Міністерство освіти і науки України

Західноукраїнський національний університет

Факультет комп’ютерних інформаційних технологій

Кафедра інформаційно-обчислювальних систем і управління

Звіт до лабораторної роботи №5

З дисципліни “Інтелектуальний аналіз даних”

Студента групи КН-31

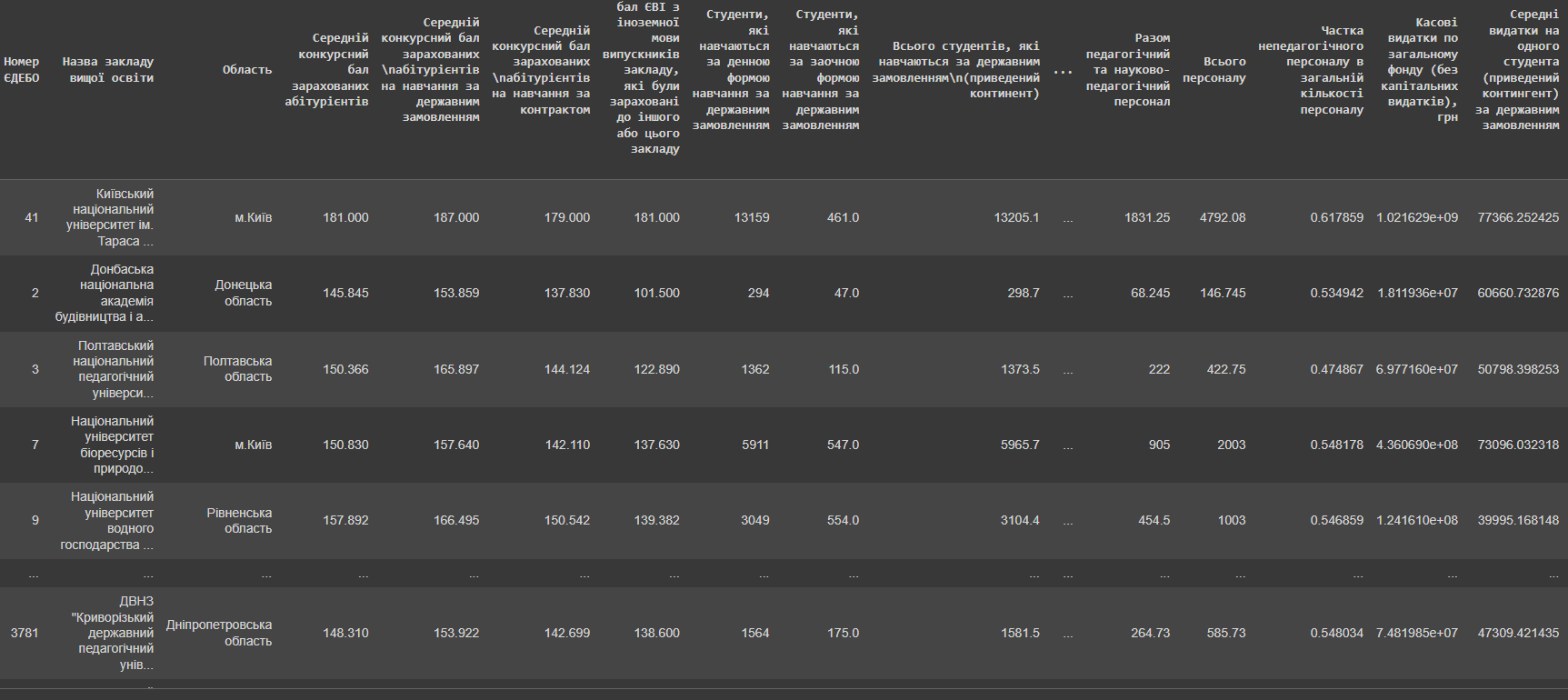
Ковальковський В. В.

Тернопіль 2023

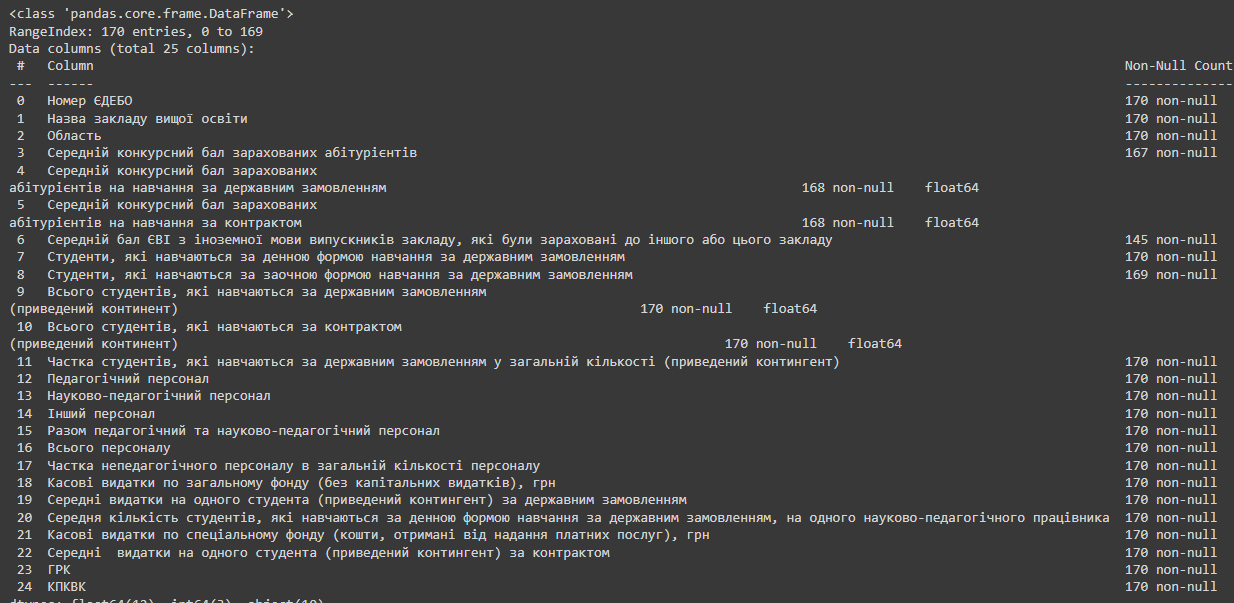
ЛАБОРАТОРНА РОБОТА 5:

ЛІНІЙНА РЕГРЕСІЯ (Витрати на ВНЗ за 2019)

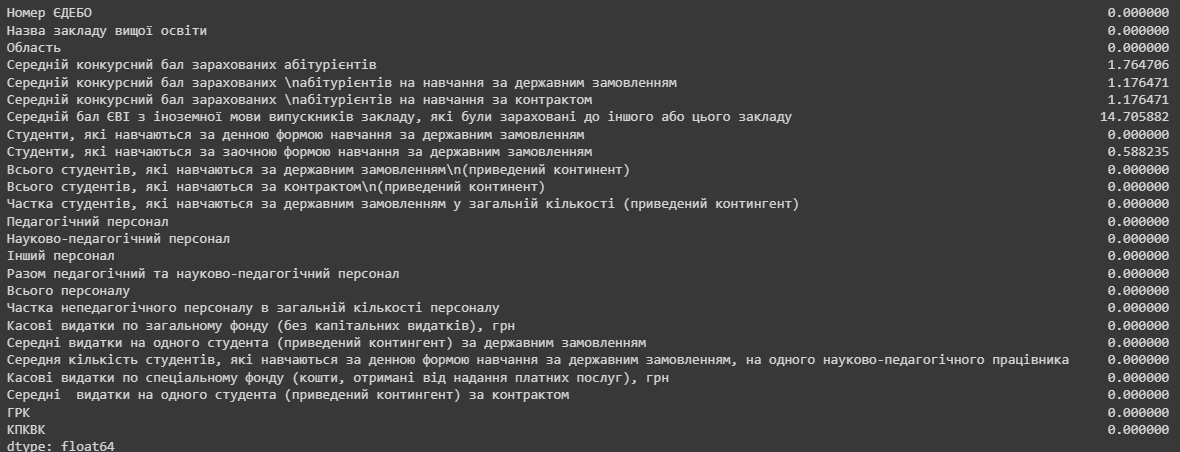
**Читання датасету:**



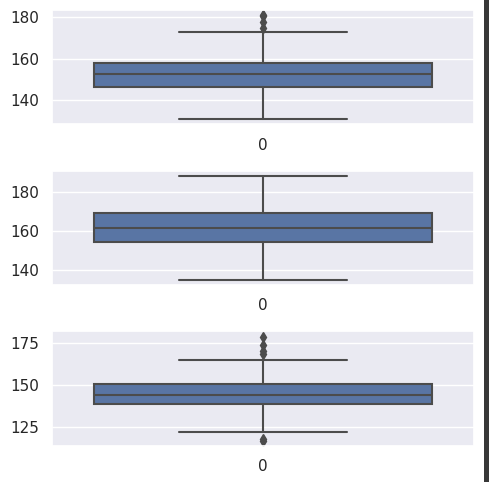
**Перевірка даних:**



**Очищення даних:**



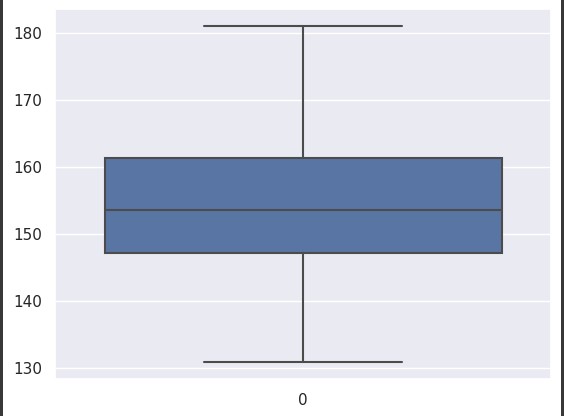
**Аналіз викидів**, для колонок - Середній конкурсний бал зарахованих абітурієнтів, Середній конкурсний бал зарахованих абітурієнтів на навчання за державним замовленням, Середній конкурсний бал зарахованих абітурієнтів на навчання за контрактом:

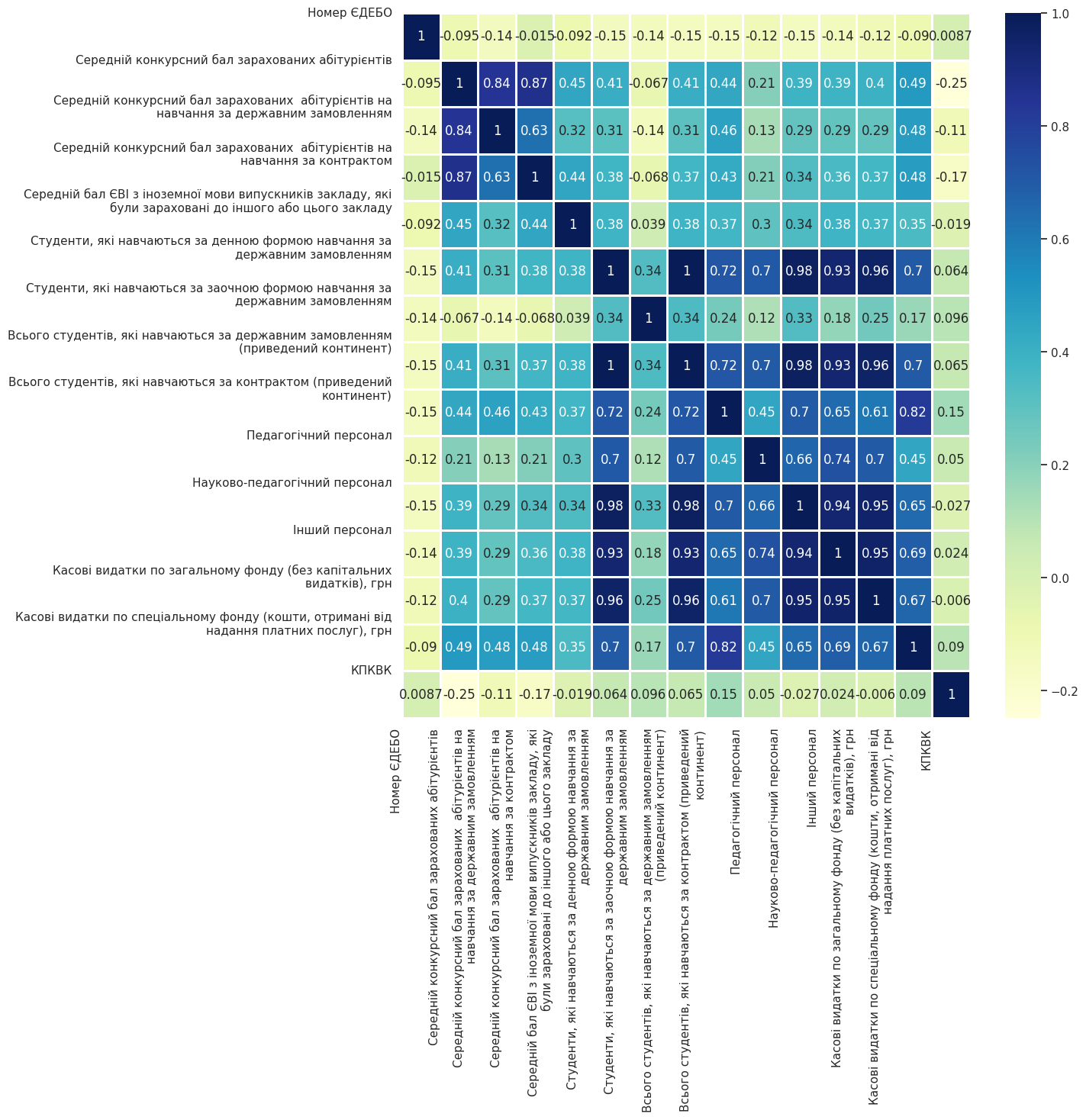


**Розвідувальний аналіз даних**

Одновимірний аналіз

Середній конкурсний бал зарахованих абітурієнтів (цільова змінна):

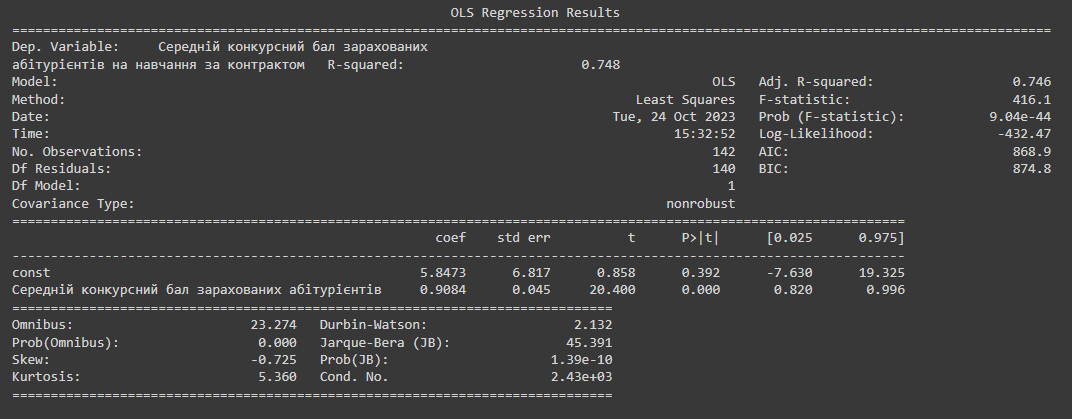


**Кореляція:** 

**Побудова моделі:**

Параметри:





**Коефіцієнти регресії:**

const: Це константа або зміщення, що вказує на значення залежної змінної, коли всі незалежні змінні рівні нулю. У цьому випадку, це 5.8473.

Середній конкурсний бал зарахованих абітурієнтів: Цей коефіцієнт (0.9084) вказує на те, наскільки змінюється залежна змінна для одиничного одиниці змінної "Середній конкурсний бал зарахованих абітурієнтів".

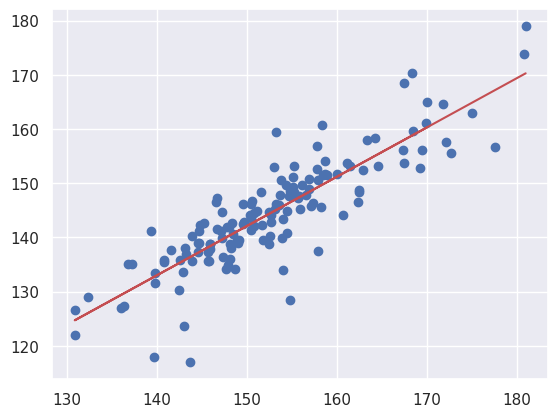
**Статистичні показники:**

R-squared: Це коефіцієнт детермінації, який вказує на те, яка частина варіації у залежній змінній пояснюється незалежною змінною у моделі. У цьому випадку, це 0.748, що означає, що 74.8% варіації у залежній змінній пояснюється змінною "Середній конкурсний бал зарахованих абітурієнтів".

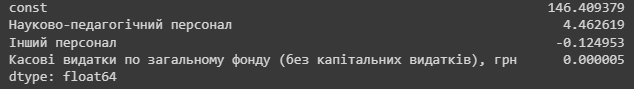
F-statistic: Це значення статистики F для перевірки значущості моделі в цілому. У цьому випадку, величина статистики F дорівнює 416.1, що є великою відносно кількості спостережень, що підтверджує значущість моделі.

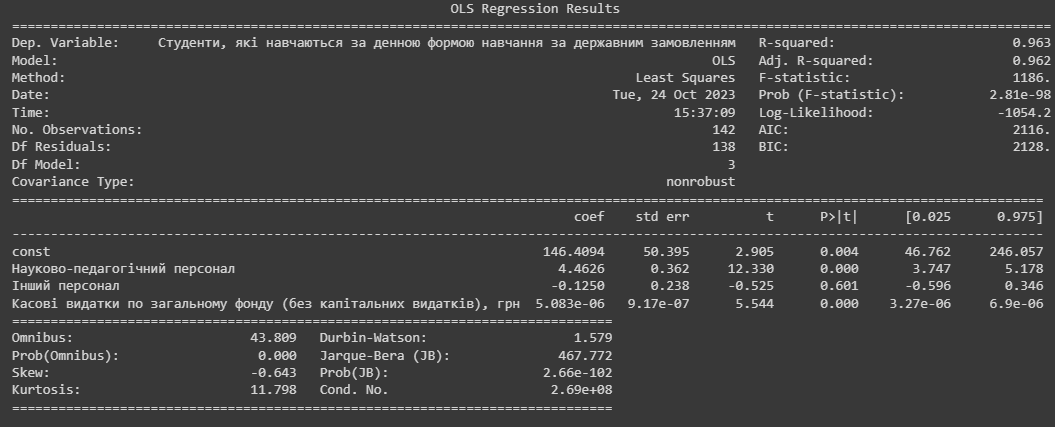
Prob (F-statistic): Це ймовірність, пов'язана зі значенням статистики F. У цьому випадку ймовірність дуже низька (9.04e-44), що свідчить про статистичну значущість моделі.

**Лінійне рівняння регресії:**



**Виконання множинної лінійної регресії:**

****

****

**Коефіцієнти регресії:**

Науково-педагогічний персонал: Коефіцієнт регресії для цієї змінної становить 4.4626, що вказує на те, що з кожним одиницею збільшення цієї змінної "Студенти, які навчаються за денною формою навчання за державним замовленням" збільшується на 4.4626 одиниць.

Інший персонал: Коефіцієнт регресії для цієї змінної становить -0.1250. Однак важливо зауважити, що цей коефіцієнт має високий p-значення (0.601), що означає, що ця змінна, можливо, не є статистично значущою для прогнозування змінної "Студенти, які навчаються за денною формою навчання за державним замовленням".

Касові видатки по загальному фонду (без капітальних видатків), грн: Коефіцієнт регресії для цієї змінної становить 5.083e-06 (5.083 \* 10^(-6)). Це величина з крапкою, і вона вказує на те, на скільки змінюється залежна змінна при збільшенні цієї змінної на одиницю.

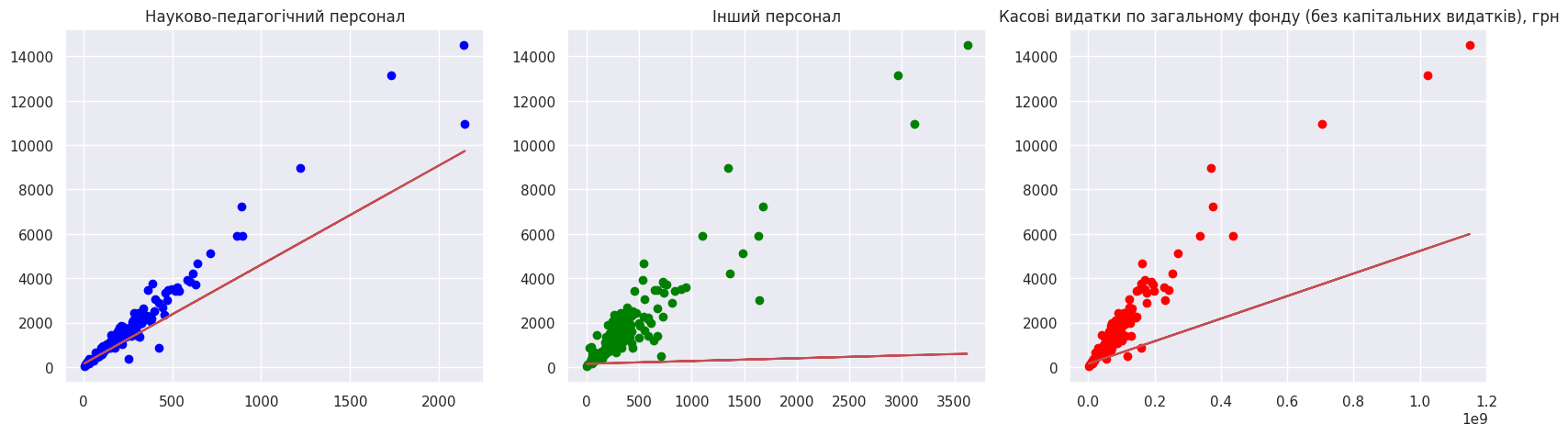
**Основні статистичні показники:**

R-squared: Це високий коефіцієнт детермінації (0.963), що означає, що більшість варіації змінної "Студенти, які навчаються за денною формою навчання за державним замовленням" пояснюється незалежними змінними в цій моделі.

F-statistic: Значення статистики F для перевірки значущості моделі в цілому є дуже великим (1186) і має дуже низьке p-значення (2.81e-98), що свідчить про статистичну значущість моделі.

Prob (F-statistic): Це ймовірність, пов'язана зі значенням статистики F. Якщо вона низька (як в цьому випадку), це свідчить про значущість моделі.

**Лінійні рівняння регресії:**

****

Висновок: під час виконання лабораторної роботи, будував моделі лінійної регресії, на основі змінних, які визначив за теплокартою кореляцій, зробив висновки по результатах моделювання.