Название проекта – программа для предварительной обработки DataFrame

Автор Проекта – Чугунов Арсений

Описание идеи:

Программа нужна для того, чтобы получить первичный анализ данных, узнать общею информацию про дата сет, про каждую переменную в отдельности и корреляцию между ними. Без этого невозможно начать полноценный анализ данных. С построением гипотез и моделей.

Описание реализации:

Основные классы: QMainWindow(основное окно и окно для вывода графиков)

При помощи PyQt5 был создан весь дизайн.

При помощи pandas были получены статистические данные

При помощи seaborn были построены графики

Интересный прием: для записи я использую вспомогательную папку, в которой хранятся графики, оттуда их потом может позаимствовать пользователь. Так же я рассчитываю все данные 1 раз при запуске.

Описание технологий:

Запуск основного окна

Получение названия файла

Обработка файла, создание словаря со всеми нужными далее переменными

Передача словаря обратно в класс окна

Создание дизайна

На первой вкладке 5 случайных строк DataFrame

На второй вкладке информация про DataFrame целиком

На третьей вкладке информация про переменные, нужно выбрать какая переменная вам интересна через ComboBox. (Реализация вывода информации: старый текст удаляется, но вый добавляется)

На четвертой вкладке график корреляции числовых столбцов

На пятой будут инструменты для удобной очистки обработки датасета.

Основные библиотеки:

Pandas, PyQt5, seaborn, matplotlib, sys, os и другие.