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 №5

Тема: Визуализация данных

|  |  |
| --- | --- |
| Руководитель | Юдина О. В. |
|  | ФИО преподавателя |
| Исполнитель |  |
| студент | 1ПИб-02-1оп-22 |
|  | группа |
|  | Харламов Денис |
|  | Алексеевич |
|  | Фамилия, имя, отчество |
| Оценка |  |
| Подпись |  |

Череповец, 2024 г.

**Цель:**

Знакомство с этапом понимание данных стандарта CRISP-DM.

Задача этапа – представить набор данных, использованный в лабораторных работах. Описать основные закономерности, которые в них содержатся, связи между этими данными, фактами, знаниями, получить что-то новое, понять, что они могут дать и визуализировать свои представления.

**Задание:**

Вам нужно представить результаты изучения этих данных руководству предприятия, чтобы решить в каком направлении строить дальнейшую кадровую политику или проводить исследование.

Представьте в диаграммах имеющиеся данные, на основании которых возможно построить исследование.

Вам нужно показать:

* Закономерности
* Структуры
* Возможные связи
* Влияние факторов друг на друга
* Если вам удалось построить модели – покажите их

Требования:

* Диаграммы должны иметь подписи по осям и заголовки, отражающие их суть
* Должны быть разными
* Их тип должен быть пригодным для представления данных
* Они должны отражать ваши данные, их связи, закономерности, которые вам удалось обнаружить.
* Для построения можно использовать любые пригодные программы – Excell, если вам удалось найти доступные программы для визуализации данных, то можно использовать их.

Ход работы.

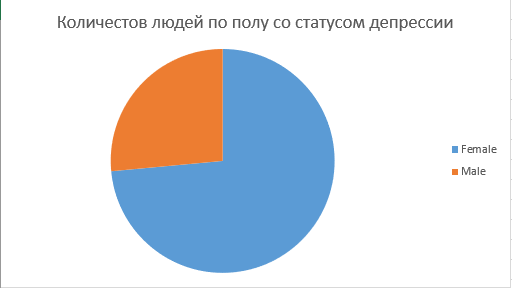
В начале исследования я ознакомилась с предложенными для анализа данными, после чего был сделан ряд предположений, которые проверялись с помощью различных тестов. Тема: влияние депрессии на студентов.

В начале проанализируем долю людей с депрессией.



Получается, что большая часть студентов иногда подвергаются с депрессией. Так же выяснилось, что большая часть из остальных уже находится в депрессии.

Далее проведем исследование на самый депрессивный пол.



На круговой диаграмме показано, что женщины гораздо более склонны к депрессии.

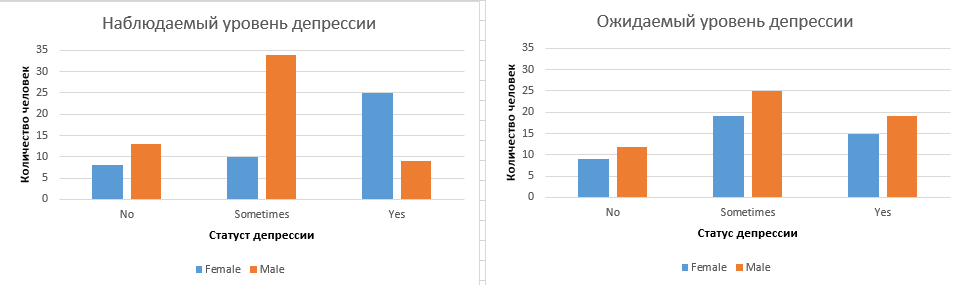
Выяснить – существует ли различие между уровнем депрессии у мужчин и женщин.

H0: не существует никакой связи между полом и наличием депрессии.

Н1: существует связь между полом и наличием депрессии.

Уровень значимости – 5%

С помощью таблицы слагаемых Хи квадрата была найдена сумма проверочной статистики = 20,45. Критическое значение = 5,99. Следовательно нулевая гипотеза отклоняется. Женский пол страдает депрессией чаще, чем мужской.



Из всего выше можно сделать вывод, что женский пол более подвержен депрессии, чем мужской.

Было выяснено, что Доля людей с депрессией готовых выполнять задачи отличается от людей без депрессии.

Гистограмма показывает, что люди с депрессией чаще готовы выполнять задачи по учебе, чем люди без депрессии.

Для более углублённого исследования взаимосвязи между депрессией и обучением было проведено ещё несколько исследований.

* 1. Влияние депрессии на записи конспектов.

На гистограмме наглядно видно, что люди с депрессией делают записи конспектов чаще, чем люди без неё.

* 1. Следующим шагом проверялось, как депрессия влияет на любовь к новым вещам.

На данных круговых диаграммах показано, что все люди без депрессии любят новые вещи и, что не всем людям с депрессией нравятся новые вещи.

* 1. Следующим шагом проверялось, как депрессия влияет на отношение к презентациям.

По данным круговым диаграммам видно, что депрессия не влияет на любовь к презентациям у студентов.

Далее было проведено исследование зависимости количества сна от статуса депрессии.

Нулевая гипотеза заключается в том, что все тратят на сон одинаковое количество времени. Альтернативная гипотеза - люди с депрессией спят меньше, чем люди, у которых нет депрессии. Выбрано испытание по выборочному среднему – одна выборка. Испытание было проведено на 5% уровне значимости.

Х = 5,6 - среднему количеству часов сна у людей с депрессией

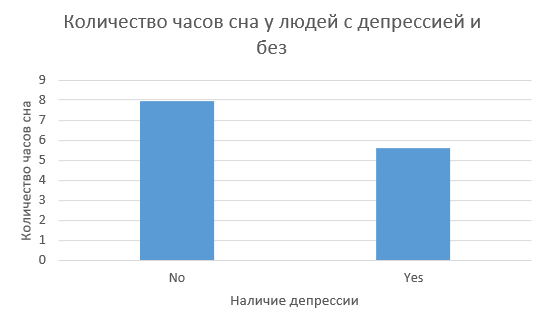
U = 8 – среднему количеству часов сна у людей без депрессии

S = 2,18 – дисперсия

t-статистика = -6,4, рассчитана по формуле (x-u)/(S/√33), где 33 это количество наблюдений – 1

Критическое значение было найдено по таблице Стьюдента = 2

В результате испытания была отвергнута нулевая гипотеза, так как t-статистика (6,4) > критического значения (2).



На рисунке представлена диаграмма, с помощью которой наглядно можно увидеть количество часов сна у людей с депрессией и без.

Далее производим исследование зависимости всех количественных переменных (возраст, количество сна и кол-во друзей) от статуса депрессии.

1) Сначала у депрессивных людей

У людей с депрессией была выявлена слабая зависимость количества друзей со сном и отсутствие зависимости между возрастом и сном. Значит, чем больше человек с депрессией спит, тем больше у него друзей.

2) У людей без депрессии

Была выявлена слабая отрицательная зависимость между количеством друзей и сном у людей с депрессией. Значит, чем больше человек без депрессии спит, тем меньше у него друзей.

Была выявлена средняя отрицательная зависимость от сна и возраста. Значит, чем больше у человека без депрессии возраст, тем меньше он спит

Далее рассматривался статус депрессии, как фактор влияния на обучение.

Ранее было доказано, что между успеваемостью и депрессией есть связь. Необходимо провести исследование для выявления связи между депрессией и готовностью выполнения академических задач.

Был проведён ХИ тест о наличии связи между успеваемостью и депрессией.

Были выдвинуты следующие гипотезы:

H0: между успеваемостью и депрессией есть связь.

H1: между успеваемостью и депрессией связи нет.

В результате исследования была принята нулевая гипотеза, что значит, между успеваемостью и депрессией есть связь.

Вывод

Исследование было проведено в соответствии с методологией CRISP-DM с целью анализа данных, выявления новых закономерностей и извлечения новых знаний.

Основной задачей исследования было выявление связей между различными наборами данных и визуализация полученных результатов для дальнейшего анализа. Для достижения этой цели мы визуализировали данные, полученные из предыдущих исследований, чтобы извлечь ценную информацию из них.

В ходе исследования были выявлены следующие закономерности:

Женский пол более подвержен депрессии. Депрессия влияет на многие факторы обучения, такие как успеваемость, готовность к задачам, записи конспектов, любовь к новым вещам. Так же выявлено, что в зависимости от статуса депрессии, она влияет на взаимосвязь сна ученика и количества его друзей.