

# Database functions for large datasets

ADVANCED EXCEL FUNCTIONS



Agata Bak-Geerinck  
Product Owner Data, Telenet

# Aggregating large data sets with conditions

## Basket CA-2014-115812

Row ID	Order ID	Returned?	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
1	CA-2016-152156	No	08/11/2016	11/11/2016	Second Class	CG-12520	Claire Gute
2	CA-2016-152156	No	08/11/2016	11/11/2016	Second Class	CG-12520	Claire Gute
3	CA-2016-138688	No	12/06/2016	16/06/2016	Second Class	DV-13045	Darrin Van Huff
4	US-2015-108966	No	11/10/2015	18/10/2015	Standard Class	SO-20335	Sean O'Donnell
5	US-2015-108966	No	11/10/2015	18/10/2015	Standard Class	SO-20335	Sean O'Donnell
6	CA-2014-115812	No	09/06/2014	14/06/2014	Standard Class	RH-11710	Brosina Hoffman
7	CA-2014-115812	No	09/06/2014	14/06/2014	Standard Class	BH-11710	Brosina Hoffman
8	CA-2014-115812	No	09/06/2014	14/06/2014	Standard Class	BH-11710	Brosina Hoffman
9	CA-2014-115812	No	09/06/2014	14/06/2014	Standard Class	BH-11710	Brosina Hoffman
10	CA-2014-115812	No	09/06/2014	14/06/2014	Standard Class	BH-11710	Brosina Hoffman
11	CA-2014-115812	No	09/06/2014	14/06/2014	Standard Class	BH-11710	Brosina Hoffman
12	CA-2014-115812	No	09/06/2014	14/06/2014	Standard Class	BH-11710	Brosina Hoffman
13	CA-2017-114412	No	19/04/2017	20/04/2017	Standard Class	AA-10480	Andrew Allen



- A record represents individual basket items
- Data can be summarized per basket, per customer, per shop etc.
- Conditional aggregation can be applied

## Conditional aggregation functions:

- SUMIFS()
- AVERAGEIFS()
- COUNTIFS()

<sup>1</sup> Image credit <https://unsplash.com/@sarascarpa>

# Limitations of conditional aggregate functions

Multiple AND conditions

*Count returned Florida orders*

A	B	C	D	E
Order	Return?	Month	Shop	State
2435	No	July	Joe's	NY
6565	Yes	August	Wago	Florida
765	No	August	Karo	Florida
65732	No	July	Karo	Ohio
1765	No	July	Karo	Florida
1332	No	August	Wago	NY

```
=COUNTIFS(B:B,"Yes",E:E,"Florida")
```

Multiple OR conditions

*Count returned Florida OR Joe's shop orders*

A	B	C	D	E
Order	Return?	Month	Shop	State
2435	No	July	Joe's	NY
6565	Yes	August	Wago	Florida
765	No	August	Karo	Florida
65732	No	July	Karo	Ohio
1765	No	July	Karo	Florida
1332	No	August	Wago	NY

```
= COUNTIFS(B:B,"Yes",E:E,"Florida") +  
COUNTIF(D:D,"Joe's")
```

# Database functions for the win!

## Family of functions:

- DSUM( )
- DCOUNT( )
- DAVERAGE( )
- DMIN( )
- DMAX( )
- ...

Syntax = the same for all!

- DSUM (**database**, field, criteria)
- DCOUNT (**database**, field, criteria)
- DAVERAGE (**database**, field, criteria)
- ...
- Database:** the data table, incl. headers.
  - Field:** what we want to sum, count...
  - Criteria:** the AND / OR conditions, incl. mathematical operators > , < , <> , and wildcards \* , ?

# Multiple AND criteria

DSUM (**database**, **field**, **criteria**)

Database						
1	A	B	C	D	E	F
	Order	Return?	Month	Shop	State	Sales
1	2435	No	July	Joe's	NY	\$ 1509
	6565	Yes	August	Wago	Florida	\$ 44
	765	No	August	Karo	Florida	\$ 1
	65732	No	July	Karo	Ohio	\$ 269
	1765	No	July	Karo	Florida	\$ 133

  

Criteria		
H	I	J
Return?	Shop	State
1		
Yes		Florida
2		

- DSUM(A1:F6, "Sales", H1:J2)
- DSUM(A1:F6, 6 , H1:J2)
- DSUM(A1:F6, F1 , H1:J2)

# AND / OR criteria!

DSUM (**database**, **field**, **criteria**)

Database						Criteria			
	A	B	C	D	E	F	H	I	J
1	Order	Return?	Month	Shop	State	Sales			
2	2435	No	July	Joe's	NY	\$ 1509			
3	6565	Yes	August	Wago	Florida	\$ 44			
4	765	No	August	Karo	Florida	\$ 1			
5	65732	No	July	Karo	Ohio	\$ 269			
6	1765	No	July	Karo	Florida	\$ 133			

- DSUM(A1:F6, "Sales", H1:J2) --> DSUM(A1:F6, "Sales", H1:J3)
- **Important!** Blank rows in criteria table equal no condition!

# Eager to practice?

ADVANCED EXCEL FUNCTIONS

# Summarizing data with database functions

ADVANCED EXCEL FUNCTIONS



**Agata Bak-Geerinck**  
Product Owner Data, Telenet

# **Let's practice!**

## **ADVANCED EXCEL FUNCTIONS**

# And it's a wrap!

ADVANCED EXCEL FUNCTIONS



**Agata Bak-Geerinck**  
Product Owner Data, Telenet

# Chapter 1

- The lookups and referencing functions

- `XLOOKUP( )`

- `INDEX ()` and `MATCH()`

- ... and their powerful combination!

- Using named ranges for efficiency

A table with 5 columns and 4 rows. The columns are labeled FEB, APR, MAY, and JUN at the top. The rows are labeled Category, Chairs, Labels, and Storage on the left. The first row (Category) is highlighted in purple. The last three rows (Chairs, Labels, Storage) are highlighted in orange. A blue box labeled "headers" points to the first row. An orange box labeled "an array" points to the last three rows.

Category	FEB	APR	MAY	JUN
Chairs	\$ 2,114	\$ 4,085	\$ 5,678	\$ 5,678
Labels	\$ 1,456	\$ 4,273	\$ 241	\$ 241
Storage	\$ 896	\$ 3,678	\$ 3,113	\$ 3,113

A table with 4 columns labeled A, B, C, and D. The rows are labeled 1, 2, 3, and 4 on the left. The first row (Category) is dark green. The second row (Chairs) is light green. The third row (Labels) is teal. The fourth row (Storage) is light blue. A yellow box highlights the cell in column C, row 3 (Labels). A blue box labeled "column 2" points to the second column. An orange box labeled "row 2" points to the second row. An orange box labeled "array B2 : D4" points to the range of cells from B2 to D4.

1	Category	FEB	APR	MAY
2	Chairs	\$ 2,114	\$ 4,085	\$ 5,678
3	Labels	\$ 1,456	\$ 4,273	\$ 241
4	Storage	\$ 896	\$ 3,678	\$ 3,113

# Chapter 2

- Building on our referencing skills

A	B	C	D	E	F	G	H	I
1				APR	MAY	JUN		
2				Sales	\$123	\$153	\$876	
3				Profit	\$54	\$12	\$345	
4								
5								

- New function: `OFFSET( )`

- ... and its powerful combination with `INDEX( )` and `MATCH( )` functions!

- Dynamic selections and ranges

A	B	C	D	E	F	G	H	
1	Starting Week	Week 2						
2		Week 2						
3		Week 3						
4								
5		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
6	Sales	\$ 84	\$ 24	\$ 34	\$ 28	\$ 34	\$ 46	\$ 138
7	Profit	\$ 50	\$ 12	\$ 20	\$ 25	\$ 20	\$ 41	\$ 110

# Chapter 3

- New family of functions: database functions

- DSUM( ), DCOUNT( ), DAVERAGE( ) ...

- Working efficiently with large datasets

- Aggregating and calculations based on AND/OR conditions

DSUM (**database**, field, criteria)

DCOUNT (**database**, field, criteria)

DAVERAGE (**database**, field, criteria)

...

Database						Criteria			
	A	B	C	D	E	F	H	I	J
1	Order	Return?	Month	Shop	State	Sales			
2	2435	No	July	Joe's	NY	\$ 1509			
3	6565	Yes	August	Wago	Florida	\$ 44			
4	765	No	August	Karo	Florida	\$ 1			
5	65732	No	July	Karo	Ohio	\$ 269			
6	1765	No	July	Karo	Florida	\$ 133			

# **Don't stop learning!**

**ADVANCED EXCEL FUNCTIONS**