МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра вычислительных технологий**

**ЛАБОРАТОРНАЯ РАБОТА №5**

**Дисциплина: Обработка больших данных**

**Тема: «Извлечение данных с WEB-страниц. Пакет rvest.»**

Работу выполнил:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Курбатский В.А.

Направление подготовки: 02.03.02 Фундаментальная информатика и информационные технологии

Направленность (профиль): Компьютерные технологии и прикладная математика.

Преподаватель: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Приходько Т. А.

Краснодар

2023

Вариант 7

Страна: **Турция, Греция, Египет, Австралия, Новая Зеландия**

**Постановка задачи**

1. Изображение выглядит как стол

   Автоматически созданное описаниеВ ходе лабораторной работы, необходимо собрать информацию об уровне жизни стран мира из таблиц сайта <https://www.numbeo.com/quality-of-life/rankings_by_country.jsp?title=2021> с 2014 по 2021гг:

## Для дополнительной информации по Рейтинг стран по уровню жизни можно использовать ссылку <https://tyulyagin.ru/ratings/rejting-stran-mira-po-urovnyu-zhizni-2021.html>

**Это** оценка общего качества жизни с использованием эмпирической формулы, которая учитывает:  
— индекс покупательной способности (чем выше, тем лучше),   
— индекс загрязнения (чем ниже, тем лучше),   
— отношение цены на жилье к доходу (ниже). лучше),   
— индекс прожиточного минимума (чем ниже, тем лучше),   
— индекс безопасности (чем выше, тем лучше),   
— индекс медицинского обслуживания (чем выше, тем лучше),   
— индекс времени движения на дороге (чем ниже, тем лучше)   
— климатический индекс (чем выше, тем лучше).

1. Каждый студент должен взять 5 стран (по варианту) .
2. Составить data.frame (возможно для каждой страны) так, чтобы иметь возможность **проанализировать с помощью графиков изменение рейтингов для всех 10 показателей для всех своих 5-ти стран, прокомменитровать в отчете результат.**

Необходимо нарисовать на одном и том же графике рейтинг всех 5 стран, проанализировать резутьтат. Проанализировать изменение во времени всех показателей указанных стран, подобрать наилучший (с вашей точки зрения) способ визуалазации.

1. С одной из страниц (<https://kudago.com/spb/list/33-luchshih-muzeya-peterburga/> или <https://tonkosti.ru/Музеи_Санкт-Петербурга> ) собрать информацию в data.frame, которя содержала бы: Название музея, его адрес и ссылку для перехода при клике на фото музея.

**Ход работы:**

Используя функции пакета «rvest», сформировал дата фрейм, взяв данные с 2020 года по 2023 с сайта «https://www.numbeo.com» для стран из моего варианта.

Для анализа полученных данных построил графики по каждому из параметров.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 1 – Индекс уровня жизни.

По данному графику видно, что самыми благоприятными для жизни странами являются Австралия и Новая Зеландия, так как за последние 4 года показатель не опускался ниже 175 баллов. Греция и Турция находятся почти на одном уровне, который не превышает 135 баллов. Хуже всего жить в Египте, его индекс не превышает 100 баллов.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 2 – Индекс покупательской способности

По данному параметру явным лидером является Австралия. Новая Зеландия не сильно отстаёт от лидера. Египет, Турция и Греция значительно отстают, их индекс не превышает 45 баллов.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 3 – Индекс безопасности

По данному графику видно, что уровень безопасности в течении 4 последних лет упал у всех стран. Также можно отметить, что Турция является самой безопасной страной из представленных.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 4 – Индекс уровня здравоохранения

Турция, Новая Зеландия и Австралия близки к друг другу по индексу, который держится выше 70 баллов последние 3 года. Самый низкий уровень зафиксирован у Египта – ниже 50 баллов. Греция имеет уровень выше, но также сильно отстает от Турции, Новой Зеландии и Австралии.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 5 – Индекс прожиточного минимума

Австралия и Новая Зеландия имеют достаточно высокий уровень жизни и безопасности, что отражается на индексе прожиточного минимума не в лучшую сторону. В Греции ситуация лучше, а Египет и Турция являются лидерами данного рейтинга.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 6 – Отношения стоимости недвижимости к заработной плате

Египет не располагает высоким качеством жизни, что негативно сказывается на данном показателе. В Новой Зеландии и Австралии все хорошо. Также на графике видно резкий скачек для Турции после 2021 года. К 2023 году Греция, Турция и Египет пришли почти к одному показателю.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 7 – Индекс времени в дороге

Самый высокий показатель у Египта, следом за ним идет Турция. Греция, Австралия и Новая Зеландия имеют показатели сильно лучше.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 8 – Индекс уровня загрязнения

Новая Зеландия и Австралия являются высока развитыми странами, где много внимания уделяют экологическим проблемам, они имеют самые низкие показатели загрязнения, что нельзя сказать об Египте, Греции и Турции. Самая плачевная ситуация в Египте, его индекс преодолел планку в 80 баллов.

Изображение выглядит как диаграмма

Автоматически созданное описание

Рисунок 9 – Климатический индекс

Новая Зеландия имеет самый высокий показатель – выше 95 баллов. Остальные странны занимают примерно один уровень. Также нужно отметить, что к 2023 году в Турции произошло падение данного параметра.

Также я собрал информацию о музеях Санкт-Петербурга с сайта «https://kudago.com/spb/list/33-luchshih-muzeya-peterburga/» и образовал из них дата фрейм.

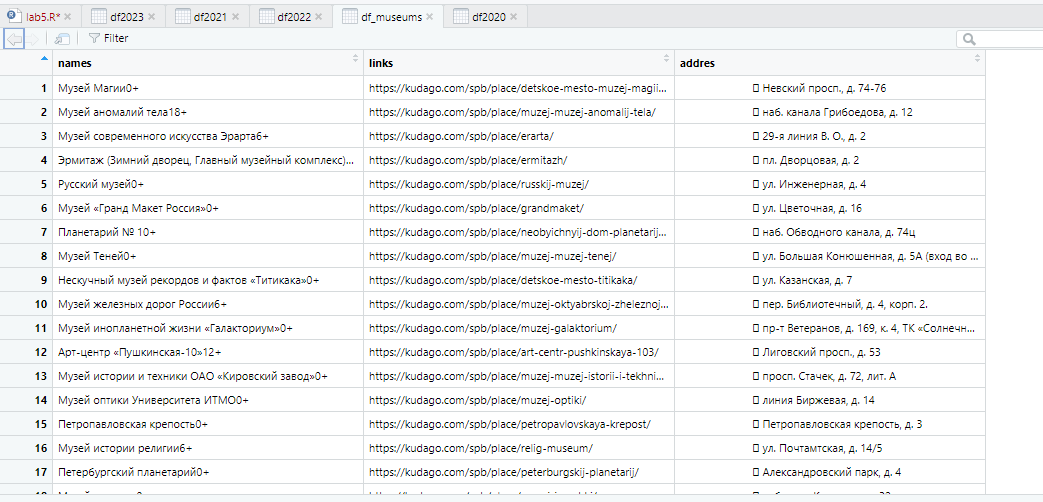


Рисунок 10 – Часть полученного дата фрейма