



# **Data analysis using Python 2024**

# Introduction

In the world of investment businesses, understanding financial figures and indicators goes beyond mere data analysis—it's a journey to uncover the story behind the numbers. This project focused on analyzing the financial performance of twenty investment companies over five years, navigating through revenues and costs to uncover strategies of success and identify challenges.

The primary goal of this project was to **understand financial patterns, reveal hidden relationships between indicators** such as revenues, profits, and the debt-to-equity ratio.

Through this analysis, we aimed to answer key questions such as:

- What factors drive revenue growth?
- How can companies improve their profitability by managing costs?
- How stable is the financial health of companies based on debt and asset ratios?

## Key Findings:

Using advanced Python tools for an in-depth analysis, we achieved the following:

1. Identified high-performing companies based on revenue and profit growth.
2. Uncovered factors that enhance profit margins and reduce operating costs.
3. Gained insights into how the debt-to-equity ratio affects corporate sustainability.
4. Predicted future trends for revenues and profits, paving the way for new investment opportunities.

# Data Analysis

## ✓ Data Description

The dataset used in this project comprises financial information from the quarterly performance of twenty investment companies over five years. The dataset includes the following columns:

1. **Revenue (USD):** Total revenue generated by each company.
2. **Net Profit (USD):** Net profit after deducting all costs and taxes.
3. **Operating Costs (USD):** Operating expenses, including daily business costs.
4. **Cash Flow (USD):** Cash flow used to evaluate liquidity.
5. **Total Debt (USD):** Total financial liabilities of the company.
6. **Shareholders' Equity (USD):** Shareholders' equity, reflecting the company's net value.
7. **Assets (USD):** The company's assets, including fixed and current assets.
8. **Liabilities (USD):** The financial obligations of the company.
9. **Profit Margin:** The profit margin calculated as the ratio of net profit to revenue.
10. **Debt to Equity Ratio:** The ratio of total debt to shareholders' equity.
11. **Return on Assets (ROA):** Return on assets.
12. **Return on Equity (ROE):** Return on shareholders' equity.

## ✓ **Financial Trends:**

Several notable financial trends were identified:

### **1. Revenue and Profit Trends Over Time:**

- Major companies showed steady increases in revenue over the five years, with a significant rise in 2021 due to market recovery after the economic slowdown.

### **2. Profit Margin:**

- Companies with relatively lower operating costs achieved higher profit margins, indicating effective cost management.

### **3. Debt-to-Equity Ratio:**

- Analysis revealed significant variation among companies. Companies with lower ratios demonstrated better financial stability compared to those with higher ratios.

### **4. Return on Assets (ROA) and Return on Equity (ROE):**

- Companies that invested in highly productive assets recorded higher returns, indicating efficient resource utilization.

### **5. Revenue and Profit Growth:**

- Certain companies exhibited consistent annual growth rates exceeding 10%, making them attractive for investment.

## 1. Loading and Exploring the Data

### Objective:

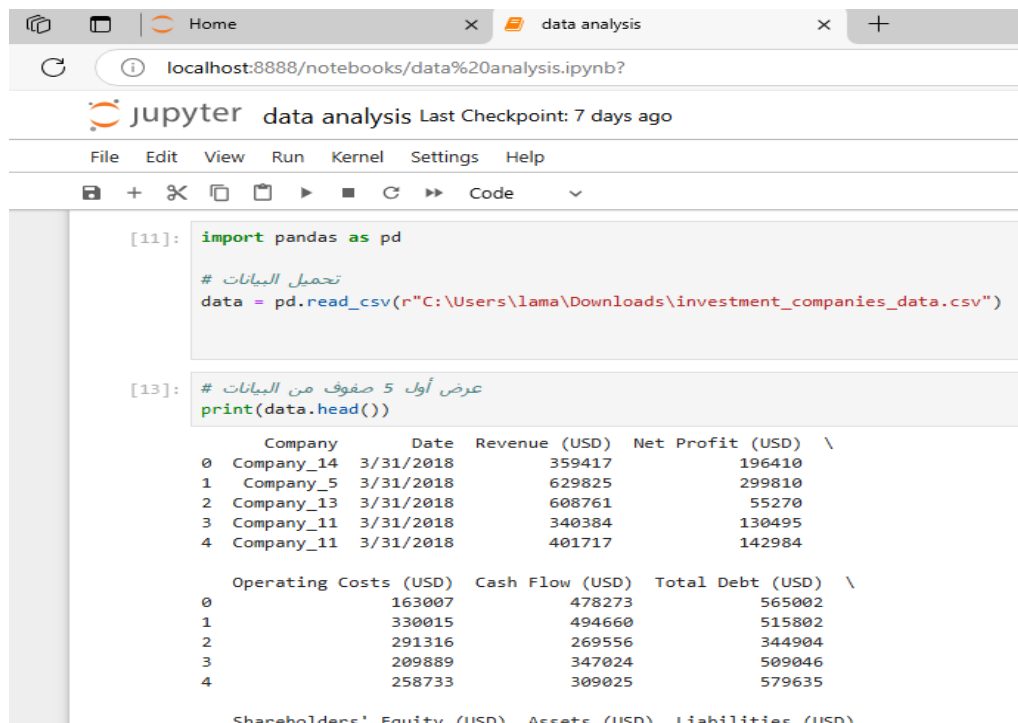
- Ensure the data is loaded correctly.
- Display its initial details.

Imagine opening a new book and skimming through the first few chapters to understand its content. Similarly, we'll begin our journey by exploring the structure and initial details of our dataset. Let's see what these numbers have to reveal.

### What will happen?

1. **First Look:** You'll see the first few rows to understand the dataset's general layout.
2. **Structure:** You'll learn about the columns and their data types (text, numbers, dates).
3. **Key Insights:** A quick glance at key metrics such as averages, minimums, and maximums.

This step allows you to get acquainted with the "story of the data" before diving deeper into the analysis.



The screenshot shows a Jupyter Notebook window titled "data analysis" with the URL "localhost:8888/notebooks/data%20analysis.ipynb?". The notebook has a menu bar (File, Edit, View, Run, Kernel, Settings, Help) and a toolbar with icons for file operations and execution. The code cell [11] contains the following Python code:

```
[11]: import pandas as pd

# تحميل البيانات
data = pd.read_csv(r"C:\Users\lama\Downloads\investment_companies_data.csv")
```

The code cell [13] contains the following Python code:

```
[13]: # عرض أول 5 صفوف من البيانات
print(data.head())
```

The output of the code is a table with 5 rows and 4 columns. The columns are Company, Date, Revenue (USD), and Net Profit (USD). The rows are indexed 0 to 4.

	Company	Date	Revenue (USD)	Net Profit (USD)
0	Company_14	3/31/2018	359417	196410
1	Company_5	3/31/2018	629825	299810
2	Company_13	3/31/2018	608761	55270
3	Company_11	3/31/2018	340384	130495
4	Company_11	3/31/2018	401717	142984

The output also includes a table with 5 rows and 3 columns. The columns are Operating Costs (USD), Cash Flow (USD), and Total Debt (USD). The rows are indexed 0 to 4.

	Operating Costs (USD)	Cash Flow (USD)	Total Debt (USD)
0	163007	478273	565002
1	330015	494660	515802
2	291316	269556	344904
3	209889	347024	509046
4	258733	309025	579635

The output also includes a table with 5 rows and 3 columns. The columns are Shareholders' Equity (USD), Assets (USD), and Liabilities (USD). The rows are indexed 0 to 4.

	Shareholders' Equity (USD)	Assets (USD)	Liabilities (USD)
0	163007	478273	565002
1	330015	494660	515802
2	291316	269556	344904
3	209889	347024	509046
4	258733	309025	579635

```
[15]: # عرض معلومات عن البيانات
print(data.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 400 entries, 0 to 399
Data columns (total 10 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   Company              400 non-null    object
 1   Date                  400 non-null    object
 2   Revenue (USD)         400 non-null    int64
 3   Net Profit (USD)      400 non-null    int64
 4   Operating Costs (USD) 400 non-null    int64
 5   Cash Flow (USD)       400 non-null    int64
 6   Total Debt (USD)      400 non-null    int64
 7   Shareholders' Equity (USD) 400 non-null    int64
 8   Assets (USD)          400 non-null    int64
 9   Liabilities (USD)     400 non-null    int64
dtypes: int64(8), object(2)
memory usage: 31.4+ KB
None
```

```
[17]: # وصف إحصائي للبيانات
print(data.describe())

Revenue (USD)  Net Profit (USD)  Operating Costs (USD)  \
count         400.000000         400.000000         400.000000
```

```
[17]: # وصف إحصائي للبيانات
print(data.describe())


Revenue (USD)  Net Profit (USD)  Operating Costs (USD)  \
count         400.000000         400.000000         400.000000
mean         556317.012500         267806.565000         230881.512500
std          248109.499119         132901.539691         193988.561568
min          103019.000000          51210.000000           0.000000
25%          345875.250000         143938.500000         67699.750000
50%          565843.500000         267071.000000         186822.500000
75%          756371.000000         385841.000000         369921.750000
max          998393.000000         495165.000000         769124.000000

Cash Flow (USD)  Total Debt (USD)  Shareholders' Equity (USD)  \
count         400.000000         400.000000         400.000000
mean          323721.127500         355360.995000         598273.230000
std          162203.339004         159157.740688         224528.258896
min           52455.000000         100434.000000         202998.000000
25%          174677.750000         224038.250000         403423.250000
50%          328702.000000         334782.500000         605469.000000
75%          467339.500000         470336.000000         781083.250000
max          596528.000000         691588.000000         996868.000000
```

## 2. Cleaning the Data

### Objective:


- Remove missing and duplicate values.
- Correct any inconsistent or invalid values

 **jupyter** data analysis Last Checkpoint: 7 days ago

---

File Edit View Run Kernel Settings Help

---

 Code ▾

---

```
max 4.953066e+06 3.982189e+06
```

```
[19]: # التحقق من القيم المفقودة
print(data.isnull().sum())
```

```
Company      0
Date          0
Revenue (USD) 0
Net Profit (USD) 0
Operating Costs (USD) 0
Cash Flow (USD) 0
Total Debt (USD) 0
Shareholders' Equity (USD) 0
Assets (USD) 0
Liabilities (USD) 0
dtype: int64
```

```
[29]: # إزالة القيم المكررة
data = data.drop_duplicates()
```

```
[31]: # إزالة الصفوف ذات القيم غير المنطقية (مثال: الإيرادات السلبية)
data = data[data['Revenue (USD)'] > 0]
```

```
[33]: # حفظ البيانات النظيفة
data.to_csv('cleaned_investment_data.csv', index=False)
```

### 3. Analyzing Temporal Trends

#### Objective:

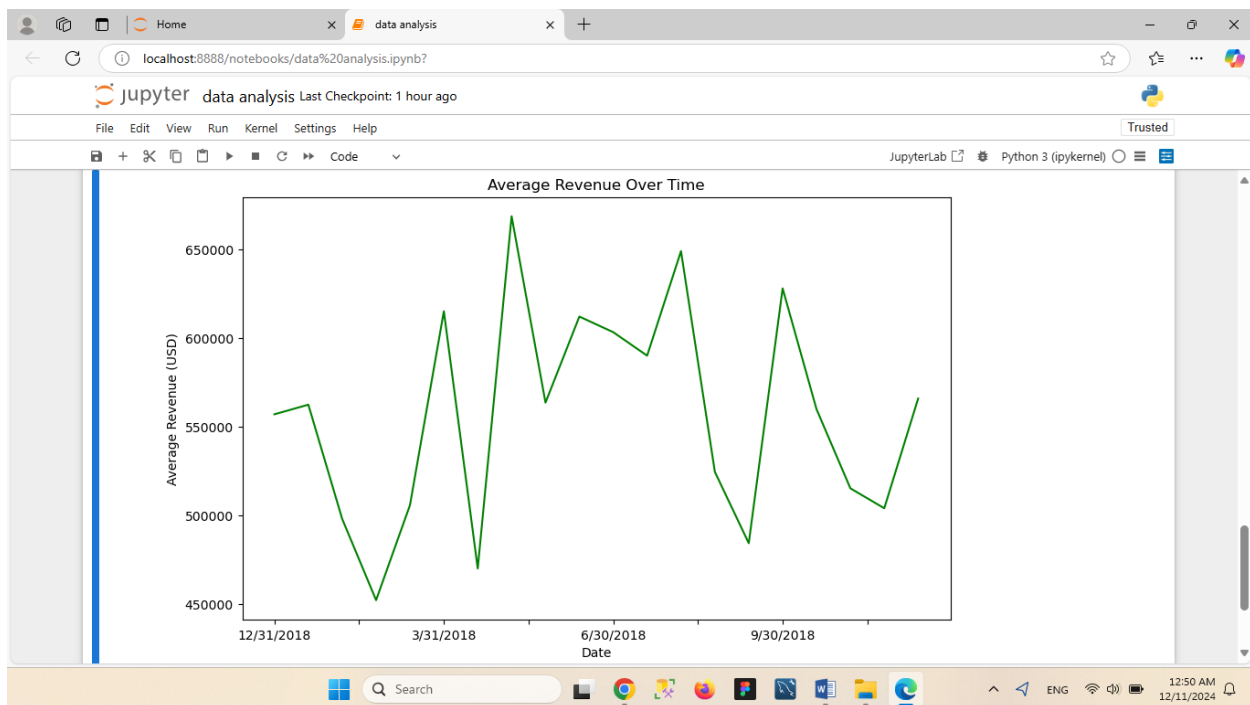
- Analyze revenue and profit over time

```
[35]: import matplotlib.pyplot as plt

# الإيرادات المتوسطة لكل ربع سنة
average_revenue = data.groupby('Date')['Revenue (USD)'].mean()

[36]: # رسم الاتجاه الزمني للإيرادات
plt.figure(figsize=(10, 6))
average_revenue.plot(kind='line', color='green', title='Average Revenue Over Time')
plt.xlabel('Date')
plt.ylabel('Average Revenue (USD)')
plt.show()
```

Average Revenue Over Time





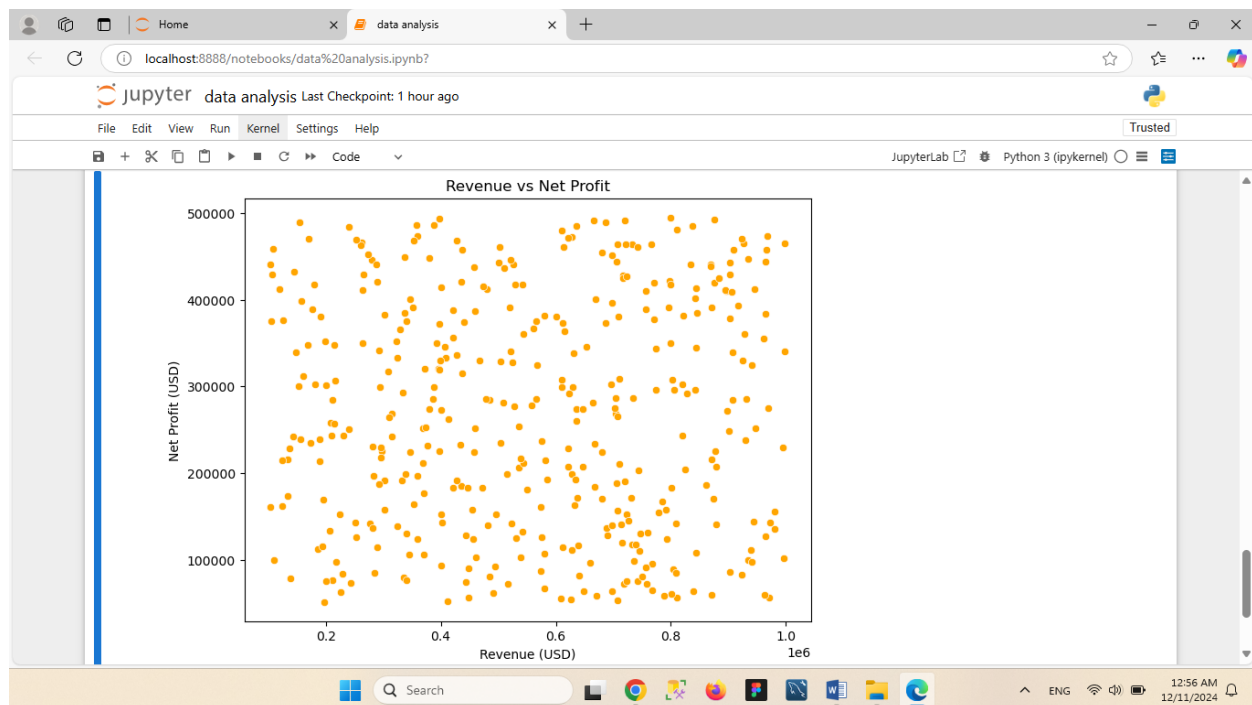
## 4. Analyzing Relationships Between Variables

### Objective:

- Uncover the relationship between revenue and net profit.

```
[41]: import seaborn as sns

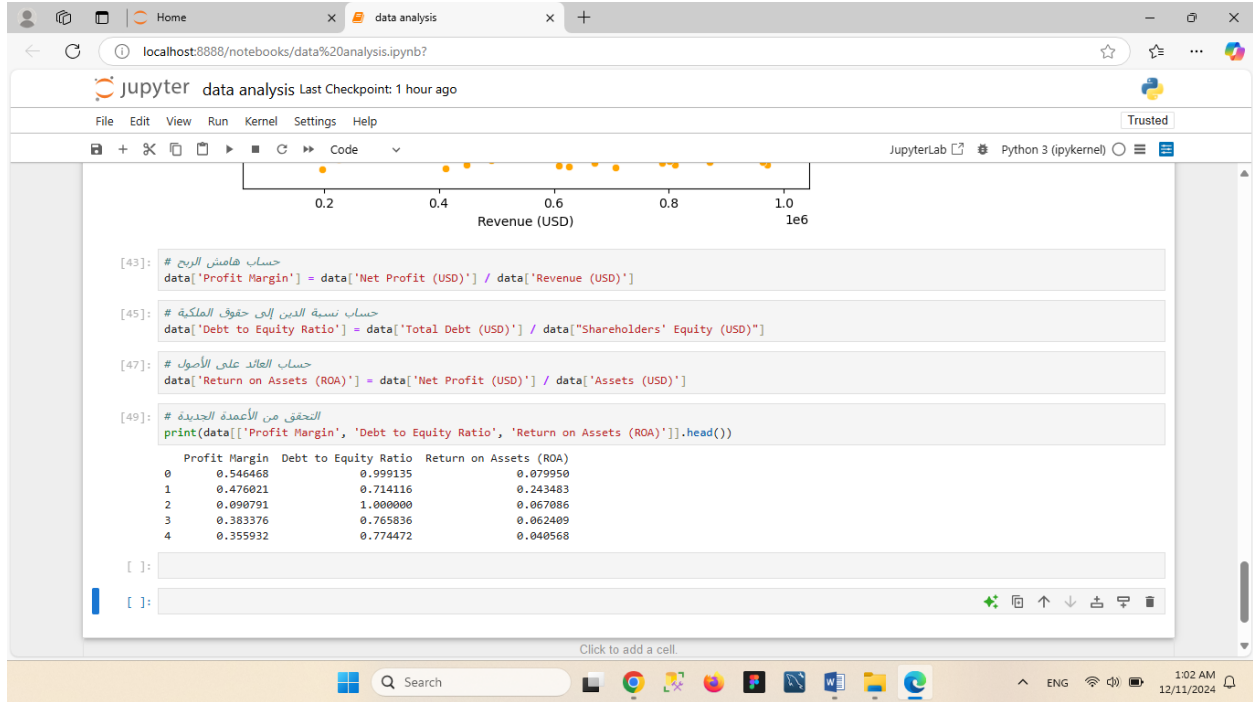
# رسم العلاقة بين الإيرادات وصافي الأرباح
plt.figure(figsize=(8, 6))
sns.scatterplot(x='Revenue (USD)', y='Net Profit (USD)', data=data, color='orange')
plt.title('Revenue vs Net Profit')
plt.xlabel('Revenue (USD)')
plt.ylabel('Net Profit (USD)')
plt.show()
```



## 5. Calculating Financial Indicators

### Objective:

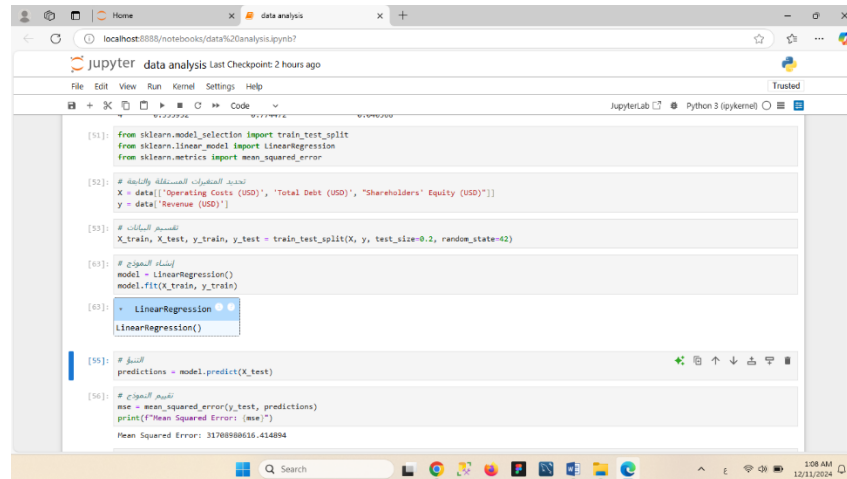
- Create new columns for financial indicators such as profit margin and debt-to-equity ratio.



## 6. Prediction Using Machine Learning

### Objective:

- Forecast future revenues using the linear regression algorithm.



```
[51]: from sklearn.model_selection import train_test_split
      from sklearn.linear_model import LinearRegression
      from sklearn.metrics import mean_squared_error

[52]: # تحديد المتغيرات المستقلة والمتابعة
      X = data[['Operating Costs (USD)', 'Total Debt (USD)', 'Shareholders' Equity (USD)']]
      y = data['Revenue (USD)']

[53]: # تقسيم البيانات
      X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

[54]: # إنشاء النموذج
      model = LinearRegression()
      model.fit(X_train, y_train)

[55]: # LinearRegression
      LinearRegression()

[56]: # التنبؤ
      predictions = model.predict(X_test)

[57]: # حساب الخطأ
      mse = mean_squared_error(y_test, predictions)
      print(f"Mean Squared Error: {mse}")
      Mean Squared Error: 31708980616.414894
```

### Analytical Dimensions of the Project

In the world of financial analysis, data is a hidden language that unveils the performance and competitive positioning of companies. This project aims to explore various dimensions of the performance of investment companies using visual analysis tools. Through a variety of charts, we shed light on key aspects such as revenue distribution, the relationship between profits and expenses, and cash flow patterns.

Each chart in this project is not merely a visual representation but a tool to uncover trends, answer critical questions, and provide actionable insights that support strategic decision-making. These analyses address questions such as: Which companies are the top performers? What is the correlation between profits and debt? How can expense management be optimized?

This project demonstrates the power of visual analysis in transforming raw data into compelling stories, enabling a deeper understanding of corporate financial performance and driving strategic thinking for improvement.

- ✓ **Which companies are generating the highest profits, and what could be the underlying factors contributing to their financial performance?**

This chart analyzes the profit generated by each company, aiming to identify the top-performing companies and explore potential strategies they employ. It provides insights into how companies are managing their resources to achieve profitability.

Imagine you are comparing the profitability of five companies. Each bar in the chart tells a story about how well a company is doing financially. The taller the bar, the more profit the company has made. By visualizing these differences, we can start asking critical questions: Why is Company C outperforming others? Is it due to lower expenses, higher revenue, or better financial management? Understanding this can guide other companies to improve their strategies and bridge the performance gap.



- ✓ **How do revenue and expenses vary across companies, and what is the relationship between them? Are there companies achieving higher profits by managing expenses more effectively?**

This chart examines the relationship between revenue and expenses for each company. The analysis aims to understand how expenses impact revenue and identify companies that achieve the best balance between the two to maximize profitability.

Imagine analyzing the performance of five different companies. The chart illustrates how revenue and expenses rise or fall for each company. Companies with a significant gap between revenue and expenses, such as Company A, show higher profitability. In contrast, Company D has a weaker balance where expenses are close to revenue. Through this chart, we can start asking critical questions: How can lower-performing companies reduce their expenses? Do the more profitable companies employ specific strategies to achieve this balance?

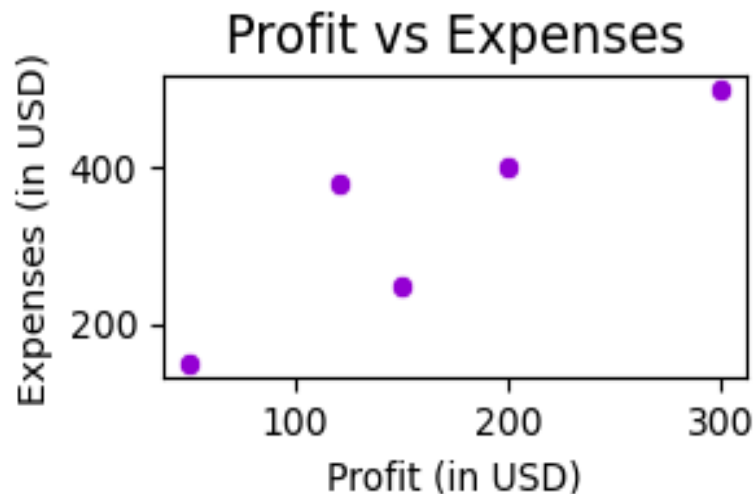


✓ **What is the relationship between profit and expenses across companies, and how do expenses influence profitability?**

This scatterplot explores the relationship between profit and expenses for different companies.

The aim is to understand whether higher expenses directly correlate with lower profits or if some companies can achieve profitability despite high expenses.

Imagine analyzing how companies manage their expenses to achieve profitability. Each point on the chart represents a company, showing its expenses and corresponding profit. Companies with lower expenses and higher profits stand out as efficient managers of their resources. For instance, does a company with higher expenses always suffer from lower profits, or can effective revenue strategies compensate? This chart invites deeper investigation into the balance between spending and earning strategies for better financial outcomes.

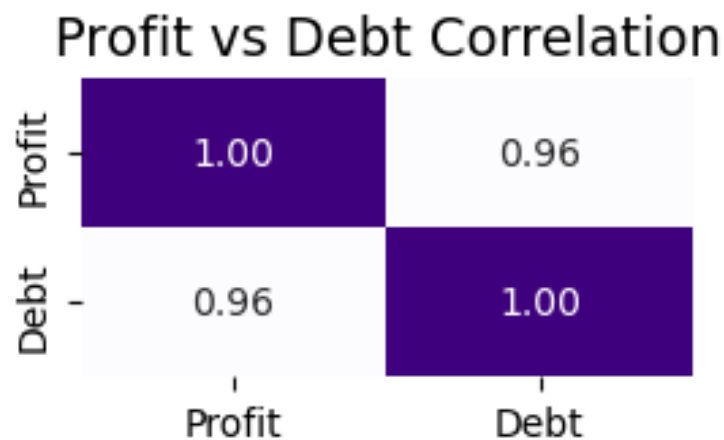


✓ **How strong is the relationship between profits and debt in companies?**  
**Can high debt levels impact a company's ability to generate profits?**

This heatmap analyzes the correlation between profits and debt using the correlation coefficient.

The goal is to understand whether there is a positive or negative relationship between profits and debt, helping to evaluate the impact of debt on financial performance.

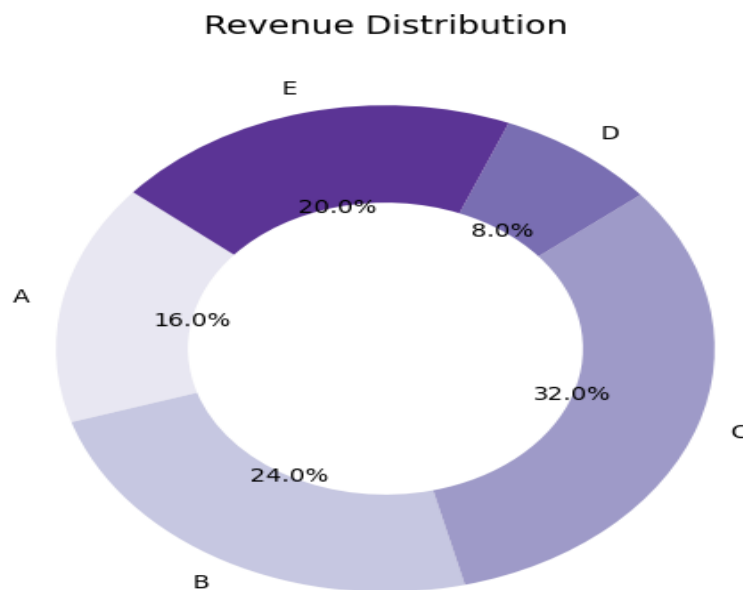
Imagine trying to understand the relationship between debt and profits. Does higher debt lead to increased profits due to effective investments, or could high debt weaken profitability? This heatmap highlights the numerical correlation between these two factors, where a coefficient close to 1 indicates a strong positive relationship. If the relationship is strong, it prompts further investigation into how companies use debt as a lever to boost profits. Conversely, if the relationship is weak or negative, it may suggest that debt acts as a burden on profitability.



✓ **How is revenue distributed among companies, and which companies contribute the most to the total revenue?**

This donut chart visualizes the distribution of revenue across different companies, highlighting the relative contribution of each company to the total revenue. The aim is to identify the dominant companies and assess their share in the market.

Imagine looking at the market shares of five companies, where each segment of the donut chart represents a company's contribution to the total revenue. Company C, with the largest slice, stands out as the leader in revenue generation, while Company D has the smallest share. This visualization allows us to pose important questions: What strategies make Company C dominant in revenue? How can smaller players like Company D increase their market share? The chart provides a foundation for deeper analysis into the factors driving revenue generation in these companies.

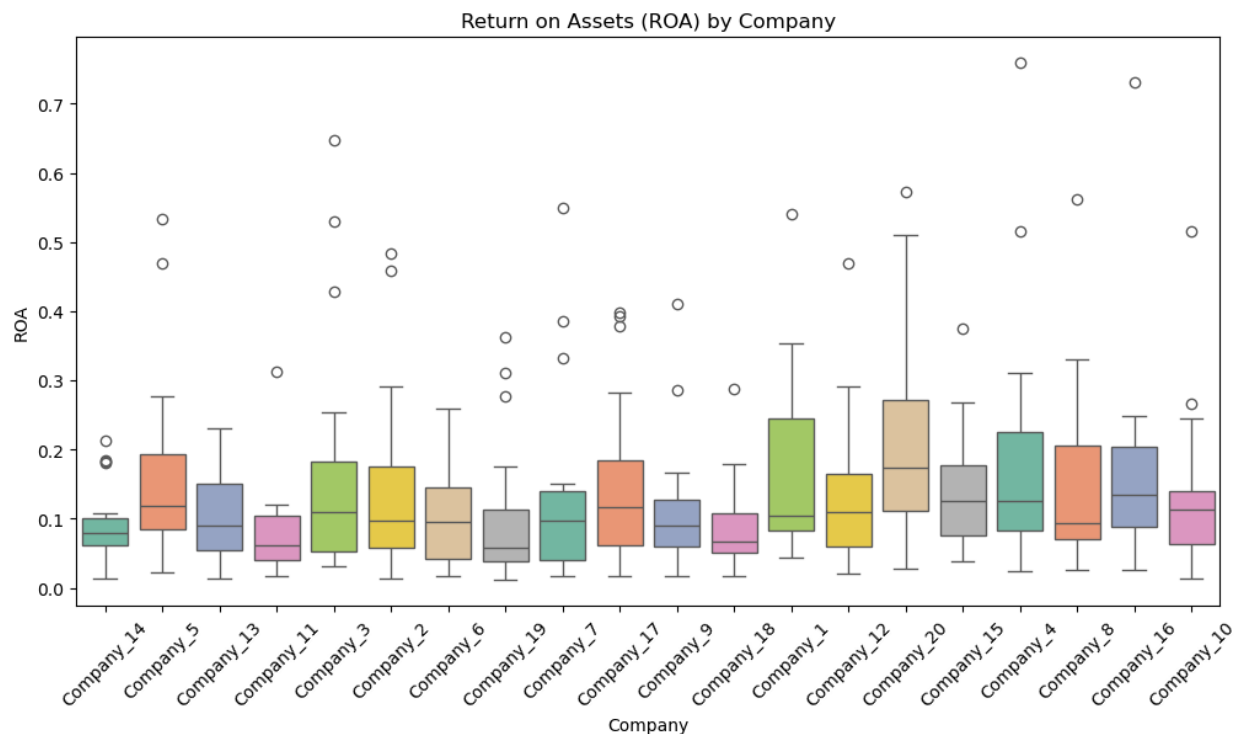




✓ **How does the return on assets (ROA) vary across different companies? Which companies demonstrate consistent performance, and which experience significant fluctuations?**

This boxplot illustrates the distribution of return on assets (ROA) for each company. The goal is to analyze the variance in financial performance among companies and identify those with stable performance versus those with significant volatility.

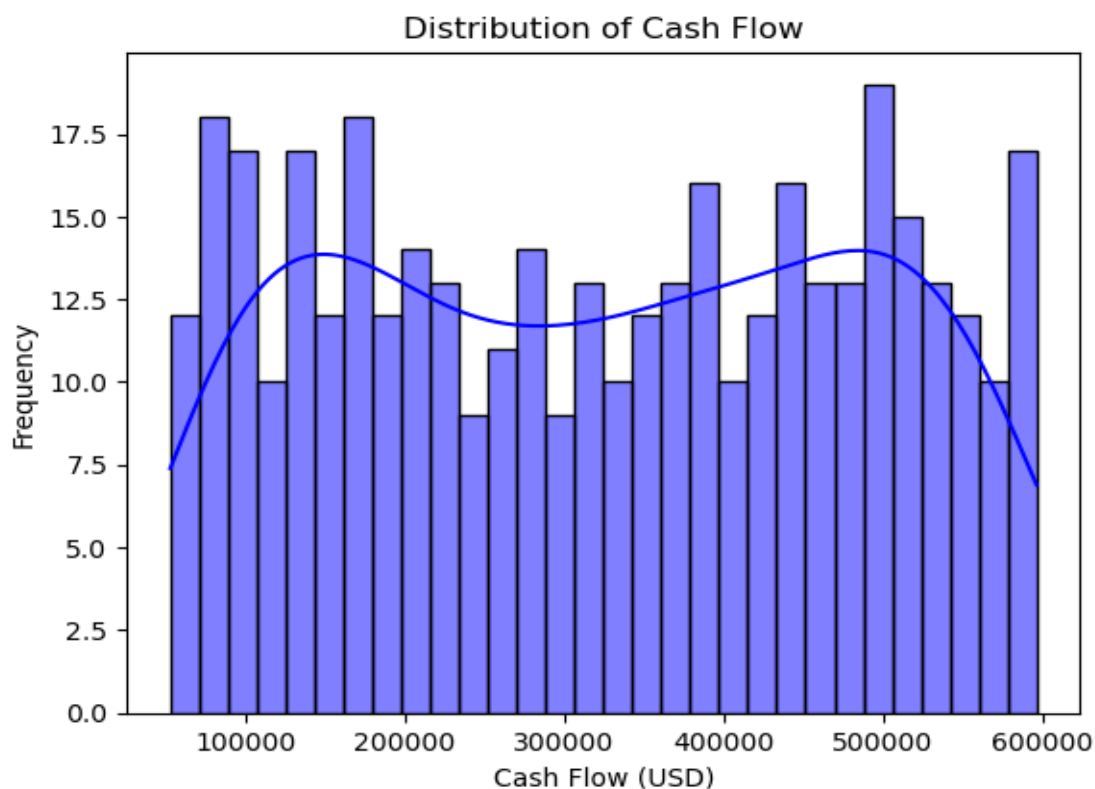
Imagine comparing the performance of 20 companies based on their return on assets (ROA). The plot highlights how performance differs across companies, with outliers indicating exceptional or weak performance. Companies with narrow boxes and high median lines demonstrate consistent and strong performance, while those with wide boxes face significant fluctuations. This visualization prompts questions: What makes some companies more stable in their performance? Is it asset diversification or effective management strategies? This analysis helps identify leading companies and those needing improvement.



✓ **What is the distribution of cash flow among companies, and are there any patterns or anomalies in the cash flow levels?**

This histogram visualizes the distribution of cash flow across various companies, highlighting the frequency of different cash flow ranges. The analysis aims to identify trends, such as whether most companies operate within a specific cash flow range or if there are significant outliers.

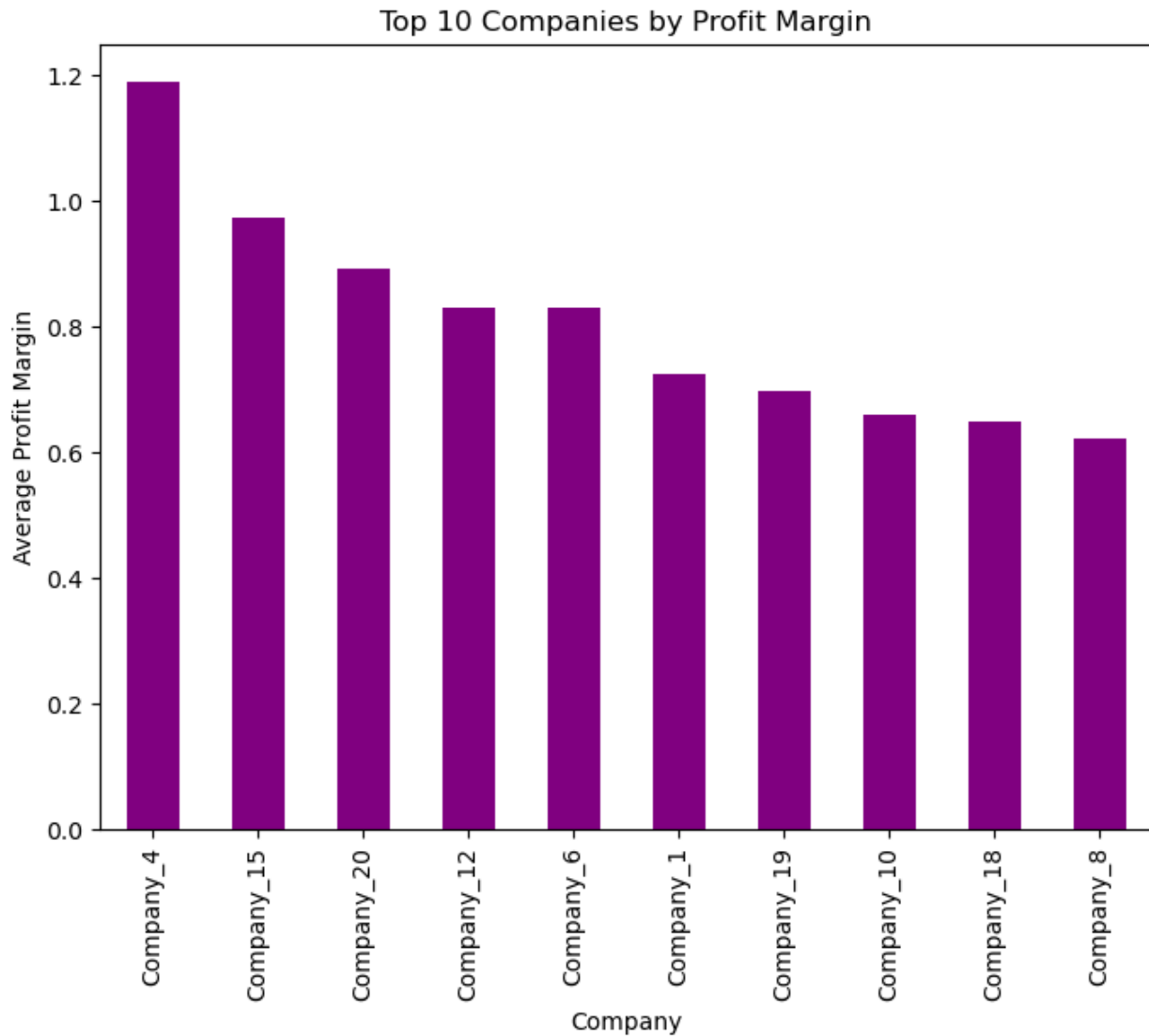
Imagine analyzing the cash flow patterns of a group of companies. Each bar in the histogram represents the number of companies falling within a specific cash flow range, while the smooth curve highlights the overall trend. Are most companies clustered around moderate cash flow levels, or do we see peaks at higher or lower ranges? This chart helps uncover whether companies maintain steady cash flow levels or if anomalies, such as extreme highs or lows, exist. By interpreting this distribution, we can gain insights into the financial health and liquidity of these companies.



✓ **Which companies achieve the highest profit margins, and what factors might contribute to this exceptional financial performance?**

This bar chart highlights the top 10 companies by average profit margin. The goal is to identify the leading companies and understand the strategies or factors that enable them to achieve higher profit margins compared to others.

Imagine analyzing company performance based on profitability. This chart shows the companies leading the list with higher profit margins, such as Company 4, indicating their efficiency in managing costs and generating greater profits. Through this analysis, we can start asking questions: Do these companies rely on effective pricing strategies? Or are they operating in less competitive industries? This chart provides insights into how other companies can improve their profit margins.

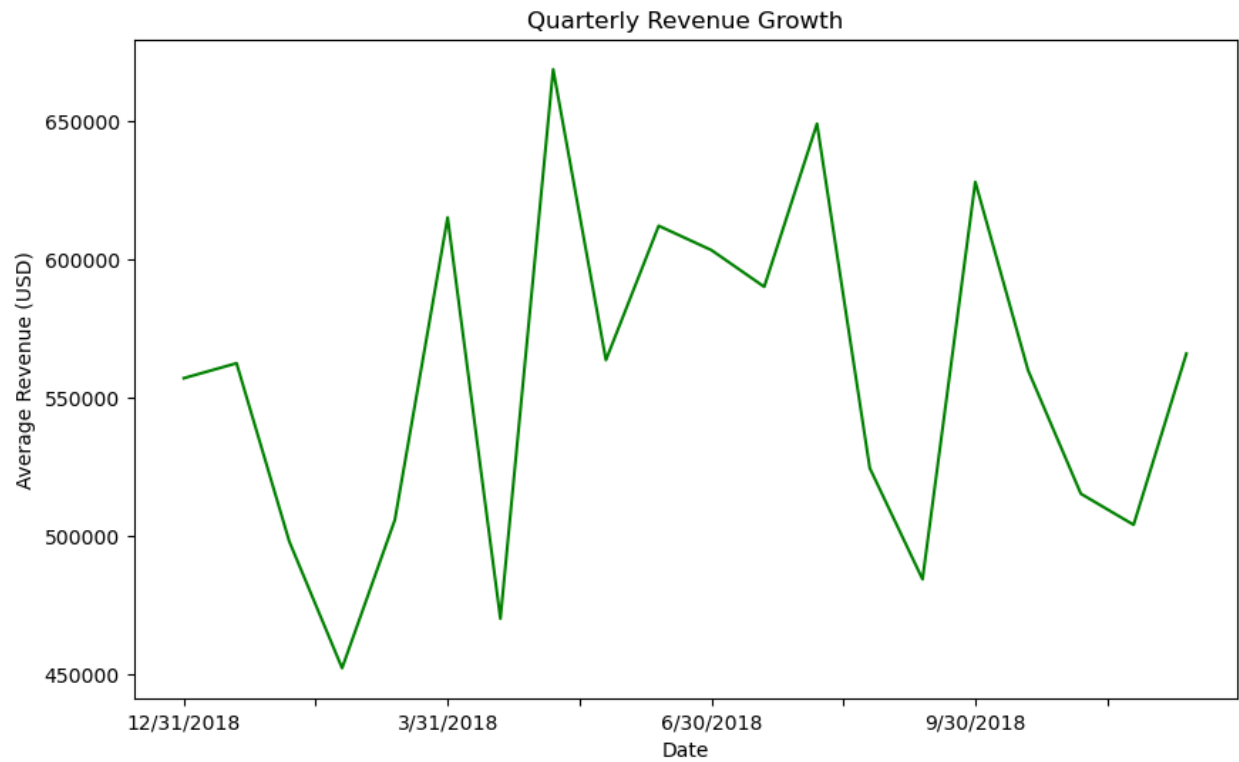


✓ **How does quarterly revenue growth fluctuate over time, and what factors might be driving these changes?**

This line chart illustrates the quarterly revenue growth over a specific period. The goal is to identify trends, patterns, or anomalies in revenue fluctuations and explore potential reasons behind significant increases or decreases.

Imagine tracking the financial journey of a company over time. This chart shows the ups and downs of quarterly revenue, highlighting moments of significant growth and decline. For instance, what might have caused the sharp increase in revenue during a specific quarter? Was it due to seasonal factors, successful

product launches, or market demand? Conversely, what led to the dips in revenue? By analyzing these patterns, we can gain insights into the company's performance and identify areas for improvement or strategic focus.

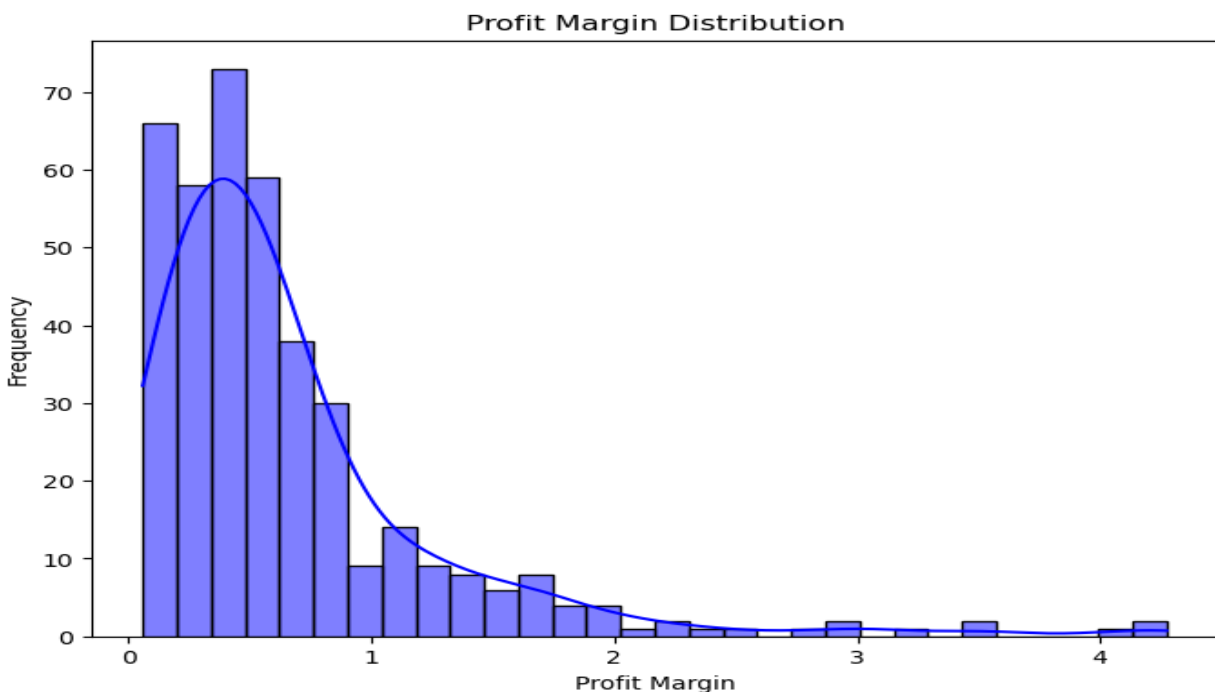


✓ **What is the overall distribution of profit margins among companies, and what does it reveal about the financial performance of most companies?**

This histogram illustrates the distribution of profit margins across companies, highlighting the frequency of different profit margin ranges. The goal is to understand whether most companies operate with high, moderate, or low profit margins and identify any unusual trends or outliers.

Imagine analyzing the profit margins of a group of companies. This chart shows how these margins are distributed, with a clear concentration around lower profit margin values. The steep decline in frequency as profit margins increase suggests that only a few companies achieve exceptionally high profitability.

This raises key questions: What differentiates these high-margin companies from the rest? Are they leveraging unique strategies, operating in less competitive markets, or achieving cost efficiencies? By understanding the distribution, we can identify patterns that contribute to financial success.



This dashboard provides a comprehensive analysis of the financial performance of several investment companies. Each chart represents a part of the story, shedding light on the relationship between various financial aspects. Through these visualizations, we can gain a deeper understanding of market dynamics and make better strategic decisions.

**1. Revenue Distribution:**

- Company C leads with 32% of the total revenue, demonstrating clear dominance. In contrast, Company D contributes only 8%. This gap raises questions: What makes Company C more successful? What can other companies do to improve their contributions?

**2. Profit by Company:**

- The chart highlights that Company C generates the highest profits, while Company D records the lowest. Focusing on operational efficiency and cost control could help underperforming companies improve their outcomes.

**3. Revenue vs Expenses:**

- The line chart reveals the balance between revenue and expenses. Companies maintaining a significant gap, such as Company C, demonstrate profitability. On the other hand, Company D struggles with a narrow margin between revenue and expenses.

**4. Profit vs Debt Correlation:**

- The strong positive correlation (0.96) shows that companies leveraging debt strategically can achieve higher profits. This invites consideration of how debt can be used as a tool for growth.

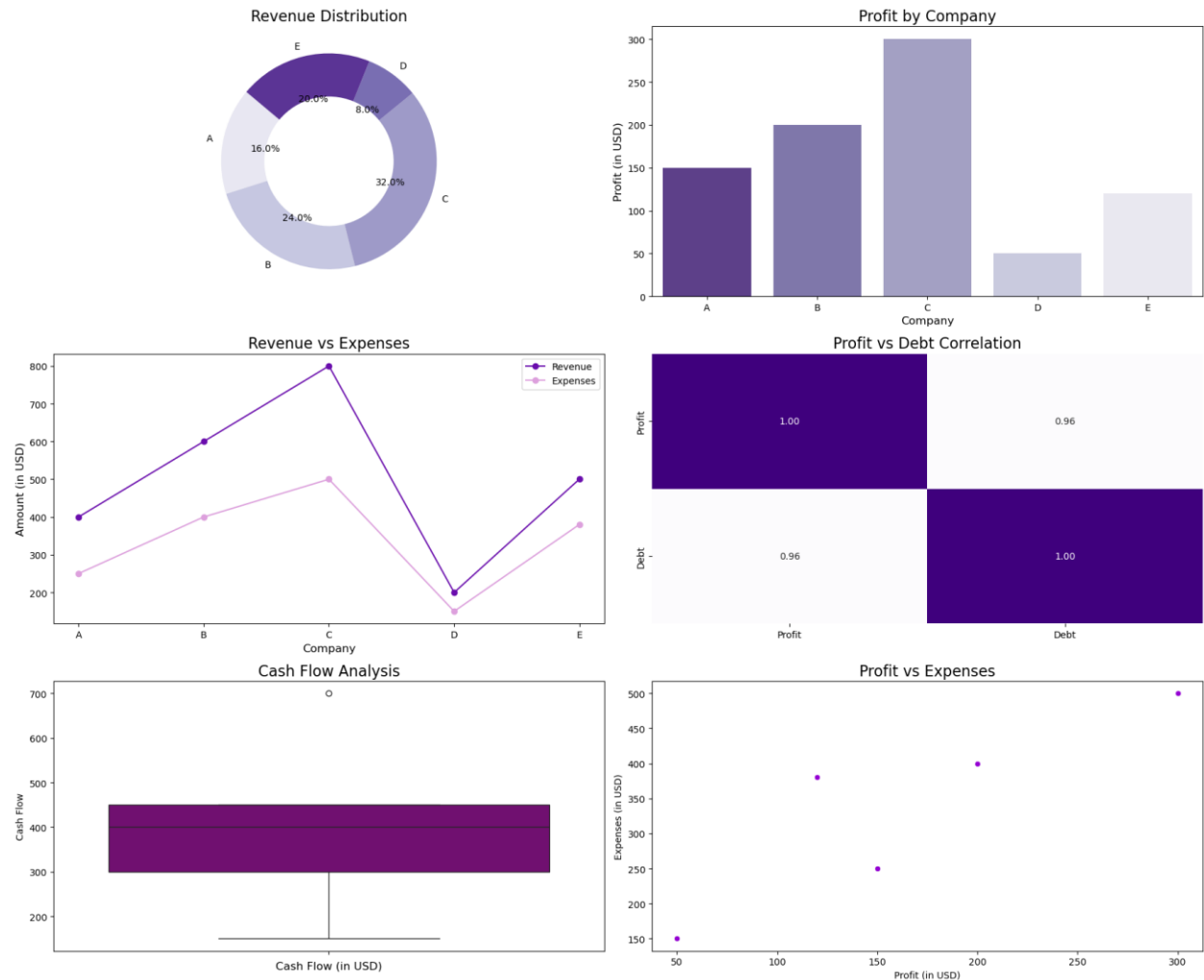
**5. Cash Flow Analysis:**

- The boxplot indicates that most companies maintain stable cash flows, with a few outliers. This stability reflects good liquidity management, though companies with outliers may need to improve their strategies.

**6. Profit vs Expenses:**

- The scatterplot shows that some companies, even with higher expenses, still achieve good profits. Analyzing these companies may uncover effective strategies worth emulating.

Financial Analysis of Investment Companies



Conclusion:

These visualizations tell a comprehensive story about the financial performance of investment companies, from revenue distribution to debt utilization and cash flow management. By analyzing this data, companies can identify their strengths and areas for improvement, enabling them to make more strategic and effective decisions.



