Student name: YIZHENG HE

Student ID:221411294

SIT123: Data Capture Technologies

# Lab Report 3.2:

# Analyze motion sensor data (30 marks)

In this task we will see how to analyse a given set of motion sensor data and infer how John’s bathroom usage compares to the rest of the population.

## Due Date Friday 5:00pm, 12th August 2022

## Hardware Required

* None.

## Software Required

Microsoft Excel

## Pre-requisites: You must do the following before this task

1. **Attend Class (Lecture) & Seminar**
2. **Read this sheet from top to bottom**

## Task Overview

Raw data is given in the Excel file. Explanation about the data in the Excel file is given in the data document file “Data document.pdf”.

We will be using the following key steps to execute this task.

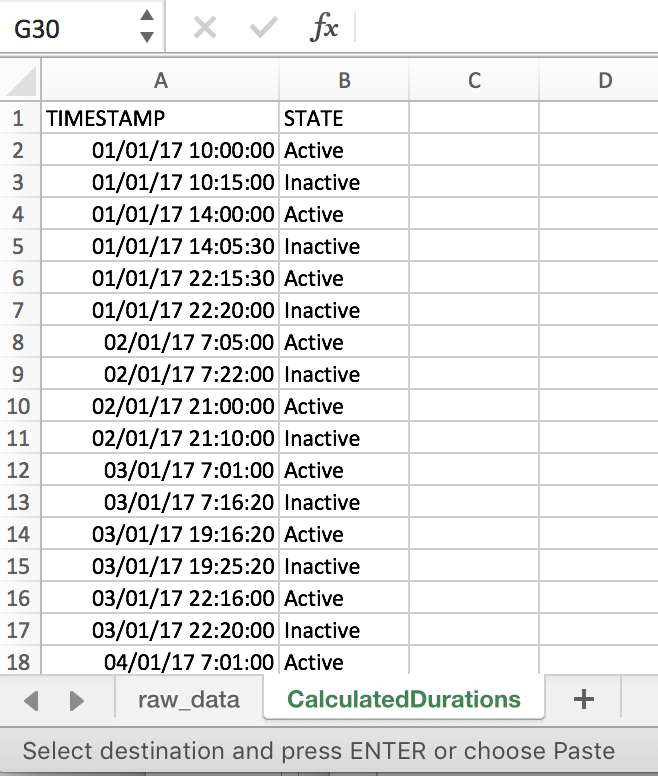
Step 1 - Read the provided Data document and inspect the provided raw data in the Excel file

Step 2 - Calculate the time durations for each bathroom visit

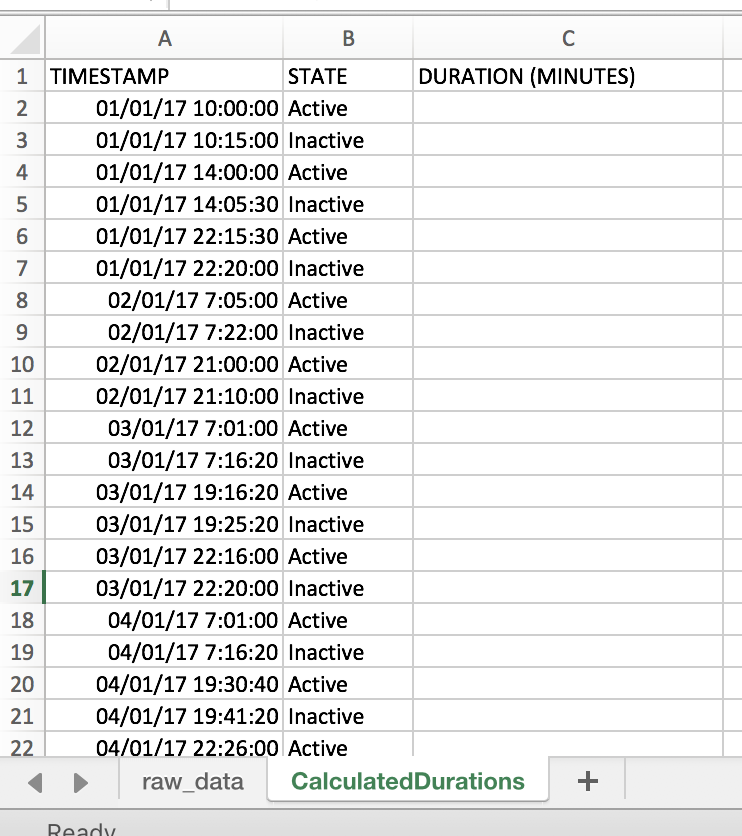
Step 3 - Calculate descriptive statistics for the data set using Excel’s built-in functions, to find out some useful information about John’s bathroom usage, such as the average time he spends in the bathroom per visit.

## Step-by-step Instructions: Calculate the time durations for each bathroom visit

1. Open the provided Excel file. You are now going to use the given data to calculate time durations for each bathroom visit. To do this, first insert new sheet. Name it ‘CalculatedDurations’
2. Copy the two columns TIMESTAMP & STATE from the first sheet to the newly created sheet



1. Add a new column with header ‘DURATION (MINUTES)’ to the right.



1. Now, what we want to do is to calculate how much time John spent in the bathroom per each visit. To do this, we need to find out the time difference between each ‘Active’ state and the next ‘Inactive’ state.

The first visit in the given data starts at 10 am and ends at 10:15 am. Let’s calculate the duration of this first visit.

The duration of the first visit is the time difference between 10:15 am and 10:00 am. To calculate this in Excel, we can write a simple formula.

Click on cell C3 and type the following:

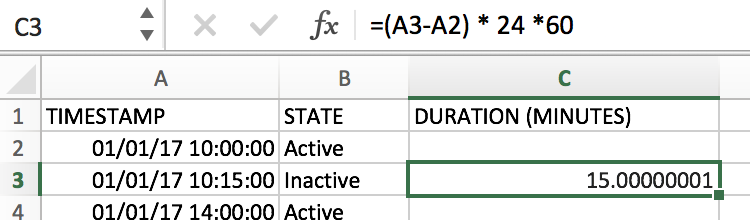
=(A3-A2)

What is displayed is the time difference, but you’ll notice that it is not in minutes! To get the answer in minutes, you need to multiply this by 24 \* 60. This is because there are 24 hours in a day, and 60 minutes to each hour. Modify the formula in C3 to be:

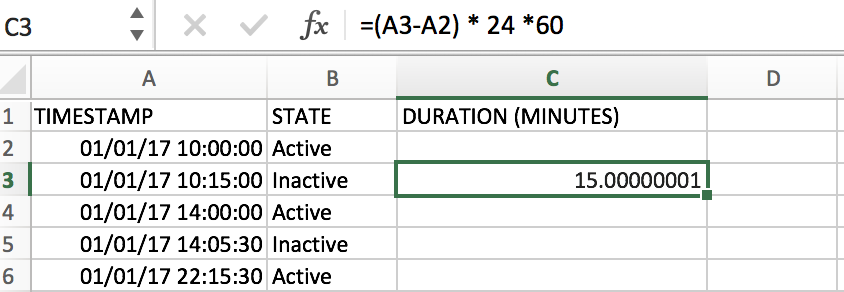
=(A3-A2) \* 24 \*60

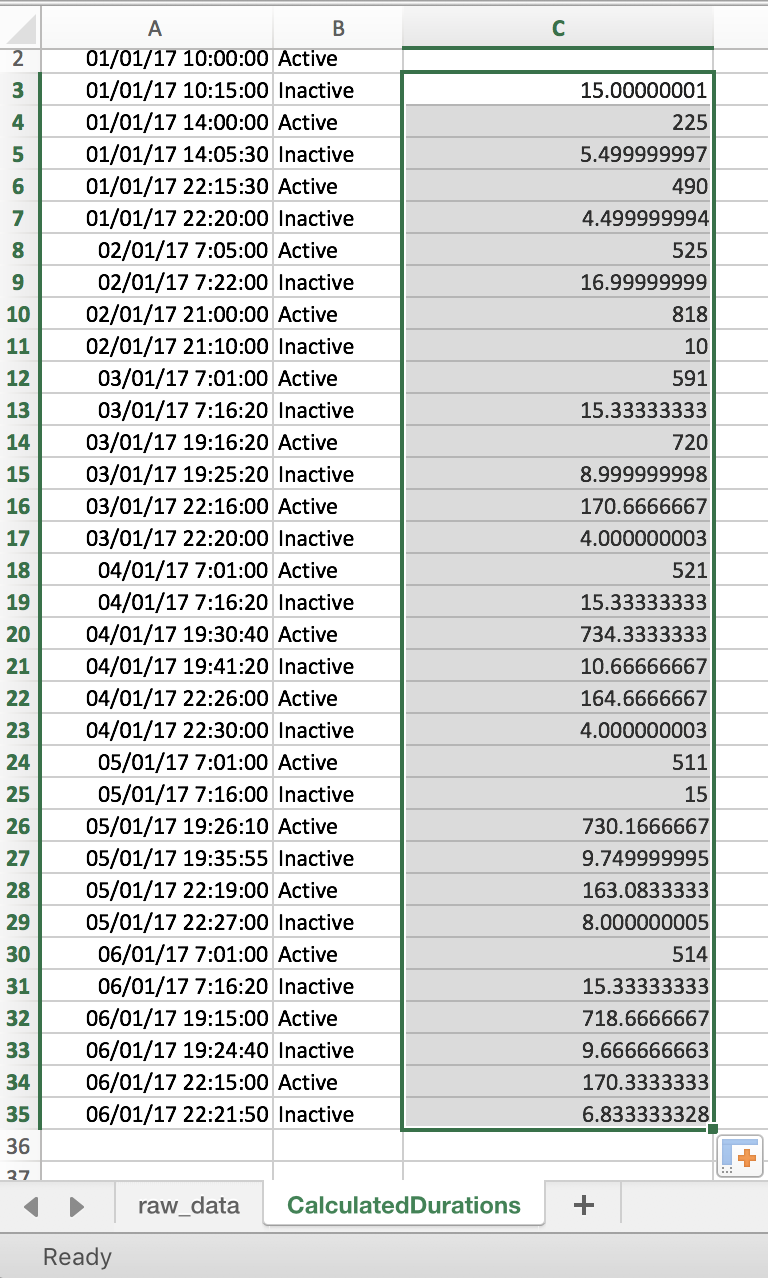
Press Enter.

You should now see 15 as the answer. As you can see, the time difference between 10:15 am and 10:00 is indeed 15 minutes, so our formula is correct!

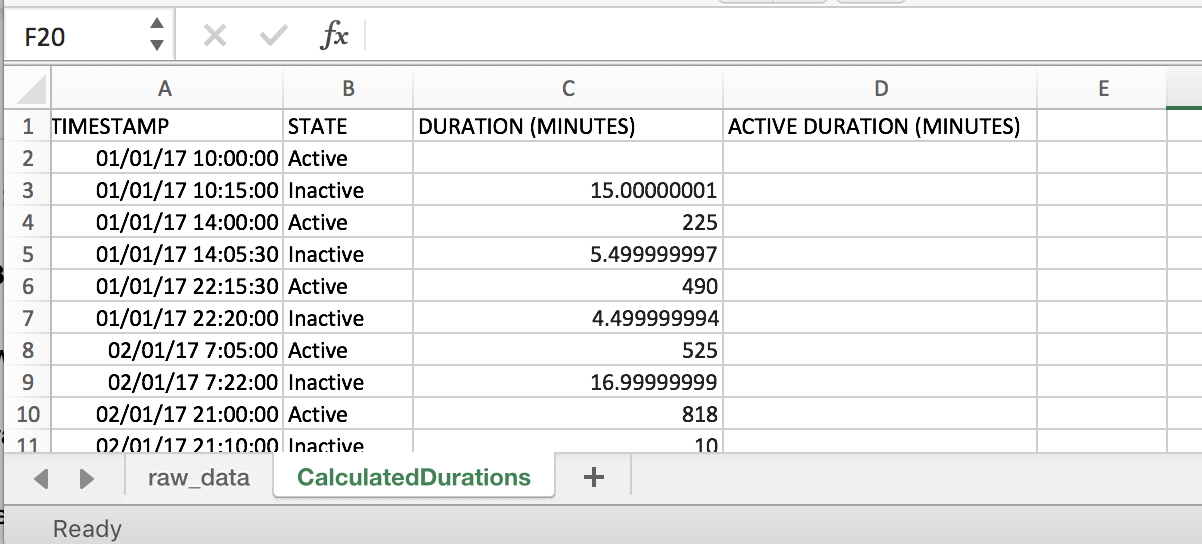


1. Now we need to apply this formula to all of the remaining rows in Column C. Select C3 and hover your pointer to the edge of the C3 cell until you see a cross hair. Then click and drag the pointer to the last cell in the range (C35)





1. Now as you can see we have the active durations in Column C. But, the active durations should be, [Time motion stopped - Time motion started], or in other words [Time motion went inactive - Time motion went active]. But our formula in step 5 applied it to ALL rows. For example, look at the result in C4. That is [Time new motion started - Time previous motion stopped]. Now we need to filter these out, and only keep durations from [Time motion went inactive - Time motion went active].
2. Add a new column header named ‘ACTIVE DURATION (MINUTES)’ to the right.



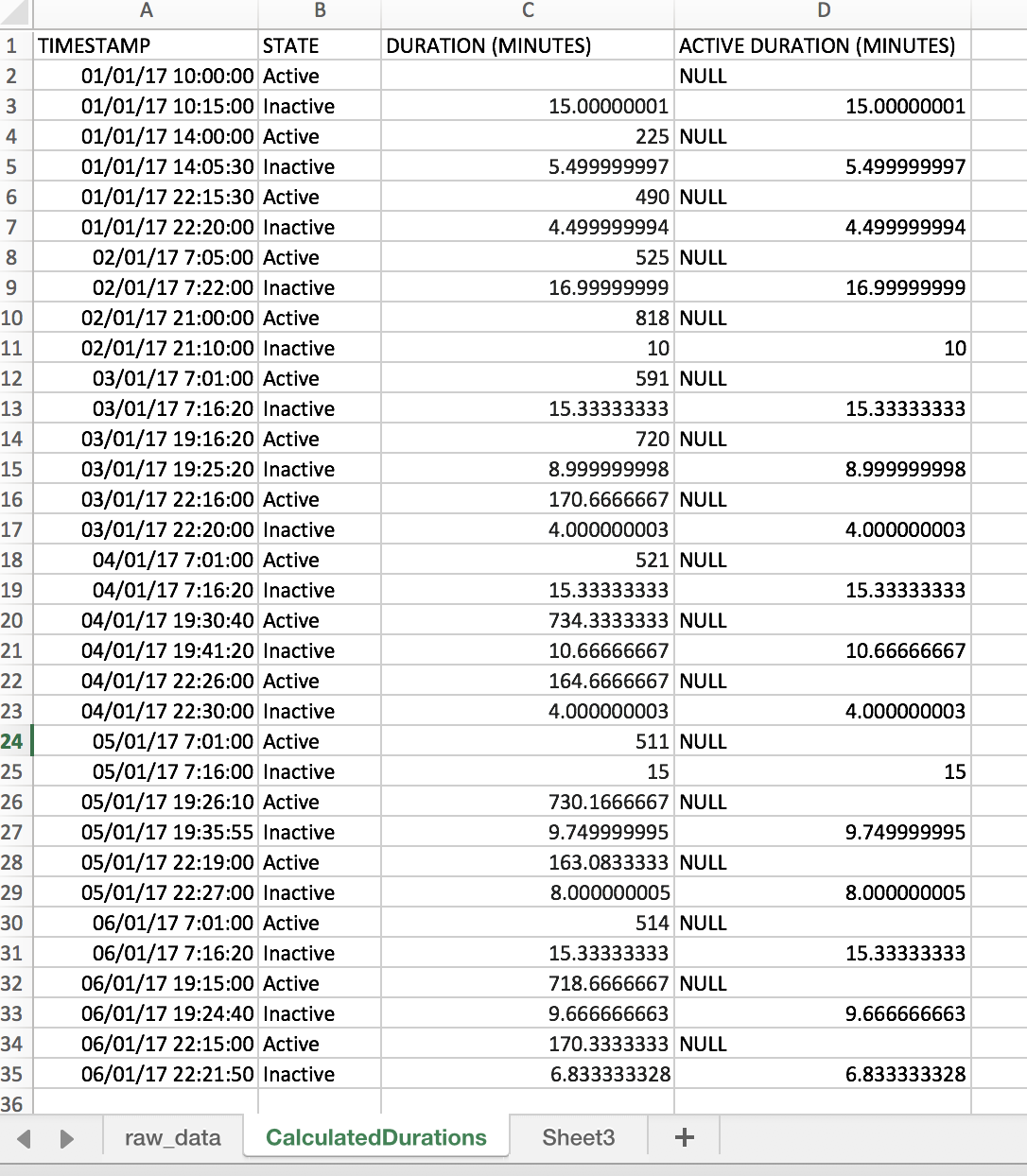
1. Of course, we can manually select which cells contain the Active Durations (time diff between Inactive - Active). But imagine if we had hundreds of rows! So we’re going to automate this by making use of Excel.
2. What we want Excel to do is to select the DURATION value from Column C if the corresponding row in Column B contains ‘Inactive’.

To do this, we write this as a formula. Type the following into D2:

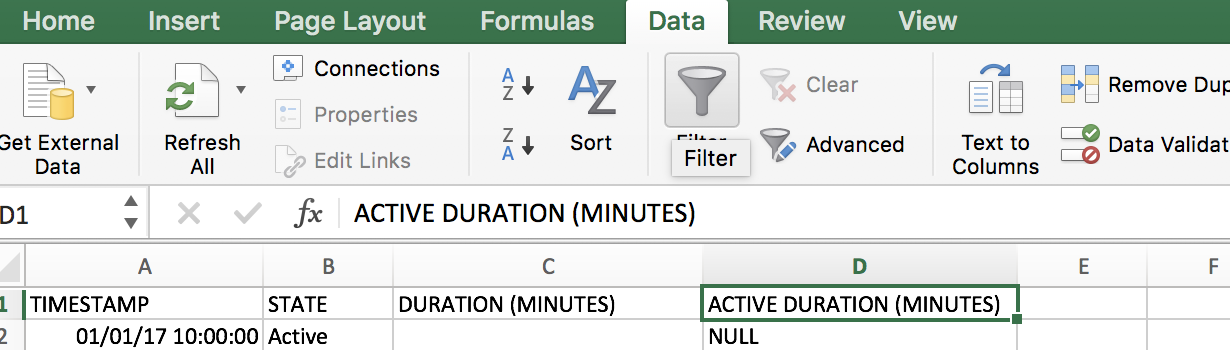
=IF(ISNUMBER(FIND("Inactive",B2)),C2,"NULL")

Here the formula is saying, IF B2 contains ‘Inactive’, then use the value in C2, else, insert NULL

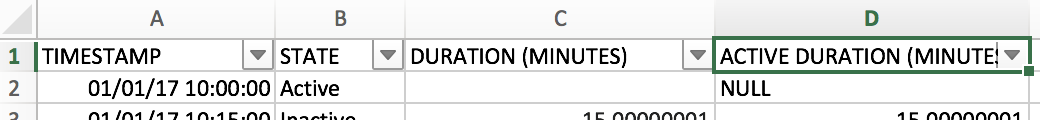
1. Apply the above formula to cells D2:D35



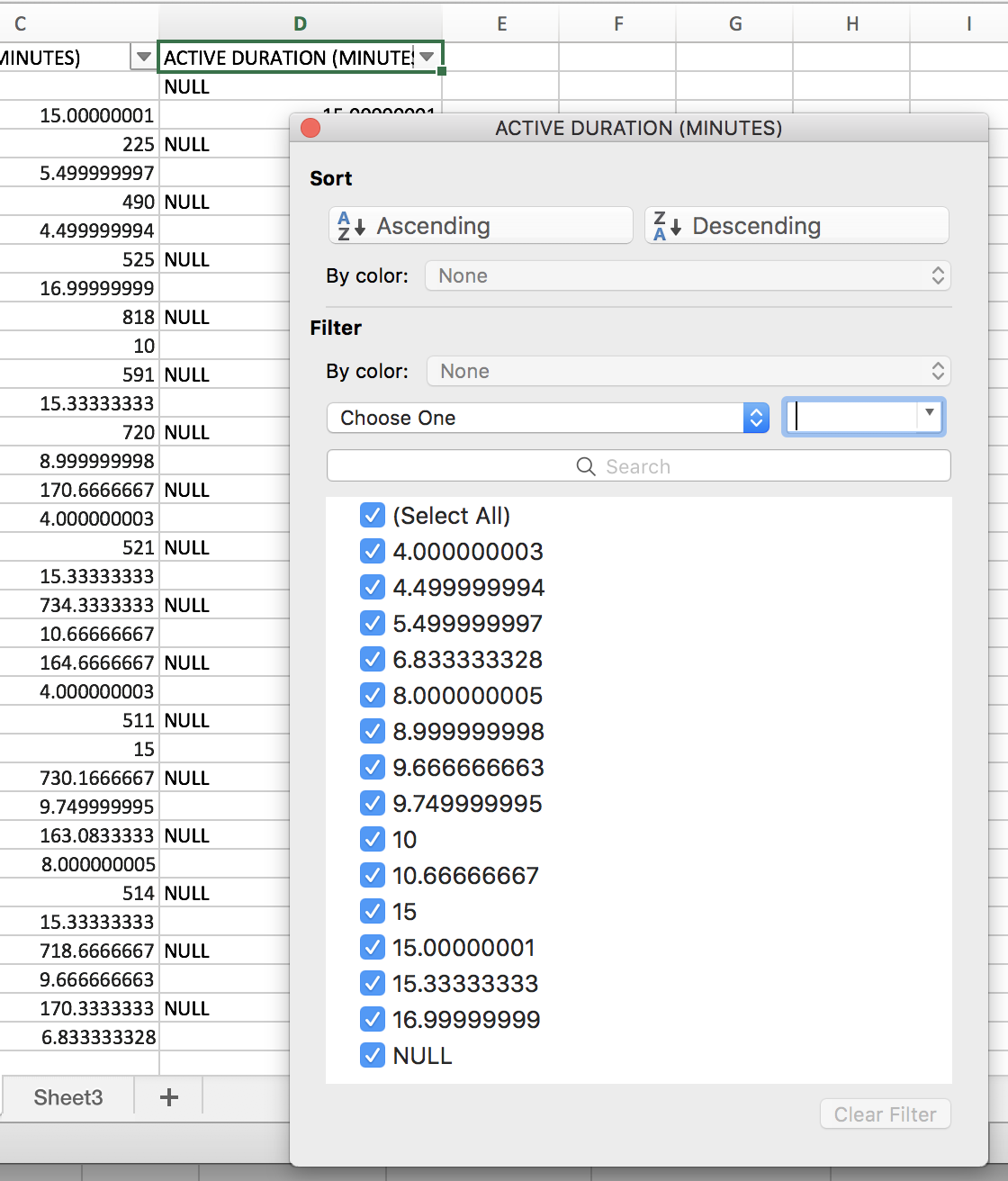
1. OK - now we have column D with only the active values, but now we’ve got NULL values which we need to filter out. Remember that our goal is to have a list of active duration values. We are going to filter these in the next step.
2. Click on D1, and on the Excel Ribbon, under ‘Data’, click on ‘Filter’



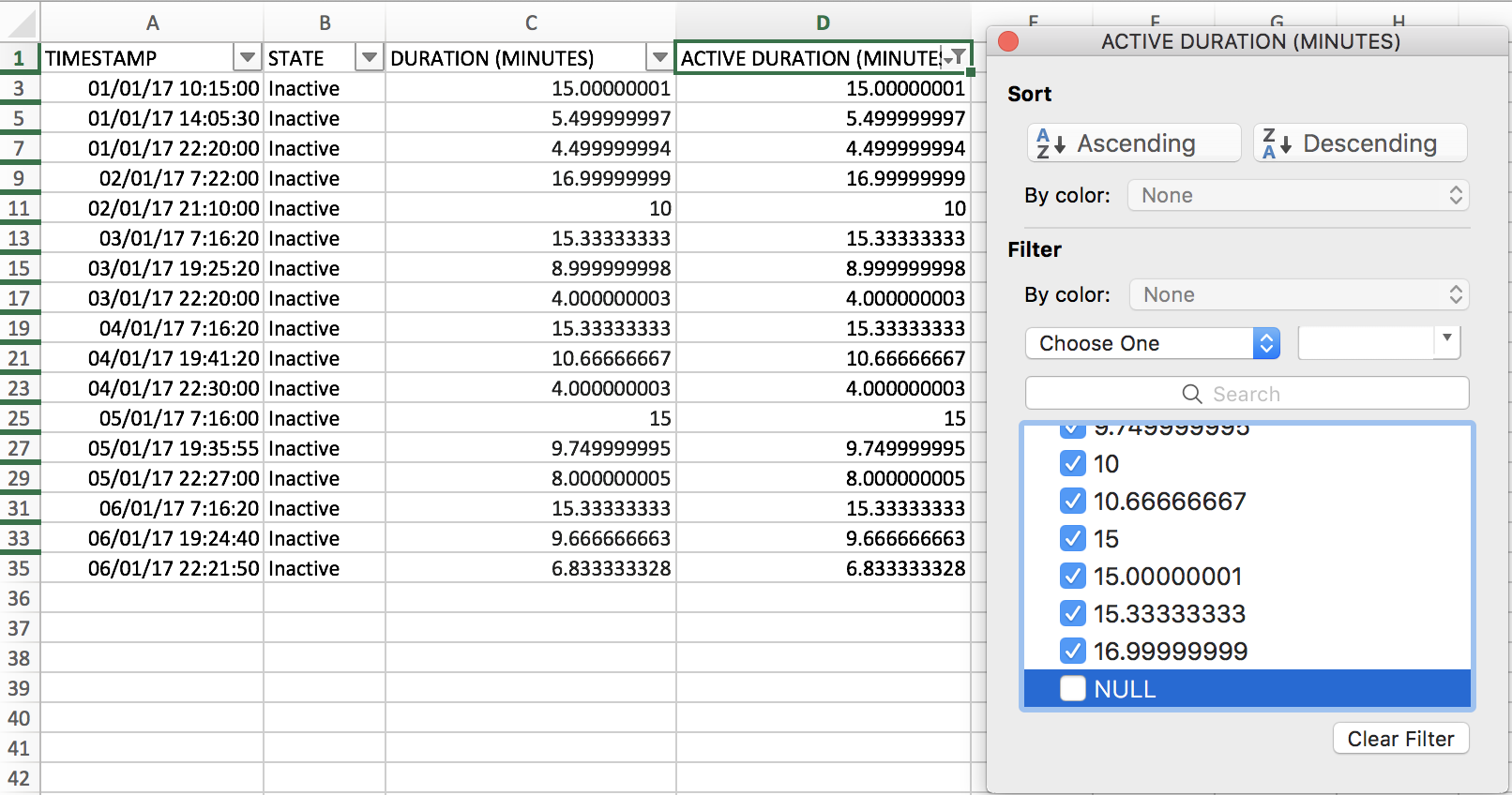
1. Now your column headers should show drop down arrows:



1. Click on the drop down arrow on D1. It will show a filter dialog box



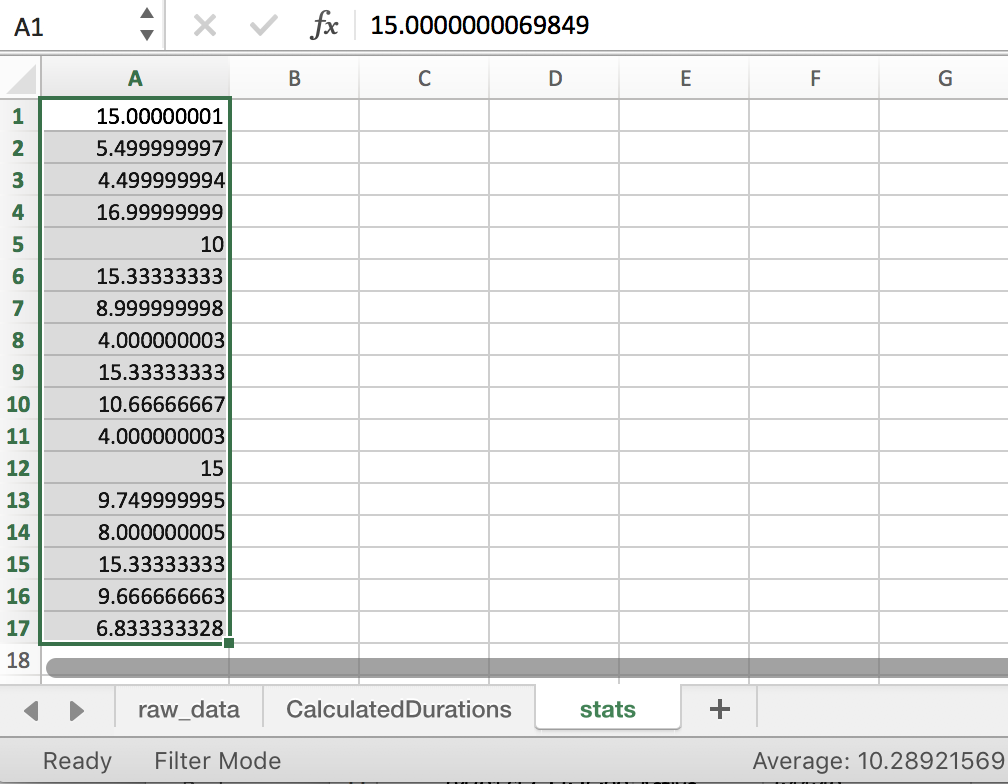
1. Untick the checkbox for ‘NULL’. This will immediately filter out the NULL values in Column D.



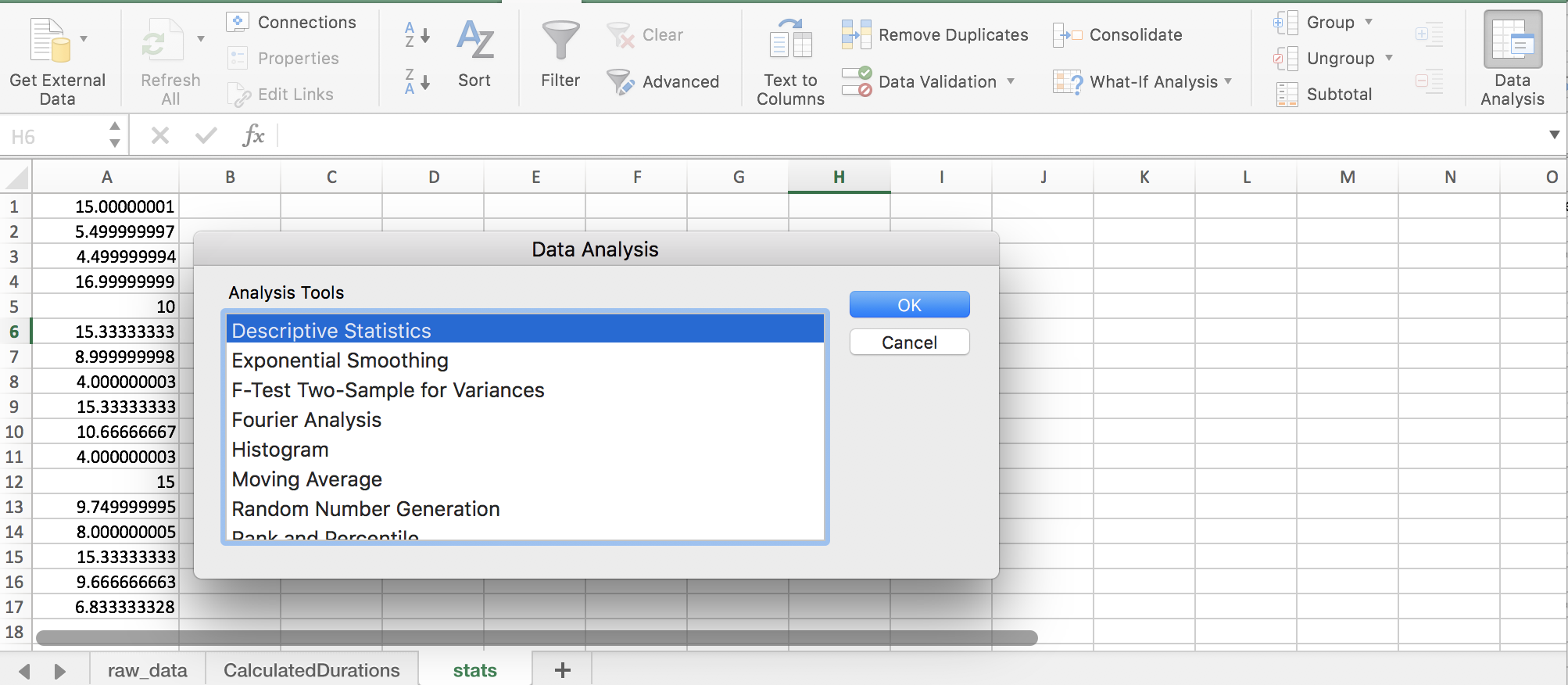
1. Have a look at column D. Now all the values that are displayed are the active durations for each bathroom visit!

## Step-by-step Instructions: Calculate descriptive statistics for the data set

1. First, create a new sheet named ‘Stats’, and copy-paste the filtered Active Duration values in column D, to the new sheet.

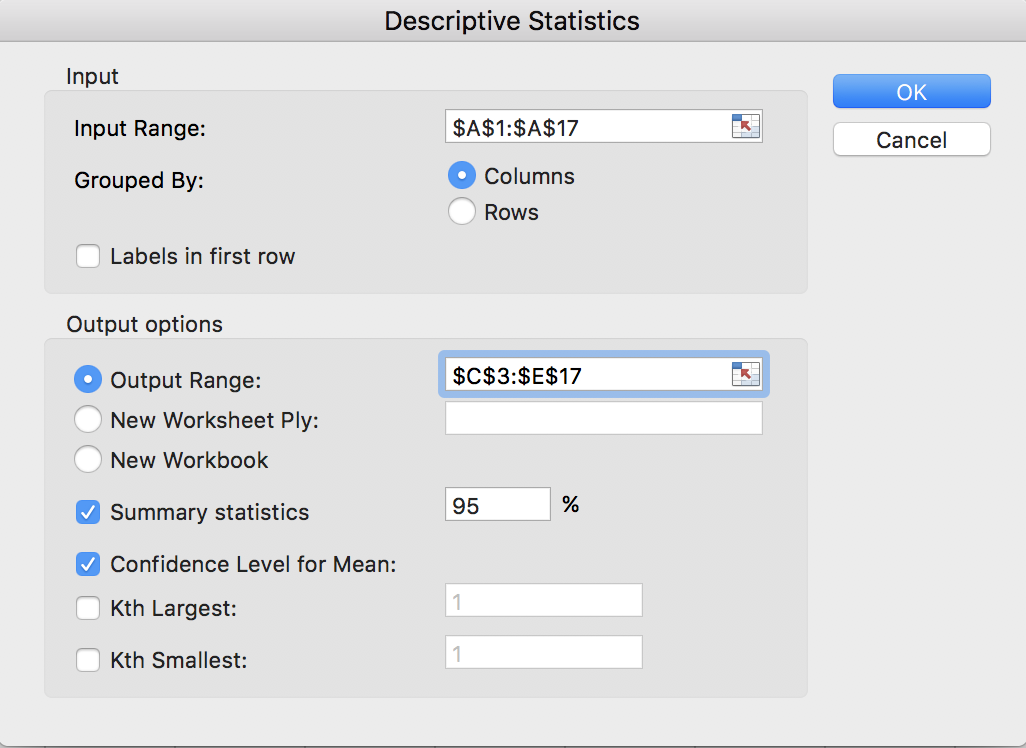


1. Next you need to turn on Excel’s Data Analysis Add-On. Follow the instructions here: <https://www.youtube.com/watch?v=mIoS7IRo36c>
2. Now, click on ‘Data Analysis’

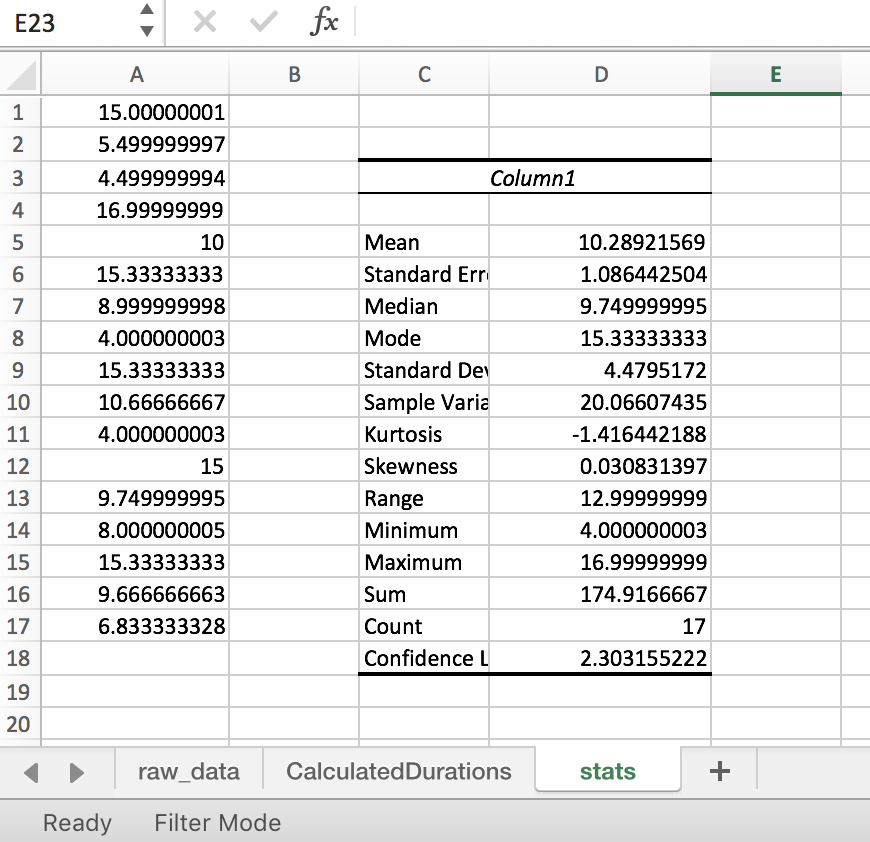


1. Select Descriptive Statistics and click OK.
2. Now, you will be shown the dialog box for Descriptive Statistics. Here, you need to select the following:
   1. Input range: this is the data range in column A
   2. Output range: this is cell range to add the descriptive statistics values
   3. Summary statistics

Type in the following values:



1. Click OK. Descriptive statistic values for the given sample should displayed as in the following screenshot:



All done! Now we have some useful information about John’s bathroom usage.

## Task Submission Details

There are 2 questions in this task. Answer all of them in this word document itself and submit to unit site.

### **Q1: Upload your completed excel workbook to unit site.**

(16 marks)

### **Q2: Look at the calculated descriptive statistics. Fill in the below table explaining what each term means in terms of John’s bathroom usage. The first row has been done for you.**

|  |  |  |
| --- | --- | --- |
| Mean | 10.289 minutes | The average time John uses the bathroom per visit is 10.289 minutes. |
| Median | 9.74999999511055 | John's median time per bathroom visit was 9.75 minutes. |
| Mode | 15.3333333344198 | John's most frequent bathroom break was 15.33 minutes. |
| Standard deviation | 4.47951720037625 | The standard deviation time for John to use the toilet was 4.48 minutes. |
| Minimum | 4.00000000256113 | The minimum time for John to use the toilet is 4 minutes. |
| Maximum | 16.9999999925494 | The maximum time John used the toilet was 17 minutes. |
| Count | 17 | A total of 17 times data of John's toilet use were recorded. |
| Sum | 174.916666651843 | John spent a total of 174.92 minutes using the toilet. |

(14 marks)

## References

<https://www.khanacademy.org/math/statistics-probability/displaying-describing-data#mean-median-basics>

<https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/pop-variance-standard-deviation/v/range-variance-and-standard-deviation-as-measures-of-dispersion>

<https://www.lynda.com/Business-Skills-tutorials/Descriptive-statistics/550747/611825-4.html?org=deakin.edu.au>