dante

Digital Area for Networking
Teachers and Educators

Co-funded by
the European Union

Data Analysis in MS Excel

Introduction

WEEK 1

Vitezslav Novak

Table of Contents

Data Analysis in MS Excel

As the name of the course suggests, the course deals with data analysis in Excel. But under the term data analysis in Excel, you can imagine a lot of things that Excel can do, and the truth is that Excel can really do a lot, and the index of its tools continues to grow with each new version of Excel. Therefore, all its functions and tools will certainly not be described, but only functions and tools usable for aggregation calculations over data stored in tables of the list type. So, we will not deal with the absolute basics of Excel such as used data types, working with workbooks, sheets, cells, formatting tables, etc., etc. We will assume that you already have this knowledge.

As aggregation calculations I mean the using basic aggregation functions as they are described in the SQL standard, a standard used by relational databases whose tables are similar to tables of the list type. These are exactly the functions that you can find in Excel in the list of functions of the AUTOSUM button on HOME tab:

- SUM,
- COUNT,
- AVERAGE,
- MIN,
- MAX.

Anyone who can use the listed functions in Excel will certainly be able to apply their knowledge to more advanced statistical functions, such as the median, variance or standard deviation. But this course will not be about these more advanced functions. Actually, it will not even be about the five mentioned basic functions, it will be more about how to implement the mentioned five types of calculations in Excel. And we will show it here in particular using:

- formulas with different ways of addressing data,
- structured tables,

- pivot tables,
- the Power Pivot add-in, i.e. with the using data models, but we will really use the database terminology then.

What Is the List in Excel?

In this chapter, you will learn about tables, which are called **lists** in Excel. The following table describing the sales of several fictitious companies which will accompany us throughout the chapter and not only this chapter is an example of a list in Excel:

	A	B	C	D	E
1	Date	Company	Product	Amount	Paid
2	23.09.2017	IMPALA	X	10000	yes
3	24.09.2017	IMPALA	Z	9450	yes
4	25.09.2017	FORTISIMO	Y	11360	yes
5	26.09.2017	BRICKMAT	Y	3250	no
6	27.09.2017	BRICKMAT	Y	19580	no
7	28.09.2017	IMPALA	X	7650	no
8	29.09.2017	IMPALA	Y	17450	no
9	30.09.2017	FORTISIMO	Z	21500	no
10	01.10.2017	FORTISIMO	X	6540	no
11	02.10.2017	FORTISIMO	Z	3990	yes
12	03.10.2017	BRICKMAT	X	5700	no
13	04.10.2017	BRICKMAT	Z	770	no
14	05.10.2017	BRICKMAT	Y	4470	no
15	06.10.2017	IMPALA	Y	6200	no
16	07.10.2017	IMPALA	Z	2340	no
17	08.10.2017	ZETEN	Z	3500	yes
18	09.10.2017	ZETEN	X	4560	yes
19	10.10.2017	IMPALA	X	8560	yes
20	11.10.2017	FORTISIMO	X	9900	no

Figure 1 A table of list type

The list in Excel is a table that has the following properties:

- The first row of the list must contain the names of the fields (columns), the name of the field must be in one cell, for a more detailed description of the fields, use a comment instead.

- Field names should not be duplicated. Tools such as structured or pivot tables always perform calculations on entire columns, they use column names instead of cell addresses in formulas, so column names must be unique.
- In the next rows of the table are the individual items of the list.
- There should be only one list per sheet, it can start in any cell of the sheet.
- There must be no empty rows in the list. To work with a list, it is not necessary to select the entire list, just click in it and Excel will select it by itself as a continuous area of data around the active cell, therefore there must not be empty rows in the list.
- There must be data of only one data type in one field (column). Data validation can be used to ensure a uniform data type in the entire column.
- Place auxiliary data (e.g., for advanced filtering) above the list. If the auxiliary data were placed next to the list, and not above the list, in the case of list filtering, where Excel hides entire worksheet rows that do not match the filter, the auxiliary data would also be hidden.
- The list must not contain merged cells.

The table in the following figure is not a list because the header consists of two rows, but mainly because this header does not contain the names of the columns, but the values that we want to analyse. These values belong in columns, not headers.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		David					Jitka					Milan	
2		Apple	Carrot	Banana	Potato	Potato	Apple	Carrot	Banana	Potato	Potato	Apple	Carrot
3	1	222	226	332	269	195	239	161	360	232		288	176
4	2	306	231	173	310	92	201	262	32	174	258	138	95
5	3	261	193	353	131	256	181	154	148	301	204	179	157
6	4	149	202	253	271	161	147	122	154	356	236	193	247
7	5	180	313	166	272	89	471	525	236	220	153	98	97
8	6	207	283	258	164	200	263	238	240	186	249	129	286
9	7	149	185	257	216	184	298	240	305	325	209	87	146
10	8	95	191	246	219	368	255	314	247	279	293	298	108
11	9	369	138	203	199	89	180	221	247	117	429	277	225
12	10	173	182	224	216	209	174	253	159	223	255	208	269
13	11	189	138	242	207	268	228	207	185	183	180	177	149
14	12	108	126	133	180	247	363	122	311	324	130	219	37

Figure 2 A table that is not a list

Such a table cannot be used as a data source for a pivot table that calculates, for example, total sales for individual products. To be able to analyse such data using a pivot table, it had to be transformed into a list using Power Query e.g., to such a table:

	A	B	C	D
1	Dealer	Product	Month	Quantity
2	David	Apple	7	37
3	Pavel	Carrot	9	36
4	David	Banana	4	32
5	David	Tomato	1	20
6	Pavel	Apple	4	34
7	David	Tomato	11	22
8	Pavel	Banana	11	21
9	Pavel	Apple	8	23
10	Jitka	Banana	8	32
11	Jitka	Banana	1	37
12	Milan	Banana	10	39
13	David	Tomato	1	39
14	Jitka	Tomato	3	29
15	Pavel	Carrot	12	31
16	David	Potato	9	32
17	Pavel	Tomato	2	39
18	Jitka	Tomato	7	28

Figure 3 Transformed table (incomplete) from the table from Figure 2 into a list

From such a table, it is no problem to use a pivot table or other tools to calculate, for example, the total sales of individual products:

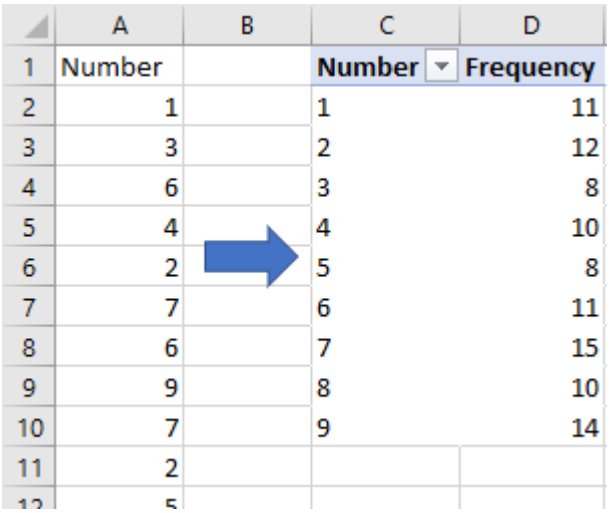
Product	Total Quantity
Apple	12016
Banana	13302
Carrot	12179
Potato	13628
Tomato	13426
Total Quantity	**64551**

Figure 4 A pivot table created based on the list in Figure 3

As you can see from the previous example, list are very important tables in Excel and we will learn how to process them using various Excel tools.

Introductory Exercise to Think About

Before you start studying working with lists in the next chapters, try to think about the following problem. Let's have a table with a single column called **Number** in the header and ninety-nine random numbers between 1 and 9 in the other rows. The number of rows 99 was chosen so that the table has 100 rows including the header and all formulas therefore have the same parameters for easier checking of your solution with the solution in the course. However, the number of lines does not matter.



	A	B	C	D
1	Number		Number	Frequency
2	1		1	11
3	3		2	12
4	6		3	8
5	4		4	10
6	2		5	8
7	7		6	11
8	6		7	15
9	9		8	10
10	7		9	14
11	2			

Figure 5 Calculation of frequencies from the list of numbers

In Excel, try to think of as many ways as possible how to find out the frequency of individual numbers 1 to 9 in the Number column.

So, how many ways of calculating frequencies have you come up with? If, after reading each chapter, you feel that the chapter shows one or more ways to calculate this frequencies, and you manage this calculation using the knowledge gained in this chapter, then this will mean that you have understood the chapter. And if you can't do it yourself, then at the end of the course, the last chapter will be devoted to the solution of this exercise.

dante

Digital Area for Networking Teachers and Educators

Project Number: 2020-1-CZ01-KA226-HE-094368

Háskólinn á Akureyri University of Akureyri

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | EKONOMICKÁ FAKULTA

TECHNICAL UNIVERSITY OF KOŠICE

Uniwersytet Ekonomiczny w Katowicach

POLITÉCNICO DE SANTARÉM