# **Лабораторная работа №5**

Работа посвящена введению в анализ данных с помощью языка Python и библиотек pandas, numpy, matplotlib. Разбор лекционного материала в примере анализа данных московских баров Вы сможете найти по ссылке: <https://colab.research.google.com/drive/1myMa_0FPWvGwX95en2BMCCYxlWNkrajm?usp=sharing>. Все датасеты для своих работ берём с предложенных ниже сайтов (или по желанию с других ресурсов, но эти ресурсы должны быть указаны в коде):

* <https://data.mos.ru/opendata>
* <https://stats.oecd.org/>
* <https://www.wto.org/english/res_e/statis_e/statis_e.htm>

**Постановка задачи:**

1. **Подготовка.** Выберите интересный для Вас датасет. Датасет должен иметь числовые характеристики и потенциал для анализа. Желательно, чтобы данные не повторялись. Во время лекционного занятия Вы записывали свои темы здесь: <https://docs.google.com/spreadsheets/d/1czHRbCRNsHdbXQ8Y-TUqerD8k2ERakjZABjUydXYsZ4/edit?usp=sharing>.
2. Следующая часть работы посвящена **анализу** выбранной предметной области с официальными данными. Для этой части Вам необходимо выполнить следующие основные шаги:
   1. Скачать датасет в формате .csv (или другом табличном формате, переформатировать с CSV при помощи Excel, Google Tables,...), загрузить данные в рабочее окружение (рекомендуем использовать интерпретаторы типа блокнот Jupyter Notebook, Google Collab).
   2. Посмотреть, какие данные нам нужны для анализа и сформулировать задачу анализа, убрать лишние данные из датасета (программно желательно).
   3. Посмотреть статистические характеристики по всем данным, выборочно по столбцам, сформулировать вывод.
   4. Выполнить группировку данных, которая поможет раскрыть потенциал данных для анализа и помочь в решении поставленной задачи.
   5. Продемонстрировать навыки агрегирования и сортировки данных. Использовать не менее двух параметров.
   6. Создать пользовательскую функцию вычисления вспомогательных значений (нахождение суммы, доли, произведения параметров). Интерпретировать полученный результат.
3. Заключительная часть работы посвящена **визуальному анализу данных**. Выполните каждое поставленное ниже задание **двумя** способами: с помощью функций построения диаграмм у датафрейма и с помощью библиотеки matplotlib.
   1. Постройте диаграмму рассеяния по имеющимся данным, интерпретируйте результат (что показывает диаграмма и для чего пригодится).
   2. Постройте круговую диаграмму, разметьте элементы диаграммы аккуратно, чтобы не было наложений. Вынесите необходимы подписи на диаграмму. Сформулируйте вывод.
   3. Постройте любую другую диаграмму (не разобранную на занятии), которая помогла бы дополнительно оценить данные и выявить их характерные особенности.