

Список 02 – Домашнее Задание

Введение о науки данных и визуализации

Предмед: Введение в профессиональную деятельность

Преподаватель: Хольгер Эспинола Ривера

1. Анализ и визуализации данных. Возьмите в качестве шаблона блокнот Jupyter из **урока 03** под названием **lab03_ds**, который вы использовали в лаборатории по науке о данных и анализу данных, и следуйте следующим инструкциям:

[1]. Выберите один набор данных с табличной структурой данных, который содержит числовые и категориальные типы данных (обязательно оба). Набор данных должен содержать не менее 500 строк и количество столбцов (признаков) мин. = 6 и макс. = 50. Вы можете использовать собственные наборы данных из своих прошлых проектов, своей компании или взять один из общедоступных наборов данных в самых популярных репозиториях, таких как:

- UCI (<https://archive.ics.uci.edu/datasets>);
- Kaggle (<https://www.kaggle.com/datasets>);
- Поиск наборов данных Google (<https://datasetsearch.research.google.com>).

Ваш набор данных может иметь формат файла, такой как .csv, .xls, .json, .sql, .data, .txt... и т. д.;

[2]. В группах по 4 студентам, разработайте и опубликуйте свой собственный проект по науке о данных в репозитории Github, выполнив статистический анализ числовых и категориальных переменных. Необходимо иметь 1 папку, содержащую файлы набора данных, и 1 файл блокнота Jupyter .ipynb, содержащий исходный код Python;

[3]. Определите метаданные, сбор данных и загрузите их как фрейм данных и позаботьтесь об отсутствующих значениях;

[4]. Визуализация данных для числовых переменных: построение гистограмм частот, бивариантных диаграмм, диаграмм с усами, расчет статистических показателей и построение таблицы корреляции числовых характеристик;

[5]. Визуализация данных для категориальных переменных: построение столбчатой диаграммы для абсолютных частот для каждого экземпляра с использованием всех категориальных переменных, построение круговой диаграммы для относительных частот для каждого экземпляра с использованием всех категориальных переменных;

[6]. Проведите статистический анализ групп категориальных переменных: выберите 2 категориальные переменные из вашего набора данных. Создайте группы, используя экземпляры этих двух категорий, и подсчитайте количество образцов для каждого из этих объединенных экземпляров. Постройте групповой анализ абсолютной частоты с помощью столбчатой диаграммы и групповой анализ относительной частоты с помощью круговой диаграммы.

[7]. Используя программное обеспечение Grafana и тот же набор данных, создайте панель мониторинга, содержащую 2 примера статистического анализа для групп категориальных переменных.