת מהו אורך הטקסט ומאיזה מקום, מאיזה מספר תו נרצה להתחיל את חילוץ התווים – על מנת לדייק את החזרת הטקסטים הרצויים,

MID(text, start_num, num_chars)

Text - מחרוזת הטקסט המכילה את התווים שברצונך לחלץ.

Start_num - מיקום התו הראשון שברצונך לחלץ ב- text. התו הראשון ב- text מכיל ערך start_num של 1, וכן הלאה.

**אם start_num גדול מהאורך של text, הפונקציה MID מחזירה "" (טקסט ריק).

**אם start_num קטן מהאורך של text, אך start_num ועוד num_chars עולה על האורך של text, הפונקציה MID מחזירה את התווים עד לסוף של text.

**אם start_num קטן מ- 1,או שלילי הפונקציה MID מחזירה את ערך השגיאה #VALUE!.

פונקציית MID

=MID(TEXT,START_NUM,NUM_CHARS)

=MID(B10,C10,D10)

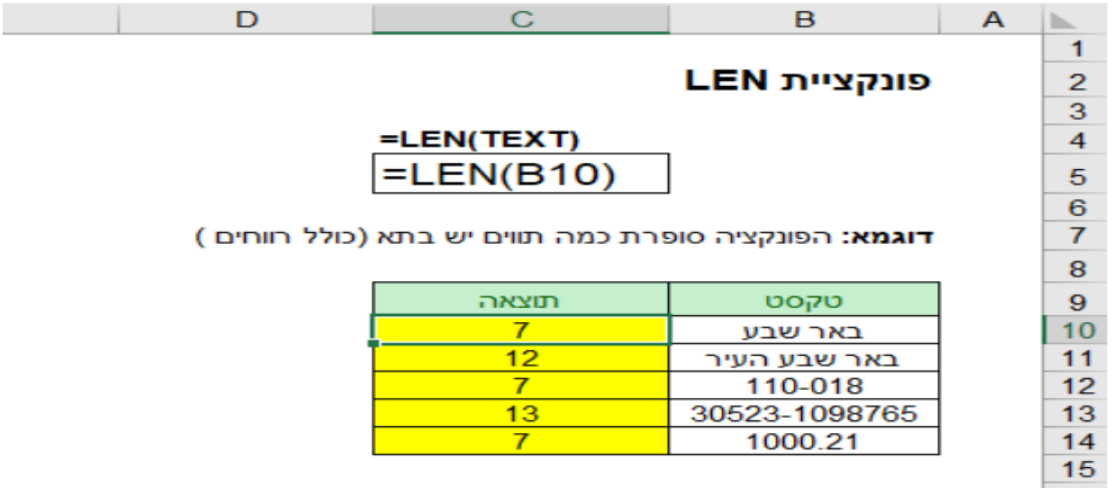
דוגמא: החזרת קטע מטקסט, לאחר שבחרנו מהיכן להתחיל בתוך הטקסט וכמה אותיות לשלוף

טקסט	תתחיל מ	אותיות/פריטים	תוצאה
אין ממה לדאוג היום	5	9	ממה לדאוג
אין ממה לדאוג היום	15	4	היום
לא רוצה לראות אות!	1	2	לא
רוצה לראות אות!	8	3	אות
No Problem?!	4	9	Problem?!

LEN - הפונקציה מחזירה את מספר התווים הנמצאים בתא. פונקציית LEN סופרת גם תווים מיוחדים כמו רווח, מכף וכד'. פונקציית LEN משמשת כפונקציית עזר בפונקציות טקסט אחרות הנדרשות למספר התווים בתא.

LEN(text)

Text- הטקסט שאת אורכו ברצונך למצוא. רווחים נחשבים לתווים.



פונקציית LEN

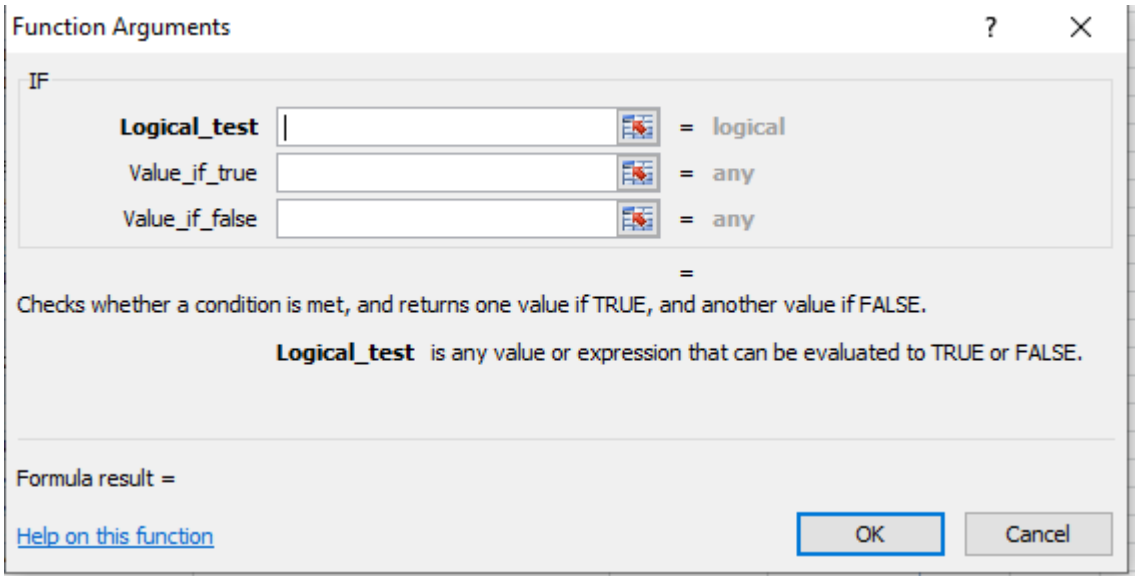
=LEN(TEXT)

=LEN(B10)

דוגמא: הפונקציה סופרת כמה תווים יש בתא (כולל רווחים)

טקסט	תוצאה
באר שבע	7
באר שבע העיר	12
110-018	7
30523-1098765	13
1000.21	7

IF – פונקציה זו תבחן תנאי לפני התוצאה של התנאי (True/False) תחזיר ערך. מאפשרת לך לבצע השוואות לוגיות בין ערך לבין התוצאה הצפויה.



Logical_test – על ערך שמגביל לנכון או לא נכון. בשורה זאת נשים את התנאי שלנו, לדוגמה גדול מתאריך מסויים, או מתחיל באות מסויימת. גדול מ, שווה ל. חשוב לציין גם את התנאי וגם את העמודה שאנחנו רוצים להכיל עליה את התנאי. ניתן לבצע עד 7 תנאים.

Value_if_true – קובע את הפלט שנקבל במידה ואכן התנאי נענה.

Value_if_false – קובע את הפלט שנקבל במידה והתנאי לא מתבצע.

IF(logical_test, value_if_true, [value_if_false])

logical_test - התנאי שברצונך לבדוק.

value_if_true - הערך שברצונך שיוחזר אם התוצאה של *logical_test* היא TRUE.

IF =IF(B2<20;"Order";"No action")

	A	B	C
1	Fruit	Amount in stock	Status
2	Apple	24	=IF(B2<20;"Order";"No action")
3	Banana	30	
4	Kiwi	22	
5	Lemon	18	
6	Melon	9	

=IF(OR(B2="delivered",B2="paid"),"Closed","Open")

	A	B	C	D	E
1	Order no.	Progress	Status		
2	101	Delivered	Closed		
3	102	In transit	Open		
4	103	Delivered	Closed		
5	104	Delivered	Closed		
6	105	Paid	Closed		
7	106	In transit	Open		
8	107	Cancelled	Open		

פונקציית IF

=IF(TEST1,VALUE1,[TEST2],[VALUE2]...)

=IF(C10>90,"מצטיין", "אינו מצטיין")

הפיכת ממוצע ציונים לטקסט לפי תנאי

שם התלמיד/ה	ממוצע ציונים	אם הציון מעל 90 לתלמיד יירשם "מצטיין"
ארבל	95	מצטיין
שירה	92	מצטיין
מאיה	100	מצטיין
לידור	87	אינו מצטיין
אלמוג	90	אינו מצטיין

LOWER/UPPER - פונקציה ממירה את האותיות הרישיות במחרוזת טקסט לאותיות קטנות – lower case או גדולות.

LOWER(text)

Text - הטקסט שברצונך להמיר לאותיות קטנות. הפונקציה LOWER אינה משנה תווים שאינם אותיות ב- text.

פונקציית LOWER

=LOWER(TEXT)

=LOWER(B10)

דוגמא: שינוי טקסט לועזי לאותיות קטנות

טקסט	תוצאה
APPLE	apple
Apple	apple
ZIP:8401	zip:8401
#HELLO_hellp!	#hello_hellp!

Concat - פונקציית CONCAT מאפשרת חיבור טקסטים מתאים שונים לתא אחד(מאפשרת לחבר עד 254 מחרוזות מתאים שונים), הארגומנטים של הפונקציה הם התאים שנרצה לחבר, כאשר במידה ונחבר למשל שם פרטי ושם משפחה שנמצאים בשני תאים שונים, נרצה להוסיף לחיבור גם רווח שהוא עוד תו בתא המחובר לו יחדיו.

**כדי לכלול מפרידים (כגון מרווח או תווי אמפרסנד (&) בין טקסטים שברצונך לשלב ולהסיר ארגומנטים ריקים שאינך מעוניין שיופיעו בתוצאת הטקסט המשולב, באפשרותך להשתמש בפונקציה TEXTJOIN.

CONCAT(text1, [text2],…)S

Text1 - פריט טקסט המיועד לצירוף. מחרוזת או מערך של מחרוזות, כגון טווח תאים.

פונקציית CONCAT

=CONCAT(TEXT1,TEXT2...)

=CONCAT(B11,C11,D11)

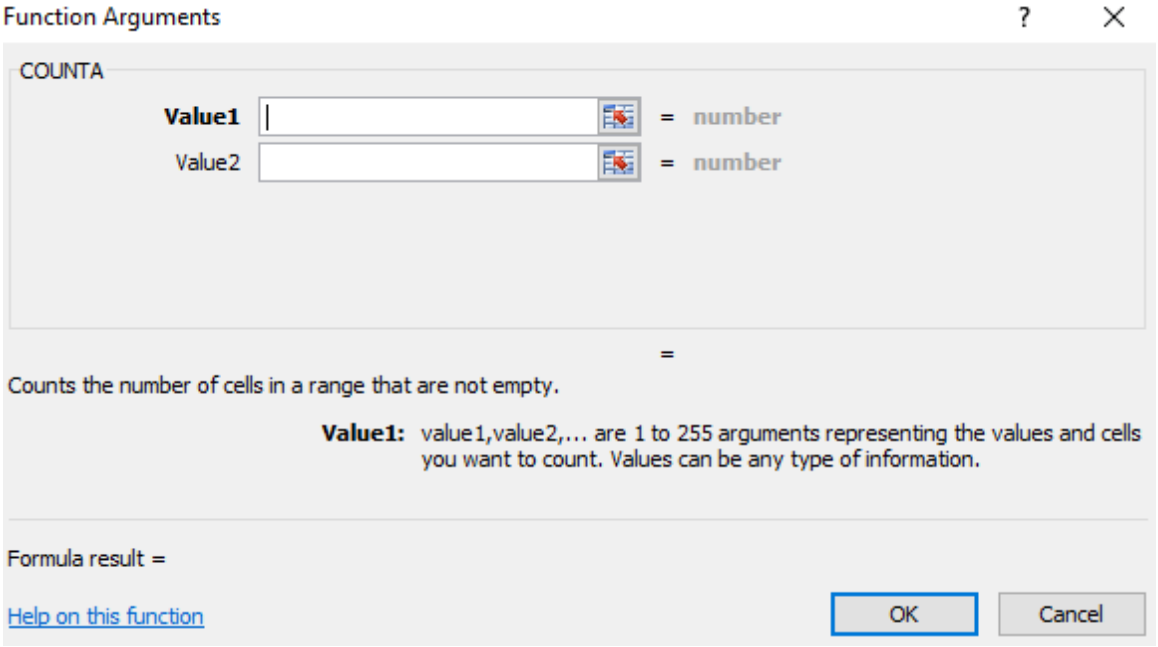
דוגמא: חיבור טקסטים מכמה תאים ו/או מחרוזות מרובות

1	2	3	תוצאת החיבור עם CONCAT
אריה		גולן	אריה גולן
A	B	C	ABC
1	2	3	123

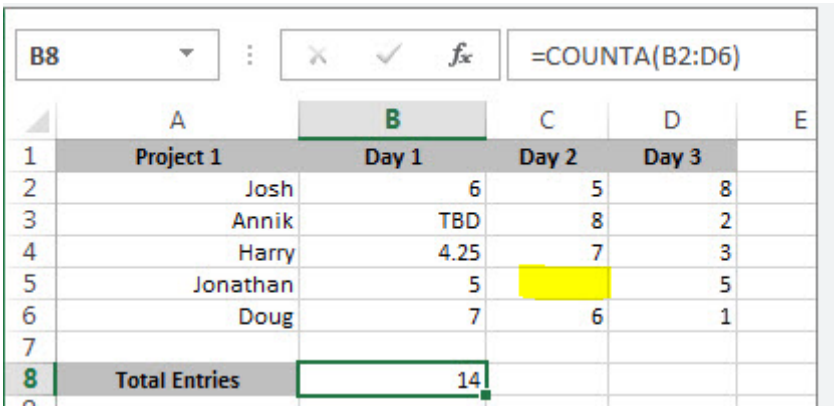
בתא C11 יש רווח

** אם אורך המחרוזת שמתקבלת עולה על 32767 תווים (המגבלה לכל תא), הפונקציה CONCAT מחזירה את השגיאה !#VALUE.

Counta - פונקציה שתחזיר את כמות התאים המכילים ערך הטווח.



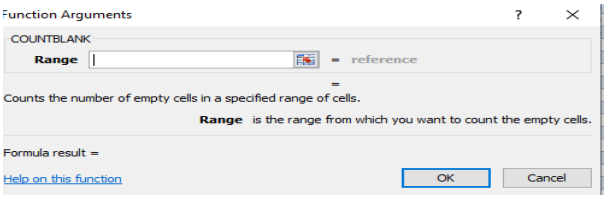
Value1 – מייצג את התאים והערכים שאני רוצה לספור, הם יכולים להיות מכל סוג שהוא כי השתמשנו בקאונט A.



	A	B	C	D
1	Project 1	Day 1	Day 2	Day 3
2	Josh	6	5	8
3	Annik	TBD	8	2
4	Harry	4.25	7	3
5	Jonathan	5		5
6	Doug	7	6	1
7				
8	Total Entries	14		

** הוא לא יספור תאים רקים.

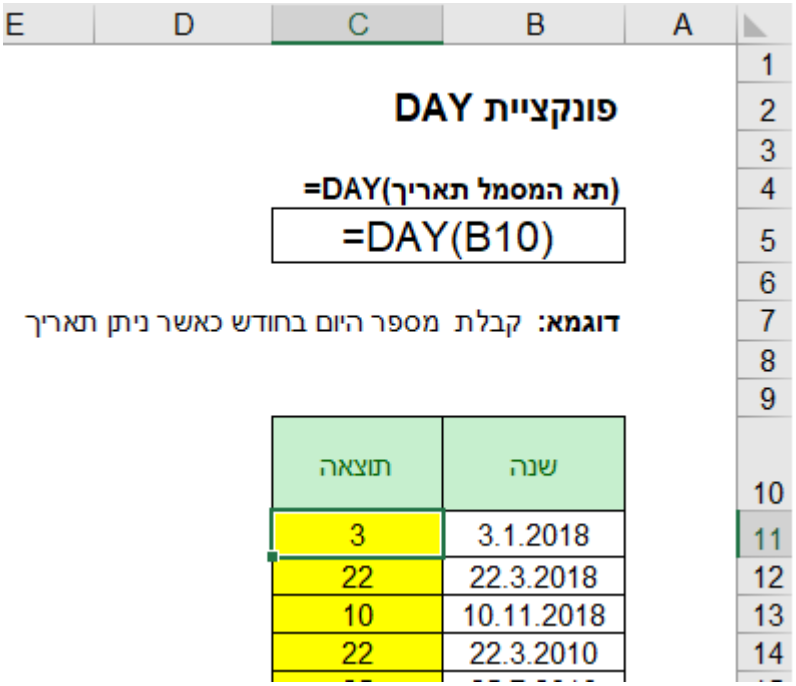
Countblank - פונקציה תבצע ספירה של התאים הריקים בטווח הנתונים.



Day – מחזיר את היום בתאריך יש לו מספרים מ 1- 31.

DAY(serial_number)

Serial_number - התאריך של היום שברצונך למצוא. יש להזין תאריכים באמצעות הפונקציה DATE, או כתוצאות של נוסחאות או פונקציות אחרות



פונקציית DAY

=DAY(תא המסמל תאריך)

=DAY(B10)

דוגמא: קבלת מספר היום בחודש כאשר ניתן תאריך

תוצאה	שנה
3	3.1.2018
22	22.3.2018
10	10.11.2018
22	22.3.2010

Month – חודש מתאריך ה -1 בינואר עד 12 – בדצמבר. בדומה לDAY

Year – נותן את השנה מתאריך 1900-9999 . בדומה לDAY

DAYS - פונקציית Days ב Excel מחזירה את מספר הימים בין שני תאריכים. ם תאריך התחלה ב- A1 ותאריך הסיום ב- B1 הפונקציה : =DAYS (B1, A1) תחזיר את מספר הימים שבין שני התאריכים.

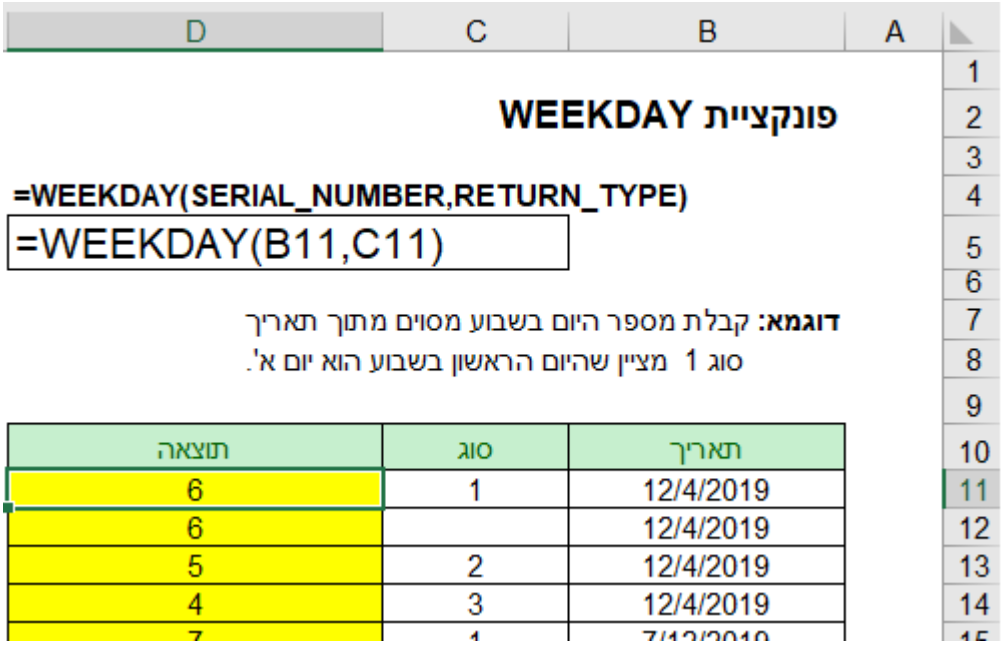
DAYS(end_date, start_date)

Start_date ו- End_date הם שני התאריכים שביניהם ברצונך לחשב את מספר הימים.

Weekday - פונקציית WeekDay מקבלת תאריך ומחזירה מספר בין 1-7 המייצג את היום בשבוע. כברירת מחדל, WEEKDAY מחזירה 1 ליום ראשון ו- 7 לשבת. נשתמש בפונקציה WEEKDAY בתוך נוסחאות אחרות כדי לבדוק את היום בשבוע ולהגיב לפי הצורך.

WEEKDAY(serial_number,[return_type])

Serial_number - מספר רציף המייצג את התאריך של היום שברצונך למצוא



פונקציית WEEKDAY

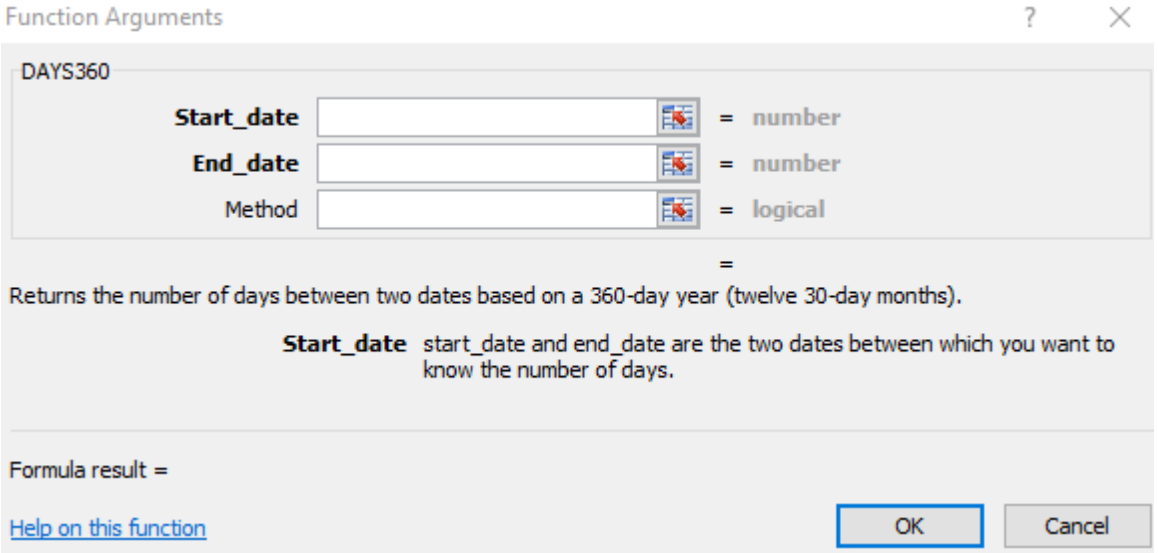
=WEEKDAY(SERIAL_NUMBER,RETURN_TYPE)

=WEEKDAY(B11,C11)

דוגמא: קבלת מספר היום בשבוע מסוים מתוך תאריך
סוג 1 מציין שהיום הראשון בשבוע הוא יום א'.

תוצאה	סוג	תאריך
6	1	12/4/2019
6		12/4/2019
5	2	12/4/2019
4	3	12/4/2019

Days360 – בוחרים שני תאריכים ואז נקבל את מספר הימים בין שני התאריכים הללו

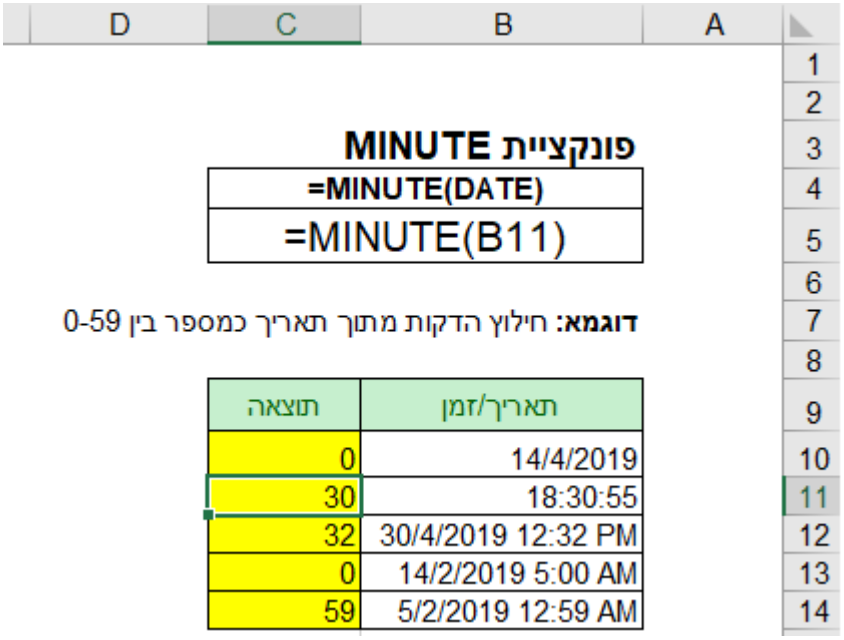


Now – מחזיר את התאריך של היום.

Minute\Hours - פונקציית minute מחלצת את רכיב הדקה מתא מסוים כמספר בין 0-59.
לדוגמה, אם השעה 9:45 בבוקר, הנוסחה תחזיר את המספר 45. אפשר להשתמש בפונקציה MINUTE כדי לחלץ את הדקה לתא.

MINUTE(serial_number)

Serial_number השעה המכילה את הדקה שברצונך לאתר.



פונקציית MINUTE

=MINUTE(DATE)
=MINUTE(B11)

דוגמא: חילוץ הדקות מתוך תאריך כמספר בין 0-59

תאריך/זמן	תוצאה
14/4/2019	0
18:30:55	30
30/4/2019 12:32 PM	32
14/2/2019 5:00 AM	0
5/2/2019 12:59 AM	59

Date - הפונקציה DATE מחזירה את המספר הסידורי הרציף המייצג תאריך מסוים. השתמש בפונקציה DATE כאשר נדרש לקחת שלושה ערכים נפרדים ולשלב אותם ליצירת תאריך.

DATE(year,month,day)

- Year - הערך של הארגומנט *year* יכול לכלול ספרה אחת עד ארבע ספרות. Excel מפרש את הארגומנט *year*בהתאם למערכת התאריכים שבה המחשב משתמש.
כברירת מחדל, Microsoft Excel for Windows משתמש במערכת התאריכים 1900, כלומר התאריך הראשון הוא 1 בינואר, 1900.
- Month - מספר שלם חיובי או שלילי שמייצג את החודש בשנה מ- 1 עד 12 (ינואר עד דצמבר).
- Day - מספר שלם חיובי או שלילי המייצג את היום בחודש מ- 1 עד 31.

פונקציית DATE

=DATE(YEAR,MONTH,DATE)

=DATE(B11,C11,D11)

דוגמא: הפונקציה עוזרת להגדיר תאריך בהינתן שנה, חודש ויום משמש בעיקר ליצירת תאריכים כאשר הנתונים לגבי השנה החודש והיום מופרדים או נוצרים בעזרת פונקציה אחרת.

תוצאה	יום	חודש	שנה
01/01/2019	1	1	2019
05/01/2019	5	1	2019
15/02/2019	15	2	2019
01/03/2019	60	1	2019

**מומלץ להשתמש בארבע ספרות בארגומנט year כדי למנוע תוצאות לא רצויות.

TIME - פונקציית time היא פונקציה מובנת המאפשרת ליצור זמן עם רכיבי שעה, דקה ושנייה בודדים. הפונקציה TIME שימושית כשרוצים להרכיב ערך זמן מתאים בתוך פונקציה אחרת או לתוך תא אחד. הפונקציה מבקשת בפרמטרים שלה: שעה, דקה ושניה

TIME(hour, minute, second)

שעה מספר בין 0 (אפס) ל- 32767 המייצג את השעה. כל ערך הגדול מ- 23 יחולק ב- 24 והשארית תיחשב כערך השעה. לדוגמה, .125 = TIME(3,0,0) = TIME(27,0,0) או 3:00 בלילה.

דקה מספר בין 0 ל- 32767 המייצג את הדקות. כל ערך הגדול מ- 59 יומר לשעות ודקות. לדוגמה, .520833 = TIME(12,30,0) = TIME(0,750,0) או 12:30 אחה"צ.

שניה מספר בין 0 ל- 32767 המייצג את השניות. כל ערך הגדול מ- 59 יומר לשעות, דקות ושניות. לדוגמה, .023148 = TIME(0,33,22) = TIME(0,0,2000) או 12:33:20 בלילה.

פונקציית TIME

=TIME(YEAR,MONTH,DATE)

=TIME(B10,C10,D10)

דוגמא: מחזירה פורמט של שעה מערכים נפרדים של שעה דקות ושניות

תוצאה	שניות	דקות	שעות
08:00:15	15	0	8
12:30:00	0	30	12
23:00:05	5	0	23
23:59:59	-1	0	24
07:15:00	0	75	6