МИНОБРАНУКИ РОССИИ

федеральное государственное бюджетное

образовательное учреждение высшего образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт информационных технологий

*наименование института (факультета)*

Математическое и программное обеспечение ЭВМ

*наименование кафедры*

Основы Data Science

*наименование дисциплины в соответствии с учебным планом*

ЛАБОРАТОРНАЯ РАБОТА №1

Тема: Описание структуры данных

|  |  |
| --- | --- |
| Руководитель | Юдина О. В. |
|  | ФИО преподавателя |
| Исполнитель |  |
| студент | 1ПИб-02-1оп-22 |
|  | группа |
|  | Рябкова Анна |
|  | Леонидовна |
|  | Фамилия, имя, отчество |
| Оценка |  |
| Подпись |  |

2024 год

**Цель:** Обзор основных стандартов работы с данными. Знакомство с этапом понимание данных стандарта CRISP-DM.

**Задача:** Понять, что именно содержится в этих данных, какие факты они описывают, понять, что они могут дать.

**Задание:**

1. Откройте файл с данными. Приведите его к формату, подходящему для обработки. Он содержит данные о студентах одного из учебных заведений.
2. Выясните структуру данных. Для этого:
3. Для каждой переменной определите возможные шкалы, выберите шкалу, которая даст возможность получить больше информации. (Для категориальных переменных укажите – номинальными или порядковыми они являются. Для количественных уточните, по какой именно из числовых шкал они могут быть измерены.)
4. Используя Excel, постройте сводные таблицы для категориальных переменных и таблицы частот для количественных
5. Постройте диаграммы, которые иллюстрируют структуру , для качественных переменных (4-6 штук, с заголовками, подписями осей и пояснениями, что они показывают)
6. Для количественных переменных постройте гистограммы и графики

(кумулятивный, частотный, простой, диаграмма рассеяния - какие будут уместны) которые иллюстрируют распределения (4-6 штук, с заголовками, подписями осей и пояснениями, что они показывают)

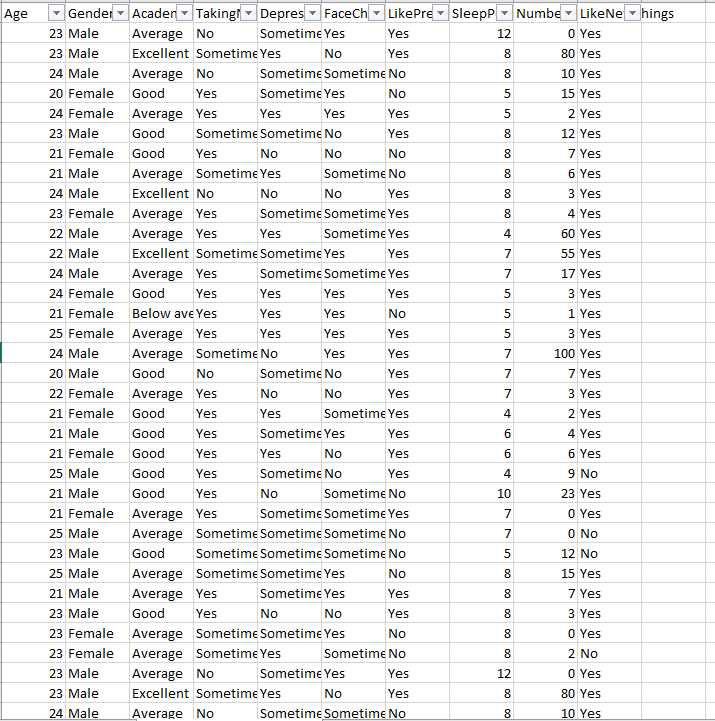
1. Сформулируйте, какие задачи могут решаться с применением этих данных. Приведите несколько формулировок

Основная часть

1. Приведем исходный файл читаемому формату

****

**Рис. 1. Исходный файл**

****

**Рис. 2. Отредактированный файл**

1. **Шкалы**

**Качественные - а) Номинальные – «Гендер», «Нравятся новые вещи», «Нравятся презентации», «Делает заметки», «Статус депрессии», «Столкнетесь с проблемами выполнения академ.задания»;**

**Б) Порядковые – «Академическая успеваемость»;**

**Количественные - а) Интервальный – «Возраст»;**

**Б) Относительные – «Кол-во часов сна в день», «Кол-во друзей»;**

1. **Сводные таблицы**

