**Отчёт**

**по лабораторной работе №1**

**по предмету «Алгоритмы и структуры данных»**

**на тему «Создание датасета»**

**Вариант – 2**

|  |  |  |
| --- | --- | --- |
| Студент гр. 24.Б16-пу |  | Музиров Б.Р. |
| Преподаватель |  | Дик А.Г. |

Оглавление

1. Цель работы.
2. Формализация задачи.
3. Теоретическая часть.
4. Основные шаги программы
5. Блок-схема пошагового выполнения алгоритма.
6. Описание программы.
7. Рекомендации пользователя.
8. Рекомендации программиста.
9. Описание контрольного примера.
10. Вывод по работе.

**Цель работы**

Разработать программу для генерации синтетического датасета посетителей платной поликлиники, имитирующего реальные данные.

**Формализация задачи**

Задача состоит в создании датасета, обладающего следующей структурой:

1. **ФИО:** только славянские варианты.
2. **Паспортные данные:** только русские, белорусские и казахские паспорта.
3. **СНИЛС:** уникальный, но привязан к клиенту, может повторяться при повторном посещении.
4. **Симптомы:** выборка минимум из 5000 симптомов, может быть комбинация итоговых симптомов (не более 10 штук).
5. **Выбор врача:** выборка минимум из 50 врачей.
6. **Дата посещения врача:** в рабочие время и дни недели, повторное посещение минимум через 24 часа после получения анализов.
7. **Анализы:** выборка минимум из 250 анализов, может быть комбинация итоговых анализов (не более 5 штук).
8. **Дата получения анализов:** в рабочие время и дни (через 24-72 часа).
9. **Стоимость анализов:** только в рублях.
10. **Карта оплаты:** максимальное количество повторов - 5 раз.
11. **Количество строк датасета**: минимум 50000.

**Теоретическая часть**

Для создания датасета использован программный модуль и набор csv таблиц:

1. main.py: Генерация всех необходимых данных с последующей их компановкой в соответствии с заданной структурой. Данные берутся из файлов names\_table.csv, medicine\_table.csv. Итог работы программы — файл dataset.csv
2. names\_table.csv — таблица, которая содержит данные для генерации ФИО посетителей поликлиники и врачей.
3. medicine\_table.csv — таблица, которая содержит данные для генерации специализации врача, симптомов, которые каждый тип врачей способен выявить, а также соответствующих им анализов и их стоимостей.

Кроме того, при организации процесса генерации следует помнить, что:

* Количество строк в датасете определяется вводом пользователя и начинается с 5000.
* Для генерации даты посещения врача необходимо учитывать, что в один день каждый доктор способен принять ограниченное количество посетителей.
* Дата получения анализов должна зависеть от даты посещения врача (быть в диапазоне 24-72 часа после)
* Симптомы не должны противоречить друг другу
* Веса банков и платежных систем определяются пользователем и должны в сумме давать положительное число