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-02-1оп-22 |
|  | *группа* |
|  | Харламов Д.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.

**Задание:**

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. Информация о шкалах приведена ниже

3. Используя Excel, постройте сводные таблицы для категориальных переменных и таблицы частот для количественных

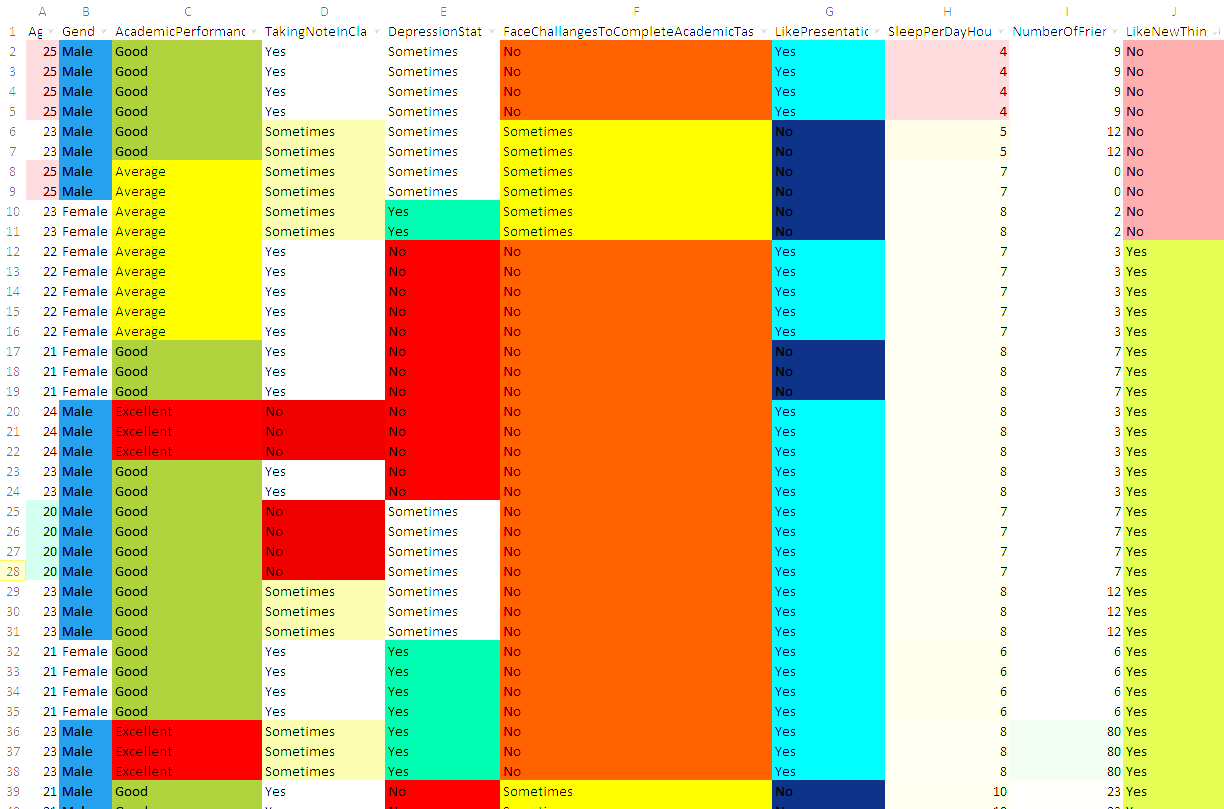
4. Постройте диаграммы, которые иллюстрируют структуру , для качественных переменных (4-6 штук, с заголовками, подписями осей и пояснениями, что они показывают)

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)

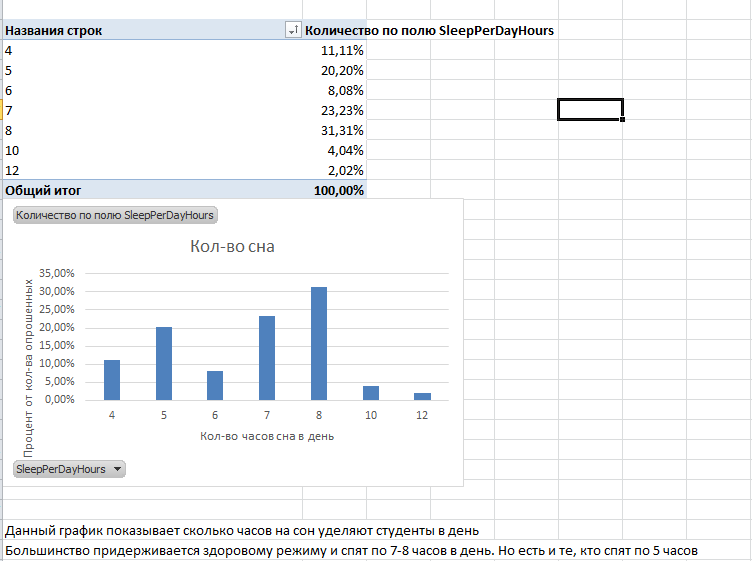
6. Сформулируйте, какие задачи могут решаться с применением этих данных. Приведите несколько формулировок

Ход работы

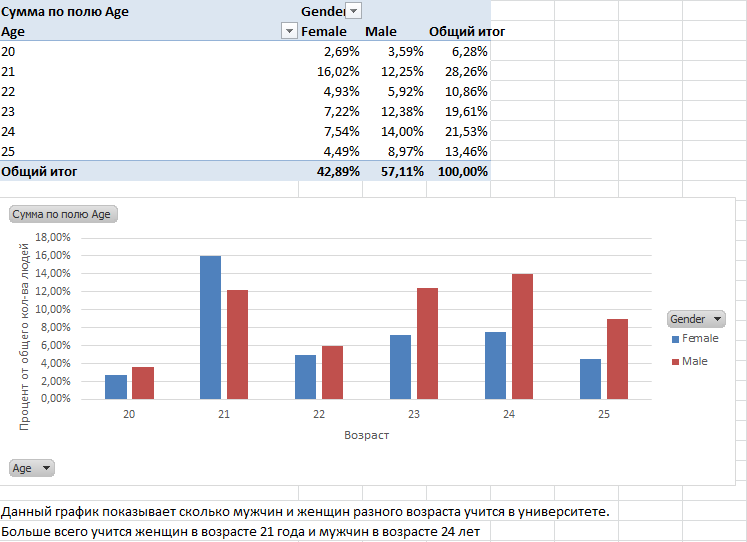
Чтобы работать с исходной таблицей было комфортно и удобно, было решено изменить форматирование и визуализировать все признаки цветовыми шкалами



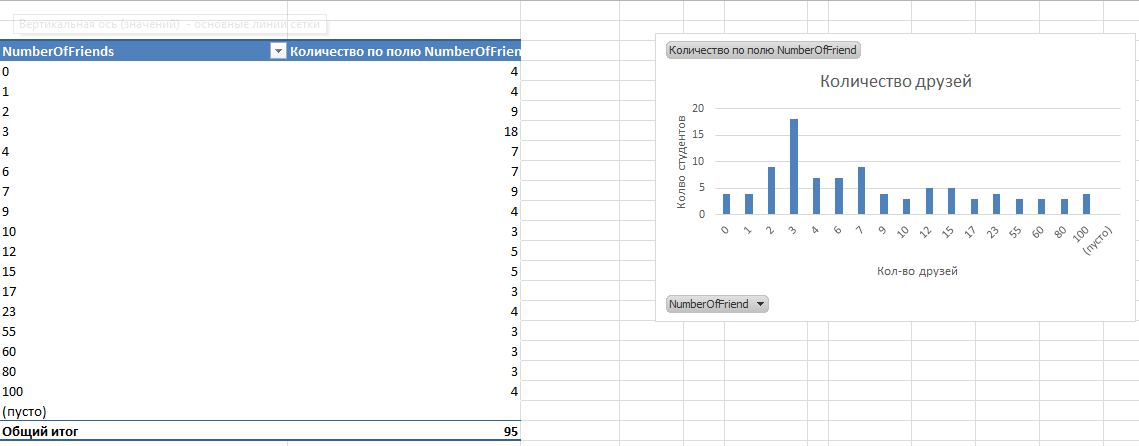
Для добавления сводных таблиц для качественных и для количественных переменных) создается лист. В них и будет происходить аналитическая работа. Из количественных признаков были взяты все три, а именно возраст ученика, его количества сна/день, количество друзей. На каждую из сводных таблиц была сформирована гистограмма с группировкой.



Данный график показывает сколько часов на сон уделяют студенты в день. Большинство придерживается здоровому режиму и спят по 7-8 часов в день. Но есть и те, кто спят по 5 часов.

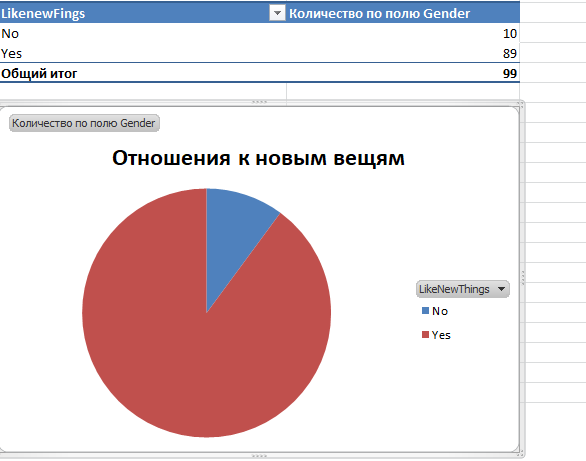


Данный график показывает сколько мужчин и женщин разного возраста учится в университете. Больше всего учится женщин в возрасте 21 года и мужчин в возрасте 24 лет

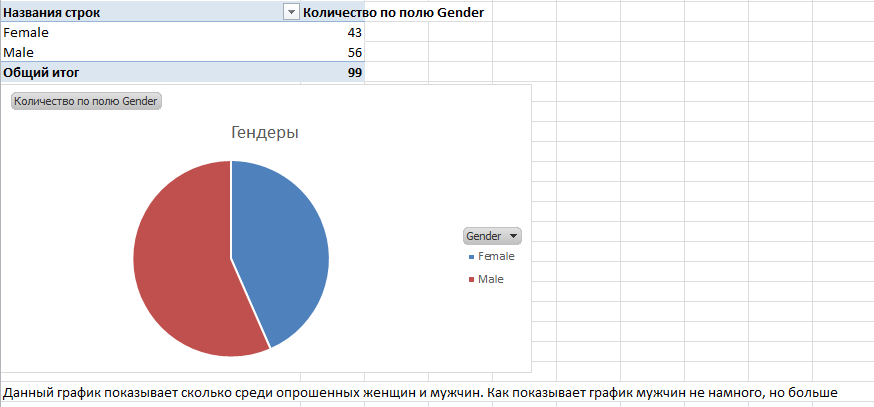


Данный график показывает, сколько у студентов друзей. В основном у студентов по 3 друга.

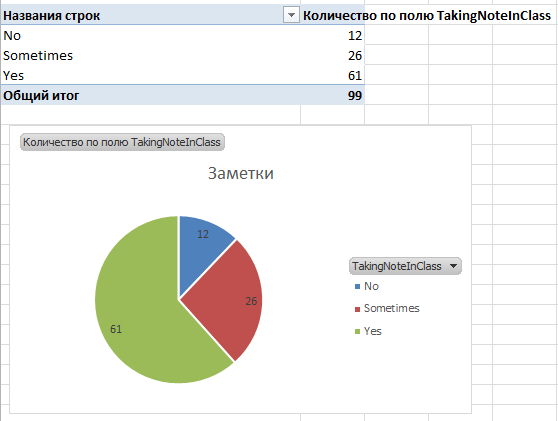
Повторим действия для качественных признаков



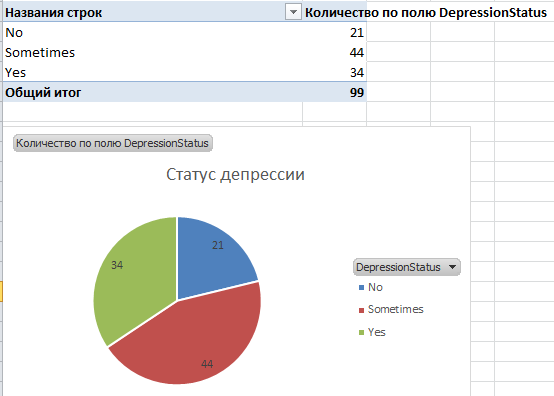
Данная диаграмма показывает насколько опрошенным нравятся новые вещи, готовы ли они приступать к новым задачам. Как показывает график, большинство за новые вещи, это значит, что они в основном адаптивны и гибки к изменениям.



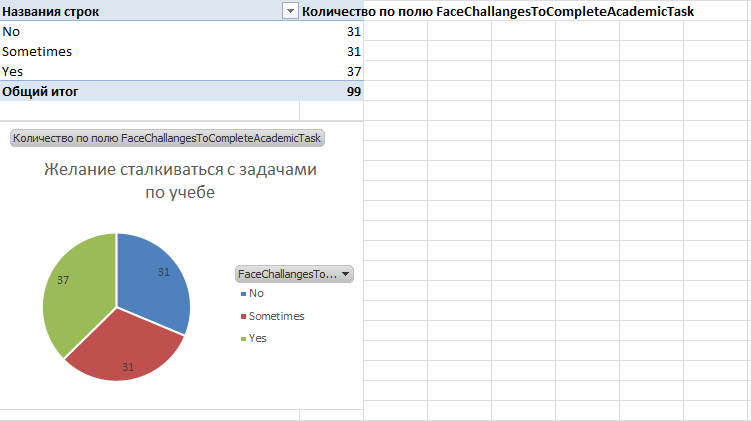
Данный график показывает, сколько среди опрошенных женщин и мужчин. Как показывает график мужчин не намного, но больше



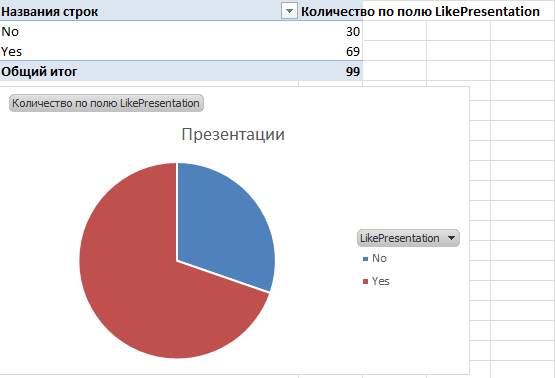
Данный график показывает делают ли опрошенные заметки в аудитории, большинство делает, но есть десятая часть, которая совсем их не делает.



Данный график показывает, какой статус депрессии у опрошенных. Большинство иногда подвержено депрессии, на втором месте те, кто постоянно находятся в депрессии.



Данный график показывает, готовы ли опрошенные сталкиваться лицом к лицу с задачами во время учебы. Все варианты примерно равны по кол-ву ответов, но все же большее количество всегда готово сталкиваться с трудностями. И пополам люди готовы иногда сталкиваться с задачами или не готовы вовсе.



Данный график показывает нравится ли людям делать презентации или нет. Большинство любят представлять графически информацию, но есть и те, кто больше предпочитает устные рассказы