

Наименование работы:

Анализ и интерпретация данных

Цель работы:

Научиться рассчитывать статистические характеристики и визуализировать данные с целью построения гипотез и формирования практически значимых выводов по ним.

Задачи:

а) В соответствие с набором данных из лаб. №2 рассчитать статистические характеристики признаков данных:

1. медиана;
2. мода;
3. среднее;
4. перцентили;
5. стандартное отклонение;
6. минимальное и максимальное значения;
7. число пропущенных и уникальных значений.

б) Визуализировать признаки, например, посмотреть распределение или сезонность, того или иного ряда;

в) Найти какие-либо зависимости в данных, например, зависимость количества проданного магазином товара за одну сделку от цены товара

Для реализации рекомендуется использовать язык программирования Python 3.x и библиотеки Pandas, Matplotlib, seaborn

Оформление результатов:

Результаты лабораторной работы оформляются в виде отчета в формате PDF.

Структура отчета:

1. Титульный лист;
2. Основная часть;
3. Заключение.

В основной части приводится описание выполнения каждой из поставленных задач в виде текста и скриншотов программного кода

В заключении приводятся практически значимые выводы по проделанной работе

Таблица 1 - наборы данных

№	Набор данных
1	Данные по штрафам за парковку https://www.kaggle.com/new-york-city/nyc-parking-tickets (любой из 4-х файлов)
2	Данные о заболеваемости COVID-19 https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset
3	Данные о скачиваниях и рейтингах приложений в Google Play https://www.kaggle.com/lava18/google-play-store-apps
4	Данные о статистике суицидов по странам с 1985 по 2016 годы https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016
5	Данные по БУ авто с Craigslist https://www.kaggle.com/austinreese/craigslist-carstrucks-data