

ESG score

November 18, 2025

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

```
[2]: # esg score.csv downloaded from S&P esg ratings
df = pd.read_csv("esg score.csv")
```

C:\Users\Dan\AppData\Local\Temp\ipykernel_42904\2442267146.py:1: DtypeWarning: Columns (29,30,31,33,35,37,38,42,49,50,52,53,54,55,60,61) have mixed types. Specify dtype option on import or set low_memory=False.
df = pd.read_csv("esg score.csv")

```
[3]: dei_aspects = [
    "Gender Pay Indicators",
    "Workforce Breakdown: Gender",
    "Employee Development Programs",
    "Employee Support Programs"
]

dei_score = (
    df[df['aspectname'].isin(dei_aspects)]
    .groupby(['gvkey', 'assessmentyear'])
    .apply(lambda g: (g['scorevalue'] * g['scoreweight']).sum() /
    g['scoreweight'].sum())
    .reset_index(name='dei_aspects_avg')
)
```

C:\Users\Dan\AppData\Local\Temp\ipykernel_42904\1410881912.py:11: DeprecationWarning: DataFrameGroupBy.apply operated on the grouping columns. This behavior is deprecated, and in a future version of pandas the grouping columns will be excluded from the operation. Either pass `include\_groups=False` to exclude the groupings or explicitly select the grouping columns after groupby to silence this warning.
.apply(lambda g: (g['scorevalue'] * g['scoreweight']).sum() /
g['scoreweight'].sum())

```
[4]: dei_aspects2 = [
    "Gender Pay Indicators"
]

dei_score2 = (
```

```

df[df['aspectname'].isin(dei_aspects2)]
  .groupby(['gvkey', 'assessmentyear'])
  .apply(lambda g: (g['scorevalue'] * g['scoreweight']).sum() /
    ↪g['scoreweight'].sum())
  .reset_index(name='gender_pay_avg')
)

```

C:\Users\Dan\AppData\Local\Temp\ipykernel_42904\1421035364.py:8:

DeprecationWarning: DataFrameGroupBy.apply operated on the grouping columns. This behavior is deprecated, and in a future version of pandas the grouping columns will be excluded from the operation. Either pass `include\_groups=False` to exclude the groupings or explicitly select the grouping columns after groupby to silence this warning.

```

  .apply(lambda g: (g['scorevalue'] * g['scoreweight']).sum() /
g['scoreweight'].sum())

```

```

[5]: result = pd.merge(dei_score, dei_score2, on=['gvkey', 'assessmentyear'],
    ↪how='outer')

```

```

[8]: result.head()

```

```

[8]:
   gvkey  assessmentyear  dei_aspects_avg  gender_pay_avg
0  1004.0             2020         11.323529           22.5
1  1004.0             2021         31.750000           11.0
2  1004.0             2022         22.800000           10.0
3  1004.0             2023         24.621359           10.5
4  1004.0             2024         19.882353           14.5

```

```

[9]: result.to_stata("20250605sp_esgscore.dta")

```

```

[ ]:

```

```

[ ]:

```