

Student Performance & Behavior Dataset

Analiza wyników uczniów

Ten zbiór danych zawiera rzeczywiste dane dotyczące 5 000 rekordów, zebranych od prywatnego dostawcy usług edukacyjnych. Zawiera kluczowe atrybuty niezbędne do eksploracji wzorców, korelacji i wyciągania wniosków związanych z wynikami w nauce.

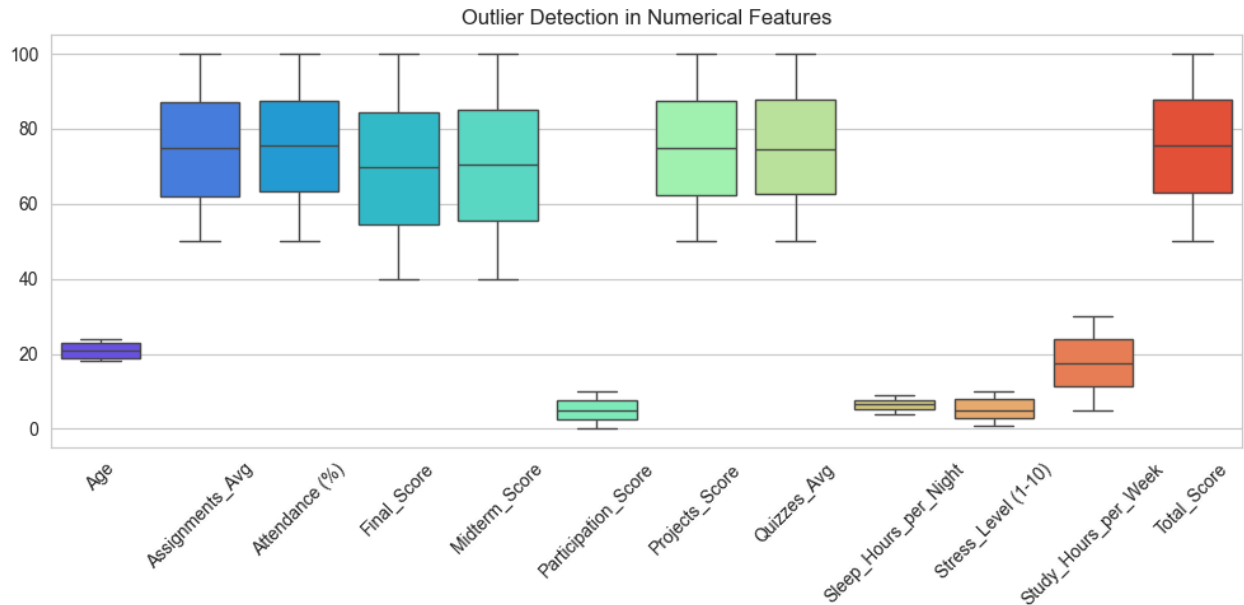
Kolumny:

1. Student_ID – unikalny identyfikator każdego studenta.
2. First_Name – imię studenta.
3. Last_Name – nazwisko studenta.
4. Email – adres e-mail (może być zanonimizowany).
5. Gender – płeć: Mężczyzna, Kobieta, Inna.
6. Age – wiek studenta.
7. Department – wydział/kierunek studiów (np. Informatyka, Inżynieria, Biznes).
8. Attendance (%) – frekwencja w procentach (0–100%).
9. Midterm_Score – wynik egzaminu śródsesemestralnego (maks. 100 pkt).
10. Final_Score – wynik egzaminu końcowego (maks. 100 pkt).
11. Assignments_Avg – średnia ocen z prac domowych (maks. 100 pkt).
12. Quizzes_Avg – średnia ocen z quizów (maks. 100 pkt).
13. Participation_Score – wynik za aktywność na zajęciach (maks. 10 pkt).
14. Projects_Score – ocena za projekt (maks. 100 pkt).
15. Total_Score – suma ważona wszystkich ocen.
16. Grade – ocena końcowa w formacie literowym (A, B, C, D, F).
17. Study_Hours_per_Week – średnia liczba godzin nauki tygodniowo.
18. Extracurricular_Activities – czy student uczestniczy w zajęciach dodatkowych (Tak/Nie).
19. Internet_Access_at_Home – czy student ma dostęp do internetu w domu (Tak/Nie).
20. Parent_Education_Level – najwyższy poziom wykształcenia rodziców (Brak, Szkoła średnia, Licencjat, Magister, Doktorat).

21. Family_Income_Level – poziom dochodu rodziny: Niski, Średni, Wysoki.
22. Stress_Level (1-10) – poziom stresu deklarowany przez studenta (1: niski, 10: wysoki).
23. Sleep_Hours_per_Night – średnia liczba godzin snu na dobę.

	A	B	C	D	E	F	G	H	I
1	Feature	Average	Median	Min	Max	Std	5th percentile	95th percentile	Missing
2	Age	21.0484	21.0	18.0	24.0	1.9897862422526627	18.0	24.0	0
3	Assignments_Avg	74.79867276377426	74.81	50.0	99.98	14.411799371836937	52.451	97.35	517
4	Attendance (%)	75.431409455843	75.725	50.01	100.0	14.372445606230213	52.6415	97.83	516
5	Final_Score	69.64078799999999	69.735	40.0	99.98	17.238744216429684	42.977999999999994	96.79	0
6	Midterm_Score	70.326844	70.50999999999999	40.0	99.98	17.213208523375233	43.29	97.12100000000001	0
7	Participation_Score	4.980024	4.955	0.0	10.0	2.890136020443905	0.53	9.49	0
8	Projects_Score	74.92485999999998	74.98	50.01	100.0	14.42341465806484	52.3195	97.23	0
9	Quizzes_Avg	74.910728	74.695	50.03	99.96	14.504280651637306	52.417500000000004	97.5705	0
10	Sleep_Hours_per_Night	6.4881400000000005	6.5	4.0	9.0	1.4522834316458912	4.2	8.8	0
11	Stress_Level (1-10)	5.4808	5.0	1.0	10.0	2.8615501138038386	1.0	10.0	0
12	Study_Hours_per_Week	17.658859999999997	17.5	5.0	30.0	7.275864288222267	6.3	28.9	0
13	Total_Score	75.121804	75.39500000000001	50.02	99.99	14.399941391825232	52.4495	97.51300000000002	0

	A	B	C	D
1	Feature	Number of unique classes	Missing	Class proportion
2	Department	4	0	CS: 40.44%; Engineering: 29.38%; Business: 20.12%; Mathematics: 10.06%
3	Extracurricular_Activities	2	0	No: 69.86%; Yes: 30.14%
4	Family_Income_Level	3	0	Low: 39.66%; Medium: 39.46%; High: 20.88%
5	Gender	2	0	Male: 51.02%; Female: 48.98%
6	Grade	5	0	A: 29.9%; B: 19.56%; D: 17.78%; F: 16.88%; C: 15.88%
7	Internet_Access_at_Home	2	0	Yes: 89.7%; No: 10.3%
8	Parent_Education_Level	4	1794	nan: 35.88%; PhD: 16.4%; Bachelor's: 16.2%; High School: 15.92%; Master's: 15.6%



Brak outlierów (lub bardzo mało):

Age: Statystyki studentów uczelni wyższych.

Assignments_Avg, Attendance (%), Final_Score, Midterm_Score,

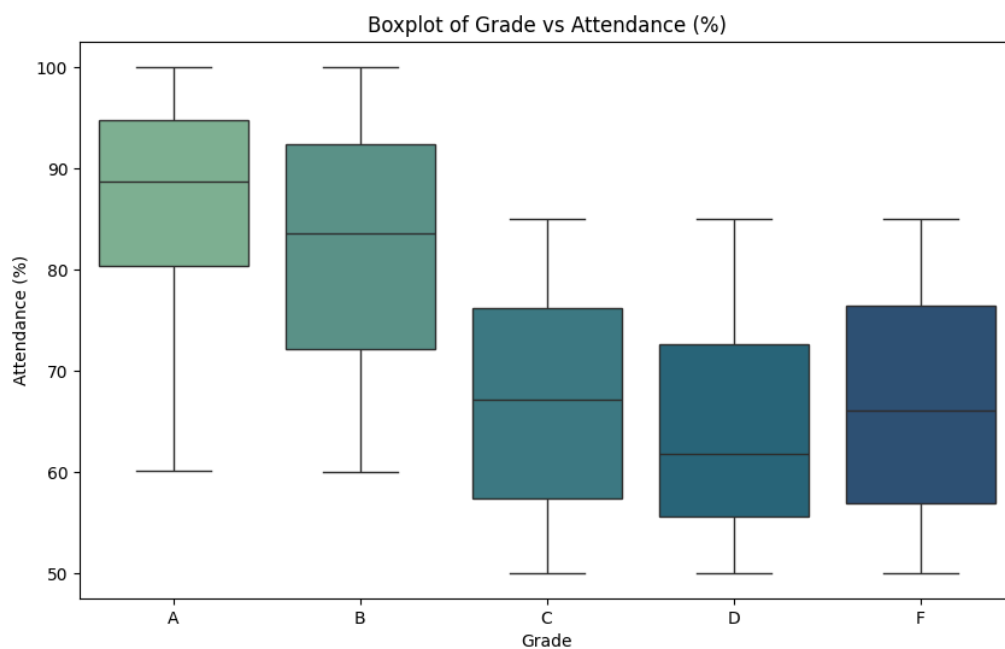
Projects_Score, Total_Score: Dane rozłożone równomiernie, bez wyraźnych wartości odstających.

Widoczne outliery:

Participation_Score: Występują wartości odstające poniżej dolnego wąsa (czyli uczniowie z bardzo niską aktywnością).

Stress_Level (1-10): Outliery po obu stronach – zarówno bardzo niski, jak i bardzo wysoki poziom stresu.

Study_Hours_per_Week: Rozrzut jest duży, widoczne outliery powyżej górnego wąsa – studenci poświęcający **znacznie więcej czasu niż przeciętni**.



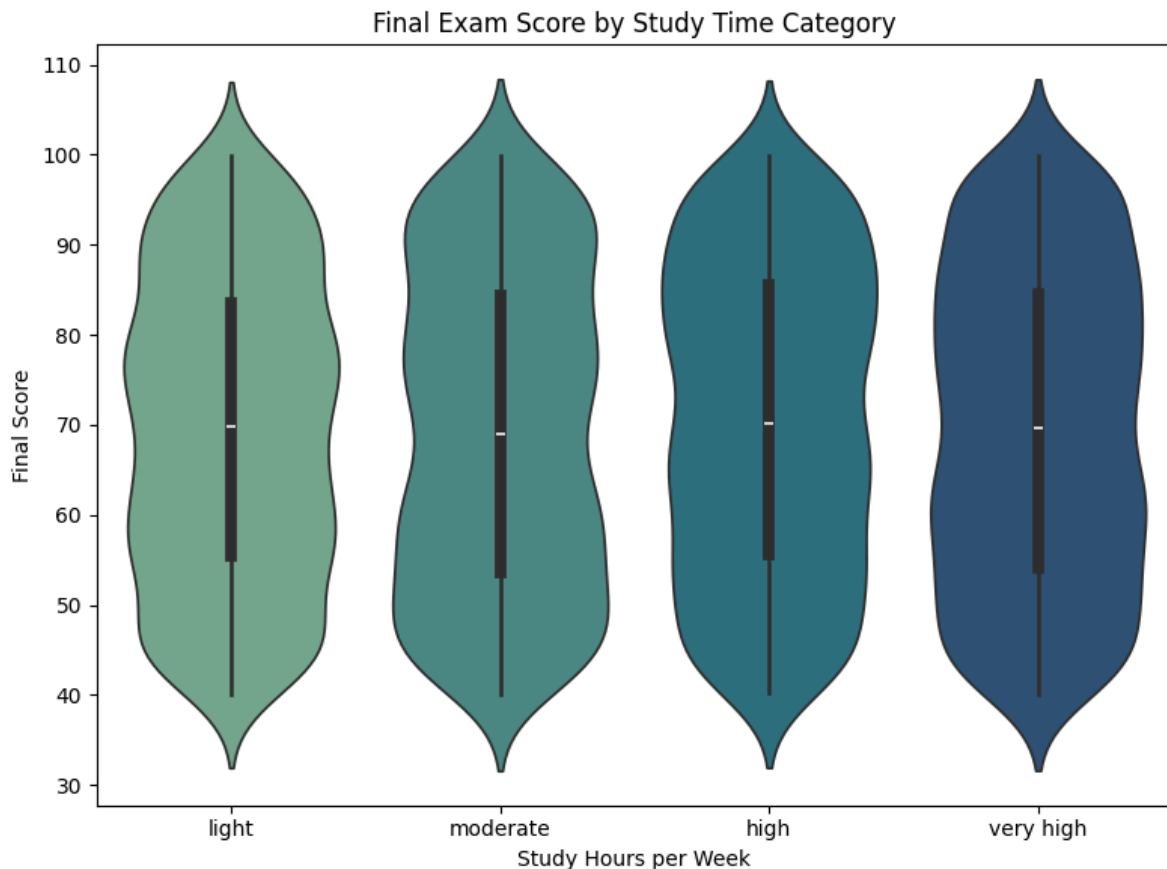
Wyższa frekwencja = lepsza ocena:

- Studenci z oceną A i B mają wyraźnie wyższą medianę frekwencji (~88–90%).
- Dla ocen C, D i F, mediana frekwencji jest niższa (~65–70%), średnia, ale także rozrzut frekwencji maleją.
- U studentów z oceną F, wartości frekwencji są najbardziej rozproszone, z wieloma przypadkami poniżej 60%.

Wniosek:

Silna zależność między frekwencją a wynikami w nauce.

Frekwencja może być jednym z **kluczowych powodów** sukcesu akademickiego.

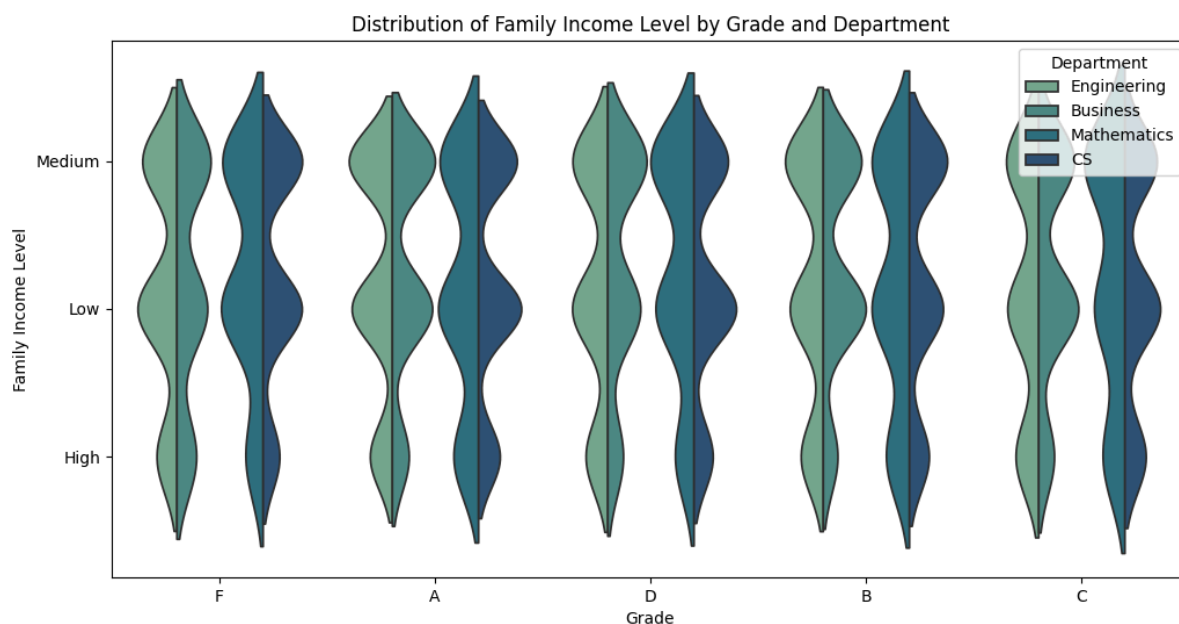


Interpretacja:

- Mediany wyników w każdej grupie (białe kreski wewnątrz) są bardzo **zbliżone**.
- Wszystkie grupy mają bardzo **podobny rozkład wyników** – ok. 40-100 punktów. Również **zagęszczenia (gęstości)** w środku wykresów są podobne.

Wniosek:

- Sama liczba godzin nauki **nie gwarantuje** wyższych wyników z egzaminu końcowego.
- Możliwe, że **efektywność nauki, jakość materiałów, styl uczenia się** lub inne czynniki (np. stres, sen, uczestnictwo) mają większy wpływ.
- W grupie „very high” jest wielu studentów z **niskimi wynikami** mimo dużej liczby godzin nauki – może wskazywać na **przemęczenie, słabą organizację** albo uczenie się na siłę.



Interpretacja:

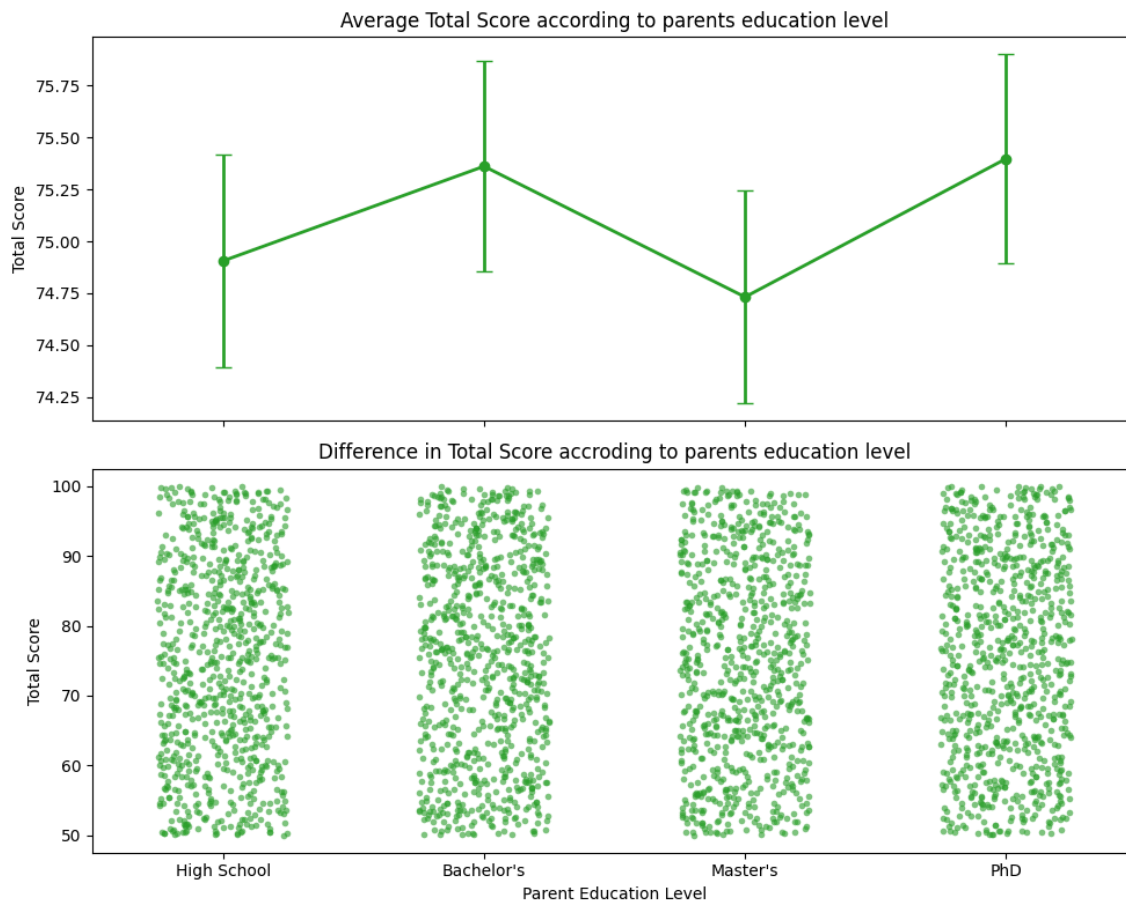
Brak silnej zależności między poziomem dochodów a ocenami

- W każdej grupie ocen (A–F) obecne są wszystkie trzy poziomy dochodów: Low, Medium, High.
- Kształt rozkładów jest dość zbliżony – oznacza to, że dochód rodziny nie różnicuje w oczywisty sposób osiąganych ocen.

Różnice między wydziałami

- Niektóre wydziały (np. CS i Mathematics) wydają się mieć bardziej **wyrównany rozkład dochodów**, podczas gdy np. Engineering może mieć lekki przechył w stronę średniego dochodu.
- Nie ma jednak drastycznych różnic – **wszystkie poziomy dochodów są reprezentowane** we wszystkich ocenach i wydziałach.

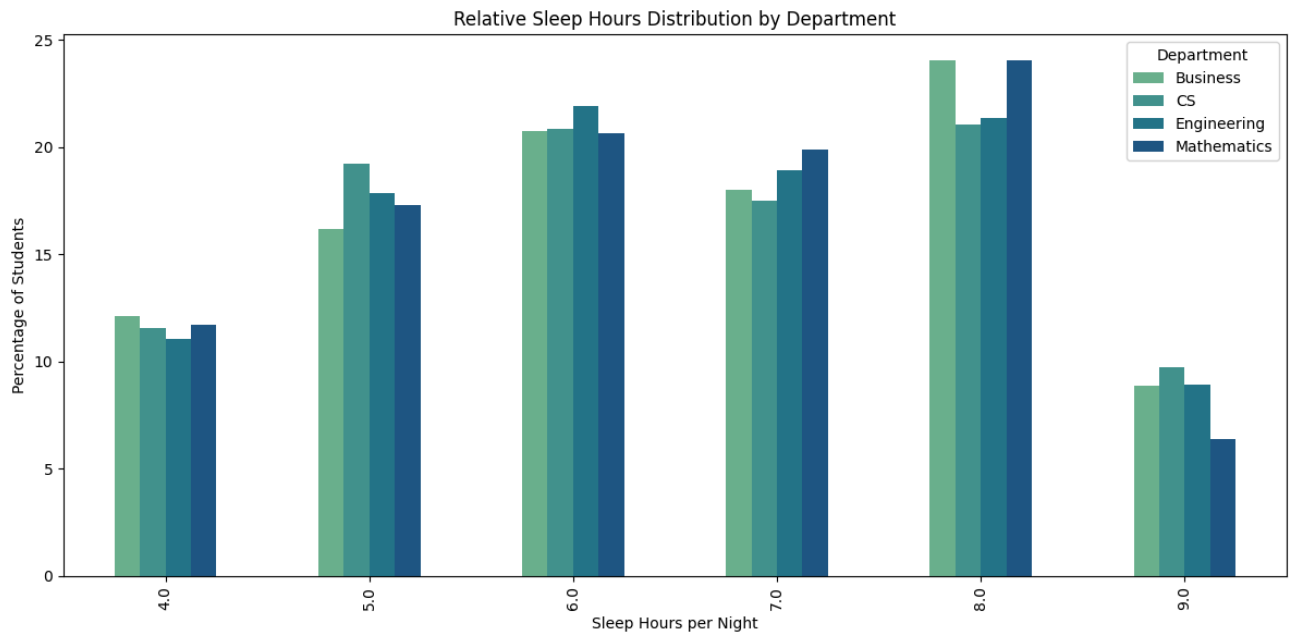
Dochód rodziny nie jest decydującym czynnikiem sukcesu akademickiego w tym zbiorze danych.



- Średni wynik całkowity (Total Score) **nie różni się** między poziomami wykształcenia rodziców.
- Wartości mieszczą się w wąskim zakresie: ~74.7–75.4.
- **Rozkład wyników jest bardzo podobny** we wszystkich grupach wykształcenia.
- Studenci z każdej grupy osiągają zarówno wysokie, jak i niskie wyniki.

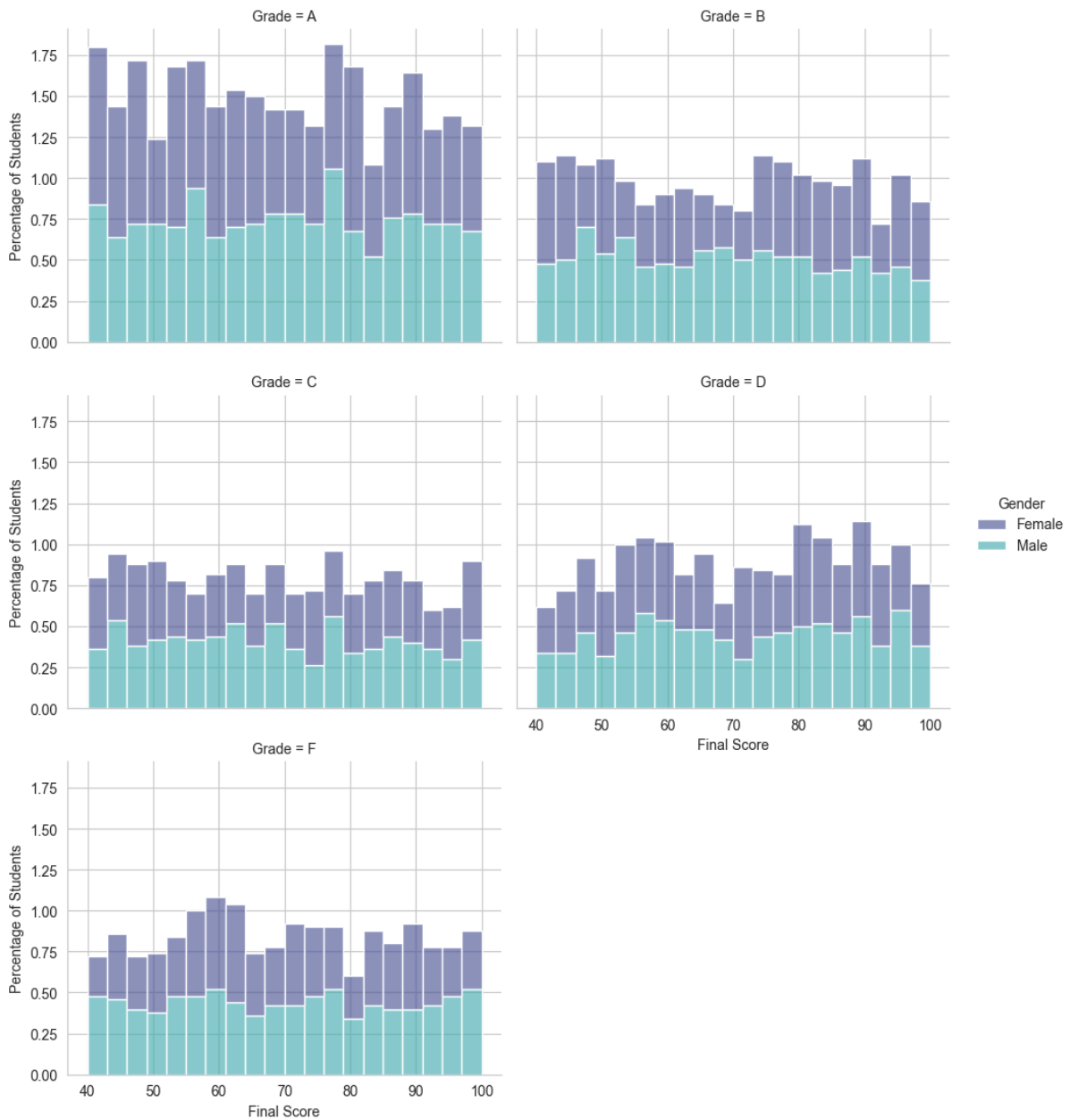
Wniosek:

Poziom wykształcenia rodziców nie ma wyraźnego wpływu na wyniki studentów w tym zbiorze danych. Potwierdzają to zarówno średnie, jak i rozrzut wyników.



- **Najwięcej studentów śpi 6–8 godzin** niezależnie od wydziału — szczególnie popularne są 6 i 8 godzin.
- **Bardzo mało studentów śpi tylko 4 godziny** (ok. 11–12%).
- **Spanie 8 godzin** to najczęstszy przypadek – ok. **24% studentów** na wszystkich wydziałach.
- Studenci **Mathematics** i **Engineering** mają **mniej osób śpiących 9h** niż inne kierunki.
- Ogólny rozkład snu jest **dość podobny** między wydziałami – **brak drastycznych różnic**.

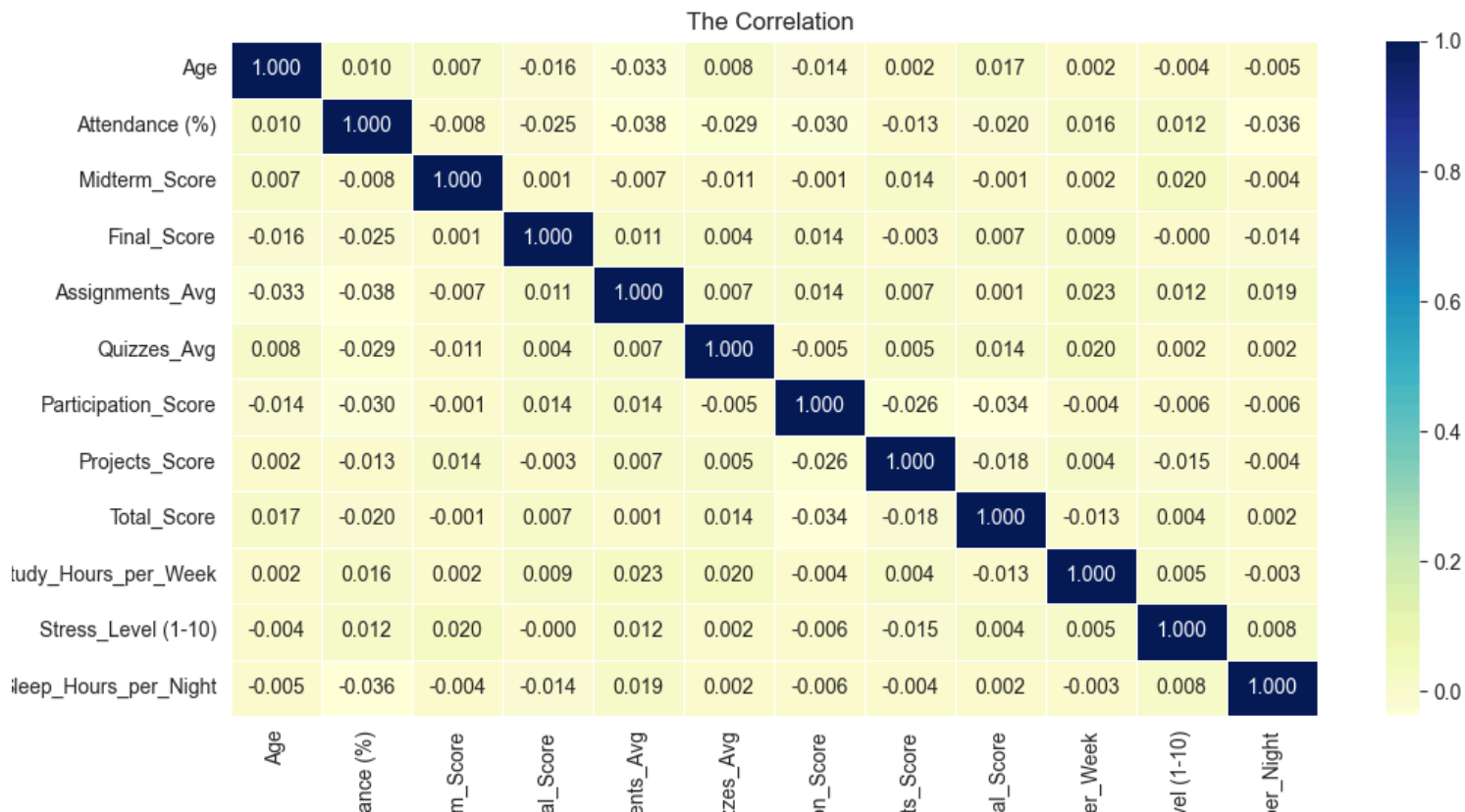
Final Score Distribution by Grade and Gender



Brak różnic płciowych – kobiety i mężczyźni mają podobne rozkłady wyników we wszystkich ocenach (A–F).

Wyniki końcowe (Final Score) są szeroko rozrzucone w każdej ocenie, co sugeruje, że **ocena końcowa zależy** nie tylko od egzaminu.

Płeć **nie wpływa istotnie** na ocenę ani wynik.



Wszystkie cechy mają **bardzo słabe korelacje** z Total_Score (poniżej ± 0.05).

Najsilniejsze korelacje są i tak **niewielkie**: Assignments_Avg (0.014), Final_Score (0.007)

Dalej narzuciłam słownik wag (określają wpływ cech w równaniu, ich suma = 1):

"Participation_Score": 0.05,

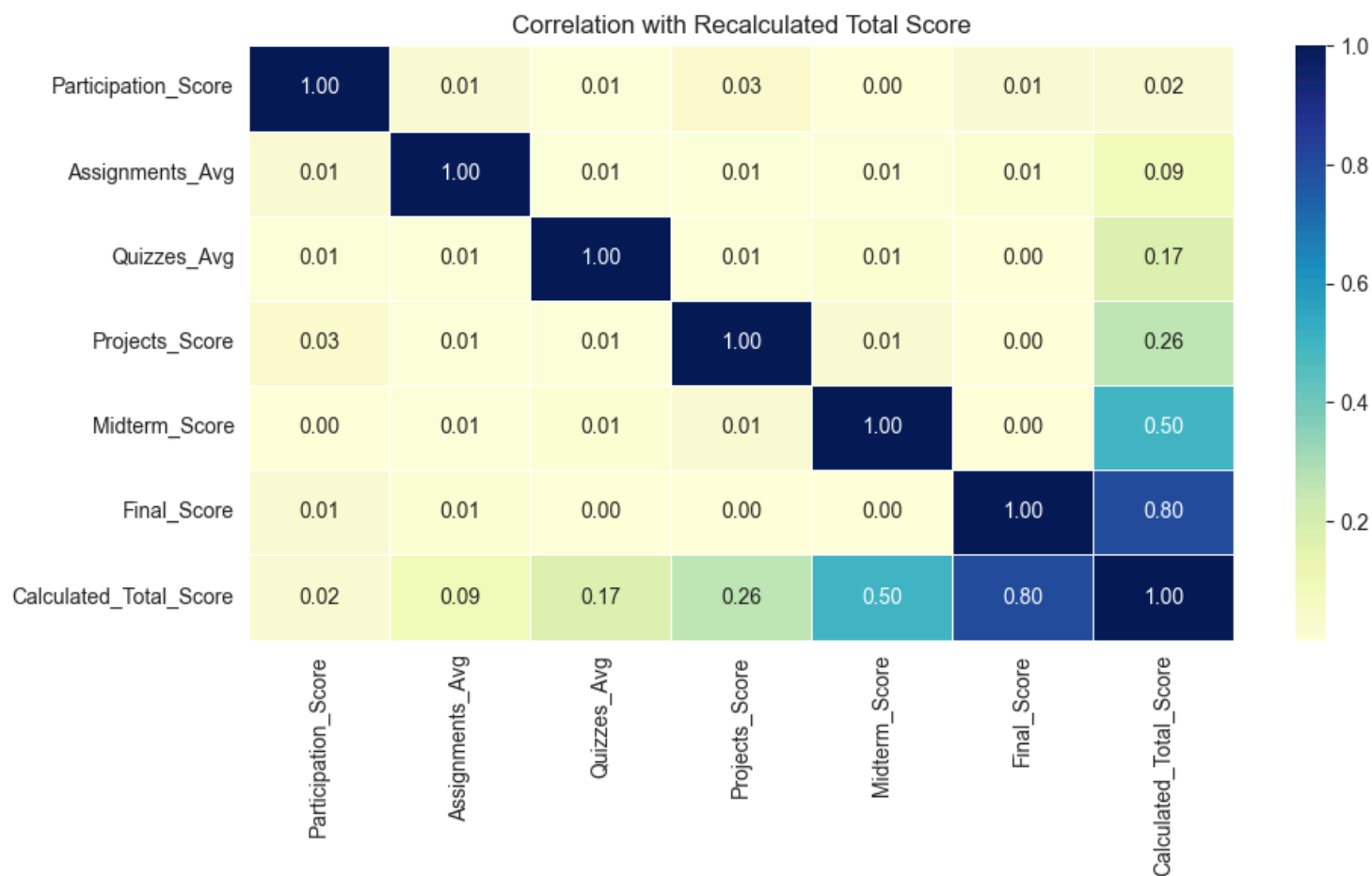
"Assignments_Avg": 0.05,

"Quizzes_Avg": 0.1,

"Projects_Score": 0.15,

"Midterm_Score": 0.25,

"Final_Score": 0.4

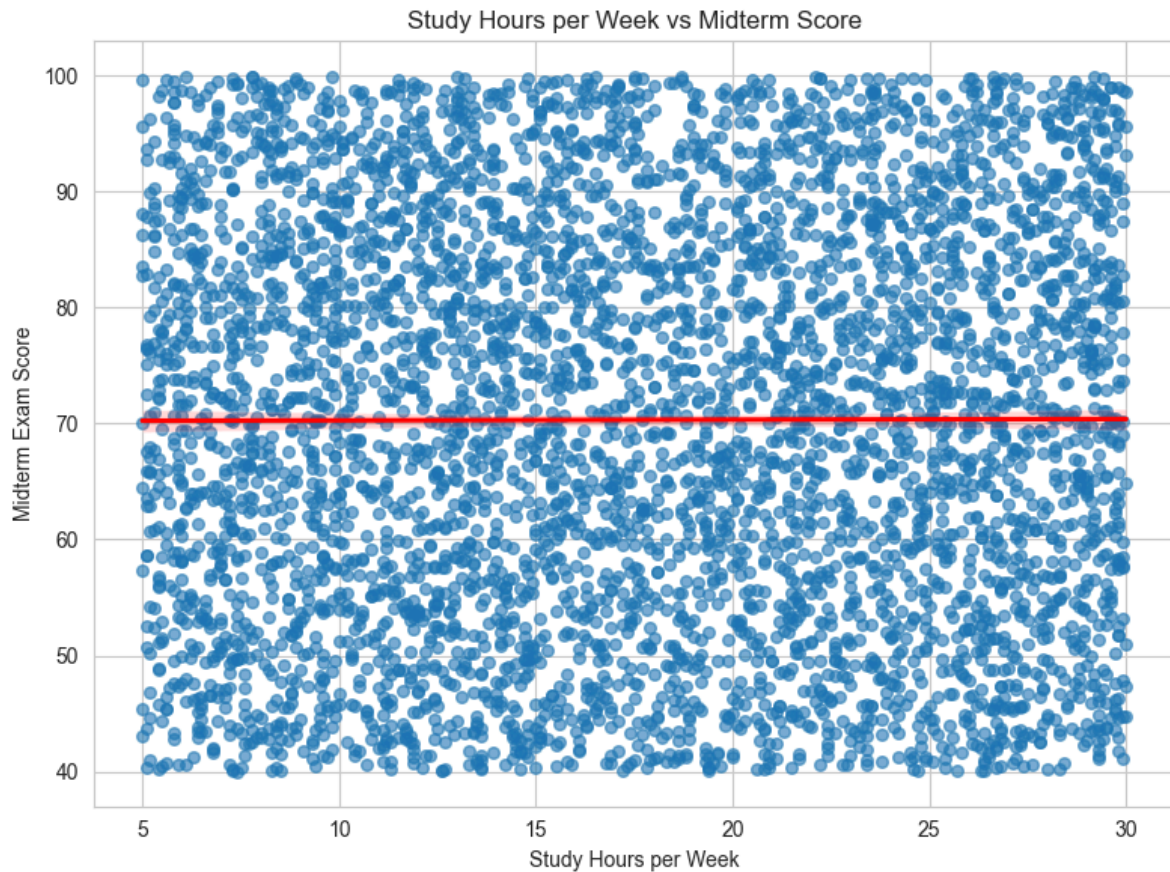


Najsilniejszy wpływ:

- Final_Score ma najwyższą korelację z Calculated_Total_Score (**0.80**) → zgodne z wagą **0.4**.
- Midterm_Score: korelacja **0.50**, również zgodna z wagą **0.25**.

Umiarkowany wpływ:

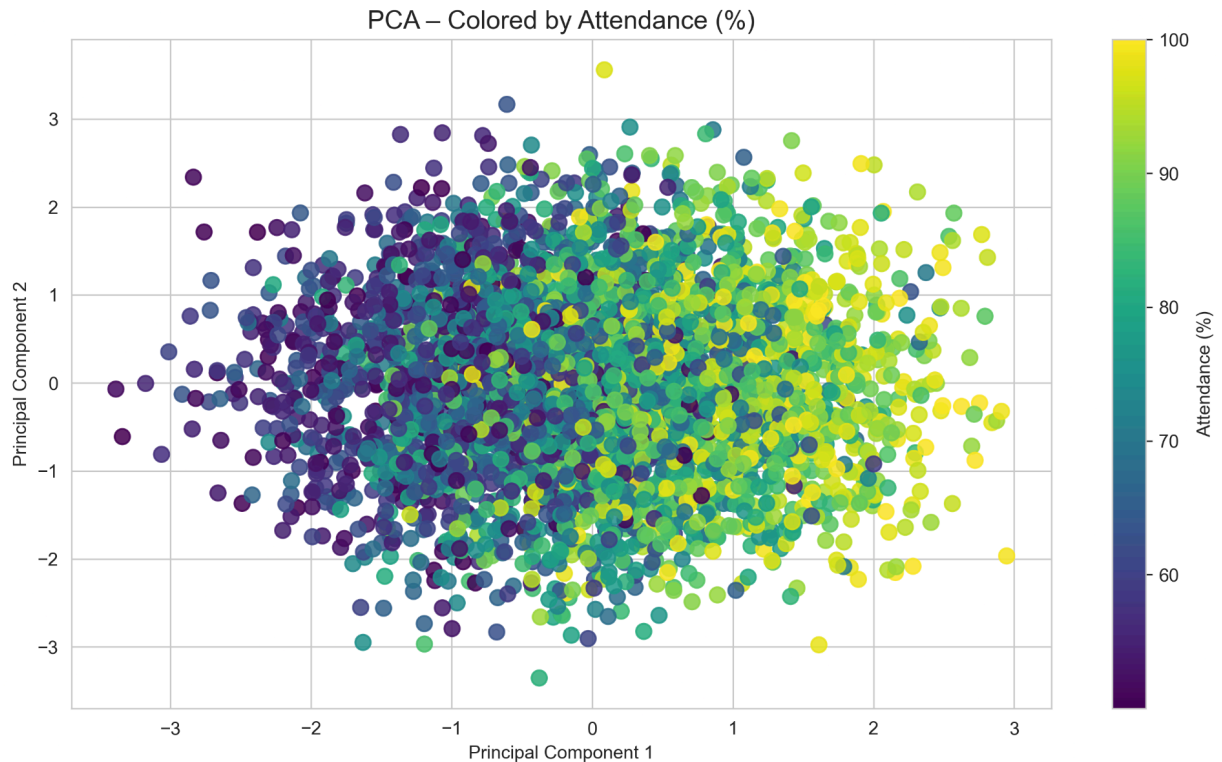
- Projects_Score – korelacja **0.26**, odpowiada wadze **0.15**.
- Quizzes_Avg – korelacja **0.17**, przy wadze **0.10**.
- Assignments_Avg – korelacja **0.09**, przy wadze **0.05**.
- Participation_Score – korelacja **0.02**, waga (0.05), najmniej wpływ



Brak zależności liniowej:

- Linia regresji jest praktycznie płaska - **brak wyraźnej korelacji** między czasem nauki a wynikiem z egzaminu.
- Studenci osiągają **różne wyniki niezależnie od liczby godzin nauki**.

Więcej godzin nauki tygodniowo nie gwarantuje lepszego wyniku z egzaminu śródsesemestralnego. Sama ilość nauki to za mało — liczy się też sposób i efektywność nauki.



- Po lewej stronie wykresu dominują ciemne kolory (niska frekwencja), a po prawej jasne (wysoka frekwencja).
- Oznacza to, że **PC1 dobrze odzwierciedla frekwencję** i prawdopodobnie opisuje **zaangażowanie studenta**.
- Wyższy PC1 = student częściej obecny, bardziej zmotywowany, zorganizowany

Ponieważ PCA uwzględniło zmienne takie jak Attendance (%),

Study_Hours_per_Week, Final_Score, Total_Score to:

PC1 może reprezentować ogólną postawę i skuteczność ucznia:

- dobra frekwencja,
- systematyczna nauka,
- wysokie wyniki,
- dbanie o sen.