

[Return to Classroom](#)[DISCUSS ON STUDENT HUB](#)

# Analyze NYSE Data

## REVIEW

## HISTORY

### Meets Specifications

Excellent project, you have a good understanding of using the correct formulas and assumptions to create a great financial model where required and creating good slides supported by visualizations with clear descriptions to explain the analysis using the measures of centre and spread to support the claims.

### Suggestions and tips for creating charts.

- Here are some suggestions for **do's and don'ts** when creating charts.
- Below are two links for creating histograms.
  - Using [Microsoft Excel](#)
  - Using [Google Sheets](#)
- Below are two links for creating box plots.
  - Using [Microsoft Excel](#)
  - Using [Google Sheets](#)

**Below are some related courses which could be of interest after completing this Nanodegree.**

- [Data Analysis](#) – Focuses on using Python, SQL and practical statistics
- [Marketing Analytics](#) – Includes work with Excel, Tableau, Google Analytics and Data Studio.
- [Predictive Analytics for Business](#) – Applying predictive analytics and business intelligence

I hope you enjoy the rest of the course.

## Submission Phase

A PDF report have been uploaded as part of a zipped folder.

Great job submitting the report in **PDF format**. It allows the reviewer to comment on your report as part of the reviewing process if required. It also shows your exceptional presentation skills. Or Excellent job with the slides/presentation!

Student provided an Excel file as part of a zipped folder or link to Google Sheet (in case the student used Google Sheets instead of Excel) necessary for review. This file should include their Profit and Loss statement and forecasts. The Google link should be included in the PDF or slides document.

The spreadsheet (Excel or Google Sheets) should contain individual tabs for the dataset, calculation of the summary statistics, dashboard for Profit and Loss statement, and Forecasting model with scenarios. There can be additional tabs in the Workbook that are needed for the dashboard and forecasting model.

Well done in this section! You have submitted an excel file with all the required tasks like **summary statistics**, **Profit and Loss Statement**, and **forecasts**. I like how you presented everything in the same excel sheet, which facilitates your management skills and professionalism.

ADBE				
Income Statement				
	Historical			
	Year 1	Year 2	Year 3	Year 4
Total Revenue	\$4,055,240,000	\$4,147,065,000	\$4,795,511,000	\$5,854,430,000
Cost of Goods Sold	\$586,557,000	\$622,080,000	\$744,317,000	\$819,908,000
Gross Profit	\$3,468,683,000	\$3,524,985,000	\$4,051,194,000	\$5,034,522,000
Sales, General and Admin	\$2,140,578,000	\$2,195,640,000	\$2,215,161,000	\$2,487,907,000
Research and Development	\$826,631,000	\$844,353,000	\$862,730,000	\$975,987,000
Other operating expenses	\$52,254,000	\$52,424,000	\$68,649,000	\$78,534,000
Total operating expenses	\$3,019,463,000	\$3,092,417,000	\$3,146,540,000	\$3,542,428,000
Operating income/ EBIT	\$449,220,000	\$432,568,000	\$904,654,000	\$1,492,094,000

Good effort with the profit and loss statement.

EBAY						
Income Statement						
	Historical				Forecast	
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Total Revenue	\$8,257,000,000	\$8,790,000,000	\$8,592,000,000	\$8,979,000,000	\$9,266,328,000	\$9,562,850,496
Cost of Goods Sold	\$1,492,000,000	\$1,663,000,000	\$1,771,000,000	\$2,007,000,000		
Gross Profit	\$6,765,000,000	\$7,127,000,000	\$6,821,000,000	\$6,972,000,000	\$10,192,960,800	\$10,519,135,546
Sales, General and Admin	\$3,260,000,000	\$3,593,000,000	\$3,660,000,000	\$3,499,000,000		
Research and Development	\$915,000,000	\$983,000,000	\$923,000,000	\$1,114,000,000		
Other operating expenses	\$136,000,000	\$75,000,000	\$41,000,000	\$34,000,000		
Total operating expenses	\$4,311,000,000	\$4,651,000,000	\$4,624,000,000	\$4,647,000,000		
Operating income/ EBIT	\$2,454,000,000	\$2,476,000,000	\$2,197,000,000	\$2,325,000,000	\$3,150,551,520	\$3,251,369,169
Operating Statistics						
	Historical			Assumptions		
Revenue growth (%)		6.5%	-2.3%	4.5%	3.2%	3.2%
Gross margin	0.82	0.81	0.79	0.78	1.10	1.10
Operating margin	0.30	0.28	0.26	0.26	0.34	0.34
Scenario	Strong case					
Operating Scenarios - sensitivity analysis						
Revenue growth (%)				Year 4	Year 5	
Strong case	1			3.2%	3.2%	
Base case	2			2.9%	2.9%	
Weak case	3			2.6%	2.6%	
Gross Margin						
Strong case	1			1.10	1.10	
Base case	2			0.80	0.80	
Weak case	3			0.50	0.50	
Operating Margin						
Strong case	1			0.34	0.34	
Base case	2			0.28	0.28	
Weak case	3			0.14	0.14	

Good effort with the financial forecast.

# Exploration of Summary Statistics

Student uses the measures of center and spread and at least one numeric summary statistic to generate insights.

Stating the summary statistics is insufficient. Please include in the written description a short insight related to each one.

For example here is an insight based on mean:

The mean total revenue for companies categorized under Pharmaceutical industry (\$26,325,440,909.09) was higher compared to mean total revenue for all healthcare industries (\$23,142,217,458.76). It looks like companies in the Pharmaceutical industry have a higher total revenue on average than all industries categorized under Health Care.

You have provided accurate **measures of center** and interpreted the insights effectively!

## Does the IT sector have similar expenditure levels for Research & Development than the industrials sector in Year 4?

In the first slide, you can find the histograms for the annual R & D expenses for companies in IT and Industrials sectors reported for all companies in Year 4.

Distributions for both sectors are right-skewed or positively skewed, as the mean for each is higher than the median.

The mean for IT (\$1.56 billion) is around 6.30 times higher the mean for Industrials (\$147.45 million). It seems companies in IT sector have a much higher R&D Expenditure on average than companies in Industrials sector.

The standard deviation for IT sector is \$2.82 billion, while the standard deviation for Industrials sector is \$749.87 million. The range for IT companies (\$12.74 billion) is around 2.75 times higher than the range for the Industrials (\$4,627 million). Therefore variability in R&D expenses for IT companies is much higher than Industrials companies.

To complete this specification, you must use the measure of center at least once to generate insights.

Good work here, you have used the **mean** and **median** to determine the distribution and average values.

### Here are some examples of other insights from fake data.

**The mode** (The most common value).

"For employed students the mode is 29, therefore based on this value the majority of the employed Udacity students are 29 years old."

**The mean** (The average).

Another example "The mean is 31.66 for students wanting start a new career and is 32.15 for students not starting a new career. From this value we can presume that the average of Udacity students is 32 years old."

**The median** (The centre value).

Another example "The median is 32, this states that 50% of the students are equal to or over the age of 32 years old."

You might find this article useful which helps illustrate the [measures of center](#) further. It contains some good explanations.

Student uses standard deviation and range to generate insights.

Stating the standard deviation and range is insufficient. Please include in the written description a short insight related to each one.

For example, please review the finished slide example in the classroom, which can be found in the Analyze NYSE S&P 500 dataset project lesson (Finished Example Slide).

Well done! Great job in meeting the specifications of this challenging rubric. It clearly shows your understanding of measures of spread. You have clearly identified the **range** and **standard deviation** of the data and provided

insights accordingly.

**Does the IT sector have similar expenditure levels for Research & Development than the industrials sector in Year 4?**

In the first slide, you can find the histograms for the annual R & D expenses for companies in IT and Industrials sectors reported for all companies in Year 4.

Distributions for both sectors are right-skewed or positively skewed, as the mean for each is higher than the median.

The mean for IT (\$1.56 billion) is around 6.30 times higher the mean for Industrials (\$147.45 million). It seems companies in IT sector have a much higher R&D Expenditure on average than companies in Industrials sector.

The standard deviation for IT sector is \$2.82 billion, while the standard deviation for Industrials sector is \$749.87 million. The range for IT companies (\$12.74 billion) is around 2.75 times higher than the range for the Industrials (\$4,627 million). Therefore variability in R&D expenses for IT companies is much higher than Industrials companies.

**Here are some examples of other insights from fake data.**

**The Standard Deviation** (The spread)  
 "The health care has a standard deviation of 15.6 whereas the energy sector is 25.8. From these values we can see that the energy sector is more spread than the health care sector."

**The Range** (The spread)  
 "The real estate sector has a higher range of 20.4 compared to the materials sector of 8.3. Its clear that the real estate sector is more spread than the materials sector."

To complete this specification, you must use the measure of spread to generate insights on how spread the data is.

Good insight for the **standard deviation** and **range** on the data variability.

This is an excellent article regarding the measures of spread which I think you will find useful.

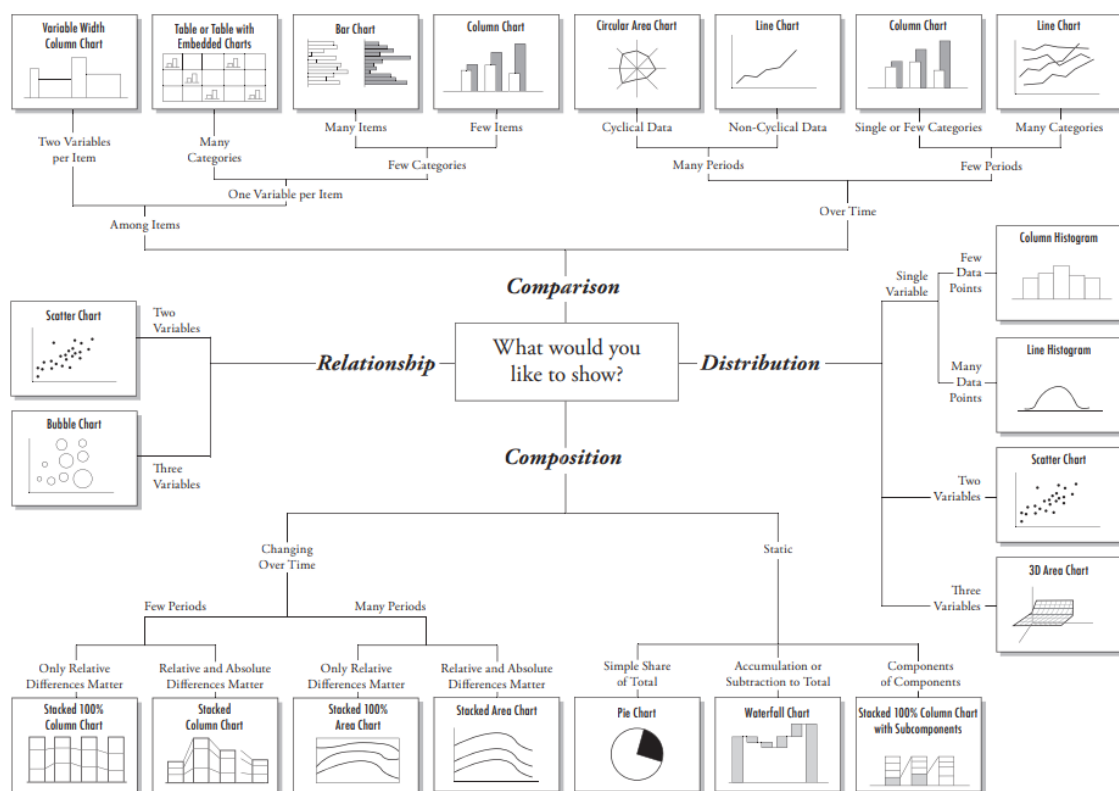
Student uses at least one plot to explore the data. The plots may include histograms, box-plots, scatterplots, and bar charts to explore data and gain insights.

All slides must contain a visualization. Screenshots of values in a table does not count.

Wonderful job! Your **histogram** plots look appropriate to understand the data more clearly.

Please see the image and [link](#) for further insights

## Chart Suggestions—A Thought-Starter



An appropriate visual is chosen to present the data. All labels are legible and the visual has appropriate axis labels.

Every visualization should have

- chart title (including which year's data the chart depicts)
- x axis title
- x axis labels
- y axis title
- y axis labels

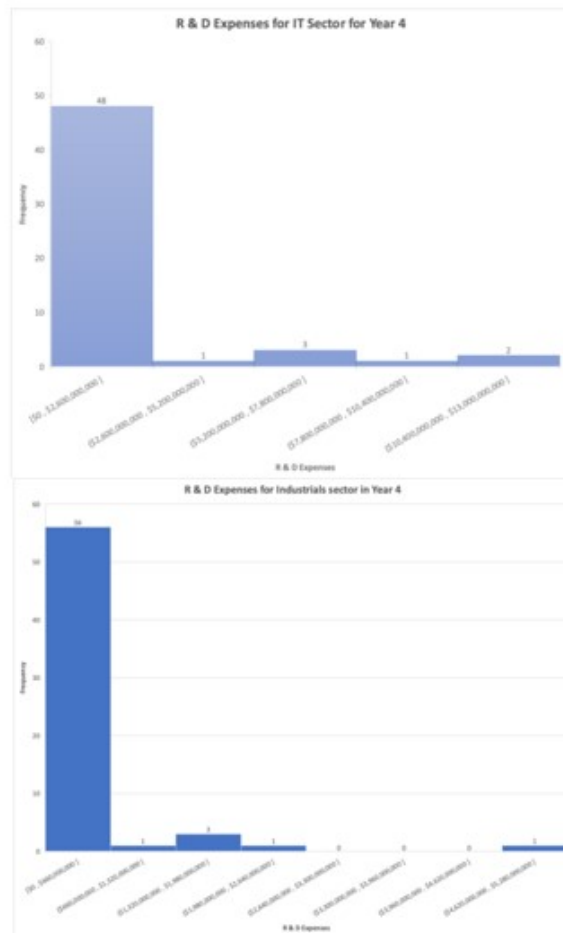
Please refer to the finished slide example page in the classroom for an example.

Good job on correctly labeling your visualizations. They convey insights very effectively.

Good histograms, well done.

You could try making the bin intervals labels smaller.

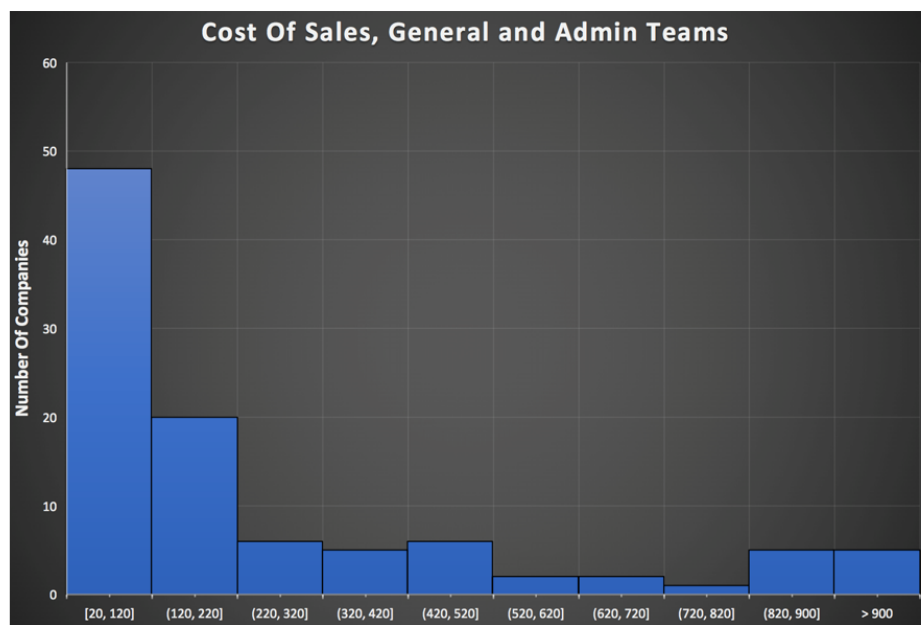
You could round the numbers down to the nearest billion such as [0, 2), [2, 4), [4, 6) etc then the axis title could be '**R&D in billions**'



Histogram example.

Here is a histogram with a quantitative variable on the x axis and the frequency on the y axis.

Notice all the intervals are rounded down to the nearest million to fit within the bins.



The link below is quite useful for general [do's and don'ts](#) when creating charts:

Below is a link for creating a line chart.

- Using [Microsoft Excel](#)

Below are two links for creating histograms.

- Using [Microsoft Excel](#)
- Using [Google Sheets](#)

Below are two links for creating box plots.

- Using [Microsoft Excel](#)
- Using [Google Sheets](#)

## Communication Phase

The results of the analysis are presented such that any limitations are clear. The analysis does not state or imply that one change causes another based solely on a correlation.

The results do not imply facts about a larger group of individuals based on descriptive values. Language is only applied to the specific data provided, unless a correct analysis beyond the course material is conducted that allows for inference.

Great work here. Your communication is on the point and clearly articulates your analysis and interpretations of data without making inferential or causal statements.

Here is a link to an article which goes into more detail regarding [correlation and causation](#)

The analysis associated with answering a particular question uses the appropriate variables, summary statistics, and plots that could provide an answer.

You have correctly provided all the descriptive statistics and interpreted them well. Your visualizations are on the mark, conveying your insights.



Below are some more images which illustrate which plots and statistics to use based on your selected variables

## Statistics

### Quantitative Variables

When describing quantitative variables, it is common to use the statistics discussed earlier:

1. Measures of center - mean, median, mode
2. Measures of spread - standard deviation, range, IQR

### Categorical Variables

However, when you are analyzing categorical variables, measures of center and spread **Do Not** make sense.

In cases of describing categorical variables, you need to use percentages or counts. Not means, medians, modes, standard deviations, or ranges.

### Important Last Thought

With this in mind, think of the variable type of the columns you are analyzing, and determine which plots and statistics make sense for your analysis.

### Plots You Can Use For Categorical Variables

If you have categorical data, here is a list of the possible univariate (one variable) plots you can make:

1. Bar Chart
2. Pie Chart

### Plots You Can Use For Quantitative Variables

If you have quantitative data, here is a list of the possible univariate plots you can make:

1. Histogram
2. Box Plot

### Plots to Compare 2 Variables

If you are interested in comparing two **quantitative** variables, then the main way to perform this comparison is with a scatterplot. However, if one of the variables is related to time, then a line plot is frequently used.

## Business Metrics

Student has input the correct formula for each business metric in the income statement and forecast model.

Student has built a forecast model for any company of choice. A dropdown for a company in the forecast



model is NOT required.

A job well done! All the formulas you used in the excel workbook are correct. I like the idea of how you used **INDEX**, **OFFSET** and **MATCH** functions to complete your **P & L statement** and **Forecast**. The dynamic company drop-down included in the **P&L statement** works well.

## Suggestions

- For further study, you could see this link for [Excel formulas](#), some formulas will be familiar while others will be new.
- For more cell formatting information please see [this link](#)

The student provides appropriate assumptions based on gross margin, revenue growth and operating margin for the financial model scenarios.

Excellent work here! Not many students understand how to make **proper assumptions**, but your assumptions in the forecast are on the mark. You calculated the assumptions based on the trends in the operating statistics. That shows your grip on the subject. Also, the forecast values change based on your assumptions - good work there!

EBAY							
Income Statement							
	Historical				Forecast		
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	
Total Revenue	\$8,257,000,000	\$8,790,000,000	\$8,592,000,000	\$8,979,000,000	\$9,266,328,000	\$9,562,850,496	
Cost of Goods Sold	\$1,492,000,000	\$1,663,000,000	\$1,771,000,000	\$2,007,000,000			
Gross Profit	\$6,765,000,000	\$7,127,000,000	\$6,821,000,000	\$6,972,000,000	\$10,192,960,800	\$10,519,135,546	
Sales, General and Admin	\$3,260,000,000	\$3,593,000,000	\$3,660,000,000	\$3,499,000,000			
Research and Development	\$915,000,000	\$983,000,000	\$923,000,000	\$1,114,000,000			
Other operating expenses	\$136,000,000	\$75,000,000	\$41,000,000	\$34,000,000			
Total operating expenses	\$4,311,000,000	\$4,651,000,000	\$4,624,000,000	\$4,647,000,000			
Operating Income/ EBIT	\$2,454,000,000	\$2,476,000,000	\$2,197,000,000	\$2,325,000,000	\$3,150,551,520	\$3,251,369,169	
Operating Statistics							
	Historical			assumptions			
Revenue growth (%)		6.5%	-2.3%	4.5%	3.2%	3.2%	
Gross margin	0.82	0.81	0.79	0.78	1.10	1.10	
Operating margin	0.30	0.28	0.26	0.26	0.34	0.34	
Scenario	Strong case						
Operating Scenarios - sensitivity analysis							
Revenue growth (%)				Year 4	Year 5		
Strong case	1			3.2%	3.2%	3.2%	
Base case	2			2.9%	2.9%	2.9%	
Weak case	3			2.6%	2.6%	2.6%	
Gross Margin							
Strong case	1			1.10	1.10	1.10	
Base case	2			0.80	0.80	0.80	
Weak case	3			0.50	0.50	0.50	
Operating Margin							
Strong case	1			0.34	0.34	0.34	
Base case	2			0.28	0.28	0.28	
Weak case	3			0.14	0.14	0.14	

Good assumptions here based on the historical data.

Another way of doing this is taking the **AVERAGE** for your base values and then using the **STDEV** for the weak and strong assumptions.

## Suggestion

- If there is **not** an obvious trend in the data then you can use the **AVERAGE** and **STDEV** formulas to help predict the year 5 and 6 cases. For example, you could do something like this:
  - Strong case = **AVERAGE**(CellRange) + **STDEV**(CellRange)
  - Base case = **AVERAGE**(CellRange)
  - Weak case = **AVERAGE**(CellRange) - **STDEV**(CellRange)

## Excel Functions and Modeling

Student demonstrates using VLOOKUP or INDEX and MATCH statements. The student can use the appropriate functions such as OFFSET and MATCH to create forecast scenarios.

Your work shows your excellent understanding of various excel functions. You have used INDEX, OFFSET and MATCH functions excellently!

## Suggestion

You might find this [link](#) useful on using INDEX and MATCH functions.

ADBE				
Income Statement				
	Historical			
	Year 1	Year 2	Year 3	Year 4
Total Revenue	\$4,055,240,000	\$4,147,065,000	\$4,795,511,000	\$5,854,430,000
Cost of Goods Sold	\$586,557,000	\$622,080,000	\$744,317,000	\$819,908,000
Gross Profit	\$3,468,683,000	\$3,524,985,000	\$4,051,194,000	\$5,034,522,000
Sales, General and Admin	\$2,140,578,000	\$2,195,640,000	\$2,215,161,000	\$2,487,907,000
Research and Development	\$826,631,000	\$844,353,000	\$862,730,000	\$975,987,000
Other operating expenses	\$52,254,000	\$52,424,000	\$68,649,000	\$78,534,000
Total operating expenses	\$3,019,463,000	\$3,092,417,000	\$3,146,540,000	\$3,542,428,000
Operating income/ EBIT	\$449,220,000	\$432,568,000	\$904,654,000	\$1,492,094,000

EBAY						
Income Statement						
	Historical			Forecast		
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Total Revenue	\$8,257,000,000	\$8,790,000,000	\$8,592,000,000	\$8,979,000,000	\$9,266,328,000	\$9,562,850,496
Cost of Goods Sold	\$1,492,000,000	\$1,663,000,000	\$1,771,000,000	\$2,007,000,000		
Gross Profit	\$6,765,000,000	\$7,127,000,000	\$6,821,000,000	\$6,972,000,000	\$10,192,960,800	\$10,519,135,546
Sales, General and Admin	\$3,260,000,000	\$3,593,000,000	\$3,660,000,000	\$3,499,000,000		
Research and Development	\$915,000,000	\$983,000,000	\$923,000,000	\$1,114,000,000		
Other operating expenses	\$136,000,000	\$75,000,000	\$41,000,000	\$34,000,000		
Total operating expenses	\$4,311,000,000	\$4,651,000,000	\$4,624,000,000	\$4,647,000,000		
Operating income/ EBIT	\$2,454,000,000	\$2,476,000,000	\$2,197,000,000	\$2,325,000,000	\$3,150,551,520	\$3,251,369,169
Operating Statistics						
	Historical			Assumptions		
Revenue growth (%)		6.5%	-2.3%	4.5%	3.2%	3.2%
Gross margin	0.82	0.81	0.79	0.78	1.10	1.10
Operating margin	0.30	0.28	0.26	0.26	0.34	0.34
Scenario	Strong case					
Operating Scenarios - sensitivity analysis						
Revenue growth (%)				Year 4	Year 5	
Strong case	1			3.2%	3.2%	
Base case	2			2.9%	2.9%	
Weak case	3			2.6%	2.6%	
Gross Margin						
Strong case	1			1.10	1.10	
Base case	2			0.80	0.80	
Weak case	3			0.50	0.50	
Operating Margin						
Strong case	1			0.34	0.34	
Base case	2			0.28	0.28	
Weak case	3			0.14	0.14	

Excellent job with using the INDEX, OFFSET and MATCH functions.

[DOWNLOAD PROJECT](#)

[RETURN TO PATH](#)

Rate this review

START



---