Министерство образования Республики Беларусь

Учреждение образования

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

КАФЕДРА ИНФОРМАТИКИ

Отчёт по лабораторному проекту

По дисциплине «Модели и методы обработки больших объёмов данных»

По теме «Формирование ленты новостей по предпочтениям (тегам)»

Выполнил:

магистрант гр. 156341

Кожемякин П.Н.

Проверил:

доцент кафедры информатики

Стержанов М.В.

Минск 2021

# **Цель работы**

Создать проект, который будет анализировать данные о аренде жилья с интернет-ресурсов (Zoopla - <https://www.zoopla.co.uk>). Полученные данные, анализируются все с помощью Pandas. Затем строит их с помощью Seaborn. Сделать выводы об особенностях и закономерностях аренды жилья на данном сайте.

# **Ход работы**

Необходимо создать проект, состоящий из 3 частей.

1. Расширяемый модуль, в который можно добавить обработчики открытых API интернет-ресурсов. Данный модуль необходим для сбора информации.

2. Модуль, который будет рассортировывать и хранить собранную и обработанную информацию.

3. Модуль, реализующий некоторый анализ на данную тему.

**Получение данных**

Для начала разработки был выбран один из самых популярных интернет-изданий Zoolpa. Данный ресурс был выбран по нескольким причинам.

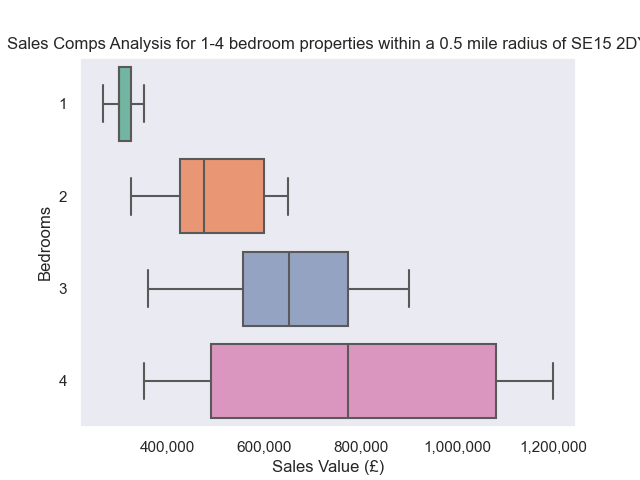
1. Очень обширная база строений, которые идет как на продажу, так и на аренду.

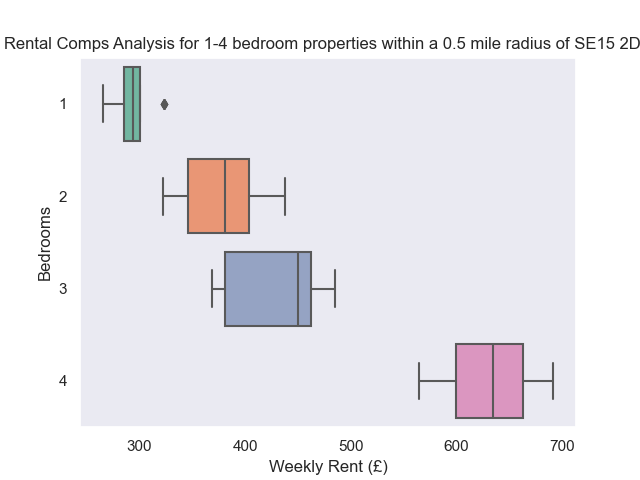
2. Имеется открытое API, для получения данных (для использования надо запросить Api-key).

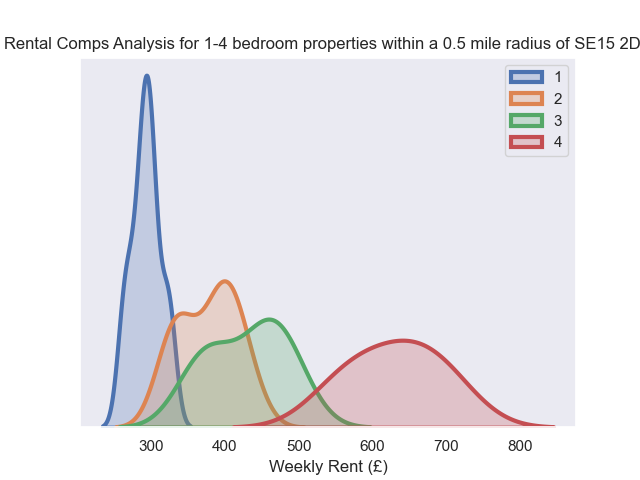
3. API запросы позволяют выбрать конкретную информацию для скачивания, что сильно ускоряет реализацию проекта, т.к. позволяет пропустить этап ручной обработки (разметки) данных.

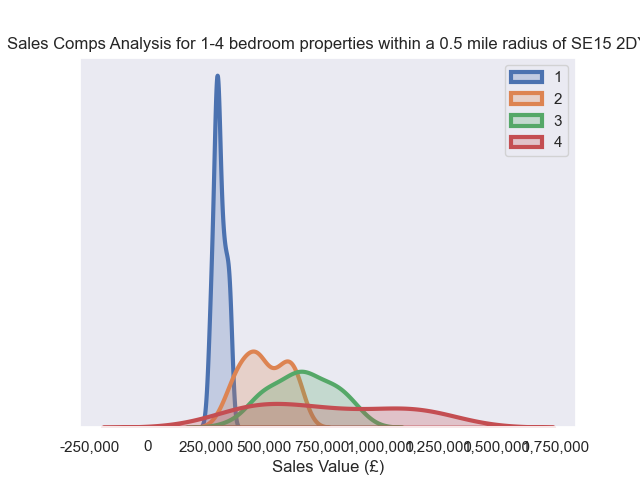
Чтобы начать использовать API, необходимо пройти по ссылке https://developer.zoopla.co.uk и запросить ключ доступа. Этим ключом мы будем помечать все наши запросы, чтобы нам была доступна не только тестовая коллекция новостей, но и остальные статьи.

**Аналитика**









**Выводы**

В ходе работы были получены данные с сайта Zoopla и проанализированы. Были использован язык разработки python, которые обладает множеством средств для ускоренного получения данных.

В ходе работы были получены базовые знания о возможностях, устройстве, области применения и интерфейсе API.

Было проанализировано данные о съеме жилья.