МИНОБРНАУКИ РОССИИ

ФГБОУ ВО Череповецкий государственный университет

Институт информационных технологий

Кафедра: МПО ЭВМ

Дисциплина: Основы Data Science

ЛАБОРАТОРНАЯ РАБОТА № 5

Тема: Визуализация данных

Выполнила:

студентка гр. 1ПИб-02-1оп-22 Харламов Д.А.

Проверил: доцент, Юдина Ольга Вадимовна

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.

Было принято решение о исследовании влияния времени сна.

Исследование было начато с выявления доли людей, основного времени сна студентов.

* 1. В среднем студенты спят по 7-8 часов в день
  2. В среднем студенты мужского пола спят больше, чем девушки.
  3. Люди с депрессией спят меньше, чем люди у которых депрессии нет

Далее проводились испытания гипотез. Для начала необходимо было сформулировать гипотезы.

Гипотеза №1:

Предположим, что студенты, спящие ровно 7 часов, реже сталкиваются с ошибками:

Н0: Студенты, которые тратят ровно 7 часов в неделю на сон, реже сталкиваются с проблемами при выполнении академических заданий.

Н1: Студенты, которые тратят более 7 часов в неделю на сон, реже сталкиваются с проблемами при выполнении академических заданий.



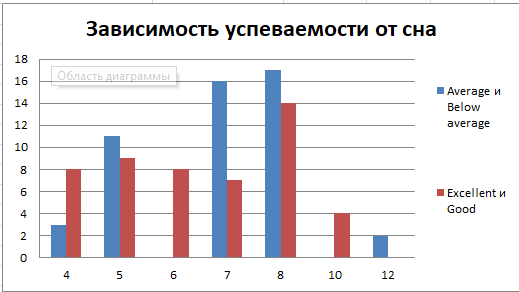
Исходя из исследований (t-статистика > t-критического), гипотеза Н0 отвергается.

Значит, студенты, спящие более 7 часов в день реже сталкиваются с проблемами.

Далее проверим зависимость успеваемости от времени сна студента.

Н0: между успеваемостью и временем сна есть связь

Н1: между успеваемостью и временем сна нет связи



Исходя из исследований, гипотеза Н0 принимается. Значит между сном и успеваемостью есть связь.

Далее было проведено исследование зависимости количественных переменных (возраст и кол-во друзей) от времени сна.

По данным можно сделать вывод, что время сна не зависит от возраста студента.

Далее по данным представленным сверху получается, что выявлена слабая зависимость количества друзей у студента от времени сна.

Далее рассматриваем время сна, как фактор влияния на дополнительные аспекты обучения.

Сначала проверим влияние времени сна на проблемы возникающие во время учебы.

Гистограмма показывает, что студенты, уделяющие мало времени сну, чаще сталкиваются с трудностями в учебе.

Далее проверим влияние сна на записи конспектов студентами.

Гистограмма показывает, что студенты, уделяющие мало времени сну, чаще ведут конспекты.

После чего проверим влияние сна на предпочтения подачи материала студенту.

На круговых диаграммах видно, что людям, которые спят меньше больше направятся презентации, нежили традиционные лекции.

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:

1. Студенты спят в основном 7-8 часов в день.
2. Время сна напрямую влияет на успеваемость студента.
3. Сон влияет на дополнительные факторы при обучении, такие как ведение конспектов или предпочтения презентации, обычным лекциям.
4. Существует слабая взаимосвязь времени сна и количества друзей у студента.