**Python Project**

**Developer Guide**

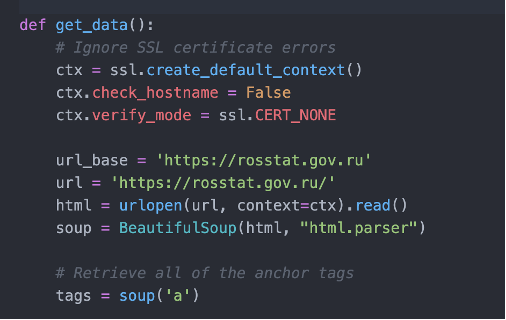
**Индекс потребительских цен на продукты и услуги с 2010 до 2021**

**Raspars.py**

Вначале мы импортируем все необходимые для программы библиотеки.



Затем мы определяем функцию, которая импортирует данные с сайта [https://rosstat.gov.ru](https://rosstat.gov.ru/).



Мы создали переменную soup, которая создает структуру данных, представляющую “распарсенный” HTML документ, используя функцию BeautifulSoup(,) для дальнейшего парсинга по тэгам.

Затем приступаем к парсингу необходимых нам данных для приложения. Под переменную х резервируется пустой список. Далее цикл перебирает все тэги и сохраняет список строк, совпадающих с паттерном “statistic”



Далее в коде идут повторяющиеся циклы, мы попадаем на нужную нам страницу, находим нужный нам тэг, подтягиваем ссылку и переходим в следующий цикл. Переходим из раздела «официальная статистика» в раздел «цены».



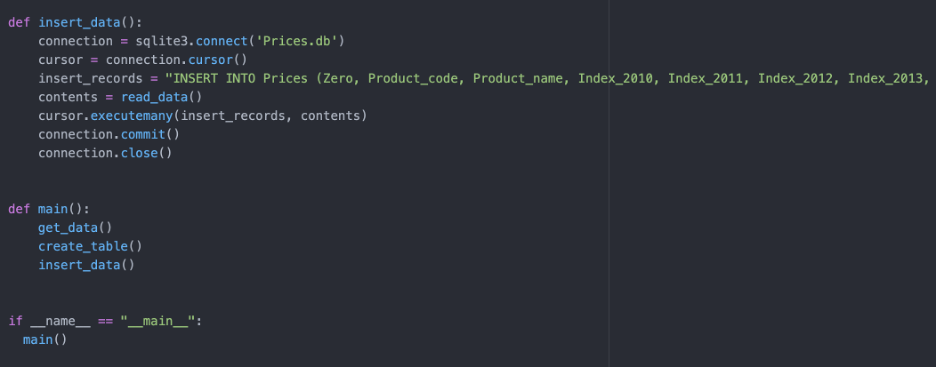
Далее выгружаем нужный нам .xls файл и трансформируем его в .csv



Далее декодируем данные и создаем таблицу для данных в SQLite.



Заполняем таблицу данными.



**Extract.py**

**Вторая часть приложения позволяет визуализировать выборку данных из базы данных Prices.sql**

**Импорт библиотек**

import sqlite3. **для работы с БД sql**

import pandas as pd **применения инструментов работы с данными pandas**

import matplotlib.pyplot as plt **применение инструментов визуализации**

**Изображение выглядит как текст

Автоматически созданное описание**

**Функция extract\_data() предназначена для извлечения выборки данных из prices.db и их включения в таблицу данных DataFrame**

**Изображение выглядит как текст

Автоматически созданное описание**

**Объект ‘year’ содержит диапозон 2010-2022**

**Цикл for запускает 4 итерации**

**Объекту ‘s’ присваиваются данные из базы данных DataFrame (из ячейки строка-столбец), где ‘I’ строка, [1:13] столбец.**

**plt.title(df[0][i]) извлекает значение из базы данных, содержащее наименование продукта и отображает в качестве заголовка при визуализации**

|  |  |
| --- | --- |
| **plt.xlabel('Год', fontsize=12).**  **plt.ylabel('Прирост цены в %', fontsize=12)** | **Значения координатной плоскости, отображаемые при визуализации** |

**Метод plt.plot(year, s) непосредственно строит график из полученных данных**

**Метод plt.savefig(str(df[0][i])+'.pdf') сохраняет визуализацию в формате pdf и присваивает наименование наименование, соответствующее наименованию анализируемого продукта**

**plt.clf() очищает из оперативной памяти информацию о созданной визуализации и готовит приложение к следующей итерации.**

**Цикл совершает 4 итерации и создает соответственно 4 pdf файла, визуализирующих данные по 4 различным продуктам из выборки**