



ИССЛЕДОВАНИЕ ЦЕНЫ ГОРНОЛЫЖНЫХ КУРОРТОВ

Индивидуальное домашнее задание №2
по дисциплине «Математическая
статистика»



**ВЫПОЛНИЛИ
ЗАДАНИЕ**

Студенты группы БИВТ-22-16

Горячева Е.К.

Князева С.А.

Пластинина Е.Д.

Николина В.М.

ЦЕЛЬ ИССЛЕДОВАНИЯ

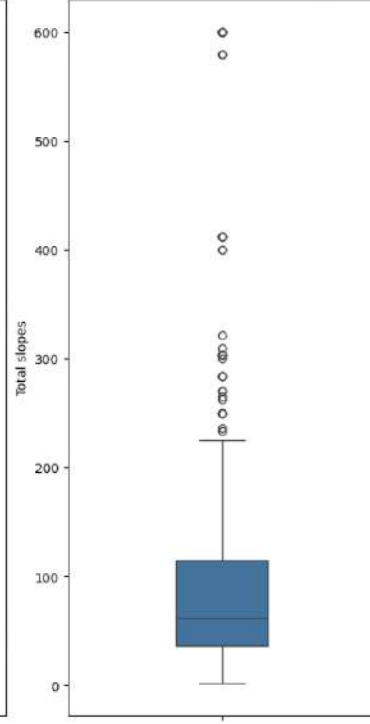
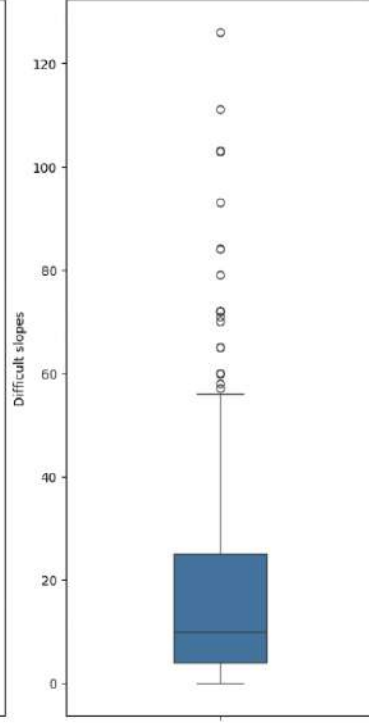
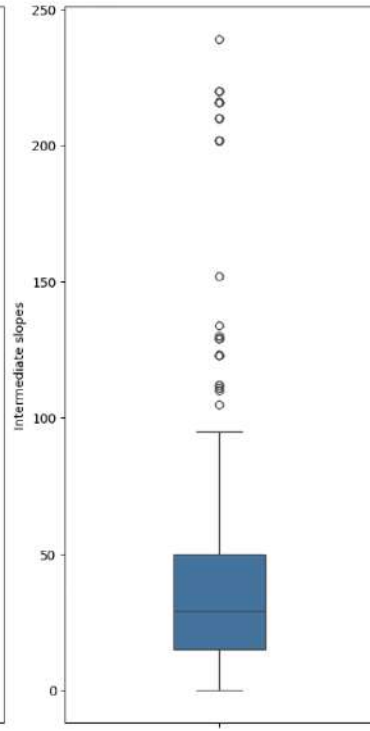
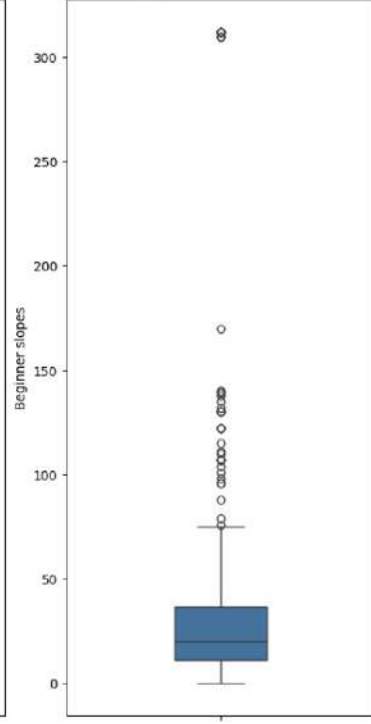
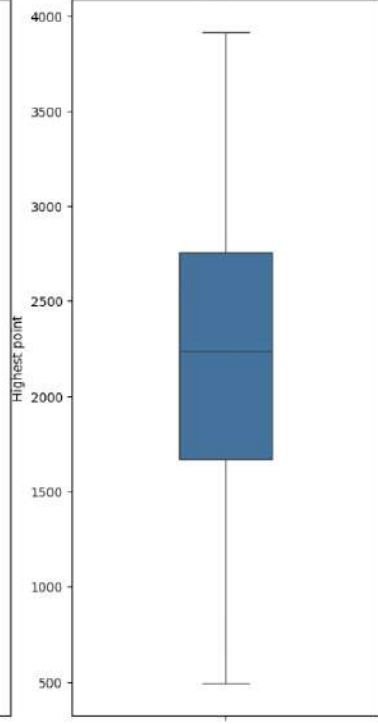
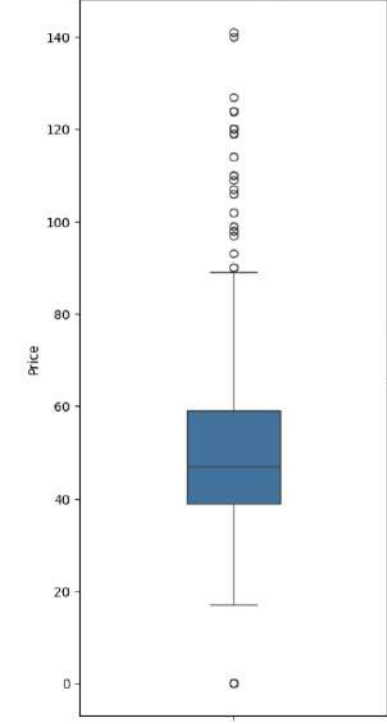
**АНАЛИЗ ФАКТОРОВ ЦЕН НА
ГОРНОЛЫЖНЫЕ КУОРТЫ**

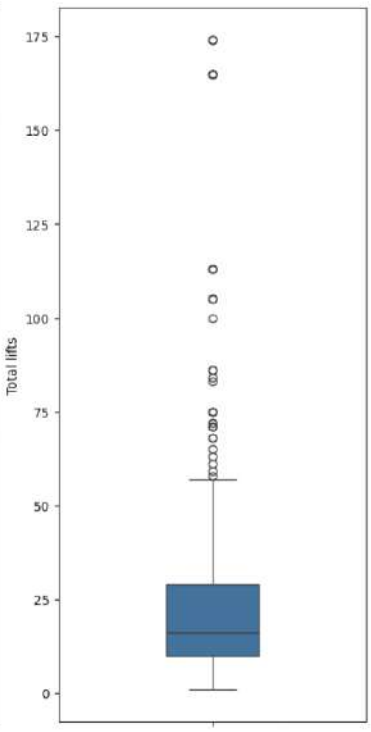
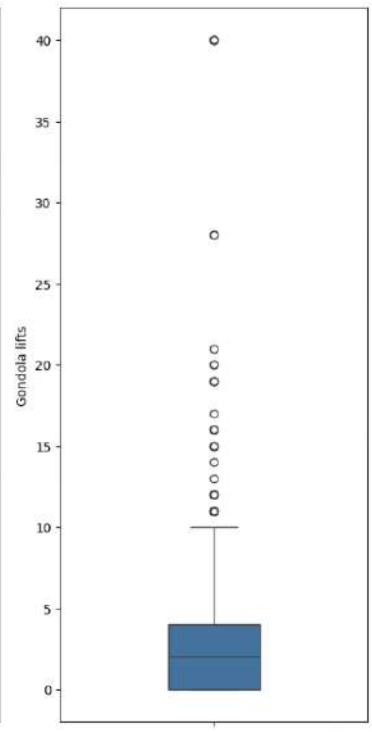
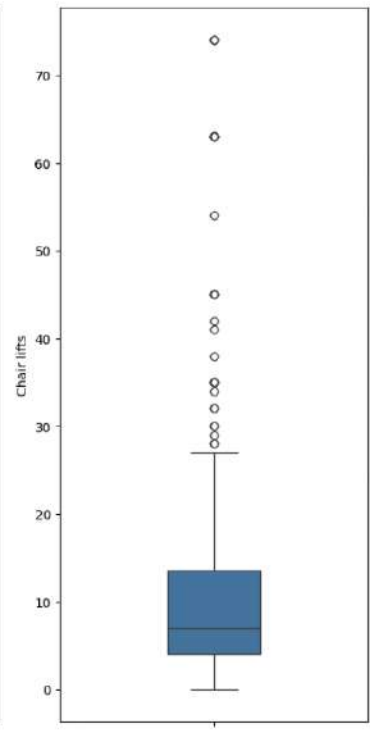
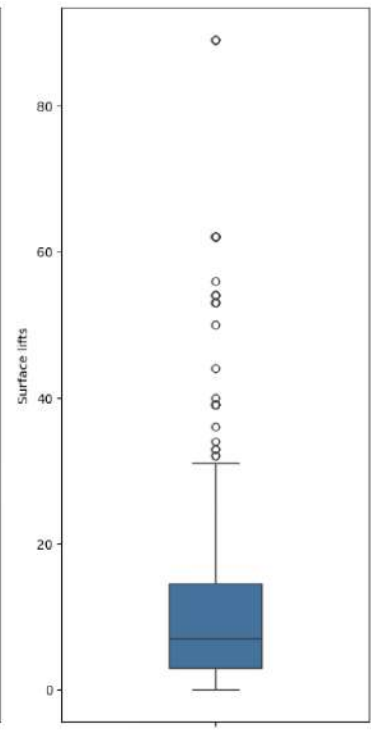
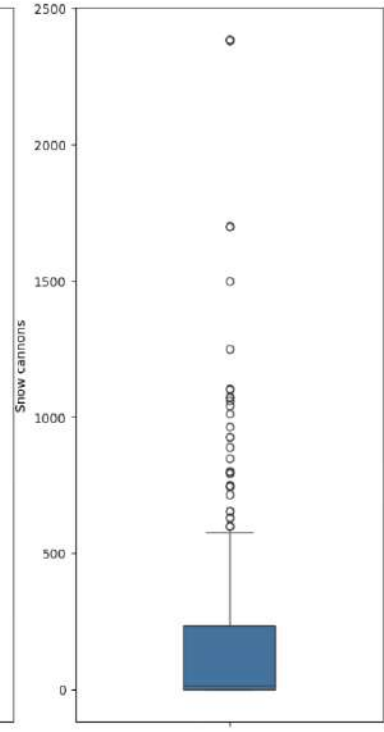
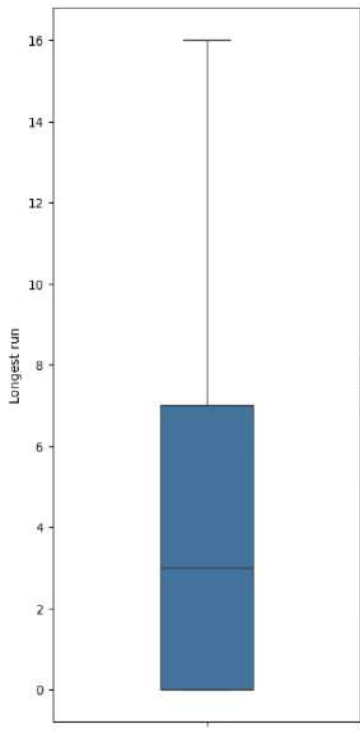
ИССЛЕДУЕТСЯ ВЛИЯНИЕ
РАЗЛИЧНЫХ ФАКТОРОВ НА ЦЕНУ
ГОРНОЛЫЖНЫХ КУОРТОВ С ЦЕЛЮ
ВЫЯВЛЕНИЯ ЗНАЧИМЫХ
ЗАВИСИМОСТЕЙ.

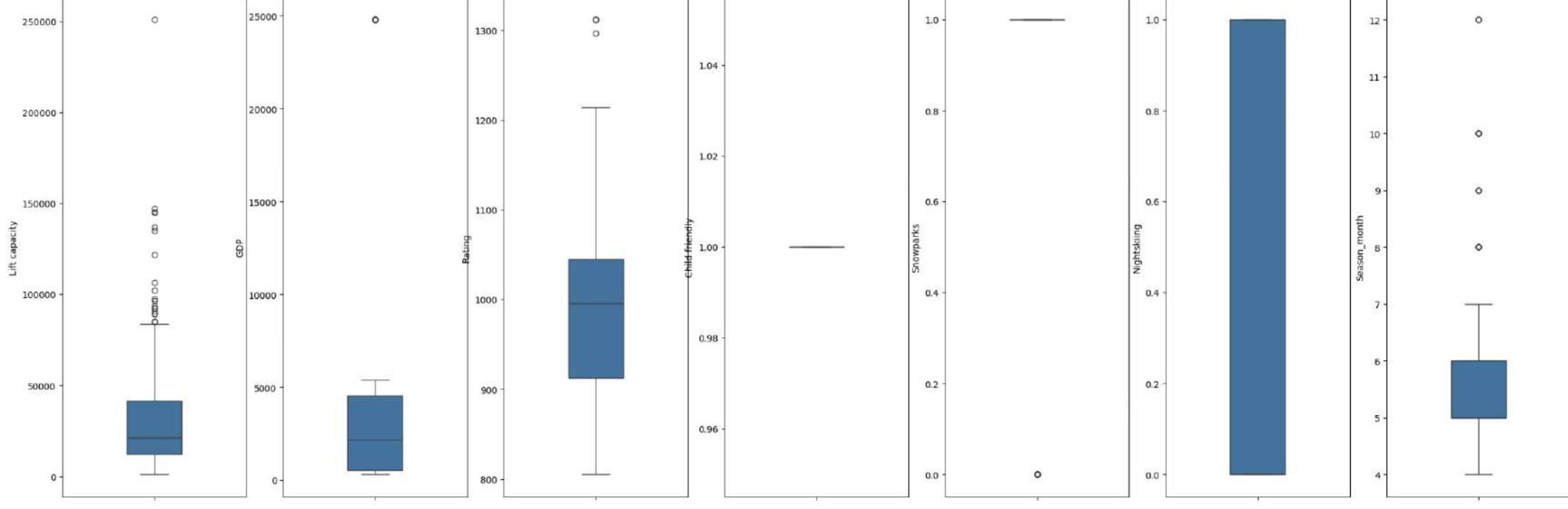


СБОР И ПРЕДОБРАБОТКА ДАННЫХ В ЧИСЛОВЫЕ ЗНАЧЕНИЯ

- стоимость отдыха (в евро)
- высота самой высокой точки курорта (в м)
- длина трасс для новичков (в км)
- длина трасс средней сложности (в км)
- длина сложных трасс (в км)
- общая длина трасс (в км)
- длина самой длинной трассы на курорте
- количество снежных пушек
- количество бугельных подъемников
- количество кресельных подъемников
- количество гондольных подъемников
- общее количество подъемников на курорте
- суммарная пропускная способность подъемников (чел/час)
- ВВП региона
- рейтинг курорта







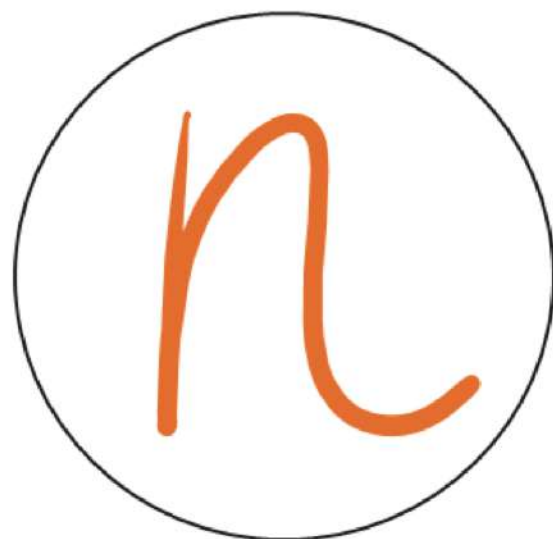
до

499

после

211

предобработка
данных



168

43

объем обучающих данных

объем тестовых данных

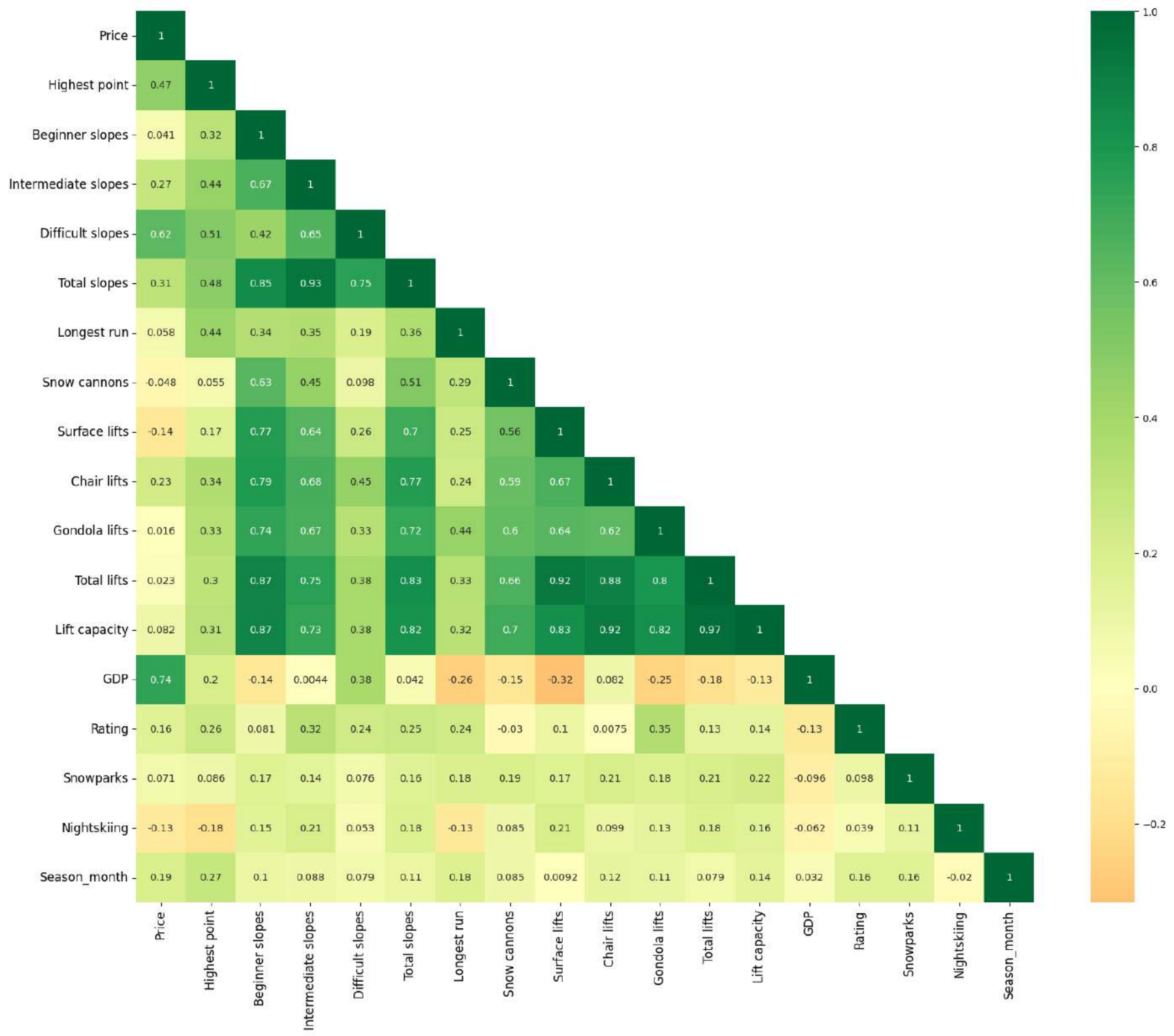
Напишите

МЕТОДЫ ИССЛЕДОВАНИЯ

**СТАТИСТИЧЕСКИЙ МЕТОД ИССЛЕДОВАНИЯ -
МНОГОФАКТОРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ**

РЕАЛИЗАЦИЯ - ЯЗЫК ПРОГРАММИРОВАНИЯ PYTHON

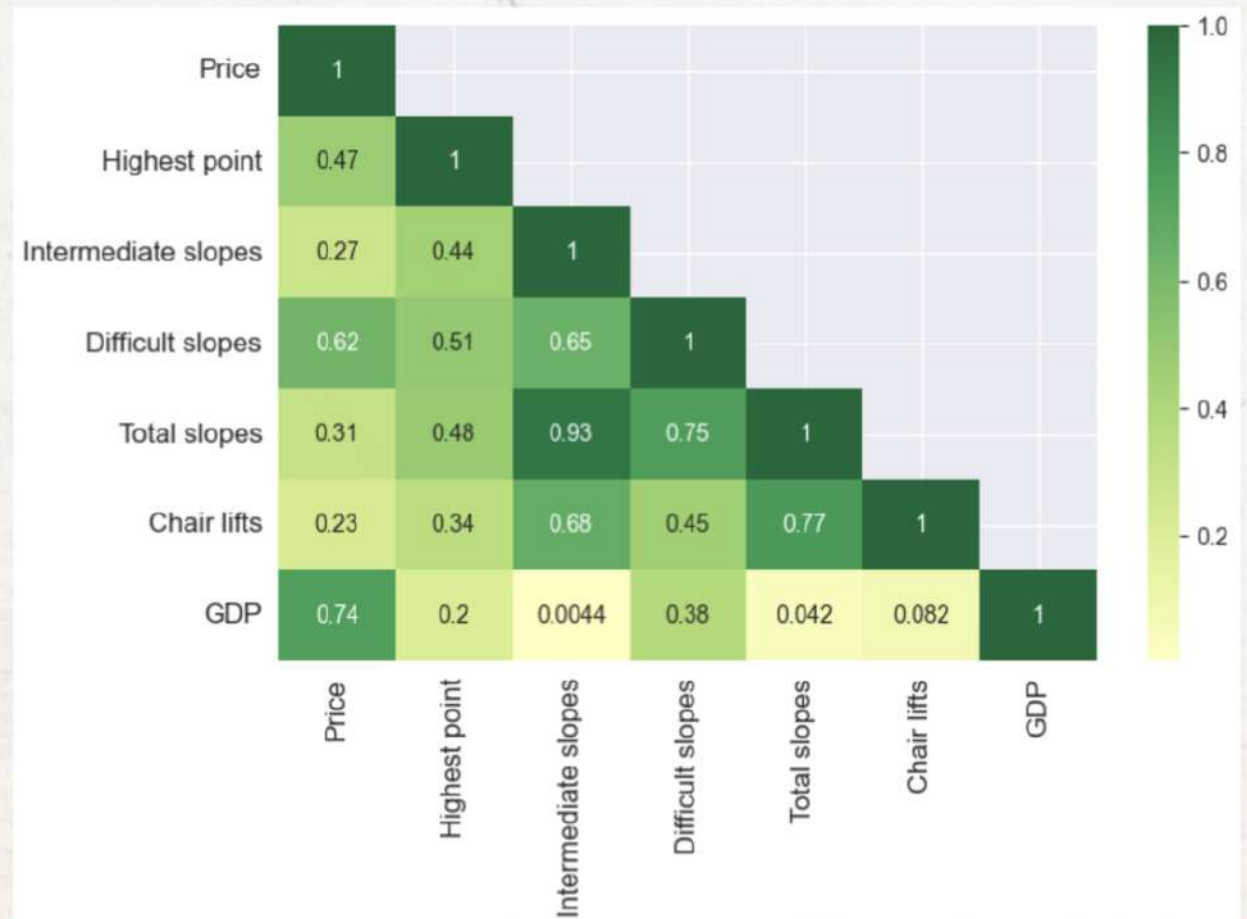
СБОР ДАННЫХ - ДАТАСЕТЫ С ПЛАТФОРМЫ KAGGLE



ОТБОР ФАКТОРОВ

Удаляем факторы, которые не значимы по шкале Чеддока:

Начинающие склоны,
Снежные пушки,
Буксировочные подъемники,
Кресельные подъемники,
Подъемники-гондолы,
Всего подъемников,
Пропускная способность подъемников,
Подходит для детей,
Сноупарки,
Ночные катания

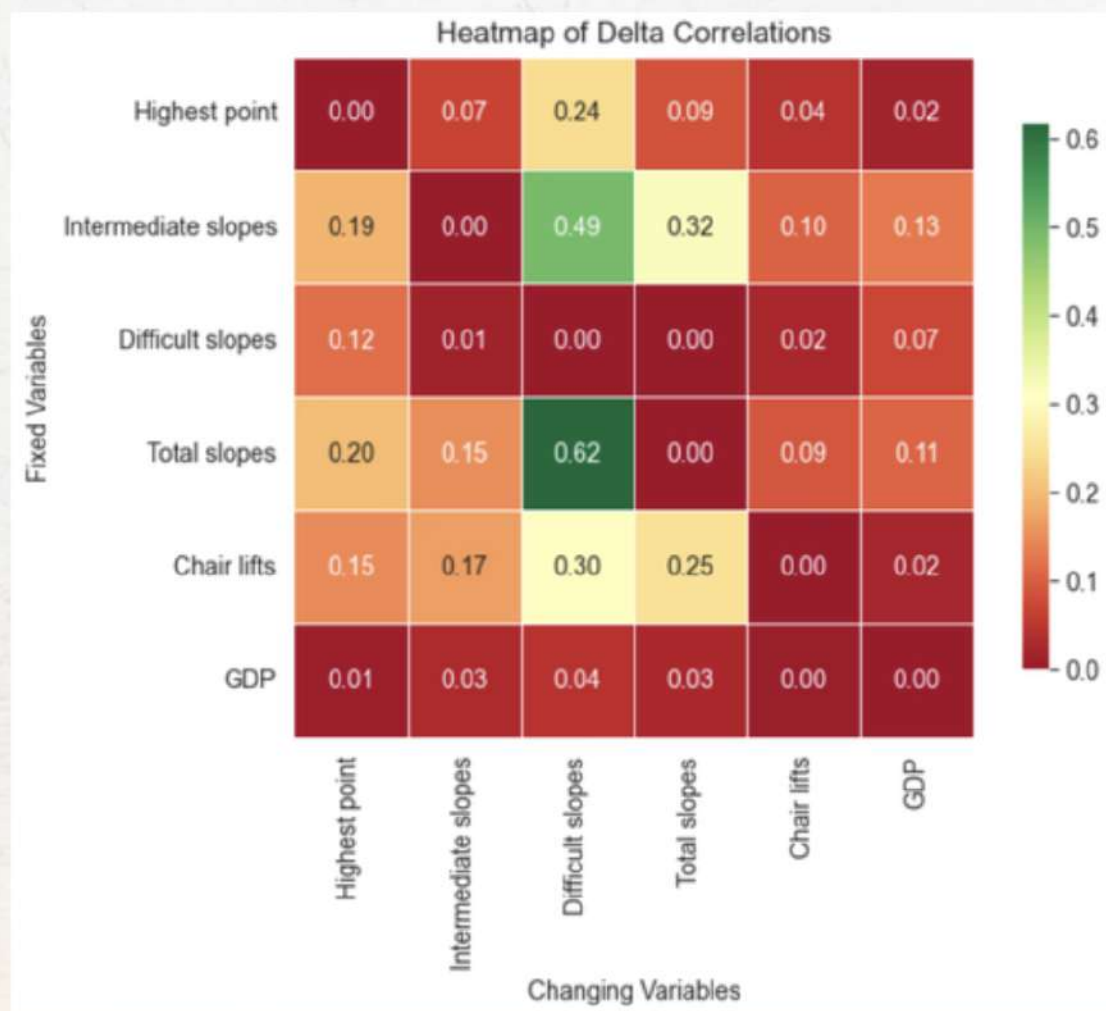


ОТБОР ФАКТОРОВ С ПОМОЩЬЮ ЧАСТНЫХ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ

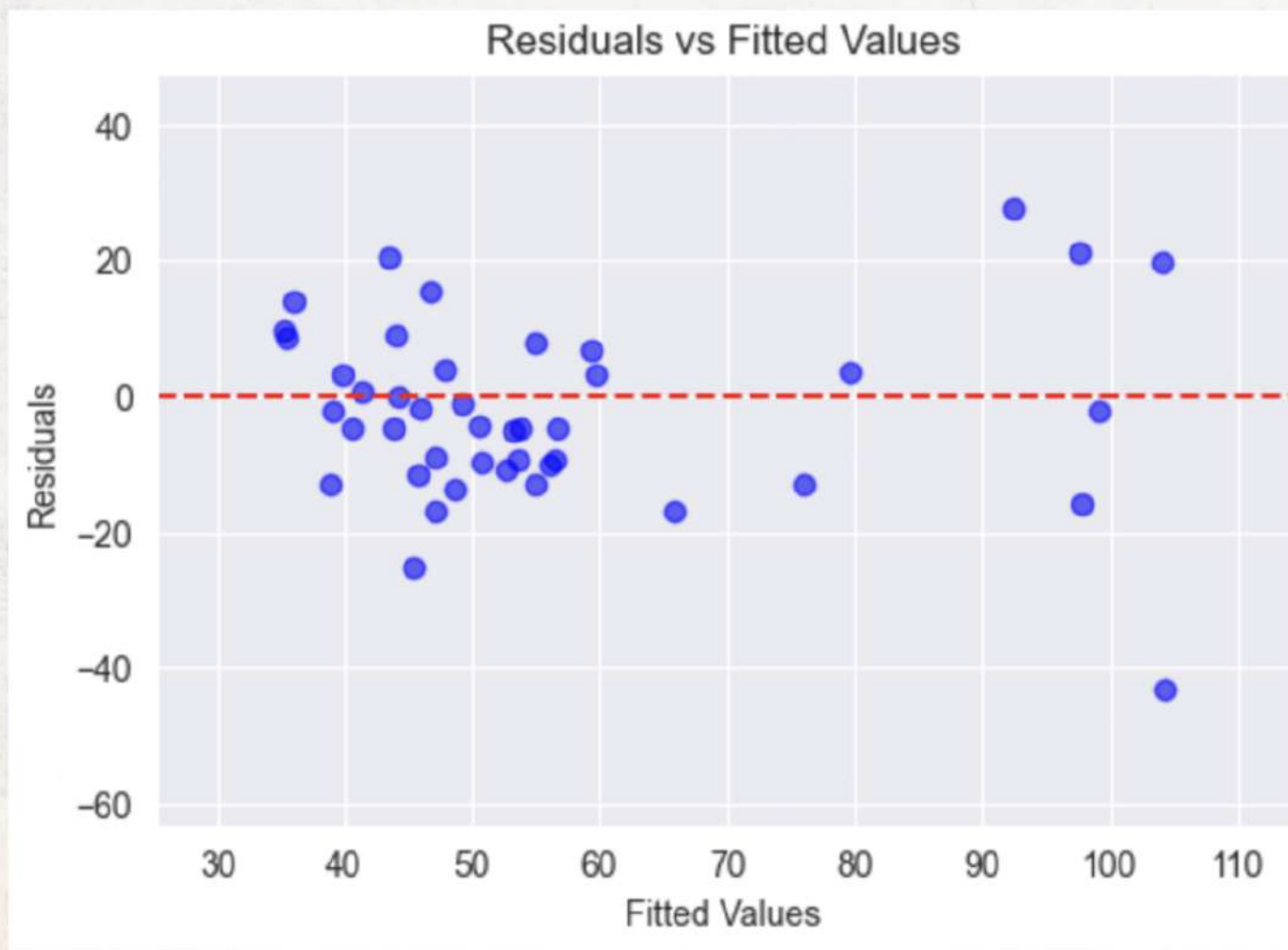
Матрица корреляций для r

Включаем в модель следующие факторы

- Total slopes
- Highest point
- GDP



ОЦЕНКА ОСТАТКОВ УСЛОВИЯ ГАУССА-МАРКОВА



ГИСТОГРАММА ОСТАТКОВ УСЛОВИЯ ГАУССА-МАРКОВА

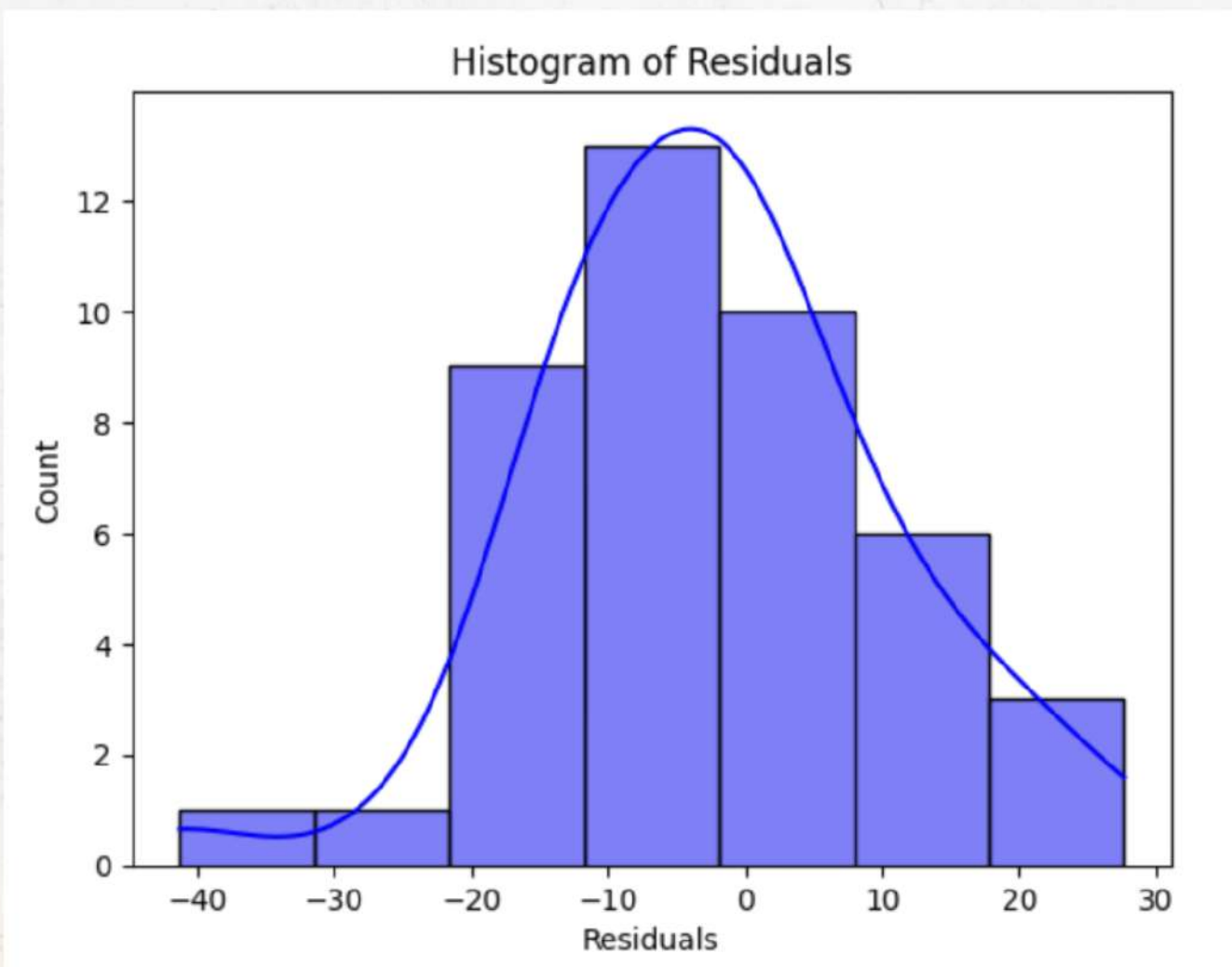
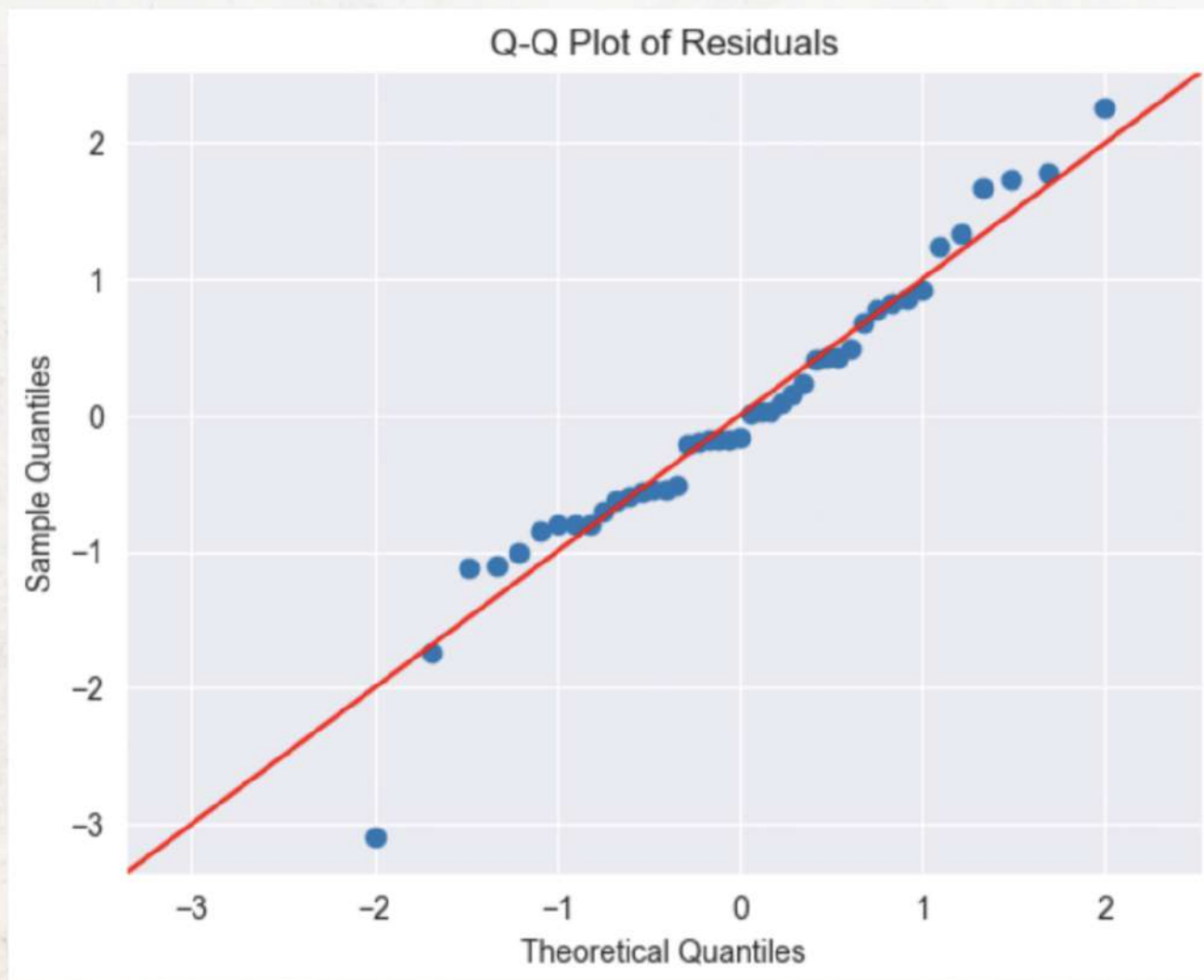


ГРАФИК КВАНТИЛЕЙ (Q-Q) УСЛОВИЯ ГАУССА-МАРКОВА



$$y = 57,27 + 3,35x_1 + 15,98x_2 + 6,53x_3$$

число трасс

ВВП страны
горнолыжного
курорта

наивысшая
точка (пик)

$$R^2 = 0,69$$

Интерпретация результатов:

При увеличении **числа трасс** на 1 - цена ski-pass на один день на взрослого в среднем увеличивается на **3,35 евро** при всех равных прочих условиях

При увеличении **ВВП страны курорта** на условную единицу - цена ski-pass на один день на взрослого в среднем увеличивается на **15,98 евро** при всех равных прочих условиях

При увеличении **высоты пика** на 1 метр - цена ski-pass на один день на взрослого в среднем увеличивается на **6,53 евро** при всех равных прочих условиях

Минимальная цена ski-pass – 57,27 евро на один день на взрослого (на 2022 год это 4333 рубля)

ПРОВЕРКА ЗНАЧИМОСТИ МОДЕЛИ В ЦЕЛОМ. КРИТЕРИЙ ФИШЕРА

F - расчетная: 40.528767680397536

F - критическая: 3.2317269928308443

F - расчетная > F - критическая,
следовательно модель значима
на уровне значимости 0.05



ПРОВЕРКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТОВ. КРИТЕРИЙ СТЬЮДЕНТА

Интерпретация значимости коэффициентов:

Фактор 'Total slopes' не статистически значим
(p-value = 0.12)

Фактор 'GDP' статистически значим
(p-value = 0.000000003)

Фактор 'Highest point' статистически значим
(p-value = 0.002)

ВЫВОД

В результате исследования было выявлено, что модель значима, условия Гаусса - Маркова выполняются.

Таким образом, наиболее значимыми факторами для цены курорта оказались ВВП региона, где находится курорт и наивысшая точка курорта.