Міністерство освіти і науки України

Західноукраїнський національний університет

Факультет комп’ютерних інформаційних технологій

Кафедра інформаційно-обчислювальних систем і управління

Звіт до лабораторної роботи №6

З дисципліни “Інтелектуальний аналіз даних”

Студента групи КН-31

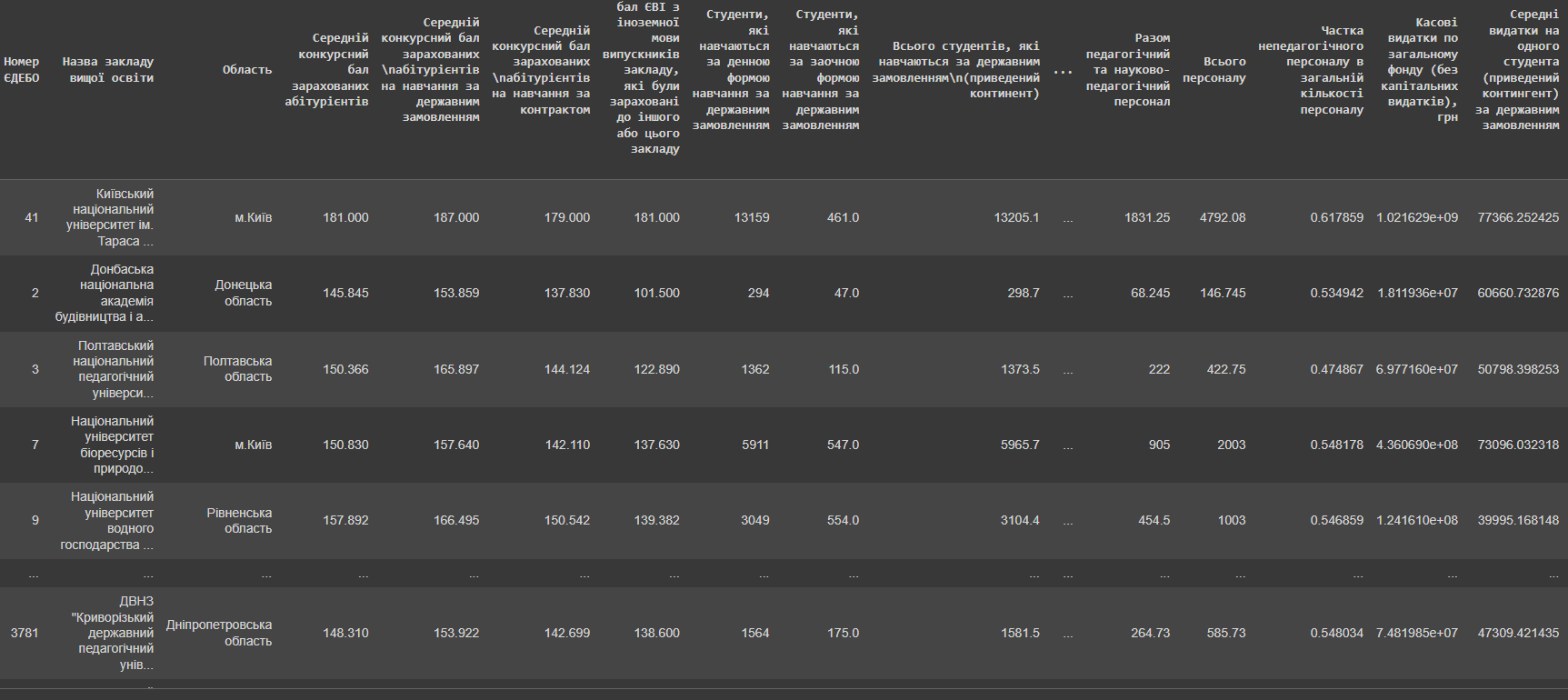
Ковальковський В. В.

Тернопіль 2023

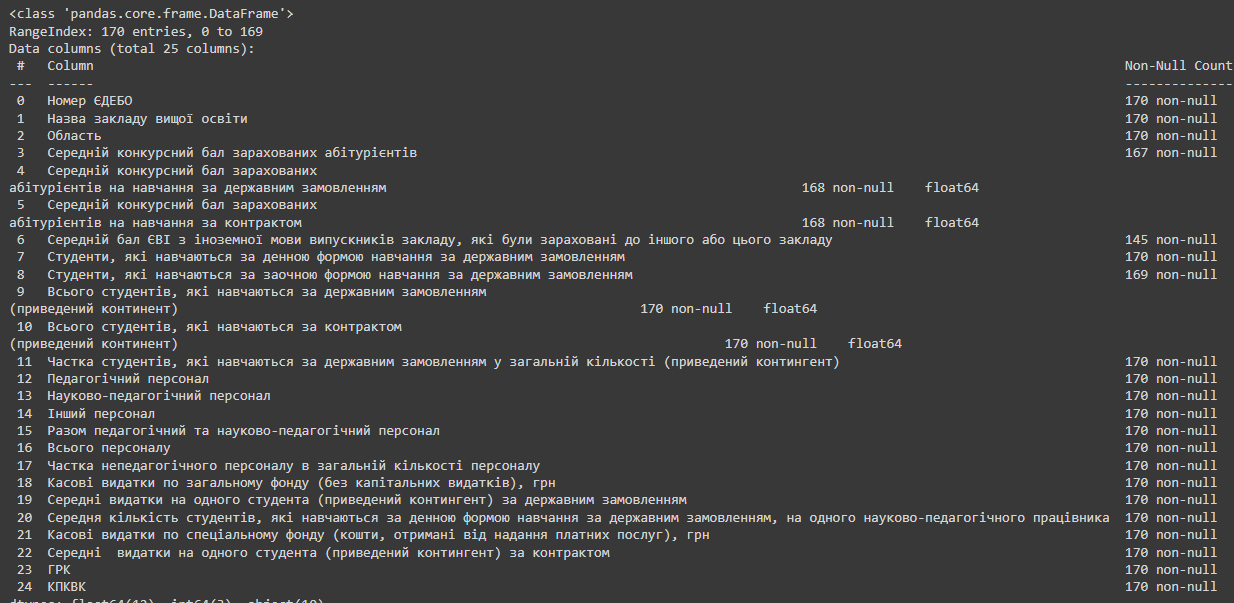
ЛАБОРАТОРНА РОБОТА 6:

НЕ ЛІНІЙНА РЕГРЕСІЯ (Витрати на ВНЗ за 2019)

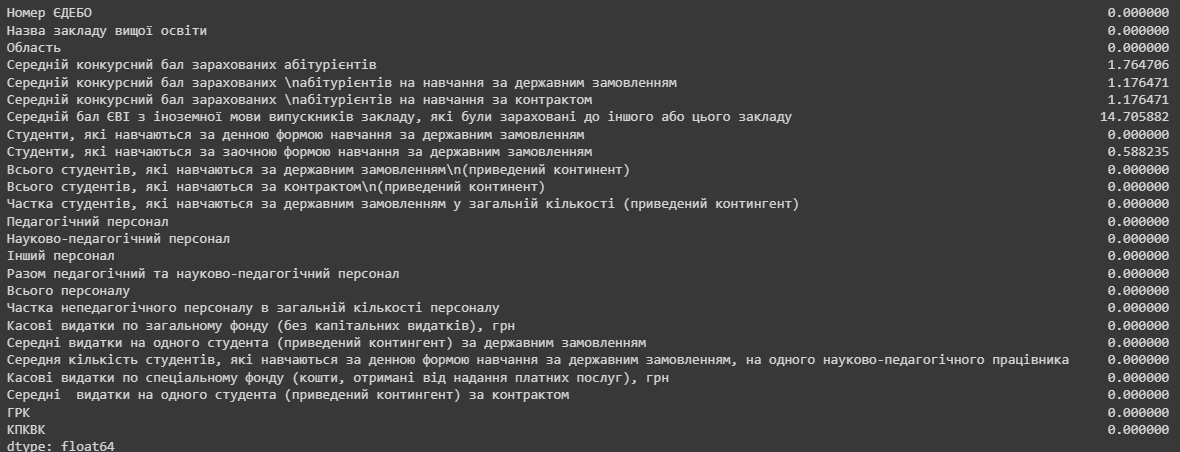
**Читання датасету:**



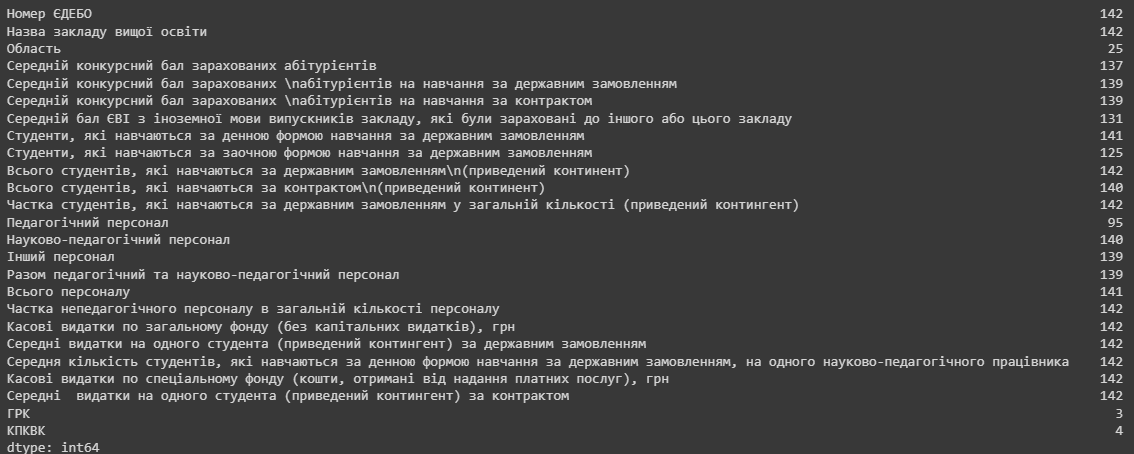
**Перевірка даних:**

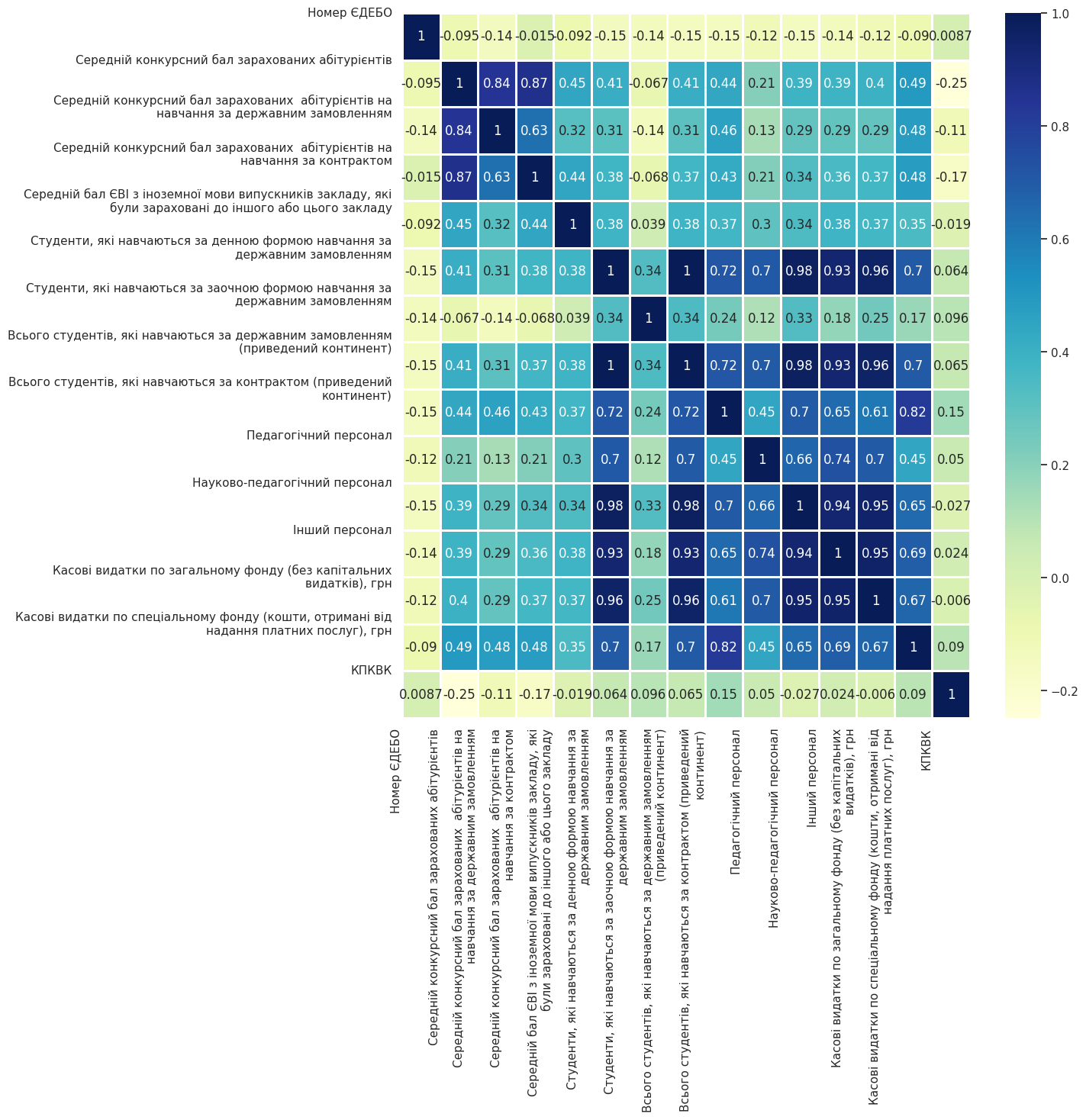


**Очищення даних:**



**Аналіз унікальності даних:**



**Кореляція:** 

****

Графік показує середні видатки на одного студента (приведений контингент) за контрактом для різних груп за середнім конкурсним балом зарахованих абітурієнтів.

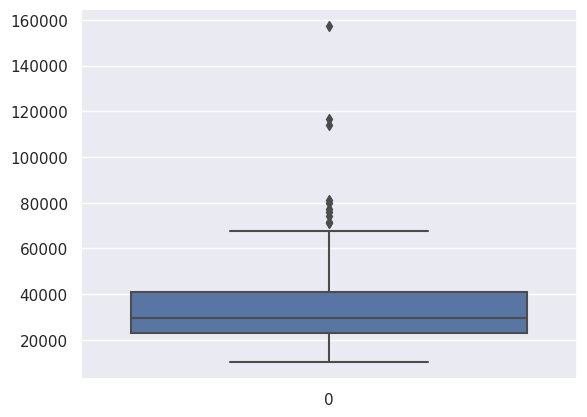
На графіку видно, що середні видатки на одного студента (приведений контингент) за контрактом збільшуються із зростанням середнього конкурсного балу зарахованих абітурієнтів. Це може бути пов'язано з тим, що в цих вузах вищий рівень вимог до вступу, і тому вони можуть залучати більш конкурентоспроможних абітурієнтів. Ці абітурієнти, як правило, мають більш високий рівень доходу, і тому вони можуть платити більше за навчання.

**Ось кілька конкретних висновків, які можна зробити на основі цього графіка:**

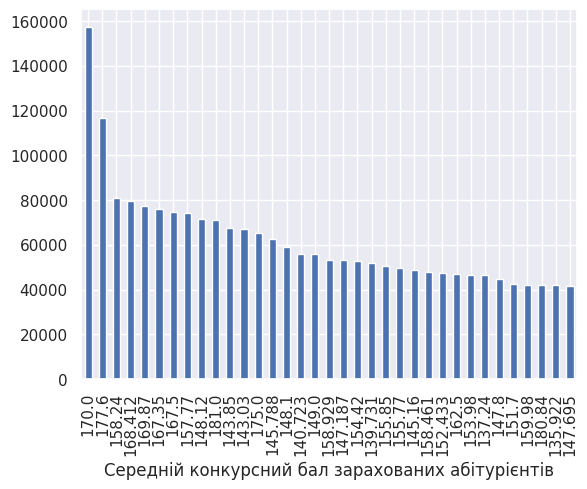
В цілому, середні видатки на одного студента (приведений контингент) за контрактом збільшуються із зростанням середнього конкурсного балу зарахованих абітурієнтів.

Ця тенденція є досить сильною, і вона спостерігається для всіх груп за середнім конкурсним балом зарахованих абітурієнтів.

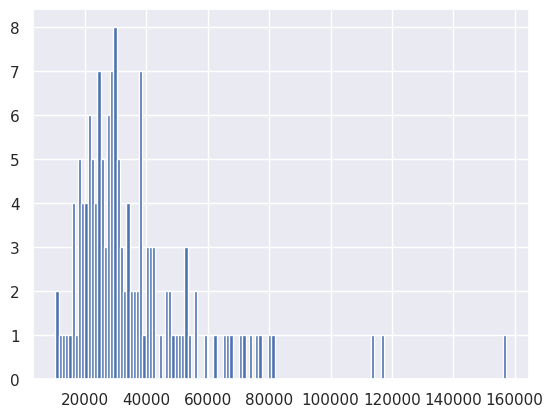
**Аналіз викидів для колонки «Середні видатки на одного студента (приведений контингент) за контрактом»:**

****

**Залежність витрат на одного студента від середнього конкурсного балу зарахованого абітурієнта:**

****

**Частотний розподіл середніх видатків на одного студента (приведений контингент) за контрактом**

****

Цей графік показує частотний розподіл середніх видатків на одного студента (приведений контингент) за контрактом. На основі цього графіка можна зробити такі висновки:

Більшість університетів мають середні видатки на одного студента за контрактом в діапазоні від 10 до 30 тисяч гривень. Цей діапазон є найпоширенішим і включає в себе близько 75% всіх університетів. Є кілька університетів з дуже високими середніми видатками на одного студента за контрактом, які перевищують 50 тисяч гривень. Це, як правило, приватні університети або університети з високим рейтингом. Є також кілька університетів з дуже низькими середніми видатками на одного студента за контрактом, які нижче 10 тисяч гривень. Це, як правило, державні університети в регіонах з низьким рівнем доходу. На основі цих висновків можна зробити висновок, що середні видатки на одного студента за контрактом в Україні в цілому є відносно низькими. Це може бути пов'язано з тим, що в Україні ще донедавна переважали державні університети, які отримують фінансування від держави. Однак в останні роки в Україні спостерігається зростання кількості приватних університетів, які, як правило, мають вищі середні видатки на одного студента.



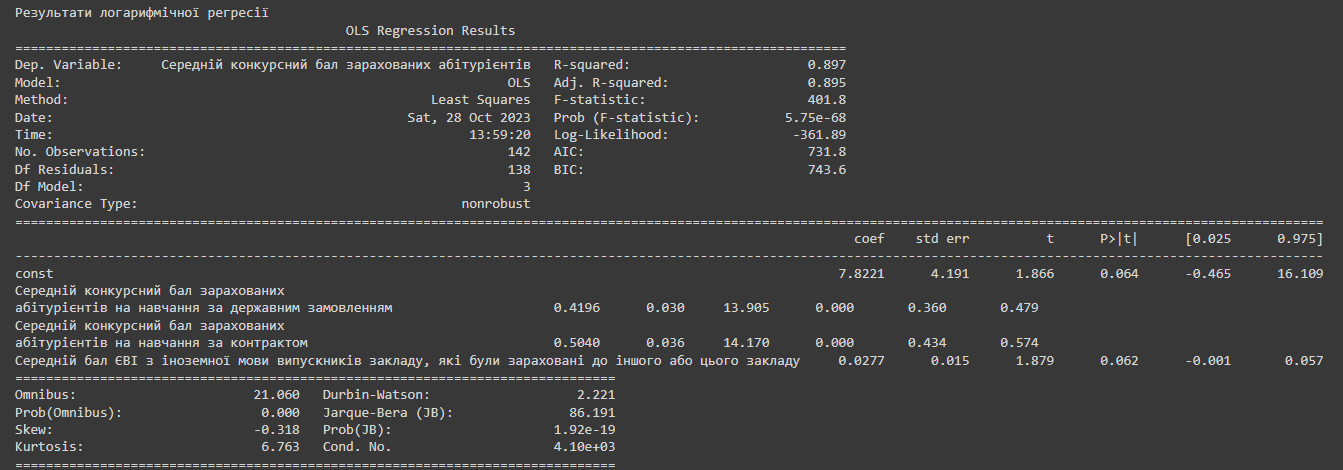
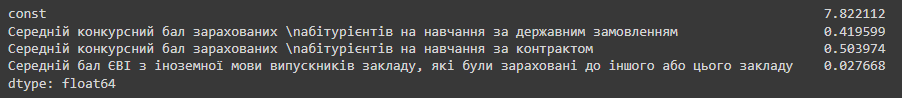
Графік показує залежність середнього конкурсного балу зарахованих абітурієнтів на навчання за контрактом від середнього конкурсного балу зарахованих абітурієнтів.

На графіку видно, що в цілому, чим вищий середній конкурсний бал зарахованих абітурієнтів, тим вищий середній конкурсний бал зарахованих абітурієнтів на навчання за контрактом. Однак, існує кілька відхилень від цієї загальної тенденції. Наприклад, є кілька точок, де середній конкурсний бал зарахованих абітурієнтів на навчання за контрактом вище середнього конкурсного балу зарахованих абітурієнтів. Це може бути пов'язано з тим, що в цих випадках у вузах були додаткові бюджетні місця, які були заповнені абітурієнтами з високими конкурсними балами.

Ось кілька конкретних висновків, які можна зробити на основі цього графіка:

В цілому, чим вищий середній конкурсний бал зарахованих абітурієнтів, тим вищий середній конкурсний бал зарахованих абітурієнтів на навчання за контрактом. Більш високі середні видатки на одного студента (приведений контингент) за контрактом, як правило, пов'язані з більш високим середнім конкурсним балом зарахованих абітурієнтів на навчання за контрактом. Існує кілька відхилень від цієї загальної тенденції, які можуть бути пов'язані з додатковими бюджетними місцями або іншими факторами.4

**Логаритмічна регресія**



На основі результатів логарифмічної регресії можна зробити такі висновки:

Середній конкурсний бал зарахованих абітурієнтів є значущим предиктором середнього конкурсного балу зарахованих абітурієнтів. Тобто, чим вищий конкурсний бал абітурієнтів, які були зараховані на навчання за державним замовленням або контрактом, тим вищий конкурсний бал абітурієнтів, які будуть зараховані в майбутньому.

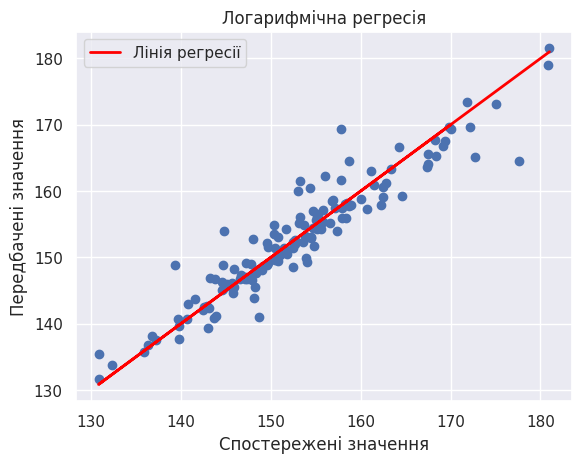
Середній бал ЄВІ з іноземної мови випускників закладу, які були зараховані до іншого або цього закладу, також є значущим предиктором середнього конкурсного балу зарахованих абітурієнтів. Тобто, чим вищий бал ЄВІ з іноземної мови випускників закладу, тим вищий конкурсний бал абітурієнтів, які будуть зараховані в майбутньому.

У цілому, результати регресії свідчать про те, що середній конкурсний бал зарахованих абітурієнтів залежить від двох основних факторів:

Середнього конкурсного балу абітурієнтів, які були зараховані на навчання за державним замовленням або контрактом у минулому році.

Середнього балу ЄВІ з іноземної мови випускників закладу, які були зараховані до іншого або цього закладу у минулому році.

Ці фактори можна використовувати для прогнозування середнього конкурсного балу зарахованих абітурієнтів у майбутньому.



**Аналіз результатів логарифмічної регресії показує наступне:**

R-squared (R^2): R-squared дорівнює 0.897, що означає, що приблизно 89.7% варіації в середньому конкурсному балу зарахованих абітурієнтів може бути пояснено цією моделлю. Це свідчить про досить високий рівень поясненої варіації.

F-statistic і Prob (F-statistic): Значення F-statistic дорівнює 401.8, і ймовірність (Prob) F-statistic дуже низька (5.75e-68), що вказує на те, що модель в цілому є статистично значущою.

Коефіцієнти (coef): Коефіцієнти перед незалежними змінними показують, що збільшення середнього конкурсного бала зарахованих абітурієнтів на навчання за державним замовленням на одиницю призводить до збільшення середнього конкурсного бала на 0.4196 балів, для навчання за контрактом - на 0.5040 бали, та для середнього балу ЄВІ з іноземної мови випускників - на 0.0277 балів.

p-value (P>|t|): P-value для кожного коефіцієнта також свідчить про їхню статистичну значущість, оскільки всі p-value дуже низькі.

Omnibus, Jarque-Bera, Skew і Kurtosis: Значення Omnibus та Jarque-Bera низькі, що вказує на можливу нормальність розподілу залишків моделі. Значення Skew близьке до 0, що означає низьку асиметрію розподілу. Значення Kurtosis також вказують на нормальність розподілу.

Cond. No. (Умовне число): Умовне число досить велике (4.10e+03), що може вказувати на можливу проблему з мультиколінеарністю.

У цілому, результати регресії свідчать про те, що модель досить добре пояснює залежність між середнім конкурсним балом зарахованих абітурієнтів і незалежними змінними. Високий R-squared і низькі p-value свідчать про статистичну значущість моделі, і коефіцієнти регресії показують важливий внесок кожної незалежної змінної в прогнозування середнього конкурсного бала. Однак слід враховувати можливу проблему з мультиколінеарністю, і, можливо, провести додатковий аналіз для підтвердження стійкості моделі.

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