# Задание. Дамп устаревшей базы

Рассмотрим ситуацию: Необходимо экспортировать данные из устаревшего приложения, проверить на ошибки и выполнить анализ данных.

Так получилось, что все данные старой БД находятся в текстовом файле, каждая строка которого обозначает запись, данные в которой разделены запятыми. В процессе экспорта произошла ошибка, поэтому данные могут быть сохранены некорректно.

Структура строки:

|  |  |
| --- | --- |
| Номер записи | Целое число |
| Имя | Одно слово |
| Инициал отчества с точкой | Буква и точка |
| Фамилия | Одно слово |
| Пол | Муж или Жен |
| Город | Одно или несколько слов |
| Область | Двухсимвольный код |
| Адрес электронной почты | [www@dddd.eee](mailto:www@dddd.eee) |
| Телефон | 8,9 или 10-значный номер, разделенный дефисами |
| Дата рождения | дата в формате м/д/гггг |
| Должность | Одно или несколько слов |
| Компания | Одно или несколько слов |
| Вес | Дробное число |
| Рост | Целое число |
| Почтовый адрес | Целое число, пробел, одно или несколько слов |
| Почтовый индекс | Целое число |
| Код страны | Двухбуквенный код |

**Вариант 1**

* Предусмотреть проверку на ошибки. Подсчитать количество ошибок для каждого поля и вывести в отчете.
  1. Email: в адресе должны присутствовать только латинские буквы и цифры, один знак @, одна точка в имени домена. Некорректные символы в адресе нужно удалить.
  2. Пол: Допустимы значения Муж и Жен, все прочие нужно удалить.
* Выполнить преобразование данных в новый формат:
  1. Номер записи дополнить ведущими нулями до 6 знаков, пример: 000001, 000352, 009999, 010000.
  2. Телефон: В зависимости от количества цифр в номере привести к виду 1-3-4, 2-3-4 или 3-3-4, т.е. 1-234-5678, 12-345-6789.
  3. Вес: Привести к целому виду, округлив дробную часть: 60,1 -> 60; 60,7 -> 61
* Переписать данные в новый файл, используя в качестве разделителя точку с запятой (;)
* Подвести статистику по файлу:
  1. Найти количество женщин и мужчин, определить средний рост, вес и возраст для каждой категории. Определить количество женщин и мужчин, которые имеют рост, вес и возраст “ниже среднего”, “средний”, “выше среднего” для каждой характеристики, в рамках каждой группы.
  2. Вывести имена людей, родившихся в праздничные дни: 01.01, 07.01, 14.02, 23.02, 08.03, 01.05, 31.12 с группировкой по дате.
* Получить методом GET название области и вывести записи всех жителей этой области из списка. Имена женщин вывести розовым цветом, мужчин – голубым, вместо даты рождения указывать возраст на текущий момент времени, почтовый адрес выводить одной строкой. При выводе отсортировать по фамилиям в алфавитном порядке.

**Вариант 2**

* Предусмотреть проверку на ошибки. Подсчитать количество ошибок для каждого поля и вывести в отчете.
  1. Номер телефона должен содержать только цифры и знак дефиса “-“. Все прочие символы нужно удалить.
  2. Адрес: Допустимы цифры только в начале адреса, до первого пробела. Далее следуют символы латинского алфавита. Некорретные символы следует удалить, сохраняя пробел между цифрами и текстом.
* Выполнить преобразование данных в новый формат:
  1. Пол: Заменить male на 1, female на 0.
  2. Дата рождения: Перевести из формата м/д/гггг в дд.мм.гггг, т.е. 1/7/1950 (7 января 1950) к 07.01.1950, 8/18/1985 -> 18.08.1985.
  3. Адрес: Перенести номер дома в конец адреса, отделив запятой: 94 Республики -> Республики, 94.
* Переписать данные в новый файл, используя в качестве разделителя точку с запятой (;)
* Подвести статистику по файлу:
  1. Вывести имя, телефон и полный адрес самого пожилого и самого молодого человека.
  2. Найти количество клиентов каждого почтового сервера.
* Получить методом GET название области и вывести записи всех жителей этой области из списка. Имена женщин вывести розовым цветом, мужчин – голубым, вместо даты рождения указывать возраст на текущий момент времени, почтовый адрес выводить одной строкой. При выводе отсортировать по фамилиям в алфавитном порядке.