Вывод:

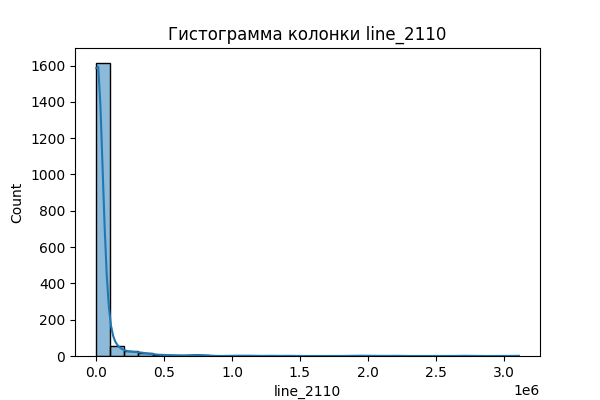
**Колонка line\_2110 (выручка компании) имеет сильно правостороннее распределение и значительное количество выбросов.**

* **Описательная статистика:**
  + Среднее значение (mean) значительно выше 50%-перцентиль, что указывает на наличие больших выбросов.
  + Максимальные значения сильно превышают 75-й перцентиль, подтверждая наличие крупных выбросов.

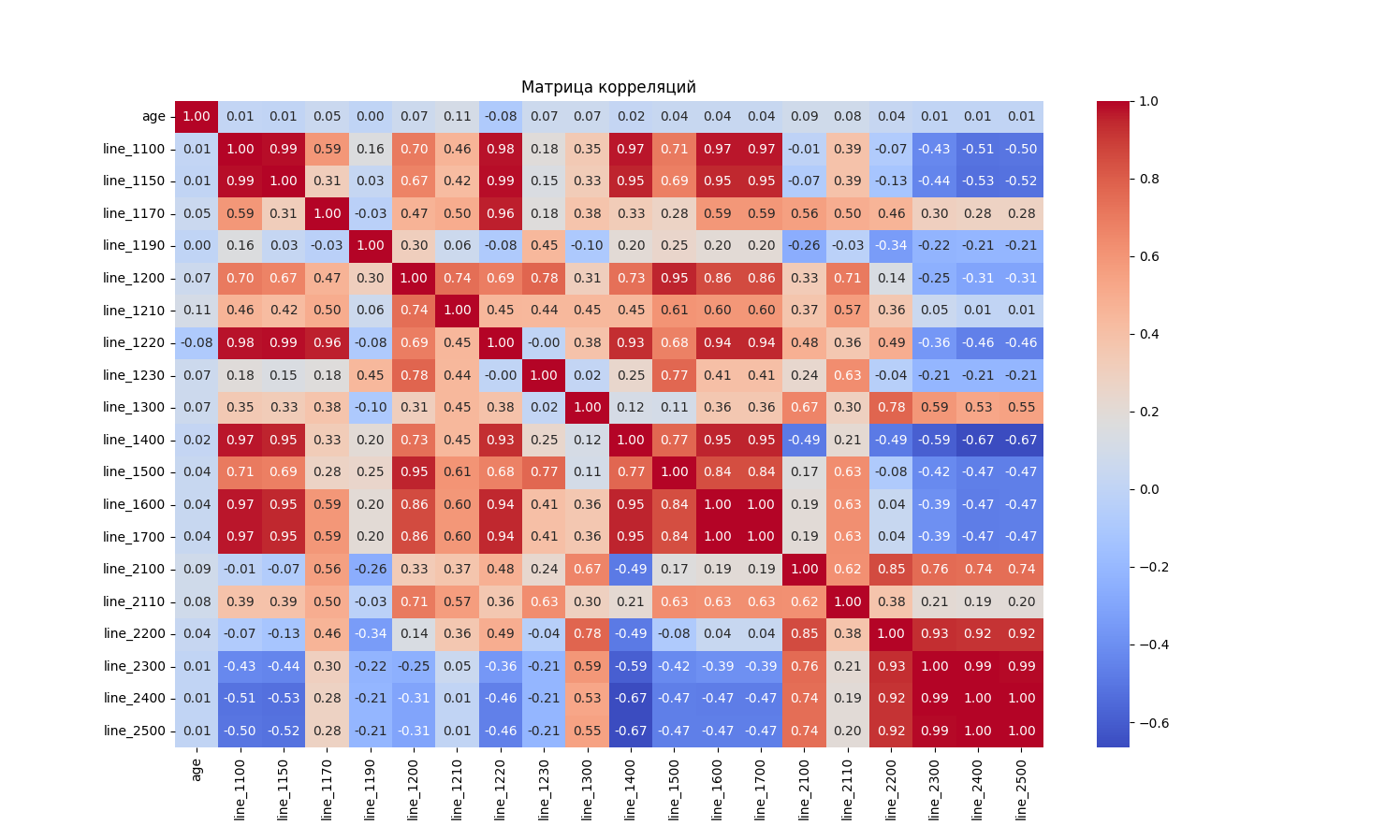
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Номер реализации** | **Mean** | **Перцентиль 50%** | **Перцентиль 75%** | **max** |
| 1. Потоковая выгрузка данных с выборкой 10К строк из выборки в 100К строк | **133 218** | **8 086** | **33 801** | **19 649 210** |
| 2. Потоковая выгрузка данных с выборкой первых 10К строк | **40 128** | **4 680** | **17 754** | **3 111 601** |
| 3. Анализ всех данных за один год. Выборка = Генеральная совокупность = +/-292К строк | **121 564** | **7 499** | **28 768** | **828 040 794** |

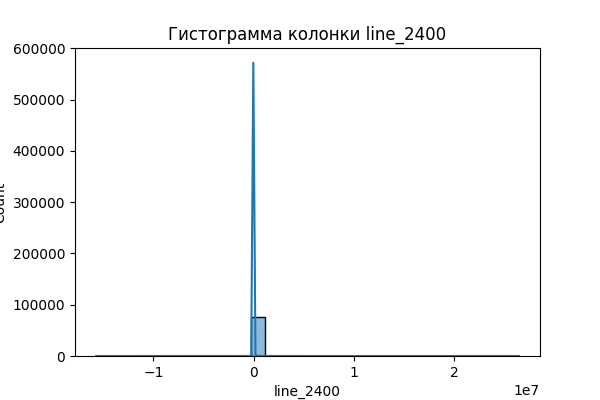
* **Гистограмма:**
  + Распределение **line\_2110** смещено вправо (положительная асимметрия).
  + Видно, что большинство компаний имеют сравнительно небольшую выручку, а есть редкие очень крупные значения.

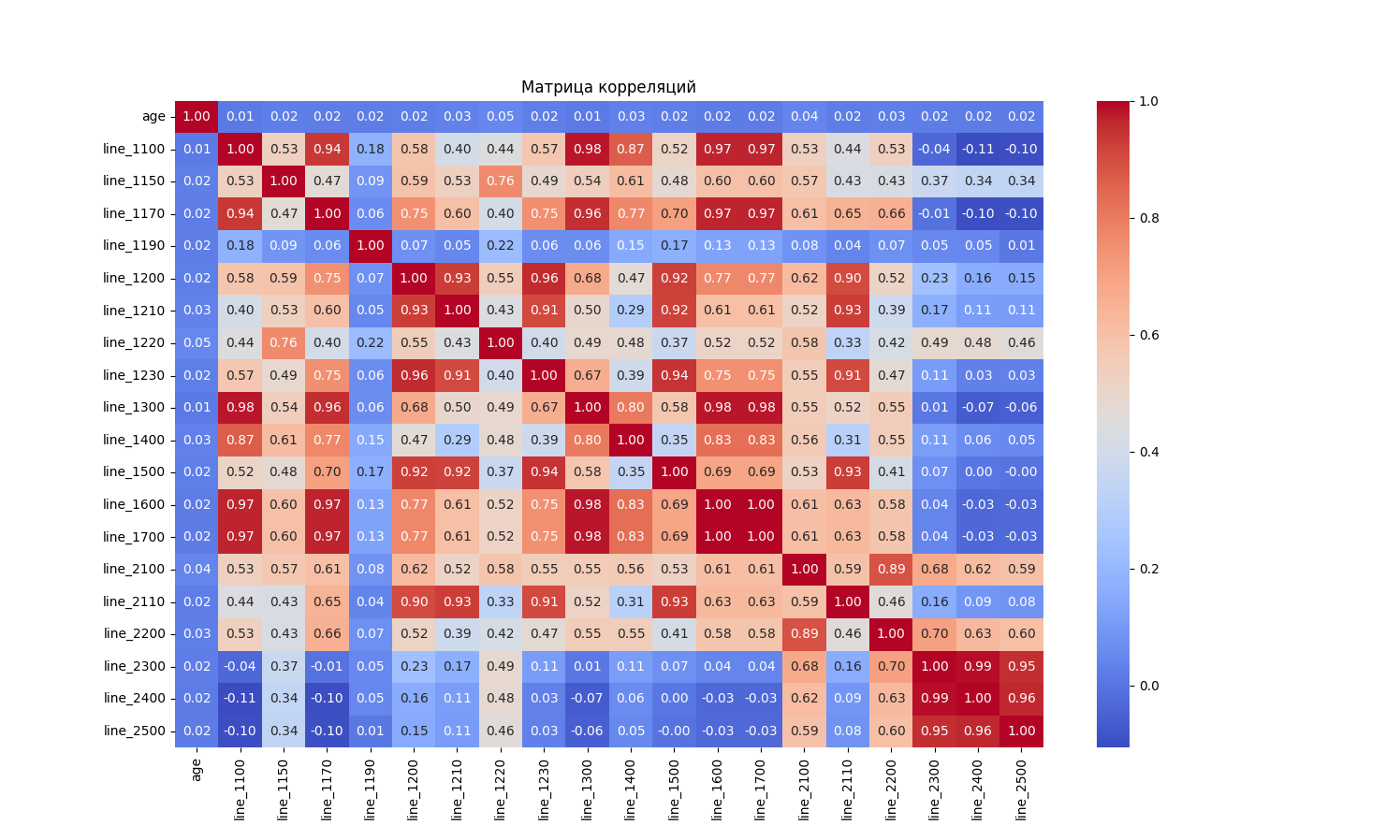
****

****

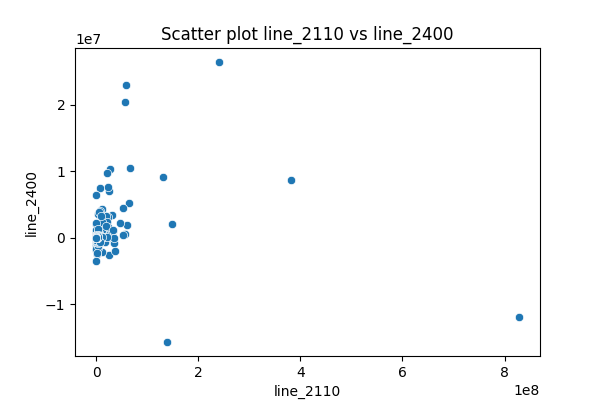
* **Корреляция:**
  + line\_2110 выручка не коррелирует с line\_2400 - (**Отчёте о финансовых результатах» в бухгалтерии обозначает «Чистая прибыль (убыток)**), что видно по корреляционной матрице (корреляция < 0.2).
    - У некоторых компаний высокая выручка, но одновременно большие расходы → низкая чистая прибыль.
    - Напротив, небольшая компания с низкой выручкой может иметь низкие расходы и, соответственно, высокую маржу.
    - Таким образом, выручка не является прямым индикатором чистой прибыли, особенно в разнотипных компаниях.

****

****

****

* + Scatter plot этих колонок подтверждает данные сильно скошены и содержат выбросы.



* **Выбросы:**
  + При проверке методом IQR (1.5\*IQR) выявлено большое количество выбросов, что особенно важно учитывать при дальнейшей статистической обработке или построении моделей.