МИНОБРНАУКИ РОССИИ  
ФГБОУ ВО Череповецкий государственный университет

Институт информационных технологий

Кафедра: МПО ЭВМ

Дисциплина: Основы Data science

ЛАБОРАТОРНАЯ РАБОТА № 1

Тема: Описание структуры данных

Выполнил:  
студент гр. «1ПИб-02-02-оп-22» Тихомиров В.В.

Проверил:

Юдина Ольга Вадимовна

Череповец, 2024 г.

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. 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

Для более комфортной работы с исходной таблицей было принято решение изменить форматирование и визуализировать все признаки цветовыми шкалами



Далее необходимо добавить сводные таблицы для качественных и количественных переменных. Для этого создадим новый лист, где будет происходить аналитическая работа. Из количественных признаков были взяты все три, а именно возраст, кол-во сна и кол-во друзей. Для каждой таблицы была сформирована гистограмма с группировкой.

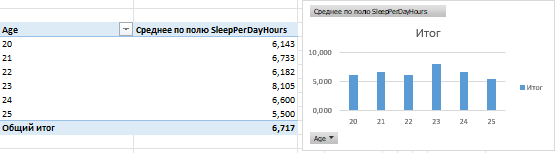


График показывает сколько часов опрошенные тратят на сон в день. И исходя из графика можно сделать вывод, что большинство студентов соблюдают здоровый режим сна и спят по 7-8 часов, но встречаются и студенты, которые спят и по 5 часов.

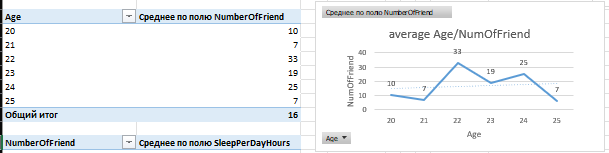
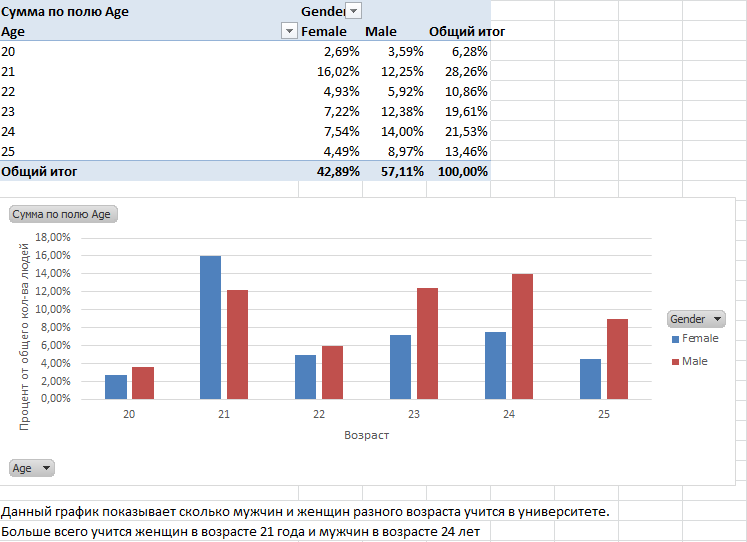
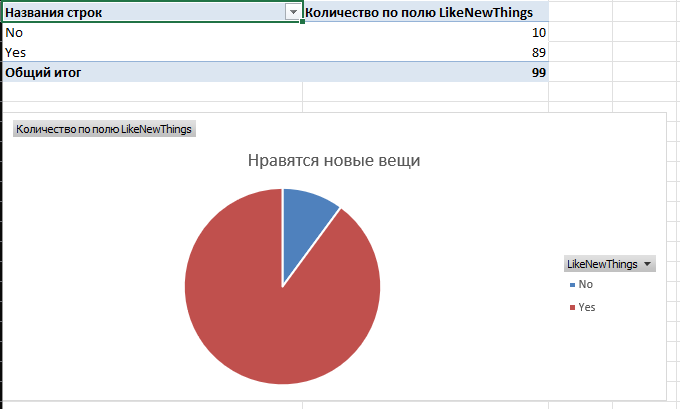


График показывает сколько в среднем друзей у опрошенных разных возрастов. Смотря на график можно сделать вывод, что в возрасте 22 года у опрошенных больше всего друзей, у ребят, которым 21 и 25 соответственно меньше всего друзей. Золотую середину забирают опрошенные, которым 23 года.

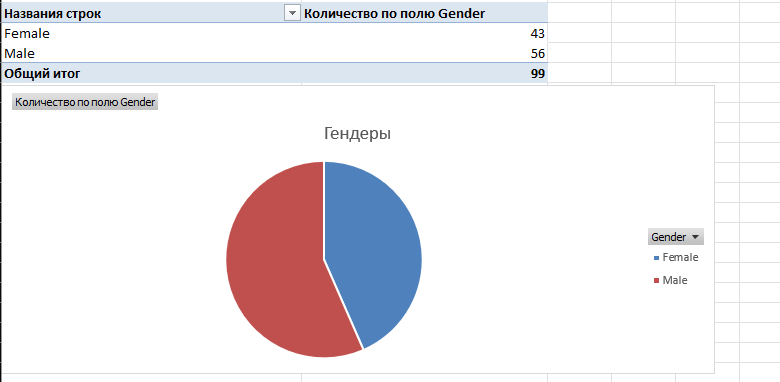


Данный график показывает сколько мужчин и женщин разного возраста учатся в университете. Больше всего учится женщин в возрасте 21 года и мужчин в возрасте 24 года.

Теперь сделаем тоже самое для качественных признаков



Данная диаграмма показывает насколько опрошенным нравятся новые вещи, готовы ли они приступать к новым задачам. Как показывает нам график, большинство людей за что-то новое, тоесть они, в основном, адаптивны и гибки к изменениям.



Этот график показывает сколько среди опрошенных мужчин и женщин. Оказалось, что женщин не сильно меньше, чем мужчин, но всё таки меньше.

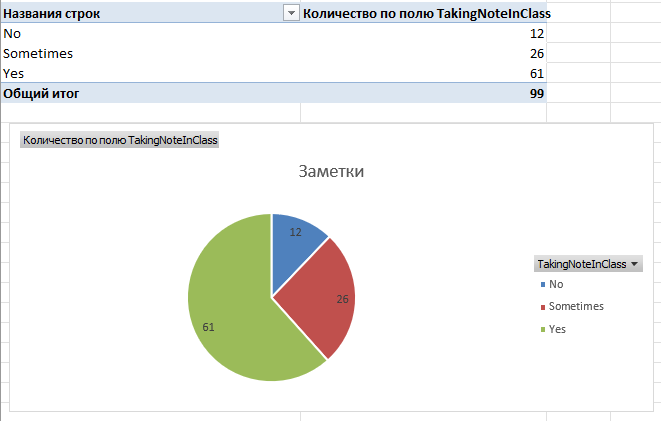


График показывает делают ли опрошенные заметки в аудитории, и исходя из данных можно сделать вывод, что большинство их делает, хотя всё таки есть небольшая часть студентов, которые их всё таки не делают (примерно десятая часть).

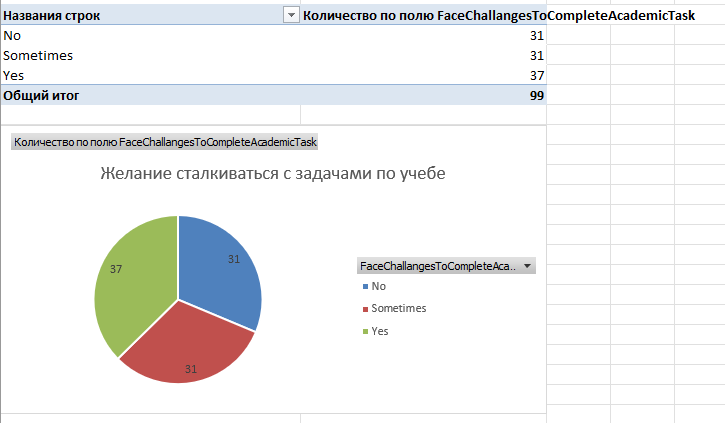
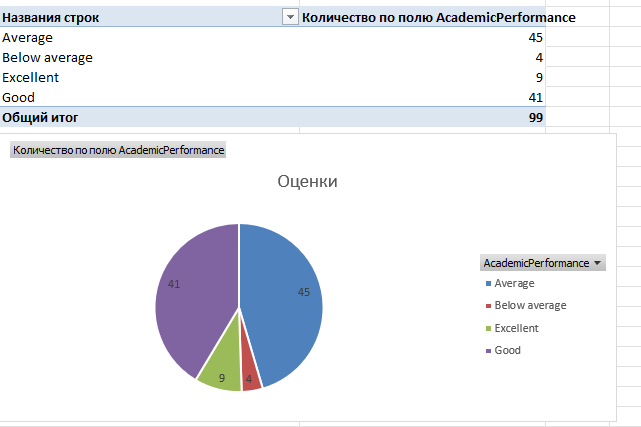


График показывает, готовы ли опрошенные сталкиваются с задачами во время учёбы. Все варианты примерно равны по кол-ву, однако студентов, которые готовы столкнуться с трудностями всё же больше, остальных поровну.



Данный график показывает на какие оценки учатся опрошенные. По данным оттуда можно понять, что большинство студентов учится либо на хорошие, либо на средние оценки, однако встречаются студенты у которых оценки либо превосходные, либо ниже среднего.