**Лабораторная работа по библиотеке Pandas**

**Общие требования и рекомендации**

Исходные данные выбрать на сайте kaggle.com/datasets. При выборе данных обратить внимание на то, чтобы задания были выполнимы с указанным набором данных. (Следует отдавать предпочтение датасетам с бОльшим количеством столбцов, как с качественными, так и с количественными значениями, а также с большим количеством строк). Избегайте пересечений датасетов с одногруппниками.

Выполнение лабораторной работы осуществляется **с применением ООП.**

Описать класс для работы с датасетами, который может включать такие методы как: формирование датасета из файла с проверкой существования файла, удобочитаемый вывод датасета, нахождение min, max, mean и других статистических значений для всех числовых полей, построение диаграмм, создание сэмплов и другие.

**Вариант 1**

1. Изменить название любых двух столбцов. А также изменить значения в любом столбце по примеру:  
   *столбец ‘Пол’ было – муж  
   столбец ‘Пол’ стало – мужской.*
2. Добавьте новый столбец в DataFrame. Например путем слияния двух связанных по смыслу столбцов с качественными значениями или путем какого-либо вычисления между столбцами с количественными значениями.
3. Посчитать минимальное, максимальное, среднее, медиану по каждому числовому полю.
4. Сгруппировать данные по одному из полей, посчитать сумму(среднее) для каждой группы, отсортировать группы по возрастанию(убыванию). Отобразить гистограмму средних значений групп.
5. Отфильтровать значения по какому-либо количественному полю и вывести информацию об уникальных значениях, попавших в диапазон, о количестве дубликатов.
6. Построить график(и) (гистограмму(ы)) для средних значений групп по отфильтрованным данным.
7. Построить перекрестную выборку по набору данных (дежурный вариант: столбцы – минимальное, максимальное, среднее, медиану по каждому числовому полю; строки – названия полей). Построить график этого выборки (3d-гистограмму в частном случае).

**Вариант 2**

1. Найти дубликаты строк в DataFrame и удалить их.
2. Вывести уникальные значения столбца и их количество.
3. Посчитать минимальное, максимальное, среднее, медиану по каждому числовому полю.
4. Сгруппировать данные по одному из полей, посчитать сумму(среднее) для каждой группы, отсортировать группы по возрастанию(убыванию). Отобразить гистограмму средних значений групп.
5. Посчитать минимальное, максимальное, среднее, медиану по какому-либо полю, предварительно отфильтровав значения по какому-либо качественному полю.
6. Построить график(и) (гистограмму(ы)) для средних значений групп по отфильтрованным данным.
7. Построить перекрестную выборку по набору данных (дежурный вариант: столбцы – минимальное, максимальное, среднее, медиану по каждому числовому полю; строки – названия полей). Построить график этого выборки (3d-гистограмму в частном случае).

**Вариант 3**

1. Определить количество пропущенных значений одного-двух столбцов. Заменить пропущенные значения в числовом столбце на среднее арифметическое его значений, в качественном – на самое популярное значение, либо придумать замену самостоятельно.
2. Создать список значений и проверить их наличие в каком-либо столбце.
3. Посчитать минимальное, максимальное, среднее, медиану по каждому числовому полю.
4. Сгруппировать данные по одному из полей, посчитать сумму(среднее) для каждой группы, отсортировать группы по возрастанию(убыванию). Отобразить гистограмму средних значений групп.
5. Посчитать минимальное, максимальное, среднее, медиану по какому-либо полю, предварительно отфильтровав значения.
6. Построить график(и) (гистограмму(ы)) для средних значений групп по отфильтрованным данным.
7. Построить перекрестную выборку по набору данных (дежурный вариант: столбцы – минимальное, максимальное, среднее, медиану по каждому числовому полю; строки – названия полей). Построить график этого выборки (3d-гистограмму в частном случае).