永續數據分析
(Sustainability and ESG Data Analytics)

ESG數據報告、企業永續報告書
(ESG Data Reporting, Corporate Sustainability Reports)

1122ESGDA07
DM4, NTPU (N4084) (Spring 2024)
Fri, 10, 11, 12 (18:30-21:15) (臺北大學民生校區 305)

https://meet.google.com/miy-fbif-max

戴敏育 教授
Min-Yuh Day, Ph.D,
Professor

Institute of Information Management, National Taipei University
https://web.ntpu.edu.tw/~myday

2024-04-26

課程大綱 (Syllabus)

週次 (Week)　日期 (Date)　內容 (Subject/Topics)

1 2024/02/23 永續數據分析概論
(Introduction Sustainability and ESG Data Analytics)

2 2024/03/01 環境、社會與治理 (ESG) 淨零數位轉型
(Environmental, Social, and Governance (ESG) in Net-Zero Digital Transformation)

3 2024/03/08 永續與ESG 資料科學
(Data Science for Sustainability and ESG)

4 2024/03/15 永續數據分析個案研究 I
(Case Study on Sustainability and ESG Data Analytics I)

5 2024/03/22 Web 3.0 和大數據分析在金融科技、綠色永續金融
(Web 3.0 and Big Data Analysis in Fintech, Green and Sustainable Finance)

課程大綱 (Syllabus)

週次 (Week)　日期 (Date)　內容 (Subject/Topics)

6 2024/03/29 TCFD 氣候相關財務揭露與En-ROADS 氣候變遷模擬
(Task Force on Climate-Related Financial Disclosures (TCFD) and En-Roads Interactive)

7 2024/04/05 放假 (No Classes)

8 2024/04/12 期中報告 (Midterm Project Report)

9 2024/04/19 ESG數據的收集、分析和視覺化
(ESG Data Gathering, Analysis, and Visualization)

10 2024/04/26 ESG數據報告 (ESG Data Reporting);
企業永續報告書 (Corporate Sustainability Reports)

課程大綱 (Syllabus)

週次 (Week)　日期 (Date)　內容 (Subject/Topics)

11 2024/05/03 ESG數據驗證 (ESG Data Verification)

12 2024/05/10 永續數據分析個案研究 II
(Case Study on Sustainability and ESG Data Analytics II)

13 2024/05/17 能源之星報告與數據揭露
(Energy Star Reporting and Data Disclosure)

14 2024/05/24 人工智慧物聯網在ESG永續應用
(Artificial Intelligence of things (AIoT) in ESG and Sustainability Applications)

15 2024/05/31 生成式AI於永續評等和報告生成
(Generative AI for ESG Rating and Reporting Generation)

16 2024/06/07 期末報告 (Final Project Report)

ESG Data Reporting, Corporate Sustainability Reports

Outline

- **ESG Data Reporting**
- **Corporate Sustainability Reports**

Sustainability and ESG Data Analytics

Importance of ESG Reporting

Why ESG Data Reporting Matters

- **Informed decision-making for investors**
- **Transparency and building trust**
- **Identifying risks and opportunities**
- **Benchmarking against peers**

Essential Python Libraries for ESG Data Reporting

- Pandas
  - Data loading, manipulation, cleaning
- NumPy
  - Numerical calculations
- Matplotlib/Seaborn
  - Data visualization

Collecting ESG Data

- Free repositories
  - MSCI ESG Ratings
  - Sustainalytics
- Paid Providers
  - Highlight specialization and more granular data
- Company Websites
  - Sustainability reports, investor relations

Processing and Analyzing ESG Data

Transforming Data into Insights

- **Cleaning and preprocessing**
  - **handling missing data**
- **Calculating ESG Scores or metrics**
- **Normalization**
  - **for cross-company comparison**

Corporate Sustainability Reports

Why Analyze Sustainability Reports?

- **The Power of Data-Driven ESG Analysis**
- **Speed and scale compared to manual reading**
- **Track performance trends more precisely**
- **Deeper insights and comparisons**
- **Identify areas for critical evaluation**

Python for Sustainability Reports Analysis

- **BeautifulSoup**
  - **Handle HTML reports**
- **pdfminer.six**
  - **Extract text from PDF reports**
- **Pandas**
  - **Store and manipulate extracted data**
- **Matplotlib/Seaborn**
  - **Data visualization**

Corporate Sustainability Reports

Finding Sustainability Data

- **Company Websites**
  - **Investor relations section, dedicated reports page**
- **Sustainability Report Repositories**
  - **GRI, etc.**

Extracting Data (HTML)

Scraping Data from Web-Based Reports

- **Finding the right HTML tags**
  - **(using browser inspection tools)**
- **BeautifulSoup to parse and extract into structured data**

Extracting Data (PDF)

Handling PDF-Based Reports

- **Using pdfminer.six for text conversion**
- **Potential use of regular expressions for cleaning**

Analysis with Pandas

Turning Data into Insights

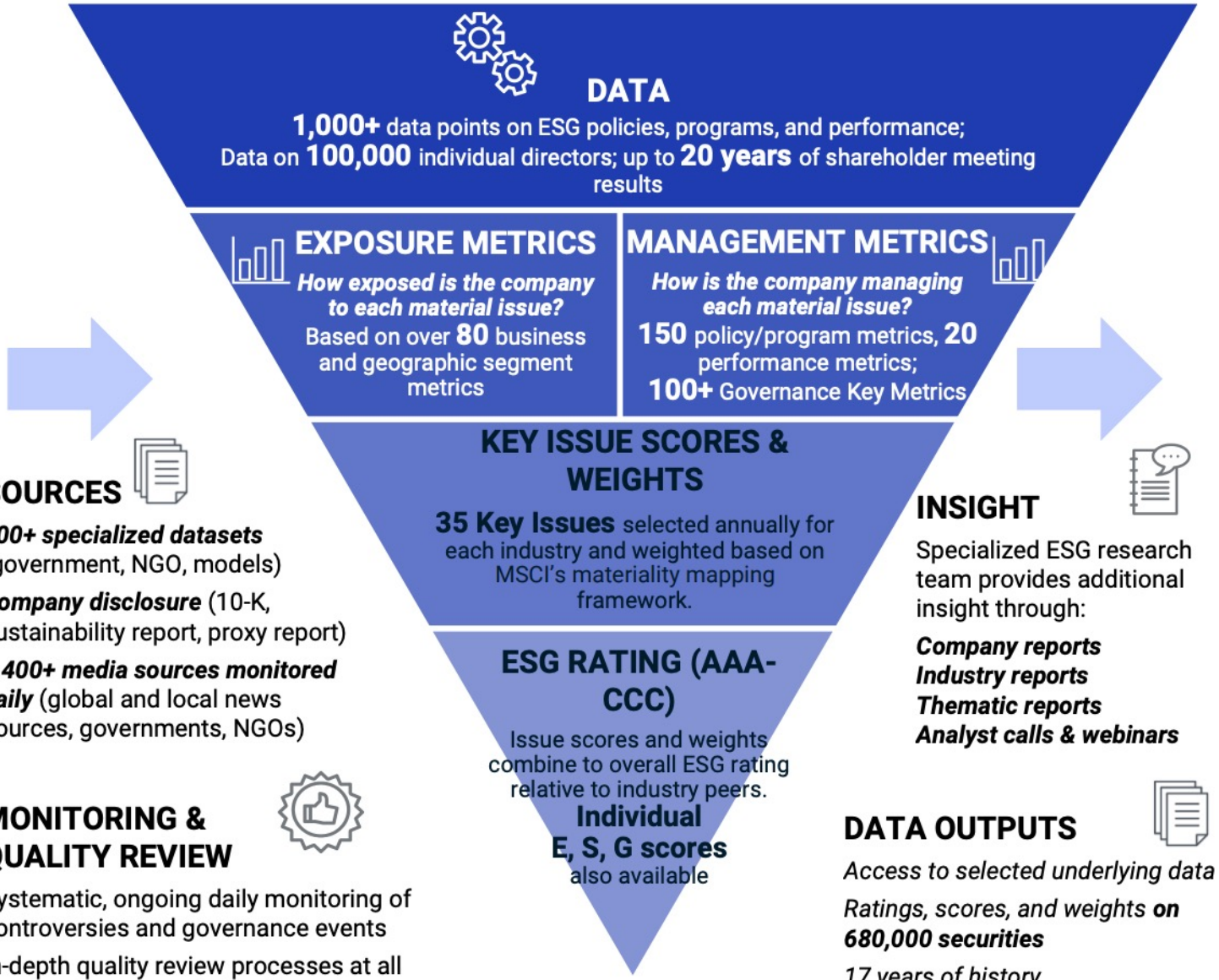
- **Loading into DataFrames**
- **Cleaning (handling missing values, formats)**
- **Calculating ESG metrics or ratios**
- **Comparing data across years**

Visualizing Results

Communicating ESG Performance

- **Choose charts that align with analysis goals**
- **Clear visuals: labeling, annotations**

MSCI ESG Rating Framework



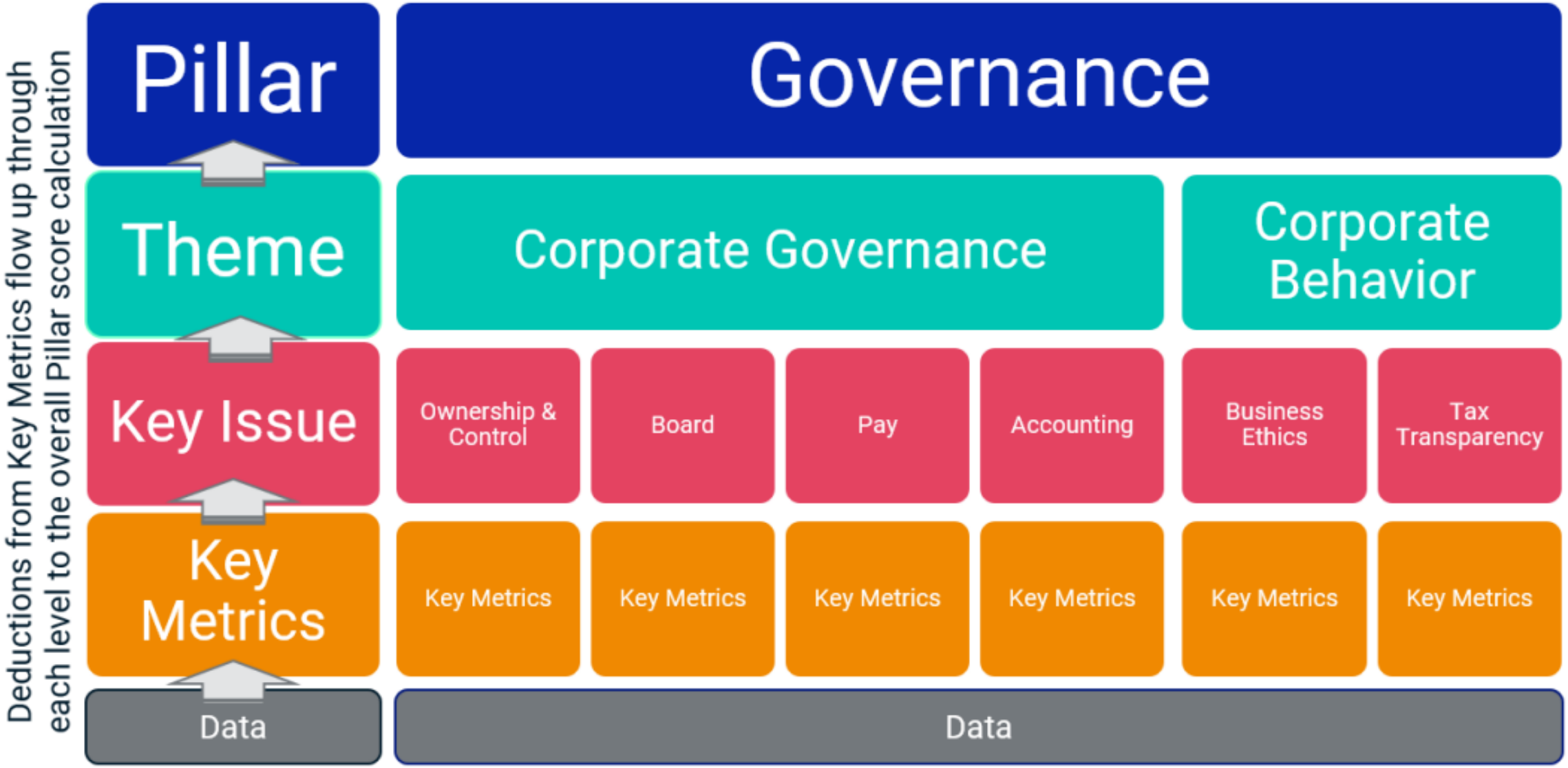
Source: https://www.msci.com/documents/1296102/21901542/ESG-Ratings-Methodology-Exec-Summary.pdf

MSCI ESG Key Issue Hierarchy

3 Pillars	10 Themes	35 ESG Key Issues	
Environment	**Climate Change**	Carbon Emissions	Financing Environmental Impact
		Product Carbon Footprint	Climate Change Vulnerability
	Natural Capital	Water Stress	Raw Material Sourcing
		Biodiversity & Land Use	
	Pollution & Waste	Toxic Emissions & Waste	Electronic Waste
		Packaging Material & Waste	
	Environmental Opportunities	Opportunities in Clean Tech	Opportunities in Renewable Energy
		Opportunities in Green Building	
Social	**Human Capital**	Labor Management	Human Capital Development
		Health & Safety	Supply Chain Labor Standards
	Product Liability	Product Safety & Quality	Privacy & Data Security
		Chemical Safety	Responsible Investment
		Consumer Financial Protection	Health & Demographic Risk
	Stakeholder Opposition	Controversial Sourcing	
		Community Relations	
	Social Opportunities	Access to Communications	Access to Health Care
		Access to Finance	Opportunities in Nutrition & Health
Governance	**Corporate Governance**	Ownership & Control	Pay
		Board	Accounting
	Corporate Behavior	Business Ethics	
		Tax Transparency	

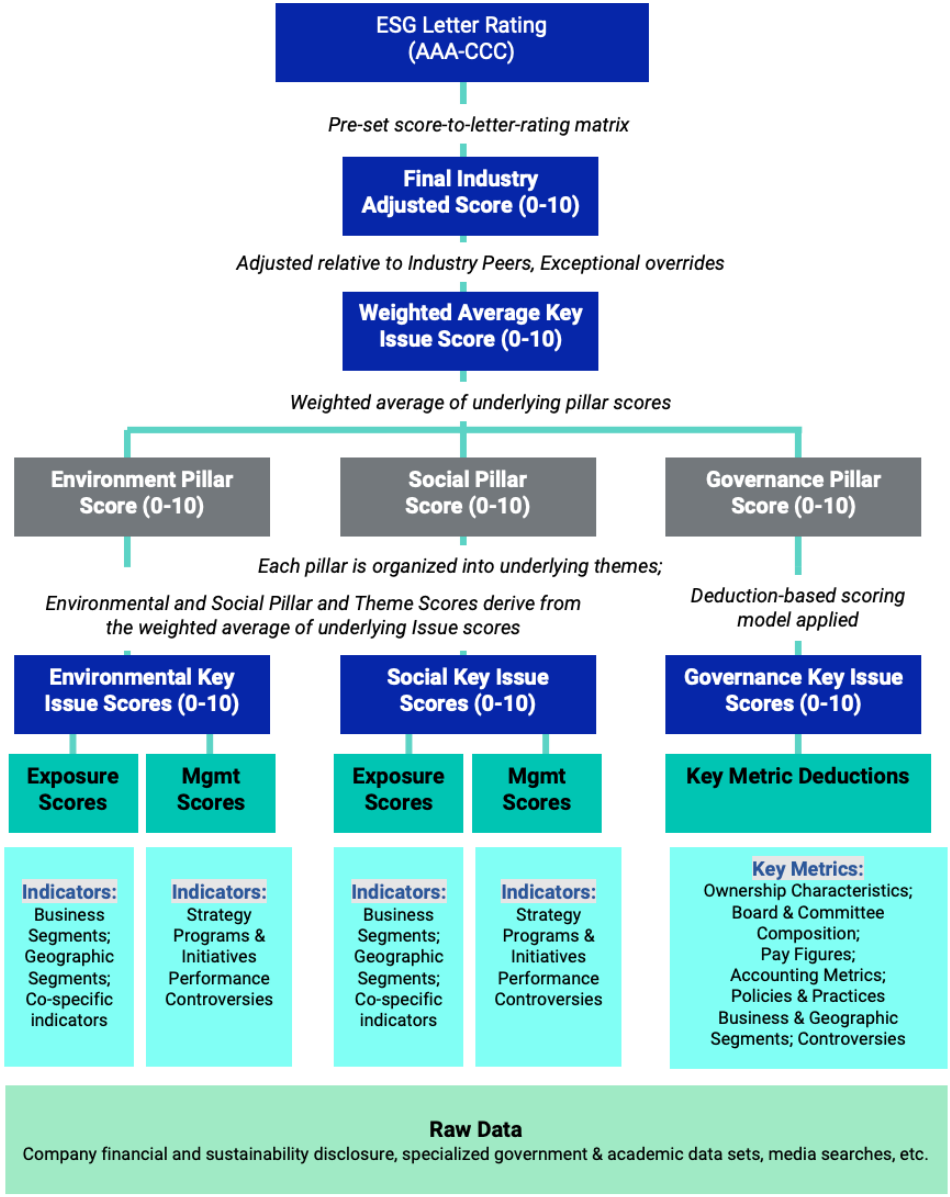
Source: https://www.msci.com/documents/1296102/21901542/ESG-Ratings-Methodology-Exec-Summary.pdf

MSCI Governance Model Structure



Source: https://www.msci.com/documents/1296102/21901542/ESG-Ratings-Methodology-Exec-Summary.pdf

MSCI Hierarchy of ESG Scores



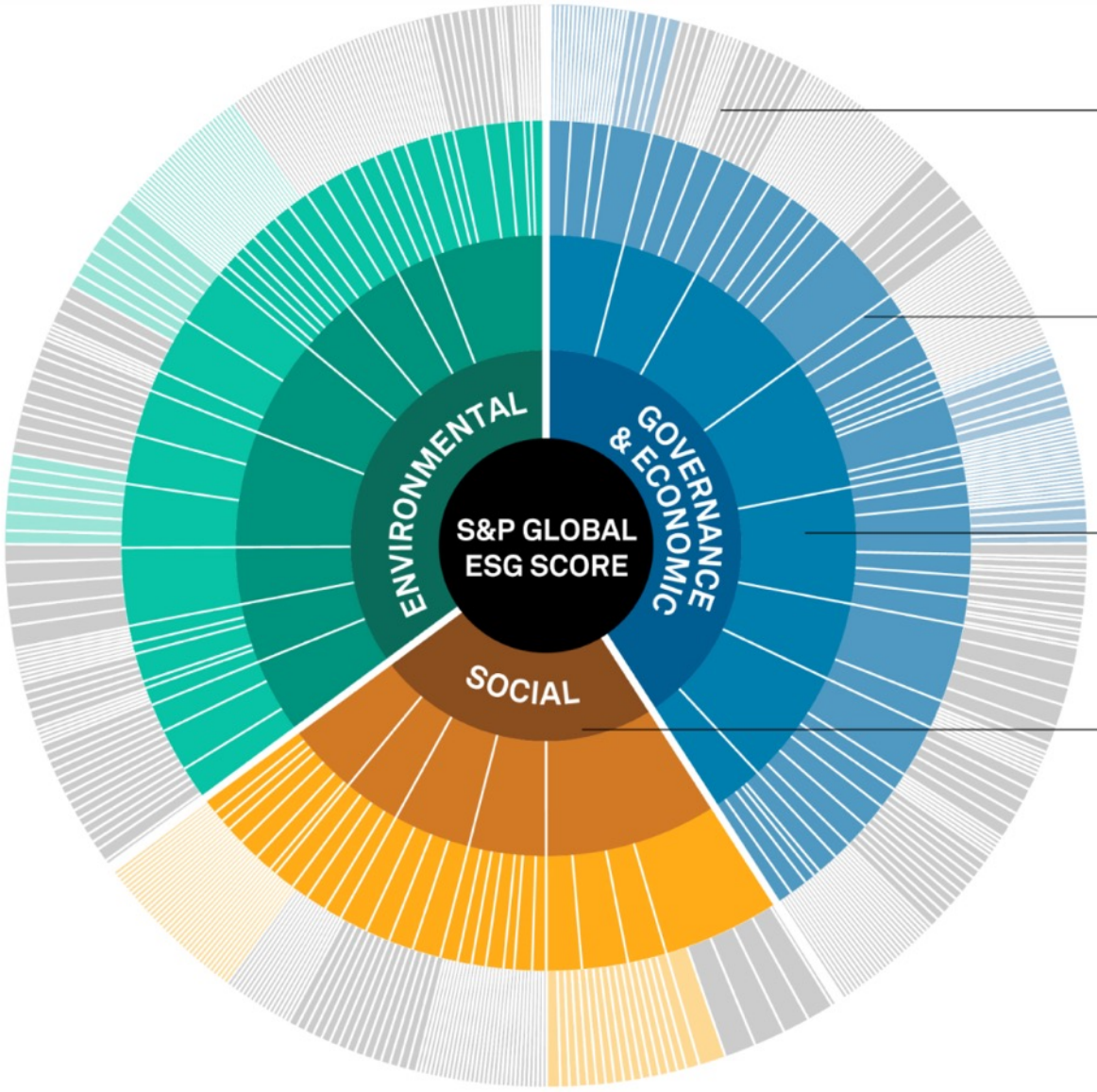
Source: https://www.msci.com/documents/1296102/21901542/ESG-Ratings-Methodology-Exec-Summary.pdf

DJSI S&P Global ESG Score

8,000
Companies

90%
Global market capitalization

340,000+
Current Research Universe and Active Securities



Approx.
1,000
Datapoints

Assessed values, text, checkboxes, documents
Sources: Web-based questionnaire and company documents

130+
Questions

Weighted data point scores
Up to 50% industry-specific

Ave.
30+
Criteria scores

Weighted question scores
61 industry specific approaches, with tailored questions, criteria and related weightings

3
Dimension scores

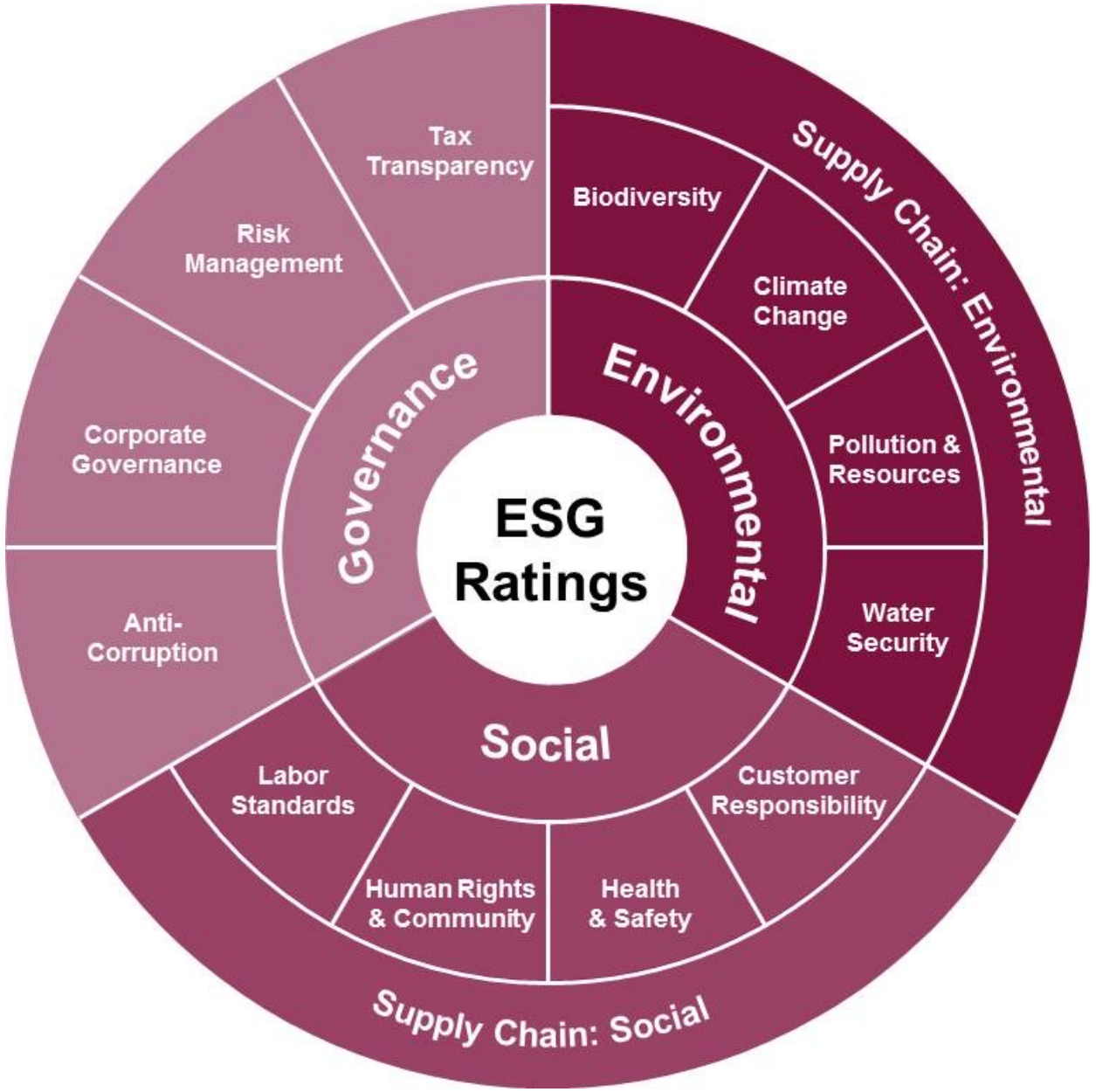
Weighted criteria scores
Adjusted for corporate ESG controversies where applicable

1
S&P Global ESG Score

Sum of weighted dimension scores

Source: DJSI, S&P Global Corporate Sustainability Assessment (CSA), https://www.spglobal.com/esg/solutions/data-intelligence-esg-scores

FTSE Russell ESG Ratings



Source: https://www.ftserussell.com/data/sustainability-and-esg-data/esg-ratings

ESG Data Analysis and Visualization

```
# ESG Data Analysis and Visualization
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from datetime import datetime
import random

# Generate synthetic data
np.random.seed(0)
data = {
    'company': ['Company A', 'Company B', 'Company C', 'Company D', 'Company E'],
    'emissions': np.random.randint(10000, 50000, 5),
    'diversity': np.random.uniform(0.2, 0.9, 5),
    'employee_satisfaction': np.random.uniform(60, 90, 5),
    'waste_type': ['Plastic', 'Organic', 'Electronic', 'Metal', 'Other'],
    'waste_amount': np.random.randint(100, 500, 5)
}
df = pd.DataFrame(data)
```

ESG Data Analysis and Visualization

```
# Separate DataFrame for time series and correlation
time_series_data = pd.DataFrame({
    'year': np.repeat(np.arange(2018, 2023), 5),
    'company': np.tile(['Company A', 'Company B', 'Company C', 'Company
    D', 'Company E'], 5),
    'energy_use': np.random.randint(1000, 5000, 25)
})

# Simulating correlation data with a slight positive trend
diversity = np.linspace(0.2, 0.9, 100)
np.random.shuffle(diversity)
employee_satisfaction = 60 + (diversity - 0.2) * 150
employee_satisfaction += np.random.normal(0, 5, 100)

correlation_data = pd.DataFrame({
    'diversity': diversity,
    'employee_satisfaction': employee_satisfaction
})
```

ESG Data Analysis and Visualization

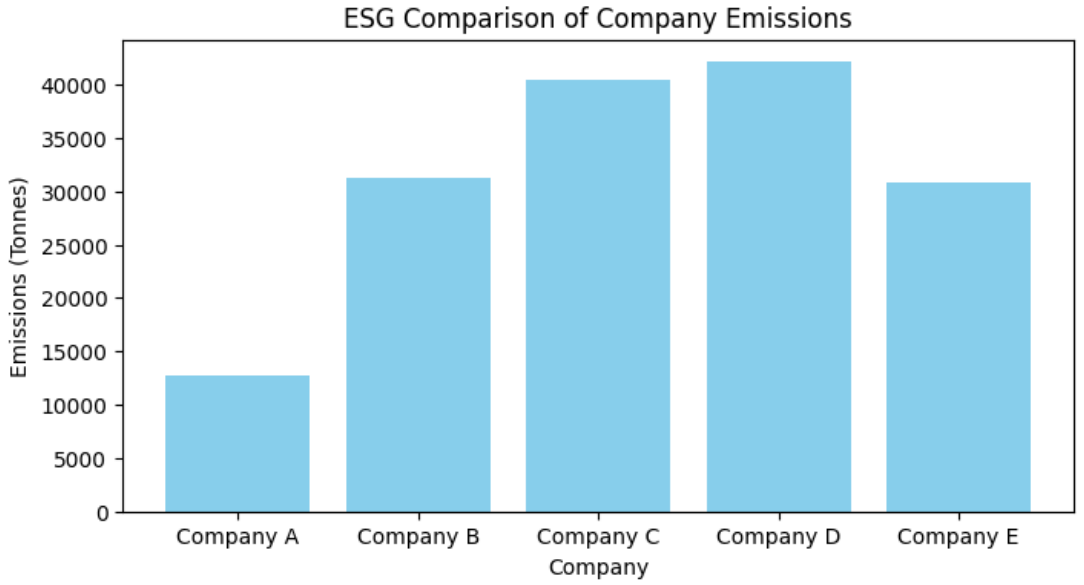
```python
# Save DataFrame to CSV
df.to_csv('ESG_Dataset.csv', index=False)

# Calculate statistics for each company
statistics = df.describe()
statistics = statistics.applymap(lambda x: format(x, '.4f'))
print(statistics)
statistics.to_csv('Company_ESG_Statistics.csv')
```

```
         emissions diversity employee_satisfaction waste_amount
count       5.0000    5.0000                5.0000       5.0000
mean    31447.6000    0.4085               77.1504     345.6000
std     11667.5748    0.1099                7.0841      94.8093
min     12732.0000    0.2397               70.1219     215.0000
25%     30757.0000    0.3909               71.7835     297.0000
50%     31243.0000    0.4083               74.3993     343.0000
75%     40403.0000    0.4691               84.3651     435.0000
max     42103.0000    0.5344               85.0824     438.0000
```

ESG Data Analysis and Visualization

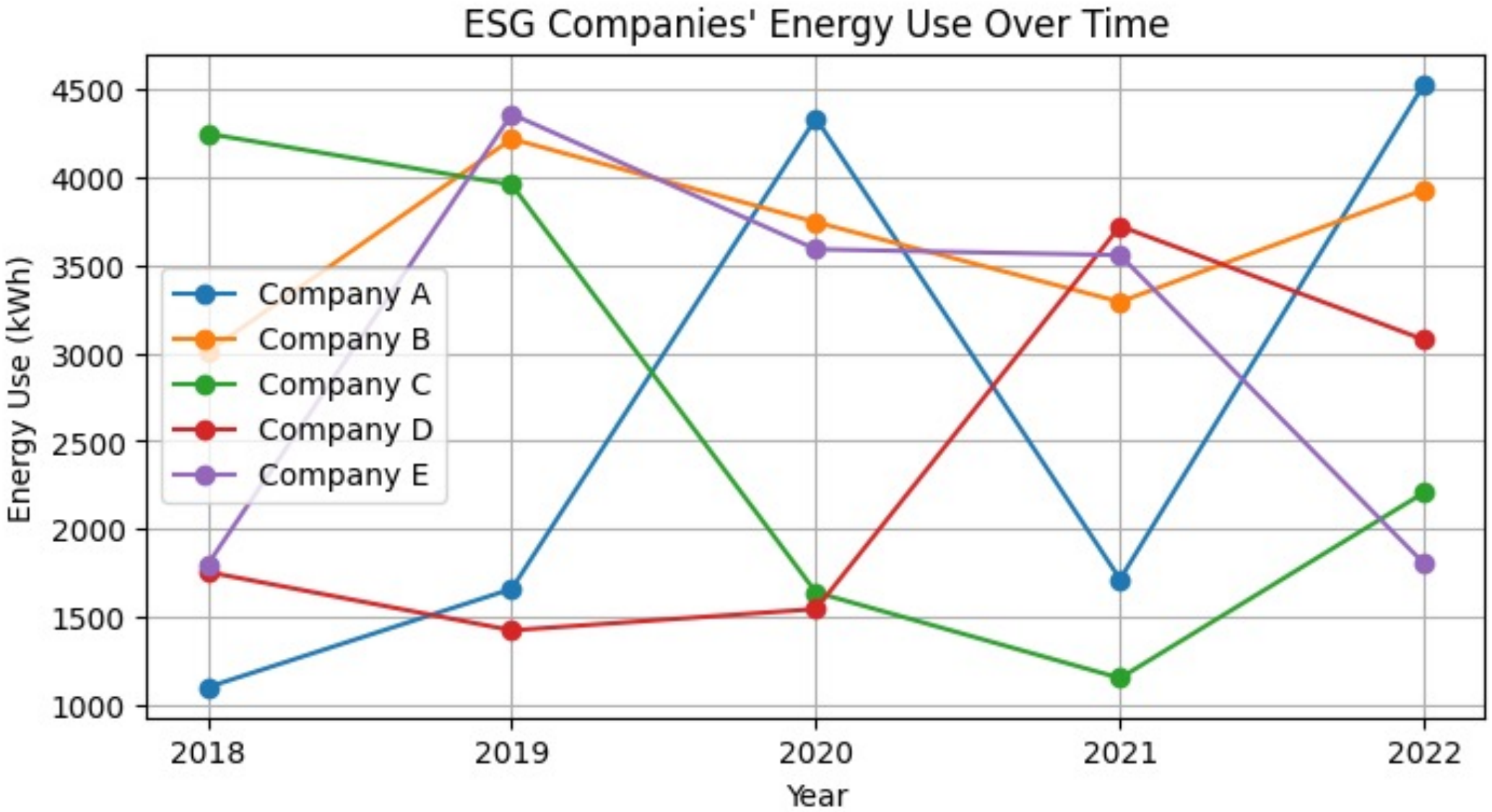
```
# Create visualizations and save them at 300 dpi
# Bar Chart for Emissions
plt.figure(figsize=(8, 4))
plt.bar(df['company'], df['emissions'], color='skyblue')
plt.xlabel('Company')
plt.ylabel('Emissions (Tonnes)')
plt.title('ESG Comparison of Company Emissions')
plt.show()
plt.savefig('ESG Company_Emissions.jpg', format='jpg', dpi=300)
```



ESG Data Analysis and Visualization

```
# Line Chart for Energy Use
plt.figure(figsize=(8, 4))
for company in time_series_data['company'].unique():
    company_data = time_series_data[time_series_data['company'] ==
    company]
    company_data = company_data.sort_values(by='year')
    plt.plot(company_data['year'], company_data['energy_use'],
    marker='o', linestyle='-', label=company)
plt.xlabel('Year')
plt.ylabel('Energy Use (kWh)')
plt.title("ESG Companies' Energy Use Over Time")
plt.xticks(company_data['year'].unique()) # Ensuring only whole years
are marked
plt.legend()
plt.grid(True)
plt.show()
```

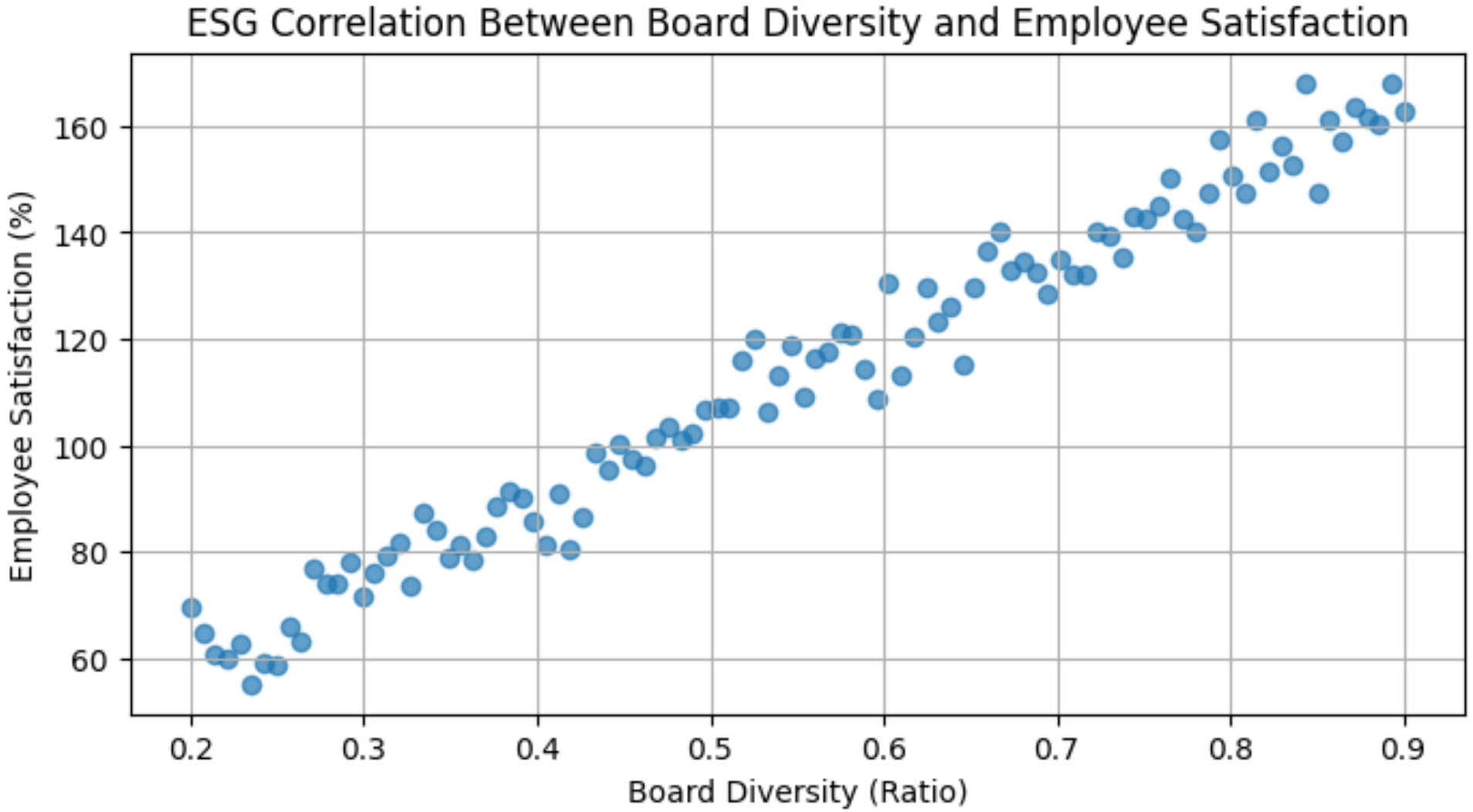
ESG Data Analysis and Visualization



ESG Data Analysis and Visualization

```
# Scatter Plot for Diversity vs. Satisfaction
plt.figure(figsize=(8, 4))
plt.scatter(correlation_data['diversity'],
correlation_data['employee_satisfaction'], alpha=0.7)
plt.xlabel('Board Diversity (Ratio)')
plt.ylabel('Employee Satisfaction (%)')
plt.title('ESG Correlation Between Board Diversity and Employee
Satisfaction')
plt.grid(True)
plt.show()
plt.savefig('ESG_Diversity_vs_Satisfaction.jpg', format='jpg', dpi=300)
```

ESG Data Analysis and Visualization

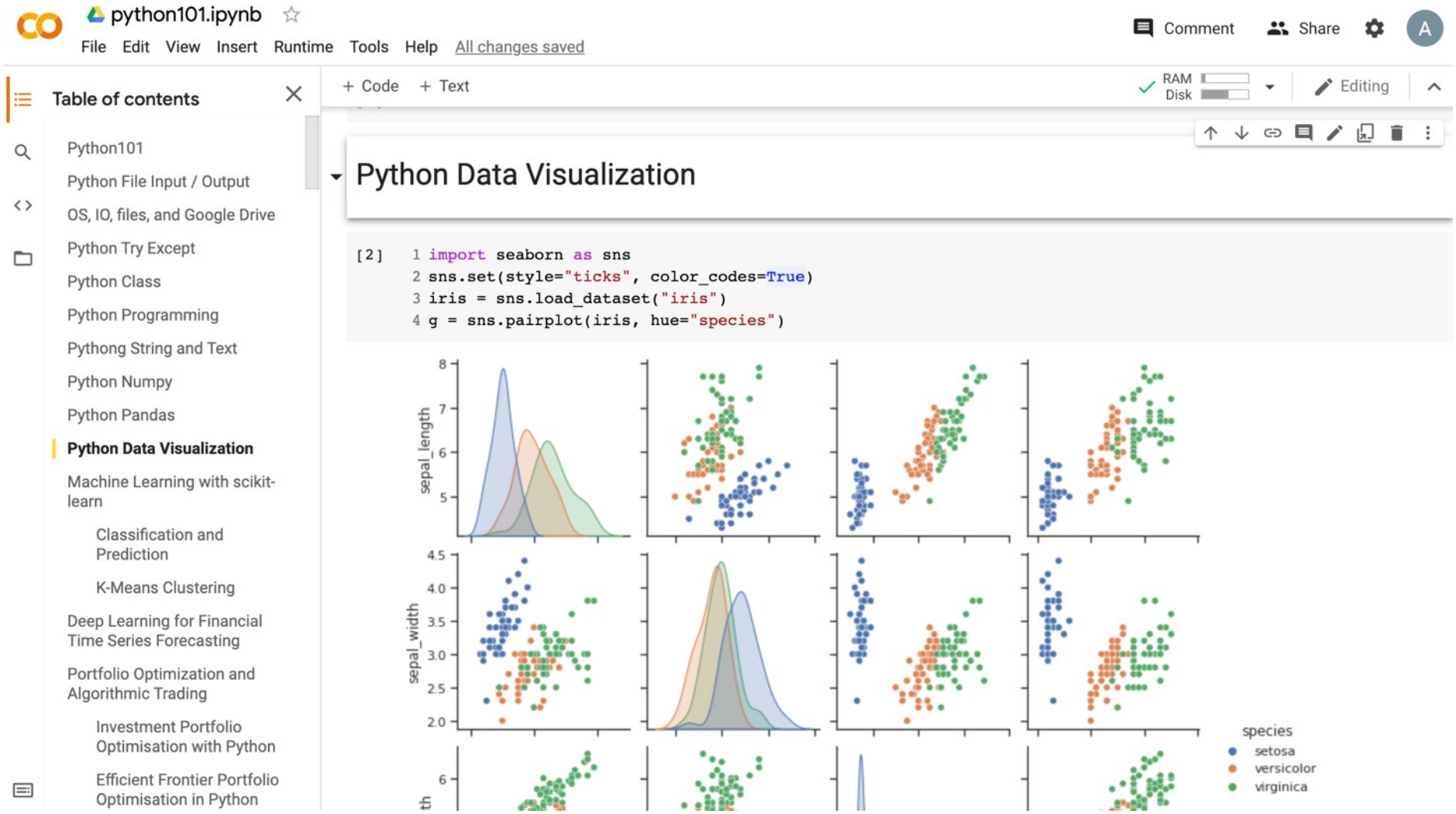


ESG Data Analysis and Visualization

```
# Pie Chart for Waste Types
plt.figure(figsize=(8, 4))
plt.pie(df['waste_amount'], labels=df['waste_type'], autopct='%1.1f%%',
startangle=140)
plt.title('ESG Waste Types')
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a
circle.
plt.show()
plt.savefig('ESG_Waste_Type_Breakdown.jpg', format='jpg', dpi=300)
```

Python in Google Colab (Python101)

https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT



https://tinyurl.com/aintpupython101

Summary

- **ESG Data Reporting**
- **Corporate Sustainability Reports**

References

- Cino Robin Castelli, Cyril Shmatov (2022), Quantitative Methods for ESG Finance, Wiley
- Simon Thompson (2023), Green and Sustainable Finance: Principles and Practice in Banking, Investment and Insurance, 2nd Edition, Kogan Page.
- Chrissa Pagitsas (2023), Chief Sustainability Officers At Work: How CSOs Build Successful Sustainability and ESG Strategies, Apress.
- Min-Yuh Day (2024), Python 101, https://tinyurl.com/aintpupython101