

Лабораторная работа №1

1. Загрузите данные из файла «wr9464.xlsx» в *DataFrame*;
2. Добавьте заголовки к столбцам: «Индекс ВМО», «Год», «Месяц», «День», «Минимальная температура воздуха», «Средняя температура воздуха», «Максимальная температура воздуха», «Количество осадков»;
3. 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*;
4. Переместите дату в индекс;
5. Удалите столбец «Индекс ВМО»;
6. Ответьте на вопросы:
 - 6.1. Содержатся ли в данных пропущенные значения?
 - 6.2. В каком столбце данных больше всего пропущенных значений?
 - 6.3. В климатических показателях за какой год содержится наибольшее количество пропусков?
7. Удалите все наблюдения (дни) до 01.01.1966 г.;
8. Для каждого наблюдения рассчитайте размах температур (разность максимальной и минимальной суточных температур) и количество предшествующих ему дней без осадков:

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

9. Запишите получившийся *DataFrame* в файл формата «.xlsx»;
10. Определите самый длинный период засухи;
11. Вычислите среднюю температуру воздуха в июне и максимальную температуру воздуха в январе каждого года;
12. Выведите наблюдения, удовлетворяющие условиям:
 - 12.1. Средняя температура воздуха ниже -30 °С,
 - 12.2. Средняя температура воздуха выше 25 °С и количество дней без осадков больше 3.

Полезные функции и методы

- [pd.read_excel\(\)](#) – загрузить данные, представленные в формате «.xlsx»;
- [pd.read_csv\(\)](#) – загрузить данные формата «.csv»;
- Заголовки столбцов могут быть переданы списком в параметр *names* функций [pd.read_excel\(\)](#) и [pd.read_csv\(\)](#);
- [.head\(\)](#) – отобразить несколько первых строк *DataFrame*;
- [.set_index\(\)](#) – изменить индексы;
- [.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!);
- [.to_excel\(\)](#), [.to_csv\(\)](#) – запись *DataFrame* в файл формата «.xlsx» или «.csv»;
- 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 характеристика 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 6.3 может быть следующая конструкция:

```
for i in range(1963, 1967):  
    print(i, df[df.index.year == i].isnull().sum().sum())
```

```
1963 2  
1964 0  
1965 96  
1966 0
```

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