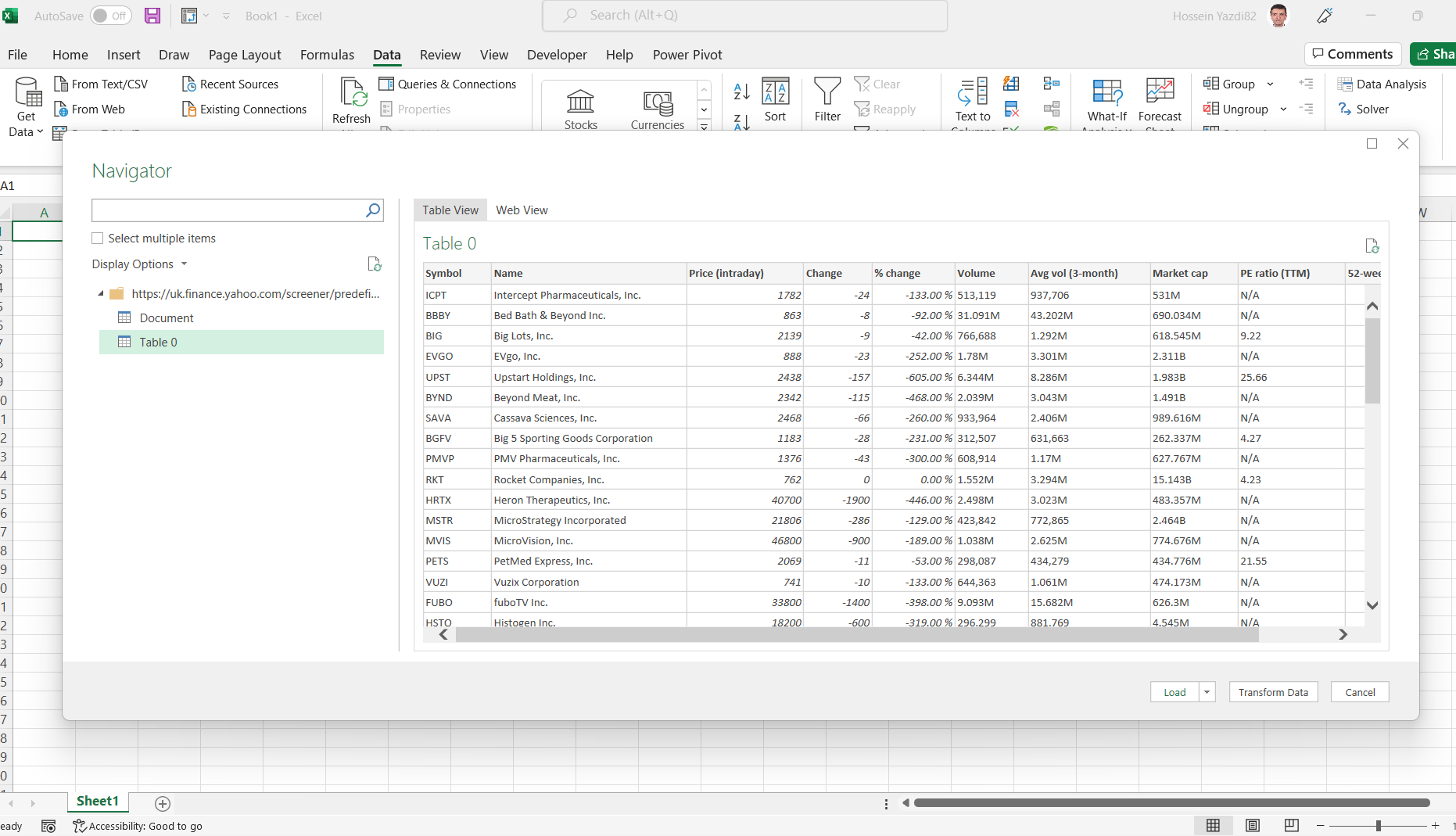
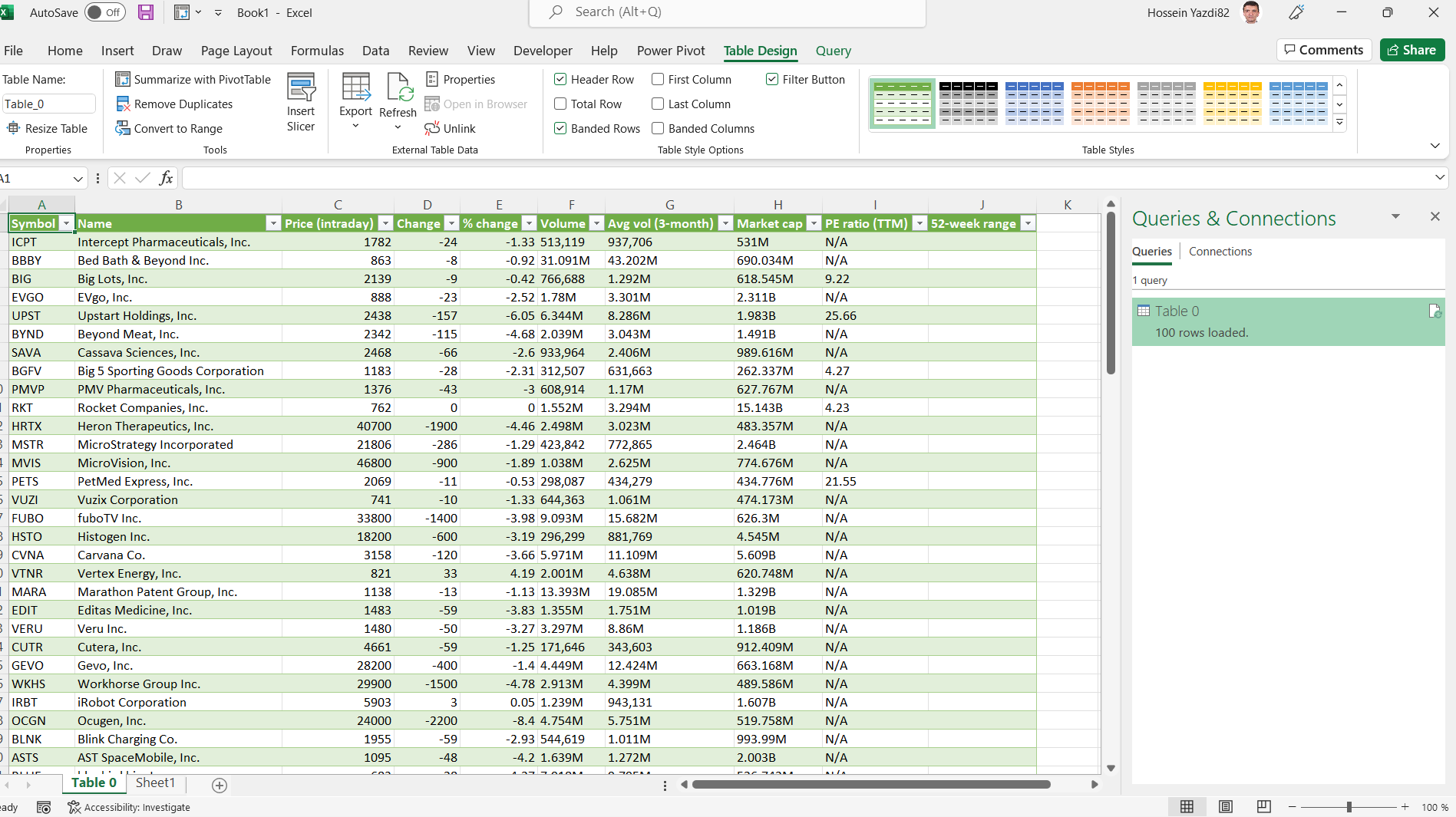
# Step 1

The data is downloaded by “Excel/Data Tab/From the Web” from this link of Yahoo Finance: <https://uk.finance.yahoo.com/screener/predefined/most_shorted_stocks?offset=0&count=100>



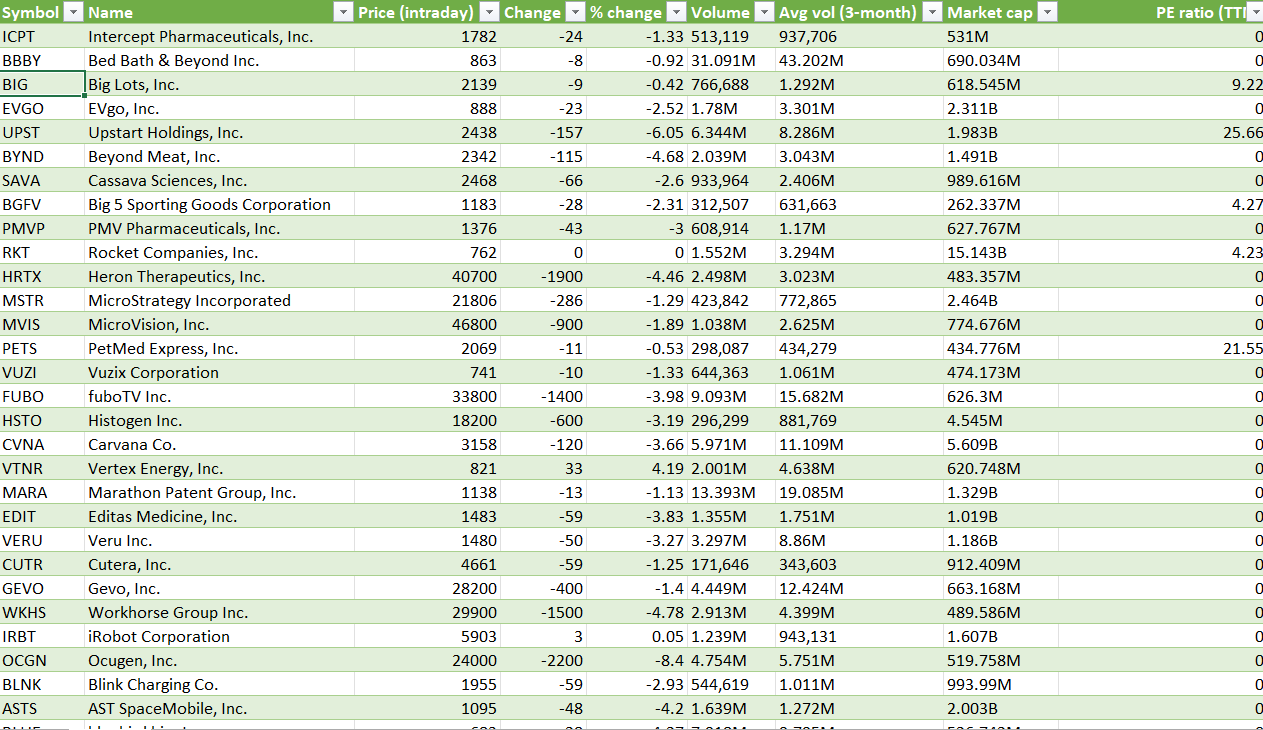


# Step 2

**Cleaning, Wrangling**

Deleting the “J” column as it is empty and useless.

Substitution of the “N/A” data in the “I” column to be zero as the relevant companies have no P/E ratio.



# Step 3

But there is a problem that some columns contain commas as thousand separators, “M” as millions, and “B” as billions so they must be changed.

To replace commas only in the selected columns from “F” to “I”, in the replace box we use Ctrl+A.

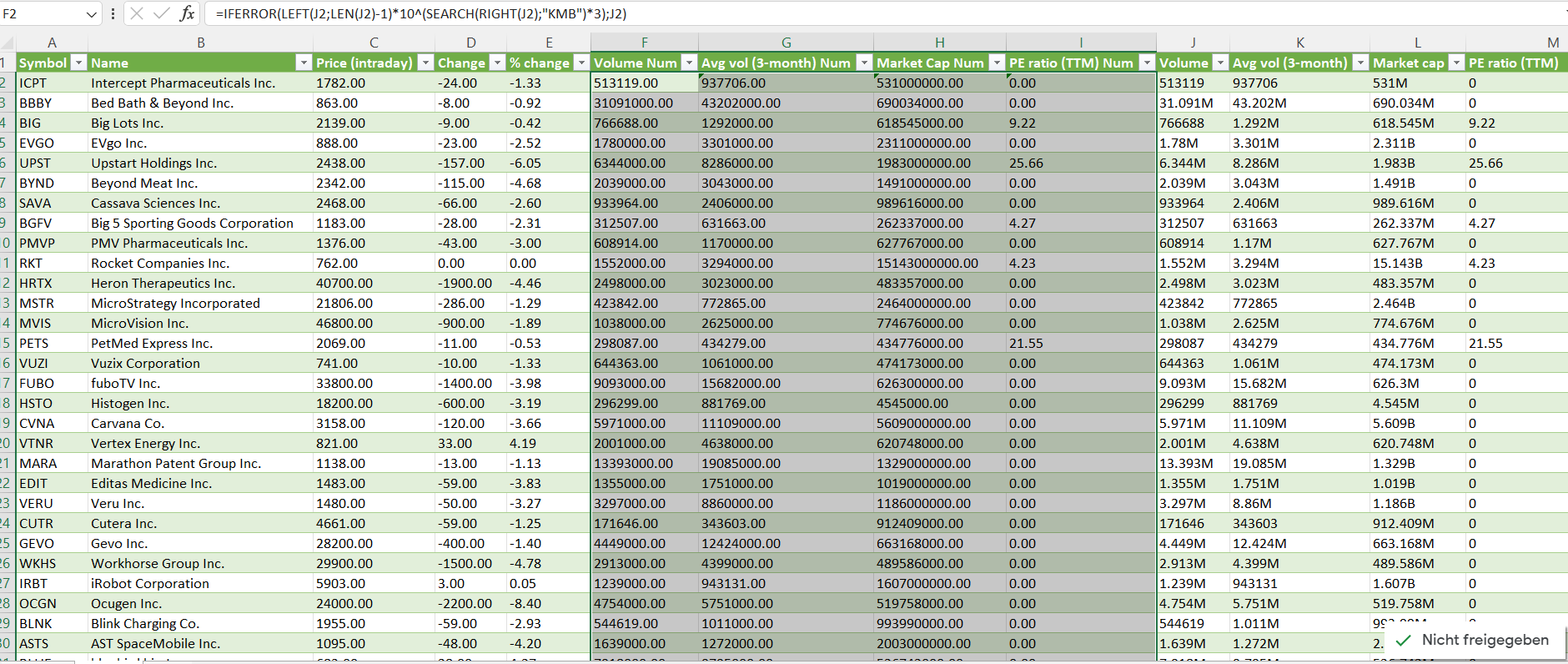
And to convert M and B to their numerical equals, after inserting new columns, the function below is helpful:

=LEFT(J2;LEN(J2)-1)\*10^(SEARCH(RIGHT(J2);"KMB")\*3)

This formula, as does not use absolute reference, can be expanded to all the columns both horizontally and vertically.

But we have another problem relating to those cells without M or B, returning the “Value” error. To solve it is possible to use “iferror” function.

=IFERROR(LEFT(J2;LEN(J2)-1)\*10^(SEARCH(RIGHT(J2);"KMB")\*3);J2)



# Step 4

**Query:** In what way and how much does the price as the dependent variable depends on the other independent variables, using the multiple linear regression?

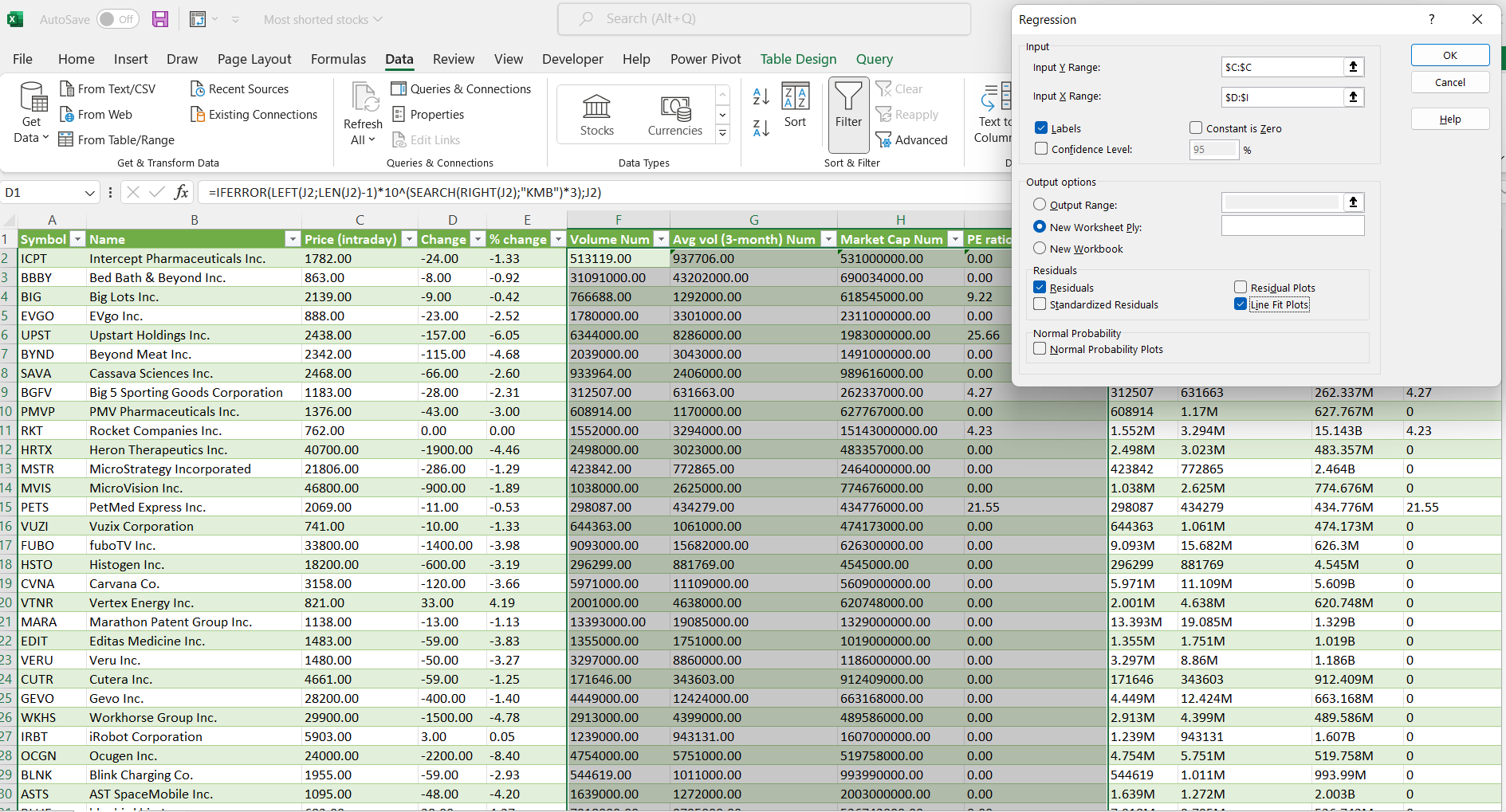
To do so, we use the “Data Tab/Analyze/Data Analysis/Regression”.

Input Y range: price in column C

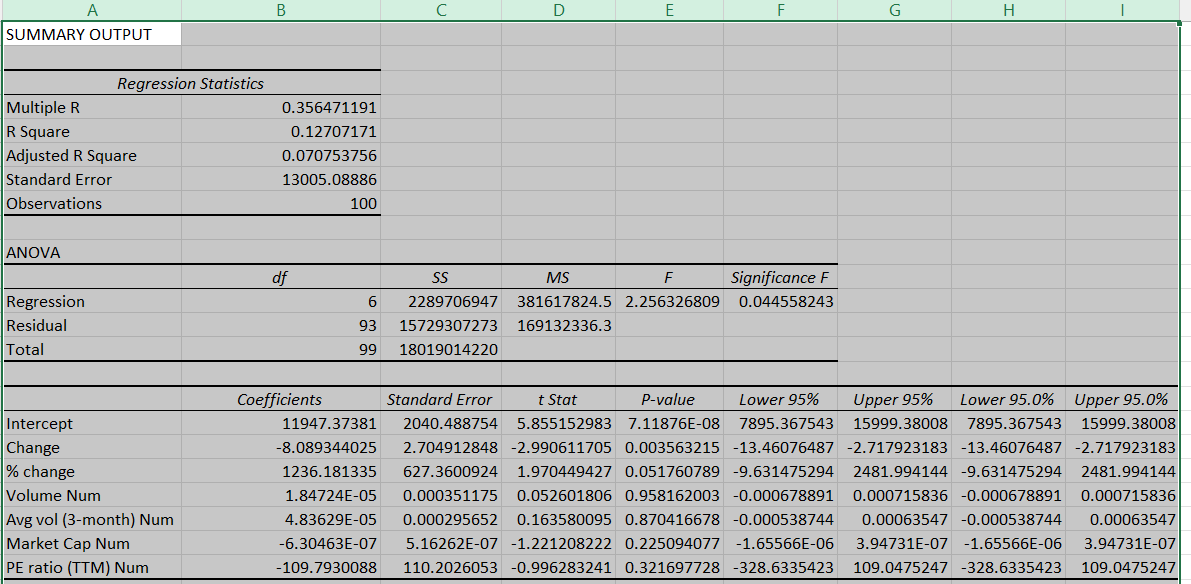
Input X range: other numerical columns

Selecting the options: Labels, Residuals, Line Fit Plots.

There are errors associated with the fact that some data are formatted as text which must be converted to numbers.



The result is as follows:



# Step 5

**Interpretation**

The figures for both “R Square” and “Adjusted R Square” are very low, indicating the fact of not a sufficient correlation between the independent and dependent variables.

The function is Y = β0 + β1 X1+ β2 X2+ β3 X3+ β4 X4+ β5 X5+ β6 X6

Y = 11947.37381 + -8.089344025 X1+ 1236.181335 X2+ 1.84724E-05 X3+ 4.83629E-05 X4+

-6.30463E-07 X5+ -109.7930088 X6

If we check the formula by use of some data from the table, it shows that the prediction is not fine, meaning that the data for “Most Shorted Stocks of screeners” are not so much correlated in total, again annotating not to be good enough predictable.

Assessing the 1 minus P-Value for all elements shows that not all of them are greatly significant.

