

Creating interactive NYSE Financial Statements with advanced Excel

This article describes how to analyze historic NYSE financial data in Excel and use advanced functions + visualizations to glean insights. The data subset is drawn from a large Kaggle dataset containing S&P 500 companies. Examples of advanced functions learned from this project include filtering through IF statements, INDEX, MATCH, dropdowns, data validation, and VLOOKUP. Additional visualizations and summary statistics were used to present business metrics and answer questions.

Questions we want answers to

- Determining the profit and loss statements of companies in the energy sector
- Evaluating the change in revenue of the energy sector between 2015 and 2014

My dataset: NYSE financial data

A subset from [Kaggle](#) NYSE from 2016 was used and further filtered to answer the questions above.

3 Step Approach: Gather Data, Clean, Get Insights

The following functions were used to clean and organize the data:

```
=DATE() used to partition out the year and group companies by year
```

```
=IF() to replace NULL values with a desired value
```

To determine the profit and loss statements the following equations were used:

```
Total Operating Expense = Selling, General and Administrative (SGA) expenses + Operating Expenses
```

```
Operating Income = Gross Profit - Total Operating Expenses
```

```
Net Income = Operating Income - (Interest and Taxes)
```

The following functions were used to determine the summary statistics of the dataset:

=AVERAGE() to determine mean for summary statistics

=MEDIAN() to determine the median

=MODE() to determine the mode

=STDEV.P() to determine the standard deviation
--

=QUARTILE() to determine the quartiles and InterQuartile Range (IQR)
--

=MIN(), =MAX() to determine the range of the data

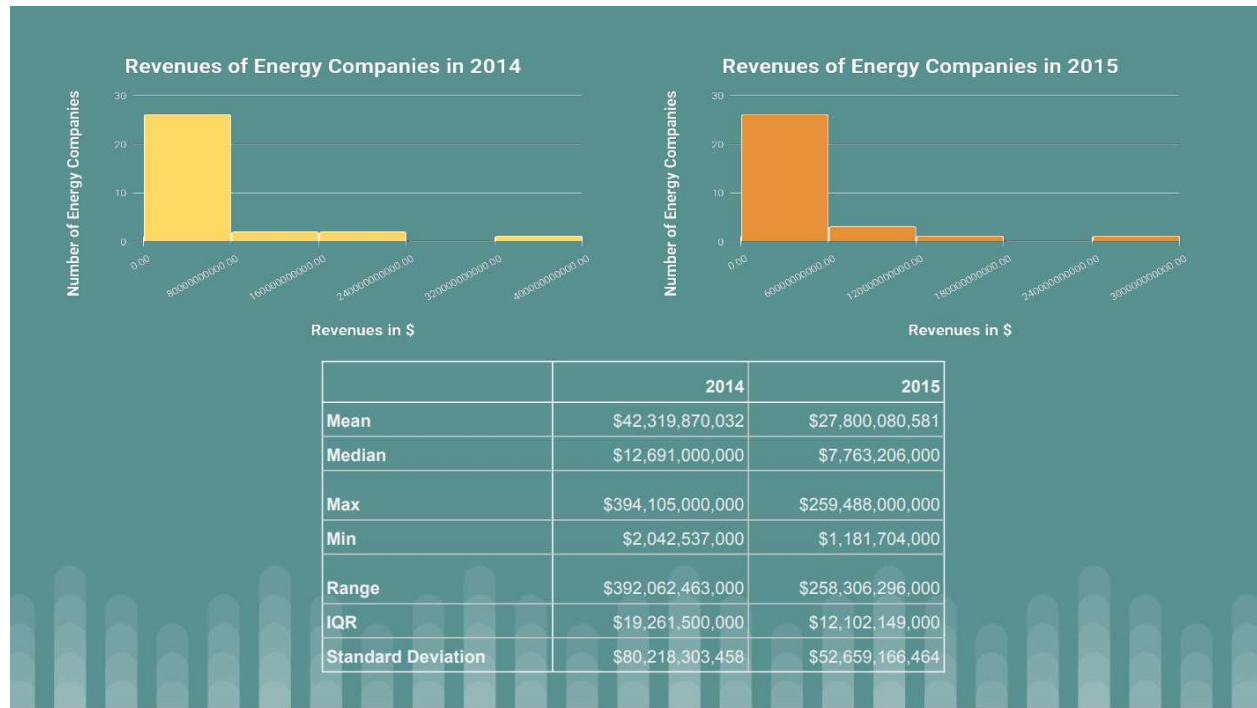
The following insights and visualizations were made as a result of the analysis:

How has the revenue changed in the energy sector from 2014 to 2015?

The revenues of companies in the energy sector averaged higher in 2014 (\$42,319,870,032) compared to revenues in 2015 (\$27,800,080,581). The median was also almost double in 2014 (\$12,691,000,000) compared to 2015 (\$7,763,206,000), showing that 50% of the companies had lower revenues in 2015. We can also see that the spread of the revenues in 2014 is larger than the spread of revenues in 2015. This can be seen through the range and standard deviation for each year. In 2014, the range and standard deviation was \$392,062,463,000 and \$80,218,303,458, respectively. In 2015, the range and standard deviation was \$258,306,296,000 and \$52,659,166,464, respectively. This shows that the energy sector did not produce as high of revenue as in 2014, but that there was a smaller difference of revenues within the sector in 2015.

Both years show an asymmetrical spread of revenues, with both being right skewed. This means that a majority of the companies have revenues within the first bin (up to \$80B in 2014 and up to \$60B in 2015), and that their means are greater than their medians.

See the histograms and table on the next slide. Additional sheets are included [here](#).



Lessons Learned

The biggest lesson learned through this project is just seeing how powerful Excel is in creating insights from small data sets. Given that the data is organized properly and is comprehensible, lots of data can be filtered, organized, and presented in an efficient manner.