

Тест начат	Понедельник, 4 декабря 2023, 17:28
Состояние	Завершены
Завершен	Понедельник, 4 декабря 2023, 17:39
Прошло времени	11 мин. 21 сек.
Оценка	Еще не оценено

Вопрос **1**

Верно

Баллов: 1,00 из 1,00

## Построение матрицы корреляции для анализа продаж супермаркетах.

В этом задании предлагаем вам поработать с датасетом [Supermarket sales](#), который содержит д продаж в трех разных супермаркетах. В LMS файл с данными `supermarket.csv` уже подгруже

Ваша задача — рассчитать матрицу корреляции и сохранить ее в переменную `corr`. Округлите з коэффициентов корреляции до двух знаков после запятой.

Обратите внимание:

- в задании **НЕ требуется** выполнять вывод полученных результатов на экран;
- рекомендуем выполнять вам задание в самостоятельно созданном Python-ноутбуке, а в L вносить требуемый код программы — так вам будет удобнее выполнять следующие зада
- для работы скачайте файл с [сайта Kaggle](#) или [по ссылке](#).

Ответ: (штрафной режим: 0%)

Сбросить ответ

```
1 import pandas as pd
2
3 # далее запишите ваш код
4 df = pd.read_csv('supermarket.csv', sep=',')
5 corr = df.corr().round(2)
```

	Тест	Ожидается	Получил
✓	print(corr)	Unit price Quantity ... gross income Rating Unit price 1.00 0.01 ... 0.63 -0.01 Quantity 0.01 1.00 ... 0.71 -0.02 Tax 5% 0.63 0.71 ... 1.00 -0.04 Total 0.63 0.71 ... 1.00 -0.04 cogs 0.63 0.71 ... 1.00 -0.04 gross margin percentage NaN NaN ... NaN NaN gross income 0.63 0.71 ... 1.00 -0.04 Rating -0.01 -0.02 ... -0.04 1.00  [8 rows x 8 columns]	Unit price Quantity ... gross inc Rating Unit price 1.00 0.01 ... 0 -0.01 Quantity 0.01 1.00 ... 0 -0.02 Tax 5% 0.63 0.71 ... 1 -0.04 Total 0.63 0.71 ... 1 -0.04 cogs 0.63 0.71 ... 1 -0.04 gross margin percentage NaN NaN ... NaN NaN gross income 0.63 0.71 ... 1 -0.04 Rating -0.01 -0.02 ... -( 1.00  [8 rows x 8 columns]

Все тесты пройдены! ✓

Спасибо за ваше решение! Вы получили нужный результат!

Предлагаем свериться с решением.

```
df = pd.read_csv('supermarket.csv', sep=',')
corr = df.corr().round(2)
```

Верно

Вопрос **2**  
Верно  
Баллов: 1,00 из 1,00

## Построение матрицы корреляции для анализа продаж супермаркетах.

Теперь давайте проанализируем получившиеся коэффициенты корреляции.

Ответьте на вопросы приведенные ниже.

1. Какие поля датафрейма имеют сильную положительную корреляцию со стоимостью проданн (cogs)?

Выберите один или несколько ответов:

- ☐ Invoice ID
- ☒ gross income ✓
- ☒ Tax 5% ✓
- ☐ Unit price
- ☐ Quantity
- ☐ Rating
- ☐ gross margin percentage
- ☒ Total ✓
- ☐ Branch

Ваш ответ верный.

Вопрос **3**  
Верно  
Баллов: 1,00 из 1,00

2. Какие параметры из приведенных в списке ниже имеют наиболее сильный отрицательный ко: корреляции?

Выберите один ответ:

- ☐ “Rating” и “Unit price”
- ☐ “Unit price” и “Quantity”
- ☐ “Rating” и “gross margin percentage”
- ☒ “Rating” и “Total” ✓

Ваш ответ верный.

Вопрос **4**  
Выполнен  
Балл: 1,00

## Анализ матрицы корреляции произвольных данных.

*Это задание не проверяется LMS.*

Но выполняя его, вы тренируете и закрепляете навыки расчета и анализа матрицы корреляции различных исходных данных.

Для этого задания вы можете использовать любой доступный вам набор данных, например, загрузив заинтересовавший вас набор данных с сайта [kaggle](https://www.kaggle.com/).

Что нужно сделать:

- создайте Python-ноутбук, загрузите в него данные из csv-файла и преобразуйте в объект
- изучите структуру данных и выполните их предобработку;
- рассчитайте матрицу корреляции;
- визуализируйте результат с применением различных способов.

По каждому пункту задания сделайте выводы и отразите их в своем Python-ноутбуке. В качестве задания прикрепите ссылку на ноутбук.

<https://colab.research.google.com/drive/1W3fctqjWexXGCw5UgCqKV19neQ91S6kw?usp=sharing>

Спасибо! Уверены, что вы хорошо постарались и поработали!

Итак, подведем итог. Матрица корреляции — полезный инструмент для анализа взаимосвязи между несколькими переменными. Она помогает определить, насколько сильно связаны между собой переменные, и какие из них имеют наибольшее влияние на другие. Матрица корреляции может быть полезна в различных областях, включая финансы, экономику, медицину и многие другие. Например, она может использоваться для анализа связи между доходом и расходами, между заболеваемостью и факторами риска, между ценами на товары и спросом на них и т.д.

**Если у вас возникли вопросы по заданию, пожалуйста, обратитесь к преподавателю на ближайшем семинаре.**

Желаем успехов в обучении!