

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Омский государственный технический университет»

Факультет информационных технологий и компьютерных систем
Кафедра «Прикладная математика и фундаментальная информатика»

Расчётно-графическая работа
по дисциплине «Практикум по программированию»

Студента	Курпенова Куата Ибраимовича
	<small>фамилия, имя, отчество полностью</small>
Курс	2, группа ФИТ-212
Направление	02.03.02 Прикладная математика и фундаментальная информатика
	<small>код, наименование</small>
Руководитель	ст. преподаватель
	<small>должность, ученая степень, звание</small>
	Шарун И. В.
	<small>фамилия, инициалы</small>
Выполнил	
	<small>дата, подпись студента</small>

Итоговый рейтинг	
------------------	--

Омск 2022

Содержание

1	Введение	3
1.1	Поиск и загрузка данных	3
1.2	Разведывательный анализ данных	4
2	Заключение	6
3	Список использованных источников	6

1 Введение

Анализ данных - область математики и информатики, занимающаяся построением и исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных (в широком смысле) данных. Анализ данных — это не просто обработка информации после ее получения и сбора, это средство проверки гипотез. Цель любого анализа данных — понимание исследуемой ситуации целиком (выявление тенденций, в том числе негативных отклонений от плана, прогнозирование и получение рекомендаций). Для достижения этой цели ставятся следующие задачи анализа данных:

- сбор информации
- структуризация информации
- выявление закономерностей, анализ
- прогнозирование и получение рекомендаций

Сейчас аналитика применяется практически во всех сферах: маркетинг, медицина, таргетированная реклама, банковская сфера и т.д. Грамотный аналитик данных в любой компании становится ключевой фигурой. Специалист по анализу данных это тот работник, который программирует лучше любого специалиста по статистике, и знает статистику лучше любого программиста, в бизнес-процессах разбирается лучше руководителя.

Самый популярный инструмент дата сайентиста это, бесспорно, язык программирования python. В данной работе он был использован, так как и ключевая для анализа данных библиотека pandas. Pandas содержит огромное количество методов работы с данными: от импорта датасета до построения графиков, вывода всевозможных характеристик.

1.1 Поиск и загрузка данных

В данной работе я использовал датасет «Dairy queen menu nutrition». Этот датасет описывает питание - продукты, калории, доли полезных веществ в них.

1.2 Разведывательный анализ данных

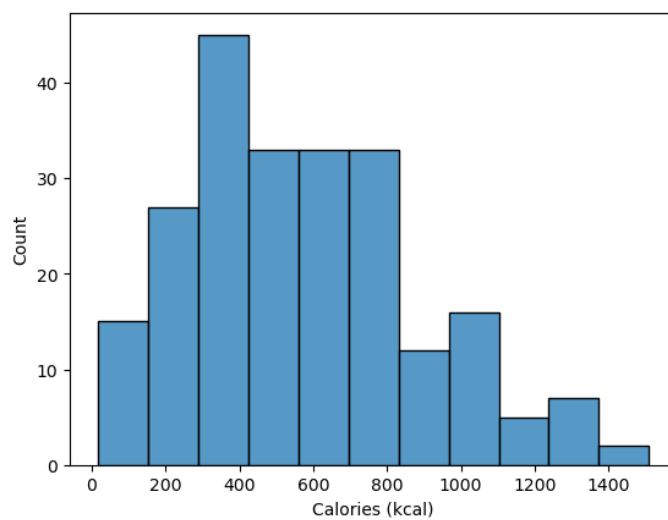


Рис. 1: Гистограмма распределения по столбцу «Calories (kcal)»

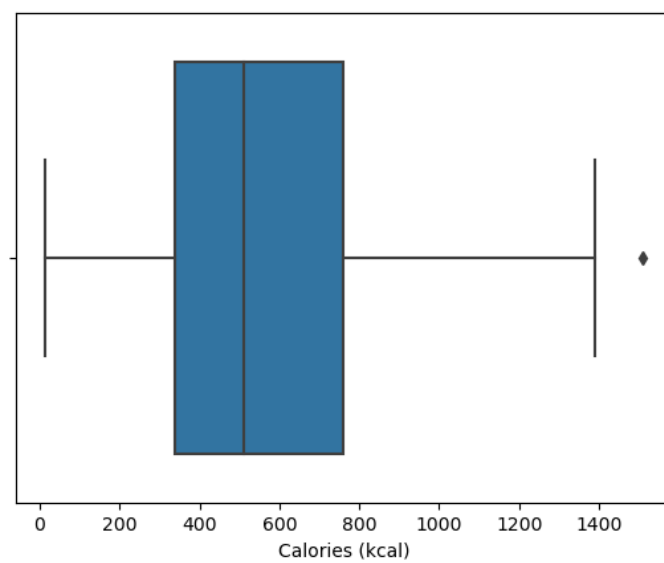


Рис. 2: Диаграмма размаха по столбцу «Calories (kcal)»



Рис. 3: Круговая диаграмма по столбцу «Fiber (g)»

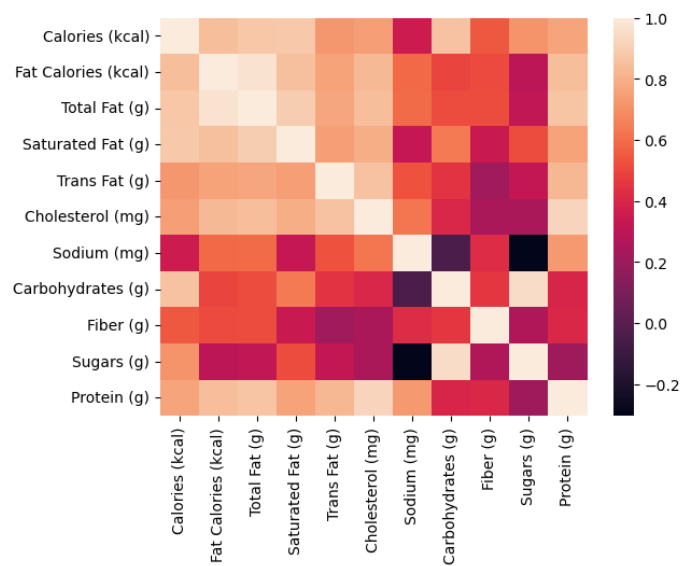


Рис. 4: Тепловая карта корреляции параметров набора данных

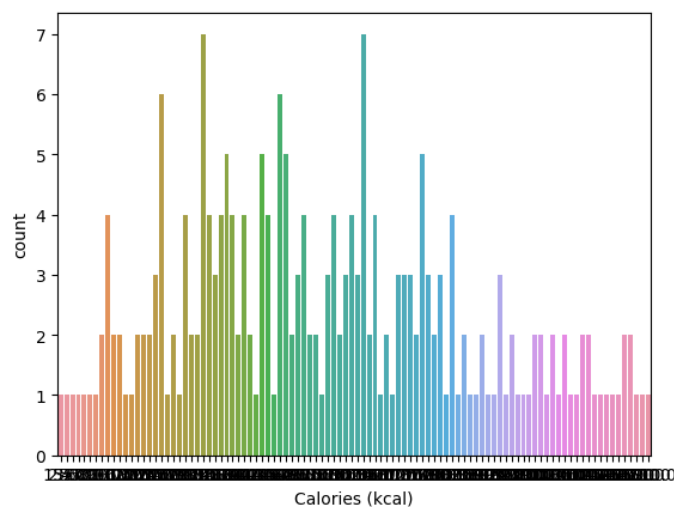


Рис. 5: Диаграмма countplot по столбцу «Calories (kcal)»

2 Заключение

В рамках задания был успешно загружен датасет. Затем производилось удаление дубликатов и вывод основной информации. Осуществлялись переименование, удаление столбцов, а также их различная визуализация и анализ.

Данная работа дала начальные навыки обработки и анализа данных с помощью различных, популярных инструментов языка python (pandas, seaborn, scipy).

3 Список использованных источников

- Pandas documentation Date: Nov 22, 2022 Version: 1.5.2
- Seaborn. User guide and tutorial
- SciPy documentation Date: October 19, 2022 Version: 1.9.3