

## Лабораторная работа №5

### Визуализация данных

По данной работе необходимо подготовить отчет в формате блокнота Jupyter Notebook (.ipynb) либо в текстовом виде (.pdf). В отчете должны быть:

- 1) исходные коды
- 2) результаты выполнения.

*В этой работе представлено 2 варианта задания: основное и альтернативное (упрощенное). За основные задания можно получить максимум 5 баллов, за альтернативные – только 3. Выберите один из вариантов заданий, которые будете выполнять. Оба варианта заданий делать не нужно, т.к. в любом случае за эту лабораторную работу максимум 5 баллов.*

#### **Основное задание (5 баллов)**

Для визуализации можно использовать любые библиотеки на свой выбор, но обязательно продемонстрировать хотя бы 2 разных (например, matplotlib и seaborn).

1. Загрузите данные из файла «weather1.csv» о погоде в Перми. Загрузите только следующие столбцы:
  - a. Местное время в Перми
  - b. Т (температура воздуха в градусах Цельсия)
  - c. Р (атмосферное давление в мм.рт.ст.)
  - d. У (относительная влажность в %)
  - e. Ff (скорость ветра в м/с)
  - f. N (облачность)
  - g. H (высота основания облаков, м)
  - h. VV (горизонтальная дальность видимости в км)
2. (0.5 балла) Постройте точечную диаграмму (диаграмму рассеяния) по признакам температуры и относительной влажности.
3. (0.5 балла) На построенной в предыдущем пункте диаграмме выделите точки разными цветами в зависимости от облачности: синим – для которых облачность составляет 100%; красным – все остальные.
4. (0.5 балла) Постройте линейную диаграмму (график) изменения температуры в зависимости от местного времени.
5. (1 балл) Посчитайте по имеющимся данным среднемесячную температуру и постройте столбчатую диаграмму (вертикальную) зависимости средней температуры от месяца. Подсказка: создайте отдельный столбец с номером месяца (вычислив его из столбца «Местное время»), а затем сгруппируйте данные по этому столбцу.
6. (0.5 балла) Постройте ленточную диаграмму (горизонтальную), отразив на ней количество имеющихся наблюдений для каждого варианта облачности.
7. (0.5 балла) Постройте гистограмму частот для температуры. На гистограмме должно быть 10 диапазонов температуры.
8. (1 балл) Разбейте данные на 3 группы по значению горизонтальной дальности видимости (одна группа – дальность видимости менее 5 км, вторая – от 5 до 15 км(включительно), третья – более 15 км). В одной области для каждой группы постройте boxplot (диаграмму «ящик с усами») для признака «атмосферное давление».
9. (0.5 балла) Постройте круговую диаграмму для признака «высота основания облаков».

### **Альтернативное задание (3 балла)**

**Для визуализации можно использовать любую библиотеку на свой выбор.**

1. Создайте серии:
  - 1) Серия X, содержащая числа от 0 до 99 (`np.arange(100)`)
  - 2) Серия Y1, содержащая 100 случайных чисел с нормальным распределением с мат. ожиданием 0 и дисперсией 1 (`np.random.normal(0,1,100)`)
  - 3) Серия Y2, содержащая 100 чисел, вычисляемых по формуле  $y2[i]=x[i]+e[i]$ , где  $e[i]$  – случайное целое число из диапазона [-2;2] с равномерным распределением ( $y2 = x + np.random.randint(-2,3,100)$ ).
2. Создайте датафрейм из построенных на предыдущем шаге серий.
3. Постройте точечную диаграмму (диаграмму рассеяния) по признакам X и Y2.
4. Постройте линейную диаграмму (график) зависимости Y1 от X.
5. Постройте гистограмму частот для признака Y1. На гистограмме должно быть 10 диапазонов значений.
6. Разбейте данные на 3 группы по значению разности X-Y2 ( $X-Y2<0$ ;  $X-Y2=0$ ;  $X-Y2>0$ ). В одной области для каждой группы постройте boxplot (диаграмму «ящик с усами») для признака Y1.