

# Task 2: Step-by-step guide

***Note:** This step-by-step guide has been prepared on Microsoft Excel. You can also do this task on Google Sheets (uploading the .xlsx file and opening in Google Sheets), and the steps below will be very similar.*

# Let's get started

Download and open the Healthy Co Data.xlsx file (and the optional Step-by-Step Guide)

Review each column heading to orient yourself to the data fields and what each is showing

Unique ID	Year	Arrival (Hour of day)	Check-in (mins)	Wait for vitals (mins)	Vitals (mins)	Wait for doctor (mins)	Doctor visit (mins)	Payment (mins)	Diagnosis group
476	2019		11	12	3	13	8.1	2	Group 3
902	2019		20	2	5.75	13	7.2	6	Group 3
1236	2019		9	3	24.7	4.6	23.4	6	Group 2
1925	2019		11	2	14	5.75	18	3.6	Group 3
1999	2019		13	3	16.5	4.2	22	7.2	Group 3
2032	2019		12	3.45	12.5	5.25	27.5	6.3	Group 3
2088	2019		13	2	11	5.7	19.8	8	Group 2
2398	2019		10	2	31.9	7.6	27.5	6	Group 1
2404	2019		16	5	12.75	4.75	18.7	10.8	Group 1
2409	2019		19	2	21.25	8.4	17.5	8.4	Group 1
2434	2019		15	2	11.7	4	8.1	7	Group 2
2553	2019		19	2	5	4.6	17.5	5.4	Group 3
2864	2019		18	5	8.4	5.75	34.8	7	Group 2
2868	2019		17	5	16.15	6	20.9	5.4	Group 3
3026	2019		9	2	19.5	5.95	35.49	4	Group 2
3184	2019		9	3	24.7	3.4	22.1	4.5	Group 3
3242	2019		12	5.75	16.25	3.3	20	4.5	Group 3
3331	2019		9	2	16.9	5.5	24.7	3.6	Group 3
3399	2019		18	5	16.8	5.75	27.6	5.4	Group 3
3454	2019		20	2	14	8.8	18	4.5	Group 3
3504	2019		9	3	37.7	6.6	20.8	3.6	Group 3
4742	2019		10	5	5.5	4.2	16.5	4.5	Group 3
4883	2019		14	4	23.75	4.6	23.75	7	Group 2
4962	2019		12	5	5	6.9	28.75	6.3	Group 3
5133	2019		20	5	15	5.25	17	8	Group 2
5307	2019		14	2	10.45	6	23.75	7.2	Group 3
5475	2019		15	4	25.2	4.6	30.42	6.3	Group 3
5536	2019		12	1	15	6	22.5	6	Group 2
5592	2019		13	2.3	25.3	5.75	28.6	5	Group 2
5632	2019		18	4	14.4	5.75	34.8	6	Group 2
5792	2019		17	5	12.35	5.75	24.7	5.4	Group 3
5849	2019		14	5	9.5	4.6	16.15	4.8	Group 1
6474	2019		16	2	15.3	4	13.6	6.3	Group 3
7034	2019		13	2	14.3	5.25	22	5.4	Group 3
7248	2019		10	5	18.7	4	31.46	7	Group 2
7348	2019		15	2	11.7	5	22.5	3.6	Group 3
7411	2019		10	4.6	22	5.75	28.6	5.4	Group 3
7414	2019		18	5	18	4.25	20.4	5.4	Group 3
7748	2019		12	2	13.75	4	32.5	7	Group 2

# Remove duplicate data

We need to make sure each row is unique, so if there are identical unique IDs, they need to be removed.

- Highlight column B (Unique ID) by clicking on “B” at the top of the spreadsheet
- Click (1) Data → (2) Click Remove Duplicates
- Make sure “Expand the selection” is selected and click “Remove Duplicates”
- Make sure all columns are selected and click “OK”
- You will get an alert that confirms duplicates have been removed

AutoSave OFF | Healthy Co\_Data — Saved to my Mac

Home Insert Draw Page Layout Formulas **Data** Review View Developer Tell me

Get Data (Power Query) Data from Picture Refresh All Queries & Connections Properties Edit Links Stocks Currencies Geography Automatic Sort Filter Clear Reapply Advanced Text to Columns What's This?

**Remove Duplicates Warning**

Microsoft Excel found data next to your selection. Because you have not selected this data, it will not be removed.

**What do you want to do?**

☒ Expand the selection

☐ Continue with the current selection

Cancel Remove Duplicates...

Unique ID	Year	Arrival (Hour of day)	Check-in (mins)	Wait for vitals (mins)	Payment (mins)	Diagnosis group
476	2019	11	1	12	8.1	Group 3
902	2019	20	2	10	7.2	Group 3
1236	2019	9	3	24.7	6	Group 2
1925	2019	11	2	14	3.6	Group 3
1999	2019	13	3	16.5	7.2	Group 3
2032	2019	12	3.45	12.5	6.3	Group 3
2088	2019	13	2	11	8	Group 2
2398	2019	10	2	31.9	6	Group 1
2404	2019	16	5	12.75	0.8	Group 1
2409	2019	19	2	21.25	12	Group 1
2434	2019	15	2	11.7	1	Group 1
2553	2019	19	2	5	4	Group 2
2864	2019	18	5	8.4	5.4	Group 3
2868	2019	17	5	16.15	7	Group 2
3026	2019	9	2	19.5	5.4	Group 3
3026	2019	9	2	19.5	5.95	Group 2
3184	2019	9	3	24.7	4	Group 2
3242	2019	12	5.75	16.25	3.4	Group 3
3331	2019	9	2	16.9	22.1	Group 3
3399	2019	18	5	16.8	3.3	Group 3
3454	2019	20	2	14	24.7	Group 3
3504	2019	9	3	37.7	5.4	Group 3
4742	2019	10	5	4.2	27.6	Group 3
4883	2019	14	4	23.75	18	Group 3
4962	2019	12	5	6.9	4.5	Group 3
5133	2019	20	5	15	7	Group 2
5307	2019	14	2	10.45	6.3	Group 3
5475	2019	15	4	25.2	8	Group 2
5536	2019	12	1	15	7.2	Group 3
5592	2019	13	2.3	25.3	6.3	Group 3
5632	2019	18	4	14.4	6	Group 2
5792	2019	17	5	12.35	5.4	Group 2
5849	2019	14	5	9.5	4.8	Group 1
6474	2019	16	2	15.3	6.3	Group 3
7034	2019	13	2	14.3	5.4	Group 3
7248	2019	10	5	18.7	7	Group 2
7348	2019	15	2	11.7	3.6	Group 3
7411	2019	10	4.6	22	5.4	Group 3
7414	2019	18	5	18	5.4	Group 3
7748	2019	12	2	13.75	7	Group 2

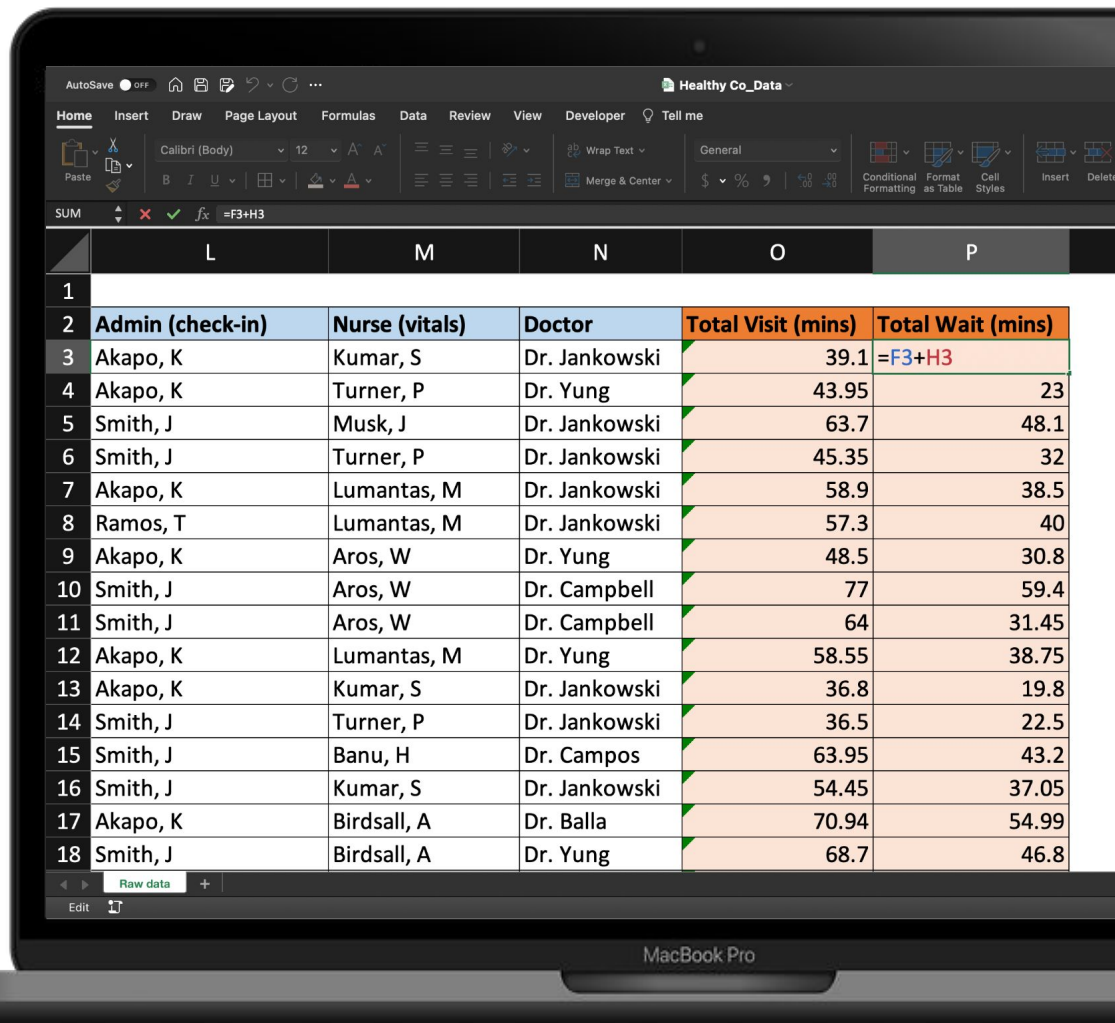
Ready | Average: 503073.3094 | Count: 5008 | Sum: 2518888060

MacBook Pro

# Create Totals columns

We need to create a column that calculates the total visit duration for each visit.


- Go to column O and give it a title in cell O2 - call it "Total Visit (mins)"
- Calculate the sum of all duration columns (E to J) in cell O3
  - Type `=sum(E3:J3)` in cell O3 and press enter
- Copy this formula for each row
  - Hover the mouse over the bottom-right corner of O3 and double click the black **+**
    - Ensure that each formula has the correct row referenced, e.g., in O5 it should read `sum(E5:J5)`
- Repeat for Total Wait (mins), except this time only add the wait columns
  - `=F3+H3` in P3 and copy down



	L	M	N	O	P
1					
2	Admin (check-in)	Nurse (vitals)	Doctor	Total Visit (mins)	Total Wait (mins)
3	Akapo, K	Kumar, S	Dr. Jankowski	39.1	=F3+H3
4	Akapo, K	Turner, P	Dr. Yung	43.95	23
5	Smith, J	Musk, J	Dr. Jankowski	63.7	48.1
6	Smith, J	Turner, P	Dr. Jankowski	45.35	32
7	Akapo, K	Lumantas, M	Dr. Jankowski	58.9	38.5
8	Ramos, T	Lumantas, M	Dr. Jankowski	57.3	40
9	Akapo, K	Aros, W	Dr. Yung	48.5	30.8
10	Smith, J	Aros, W	Dr. Campbell	77	59.4
11	Smith, J	Aros, W	Dr. Campbell	64	31.45
12	Akapo, K	Lumantas, M	Dr. Yung	58.55	38.75
13	Akapo, K	Kumar, S	Dr. Jankowski	36.8	19.8
14	Smith, J	Turner, P	Dr. Jankowski	36.5	22.5
15	Smith, J	Banu, H	Dr. Campos	63.95	43.2
16	Smith, J	Kumar, S	Dr. Jankowski	54.45	37.05
17	Akapo, K	Birdsall, A	Dr. Balla	70.94	54.99
18	Smith, J	Birdsall, A	Dr. Yung	68.7	46.8

## Create wait cost columns

We need to create a column that calculates the vital and doctor wait cost of each visit.

- Add 3 rows at the top of the table to make room for a *costs per wait type* table
- Create a “Vital wait cost” title in cell Q5 - call it “Total Visit (mins)”
- Calculate the cost of the Vital wait in cell Q6
  - Type “=F6\*\$R\$2” in cell Q6 and press enter (the \$ before the row and cell reference locks this cell so it is always referenced, even when you copy your formula down)
- Copy this formula for each row
  - Hover the mouse over the bottom-right corner of Q6 and double click the black 
  - Ensure that each formula has the correct row referenced, e.g., in Q9 it should read “F9\*\$R\$2”
- Repeat for Doctor wait cost, except this time you’ll use column I and cost cell \$R\$3

**Healthy Co\_Model Answer**

Home Insert Draw Page Layout Formulas Data Review View Developer Tell me

Paste Calibri (Body) 12 B I U Merge & Center \$ % , .00 →0 Conditional Formatting

Q6 =F6\*\$R\$2

	L	M	N	O	P	Q	R
1							
2	<b>Vital wait cost (\$/min)</b>						\$10
3	<b>Doctor wait cost (\$/min)</b>						\$50
4							
5	<b>Admin (check-in)</b>	<b>Nurse (vitals)</b>	<b>Doctor</b>	<b>Total Visit (mins)</b>	<b>Total Wait (mins)</b>	<b>Vital wait cost</b>	<b>Doctor wait cost</b>
6	Akapo, K	Kumar, S	Dr. Jankowski	39.1	25	\$120	\$650
7	Akapo, K	Turner, P	Dr. Yung	43.95	23	\$100	\$650
8	Smith, J	Musk, J	Dr. Jankowski	63.7	48.1	\$247	\$1,170
9	Smith, J	Turner, P	Dr. Jankowski	45.35	32	\$140	\$900
10	Akapo, K	Lumantas, M	Dr. Jankowski	58.9	38.5	\$165	\$1,100
11	Ramos, T	Lumantas, M	Dr. Jankowski	57.3	40	\$125	\$1,375
12	Akapo, K	Aros, W	Dr. Yung	48.5	30.8	\$110	\$990
13	Smith, J	Aros, W	Dr. Campbell	77	59.4	\$319	\$1,375
14	Smith, J	Aros, W	Dr. Campbell	64	31.45	\$128	\$935
15	Akapo, K	Lumantas, M	Dr. Yung	58.55	38.75	\$213	\$875
16	Akapo, K	Kumar, S	Dr. Jankowski	36.8	19.8	\$117	\$405
17	Smith, J	Turner, P	Dr. Jankowski	36.5	22.5	\$50	\$875
18	Smith, J	Banu, H	Dr. Campos	63.95	43.2	\$84	\$1,740
19	Smith, J	Kumar, S	Dr. Jankowski	54.45	37.05	\$162	\$1,045
20	Akapo, K	Birdsall, A	Dr. Balla	70.94	54.99	\$195	\$1,775
21	Smith, J	Birdsall, A	Dr. Yung	68.7	46.8	\$247	\$1,105
22	Ramos, T	Knight, F	Dr. Yung	52.1	36.25	\$163	\$1,000
23	Akapo, K	Knight, F	Dr. Jankowski	55.7	41.6	\$169	\$1,235
24	Smith, J	Turner, P	Dr. Campos	61.55	44.4	\$168	\$1,380
25	Smith, J	Knight, F	Dr. Campbell	52.3	32	\$140	\$900
26	Akapo, K	Knight, F	Dr. Campos	83.7	58.5	\$377	\$1,040

MacBook Pro

# Pulling out the insights

You are about to create pivot tables to find the answers to a number of questions. You should return to the Raw Data worksheet and create a new pivot table for each question.

Once you have created a new worksheet, stay organized by renaming each sheet with a question-related title (double-click the worksheet name in the bottom-left to rename).

Repeat the instructions on the next 3 slides to answer each of the following:

- a. Does the hour of the day in which patients arrive have an impact on total wait time per patient?
- b. Does the time taken to complete the vitals (ignoring the wait) vary by nurse?
- c. Does the check-in duration vary by diagnosis group?
- d. Does the total time spent vary by diagnosis group?
- e. Does the time between vitals being completed and the doctor visit starting vary by doctor?
- f. Which nurse is losing the most revenue for the company due to their average wait times?
- g. Which doctor is losing the most revenue for the company due to their average wait times?

There will be further instruction provided to help you answer the last question:

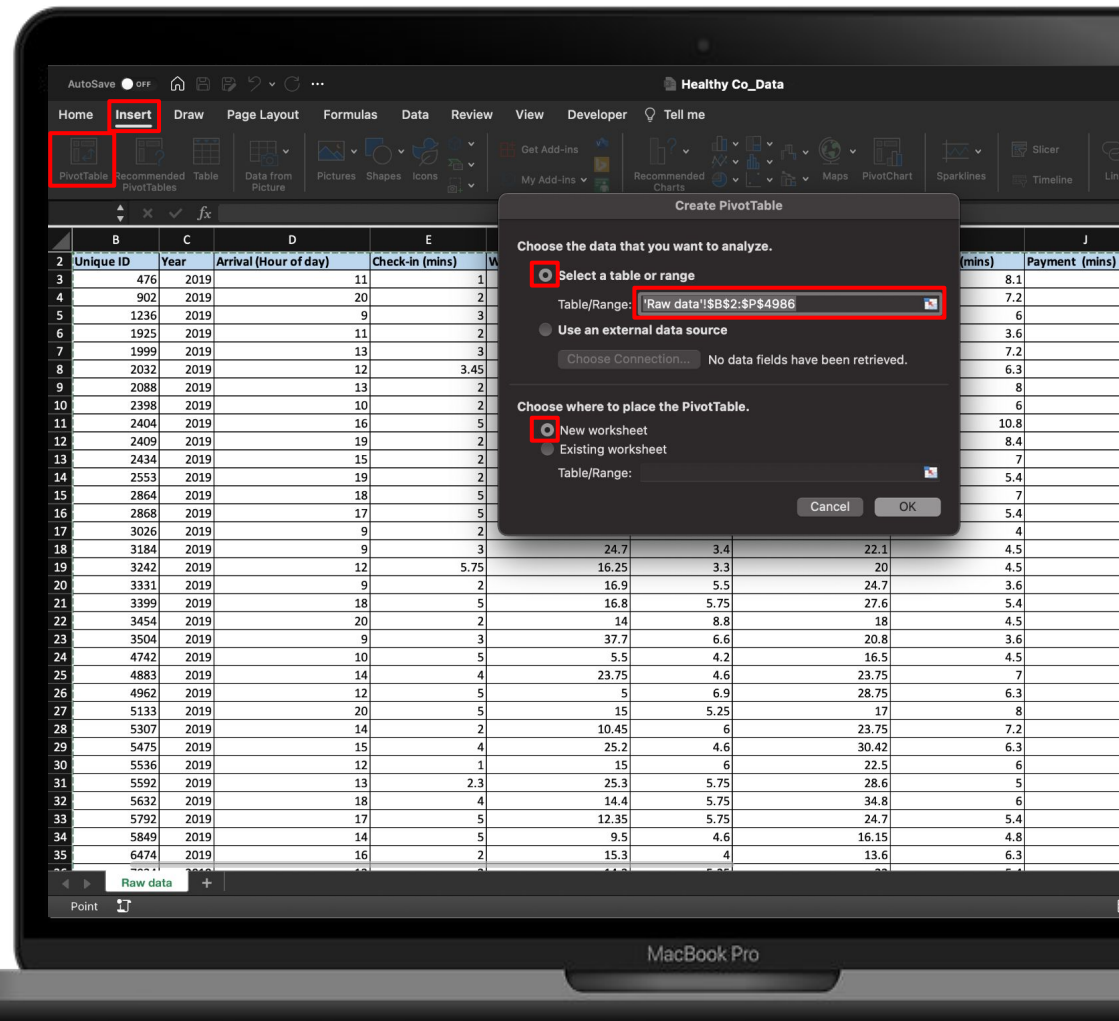
- h. Does each doctor spend a similar amount of time with each patient? Does this vary by diagnosis group?

# Create a pivot table

A pivot table speeds up data analysis significantly

- Click on Insert → Pivot Table
- Ensure “Select a table or range” is selected, and by default Table/Range should say the following:
  - 'Raw data'!\$B\$2:\$P\$4986
- Make sure “New worksheet” is selected, and click “OK”

**Note:** you will repeat this step to create pivot tables for each of the questions you are asked



# Use your pivot table

Does the hour of day in which patients arrive have an impact on total wait time per patient?

- Click in the area on the new sheet that says “Click in this area to work with the pivot table report”
- In the PivotTable Fields list, drag “Arrival (Hour of day)” into the “Rows” box (each unique value will have its own row)
- Find “Total Wait (mins)” in the list, and drag it into the “Values” box

This now shows the sum of total wait times by hour of the day, but we need the average wait time per patient by hour of the day, so:

- Click the (i) next to “Sum of total...” in the “Values” box
- Change the “Summarize by” option to “Average” and hit OK
- The numbers become more readable if you click and drag to highlight them all, then click the “Decrease Decimal” button in the “Home” tab until you are left with just 1 decimal

**Note:** you will repeat this step for each of the questions you are asked

The screenshot shows a Microsoft Excel spreadsheet on a MacBook Pro. The spreadsheet is titled 'Healthy Co\_Data' and is on 'Sheet9'. The pivot table is located in the range B4:B16. The pivot table has 'Arrival (Hour of day)' in the Rows field and 'Average of Total Wait (mins)' in the Values field. The 'Decrease Decimal' button in the Home tab is highlighted. The task pane on the right shows the 'PivotTable Fields' list with 'Arrival (Hour of day)' in the Rows field and 'Average of Total Wait (mins)' in the Values field. The 'Decrease Decimal' button in the task pane is also highlighted.

Arrival (Hour of day)	Average of Total Wait (mins)
9	47.6
10	40.1
11	37.4
12	48.3
13	40.7
14	34.9
15	33.4
16	31.5
17	35.5
18	43.7
19	46.4
20	36.7
Grand Total	39.7

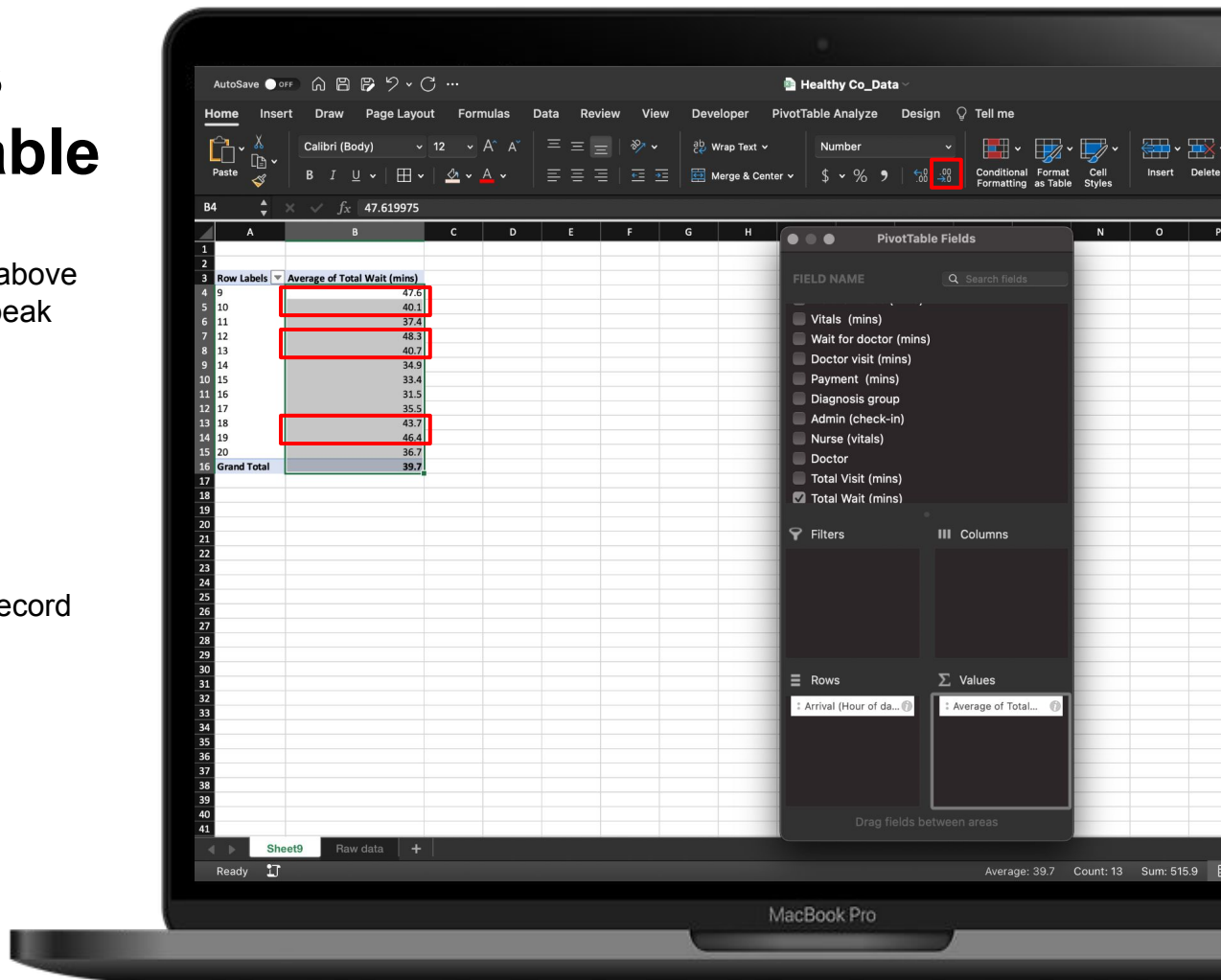


# Generate insights from your pivot table

The data shows that the wait is higher (above 40 mins) for arrivals in three groups of peak hours:

- **Morning:** 9 & 10 am
- **Lunch:** 12 & 1 pm
- **Evening:** 6 & 7 pm

**Tip:** Create an “Insights” worksheet to record the answer to each question



# Visualize insights with a column chart

Highlight the data (A3:B15), click Insert at the top, then look for the column option in the top ribbon. Then choose your graph type.

You can use these graphs when you put a presentation together in Task 4.

**Tip:** Edit the look and feel of your graph! Color specific columns to highlight them. Double click the column and look for the Fill color option.

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable. The PivotTable has 'Row Labels' in column A and 'Average of Total Wait (mins)' in column B. The data is as follows:

Row Labels	Average of Total Wait (mins)
9	47.6
10	40.1
11	37.4
12	48.3
13	40.7
14	34.9
15	33.4
16	31.5
17	35.5
18	43.7
19	46.4
20	36.7
Grand Total	39.7

The Excel ribbon is set to 'Insert', and the 'Charts' group is open, showing various chart options. The '2-D Column' chart type is selected, and the 'Clustered Bar' option is highlighted. The status bar at the bottom shows 'Average: 27.09400627' and 'Count: 26'.

# Using your pivot table: multiple criteria

Each of the previous questions were straightforward; essentially, what is the average of all values in a column (e.g., wait time) for each of the unique values in another column (e.g., arrival hour of day). In pivot table set-up terms, we've only used the "Rows."

Pivot tables become really useful when finding the average of all values in a column (e.g., doctor visit duration) based on a combination of unique values in multiple columns (e.g., diagnostic group AND doctor). We are about to set up a pivot table with both "Rows" and "Columns."

Being able to pivot this data will allow you to see the average doctor visit duration in minutes for the following:

- Dr. Campbell visiting with Diagnostic group 1
- Dr. Campbell visiting with Diagnostic group 2
- Dr. Campbell visiting with Diagnostic group 3
- Dr. Campos visiting with Diagnostic group 1
- Dr. Campos visiting with Diagnostic group 2
- etc.

# Use your pivot table

Does each doctor spend a similar amount of time with each patient? Does this vary by diagnosis group?

- In the PivotTable Fields list, drag “Doctor” into the “Rows” box
- Drag “Diagnosis” into the “Columns” box
- Find “Doctor Visit (mins)” in the “Values” box, and change it from Sum to Average if needed

Each cell in the pivot table is now a unique combination. E.g., for Dr. Campos visiting with a patient in Group 1, we look at cell B7 and see 10.0 mins.

We can still see the average doctor visit duration of all patients, regardless of diagnosis group, separated by doctor, in column E, and we can see the average doctor visit duration of each diagnosis group, regardless of doctor, in row 10.

The screenshot shows a MacBook Pro screen displaying a Microsoft Excel spreadsheet. The spreadsheet has a PivotTable with the following data:

Row Labels	Group 1	Group 2	Group 3	Grand Total
Dr. Balla	14.2	6.8	5.4	6.8
Dr. Campbell	10.4	6.9	5.3	6.4
Dr. Campos	10.0	6.9	5.3	6.3
Dr. Jankowski	10.9	6.7	5.3	6.4
Dr. Yung	10.5	6.7	5.4	6.3
Grand Total	11.2	6.8	5.3	6.4

The PivotTable Fields task pane is open on the right side of the screen. It shows the following fields:

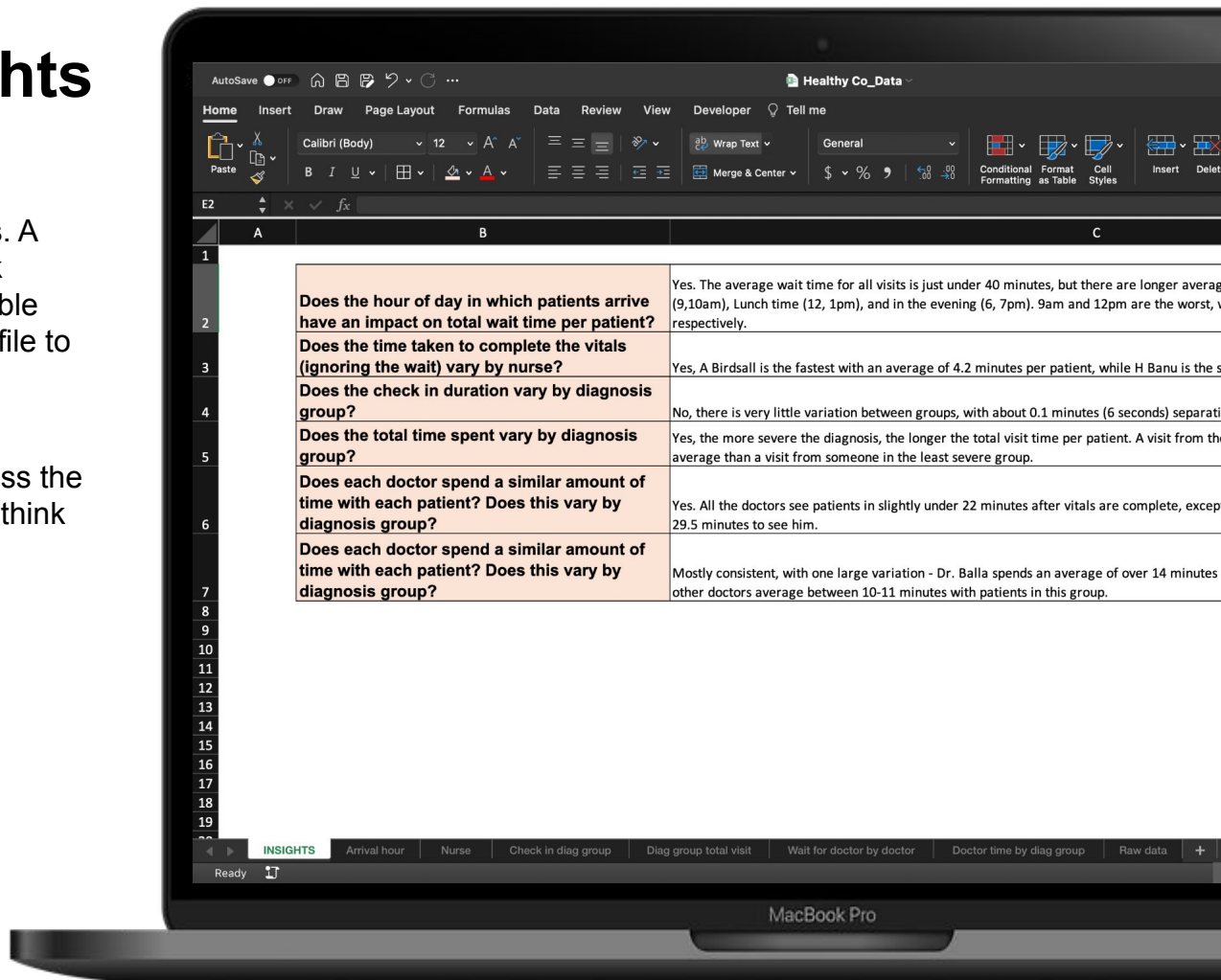
- FIELD NAME**: Search fields
- Filters**: None
- Columns**: Diagnosis group
- Rows**: Doctor
- Values**: Average of Doctor Visit (mins)

The task pane also includes a section for "Drag fields between areas" at the bottom.

# Consolidate insights

Create a sheet to organize your insights. A simple table with the question and quick summary of the insight from the pivot table helps to orient the people you send the file to and get them up to speed quickly.

It also helps you to step back and process the answers to your work, and helps you to think about the so-what of the task.



# Cheat sheet for model answer

Set-up for each pivot table

Question	Row	Column	Value
Does the hour of the day in which patients arrive have an impact on total wait time per patient?	Arrival (Hour of day)	-	[AVERAGE] Total Wait (mins)
Does the time taken to complete the vitals (ignoring the wait) vary by nurse?	Nurse (Vitals)	-	[AVERAGE] Vitals (mins)
Does the check-in duration vary by diagnosis group?	Diagnosis group	-	[AVERAGE] Check-in (mins)
Does the total time spent vary by diagnosis group?	Diagnosis group	-	[AVERAGE] Total Visit (mins)
Does the time between vitals being completed and the doctor visit starting vary by doctor?	Doctor	-	[AVERAGE] Wait for doctor (mins)
Which nurse/doctor is losing the most revenue for the company due to their average wait times?	Nurse (Vitals) or Doctor	-	[AVERAGE] Vital/Doctor wait cost
Does each doctor spend a similar amount of time with each patient? Does this vary by diagnosis group?	Doctor	Diagnosis group	[AVERAGE] Doctor Visit (mins)