

### Практическое задание №1.3.3

1. Загрузите данные из файла в объект *DataFrame*, Добавьте заголовки к столбцам: «index», «year», «month», «day», «min\_t», «average\_t», «max\_t», «rainfall».

Расшифровка:

- index – индекс ВМО,
  - year – год,
  - month – месяц,
  - day – день,
  - min\_t – минимальная температура воздуха,
  - average\_t – средняя температура воздуха,
  - max\_t – максимальная температура воздуха,
  - rainfall – количество осадков.
2. Удалите столбец index.
  3. Используя метод info(), ответьте на вопросы:
    - 3.1. Есть ли в данных пропущенные значения?
    - 3.2. В каком столбце данных больше всего пропущенных значений?
  4. В данных за какой год больше всего пропусков?
  5. Объедините столбцы «Год», «Месяц» и «День» в один столбец «Дата» в формате гггг-мм-дд (2000-01-20). Данные в новом столбце должны иметь формат *datetime*;
  6. Для каждого наблюдения рассчитайте размах температур (разность максимальной и минимальной суточных температур) и количество предшествующих ему дней без осадков (используйте циклы Python и условный оператор):

Мин. темп. воздуха	Сред. темп. воздуха	Макс. темп. воздуха	Кол-во осадков	Размах темп.	Кол-во дней без осадков
-12.4	-11.0	-9.9	3.9	2.5	0
-28.1	-14.8	-9.8	3.8	18.3	0
-38.5	-34.6	-26.6	0.0	11.9	0
-34.6	-30.1	-23.4	0.0	11.2	1
-26.8	-21.4	-16.6	1.1	10.2	2
-28.6	-24.2	-17.4	0.8	11.2	0
-31.0	-27.0	-24.0	0.0	7.0	0
-33.3	-30.3	-24.6	0.0	8.7	1

7. Определите самый длинный период засухи.
8. Для каждого года вычислите среднегодовую температуру и общее количество осадков. Запишите результаты в объекты *Series*.
  - 8.1. Какой год можно считать самым теплым? Какой самым холодным?
  - 8.2. В какой год выпало больше всего осадков? В какой меньше всего?

Используя запись `имя_серии.plot()` вы можете построить график и посмотреть как изменялась температура. С помощью `имя_серии.plot.bar()` можно отобразить на столбиковой диаграмме количество осадков, выпавших в каждый год.

9. Выведите наблюдения, удовлетворяющие условиям:

9.1. Средняя температура воздуха ниже  $-30^{\circ}\text{C}$  (для некоторых регионов можно использовать  $-10^{\circ}\text{C}$ ,  $-35^{\circ}\text{C}$ ,  $-40^{\circ}\text{C}$ ).

9.2. Средняя температура воздуха выше  $27^{\circ}\text{C}$  и количество дней без осадков больше 3.

### Полезные функции и методы

Одну и ту же задачу можно решить несколькими способами. Эти функции и методы могут вам понадобиться:

- [`read\_table\(\)`](#) – загрузить данные из «.txt»;
- [`.head\(\)`](#) – отобразить несколько первых строк *DataFrame*;
- [`.info\(\)`](#) – информация о *DataFrame*;
- [`.drop\(\)`](#) – удалить строки или столбцы;
- [`.dtypes`](#) – узнать тип данных в столбце;
- [`.astype\(\)`](#), [`to\_datetime\(\)`](#), [`.to\_numeric\(\)`](#) – изменить тип данных;
- [`.isnull\(\).sum\(\)`](#) – вычислить количество пропущенных значений в каждом столбце;
- [`.max\(\)`](#), [`.min\(\)`](#), [`.mean\(\)`](#) – максимум, минимум, среднее значение;
- [`pd.Grouper\(\)`](#), [`.groupby\(\)`](#) – группировка наблюдений;
- [`.agg\(\)`](#) – агрегирование наблюдений;
- [`.truncate\(\)`](#) – логическая индексация (можно использовать даты!);
- Уже знакомые вам операторы тоже работают с *pandas*. Действие (или условие) выполняется (или проверяется) для каждого наблюдения. Так, например, чтобы найти разность между двумя числовыми характеристиками (столбцами) по всему набору данных, используйте оператор «-»:  
имя\_ DF [«новый\_столбец»] = имя\_ DF [«столбец\_1»] - имя\_ DF [«столбец\_2»]  
Если необходимо найти сумму двух числовых характеристик или склеить строковые значения двух столбцов – используйте «+». Если вы хотите получить значения, которые больше заданного числа, например 5, используйте запись  
имя\_ DataFrame [«имя\_столбца»] > 5
- Одним из вариантов выполнения задания 4 может быть следующая конструкция:

```
# В данных за какой год больше всего пропусков
for i in range(1960, 2020):
    print(i, df[df['year'] == i].isnull().sum().sum())
```

- Шпаргалка: <https://smysl.io/blog/pandas/>.