

Тест начат	Суббота, 21 октября 2023, 09:28
Состояние	Завершены
Завершен	Суббота, 21 октября 2023, 16:57
Прошло времени	7 час. 28 мин.
Оценка	Еще не оценено

Вопрос 1

Верно

Баллов: 1,00 из 1,00

Расчет статистических показателей подписчиков социальной сети

Вам даны показатели прироста подписчиков за месяц в 300 аккаунтах социальной сети, которые сохранены в объект DataFrame с именем list_metrics:

Прирост	
0	-34
1	778
2	888
3	-70
4	322
...	...
295	1
296	261
297	104
298	638
299	280

300 rows x 1 columns

Выполните следующие задания применительно к исходным данным:

1. Рассчитайте стандартное отклонение прироста подписчиков. Результат округлите до второго знака после запятой и сохраните в переменную result1.
2. Рассчитайте размах прироста подписчиков. Результат округлите до второго знака после запятой и сохраните в переменную result2.

Обратите внимание: в заданиях **не требуется** выполнять вывод полученных результатов на экран.

* **Дополнительное задание** (не проверяется LMS, но предлагаем проделать его самостоятельно). Постройте гистограмму частот по значению прироста.

(Подсказка: для построения гистограммы используйте [pandas.DataFrame.hist](#))

Ответ: (штрафной режим: 0%)

Сбросить ответ

```
1 # начальный код
2
3 import random
4 import pandas as pd
5
6 random.seed(10)
7
8 list_metrics = []
9
10 for i in range(0,300):
11     n = random.randint(-100,1000)
12     list_metrics.append(n)
13
14 list_metrics = pd.DataFrame({'Прирост': list_metrics})
15 list_metrics['Прирост'] = list_metrics['Прирост'].astype('float')
16
17
18 # далее будет код шаблона
19 result1 = round(list_metrics['Прирост'].std(), 2)
20 result2 = round(list_metrics['Прирост'].max() - list_metrics['Прирост'].min(),2)
21
```

	Тест	Ожидается	Получил	
✓	print("Задание 2.1") print(result1)	Задание 2.1 316.26	Задание 2.1 316.26	✓
✓	print("Задание 2.2") print(result2)	Задание 2.2 1094.0	Задание 2.2 1094.0	✓

Все тесты пройдены! ✓

Спасибо за ваше решение! Вы получили нужный результат!

Предлагаем свериться с решением.

```
# Задание 1
result1 = round(float(list_metrics.std()),2)

# Задание 2
result2 = round(float(list_metrics.max()- list_metrics.min()),2)

# Задание 3 – LMS это задание не проверяет
hist = list_metrics['Прирост'].hist()
figure = hist.get_figure()
```

Верно

Баллы за эту попытку: 1,00/1,00.

Вопрос **2**

Верно

Баллов: 1,00 из 1,00

Исследование датасета

В этом задании предлагаем вам поработать с датасетом [Pandas practices](#) , который содержит данные об автомобилях, остановленных полицией.

Исследование датасета будет состоять из нескольких частей. Чтобы проверить, как вы справились с заданием, в некоторых заданиях мы будем просить вас записать получившийся результат, в других — попросим записать код.

Рекомендуем вам [скачать датасет](#) , самостоятельно создать Python-ноутбук для его выполнения, а в LMS прикладывать лишь требуемые результаты.

Итак, для начала предлагаем проверить набор данных на наличие пропусков.

Определите название колонки, в которой больше всего пропусков, и запишите ее название в поле ответов.

Ответ: county_name



- Чтобы начать работу с датасетом, скачайте его и подгрузите в Python-ноутбук. Для этого используйте библиотеку Pandas.
- Метод `isnull()` в Pandas используется для определения пропущенных значений (NaN) в датафрейме или серии. В результате вы получите набор данных, содержащий значения `True` (значение пропущено) и `False` (значение не пропущено). Далее вам останется подсчитать количество пропущенных значений.
- Чтобы было удобнее работать с результатом, можно выполнить сортировку — функция `sort_values()`.

Вопрос **3**

Верно

Баллов: 1,00 из 1,00

Исследование датасета (продолжение)

Мы продолжаем работать с датасетом [Pandas practices](#) . В этом задании вам необходимо написать программу, которая удалит столбец с наибольшим количеством пропусков.

Подсказка: используйте результаты ранее выполненного задания, выводить на экран ничего не требуется.

Ответ: (штрафной режим: 0%)

Сбросить ответ

```
1 import pandas as pd
2 police = pd.read_csv('police.csv', sep=',')
3
4 # далее запишите ваш код
5 b= police.isnull().sum().sort_values(ascending = False).keys()[0]
6 police=police.drop(b, axis=1)
```

	Тест	Ожидается	Получил	
✓	print (police.columns.values.tolist ())	['stop_date', 'stop_time', 'driver_gender', 'driver_age_raw', 'driver_age', 'driver_race', 'violation_raw', 'violation', 'search_conducted', 'search_type', 'stop_outcome', 'is_arrested', 'stop_duration', 'drugs_related_stop']	['stop_date', 'stop_time', 'driver_gender', 'driver_age_raw', 'driver_age', 'driver_race', 'violation_raw', 'violation', 'search_conducted', 'search_type', 'stop_outcome', 'is_arrested', 'stop_duration', 'drugs_related_stop']	✓

Все тесты пройдены! ✓

Спасибо за ваше решение! Вы получили нужный результат!

Предлагаем свериться с решением.

```
import pandas as pd
police =
pd.read_csv('police.csv', sep=',')
police.drop(['county_name'],
axis = 1 , inplace = True)
```

Обратите внимание, имя столбца county_name мы взяли из результатов ранее выполненного задания.

Верно

Баллы за эту попытку: 1,00/1,00.

Вопрос **4**

Верно

Баллов: 1,00 из 1,00

Исследование датасета (продолжение)

Мы продолжаем работать с датасетом [Pandas practices](#) . Исследуйте данные и ответьте на вопросы:

1. Водителей какого пола чаще останавливают?

Выберите один ответ:

- ☐ женщин
- ☒ мужчин ✓
- ☐ невозможно определить

Спасибо за ваше решение! Вы получили нужный результат!

Чтобы получить ответ на вопрос, водителей какого пола чаще останавливают, используйте код:

```
police['driver_gender'].value_counts(normalize=True)
```

Вопрос **5**

Верно

Баллов: 1,00 из 1,00

Можно ли по предыдущему пункту сделать вывод, что к одному из полов относятся предвзято, чаще останавливают?

Выберите один ответ:

☒ Нет нельзя, так как мы не знаем пропорцию по водителям (какую долю составляют мужчины и женщины)

☐ Да. Останавливают чаще мужчин, причем в среднем на больший промежуток времени

☐ Нет, нельзя, потому что такой вывод можно сделать только из выборки, где присутствуют молодые мужчины (до 21 года)

Спасибо за ваше решение! Вы получили нужный результат!

Чтобы получить ответ на вопрос, водителей какого пола чаще останавливают, используйте код:

```
police['driver_gender'].value_counts(normalize=True)
```

Вопрос **6**

Выполнен

Балл: 1,00

Исследование произвольных данных

Это задание не проверяется LMS, но выполняя его, вы тренируете и закрепляете навыки работы с функциями библиотеки Pandas на различных исходных данных.

Для этого задания вы можете использовать любой доступный вам набор данных, например, загрузите заинтересовавший вас набор данных с сайта [kaggle](#).

Что нужно сделать:

- создайте Python-ноутбук, загрузите в него данные из csv-файла и преобразуйте в объект DataFrame;
- посмотрите полученный объект;
- вычислите описательные статистики для DataFrame;
- поразмышляйте, как вы можете использовать срезы для анализа вашего набора данных, получите необходимый срез(ы) DataFrame;
- определите, какую полезную информацию, вы можете получить о данных, используя отбор и фильтрацию, сформируйте необходимый поднабор данных и проанализируйте его;
- выполните сортировку применительно к исследуемому набору данных;
- потренируйтесь в переименование колонок исследуемого DataFrame;
- проверьте DataFrame на наличие пропущенных значений, выберите один из методов работы с пропусками и примените его;
- проверьте DataFrame на наличие дубликатов;
- сделайте предположение, какой новый признак вы могли бы ввести для исследуемого набора данных, и реализуйте его в созданном DataFrame.

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

```
# набор данных с сайта kaggle
# [International football results from 1872 to 2023](https://www.kaggle.com/code/fedescarl/starter-international-football-results-d4b7fd1e-b)
#
import random
import pandas as pd
goalscorers = pd.read_csv('assets/goalscorers.csv', sep=',')
goalscorers.sample(5)
goalscorers.describe()
goalscorers.info()
goalscorers.isnull().sum().sort_values(ascending = False)
goalscorers.columns
goalscorers.groupby('date').value_counts(normalize=True).sort_values()
goalscorers.rename(columns={'team': 'superteam'}, inplace=True)
goalscorers[goalscorers.duplicated()] # 81 rows x 8 columns
```

Спасибо! Уверены, что вы хорошо постарались и поработали!

[Библиотека Pandas](#) предоставляет множество полезных функций для работы с данными. Вот некоторые из них.

- Чтение и запись данных в различных форматах.
- Обработка данных: функции для фильтрации, сортировка, группировка, агрегация, преобразование и объединение данных.
- Индексация и выборка данных: функции для выборки данных по условиям, индексации по метке и индексу.
- Работа с пропущенными данными и дубликатами: функции для обнаружения, удаления и заполнения пропущенных значений, функции для проверки и удаления дубликатов.

Если у вас возникли вопросы по заданию, пожалуйста, задайте их в форму для ближайшего семинара.

Существуют множество других функций и возможностей Pandas, которые мы не успели рассмотреть в рамках нашего курса. Самостоятельное изучение оставшихся функций библиотеки может быть полезным и интересным опытом для вас. Желаем успехов в обучении!

