

Санкт-Петербургский государственный университет  
Факультет прикладной математики и процессов управления

## Лабораторная работа №1

Выполнил:  
Дрозин Николай Андреевич  
Группа: 22.Б03-пу

## Анализ набора данных "Airport"

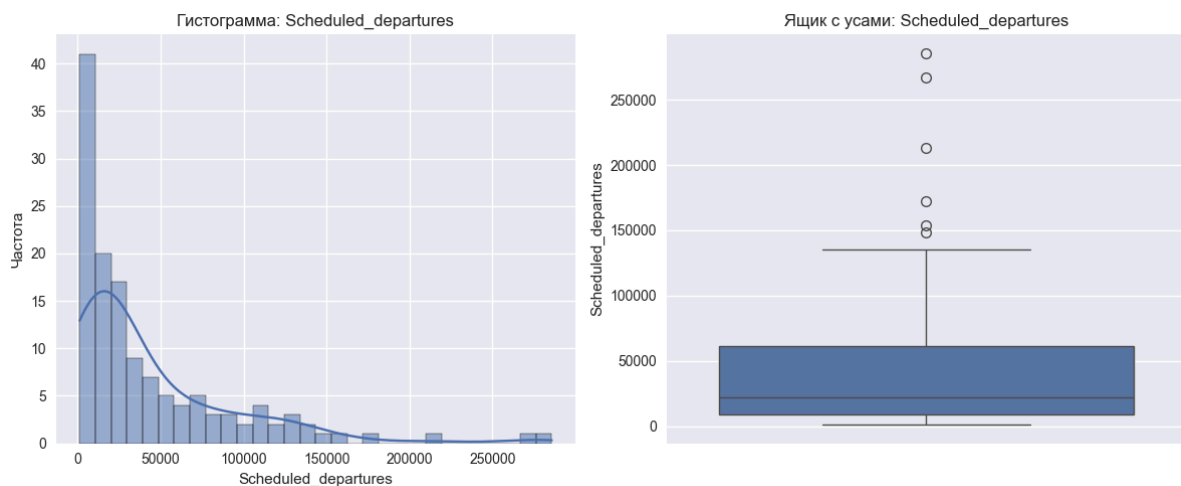
В этом разделе выполнен анализ данных из набора "Airport" по следующим параметрам:

- Запланированные рейсы (Scheduled departures)
- Выполненные рейсы (Performed departures)
- Количество перевезённых пассажиров (Enplaned passengers)
- Объём перевезённых грузов (Enplaned revenue tons of freight)
- Объём перевезённой почты (Enplaned revenue tons of mail)

### Визуализация данных: гистограммы и диаграммы "ящик с усами"

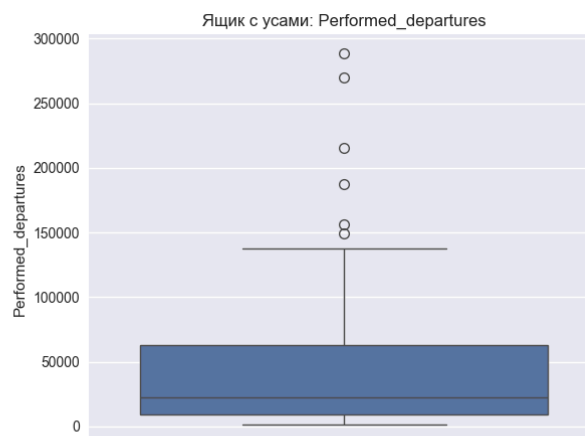
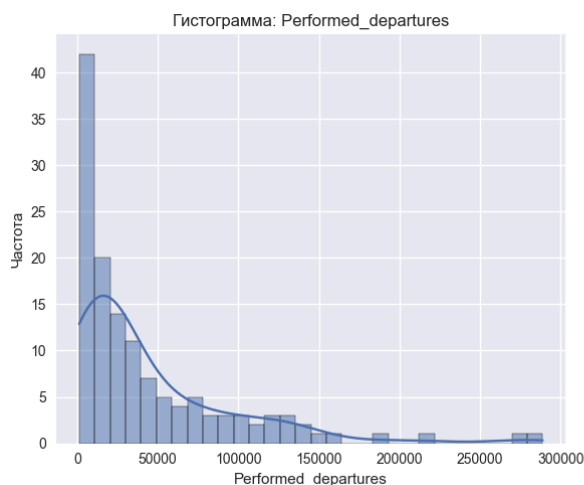
Для каждого параметра были созданы гистограммы и диаграммы "ящик с усами" с использованием кода Python.

- Запланированные рейсы



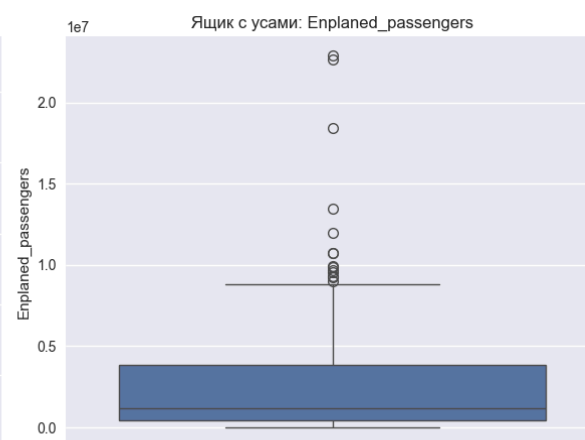
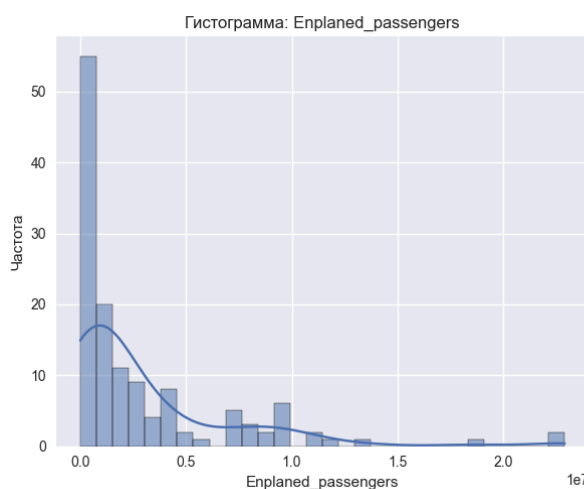
- Гистограмма показывает распределение с явной правой асимметрией. Большинство аэропортов планируют небольшое число рейсов. Пик гистограммы находится в области низких значений.
- Вывод о распределении: Данные не подчиняются нормальному распределению, вероятно, соответствуют экспоненциальному или логнормальному из-за длинного правого "хвоста".
- Диаграмма "ящик с усами": Медиана ближе к нижнему квартилю (Q1). Верхний "ус" длиннее нижнего, а справа наблюдаются выбросы, указывающие на аэропорты с аномально большим числом рейсов. Квартили Q1 и Q3 определяют диапазон, охватывающий 50% значений.

- Выполненные рейсы



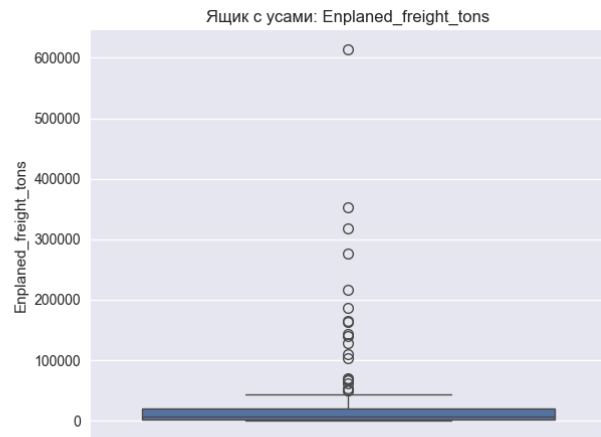
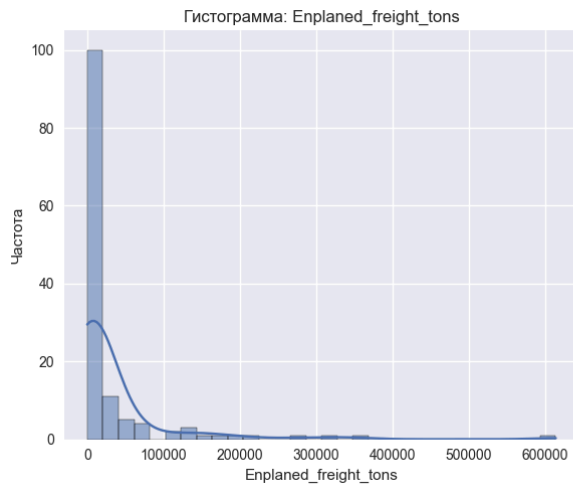
- Гистограмма демонстрирует сильную правую асимметрию, схожую с запланированными рейсами. Большинство аэропортов выполняют мало рейсов.
- Вывод о распределении: Распределение не является нормальным и имеет выраженную правую асимметрию.
- Диаграмма "ящик с усами": Медиана смещена к нижнему квартилю, правый "ус" длинный, а выбросы справа указывают на аэропорты с высоким числом выполненных рейсов.

#### ● Перевезённые пассажиры



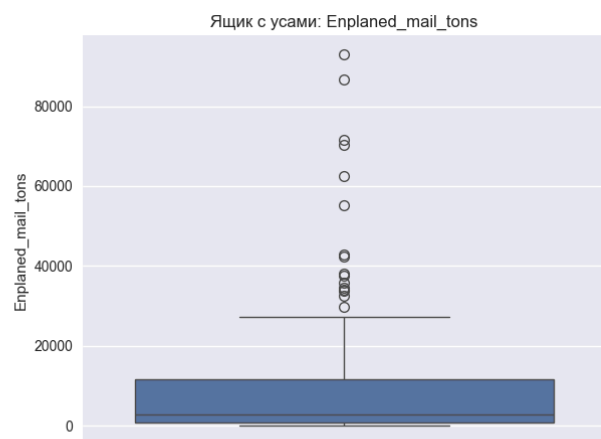
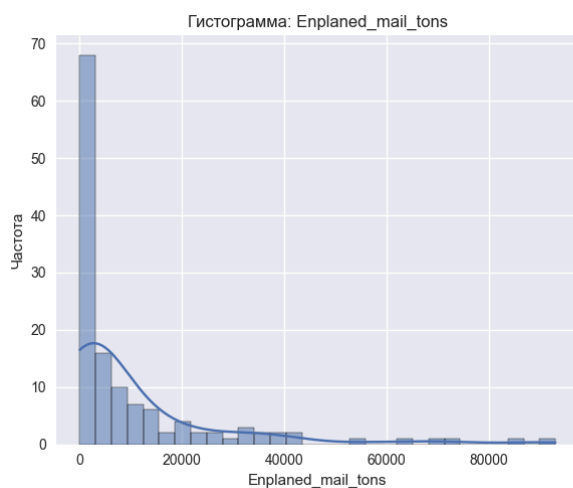
- Гистограмма отражает сильную правую асимметрию, где большинство аэропортов обслуживают небольшое число пассажиров.
- Вывод о распределении: Не нормальное, вероятно, логнормальное или схожее.
- Диаграмма "ящик с усами": Медиана низкая, правый "ус" длинный, а выбросы справа подтверждают асимметрию данных.

#### ● Перевезённые тонны грузов



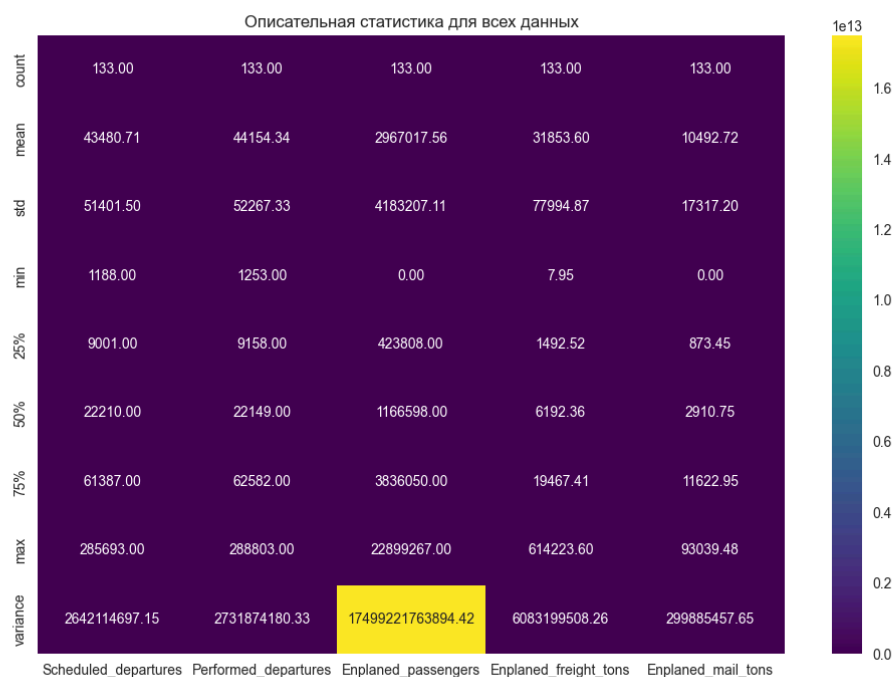
- Гистограмма показывает, что большинство аэропортов обрабатывают небольшие объёмы грузов, с выраженной правой асимметрией.
- Вывод о распределении: Распределение не нормальное, с сильной правой асимметрией.
- Диаграмма "ящик с усами": Медиана близка к Q1, правый "ус" длинный, с множеством выбросов справа.

#### ● Перевезённые тонны почты



- Гистограмма демонстрирует экстремальную правую асимметрию: у большинства аэропортов объёмы почты минимальны, и лишь отдельные аэропорты обрабатывают значительные объёмы.
- Вывод о распределении: Распределение не нормальное, с выраженной правой асимметрией.
- Диаграмма "ящик с усами": Медиана почти совпадает с Q1 и близка к нулю. Правый "ус" очень длинный, а выбросы справа подчёркивают сильную асимметрию.

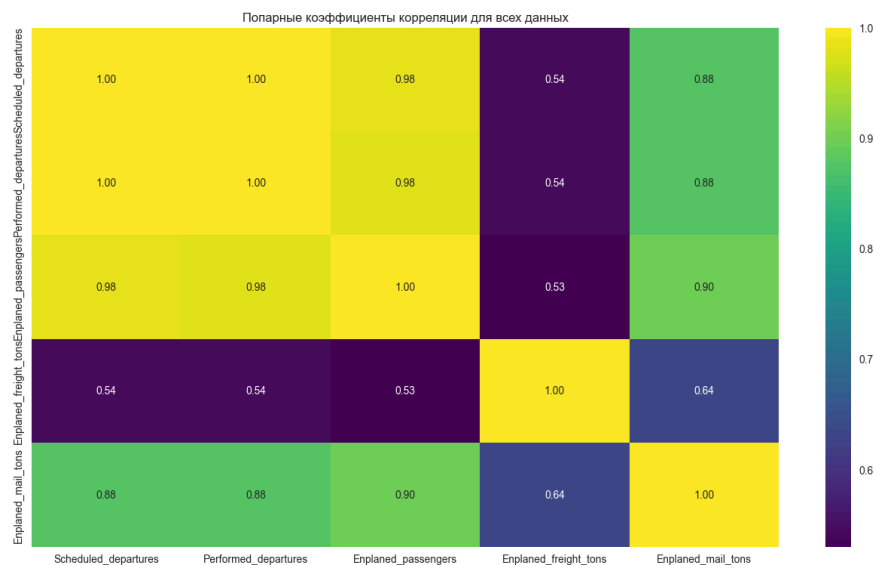
#### Описательная статистика



Числовые характеристики параметров были рассчитаны с использованием Python и представлены в таблице.

- Интерпретация: Средние значения всех параметров значительно превышают медианы, что подтверждает правую асимметрию распределений. Высокие значения стандартного отклонения и дисперсии указывают на большой разброс данных.

## Корреляционный анализ



Для изучения взаимосвязей между параметрами была рассчитана матрица парных коэффициентов корреляции Пирсона.

### Выводы:

- Запланированные и выполненные рейсы имеют очень высокую положительную корреляцию, так как эти параметры тесно связаны.
- Запланированные и выполненные рейсы имеют почти идеальную корреляцию, что указывает на высокую надёжность планирования в большинстве аэропортов.
- Количество рейсов (запланированных и выполненных) сильно коррелирует с числом перевезённых пассажиров, что указывает на зависимость пассажиропотока от числа рейсов.
- Корреляция между числом рейсов и объёмами грузов или почты положительная, но менее выраженная, чем с пассажирами.
- Пассажиропоток тесно связан с числом рейсов, но объёмы грузов и почты показывают меньшую корреляцию, что отражает специализацию некоторых аэропортов на грузовых перевозках.

---

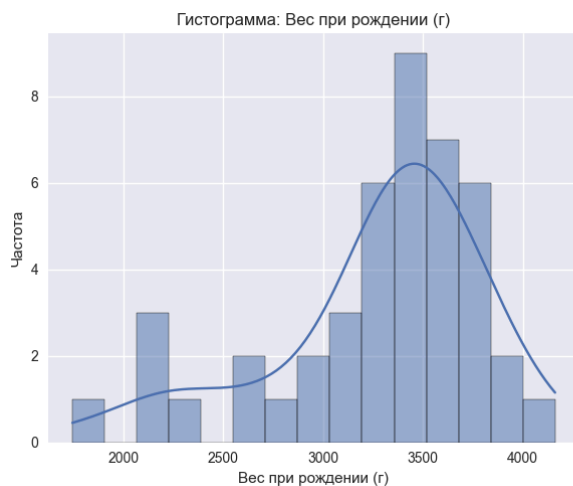
### Анализ набора данных "babyboom"

В этом разделе анализируется набор данных "babyboom" по следующим параметрам:

- Вес новорождённых (в граммах)
- Время рождения (в минутах после полуночи)

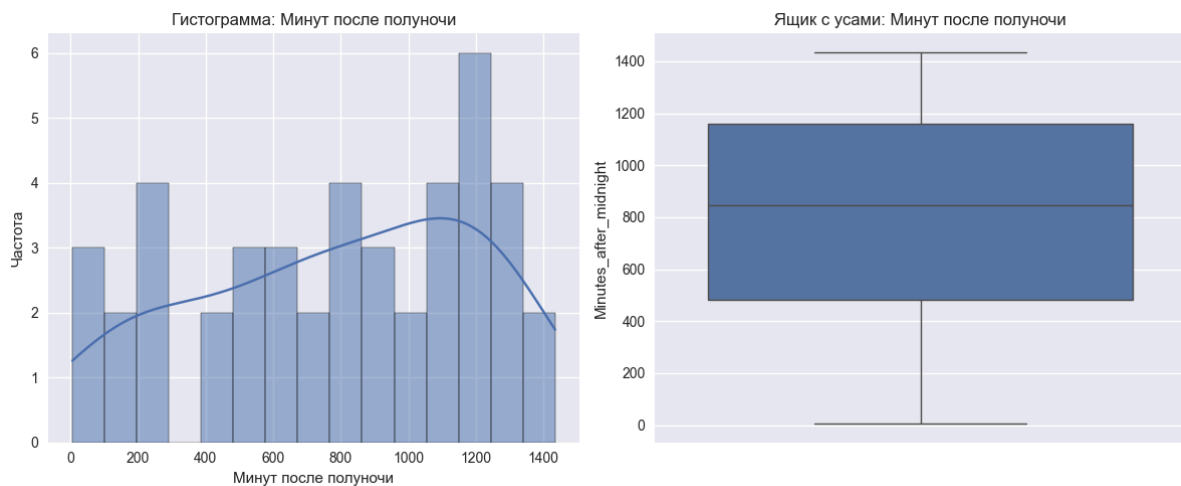
### Визуализация данных: гистограммы и диаграммы "ящик с усами"

#### ● Вес новорождённых



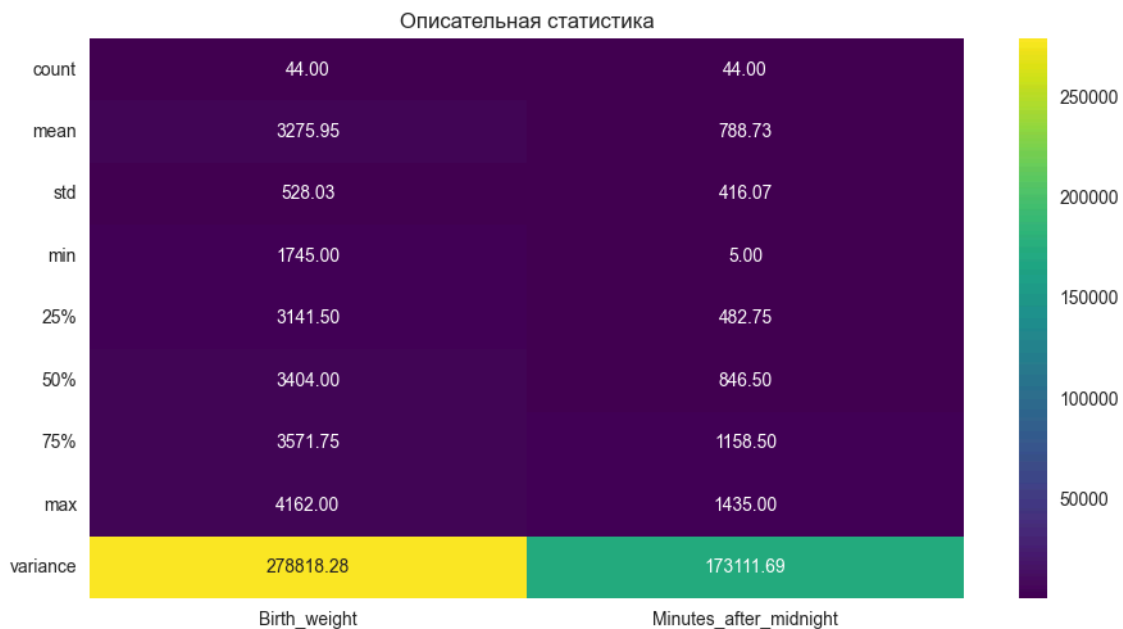
- Гистограмма: Распределение веса новорождённых близко к симметричному, с возможной лёгкой левой асимметрией из-за детей с низким весом. Значения сосредоточены вокруг центрального пика.
- Вывод о распределении: Распределение похоже на нормальное, но для подтверждения нужен формальный тест.
- Диаграмма "ящик с усами": Медиана находится примерно посередине между Q1 и Q3. Усы схожей длины, но могут быть выбросы, особенно слева, связанные с низким весом.

- **Время рождения**



- Гистограмма: Распределение времени рождения выглядит почти равномерным, без чётких пиков, что указывает на случайное распределение рождений в течение суток. Возможны небольшие колебания.
- Вывод о распределении: Распределение близко к равномерному на интервале  $[0, 1440)$ .
- Диаграмма "ящик с усами": Ящик охватывает большую часть диапазона  $[0, 1440]$ , с медианой около 720 минут (середина суток). Усы простираются почти до границ диапазона, а выбросы маловероятны при корректных данных.

## Описательная статистика



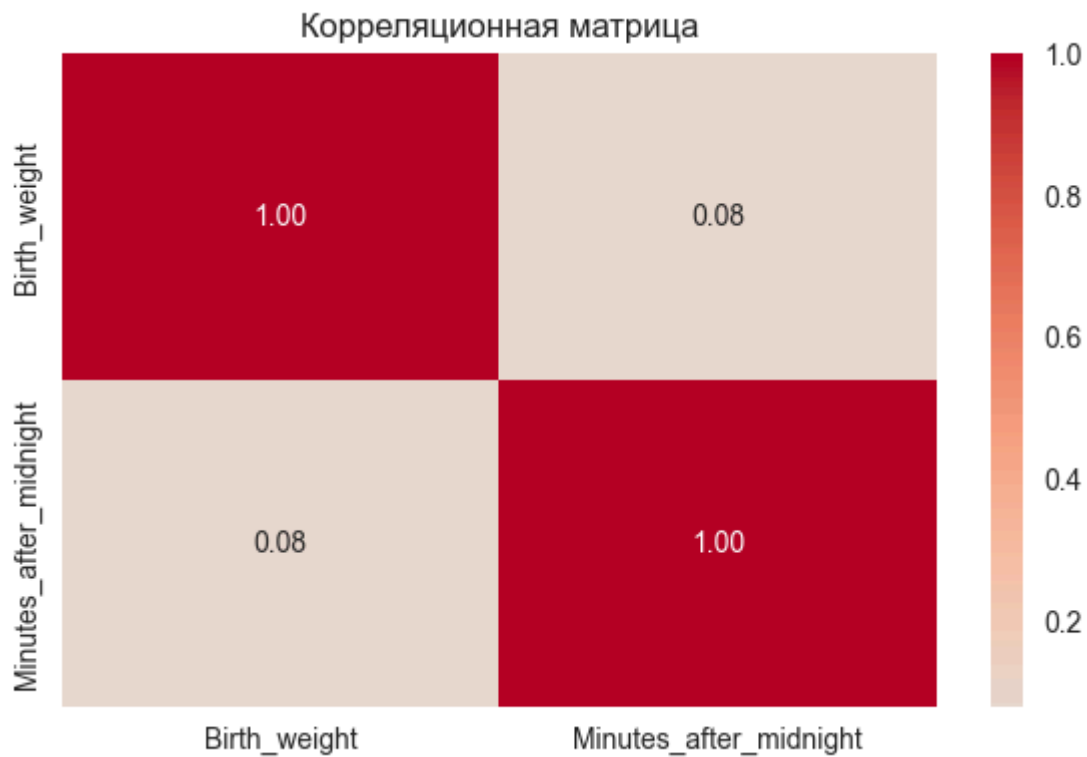
Числовые характеристики параметров набора "babyboom" были рассчитаны.

Интерпретация:

- Для веса новорождённых среднее и медиана близки, что соответствует почти симметричному распределению. Стандартное отклонение показывает типичный разброс.

- Для времени рождения среднее и медиана близки к 720 минутам (середина суток).  
Высокое стандартное отклонение подтверждает равномерность распределения.

#### Корреляционный анализ



Был рассчитан коэффициент корреляции Пирсона между весом новорождённых и временем рождения.

Выводы:

- Коэффициент корреляции, вероятно, близок к нулю, что указывает на отсутствие линейной связи между весом ребёнка и временем его рождения.