

Task_1_Analysis

April 4, 2024

Importing pandas library to manipulate data to extract meaningful data from the extracted data

```
[1]: import pandas as pd
```

```
[2]: df = pd.read_excel(r"C:\Users\Bikash shah\Desktop\Data_□  
↳Extraction\Task-1\Final_Extracted_Data.xlsx")
```

```
[3]: df
```

```
[3]:   Year  Company  Total Revenue  Net Income  Total Assets \  
0  2023  Microsoft      211915      72361      411976  
1  2022  Microsoft      198270      72738      364840  
2  2021  Microsoft      168088      61271      333779  
3  2023      Tesla       96773      14974      106618  
4  2022      Tesla       81462      12556       82338  
5  2021      Tesla       53821       5519       62131  
6  2023      Apple      383285      96995      352583  
7  2022      Apple      394328      99803      352755  
8  2021      Apple      365817      94680      351002
```

	Total Liabilities	Cash Flow from Operating Activities
0	205753	87582
1	198298	89035
2	191791	76740
3	43009	13256
4	36440	14724
5	30548	11497
6	290437	110543
7	302083	122151
8	287912	104038

Checking is there any null value

```
[5]: df.isnull().sum()
```

```
[5]: Year          0  
Company          0  
Total Revenue    0  
Net Income       0
```

```
Total Assets          0
Total Liabilities      0
Cash Flow from Operating Activities  0
dtype: int64
```

Calculating Year-by-Year growth rates for Total Revenue and Net Income with the help of pct_change() method which returns a DataFrame with the percentage difference between the values for each row and, by default, the previous row

```
[6]: df['Revenue Growth (%)'] = df.groupby(['Company'])['Total Revenue'].
      ↪pct_change() * 100
df['Net Income Growth (%)'] = df.groupby(['Company'])['Net Income'].
      ↪pct_change() * 100
```

```
[7]: df
```

```
[7]:   Year  Company  Total Revenue  Net Income  Total Assets  \
0  2023  Microsoft      211915      72361      411976
1  2022  Microsoft      198270      72738      364840
2  2021  Microsoft      168088      61271      333779
3  2023    Tesla       96773      14974      106618
4  2022    Tesla       81462      12556       82338
5  2021    Tesla       53821       5519       62131
6  2023    Apple      383285      96995      352583
7  2022    Apple      394328      99803      352755
8  2021    Apple      365817      94680      351002
```

```
      Total Liabilities  Cash Flow from Operating Activities  Revenue Growth (%)  \
0          205753              87582                NaN
1          198298              89035             -6.438902
2          191791              76740            -15.222676
3           43009              13256                NaN
4           36440              14724            -15.821562
5           30548              11497            -33.931158
6          290437             110543                NaN
7          302083             122151              2.881146
8          287912             104038             -7.230275
```

```
      Net Income Growth (%)
0                NaN
1           0.520999
2          -15.764800
3                NaN
4          -16.147990
5          -56.044919
6                NaN
7           2.894995
8          -5.133112
```

```
[8]: df.isnull().sum()
```

```
[8]: Year                0
     Company            0
     Total Revenue      0
     Net Income         0
     Total Assets       0
     Total Liabilities  0
     Cash Flow from Operating Activities  0
     Revenue Growth (%)  3
     Net Income Growth (%)  3
     dtype: int64
```

Removing Null/NaN values with 0

```
[9]: df.fillna(0, inplace=True)
```

Calculating Year-by-Year growth rates for Total Assets, Total Liabilities and Cash flow from Operations Activities

```
[10]: df['Assets Growth (%)'] = df.groupby(['Company'])['Total Assets'].pct_change()
      ↪ * 100
      df['Liabilities Growth (%)'] = df.groupby(['Company'])['Total Liabilities'].
      ↪ pct_change() * 100
      df['Cash Flow from Operations Growth(%)'] = df.groupby(['Company'])['Cash Flow
      ↪ from Operating Activities'].pct_change() * 100
```

```
[11]: df
```

```
[11]:   Year  Company  Total Revenue  Net Income  Total Assets  \
0  2023  Microsoft      211915      72361      411976
1  2022  Microsoft      198270      72738      364840
2  2021  Microsoft      168088      61271      333779
3  2023      Tesla       96773      14974      106618
4  2022      Tesla       81462      12556       82338
5  2021      Tesla       53821       5519       62131
6  2023      Apple      383285      96995      352583
7  2022      Apple      394328      99803      352755
8  2021      Apple      365817      94680      351002

      Total Liabilities  Cash Flow from Operating Activities  Revenue Growth (%)  \
0              205753                87582              0.000000
1              198298                89035             -6.438902
2              191791                76740            -15.222676
3               43009                13256              0.000000
4               36440                14724            -15.821562
5               30548                11497            -33.931158
6              290437               110543              0.000000
```

7	302083	122151	2.881146
8	287912	104038	-7.230275

	Net Income Growth (%)	Assets Growth (%)	Liabilities Growth (%)	\
0	0.000000	NaN	NaN	
1	0.520999	-11.441443	-3.623276	
2	-15.764800	-8.513595	-3.281425	
3	0.000000	NaN	NaN	
4	-16.147990	-22.772890	-15.273547	
5	-56.044919	-24.541524	-16.169045	
6	0.000000	NaN	NaN	
7	2.894995	0.048783	4.009820	
8	-5.133112	-0.496945	-4.691095	

	Cash Flow from Operations Growth(%)
0	NaN
1	1.659017
2	-13.809176
3	NaN
4	11.074231
5	-21.916599
6	NaN
7	10.500891
8	-14.828368

```
[12]: df.isnull().sum()
```

```
[12]: Year          0
      Company      0
      Total Revenue  0
      Net Income    0
      Total Assets  0
      Total Liabilities  0
      Cash Flow from Operating Activities  0
      Revenue Growth (%)  0
      Net Income Growth (%)  0
      Assets Growth (%)  3
      Liabilities Growth (%)  3
      Cash Flow from Operations Growth(%)  3
      dtype: int64
```

```
[13]: df.fillna(0, inplace=True)
```

```
[14]: df
```

```
[14]:   Year  Company  Total Revenue  Net Income  Total Assets  \
0  2023  Microsoft         211915         72361         411976
```

1	2022	Microsoft	198270	72738	364840
2	2021	Microsoft	168088	61271	333779
3	2023	Tesla	96773	14974	106618
4	2022	Tesla	81462	12556	82338
5	2021	Tesla	53821	5519	62131
6	2023	Apple	383285	96995	352583
7	2022	Apple	394328	99803	352755
8	2021	Apple	365817	94680	351002

	Total Liabilities	Cash Flow from Operating Activities	Revenue Growth (%)	\
0	205753	87582	0.000000	
1	198298	89035	-6.438902	
2	191791	76740	-15.222676	
3	43009	13256	0.000000	
4	36440	14724	-15.821562	
5	30548	11497	-33.931158	
6	290437	110543	0.000000	
7	302083	122151	2.881146	
8	287912	104038	-7.230275	

	Net Income Growth (%)	Assets Growth (%)	Liabilities Growth (%)	\
0	0.000000	0.000000	0.000000	
1	0.520999	-11.441443	-3.623276	
2	-15.764800	-8.513595	-3.281425	
3	0.000000	0.000000	0.000000	
4	-16.147990	-22.772890	-15.273547	
5	-56.044919	-24.541524	-16.169045	
6	0.000000	0.000000	0.000000	
7	2.894995	0.048783	4.009820	
8	-5.133112	-0.496945	-4.691095	

	Cash Flow from Operations Growth(%)
0	0.000000
1	1.659017
2	-13.809176
3	0.000000
4	11.074231
5	-21.916599
6	0.000000
7	10.500891
8	-14.828368

Exporting the dataframe to excel file

```
[16]: df.to_excel("Final_Data_Report.xlsx")
```

Dataframe of all the summary of all the keys Year-by-Year growth rates

```
[19]: summary = df.groupby(['Company']).agg({
    'Revenue Growth (%)': 'mean',
    'Net Income Growth (%)': 'mean',
    'Assets Growth (%)' : 'mean',
    'Liabilities Growth (%)' : 'mean',
    'Cash Flow from Operations Growth(%)' : 'mean'
}).reset_index()
```

```
[20]: summary
```

```
[20]:
```

	Company	Revenue Growth (%)	Net Income Growth (%)	Assets Growth (%)	\
0	Apple	-1.449710	-0.746039	-0.149388	
1	Microsoft	-7.220526	-5.081267	-6.651679	
2	Tesla	-16.584240	-24.064303	-15.771471	

	Liabilities Growth (%)	Cash Flow from Operations Growth(%)
0	-0.227092	-1.442492
1	-2.301567	-4.050053
2	-10.480864	-3.614123

```
[21]: summary.to_excel("Summary_of_Final_Data_Report.xlsx")
```

```
[ ]:
```