МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное

образовательное учреждение высшего образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт информационных технологий

*наименование института(факультета)*

Кафедра математического и программного обеспечения ЭВМ

*наименование кафедры*

Основы Data Science

*наименование дисциплины в соответствии с учебным планом*

ЛАБОРАТОРНАЯ РАБОТА №1

ОПИСАНИЕ СТРУКТУРЫ ДАННЫХ

|  |  |
| --- | --- |
| Исполнитель |  |
| Студент | 1ПИб-02-2оп-22 |
|  | *группа* |
|  | Зернов В.А. |
|  | *ФИО* |
| Руководитель | Юдина О.В. |
|  | *ФИО преподавателя* |
| Оценка |  |
| Подпись |  |

2024 год

Цель:

Обзор основных стандартов работы с данными. Знакомство с этапом понимание данных стандарта CRISP-DM.  
Задача этапа – понять, что именно содержится в этих данных, какие факты они  
описывают, понять, что они могут дать.

Задания

Откройте файл с данными. Приведите его к формату, подходящему для обработки. Он содержит данные о студентах одного из учебных заведений.

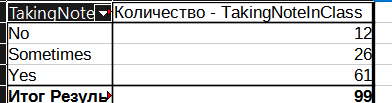
1. Выясните структуру данных. Для этого:
2. Для каждой переменной определите возможные шкалы, выберите шкалу, которая даст возможность получить больше информации.  
   Для категориальных переменных укажите – номинальными или порядковыми они являются. Для количественных уточните, по какой именно из числовых шкал они могут быть измерены.
3. Используя Excel, постройте сводные таблицы для категориальных переменных и таблицы частот для количественных
4. Постройте диаграммы, которые иллюстрируют структуру, для качественных переменных (4-6 штук, с заголовками, подписями осей и пояснениями, что они показывают)
5. 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)

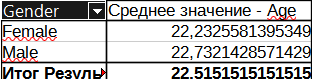
Сформулируйте, какие задачи могут решаться с применением этих данных. Приведите несколько формулировок.

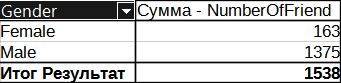
Практическая часть

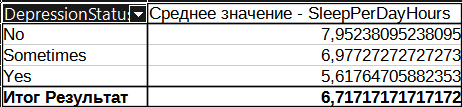
|  |  |
| --- | --- |
| **Наименование переменной** | **Шкала** |
| Age | Количественный, интервальный |
| Gender | Качественный, номинальный |
| AcademicPerformance | Качественный, порядковый |
| TakingNoteInClass | Качественный, номинальный |
| DepressionStatus | Качественный, номинальный |
| FaceChallangesToCompleteAcademicTask | Качественный, номинальный |
| LikePresentation | Качественный, номинальный |
| SleepPerDayHours | Количественный, относительный |
| NumberOfFriend | Количественный, относительный |
| LikeNewThings | Качественный, номинальный |

Сводные таблицы для категориальных переменных

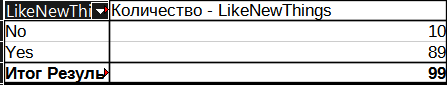












Диаграммы для качественных переменных

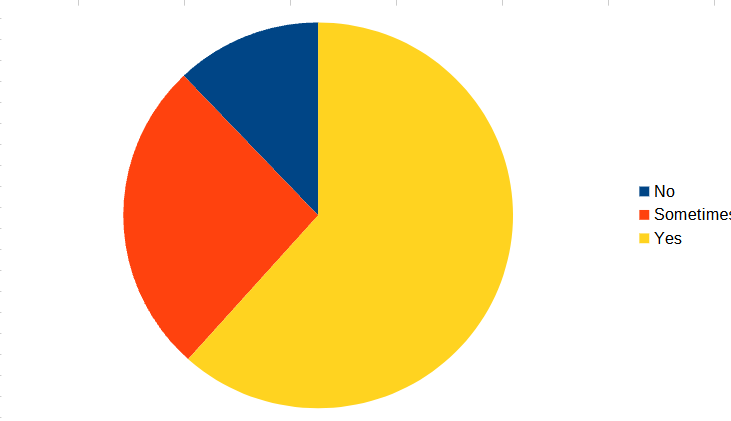


Диаграмма показывает кол-во сколько людей делают записи, иногда делают записи и не делают записи. Видно, что в основном люди записывают.

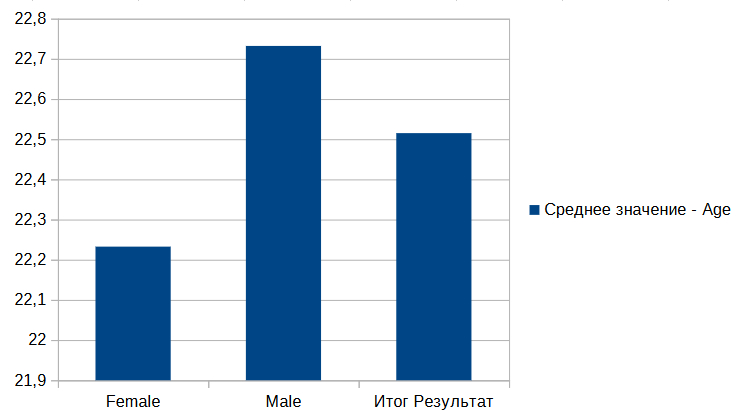


Диаграмма показывает средний возраст среди мужчин и женщин. Видно, что в основном возраст у обоих полов одинаковый, но у мужчин немного больше.

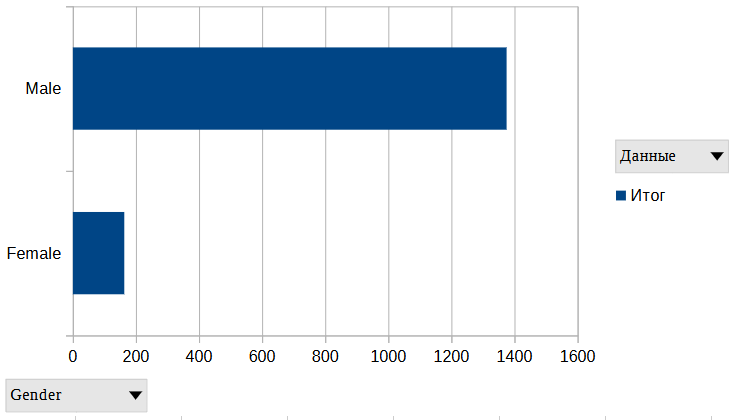


Диаграмма показывает общее кол-во друзей у мужчин и женщин. Видно, что у мужчин друзей больше, чем у женщин.

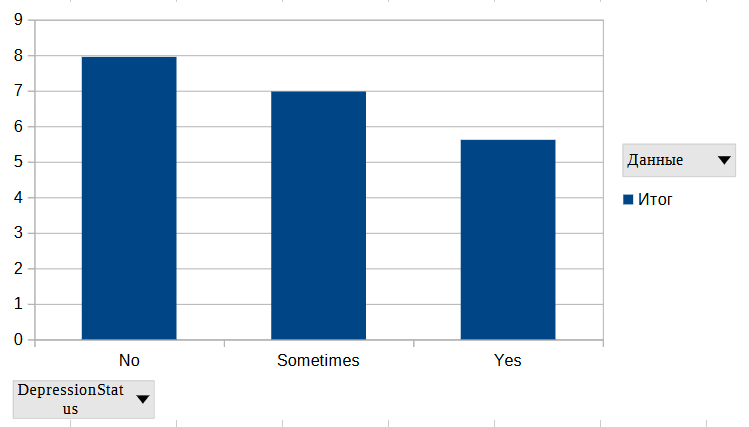


Диаграмма показывает среднее часы сна и в зависимости от статуса депрессии. Видно, что в среднем те, у кого депрессии нет спят около 8 часов, а те, у кого есть депрессия спят около 5 часов.

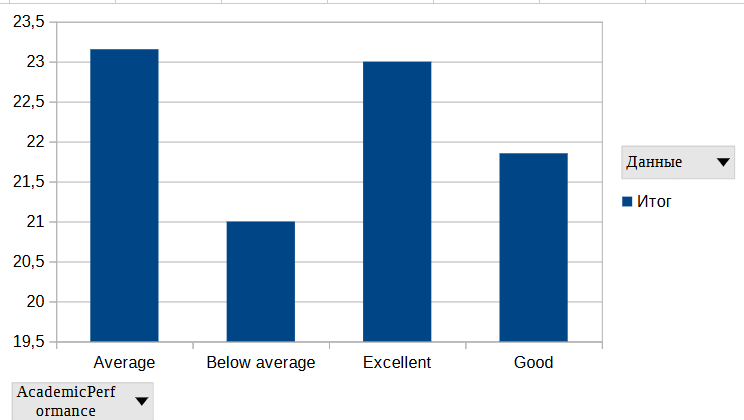


Диаграмма показывает зависимость возраста от успеваемости в учебе. Видно, что учеба дается тем, кто старше.

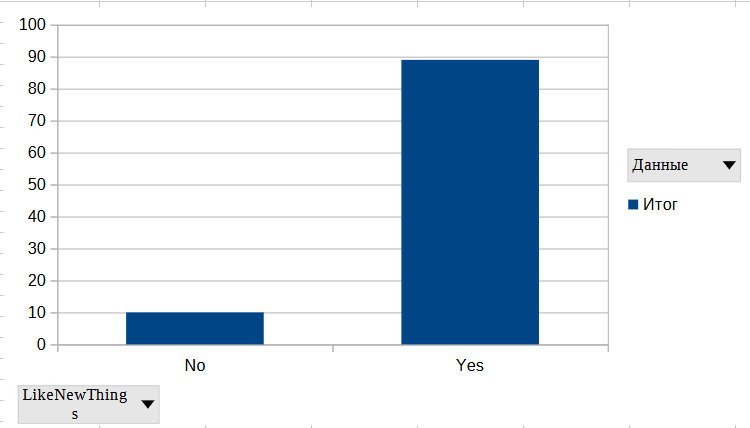
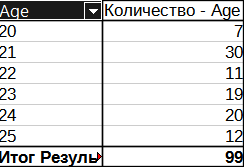
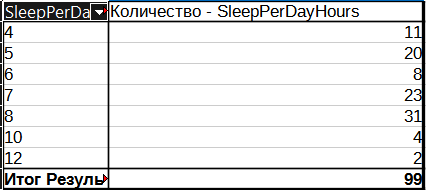


Диаграмма показывает кол-во людей, которым нравится изучать новое.

Таблицы частот для количественных переменных





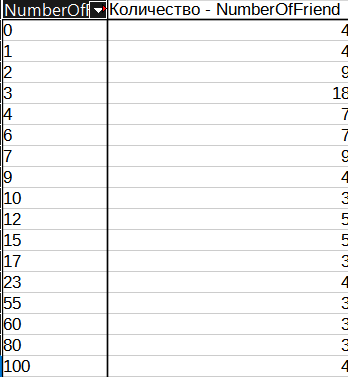
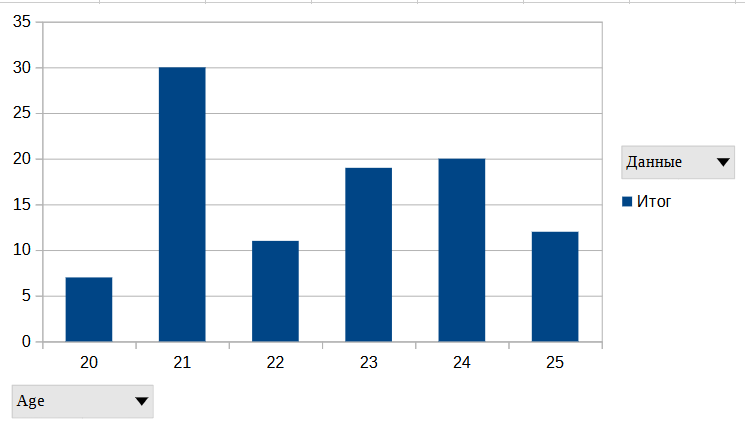
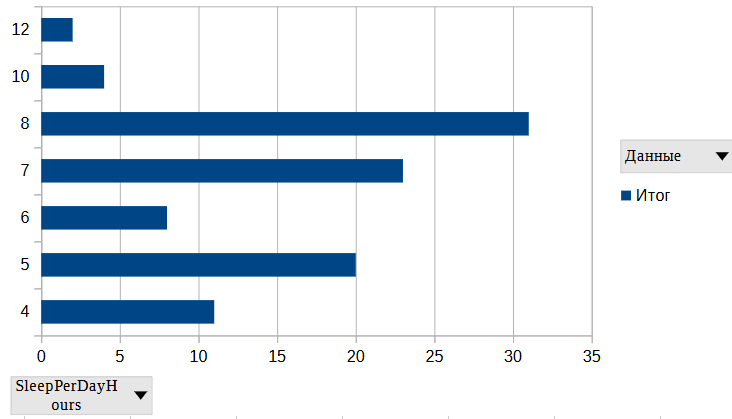


Диаграмма для количественных переменных:

Диаграмма показывает кол-во возрастов. Видно, что людей с возрастом 21 больше всего.

Диаграмма показывает сколько людей спят по часам в диапазоне от 4 до 12 часов. Видно, что люди спят в среднем по 8 часов.

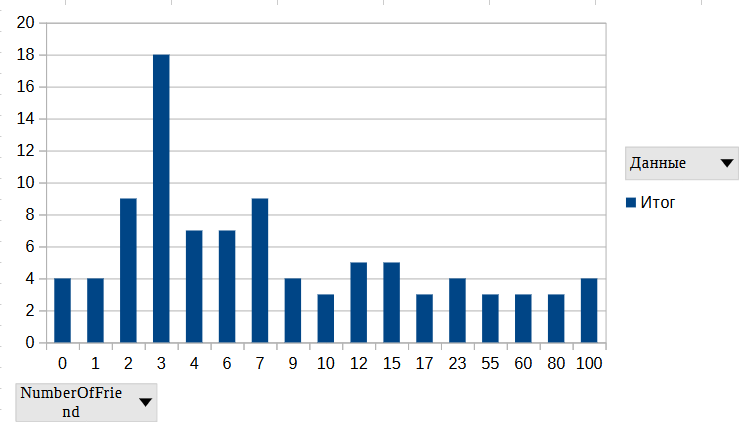


Диаграмма показывает сколько у людей друзей. Видно, что в среднем у людей в основном у каждого три друга и не больше.

6) Задачи, возможные с представленными данными

Влияние уровня депрессии на успеваемость:

Изучить связь между уровнем депрессии и академическими показателями.

Оценить, насколько часто студенты с высоким уровнем депрессии имеют низкие оценки.

Анализ факторов, влияющих на выполнение учебных задач:

Изучить, как студенты справляются с выполнением учебных задач в зависимости от различных факторов, таких как возраст, пол и уровень депрессии.

Оценить, какие факторы чаще всего затрудняют выполнение учебных задач.

Оценка влияния количества часов сна на успеваемость:

Исследовать связь между количеством часов сна и академической успеваемостью.

Оценить, как количество часов сна влияет на оценки/успех в учебе.

Вывод

В результате детального анализа данных о студентах, проведенного с использованием различных методик и инструментов, включая сводные таблицы и графики в Excel, были получены ценные сведения о характеристиках студентов и факторах, влияющих на их успеваемость. Эти данные могут быть применены для оптимизации учебного процесса и поддержки студентов.