Data Analysis: A Practical Introduction for Absolute Beginners

Module 5, Lab 1: Healthcare Data

Learning Objectives

- Use the AVERAGE function to find the mean/average of a single variable in a data set.
- Use the AVERAGEIF function to find the mean/average of one variable based on certain criteria in a second variable.
- Create a bar graph to visualize data.

Data Set

Mod5Lab1.csv

What You'll Need

To complete the lab, you will need the online version of Microsoft Excel.

Overview

Imagine that you're a data analyst who's been hired by a small medical practice to analyze the efficiency of four different doctors. The data set in question shows the morning schedules of these four doctors — Jessica, Fatima, Haruto, and Benjamin — over the course of a single week, as well as the duration of their appointments with each patient.

In this lab, your job is to analyze each doctor's average time per patient, decide which doctor is the most efficient in terms of time, and create a visual aid to show these averages.

Exercise 1: Total Average Duration

1. Open the data set in Excel, which shows the appointment times and durations for the patients of four different doctors. Here's a snapshot of the data:

	A	B	C	D	E
1		day	time	provider	duration (min)
2	1	M	8	Jessica	17
3	2	T	8	Jessica	14.8
4	3	W	8	Jessica	13.3
5	4	R	8	Jessica	15.1
6	5	F	8	Jessica	20.9
7	6	M	8:30	Jessica	15.3
8	7	T	8:30	Jessica	14.4
9	8	W	8:30	Jessica	18.9
10	9	R	8:30	Jessica	18.5
11	10	F	8:30	Jessica	14.5

Each row represents one patient (there should be 160 different patients). Here's what each column represents:

day = the day of the week (note: T = Tuesday, and R = Thursday)
time = the appointment time (a.m.)
provider = the doctor's name (either Jessica, Fatima, Haruto, or Benjamin)
duration (min) = the length of the appointment, in minutes

The untitled column A values give the patient number for ID purposes. Notice that these patient numbers are slightly different than the actual row numbers — so Patient 1 is actually in row 2 of the spreadsheet, because all the column titles are in row 1.

2. Start out by creating a new mini-table off to the side of the data (in the same spreadsheet), with columns for each provider, the total average appointment duration from that provider, and the average appointment from that provider for each day of the week (M, T, W, R, F). It should look something like this:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		day	time	provider	duration (min)		**provider**	**total avg**	**M avg**	**T avg**	**W avg**	**R avg**	**F avg**
2	1	M	8	Jessica	17		Jessica						
3	2	T	8	Jessica	14.8		Fatima						
4	3	W	8	Jessica	13.3		Haruto						
5	4	R	8	Jessica	15.1		Benjamin						
6	5	F	8	Jessica	20.9								

3. Next, click into cell H2 and find Jessica's average appointment duration using the AVERAGE function in Excel. The syntax here is **=AVERAGE(first cell:last cell)**. The duration data are in column E.

The first 40 patients were Jessica's, so your first cell is E2 and your last cell is E41. (Remember, the cell numbers are 1 digit off from the patient number, so Patient 1 is in row 2.) You can either type those cells directly into the AVERAGE function (with a colon in between), or just type in

=AVERAGE( ), click inside the parentheses, and highlight the whole range of cells between E2 and E41. Either way is fine.

Hit Enter to run the calculation.

G	H	I	J	K	L	M
provider	**total avg**	**M avg**	**T avg**	**W avg**	**R avg**	**F avg**
Jessica	16.68					
Fatima						
Haruto						
Benjamin						

Boom: The average amount of time that Jessica spent with each patient was 16.68 minutes.

4. Now find the total average for the second doctor, Fatima. Click into cell H3 and use the AVERAGE function again. You'll still use the duration data from column E, but this time, you want Fatima's patients, who run from cell E42 down to E81.

Hit Enter.

G	H	I	J	K	L	M
provider	**total avg**	**M avg**	**T avg**	**W avg**	**R avg**	**F avg**
Jessica	16.68					
Fatima	12.42					
Haruto						
Benjamin						

Fatima spent an average of 12.42 minutes per patient.

5. Haruto's up next. Click into cell H4 and use the AVERAGE function on Haruto's patients, who run from E82 to E121.

	G	H	I	J	K	L	M
	provider	total avg	M avg	T avg	W avg	R avg	F avg
	Jessica	16.68					
	Fatima	12.42					
	Haruto	9.325					
	Benjamin						

Haruto was really hustling, with an average duration of 9.325 minutes per patient.

6. Finish up with Benjamin's average in cell H5. Once again, use the AVERAGE function. Benjamin's patients run from E122 all the way down to E161.

	G	H	I	J	K	L	M
	provider	total avg	M avg	T avg	W avg	R avg	F avg
	Jessica	16.68					
	Fatima	12.42					
	Haruto	9.325					
	Benjamin	10.96					

Benjamin spent an average of 10.96 minutes with each patient.

Exercise 2: Daily Average Duration

Now we'll track down each doctor's average duration by the day of the week.

1. Using that same new table you created in Exercise 1, click into cell I2, which you'll use for Jessica's average duration per patient *on Monday only*. This might sound tricky, but Excel Online has another fancy function that'll do it nicely: AVERAGEIF, which finds the average based on certain criteria in the data.

   The syntax is slightly more complicated for this function:

   =AVERAGEIF(criteria range, criteria, average range)

   Here's what each part means:

   criteria range = the range of cells to test with the criteria, with a colon between them (first cell:last cell)

criteria = the condition you want to use to narrow down the variable in the criteria range (if the criteria involves text instead of numbers, stick the text in quotation marks)
average range = the range of cells that you actually want to average, with a colon between them (first cell:last cell)

For example, in cell I2 you want to find Jessica's average duration on Monday only, so the criteria range is the "day" variable (column B, but only the cells from Jessica's patients). The criteria is Monday ("M"). The average range is the "duration" variable (column E, but again, only the cells from Jessica's patients). Put it all together, and you get:

Don't forget to put the "M" in quotation marks, since it's text instead of a number. Once you hit Enter, Excel will calculate the average duration from Jessica's patients, but only the ones she saw on Monday.

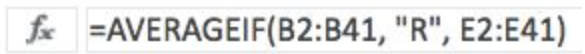
G	H	I	J	K	L	M
provider	**total avg**	**M avg**	**T avg**	**W avg**	**R avg**	**F avg**
Jessica	16.68	15.575				
Fatima	12.42					
Haruto	9.325					
Benjamin	10.96					

Beautiful. On Monday, Jessica spent an average of 15.575 minutes with each patient.

2. Now use this same formula for the other days of the week for Jessica. The AVERAGEIF function will stay nearly the same for Jessica's other days: Her criteria range is always B2:B41, and her average range is always E2:E41. The only thing that changes is the criteria itself: Tuesday = "T" instead of "M." This'll go in the "T avg" column (cell J2).

Wednesday is "W." This goes in cell K2.

Thursday is "R." This goes in cell L2.



And Friday is “F,” which goes in cell M2.

Once you’ve entered all the formulas correctly, here’s what you should get for Jessica’s daily averages:

G	H	I	J	K	L	M
provider	**total avg**	**M avg**	**T avg**	**W avg**	**R avg**	**F avg**
Jessica	16.68	15.575	17.175	16.1375	18.6875	15.825
Fatima	12.42					
Haruto	9.325					
Benjamin	10.96					

3. Repeat Steps 1 and 2 for Fatima. This time, Fatima’s criteria range runs from B42:B81, and her average range runs from E42:E81.

   Monday (this goes in cell I3 for Fatima):

   Tuesday (this goes in cell J3):

   Wednesday (this goes in cell K3):

   Thursday (this goes in cell L3):

   And Friday (this goes in cell M3):

f_x =AVERAGEIF(B42:B81, "F", E42:E81)

With all the formulas entered, your table should now look like this:

G	H	I	J	K	L	M
provider	**total avg**	**M avg**	**T avg**	**W avg**	**R avg**	**F avg**
Jessica	16.68	15.575	17.175	16.1375	18.6875	15.825
Fatima	12.42	10.1	16.3	12.525	10.7	12.475
Haruto	9.325					
Benjamin	10.96					

4. Repeat those steps again for Haruto. The criteria range runs from B82:B121, and the average range runs from E82:E121.

   Monday (this goes in cell I4):

   f_x =AVERAGEIF(B82:B121, "M", E82:E121)

   Tuesday (this goes in cell J4):

   f_x =AVERAGEIF(B82:B121, "T", E82:E121)

   Wednesday (this goes in cell K4):

   f_x =AVERAGEIF(B82:B121, "W", E82:E121)

   Thursday (this goes in cell L4):

   f_x =AVERAGEIF(B82:B121, "R", E82:E121)

   Friday (this goes in cell M4):

   f_x =AVERAGEIF(B82:B121, "F", E82:E121)

   With Haruto's stuff all entered, your table should look like this:

G	H	I	J	K	L	M
provider	**total avg**	**M avg**	**T avg**	**W avg**	**R avg**	**F avg**
Jessica	16.68	15.575	17.175	16.1375	18.6875	15.825
Fatima	12.42	10.1	16.3	12.525	10.7	12.475
Haruto	9.325	7.1875	10.6625	8.1	10.2125	10.4625
Benjamin	10.96					

5. Repeat those steps one last time for Benjamin. Don't give up! Almost there! Benjamin's criteria range is B122:B161, and his average range is E122:E161.

   Here's Monday (cell I5):

   f_x =AVERAGEIF(B122:B161, "M", E122:E161)

   Here's Tuesday (cell J5):

   f_x =AVERAGEIF(B122:B161, "T", E122:E161)

   Here's Wednesday (cell K5):

   f_x =AVERAGEIF(B122:B161, "W", E122:E161)

   Here's Thursday (cell L5):

   f_x =AVERAGEIF(B122:B161, "R", E122:E161)

   And finally, at long last, here's Friday (cell M5):

   f_x =AVERAGEIF(B122:B161, "F", E122:E161)

   Our table is finally done!

G	H	I	J	K	L	M
provider	total avg	M avg	T avg	W avg	R avg	F avg
Jessica	16.68	15.575	17.175	16.1375	18.6875	15.825
Fatima	12.42	10.1	16.3	12.525	10.7	12.475
Haruto	9.325	7.1875	10.6625	8.1	10.2125	10.4625
Benjamin	10.96	10.75	10.475	9.525	11.5375	12.5125

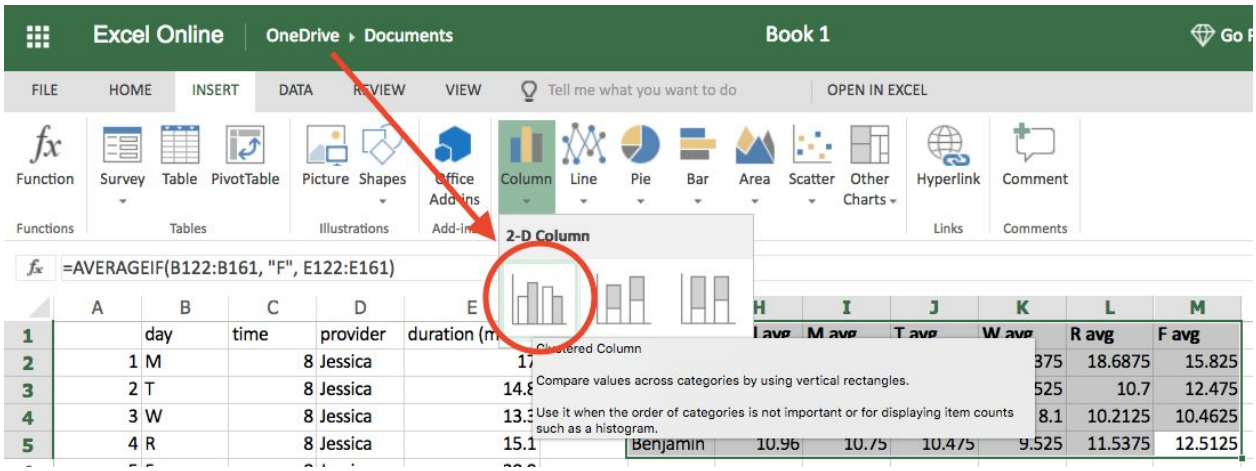
Now you can see how everyone's weekly and daily average times stack up against each other. But a visual aid will make these comparisons even easier to see.

Exercise 3: Graph It!

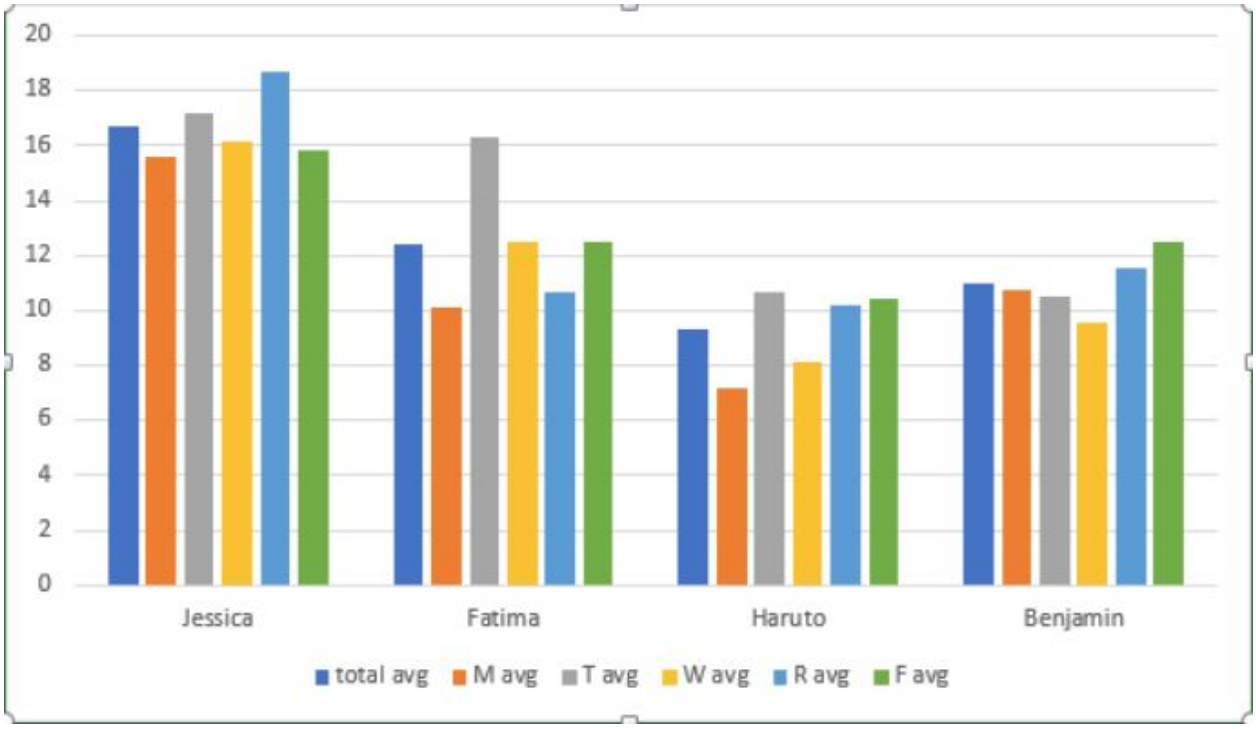
1. It's visual aid time. Highlight the entire data table you created in the last two exercises. (But do *not* highlight the entire spreadsheet — just the stuff on each doctor's weekly and daily averages, from columns G to M.)

G	H	I	J	K	L	M
provider	total avg	M avg	T avg	W avg	R avg	F avg
Jessica	16.68	15.575	17.175	16.1375	18.6875	15.825
Fatima	12.42	10.1	16.3	12.525	10.7	12.475
Haruto	9.325	7.1875	10.6625	8.1	10.2125	10.4625
Benjamin	10.96	10.75	10.475	9.525	11.5375	12.5125

2. With those data highlighted, click Insert > Column > Clustered Column (it's usually the icon on the left under 2-D Column).



3. A fancy new bar graph should pop up, like so:



Looking good, right? Now it's very easy to see how our four doctors compare to each other in terms of their weekly average time with each patient (the blue bar) and their average time per day (the other bars).

From the graph, you can see that Haruto is the most efficient doctor in terms of time: He runs through his patients more quickly, on average, than the other doctors.

Jessica, on the other hand, spends the most time with each patient, which might also be a good thing. Maybe she's the most careful and methodical doctor.

Benjamin appears to be the most consistent in his daily average durations: The amount of time he spends with each patient varies less by day than the other docs.