**HW2 Submission**

Megan Moore

[Mmmoore2@buffs.wtamu.edu](mailto:Mmmoore2@buffs.wtamu.edu)

**1) Integrate data in SQL database (34 points)**

a) How many records do you get after you run your query? 518

b) How many columns or attributes do you have in your dataset? 12

SELECT suppliername,suppliers.country AS suppliercountry, categoryname, productname, price, quantity,

orderdate, customername, customers.country AS customercountry, shippername, lastname, firstname

FROM suppliers

JOIN products ON suppliers.supplierid=products.supplierid

JOIN categories ON categories.categoryid=products.categoryid

JOIN orderdetails ON products.productid=orderdetails.productid

JOIN orders ON orderdetails.orderid=orders.orderid

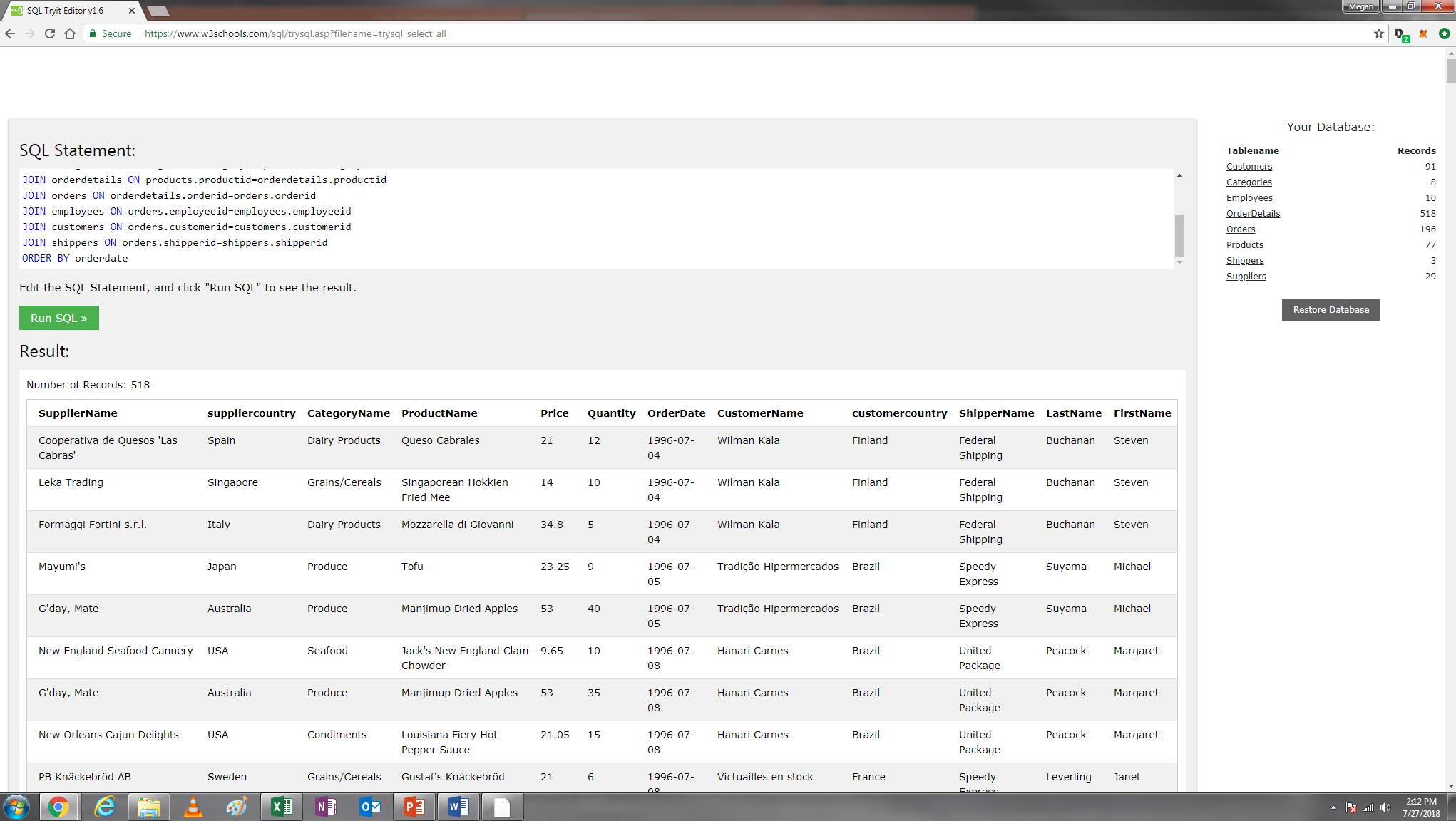
JOIN employees ON orders.employeeid=employees.employeeid

JOIN customers ON orders.customerid=customers.customerid

JOIN shippers ON orders.shipperid=shippers.shipperid

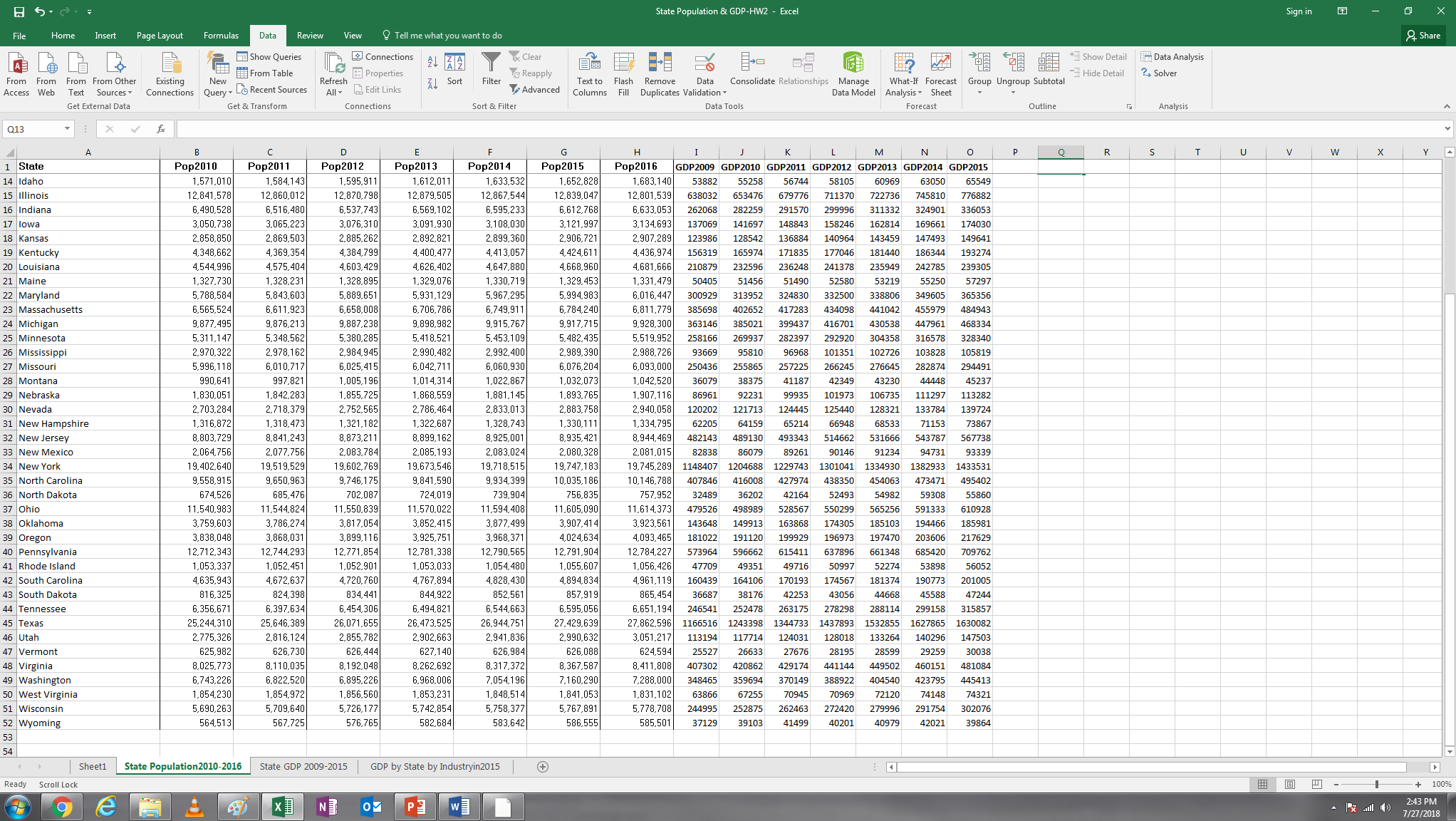
ORDER BY orderdate

Screenshot 1



2) Combine State Data & Calculate Correlation (12 points)

Screenshot 2



a) What’s the correlation coefficient between population in 2014 and GDP in 2014 (round your result to

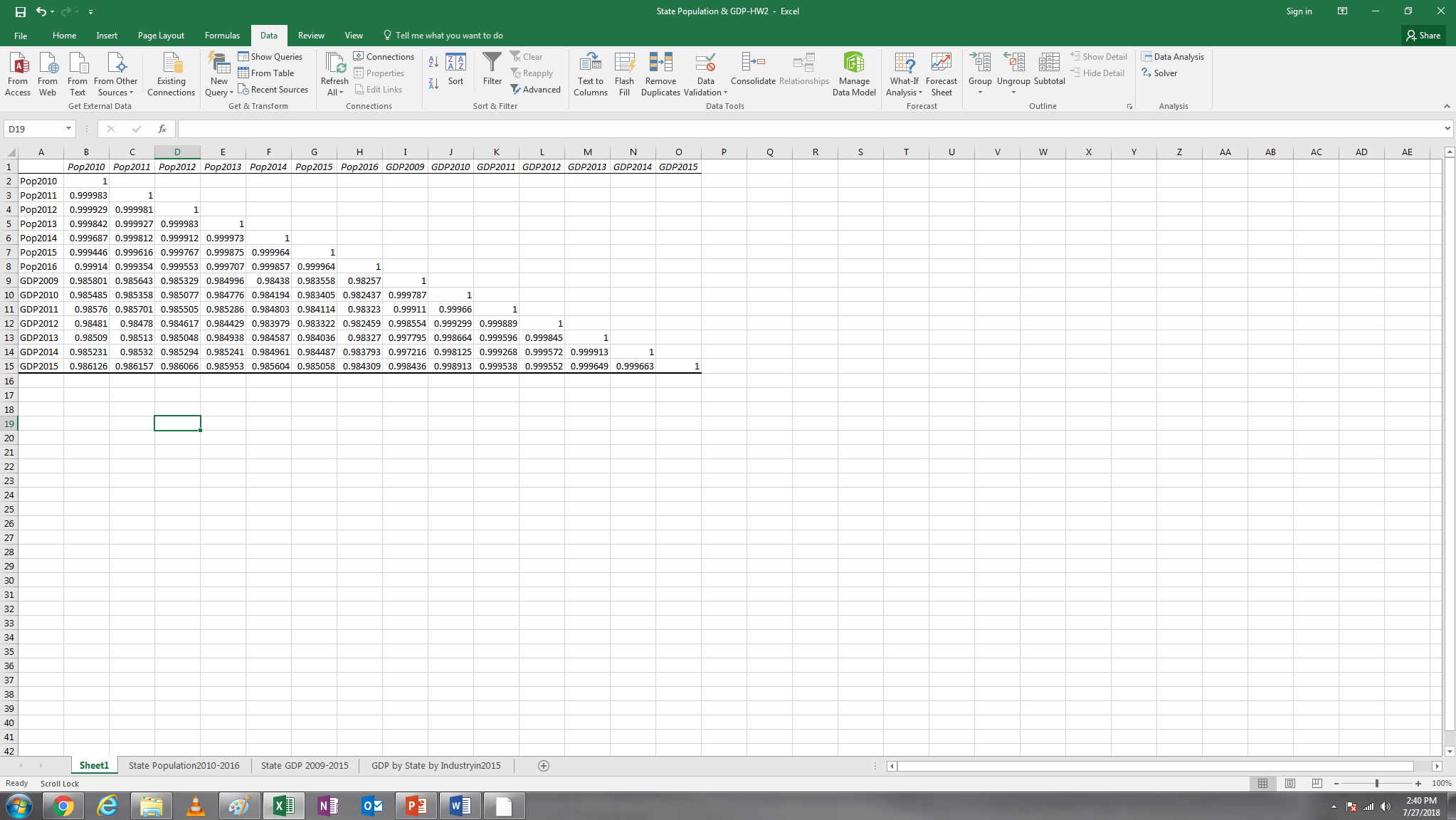
the four decimal places such as 0.1111 or -0.1111) .984961, rounds to .9850

b) What’s the correlation coefficient between population in 2011 and GDP in 2012 (round your result to

the four decimal places such as 0.1111 or -0.1111) .98478, rounds to .9848

c) Based on Slide 16, can we consider one of state population and state GDP as a redundant feature? Type Yes or No. YES

Screenshot 3



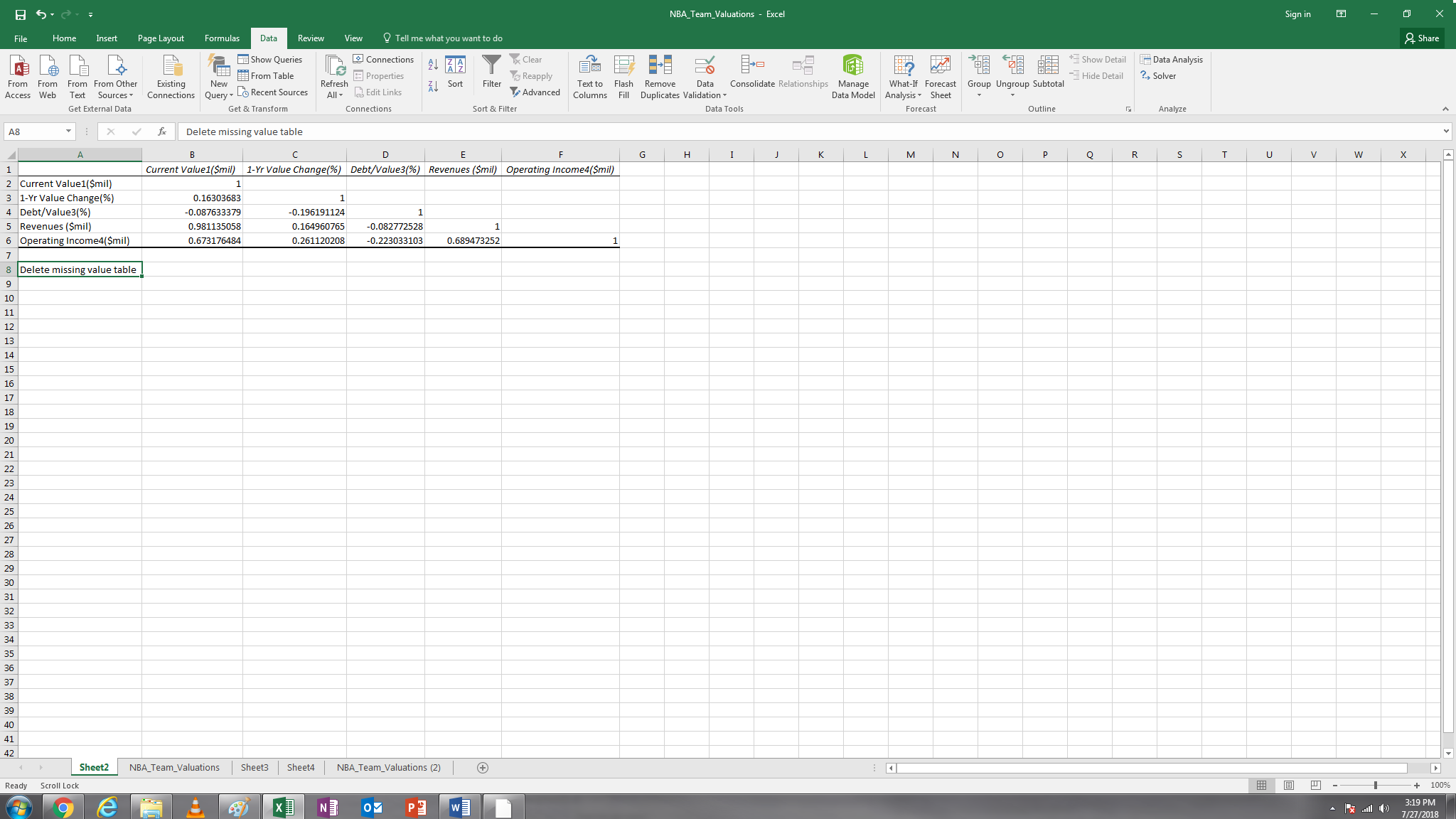
**Practice 2: Deal with missing values (21 points)**

1) Method 1: Delete the record with the missing value

a) What’s the correlation coefficient between 1-Yr Value Change(%) and Debt/Value3(%)? -0.1962

b) What’s the correlation coefficient between Current Value1($mil) and Revenues ($mil)? 0.9811

Please take a screenshot of the correlation coefficient matrix you generated for Method 1 (Screenshot 4).

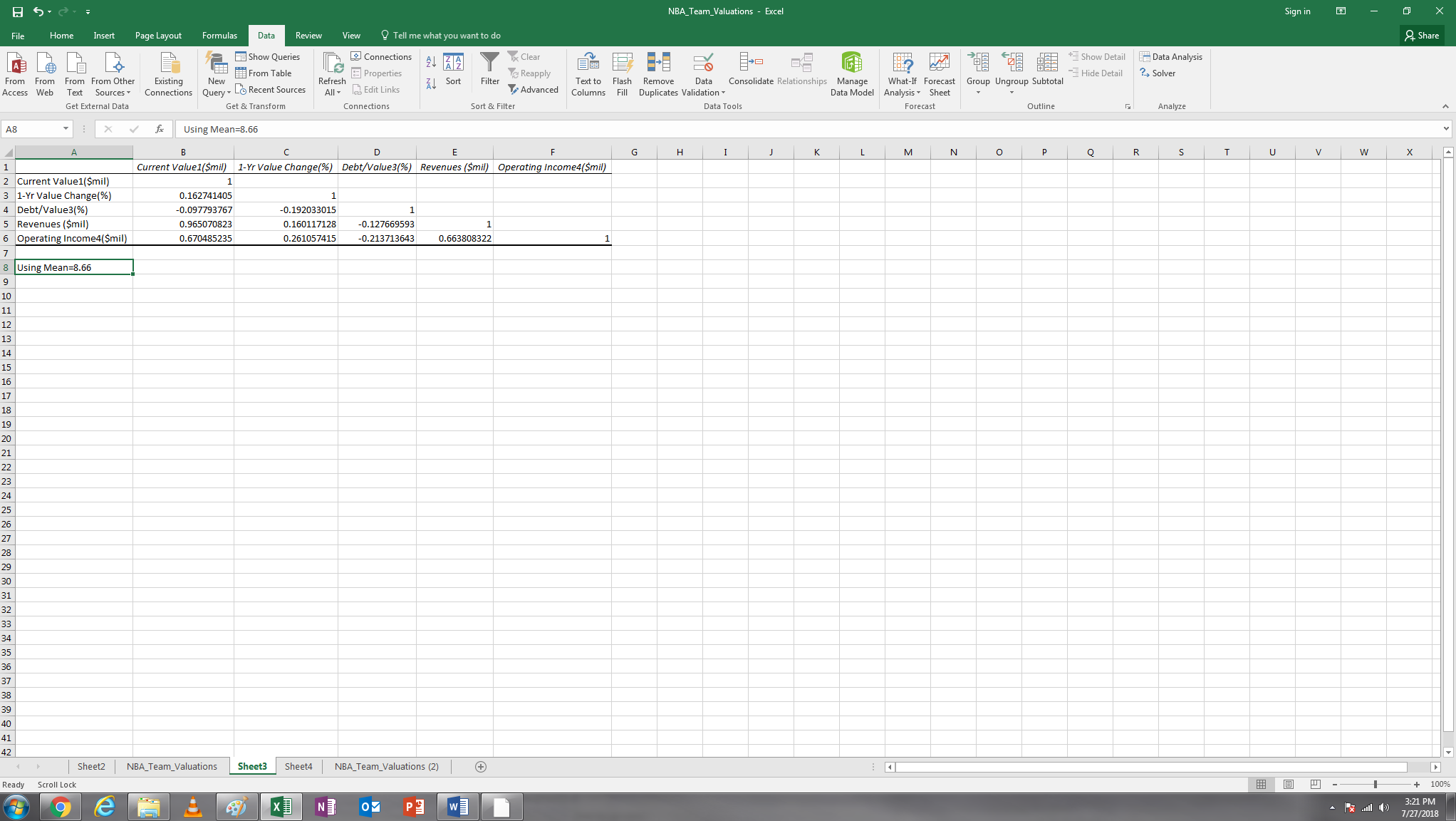


2) Method 2: Replace the missing value with the mean

a) What’s the correlation coefficient between 1-Yr Value Change(%) and Debt/Value3(%)? -0.1920

b) What’s the correlation coefficient between Current Value1($mil) and Revenues ($mil)? 0.9651

Please take a screenshot of the correlation coefficient matrix you generated for Method 2 (Screenshot 5).

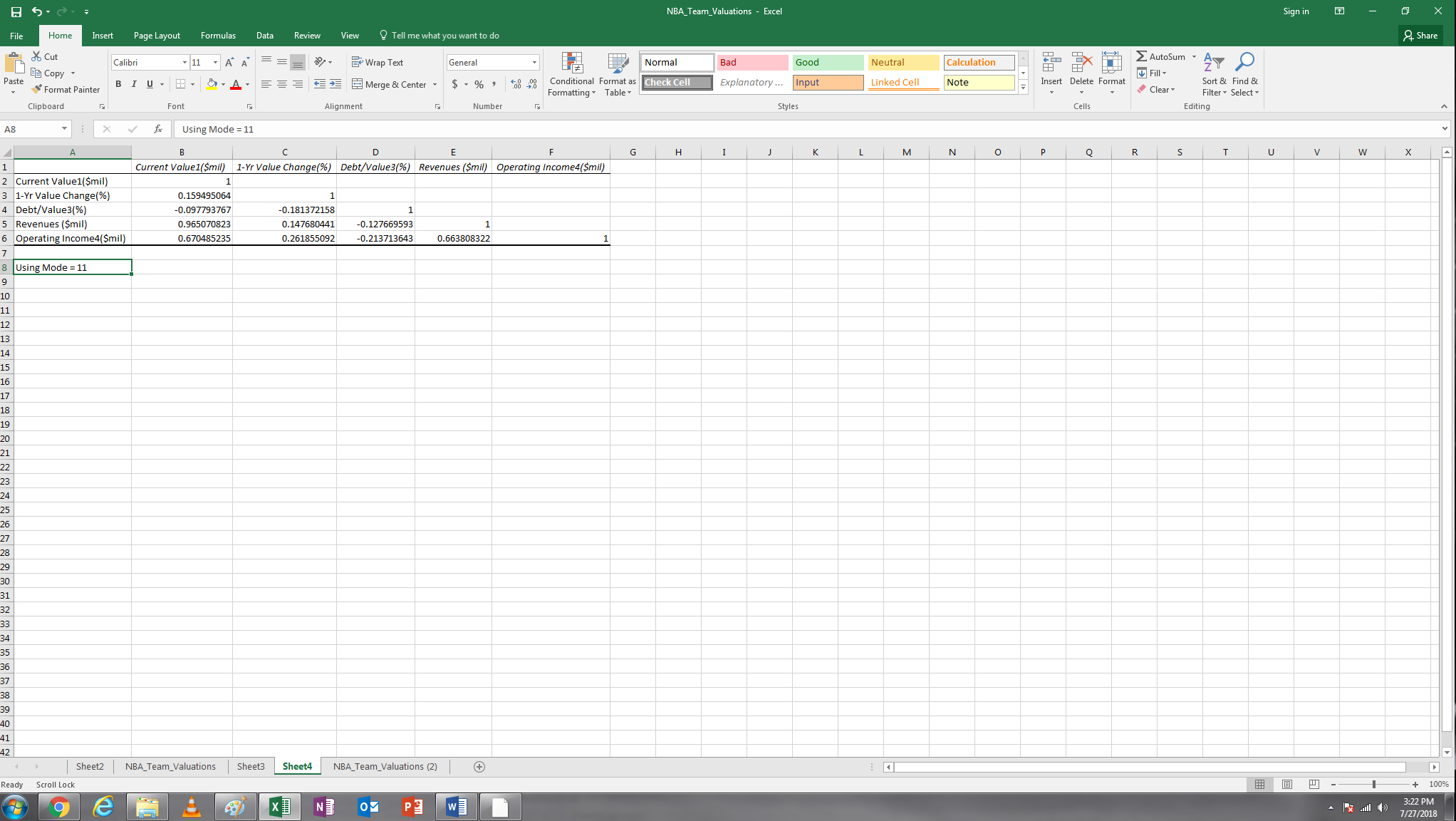


3) Method 3: the missing value with the mode

a) What’s the correlation coefficient between 1-Yr Value Change(%) and Debt/Value3(%)? -0.1814

b) What’s the correlation coefficient between Current Value1($mil) and Revenues ($mil)? 0.9651

Please take a screenshot of the correlation coefficient matrix you generated for Method 3 (Screenshot 6).



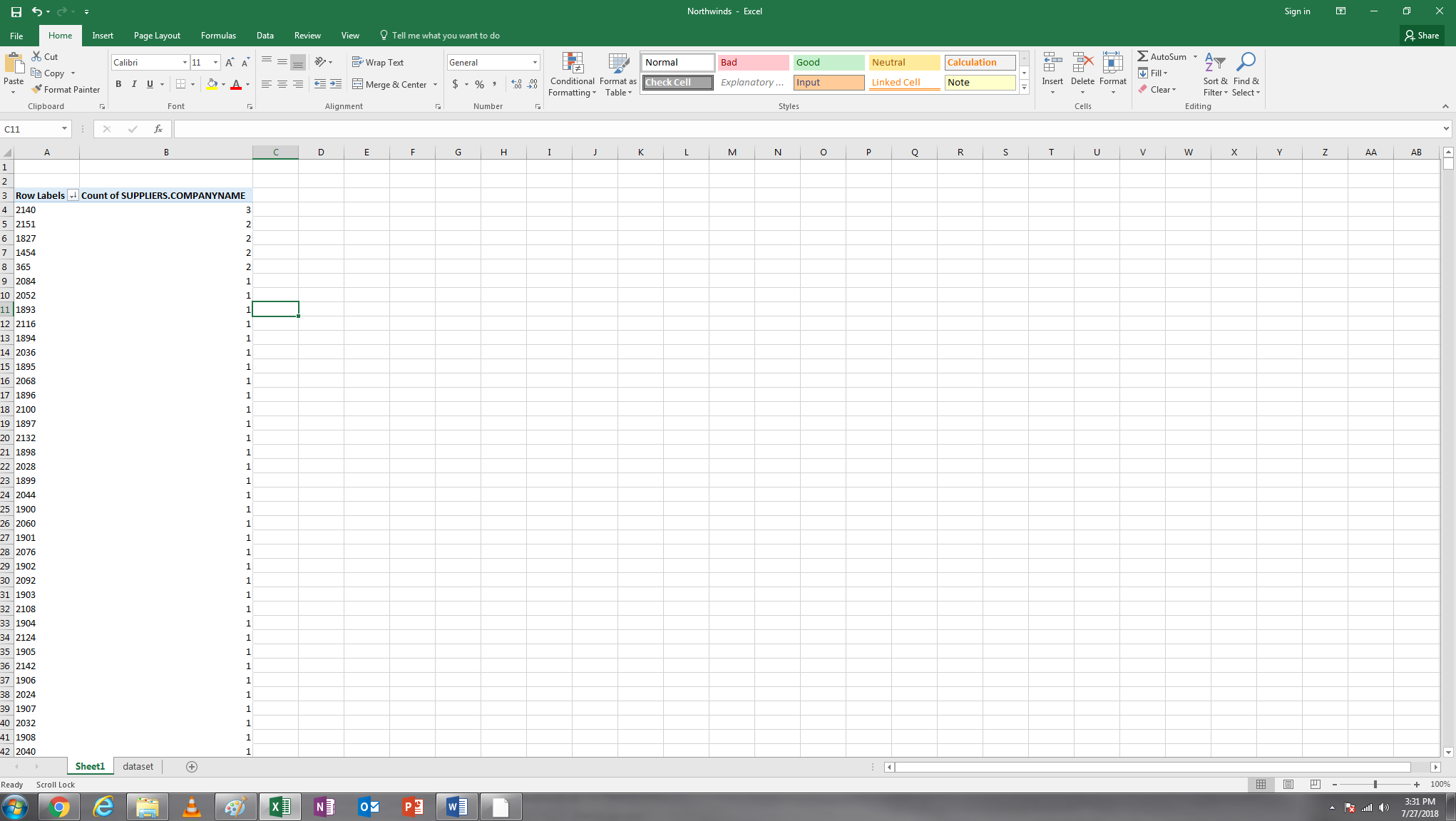
4) Based on the analysis, do you think different methods to deal with missing values will always generate the same results? Please type Yes or No here. No. Different methods generate different results.

**Practice 3: Deal with duplicated values (9 points)**

a) One record appears three times in this dataset. Please indicate the ID number of this record . 2140

b) How many records appears twice in this dataset? 4

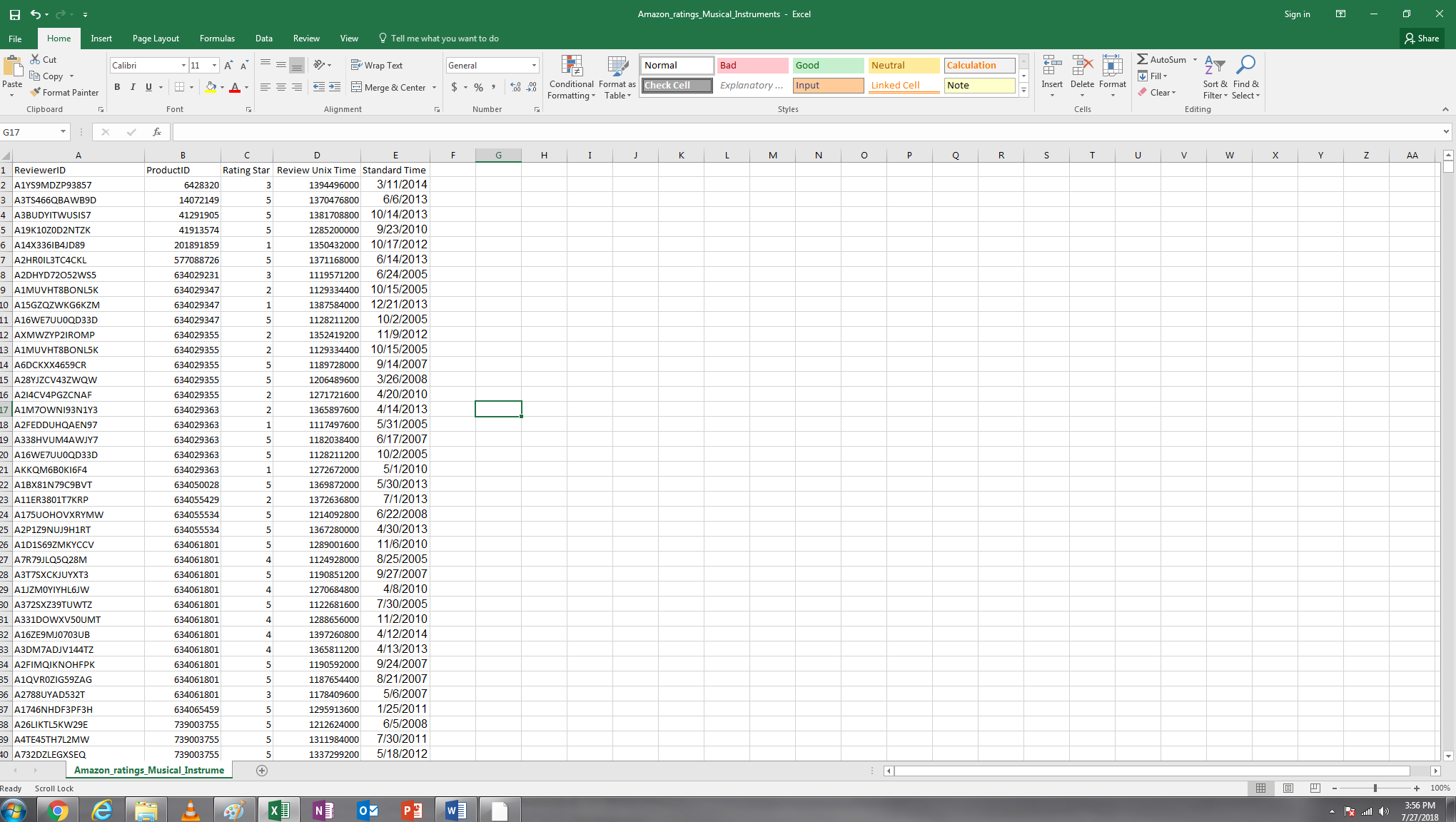
Please take a screenshot of the PivotTable you generated for this question (Screenshot 7).



c) How many duplicated records are removed from the dataset after you use the function “Remove Duplicates” in Excel? 6

**Practice 4: Data Transformation & Basic Analysis (24 points)**

After completing this data transformation, please take a screenshot of the transformed dataset, including all the attributes (Screenshot 8).



Now, please insert a Pivot Table to answer a few questions below:

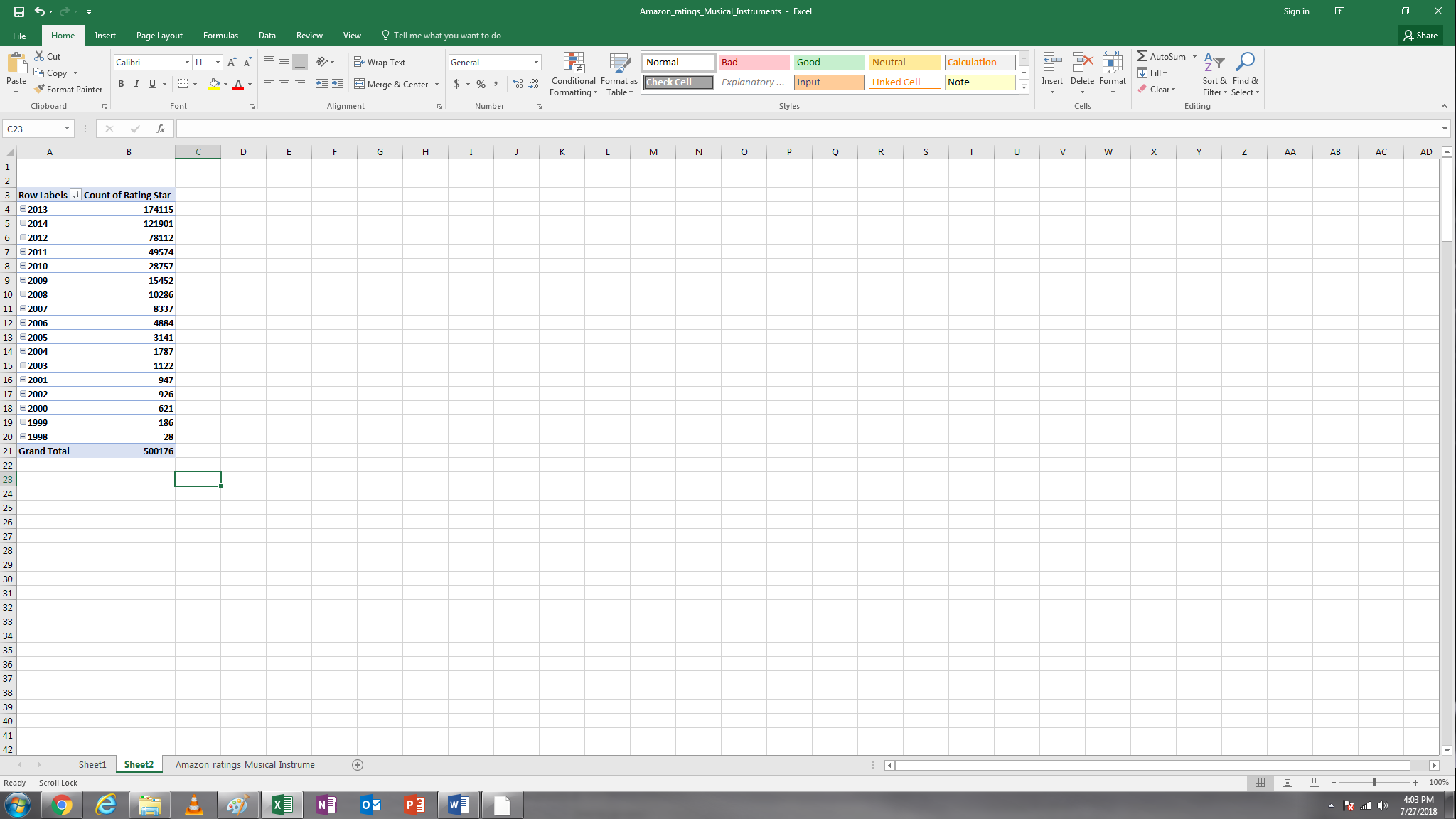
a) Which year generated the lowest average ratings (the attribute: rating star) of online reviews in the musical instrument category? 2003

What is the average (round to two decimal places)? 4.02

b) Which year generated the largest number of online reviews in the musical instrument category? 2013

How many reviews this year generated (please type the number here, please do not include any symbols such as commas)? 174115

Please take a screenshot of the PivotTable for this question (Screenshot 9).



c) In the musical instrument category, which product had the highest number of online reviews?

Copy and Paste the product ID here. B000ULAP4U

How many online reviews this product received (please type the number here, please do not include any symbols such as commas)? 3523

d) In year 2014, which product received the highest number of online reviews? B00FPPQYXM

How many reviews it received? 1168

Please take a screenshot of the PivotTable for this question (Screenshot 10).

