

Методи та технології аналітики даних

Недождій Олексій ФФ-93

Пояснення до обрахунків

Обрахунки проводились тільки числовими даними і відповідно були записані тільки вони.

Пропуски я видалив за допомогою метода *dropna()*.

На рахунок викидів потрібно зазначити, що порогові значення я вводив тільки знизу, так як зверху і так маю обмеження на максимальну кількість балів за тест(100). Порогові значення я задав орієнтуючись на графік та медіану. Вони були наступними:

1. Math score: (32, 100)
2. Reading score: (30, 100)
3. Writing score: (28, 100)

При середній нормалізації не створюються допустимі межі значень, натомість ми гарантуємо, що середнє значення завжди буде рівне нулю, а стандартне відхилення одиниці.

Файли

- Очищенні від пустих рядків данні зберіг в файл *exam_cleared.csv*.
- обраховані данні збереженні в файлі *calculated.csv*.
- код для обрахунку і запису файл називається *calculation.py*.
- нормалізовані данні збереженні в файлі *minmax_normalized.csv* та *mean_normalized.csv* відповідно до типу нормалізації.
- графікі збереженні в файлах *reading score.jpg*, *writing score.jpg* та *math score.jpg*, відповідно до даних які вони відображають.
- код для виведення графіків знаходиться в файлі *visualisation.py*.
- за допомогою файлу *main.py* можна запустити обрахунок і візуалізацію одночасно.