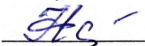


Министерство образования и науки Республики Башкортостан
ГБПОУ Октябрьский многопрофильный профессиональный колледж

Утверждено
на заседании МС
протокол № 1
от 30.08.2024 г.

Рассмотрено и рекомендовано к утверждению
на заседании ПЦК преподавателей
общеобразовательных дисциплин,
воспитателей
протокол № 1 от «30» августа 2024 г.
Председатель ПЦК  Фаттахова Н.Г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
ПО УЧЕБНОЙ ДИСЦИПЛИНЕ ОДП.08 ИНФОРМАТИКА

ОСНОВНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ
ПРОГРАММЫ ПОДГОТОВКИ КВАЛИФИЦИРОВАННЫХ РАБОЧИХ (СЛУЖАЩИХ)
ПО ПРОФЕССИИ СРЕДНЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
09.01.03 ОПЕРАТОР ИНФОРМАЦИОННЫХ СИСТЕМ И РЕСУРСОВ

Разработала преподаватель:

Исаргакова И.Н

2024г.

СОДЕРЖАНИЕ

1. Результаты обучения, регламентированные ФГОС СОО с учетом ФГОС СПО.....	3
2. Модельные примеры оценочных средств для проведения промежуточного и рубежного контроля по основному содержанию	6
3. Модельные примеры оценочных средств для проведения промежуточного и рубежного контроля по прикладному модулю 1 «Основы аналитики и визуализации данных»	34
4. Модельные примеры оценочных средств для проведения промежуточного и рубежного контроля по прикладному модулю 2 «Аналитика и визуализация данных на Python».....	37

1. Результаты обучения, регламентированные ФГОС СОО с учетом ФГОС СПО

Результаты обучения должны быть ориентированы на получение компетенций для последующей профессиональной деятельности как в рамках данной предметной области, так и в смежных с ней областях. Они включают в себя результаты освоения общеобразовательной дисциплины в соответствии с ФГОС СПО и на основе ФГОС СОО:

Базовый уровень

Б1 владение представлениями о роли информации и связанных с ней процессов в природе, технике и обществе; понятиями «информация», «информационный процесс», «система», «компоненты системы» «системный эффект», «информационная система», «система управления»; владение методами поиска информации в сети Интернет; умение критически оценивать информацию, полученную из сети Интернет; умение характеризовать большие данные, приводить примеры источников их получения и направления использования;

Б2 понимание основных принципов устройства и функционирования современных стационарных и мобильных компьютеров; тенденций развития компьютерных технологий; владение навыками работы с операционными системами и основными видами программного обеспечения для решения учебных задач по выбранной специализации;

Б3 наличие представлений о компьютерных сетях и их роли в современном мире; об общих принципах разработки и функционирования интернет-приложений;

Б4 понимание угроз информационной безопасности, использование методов и средств противодействия этим угрозам, соблюдение мер безопасности, предотвращающих незаконное распространение персональных данных; соблюдение требований техники безопасности и гигиены при работе с компьютерами и другими компонентами цифрового окружения; понимание правовых основ использования компьютерных программ, баз данных и работы в сети Интернет;

Б5 понимание основных принципов дискретизации различных видов информации; умение определять информационный объем текстовых, графических и звуковых данных при заданных параметрах дискретизации;

Б6 умение строить неравномерные коды, допускающие однозначное декодирование сообщений (префиксные коды); использовать простейшие коды, которые позволяют обнаруживать и исправлять ошибки при передаче данных;

Б7 владение теоретическим аппаратом, позволяющим осуществлять представление заданного натурального числа в различных системах счисления; выполнять преобразования логических выражений, используя законы алгебры логики; определять кратчайший путь во взвешенном графе и количество путей между вершинами ориентированного ациклического графа;

Б8 умение читать и понимать программы, реализующие несложные алгоритмы обработки числовых и текстовых данных (в том числе массивов и символьных строк) на выбранном для изучения универсальном языке программирования высокого уровня (Паскаль, Python, Java, C++, C#); анализировать алгоритмы с использованием таблиц трассировки; определять без использования компьютера результаты выполнения несложных программ, включающих циклы, ветвления и подпрограммы, при заданных исходных данных; модифицировать готовые программы для решения новых задач, использовать их в своих программах в качестве подпрограмм (процедур, функций);

Б9 умение реализовать этапы решения задач на компьютере; умение реализовывать на выбранном для изучения языке программирования высокого уровня (Паскаль, Python, Java, C++, C#) типовые алгоритмы обработки чисел, числовых последовательностей и массивов: представление числа в виде набора простых сомножителей; нахождение максимальной (минимальной) цифры натурального числа, записанного в системе счисления с основанием, не превышающим 10; вычисление обобщенных характеристик элементов массива или числовой последовательности (суммы, произведения среднего арифметического, минимального и максимального элементов, количества элементов, удовлетворяющих заданному условию); сортировку элементов массива;

Б10 умение создавать структурированные текстовые документы и демонстрационные материалы с использованием возможностей современных программных средств и облачных сервисов; умение использовать табличные (реляционные) базы данных, в частности, составлять запросы в базах данных (в том числе вычисляемые запросы), выполнять сортировку и поиск записей в базе данных; наполнять разработанную базу данных; умение использовать электронные таблицы для анализа, представления и обработки данных (включая вычисление суммы, среднего арифметического, наибольшего и наименьшего значений, решение уравнений);

Б11 умение использовать компьютерно-математические модели для анализа объектов и процессов: формулировать цель моделирования, выполнять анализ результатов, полученных в ходе моделирования; оценивать адекватность модели моделируемому объекту или процессу; представлять результаты моделирования в наглядном виде;

Б12 умение организовывать личное информационное пространство с использованием различных средств цифровых технологий; понимание возможностей цифровых сервисов государственных услуг, цифровых образовательных сервисов; понимание возможностей и ограничений технологий искусственного интеллекта в различных областях; наличие представлений об использовании информационных технологий в различных профессиональных сферах.

Перечисленные результаты освоения соотносятся с формируемыми компетенциями:

ОК 01. Выбирать способы решения задач профессиональной деятельности применительно к различным контекстам

ОК 02. Использовать современные средства поиска, анализа и интерпретации информации и информационные технологии для выполнения задач профессиональной деятельности

Общая/профессиональная компетенция	Раздел/Тема	Тип оценочных мероприятий
ОК 01	Тема 1.6 Тема 1.9 Тема 3.5	Тестирование
ОК 02	Тема 1.1 Тема 1.3 Тема 3.1 Тема 3.2 Тема 1.6 Тема 1.9	
ОК 01	Тема 1.7 Тема 1.8 Тема 2.2 Тема 3.4	Выполнение практических заданий
ОК 02	Тема 1.2 Тема 1.4 Тема 1.5 Тема 2.1 Тема 2.3 Тема 2.4	

	Тема 2.5 Тема 2.6 Тема 2.7 Тема 3.3 Тема 1.7 Тема 1.8 Тема 2.2 Тема 3.6 Тема 3.7 Тема 3.8 Тема 3.9 Тема 3.10 Тема 3.11 Тема 3.12 Тема 3.13	
ОК 02	Прикладные модули 1-2	Контрольная работа
ОК 02	Прикладные модули 2-8	Проектная работа
ОК 01, ОК 02	Все модули	Выполнение заданий дифференцированного зачета

2. Модельные примеры оценочных средств для проведения промежуточного и рубежного контроля по основному содержанию

Фонд оценочных средств для входного контроля

Входной контроль состоит из заданий, взятых из открытого банка ОГЭ и ВПР по информатике. На выполнение заданий входного контроля дается 1 академический час (30 минут). Входной контроль состоит из 2-х частей: обязательной и дополнительной. Обязательная часть содержит задания минимального обязательного уровня. При выполнении заданий требуется представить ход решения и указать полученный ответ. Правильно выполненное задание из обязательной части оценивается в один балл. Баллы, полученные за все выполненные задания, суммируются.

Критерии оценивания КИМ

Оценка	Количество баллов
5 (отлично)	от 90-100 %
4 (хорошо)	от 70-90 %
3 (удовлетворительно)	от 50-70%
2 (неудовлетворительно)	менее 50 %

Задания входного контроля (один из возможных вариантов)

1. В одной из кодировок Unicode каждый символ кодируется 16 битами. Вова написал текст (в нём нет лишних пробелов): «Бор, азот, гелий, натрий, водород, кислород, рентгений, менделевий, резерфордий – химические элементы». Ученик вычеркнул из списка название одного химического элемента. Заодно он вычеркнул ставшие лишними запятую и пробел – два пробела не должны идти подряд. При этом размер нового предложения в данной кодировке оказался на 18 байт меньше, чем размер исходного предложения. Напишите в ответе вычеркнутое название элемента.

Ответ _____

2. Вася и Петя играли в шпионов и кодировали сообщение собственным шифром.

Фрагмент кодовой таблицы приведён ниже.

А	Б	В	Г	Д	Е
~	#	#+	+~#	+ #	~#

Расшифруйте сообщение. Получившееся слово (набор букв) запишите в качестве ответа: # ~ # + + ~ #

Ответ _____

3. Напишите наибольшее натуральное число x , для которого ИСТИННО высказывание:
 $\neg (x < 3) \vee (x < 4)$

Ответ _____

4. Между населёнными пунктами А, В, С, D, Е построены дороги, протяжённость которых (в километрах) приведена в таблице:

	А	В	С	D	Е
А		3			

В	3		1	2	6
С		1			3
Д		2			3
Е		6	3	3	

Определите длину кратчайшего пути между пунктами А и Е. Передвигаться можно только по дорогам, протяжённость которых указана в таблице. Каждый пункт можно посетить только один раз.

Ответ _____

5. У исполнителя Вычислитель две команды, которым присвоены номера:

1. прибавь 3
2. умножь на 2

Первая из них увеличивает число на экране на 3, вторая удваивает его.

Составьте алгоритм получения из числа 1 числа 25, содержащий не более 5 команд.

В ответе запишите только номера команд.

(Например, 11221 – это алгоритм:

прибавь 3

прибавь 3

умножь на 2

умножь на 2

прибавь 3

который преобразует число 4 в 43.)

Если таких алгоритмов более одного, то запишите любой из них.

Ответ _____

6. Ниже приведена программа, записанная на четырех языках программирования

Алгоритмический язык	Паскаль	Python	C++
алг нач цел s, t, A ввод s ввод t ввод A если s > 10 или t > A то вывод "YES" иначе вывод "NO" все кон	<pre> var s, t, A: integer; begin readln(s); readln(t); readln(A); if (s > 10) or (t > A) then writeln("YES") else writeln("NO") end. </pre>	<pre> s = int(input()) t = int(input()) A = int(input()) if (s > 10) or (t > A): print("YES") else: print("NO") </pre>	<pre> #include <iostream> using namespace std; int main() { int s, t, A; cin >> s; cin >> t; cin >> A; if (s > 10) or (t > A) cout << "YES" << endl; else cout << "NO" << endl; return 0; } </pre>

Было проведено 9 запусков программы, при которых в качестве значений переменных s и t вводились следующие пары чисел:

(1, 2); (11, 2); (1, 12); (11, 12); (−11, −12); (−11, 12); (−12, 11); (10, 10); (10, 5).

Укажите наименьшее целое значение параметра A , при котором для указанных входных данных программа напечатает «NO» семь раз.

Ответ _____

7. Доступ к файлу `foto.jpg`, находящемуся на сервере `email.ru`, осуществляется по протоколу `http`. В таблице фрагменты адреса файла закодированы цифрами от 1 до 7. Запишите последовательность этих цифр, кодирующую адрес указанного файла в сети Интернет.

- 1) foto
- 2) email
- 3) .ru
- 4) ://
- 5) http
- 6) /
- 7) .jpg

Ответ _____

8. В языке запросов поискового сервера для обозначения логической операции «ИЛИ» используется символ «|», а для логической операции «И» – символ «&». В таблице приведены запросы и количество найденных по ним страниц некоторого сегмента сети Интернет.

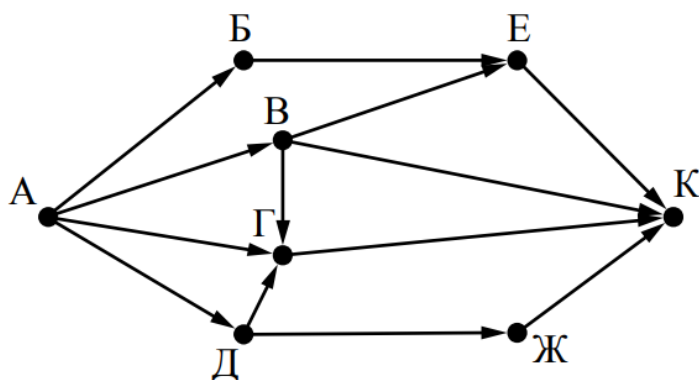
Запрос	Найдено страниц (в тысячах)
Москва	4220
Санкт-Петербург	3600
Москва Санкт-Петербург	5900

Какое количество страниц (в тысячах) будет найдено по запросу Москва & Санкт-Петербург?

Считается, что все запросы выполнялись практически одновременно, так что набор страниц, содержащих все искомые слова, не изменялся за время выполнения запросов

Ответ _____

9. На рисунке – схема дорог, связывающих города А, Б, В, Г, Д, Е, Ж и К. По каждой дороге можно двигаться только в одном направлении, указанном стрелкой. Сколько существует различных путей из города А в город К?



Ответ _____

10. Переведите число 110 из десятичной системы счисления в двоичную систему счисления. Сколько единиц содержит полученное число?

В ответе укажите одно число – количество единиц

Ответ _____

Эталоны ответов:

№	1	2	3	4	5	6	7	8	9	10
Ответ	водород	БАВГ	3	7	12121	12	5423617	1920	7	5

Фонд оценочных средств для текущего контроля

Текущий контроль проводится в форме домашнего задания для самостоятельного выполнения, используя материалы ЭОР на выбор (ЯКласс, РЭШ, 1СУрок, МЭО и др.). Результаты фиксируются в образовательной платформе, на которой зарегистрированы студенты и преподаватель.

Фонд оценочных средств для рубежного контроля

(Тема 1.1, Тема 1.3, Тема 1.6, Тема 1.9, Тема 3.1, Тема 3.2, Тема 3.5)

Образовательные результаты, подлежащие проверке (элементы):

Б1, Б2, Б3, Б4, Б7, Б9, Б11

ОК 01, ОК 02

Тема.1.1. Информация и информационные процессы

Ближе всего раскрывается смысл понятия «информация, используемая в бытовом общении» в утверждении:

последовательность знаков некоторого алфавита;

сообщение, передаваемое в форме знаков или сигналов;

сообщение, уменьшающее неопределенность знаний;

сведения об окружающем мире, воспринимаемые человеком

сведения, содержащиеся в научных теориях

Информацию, не зависящую от личного мнения, называют:

достоверной;

актуальной;

объективной;

полезной;

понятной

Информацию, отражающую истинное положение дел, называют:

понятной;

достоверной;

объективной;

полной;

полезной

Информацию, существенную и важную в настоящий момент, называют:

полезной;

актуальной;

достоверной;

объективной;

полной

Информацию, дающую возможность, решать поставленную задачу, называют:

понятной;

актуальной;

достоверной;

полезной;

полной

Информацию, достаточную для решения поставленной задачи, называют:

полезной;

актуальной;

полной;

достоверной;

понятной

Информацию, изложенную на доступном для получателя языке, называют:

полной;

полезной;

актуальной;

достоверной;

понятной

По способу восприятия информации человеком различают следующие виды информации:

текстовую, числовую, символьную, графическую, табличную и пр.;

научную, социальную, политическую, экономическую, религиозную пр.;

обыденную, производственную, техническую, управленческую;

визуальную, звуковую, тактильную, обонятельную, вкусовую;

математическую, биологическую, медицинскую, психологическую и пр.

По форме представления информации можно условно разделить на следующие виды:

социальную, политическую, экономическую, техническую, религиозную и пр.;

техническую, числовую, символьную, графическую, табличную пр.;

обыденную, научную, производственную, управленческую;

визуальную звуковую, тактильную, обонятельную, вкусовую;

математическую, биологическую, медицинскую, психологическую.

Укажите лишний объект с точки зрения способа представления информации:

школьный учебник;
фотография;
телефонный разговор;
картина;
чертеж

По области применения информацию можно условно разделить на:

текстовую и числовую;
визуальную и звуковую;
графическую и табличную;
научную и техническую;
тактильную и вкусовую

Какое из высказываний ложно?

получение и обработка информации являются необходимыми условиями жизнедеятельности любого организма.

для обмена информацией между людьми служат языки.

информацию условно можно разделить на виды в зависимости от формы представления.

процесс обработки информации техническими устройствами носит осмысленный характер.

процессы управления – это яркий пример информационных процессов, протекающих в природе, обществе, технике.

Каждая знаковая система строится на основе:

естественных языков, широко используемых человеком для представления информации;
двоичной знаковой системы, используемой в процессах хранения, обработки и передачи информации в компьютере;
определенного алфавита (набора знаков) и правил выполнения операций над знаками;
правил синтаксиса алфавита.

Выбери из списка все языки, которые можно считать формальными языками:

двоичная система счисления
языки программирования
кириллица
китайский язык
музыкальные ноты
русский язык
дорожные знаки
код азбуки Морзе.

Производится бросание симметричной восьмигранной пирамидки. Какое количество информации мы получаем в зрительном сообщении о ее падении на одну из граней?

1 бит
1 байт
3 бит
3 бита.

Какое количество информации (с точки зрения алфавитного подхода) содержит двоичное число 10101001?

1 байта
2 байта

3 байта

3 бита.

Что из нижеперечисленного не является основой формирования информационной культуры?

знания о законах функционирования информационной среды

принцип узкой специализации

знания об информационной среде

умение ориентироваться в информационных потоках

Установите соответствие:

А Полнота	1 Язык понятен получателю
Б Достоверность	2 Достаточность для понимания, принятия решения
В Актуальность	3 Важность, значимость
Г Понятность	4 Неискажение истинного положения дел
Д Релевантность	5 Вовремя, в нужный срок

Ключ к тесту

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Ответ	г	в	б	б	г	в	д	г	б	в	г	г	в	абджз	в	а	б	а2 б4 в3 г1 д3

Тема.1.3 Компьютер и цифровое представление информации. Устройство компьютера

Тактовая частота процессора – это

число двоичных операций, совершаемых за единицу времени

число обращений процессора к оперативной памяти за единицу времени

скорость обмена информацией между процессор и устройствами ввода- вывода

скорость обмена информацией между процессором и постоянным запоминающим

устройством (ПЗУ)

Через какие устройства взаимодействуют устройства внешней памяти и ввода/вывода с процессором

оперативную память

контроллеры

материнскую плату

системный блок

Часть магистрали, по которой передаются управляющие сигналы

шина управления

шина адреса

шина данных

шина контроллеров

Оперативная память ПК работает...

быстрее, чем внешняя
медленнее, чем внешняя
одинаково по скорости с внешней памятью
Внешняя память компьютера является...

энергозависимой
постоянной
оперативной
энергонезависимой

Основная характеристика процессора - это...

производительность
размер
температура
цена

Общим свойством машины Беббиджа, современного компьютера и человеческого мозга является способность обрабатывать:

числовую информацию;
текстовую информацию;
звуковую информацию;
графическую информацию.

В _____ г. Лейбниц изготовил механический калькулятор.

1643
1673
1642
1700

_____ октября — день рождения Интернета.

19
27
17
29

Выбери к какому поколению относится данная особенность: Габариты — ЭВМ выполнена в виде громадных шкафов.

1 поколение
2 поколение
3 поколение
4 поколение

Как назывался первый офисный компьютер, управляемый манипулятором «мышь»?

Altair 8800
IBM/370
Apple Lisa
Apple – 1

Массовое производство персональных компьютеров началось в:

40-е годы XX в.
50-е годы XX в.
80-е годы XX в.
90-е годы XX в.

Укажите верное высказывание:

компьютер состоит из отдельных модулей, соединенных между собой магистралью;
компьютер представляет собой единое, неделимое устройство;
составные части компьютерной системы являются незаменимыми;
компьютерная система способна сколь угодно долго соответствовать требованиям современного общества и не нуждается в модернизации.

Наименьшим адресуемым элементом оперативной памяти является:

машинное слово;

регистр;

байт;

файл.

При выключении компьютера вся информация стирается:

на флешке;

в облачном хранилище;

на жестком диске;

в оперативной памяти

Производительность работы компьютера зависит от:

типа монитора;

частоты процессора;

напряжения питания;

объема жесткого диска.

Укажите верное высказывание:

На материнской плате размещены только те блоки, которые осуществляют обработку информации, а схемы, управляющие всеми остальными устройствами компьютера, реализованы на отдельных платах и вставляются в стандартные разъемы на материнской плате;

На материнской плате размещены все блоки, которые осуществляют прием, обработку и выдачу информации с помощью электрических сигналов и к которым можно подключить все необходимые устройства ввода-вывода;

На материнской плате находится системная магистраль данных, к которым подключены адаптеры и контроллеры, позволяющие осуществлять связь ЭВМ с устройствами ввода-вывода;

На материнской плате расположены все устройства компьютерной системы и связь между ними осуществляется через магистраль.

Системное программное обеспечение – это

Программы для организации совместной работы устройств компьютера как единой системы

Программы для организации удобной системы размещения программ на диске
набор программ для работы устройств системного блока компьютера
программы, ориентированные на решение конкретных задач, рассчитанные на взаимодействие с пользователем

Ключ к тесту

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Ответ	б	б	а	а	б	а	а	б	г	а	в	в	а	б	г	б	а	а

Тема 1.6. Компьютерные сети: локальные сети, сеть Интернет

Компьютерная сеть это...

группа компьютеров и линии связи

группа компьютеров в одном помещении

группа компьютеров в одном здании

группа компьютеров, соединённых линиями связи

Укажите преимущества, использования компьютеров в сети

совместное использование ресурсов

обеспечение безопасности данных

использование сетевого оборудования

быстрый обмен данными между компьютерами

Установите соответствие типов компьютерных сетей по "радиусу охвата"

1 Сети, объединяющие компьютеры в пределах города	А Персональные сети
2 Сети компьютеров одной организации (возможно, находящиеся в разных районах города или даже в разных городах)	Б Глобальные сети
3 Сети, объединяющие компьютеры в разных странах; типичный пример глобальной сети – Интернет	В Локальные сети
4 Сети, объединяющие, как правило, компьютеры в пределах одного или нескольких соседних зданий	Г Городские сети
5 Сети, объединяющие устройства одного человека (сотовые телефоны, карманные компьютеры, смартфоны, ноутбук и т. п.) в радиусе не более 30 м	Д Корпоративные сети

Установите соответствие между типом сервера и его назначением.

1 Обеспечивает доступ к общему принтеру	А Почтовый сервер
2 Хранит данные и обеспечивает доступ к ним	Б Файловый сервер
3 Управляет электронной почтой	В Сервер печати
4 Выполняют обработку информации по запросам клиента	Г Сервер приложений

Укажите наиболее полное верное назначение шлюза.

Преобразование данных в формат нужного протокола.

верны все варианты

Передача информации по сети.

Дублирование пакетов при их передаче в сетях

Выберете наиболее верное утверждение о сервере

это компьютер, использующий ресурсы сервера

это самый большой и мощный компьютер

это компьютер, предоставляющий свои ресурсы в общее использование

сервером является каждый компьютер сети

Определите топологии

1 Все рабочие станции подключены в сеть через	А Кольцо
---	----------

центральное устройство (коммутатор).	
2 Все рабочие станции подключены к одному кабелю с помощью специальных разъёмов	Б Звезда
3 Каждый компьютер соединён с двумя соседними, причём от одного он только получает данные, а другому только передаёт. Таким образом, пакеты движутся в одном направлении.	В Шина

Укажите достоинства топологии "Шина"

при выходе из строя любого компьютера сеть продолжает работать

легко подключать новые рабочие станции

высокий уровень безопасности

самая простая и дешёвая схема

простой поиск неисправностей и обрывов

небольшой расход кабеля

Укажите недостатки топологии "Звезда"

большой расход кабеля, высокая стоимость

для подключения нового узла нужно останавливать сеть

при выходе из строя коммутатора вся сеть не работает

количество рабочих станций ограничено количеством портов коммутатора

низкий уровень безопасности

Укажите достоинства топологии "Кольцо"

не нужно дополнительное оборудование (коммутаторы)

при выходе из строя любой рабочей станции сеть остаётся работоспособной

легко подключать новые рабочие станции

большой размер сети (до 20 км)

надёжная работа при большом потоке данных, конфликты практически невозможны

Укажите особенности организации одноранговой сети

каждый компьютер может выступать как в роли клиента, так и в роли сервера

повышенный уровень безопасности

все компьютеры в сети равноправны

пользователь сам решает какие ресурсы своего компьютера сделать совместными

основная обработка данных выполняется на серверах

Выберете верные утверждения

Серверная операционная система устанавливается на каждую рабочую станцию, входящую в сеть.

Серверная операционная система устанавливается на мощный компьютер, отвечающий за работу всей сети.

Современные технологии позволяют создавать сложные сети без использования серверной операционной системы.

Терминальный доступ - важная особенность сетевой операционной системы.

Для объединения компьютеров в беспроводную сеть чаще всего используют специальное устройство...

Адаптер

Коммутатор

Шлюз

Точка доступа

Восьмиконтактный разъём с защёлкой часто называют



Витая пара

RJ-45

RJ

шлюз

Для связи локальной сети с Интернетом необходимо такое устройство как...

коммутатор

концентратор

адаптер

маршрутизатор

Установите соответствие между устройствами и их назначением

1 Устройство для передачи пакета данных только тому узлу, которому он предназначен.	А Шлюз
2 Дублирует пакеты на все подключенные к нему рабочие станции	Б Коммутатор
3 Используется для объединения в сеть устройств, использующих разные протоколы обмена данными	В Точка доступа
4 Используется для объединения компьютеров в беспроводную сеть	Г Концентратор

Установите соответствие определений и понятий

1 Программа, удаляющая из текста страницы всю служебную информацию -	А Поисковая система
2 Текст, в котором есть активные ссылки на другие документы -	Б Веб-сайт
3 Группа веб-страниц, расположенных на одном сервере, связанных с помощью гиперссылок -	В Индексный робот
4 Веб-сайт, предназначенный для поиска информации в Интернете -	Г Гипертекст

Укажите протокол, используемый для скачивания файлов с сервера на компьютер пользователя.

HTTP

FTP

SMTP

FAIL

Ключ к тесту

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Ответ	г	а г	1г 2д 3б 4в 5а	1в 2б 3а 4г	а	в	1б 2в 3а	а б г е	а в г	а г д	а в г	б г	г	б	г	1б 2г 3а 4в	1в 2г 3б 4а	б

Тема 1.9. Информационная безопасность

Обеспечение какого из свойств информации не является задачей информационной безопасности?

актуальность

аутентичность

целостность

конфиденциальность

Воздействие на информацию, которое происходит вследствие ошибок ее пользователя, сбоя технических и программных средств информационных систем, природных явлений или иных нецеленаправленных на изменение информации событий, называется...

Ответ _____

Заполните пропуски в предложении.

... информации – субъект, пользующийся информацией, полученной от ее собственника, владельца или ... в соответствии с установленными правами и правилами доступа к информации либо с их

пользователь, разработчика, модификациями

пользователь, посредника, нарушением

владелец, разработчика, нарушением

владелец, посредника, модификациями

К показателям информационной безопасности относятся:

дискретность

целостность

конфиденциальность

доступность

актуальность

Установите соответствие

1 право пользования	А только собственник информации имеет право определять, кому эта информация может быть предоставлена
2 право распоряжения	Б собственник информации имеет право использовать ее в своих интересах
3 право владения	В никто, кроме собственника информации, не может ее изменять

Лицензия на программное обеспечение – это

документ, определяющий порядок распространения программного обеспечения, защищённого авторским правом

документ, определяющий порядок использования и распространения программного обеспечения, незащищённого авторским правом

документ, определяющий порядок использования и распространения программного обеспечения, защищённого авторским правом

документ, определяющий порядок использования программного обеспечения, защищённого авторским правом

Как называется совокупность условий и факторов, создающих потенциальную или реально существующую опасность нарушения безопасности информации?

уязвимость

слабое место системы

угроза

атака

Пароль пользователя должен

Содержать цифры и буквы, знаки препинания и быть сложным для угадывания

Содержать только буквы

Иметь явную привязку к владельцу (его имя, дата рождения, номер телефона и т.п.)

Быть простым и легко запоминаться, например «123», «111», «qwerty» и т.д.

Каким требованиям должен соответствовать пароль, чтобы его было трудно взломать?

Пароль должен состоять из цифр

Символы в пароле не должны образовывать никаких слов, чисел, аббревиатур, связанных с пользователем

Пароль не должен быть слишком длинным

Пароль должен быть достаточно простым, чтобы вы его могли запомнить

Пароль не должен состоять из одного и того же символа или повторяющихся фрагментов

Пароль не должен совпадать с логином

Пароль должен состоять не менее чем из 6 символов

Пароль должен совпадать с логином

Что требуется ввести для авторизованного доступа к сервису для подтверждения, что логином хочет воспользоваться его владелец

Ответ _____

Как называется программа для обнаружения компьютерных вирусов и вредоносных файлов, лечения и восстановления инфицированных файлов, а также для профилактики?

Ответ _____

Установите соответствие между средством или способом защиты и проблемой, для решения которой данный способ применяется:

1 использование тонкого клиента	А передача секретной информации сотрудникам компании (человеческий фактор)
2 шифрование с открытым ключом	Б доступ посторонних к личной информации
3 Антивирусы	В несанкционированный доступ к компьютеру и части сети
4 Авторизация пользователя	Г доступ посторонних к личной информации при хранении и передаче по открытым каналам связи

5 Межсетевые экраны	Д вредоносные программы

Виды информационной безопасности:

Персональная, корпоративная, государственная

Клиентская, серверная, сетевая

Локальная, глобальная, смешанная

Что называют защитой информации?

Все ответы верны

Называют деятельность по предотвращению утечки защищаемой информации

Называют деятельность по предотвращению несанкционированных воздействий на защищаемую информацию

Называют деятельность по предотвращению непреднамеренных воздействий на защищаемую информацию

Шифрование информации это

Процесс ее преобразования, при котором содержание информации становится непонятным для не обладающих соответствующими полномочиями субъектов

Процесс преобразования, при котором информация удаляется

Процесс ее преобразования, при котором содержание информации изменяется на ложную

Процесс преобразования информации в машинный код

Функция защиты информационной системы, гарантирующая то, что доступ к информации, хранящейся в системе, может быть осуществлен только тем лицам, которые на это имеют право

управление доступом

конфиденциальность

аутентичность

целостность

доступность

Элемент аппаратной защиты, где используется резервирование особо важных компьютерных подсистем

защита от сбоев в электропитании

защита от сбоев серверов, рабочих станций и локальных компьютеров

защита от сбоев устройств для хранения информации

защита от утечек информации электромагнитных излучений

Что можно отнести к правовым мерам ИБ?

разработку норм, устанавливающих ответственность за компьютерные преступления, защиту авторских прав программистов, совершенствование уголовного и гражданского законодательства, а также судопроизводства

охрану вычислительного центра, тщательный подбор персонала, исключение случаев ведения особо важных работ только одним человеком, наличие плана восстановления работоспособности центра и т.д.

защиту от несанкционированного доступа к системе, резервирование особо важных компьютерных подсистем, организацию вычислительных сетей с возможностью перераспределения ресурсов в случае нарушения работоспособности отдельных звеньев, установку оборудования обнаружения и тушения пожара, оборудования обнаружения воды, принятие конструктивных мер защиты от хищений, саботажа, диверсий, взрывов,

установку резервных систем электропитания, оснащение помещений замками, установку сигнализации и многое другое
охрану вычислительного центра, установку сигнализации и многое другое

Ключ к тесту

№	Ответ
1	а
2	непреднамеренным
3	б
4	б в г
5	1б 2а 3в
6	в
7	в
8	а
9	б г д е ж
10	пароль
11	антивирус
12	1а 2г 3д 4б 5в
13	а
14	а б в г
15	а
16	б
17	б
18	а

Тема 3.1. Модели и моделирование. Этапы моделирования

Отметьте те задачи, которые могут решаться с помощью моделирования:

разработка объекта с заданными свойствами

оценка влияния внешней среды на объект

разрушение объекта

перемещение объекта

выбор оптимального решения

Отметьте все «плохо поставленные» задачи?

задача, которую вы не умеете решать

задача, в которой не хватает исходных данных

задача, в которой может быть несколько решений

задача, для которой неизвестно решение

задача, в которой неизвестны связи между исходными данными и результатом

Какие из этих высказываний верны?

Для каждого объекта можно построить только одну модель.

Для каждого объекта можно построить много моделей.

Разные модели отражают разные свойства объекта.

Модель должна описывать все свойства объекта.

Модель может описывать только некоторые свойства объекта.

Отметьте все пары объектов, которые в каких-то задачах можно рассматривать как пару «оригинал-модель».

страна — столица

болт — чертёж болта

курица — цыпленок

самолёт — лист металла

учитель — ученик

Как называется модель в форме словесного описания (в ответе введите прилагательное)?

Ответ _____

Как называется модель сложного объекта, предназначенная для выбора оптимального решения методом проб и ошибок (в ответе введите прилагательное)?

Ответ _____

Как называется модель, в которой используются случайные события?

Ответ _____

Как называется модель, которая описывает изменение состояния объекта во времени (в ответ введите прилагательное)?

Ответ _____

Как называется проверка модели на простых исходных данных с известным результатом?

Ответ _____

Как называется четко определенный план решения задачи?

Ответ _____

Какие из перечисленных моделей относятся к информационным?

рисунок дерева

модель ядра атома из металла

уменьшенная копия воздушного шара

таблица с данными о населении Земли

формула второго закона Ньютона

Какие из этих фраз можно считать определением модели?

это уменьшенная копия оригинала

это объект, который мы исследуем для того, чтобы изучить оригинал

это копия оригинала, обладающая всеми его свойствами

это словесное описание оригинала

это формулы, описывающие изменение оригинала

Какими свойствами стального шарика можно пренебречь, когда мы исследуем его полет на большой скорости?

массой шарика

объемом шарика

изменением формы шарика в полете

изменением ускорения свободного падения

сопротивлением воздуха

Какой из этапов моделирования может привести к самым трудноисправимым ошибкам?

Тестирование

Эксперимент

постановка задачи

разработка модели

анализ результатов моделирования

Какую фразу можно считать определением игровой модели?

это модель для поиска оптимального решения

это модель, учитывающая действия противника

это модель компьютерной игры

это модель объекта, с которой играет ребенок

это компьютерная игра

Какая фраза может служить определением формальной модели?

модель в виде формулы

словесное описание явления

модель, записанная на формальном языке

математическая модель

Модель – это:

фантастический образ реальной действительности

материальный или абстрактный заменитель объекта, отражающий его пространственно-временные характеристики

материальный или абстрактный заменитель объекта, отражающий его существенные характеристики

описание изучаемого объекта средствами изобразительного искусства

информация о несущественных свойствах объекта

Файловая система персонального компьютера наиболее адекватно может быть описана в виде:

Табличной модели

Графической модели

Иерархической модели

Математической модели

Ключ к тесту

№	Ответ
1	а б д
2	б в д
3	б в д
4	а б в г д
5	вербальная
6	имитационная
7	вероятностная
8	динамическая
9	тестирование
10	алгоритм
11	а г д
12	б
13	в г
14	в
15	б

16	В
17	В
18	В

Тема 3.2. Списки, графы, деревья

Между населёнными пунктами А, В, С, D, E, F построены дороги, протяжённость которых приведена в таблице. Определите длину кратчайшего маршрута из А в F:

	A	B	C	D	E	F
A		2	4			
B	2		1		7	
C	4	1		3	4	
D			3		3	
E		7	4	3		2
F					2	

Ответ _____

Между населёнными пунктами А, В, С, D, E построены дороги, протяжённость которых приведена в таблице. Определите длину кратчайшего маршрута из А в E:

Ответ _____

	A	B	C	D	E
A			3	1	
B			4		2

	A	B	C	D	E
A		4	3		7
B	4			2	
C	3			6	
D		2	6		1
E	7			1	

Между населёнными пунктами А, В, С, D, E построены дороги, стоимость перевозки по которым приведена в таблице. Определите МАКСИМАЛЬНУЮ стоимость перевозки груза из С в В при условии, что маршрут не может проходить через какой-то пункт более одного раза:

Ответ _____

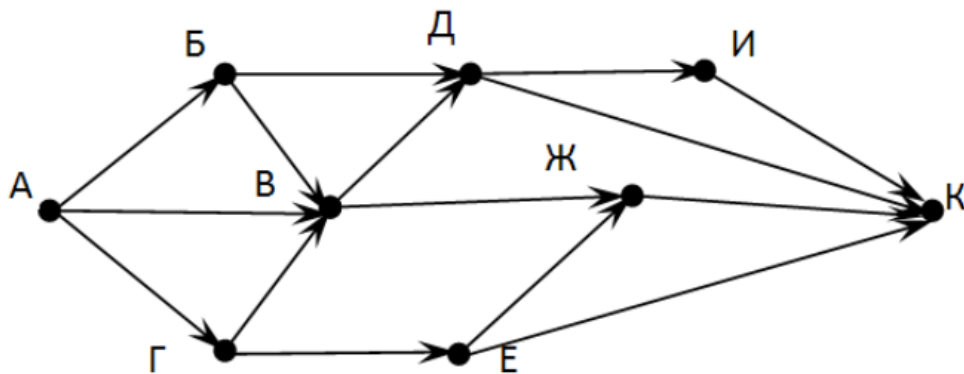
Между населёнными пунктами А, В, С, D, E построены дороги, стоимость перевозки по

	A	B	C	D	E
A			2	2	6
B				2	
C	2			2	
D	2	2	2		
E	6				

которым приведена в таблице. Определите МАКСИМАЛЬНУЮ стоимость перевозки груза из С в В при условии, что маршрут не может проходить через какой-то пункт более одного раза:

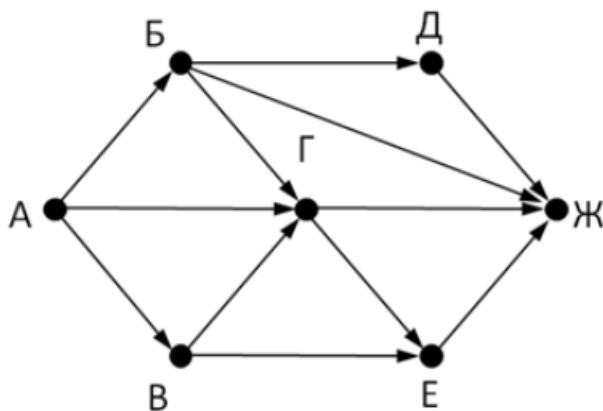
Ответ _____

На рисунке - схема дорог, связывающих города А, Б, В, Г, Д, Е, Ж, З, И, К. По каждой дороге можно двигаться только в одном направлении, указанном стрелкой. Сколько существует различных путей из города А в город К



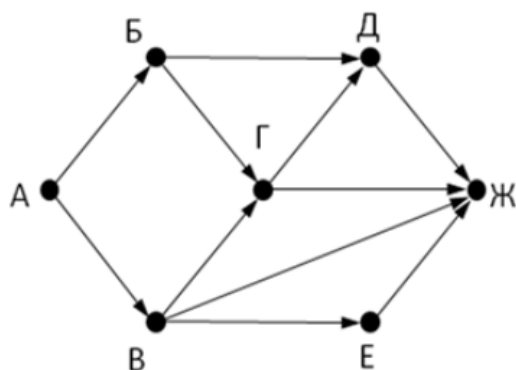
Ответ _____

На рисунке - схема дорог, связывающих города А, Б, В, Г, Д, Е, Ж. По каждой дороге можно двигаться только в одном направлении, указанном стрелкой. Сколько существует различных путей из города А в город Ж



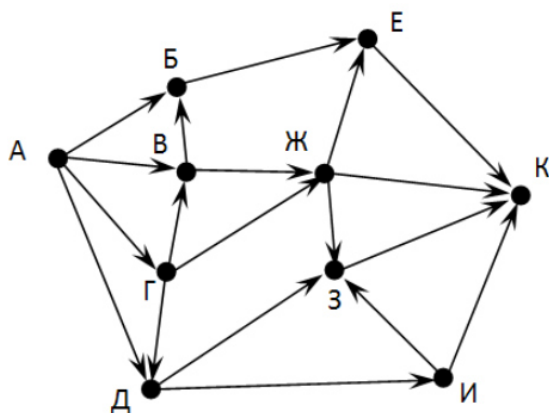
Ответ _____

На рисунке - схема дорог, связывающих города А, Б, В, Г, Д, Е, Ж. По каждой дороге можно двигаться только в одном направлении, указанном стрелкой. Сколько существует различных путей из города А в город Ж



Ответ _____

На рисунке - схема дорог, связывающих города А, Б, В, Г, Д, Е, Ж, З, И, К. По каждой дороге можно двигаться только в одном направлении, указанном стрелкой. Сколько существует различных путей из города А в город К



Ответ _____

Вычислите выражение, записанное в постфиксной форме: 5 13 7 - *

Ответ _____

Вычислите выражение, записанное в префиксной форме: * + 5 7 - 6 3

Ответ _____

Запишите выражение $c \cdot (a+b)$ в постфиксной форме (без пробелов!)

Ответ _____

Запишите выражение $5 \cdot (d-3)$ в префиксной форме (без пробелов!)

Ответ _____

Запишите выражение $(c-d) \cdot (a-b)$ в постфиксной форме (без пробелов!)

Ответ _____

Запишите выражение $3 \cdot a + 2 \cdot d$ в префиксной форме (без пробелов!)

Ответ _____

Вычислите выражение, записанное в префиксной форме: * - + a 3 b c при $a = 6$, $b = 4$ и $c = 2$

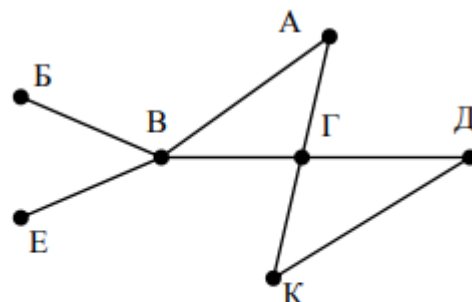
Ответ _____

Вычислите выражение, записанное в постфиксной форме: a b c 7 + * - при $a = 28$, $b = 2$ и $c = 1$

Ответ _____

На рисунке схема дорог N-ского района изображена в виде графа, в таблице содержатся сведения о протяжённости каждой из этих дорог (в километрах).

		Номер пункта						
		1	2	3	4	5	6	7
Номер пункта	1		9	6	3	10		
	2	9						
	3	6			8			
	4	3		8			7	1
	5	10						
	6				7			5
	7				1		5	

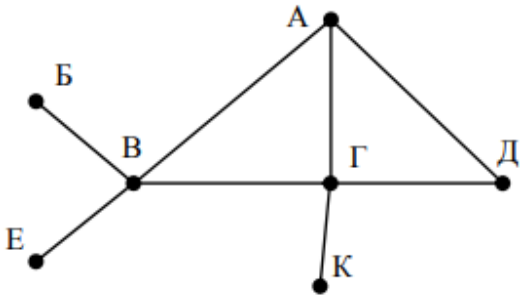


Так как таблицу и схему рисовали независимо друг от друга, то нумерация населённых пунктов в таблице никак не связана с буквенными обозначениями на графе. Определите, какова протяжённость дороги из пункта А в пункт Г. В ответе запишите целое число – так, как оно указано в таблице

Ответ _____

На рисунке схема дорог N-ского района изображена в виде графа, в таблице содержатся сведения о протяжённости каждой из этих дорог (в километрах).

		Номер пункта						
		1	2	3	4	5	6	7
Номер пункта	1		15	15	9	7		
	2	15						
	3	15			12			20
	4	9		12			14	10
	5	7						
	6				14			
	7			20	10			



Так как таблицу и схему рисовали независимо друг от друга, то нумерация населённых пунктов в таблице никак не связана с буквенными обозначениями на графе. Определите, какова протяжённость дороги из пункта А в пункт Г. В ответе запишите целое число – так, как оно указано в таблице

Ответ _____

Ключ к тесту

№	Ответ
1	9
2	5
3	18
4	6
5	13
6	9
7	7
8	18
9	30
10	36
11	cab+*
12	*5-d3
13	cd-ab-*
14	+*3a*2d
15	10
16	12
17	8
18	12

Тема 3.5 Анализ алгоритмов в профессиональной области

Предлагается некоторая операция над двумя произвольными трехзначными десятичными числами:

- 1) записывается результат сложения старших разрядов этих чисел;
- 2) к нему дописывается результат сложения средних разрядов по такому правилу: если он меньше первой суммы, то полученное число приписывается к первому слева, иначе — справа;
- 3) итоговое число получают приписыванием справа к числу, полученному после второго шага, суммы значений младших разрядов исходных чисел.

Какое из перечисленных чисел могло быть построено по этому правилу?

141310

102113

101421

101413

Определите значение целочисленных переменных x , y и t после выполнения фрагмента программы:

$x := 5;$

$y := 7;$

$t := x;$

$x := y \bmod x;$

$y := t;$

$x=2, y=5, t=5$

$x=7, y=5, t=5$

$x=2, y=2, t=2$

$x=5, y=5, t=5$

Определите значение переменной c после выполнения следующего фрагмента программы:

$a := 6;$

$b := 15;$

$a := b - a * 2;$

if $a > b$

then $c := a + b$

else $c := b - a;$

–3

33

18

12

У исполнителя Калькулятор две команды, которым присвоены номера:

1. прибавь 2

2. умножь на 3

Выполняя первую из них, Калькулятор прибавляет к числу на экране 2, а выполняя вторую, утраивает его. Запишите порядок команд в программе получения из 0 числа 28, содержащей не более 6 команд, указывая лишь номера команд.

Например, программа 21211 — это программа:

умножь на 3

прибавь 2

умножь на 3
прибавь 2
прибавь 2
которая преобразует число 1 в 19.

Ответ _____

Определите значение переменной y, которое будет получено в результате выполнения следующей программы:

```
var i, y: integer;  
begin  
  y := 0;  
  for i := 1 to 4 do  
    begin  
      y := y * 10;  
      y := y + i;  
    end  
  end.
```

Ответ _____

Определите значение переменной y, которое будет получено в результате выполнения следующей программы:

```
var y : real; i : integer;  
begin  
  y := 0;  
  i := 1;  
  repeat  
    i := 2*i;  
    y := y + i  
  until i > 5;  
end.
```

Ответ _____

Определите значение переменной y, которое будет получено в результате выполнения следующей программы:

```
var y : real; i : integer;  
begin  
  y := 0;  
  i := 5;  
  while i > 2 do  
    begin  
      i := i - 1;  
      y := y + i * i  
    end;  
  end.
```

Ответ _____

В программе описан одномерный целочисленный массив с индексами от 0 до 10. В приведенном ниже фрагменте программы массив сначала заполняется, а потом изменяется:

for i:=0 to 10 do

A[i]:= i + 1;

for i:=0 to 10 do

A[i]:= A[10-i];

Чему будут равны элементы этого массива?

10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0

11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1

11, 10, 9, 8, 7, 6, 7, 8, 9, 10, 11

10, 9, 8, 7, 6, 5, 6, 7, 8, 9, 10

Все элементы двумерного массива A размером 5x5 равны 0. Сколько элементов массива после выполнения фрагмента программы будут равны 1?

for n:=1 to 5 do

for m:=1 to 5 do

A[n,m] := (m - n)*(m - n);

2

5

8

14

В программе описан одномерный целочисленный массив A с индексами от 0 до 10. Ниже представлен фрагмент этой программы, в котором значения элементов массива сначала задаются, а затем меняются.

for i:=0 to 10 do

A[i]:=i-1;

for i:=1 to 10 do

A[i-1]:=A[i];

A[10]:=10;

Как изменятся элементы этого массива после выполнения фрагмента программы?

все элементы, кроме последнего, окажутся равны между собой

все элементы окажутся равны своим индексам

все элементы, кроме последнего, будут сдвинуты на один элемент вправо

все элементы, кроме последнего, уменьшатся на единицу

Алгоритм вычисления значения функции F(n), где n – натуральное число, задан следующими соотношениями:

$F(1) = 1$

$F(n) = F(n-1) * (2*n + 1)$, при $n > 1$

Чему равно значение функции F(4)?

27

9

105

315

Дан рекурсивный алгоритм:

procedure F(n: integer);

begin

writeln('*');

if n > 0 then begin

```

F(n-3);
F(n div 2);
end
end;

```

Сколько символов «звездочка» будет напечатано на экране при выполнении вызова F(7)?

7
10
13
15

Дан рекурсивный алгоритм:

```

procedure F(n: integer);
begin
  writeln(n);
  if n < 5 then begin
    F(n+3);
    F(n*3)
  end
end;

```

Найдите сумму чисел, которые будут выведены при вызове F(1).

Ответ _____

Ниже записан рекурсивный алгоритм F:

```

function F(n: integer): integer;
begin
  if n > 2 then
    F := F(n-1)+F(n-2)+F(n-3)
  else
    F := n;
  end;
end;

```

Чему будет равно значение, вычисленное алгоритмом при выполнении вызова F(6)?

Ответ _____

Запишите число, которое будет напечатано в результате выполнения следующей программы. Для Вашего удобства программа представлена на четырех языках программирования

Алгоритмический язык	Паскаль	Python	C++
алг нач цел n, s s := 0 n := 95 нц пока s + n < 177 s := s + 10 n := n - 5 кц вывод n	<pre> var s, n: integer; begin s := 0; n := 95; while s + n < 177 do begin s := s + 10; n := n - 5 end; writeln(n) </pre>	<pre> s = 0 n = 95 while s + n < 177: s = s + 10 n = n - 5 print(n) </pre>	<pre> #include <iostream> using namespace std; int main() { int s = 0, n = 95; while (s + n < 177) { s = s + 10; n = n - 5; } cout << n << endl; return 0; } </pre>

кон	end.		}
-----	------	--	---

Ответ _____

Запишите число, которое будет напечатано в результате выполнения следующей программы. Для Вашего удобства программа представлена на четырех языках программирования

Алгоритмический язык	Паскаль	Python	C++
алг нач цел n, s s := 0 n := 90 нц пока s + n < 145 s := s + 15 n := n - 5 кц вывод n кон	<pre> var s, n: integer; begin s := 0; n := 90; while s + n < 145 do begin s := s + 15; n := n - 5; end; writeln(n) end.</pre>	<pre> s = 0 n = 90 while s + n < 145: s = s + 15 n = n - 5 print(n)</pre>	<pre> #include <iostream> using namespace std; int main() { int s = 0, n = 90; while (s + n < 145) { s = s + 15; n = n - 5; } cout << n << endl; return 0; }</pre>

Ответ _____

Ниже приведена программа, записанная на пяти языках программирования

Алгоритмический язык	Паскаль	Python	C++
алг нач цел s, t, A ввод s ввод t ввод A если s > 10 или t > A то вывод "YES" иначе вывод "NO" все кон	<pre> var s, t, A: integer; begin readln(s); readln(t); readln(A); if (s > 10) or (t > A) then writeln("YES") else writeln("NO") end.</pre>	<pre> s = int(input()) t = int(input()) A = int(input()) if (s > 10) or (t > A): print("YES") else: print("NO")</pre>	<pre> #include <iostream> using namespace std; int main() { int s, t, A; cin >> s; cin >> t; cin >> A; if (s > 10) or (t > A) cout << "YES" << endl; else cout << "NO" << endl; return 0; }</pre>

Было проведено 9 запусков программы, при которых в качестве значений переменных s и t вводились следующие пары чисел: (1, 2); (11, 2); (1, 12); (11, 12); (−11, −12); (−11, 12); (−12, 11); (10, 10); (10, 5). Укажите количество целых значений параметра A, при которых для указанных входных данных программа напечатает «YES» шесть раз.

Ответ _____

Ниже приведена программа, записанная на пяти языках программирования

Алгоритмический язык	Паскаль	Python	C++
----------------------	---------	--------	-----

алг нач цел s, t, A ввод s ввод t ввод A если s > 10 или t > A то вывод "YES" иначе вывод "NO" все кон	var s, t, A: integer; begin readln(s); readln(t); readln(A); if (s > 10) or (t > A) then writeln("YES") else writeln("NO") end.	s = int(input()) t = int(input()) A = int(input()) if (s > 10) or (t > A): print("YES") else: print("NO")	#include <iostream> using namespace std; int main() { int s, t, A; cin >> s; cin >> t; cin >> A; if (s > 10) or (t > A) cout << "YES" << endl; else cout << "NO" << endl; return 0; }
--	---	---	---

Было проведено 9 запусков программы, при которых в качестве значений переменных s и t вводились следующие пары чисел: (1, 2); (11, 2); (1, 12); (11, 12); (–11, –12); (–11, 12); (–12, 11); (10, 10); (10, 5). Укажите наименьшее целое значение параметра A, при котором для указанных входных данных программа напечатает «NO» семь раз.

Ответ _____

Ключ к тесту

№	Ответ
1	г
2	а
3	г
4	121211
5	1234
6	14
7	29
8	в
9	в
10	б
11	г
12	г
13	42
14	20
15	10
16	60
17	5
18	12

3. Модельные примеры оценочных средств для проведения промежуточного и рубежного контроля по прикладному модулю 1 «Основы аналитики и визуализации данных»

Промежуточный контроль

Промежуточный контроль по прикладному модулю 1 “Основы аналитики и анализа данных проходит в форме разработки и презентации дашборда после изучения темы 1.3. “Потоки данных”.

Создание презентации на основе данных датасета Анализ продаж и сети магазинов из БД Clickhouse и Marketplace.

Постановка задачи:

Самостоятельно создайте дашборд с визуализацией данных по продажам в сети магазинов из БД Clickhouse и Marketplace в соответствии с рекомендациями -

<https://cloud.yandex.ru/docs/datalens/tutorials/data-from-ch-visualization>

-по дням и месяцам

-по продуктовым категориям

Требования к презентации:

Презентация должна содержать описание этапов работы над дашбордом и скриншот самого получившегося дашборда с необходимыми чартами (диаграммами).

Шкала оценки

Шкала	Критерии
Зачтено	Выполнены все этапы разработки дашборда. Все чарты отображаются корректно. Презентация включает этапы работы над проектом. Оформление презентации соответствует общим требованиям к дизайну визуальных продуктов.
Не зачтено	Выполнены не все этапы разработки дашборда. Не все чарты отображаются корректно. Презентация не содержит этапы работы над проектом. Оформление не презентации соответствует общим требованиям к дизайну визуальных продуктов.

Рубежный контроль

Рубежный контроль проходит в форме выполнения и защиты проектной работы **Анализ открытых данных ДТП на дорогах России.**

Выполнение проекта по теме Анализ открытых данных ДТП на дорогах России.

Постановка задачи: провести анализ набора данных (датасета) о дорожно-транспортных происшествиях на дорогах в регионах РФ. Необходимо проанализировать данные, построить дашборд по ключевым диаграммам, сделать выводы о том, в каких регионах и в какое время суток ДТП происходят чаще. Сформулировать свои предложения по улучшению ситуации на дорогах. Предложить свои параметры и метрики, по которым будет проверяться эффективность сделанных предложений.

Содержание самостоятельной деятельности обучающегося:

- Авторизоваться на сервисе Yandex DataLens
- Создать подключение к датасету Анализ открытых данных ДТП на дорогах России в соответствии с рекомендациями - (см. <https://cloud.yandex.ru/docs/datalens/tutorials/data-from-csv-geo-visualization>)
- Создать карты геоданных и зависимости ДТП от времени суток
- Создать дашборд из полученных чартов
- Проанализировать полученные данные
- Создать презентацию, включив в нее построенные карты и зафиксировав выводы о выявленных зависимостях
- Включить в презентацию свои предложения для снижения количества ДТП
- Включить в презентацию описание параметров и метрик для оценки эффективности сделанных предложений

Требования к результату проектной деятельности

Разработка дашборда

Выполнены все этапы разработки дашборда.

Все карты отображаются корректно

Все переключатели и фильтры переключаются корректно

Презентация и выступление

Презентация включает:

Этапы работы над проектом:

анализ данных и принцип их отбора и обработки;

аргументацию выбора типа диаграммы и процесс ее создания;

визуализацию итогового дашборда

Формулировку предложений и метрик по оценке их эффективности.

Рефлексия: с какими проблемами столкнулись участники команды во время разработки проекта, как они их преодолели.

Оформление презентации соответствует общим требованиям к дизайну визуальных продуктов.

Выступление содержательно, при этом выдержан регламент выступления (5-7 минут).

Шкала итоговой оценки

Шкала	Критерии
Зачтено	Выполнены все этапы разработки дашборда. Все карты отображаются корректно. Все переключатели и фильтры переключаются корректно. Презентация включает этапы работы над проектом: анализ данных и принцип их отбора и обработки; аргументацию выбора типа диаграммы и процесс ее создания; визуализацию итогового дашборда формулировку предложений и метрик по оценке их эффективности. рефлексию работы над проектом. Оформление презентации соответствует общим требованиям к дизайну визуальных продуктов. Выступление содержательно, при этом выдержан регламент

	выступления (5-7 минут).
Не зачтено	Выполнены не все этапы разработки дашборда. Не все чарты отображаются корректно. Не все переключатели и фильтры переключаются корректно. Презентация включает не все этапы работы над проектом. Оформление презентации не соответствует общим требованиям к дизайну визуальных продуктов. Выступление недостаточно содержательно, регламент выступления (5-7 минут) не соблюден.

4. Модельные примеры оценочных средств для проведения промежуточного и рубежного контроля по прикладному модулю 2 «Аналитика и визуализация данных на Python»

Выполнение контрольной работы предусмотрено после изучения темы 2.3. Работа со списками и словарями

1. Определите, что будет напечатано в результате выполнения программы:

```
z = 30
for n in range(10):
    if n < 0:
        z = z - 2 * n
    else:
        z = n - z
```

```
print(z)
```

Ответ: -35

2. Определите, что будет напечатано в результате выполнения программы:

```
a = 23
b = 4
while a > b:
    if a % 2 == 0:
        b = b + a
    else:
        a = a - 2 * b + 1
```

```
print(b)
```

Ответ: 20

3. Определите, что будет напечатано в результате выполнения программы:

```
s = 0
m = 123
while m > 0:
    d = m % 10
    s = s + d
    m = m // 10
```

```
print(s)
```

Ответ: 6

4. Дан список из чисел и индекс элемента в списке k. Удалите из списка элемент с индексом k, сдвинув влево все элементы, стоящие правее элемента с индексом k. Программа получает на вход список, затем число k. Программа сдвигает все элементы, а после этого удаляет последний элемент списка при помощи метода pop() без параметров.

Программа должна осуществлять сдвиг непосредственно в списке, а не делать это при выводе элементов. Также нельзя использовать дополнительный список. Также не следует использовать метод pop(k) с параметром.

Решение:

```

a = [int(s) for s in input().split()]
k = int(input())
for i in range(k, len(a) - 1):
    a[i] = a[i + 1]
a.pop()
print(' '.join([str(i) for i in a]))

```

5. Предприятие производит оптовую закупку некоторых изделий А и В, на которую выделена определённая сумма денег. У поставщика есть в наличии партии этих изделий различных модификаций по различной цене. На выделенные деньги необходимо приобрести как можно больше изделий А независимо от модификации. Если у поставщика закончатся изделия А, то на оставшиеся деньги необходимо приобрести как можно больше изделий В. Известны выделенная для закупки сумма, а также количество и цена различных модификаций данных изделий у поставщика. Необходимо определить, сколько будет закуплено изделий В и какая сумма останется неиспользованной.

Входные данные.

Первая строка входного файла содержит два целых числа: N — общее количество партий изделий у поставщика и M — сумма выделенных на закупку денег (в рублях). Каждая из следующих N строк описывает одну партию и содержит два целых числа (цена одного изделия в рублях и количество изделий в партии) и один символ (латинская буква А или В), определяющий тип изделия. Все данные в строках входного файла отделены одним пробелом.

В ответе запишите два целых числа: сначала количество закупленных изделий типа В, затем оставшуюся неиспользованной сумму денег.

Пример входного файла:

```

4 1000
30 8 A
50 12 B
40 14 A
30 60 B

```

В данном случае сначала нужно купить изделия А: 8 изделий по 30 рублей и 14 изделий по 40 рублей. На это будет потрачено 800 рублей. На оставшиеся 200 рублей можно купить 6 изделий В по 30 рублей. Таким образом, всего будет куплено 6 изделий В и останется 20 рублей. В ответе надо записать числа 6 и 20.

Решение

```

f = open('26.txt')
x, y = f.readline().split()
y = int(y)
for _B_price = []
for _B_kol = []
for i in f:
    if 'A' in i:
        a, b, c = i.split()
        y -= int(a) * int(b)
    else:

```

```

a2, b2, c2 = i.split()
for_B_price.append(int(a2))
for_B_kol.append(int(b2))
mini = min(for_B_price)
index_mini = 0
for i in range(len(for_B_price)):
    if mini == for_B_price[i]:
        index_mini = i
kol_B = 0
while y > for_B_price[index_mini]:
    y -= for_B_price[index_mini]
    for_B_kol[index_mini] -= 1
    kol_B += 1
if for_B_kol[index_mini] == 0:
    for_B_price[index_mini] = 1000000000
    mini = min(for_B_price)
    for i in range(len(for_B_price)):
        if mini == for_B_price[i]:
            index_mini = i
print(kol_B, y)

```

Отметка	Число баллов, необходимое для получения отметки
«3» (удов.)	Выполнены три задания (№ 1, 2, 3)
«4» (хорошо)	Выполнены четыре задания (№ 1, 2, 3, 4) Выполнены два задания (№ 4, 5)
«5» (отлично)	Выполнены все пять заданий

Рубежный контроль

Рубежный контроль проходит в форме выполнения и защиты проектной работы
Проектная работа «Анализ больших данных в профессиональной сфере»

Основные этапам работы над проектом: определение проблемы в предметной области; извлечение данных; подготовка данных — очистка данных и преобразование данных; исследование и визуализация данных; построение предсказательной модели; интерпретация результатов.

Проект выполняется на примере открытых данных, например, Сбербанкка <https://www.kaggle.com/code/umkkka/eda-sberbank-open-data-russian-version/data>. Обучающиеся самостоятельно подбирают набор данных на <https://www.kaggle.com/> в профессиональной сфере и выполняют исследование и визуализацию данных.

Пример выполнение проекта «Прогнозирование инсульта»

Определение проблемы в предметной области. По данным Всемирной организации здравоохранения (ВОЗ) инсульт является 2-й ведущей причиной смерти в мире, ответственной примерно за 11% всех смертей. Определим вероятность развития инсульта

среди пациентов. Узнаем, что является фактором возникновения инсульта в большей степени.

Каждая строка данных содержит соответствующую информацию о пациенте.

id: уникальный идентификатор

gender: «Male», «Female» or «Other» - пол: «Мужской», «Женский» или «Другой»

age: age of the patient - возраст пациента

hypertension: 0, если у пациента нет гипертонии, 1, если у пациента гипертония

heart disease: 0, если у пациента нет сердечных заболеваний, 1, если у пациента сердечное заболевание

ever married: «No» or «Yes» – состоят ли в браке: «нет», «да»

work type: «children», «Govt job», «Never_worked», «Private» or «Self-employed» – тип работы: «с детьми», «работа в правительстве», «никогда не работал», «частное предпринимательство», «самозанятый».

Residence type: «Rural» or «Urban» – тип местности: сельский, городской.

Avg glucose level: average glucose level in blood - средний уровень глюкозы в крови.

bmi: body mass index – индекс массы тела

smoking status: «formerly smoked», «never smoked», «smokes» or «Unknown»* - «ранее курил», «никогда не курил», «курит» или «Неизвестно»*

stroke: 1, если у пациента был инсульт, или 0, если нет.

Note: «Unknown» (неизвестный) в состоянии курения означает, что информация недоступна для данного пациента.

С использованием набора данных проанализирована вероятность появления инсульта у пациента на основе входных параметров, таких как пол (gender), возраст (age), различные заболевания: гипертония (hypertension), сердечные заболевания (heart disease), а также статус курения (smoking status).

Извлечение данных.

В качестве основной библиотеки графики используется Plotly. Так как этот пакет с открытым исходным кодом, который предлагает сложные графики, которые могут удовлетворить требования любого проекта. Ведь одно из достоинств Plotly - это интерактивность исследования данных с помощью графиков и диаграмм.

Seaborn (sns) - библиотека визуализации данных Python, основанная на matplotlib. Она предоставляет высокоуровневый интерфейс для рисования привлекательной и информативной статистической графики. (Рис. 1)

```
B [2]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
plt.style.use('seaborn-whitegrid')

import plotly.express as px
import plotly.graph_objects as go
```

Рис. 1

После того, как скачаны и загружены данные из kaggle.com, следующим шагом стало создание DataFrame для панд на основе данных CSV. (Рис. 2)

```
B [3]: dataset = df=pd.read_csv(r'C:\Users\Светлана\Desktop\healthcare-dataset-stroke-data.csv')
```

Рис. 2

Необходимо, чтобы все столбцы (и числовые, и строковые) входных данных были включены в выходные данные. А выход будет варьироваться в зависимости от того, что предусмотрено. (Рис. 3)

В [4]: `dataset.describe(include = 'all')`

	я бы	Пол	возраст	гипертензия	болезнь сердца	когда-либо женат	work_type	Residence_type	avg_glucose_level	bmi	Курение_статус
считать	5110.000000	5110	5110.000000	5110.000000	5110.000000	5110	5110	5110	5110.000000	4909.000000	5110
никальный	NaN	3	NaN	NaN	NaN	2	5	2	NaN	NaN	4
вершина	NaN	женский	NaN	NaN	NaN	да	Частный	Городской	NaN	NaN	никогда не курит
частота	NaN	2994	NaN	NaN	NaN	3353	2925	2596	NaN	NaN	1894
иметь в виду	36517.829354	NaN	43.226614	0.097456	0.054012	NaN	NaN	NaN	106.147677	28.893237	NaN
часы	21161.721625	NaN	22.612647	0.296607	0.226063	NaN	NaN	NaN	45.283560	7.854067	NaN
мин	67.000000	NaN	0.080000	0.000000	0.000000	NaN	NaN	NaN	55.120000	10.300000	NaN
25%	17741.250000	NaN	25.000000	0.000000	0.000000	NaN	NaN	NaN	77.245000	23.500000	NaN
50%	36932.000000	NaN	45.000000	0.000000	0.000000	NaN	NaN	NaN	91.885000	28.100000	NaN
75%	54682.000000	NaN	61.000000	0.000000	0.000000	NaN	NaN	NaN	114.090000	33.100000	NaN
Максимум	72940.000000	NaN	82.000000	1.000000	1.000000	NaN	NaN	NaN	271.740000	97.600000	NaN

Рис. 3

Подготовка данных.

Проверим качество данных и типы каждого столбца. В данном коде содержится информация о фрейме данных, а также количество памяти, используемой данным кадром. (Рис. 4)

В [5]: `dataset.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5110 entries, 0 to 5109
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    5110 non-null   int64
1   gender                5110 non-null   object
2   age                   5110 non-null   float64
3   hypertension          5110 non-null   int64
4   heart_disease         5110 non-null   int64
5   ever_married          5110 non-null   object
6   work_type             5110 non-null   object
7   Residence_type        5110 non-null   object
8   avg_glucose_level     5110 non-null   float64
9   bmi                   4909 non-null   float64
10  smoking_status        5110 non-null   object
11  stroke                5110 non-null   int64
dtypes: float64(3), int64(4), object(5)
memory usage: 479.2+ KB
```

Рис. 4

С помощью метода `head()` можно запросить отображения первых пяти строк кадра данных. (Рис. 5)

```
В [7]: dataset.head()
```

	я бы	Пол	возраст	гипертония	болезнь сердца	когда-либо женат	work_type	Residence_type	avg_glucois_level	bmi	Курение_статус	stroke
0	9046	Мужчина	67.0	0	1	да	Частный	Городской	228.69	36.6	ранее курил	1
1	51676	женский	61.0	0	0	да	Частный предприниматель	Деревенский	202.21	NaN	никогда не курил	1
2	31112	Мужчина	80.0	0	1	да	Частный	Деревенский	105.92	32.5	никогда не курил	1
3	60182	женский	49.0	0	0	да	Частный	Городской	171.23	34.4	курит	1
4	1665	женский	79.0	1	0	да	Частный предприниматель	Деревенский	174.12	24.0	никогда не курил	1

Рис. 5

В проекте нужно узнать общее количество людей и их особенностей. (Рис. 6)

```
В [12]: print("Number of records (patients): ", df.shape[0])
        print("Number of columns (patients): ", df.shape[1])
```

```
Number of records (patients): 5110
Number of columns (patients): 12
```

Рис. 6

Исследование и визуализация данных.

Рассмотрим распределение пациентов по полу.

Основным элементом изображения, которое строит pyplot является фигура (Figure), на нее накладывается график, оси, надписи и т.д.

Для построения графика используется команда countplot() библиотеки Seaborn. В коде использовался размер графического элемента, был выбран столбец "Пол" в dataset, подписан заголовок и обозначен размер текста, введено название для x и y, и выведено на экран. (Рис. 7)

```
В [13]: plt.figure(figsize = (10, 6))
        sns.countplot(x = dataset['gender'])
        plt.title('Распределение пациентов по полу', fontsize = 30)
        plt.xlabel('Пол пациента')
        plt.ylabel('Количество пациентов')
        plt.show()
```

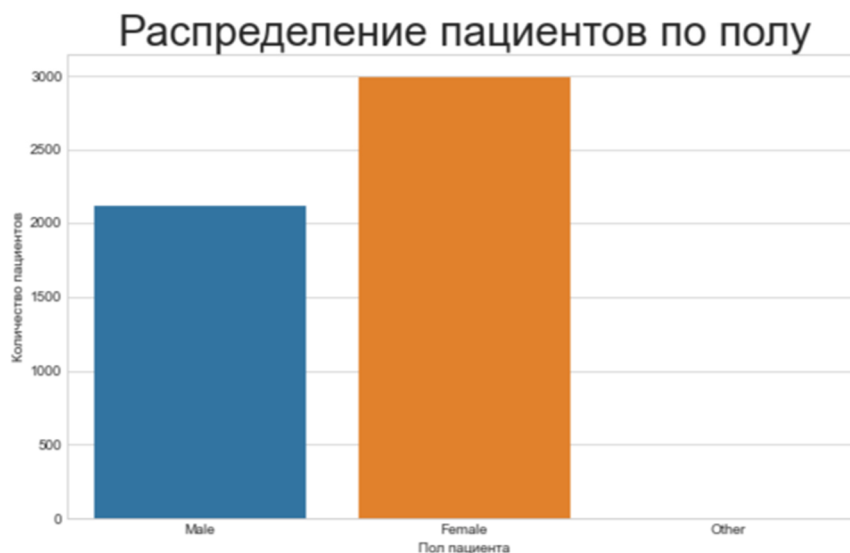


Рис. 7

По диаграмме можно увидеть, что женский пол преобладает, но в соотношении 100% не совсем понятно. Поэтому воспользуемся библиотекой `plotly.graph_objects` для создания круговой диаграммы. Для того, чтобы ее сделать необходимо использовать функцию `go.Pie`.

Для создания кода применяем атрибут `labels` значений "Пола пациента" из `dataset`. Задано наглядное поведение графика с помощью `update_traces`: а именно размер текста значений, цвет и жирность границ. А с помощью `update_layout` выведен текст заголовка. (Рис. 8)

```
In [14]: colors = ['RoyalBlue', 'crimson', 'green']
fig = go.Figure(data = [go.Pie(labels = dataset['gender'])])
fig.update_traces(textfont_size = 25, marker = dict(colors = colors, line = dict(color = '#000000', width = 3)))
fig.update_layout(title_text = 'Соотношение пациентов по полу')
fig.show()
```

Соотношение пациентов по полу

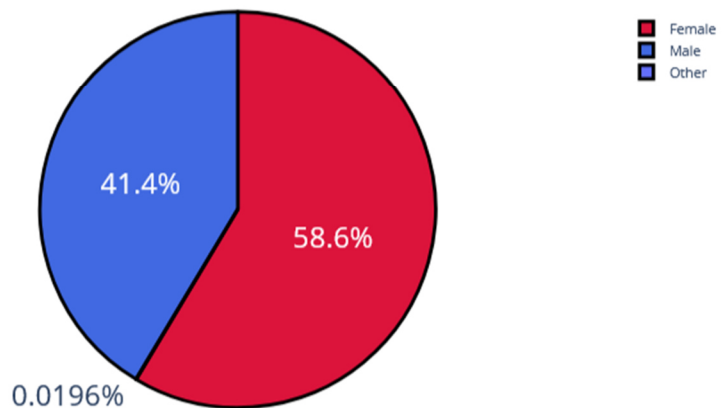


Рис. 8

Рассмотрим пациентов, у которых был и не был инсульт, с помощью графика. Вероятность появления второго инсульта очень велика, и он намного опаснее. (Рис. 9)

```
B [15]: plt.figure(figsize = (10, 6))
sns.countplot(x = dataset['stroke'])
plt.title('Был ли инсульт у пациентов?', fontsize = 30)
plt.xlabel('Инсульт')
plt.ylabel('Количество')
plt.show()
```



Рис. 9

Изучив столбчатую диаграмму, можно сделать вывод, что у большего количества пациентов не было инсульта. В группу риска опасного инсульта могут попасть пациенты,

у которых инсульт уже был. Рассмотрим этот же вопрос с помощью круговой диаграммы. (Рис. 10)

```
B [16]: colors = ['crimson', 'green']
fig = go.Figure(data = [go.Pie(labels = dataset['stroke'])])
fig.update_traces(textfont_size = 25, marker = dict(colors = colors, line = dict(color = '#000000', width = 3)))
fig.update_layout(title_text = 'Был ли инсульт у пациентов?')
fig.show()
```

Был ли инсульт у пациентов?

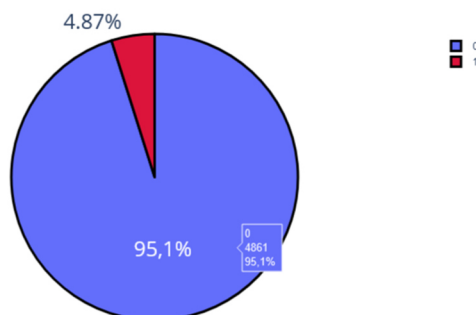


Рис. 10

Проанализируем данные о сердечных заболеваниях сразу с помощью диаграмм, так как они более лучше помогают визуализировать примеры таких задач. (Рис. 11)

```
fig = go.Figure(data = [go.Pie(labels = dataset['heart_disease'])])
fig.update_traces(textfont_size = 25, marker = dict(colors = colors, line = dict(color = '#000000', width = 3)))
fig.update_layout(title_text = 'Было ли сердечное заболевание?')
fig.show()
```

Было ли сердечное заболевание?

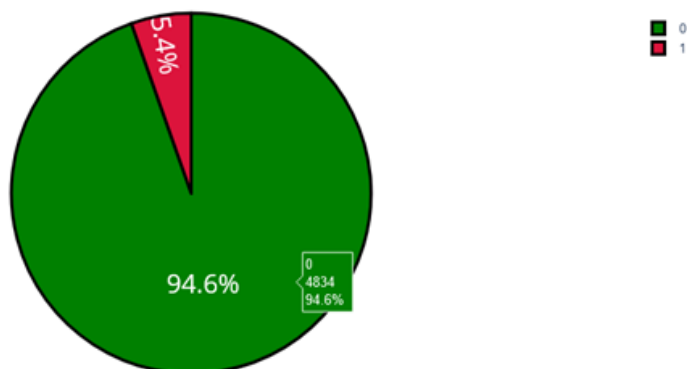


Рис. 11

Одна из форм инсульта может быть вызвана гипертонией, проверим с помощью диаграммы количество людей с наличием гипертонии. (Рис. 12)

```
fig = go.Figure(data = [go.Pie(labels = dataset['hypertension'])])
fig.update_traces(textfont_size = 25, marker = dict(colors = colors, line = dict(color = '#000000', width = 3)))
fig.update_layout(title_text = 'Была ли гипертония?')
fig.show()
```

Была ли гипертония?

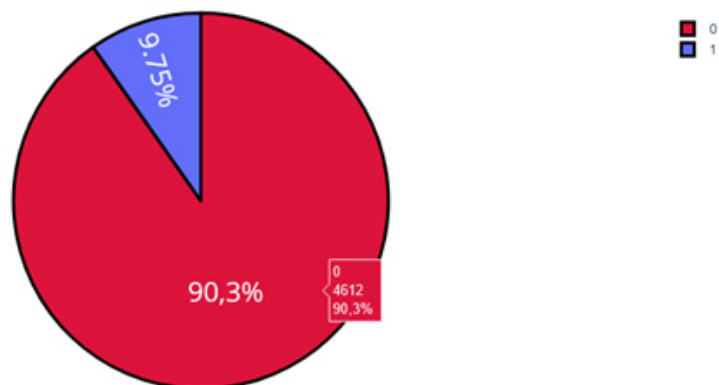


Рис. 12

Ученые из Хельсинкского университета выяснили, что одинокие люди страдают от инсультов на 39% чаще, а люди в браке, у которых даже был инсульт, восстанавливаются после инсульта стабильно хорошо и быстрее, в отличие, от одиноких людей. Можно проверить: может ли брак повлиять на появление инсульта.

Рассмотрим, количество людей женаты/замужем. (Рис. 13)

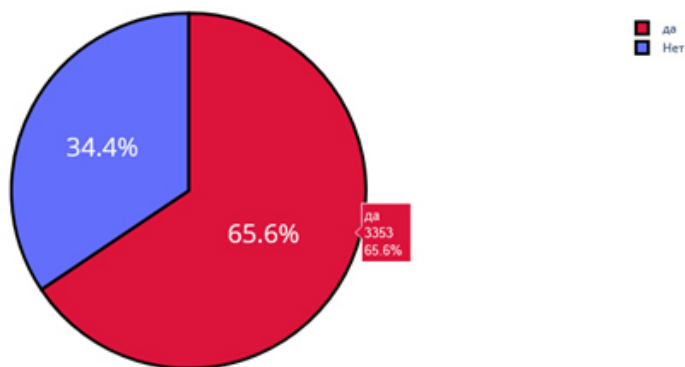


Рис. 13

Проверим тип работы у пациентов. (Рис. 14)

```

fig = go.Figure(data = [go.Pie(labels = dataset['work_type'])])
fig.update_traces(textfont_size = 25, marker = dict(colors = colors, line = dict(color = '#000000', width = 3)))
fig.update_layout(title_text = 'Тип работы у пациентов')
fig.show()

```

Тип работы у пациентов

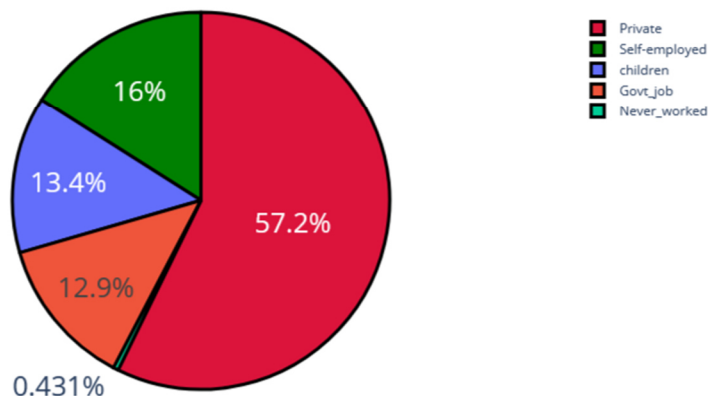


Рис. 14

Проверим: влияет ли тип местности на возникновение инсульта.

Ведь с началом летнего периода многие люди, живущие в квартирах, перебираются на дачные участки, в том числе люди старшего поколения. Большинство не только отдыхают на природе, но и поддерживают порядок, ухаживают за растениями. И в результате, особенно для пожилых, с не привыкшим к таким усилиям после зимнего периода, существенно возрастает опасность внезапных инсультов. (Рис. 15)

```
fig = go.Figure(data = [go.Pie(labels = dataset['Residence_type'])])
fig.update_traces(textfont_size = 25, marker = dict(colors = colors, line = dict(color = '#000000', width = 3)))
fig.update_layout(title_text = 'Тип местности у пациентов')
fig.show()
```

Тип местности у пациентов

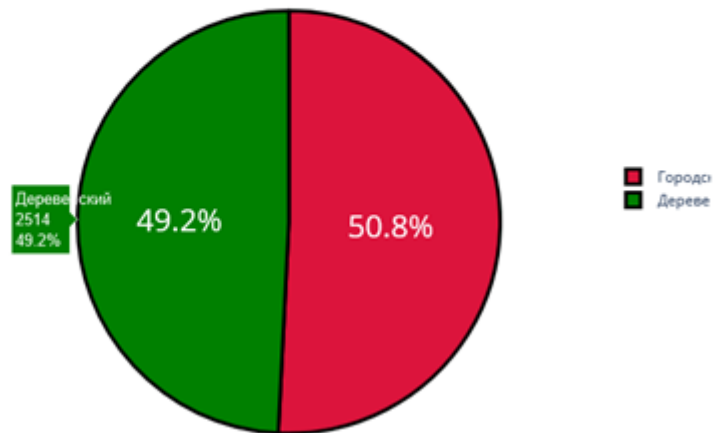


Рис. 15

Согласно результатам последних научных исследований, отказ от сигарет помогает снизить риск инсульта на 12,4%. Пассивное курение повышает риск инсульта на 20–30%. Ежегодно от инсульта погибают 8 000 пассивных курильщиков. Проверим отношение людей к курению. (Рис. 16)

```
В [23]: plt.figure(figsize = (10, 4))
sns.countplot(x = dataset['smoking_status'])
plt.title('Отношение пациентов к курению', fontsize = 30)
plt.xlabel('Курение')
plt.ylabel('Количество')
plt.show()
```

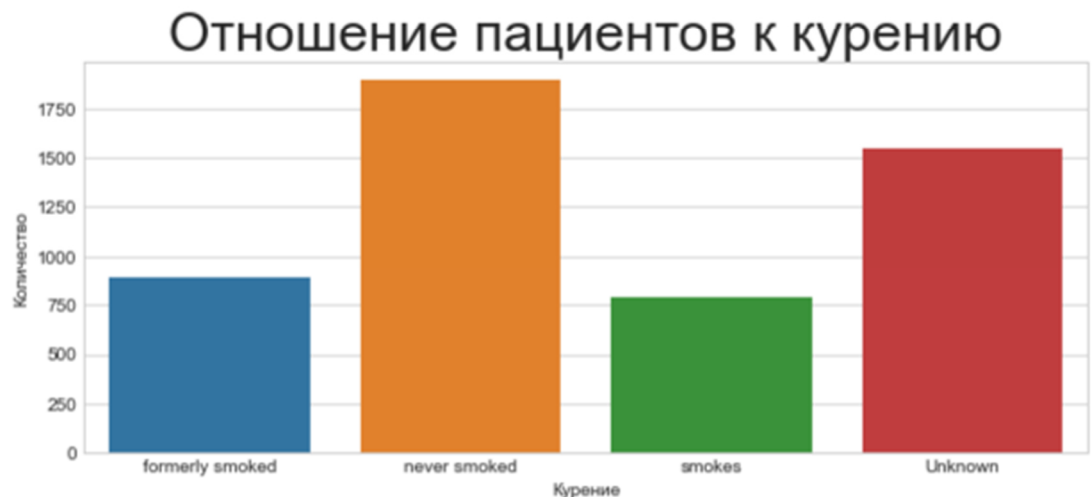


Рис. 16

В графике "Отношение пациентов к курению" наглядно было видно кого больше, кого меньше. Но необходимо узнать, сколько людей "никогда не курил", "неизвестная информация", "курил ранее/бросил", "курит".

Добавочный код `.value_counts` позволяет выполнить подсчет количества элементов в столбце. А при написании кода `px.bar`: каждая строка фрейма данных будет представлена в виде прямоугольной метки. (Рис. 17)

```

В [24]: df = pd.DataFrame(dataset['smoking_status'].value_counts())
px.bar(x = df.index, y = df.smoking_status, height = 400, text = df.smoking_status, labels = {'x': 'Статус курения', 'y': 'Количество'},
       title = 'Распределение пациентов по отношению к курению')

```



Рис. 17

Определим по половому признаку и наличию сердечного заболевания, зависимость от типа работы пациента. Для этого добавим в обычный код для построения гистограммы функцию `seaborn sns.barplot()`, который очень удобно рисует барплот с разными по размеру линиями погрешностями. (Рис. 18)

```

В [25]: plt.figure(figsize = (10, 8))
sns.barplot(x = dataset['gender'], y = dataset['heart_disease'], hue = dataset['work_type'])
plt.legend(loc = 'upper right')
plt.title('Зависимость типа работы от пола пациента и наличия у него сердечного заболевания', fontsize = 20)
plt.xlabel('Пол человека')
plt.ylabel('Наличие сердечного заболевания')
plt.show()

```

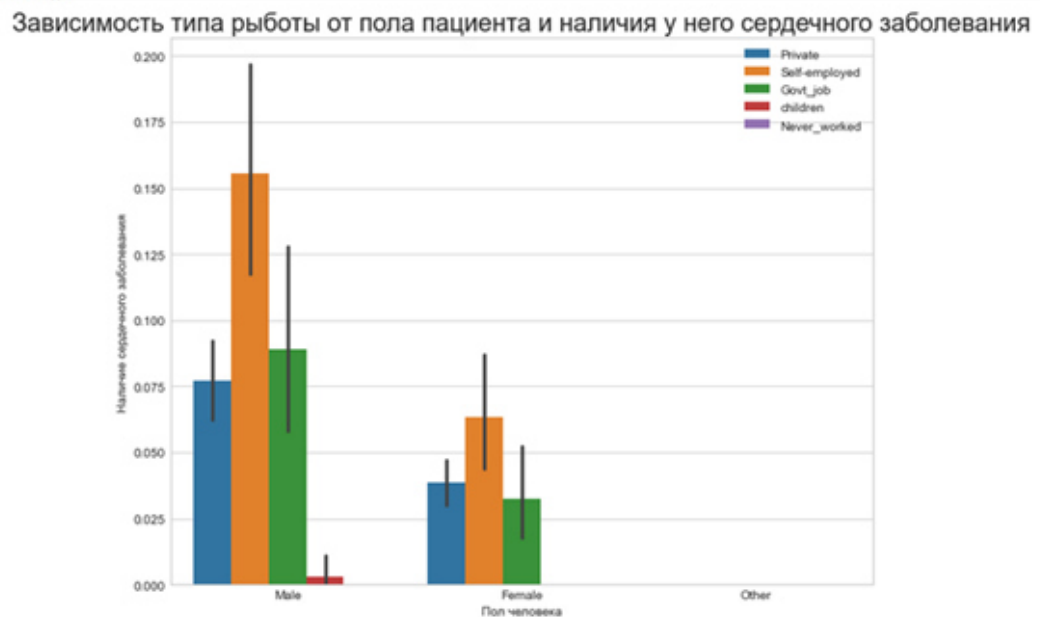


Рис. 18

Определим по половому признаку и наличию сердечного заболевания, зависимость от отношения к курению пациента. (Рис. 19)

```
[26]: plt.figure(figsize = (10, 7))
sns.barplot(x = dataset['gender'], y = dataset['heart_disease'], hue = dataset['smoking_status'])
plt.legend(loc = 'upper right')
plt.title('Зависимость отношения к курению от пола пациента и наличия сердечного заболевания', fontsize = 20)
plt.xlabel('Пол человека')
plt.ylabel('Наличие сердечного заболевания')
plt.show()
```

Зависимость отношения к курению от пола пациента и наличия сердечного заболевания

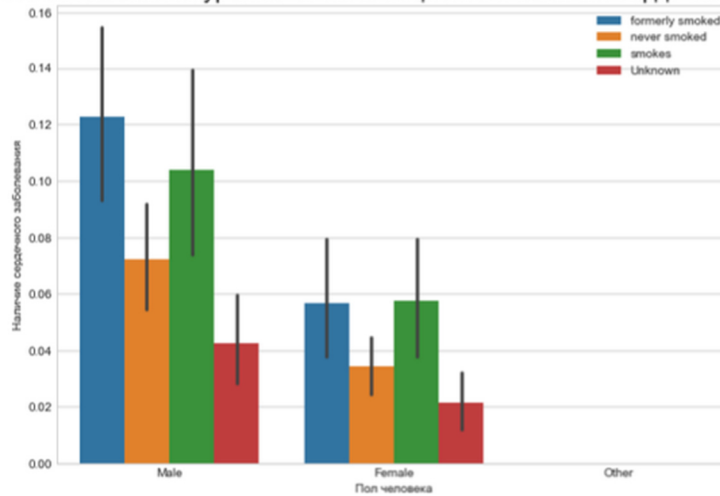


Рис. 19

Представим визуализацию парных отношений, которая является самой наглядной. Для того, чтобы нарисовать совместное распределение всех переменных в наборе данных воспользуемся функцией `pairplot()`.

По умолчанию, парные распределения (те что под и над главной диагональю сетки) изображаются графиками разброса, а одномерные распределения (те что на главной диагонали) изображаются гистограммами.

Функция `pairplot()` содержит класс `PairGrid`, который строит сетку для решетчатого графика (Рис. 20)

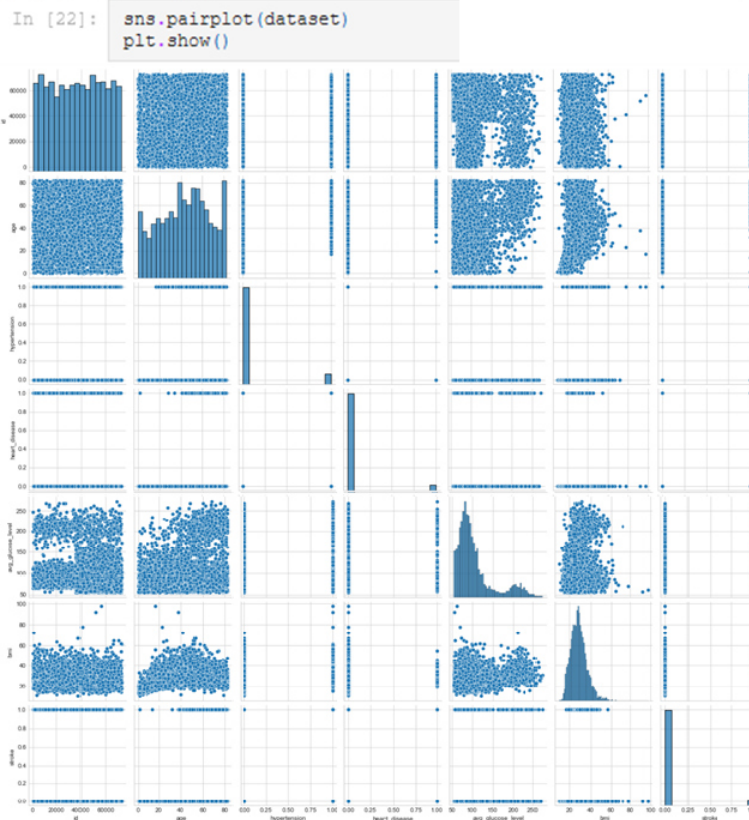


Рис. 20

Индекс массы тела влияет на состояние здоровья, а также может влиять на развитие появления инсульта.

Классификация массы тела, ИМТ и риск развития инсульта:

1. Дефицит массы тела с ИМТ $<18,5$ имеет низкий риск.
2. Нормальная масса тела с ИМТ $18,5-24,9$ имеет обычный риск.
3. Избыточная масса тела с ИМТ $25,0-29,9$ имеет повышенный риск.
4. Ожирение I степени с ИМТ $30,0-34,9$ имеет высокий риск.
5. Ожирение II степени с ИМТ $35,0-39,9$ имеет очень высокий риск.
6. Ожирение III степени с ИМТ >40 имеет чрезвычайно высокий риск.

С помощью функции `histogram()` можно вычислить гистограмму набора данных. Воспользуемся последовательностью целых чисел, для этого выберем параметр выбрала `bins`, который используем вместе с методом расчета ширины `range`.

Для наглядности сформируем формулу для среднего значения ИМТ. А при написании кода `px.bar`: каждая строка фрейма данных будет представлена в виде прямоугольной метки. (Рис. 21)

```

: counts, bins = np.histogram(dataset.bmi, bins=range(0, 100, 5))
bins = 0.5 * (bins[:-1] + bins[1:])

fig = px.bar(x = bins, y = counts, labels = {'x': 'Индекс массы тела', 'y': 'Количество'}, title = 'Индекс массы т
fig.show()

```

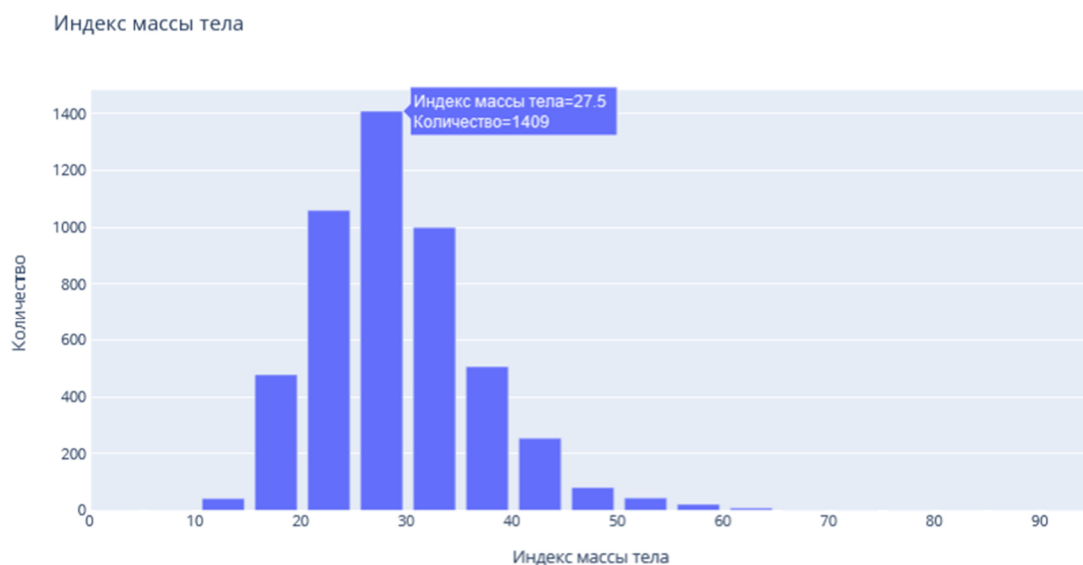


Рис. 21

Определим возрастную категорию людей, которые являются пациентами. (Рис. 22)

```

counts, bins = np.histogram(dataset.age, bins=range(0, 90, 5))
bins = 0.5 * (bins[:-1] + bins[1:])

fig = px.bar(x = bins, y = counts, labels = {'x': 'Возраст', 'y': 'Количество'}, title = 'Возраст пациентов')
fig.show()

```

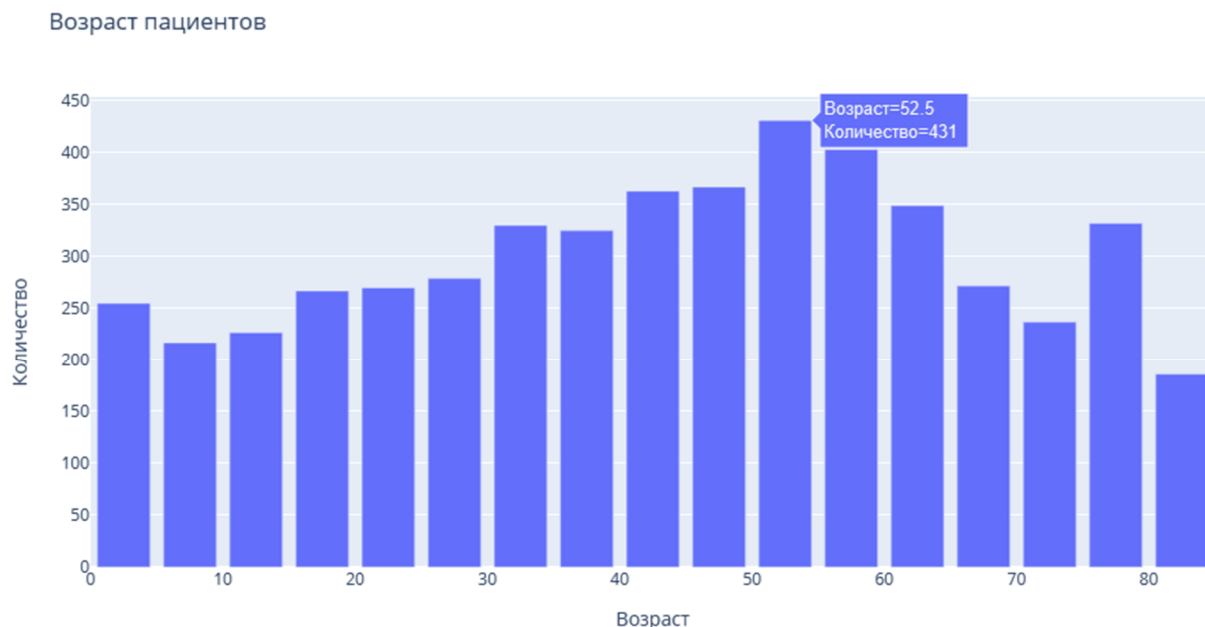


Рис. 22

Чаще всего пациентами становятся люди от 50 до 55 лет.

Определим, зависит тип проживания от ИМТ и типа работы. (Рис. 23)

```
px.box(x = dataset['bmi'], y = dataset['work_type'], color = dataset['Residence_type'], labels = {'x': 'Индекс массы тела', 'y': 'Тип работы'}, title = 'Зависимость типа проживания от индекса массы тела и типа работы')
```



Рис. 23

Определим, зависит пол пациента от ИМТ и типа работы. (Рис. 24)

```
px.box(x = dataset['bmi'], y = dataset['work_type'], color = dataset['gender'], labels = {'x': 'Индекс массы тела', 'y': 'Тип работы'}, title = 'Зависимость пола пациента от индекса массы тела и типа работы')
```

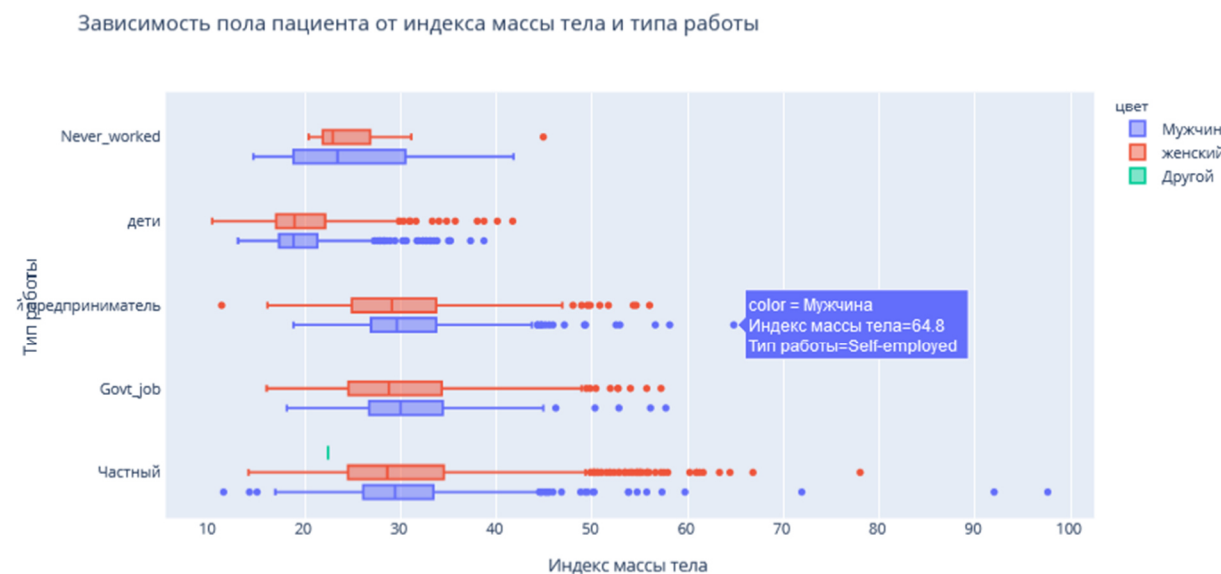


Рис. 24

Функцию Plotly Express `density_contour()` можно использовать для создания контуров плотности (2д контурная гистограмма).

Она напоминает контур участка, вычисляется путем группирования множества точек и применения статистической функции.

Этот вид визуализации часто используется для управления чрезмерным построением графика или ситуациями, когда большие наборы данных отображаются в

виде точечных диаграмм, которые приводят к тому, что точки будут перекрывать друг друга и скрывать узоры.

Построим график индекса массы тела на основе сердечных заболеваний. (Рис. 25)

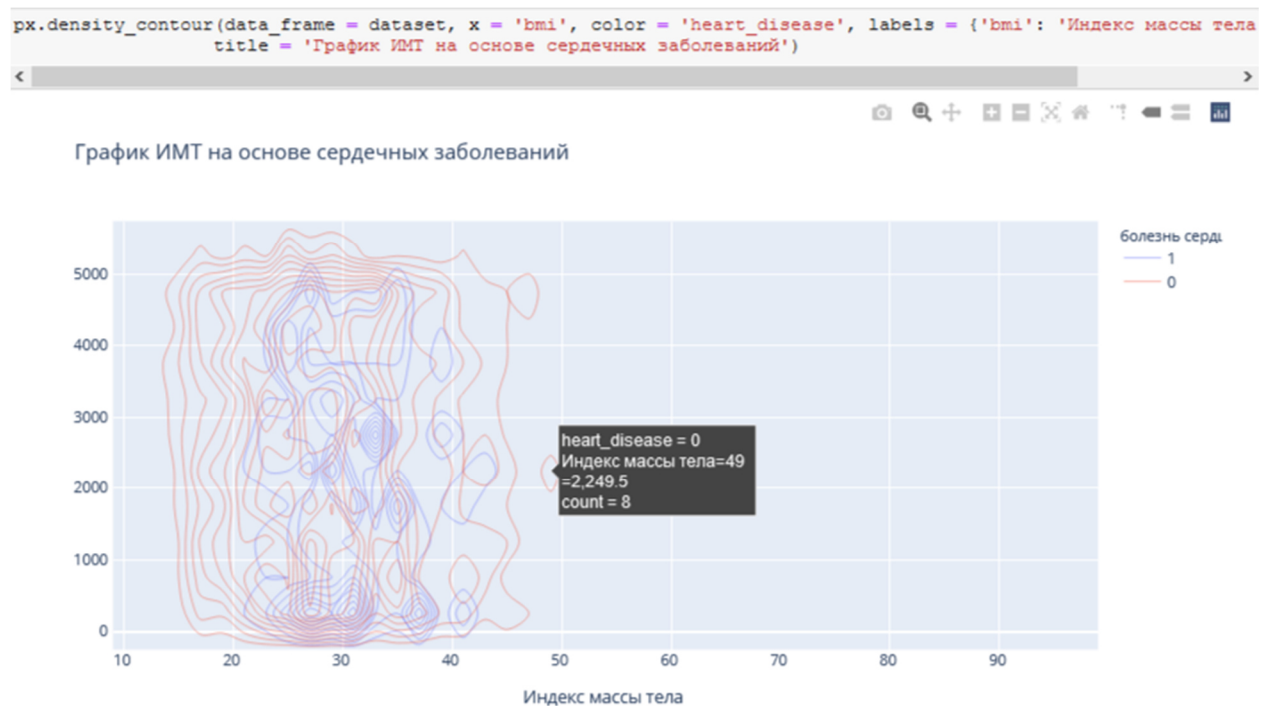


Рис. 25

Построим график индекса массы тела на основе сердечных заболеваний. (Рис. 26)

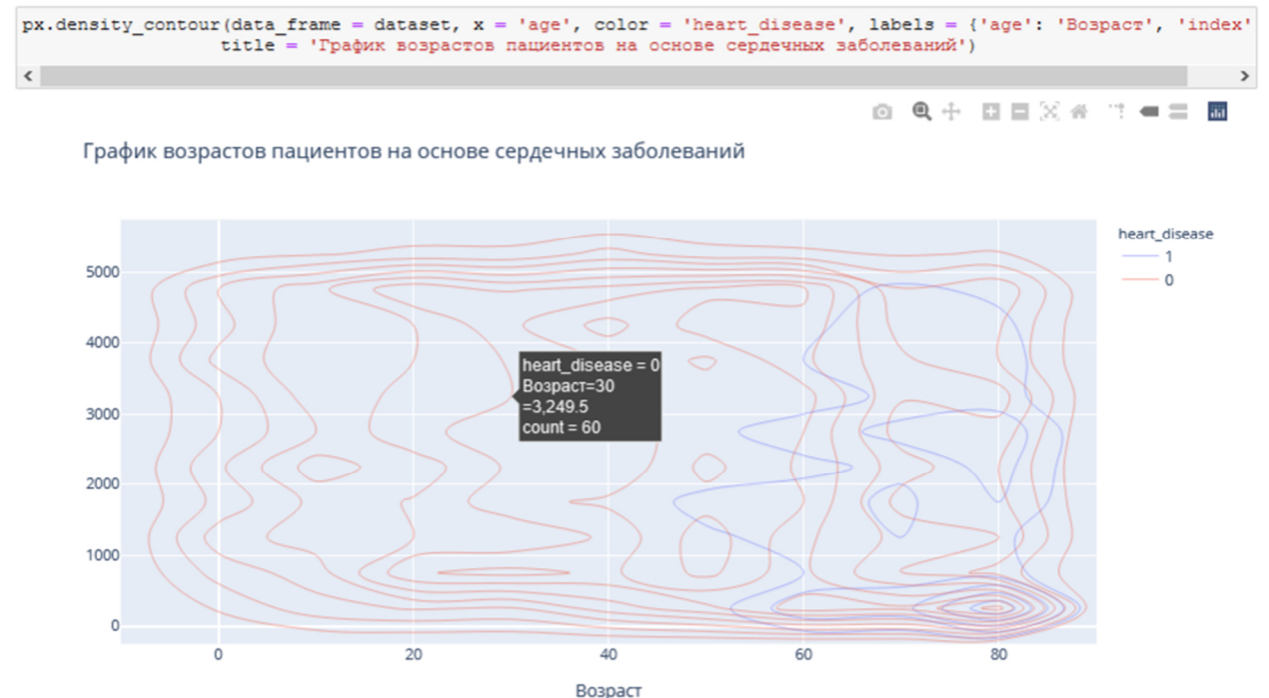


Рис. 26

Преобразуем переменные в фиктивные коды для значений в столбцах: Пол, состоит в браке, тип работы, тип местности, отношение к курению. Функция concat выполняет всю тяжелую работу по выполнению операций конкатенации вдоль оси. (Рис. 27)

```

: gender = pd.get_dummies(dataset['gender'], drop_first = True)
married = pd.get_dummies(dataset['ever_married'], drop_first = True)
work = pd.get_dummies(dataset['work_type'], drop_first = True)
residence = pd.get_dummies(dataset['Residence_type'], drop_first = True)
smoking = pd.get_dummies(dataset['smoking_status'], drop_first = True)
dataset = pd.concat([gender, married, work, residence, smoking, dataset], axis = 1)

```

Рис. 27

Удалим указанные метки из указанных столбцов и выведем первые 5 записей. Будем использовать функцию head(), чтобы посмотреть, как это будет выглядеть. (Рис. 28)

```
dataset.drop(['gender', 'ever_married', 'work_type', 'Residence_type', 'smoking_status'], axis = 1, inplace = True)
dataset.head()
```

	Male	Other	Yes	Never_worked	Private	Self-employed	children	Urban	formerly smoked	never smoked	smokes	id	age	hypertension	heart_disease	avg_glucose_level	bmi	stroke
0	1	0	1	0	1	0	0	1	1	0	0	9046	67.0	0	1	228.69	36.6	1
1	0	0	1	0	0	1	0	0	0	1	0	51676	61.0	0	0	202.21	NaN	1
2	1	0	1	0	1	0	0	0	0	1	0	31112	80.0	0	1	105.92	32.5	1
3	0	0	1	0	1	0	0	1	0	0	1	60182	49.0	0	0	171.23	34.4	1
4	0	0	1	0	0	1	0	0	0	1	0	1665	79.0	1	0	174.12	24.0	1

Рис. 28

Проверим, есть ли нулевые значения в каждом столбце. (Рис. 29)

```
dataset.isnull().sum()
```

```

Male                0
Other               0
Yes                 0
Never_worked        0
Private             0
Self-employed       0
children            0
Urban               0
formerly smoked     0
never smoked        0
smokes              0
id                  0
age                 0
hypertension        0
heart_disease        0
avg_glucose_level   0
bmi                 201
stroke              0
dtype: int64

```

Рис. 29

В данном наборе данных нет нулевых значений.

Функции содержат ошибки ввода (3-я запись сверху). Поэтому переименуем имя столбца с ошибкой "Yes". (Рис. 30)

```
dataset.rename(columns = {'Yes': 'ever_married'}, inplace = True)
dataset.head()
```

	Male	Other	ever_married	Never_worked	Private	Self-employed	children	Urban	formerly smoked	never smoked	smokes	id	age	hypertension	heart_disease	avg_glucose_level	bmi	stroke
0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	9046.0	67.0	0.0	1.0	228.69	36.60	1.0
1	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	51676.0	61.0	0.0	0.0	202.21	28.08	1.0
2	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	31112.0	80.0	0.0	1.0	105.92	32.50	1.0
3	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	60182.0	49.0	0.0	0.0	171.23	34.40	1.0
4	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1665.0	79.0	1.0	0.0	174.12	24.00	1.0

Рис. 30

Выведем визуализацию матричной модели. Чтобы наглядно посмотреть вероятность появления инсульта при всех входных параметрах. (Рис. 31)

Для этого воспользуюсь библиотекой Seaborn с функцией: `heatmap()`, которая используется для создания тепловых карт. В качестве первого аргумента, используя dataset воспользуемся методом `corr()`, и для точной интерпретации добавим аргумент `annot=True`, который помогает отобразить коэффициент корреляции. Добавим цвет, используя аргумент `cmap`, добавим цвет разделяющей линии, используя `linecolor`, добавим жирность разделяющей линии, используя `linewidths`.

```
plt.figure(figsize = (12, 8))
sns.heatmap(dataset.corr(), linecolor = 'white', linewidths = 3, cmap= 'RdPu', annot = True)
plt.show()
```

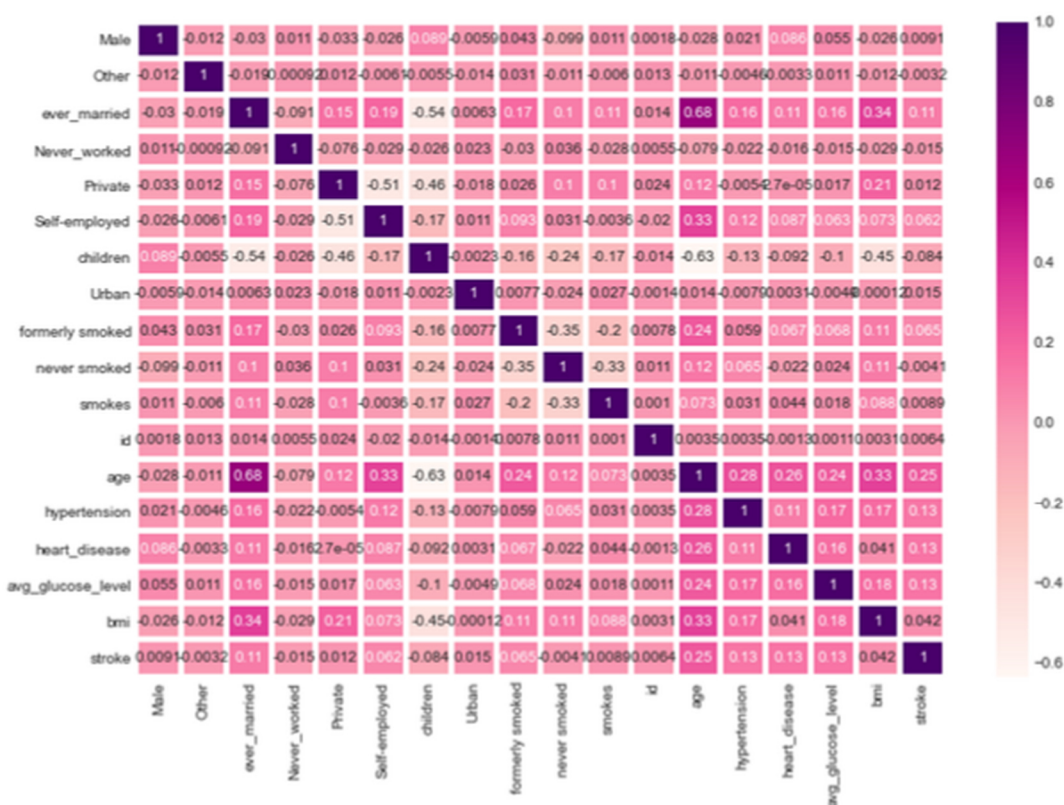


Рис. 31

Таким образом, по матричной модели можно выделить, например, три группы, которые входят в зону риска возникновения инсульта:

1. Возраст пациента от 50 до 90 лет и индекс массы тела которого равен от избыточной массы тела до ожирения III степени.
2. Возраст пациента от 50 до 90 лет и наличие гипертонии.
3. Возраст пациента от 50 до 90 лет и наличие сердечных заболеваний.

Благодаря этому проекты можно сделать выводы:

1. По диаграмме можно увидеть, что пациентов женщин – 58,6% (2994 человека) больше, чем мужчин – 41,4% (2115 человек).
2. Рассмотрели пациентов, у которых был инсульт, их – 4,87% (249 человек), а тех, у кого не было инсульта – 95,1% (4861 человек).
3. Рассмотрели пациентов, у которых есть сердечные заболевания, их оказалось 5,4% (276 человек), у кого нет – 94,6% (4834 человек).
4. Одна из форм инсульта может быть вызвана гипертонией, поэтому проверили количество людей с наличием гипертонии – 9,75% (498 человек), людей, которые не страдают гипертонией – 90,3% (4612 человек)

5. Ученые из Хельсинкского университета выяснили, что одинокие люди страдают от инсультов чаще, а люди в браке, у которых даже был инсульт, восстанавливаются после инсульта быстрее. Поэтому проверили: какое количество людей женаты/замужем – 65,6% (3353 человека), а люди, которые не состояли в браке – 34,4% (1757 человек).

6. Тип работы также влияет на появление инсульта. Проверив тип работы, выяснили, что людей, работающих с детьми – 13,4% (687 человек), которые работают в правительстве – 12,9% (657 человек), тех, кто никогда не работал – 0,4% (22 человека), частные предприниматели – 57,2% (2925 человека), самозанятых – 16% (819 человек).

7. Проверили, влияет ли тип местности на возникновение инсульта, ведь с началом летнего периода многие люди, живущие в квартирах, перебираются на дачные участки, в том числе люди старшего поколения. Большинство не только отдыхают на природе, но и поддерживают порядок, ухаживают за растениями. И в результате, особенно для пожилых, с не привыкшим к таким усилиям после зимнего периода, существенно возрастает опасность внезапных инсультов. Людей, живущих в сельской местности – 49,2% (2514 человека), а в городской – 50,8% (2596 человек).

8. Согласно результатам последних научных исследований, отказ от сигарет помогает снизить риск инсульта на 12,4%. Пассивное курение повышает риск инсульта на 20–30%. Ежегодно от инсульта погибают 8 000 пассивных курильщиков. Проверив отношение людей к курению, выяснили, что людей, которые никогда не курили больше (1892 человека), на втором месте – «неизвестно» (то есть информация недоступна) (1544 человека), на третьем месте – раньше курил (885 человек) и на последнем месте – курит (789 человек).

9. Определим по половому признаку и наличию сердечного заболевания, зависимость от типа работы пациента, выяснили, что наличие сердечного заболевания чаще встречается у мужчин, которые самозаняты – 1 место, работающих в правительстве – 2 место, частные предприниматели – 3 место и работа с детьми – 4 место. У женщин наличие сердечного заболевания, встречается намного реже. Но все равно оно есть у девушек, которые самозаняты – 1 место, частные предприниматели – 2 место, работающих в правительстве – 3 место.

10. Определим по половому признаку и наличию сердечного заболевания, зависимость от отношения к курению пациента, выяснили, что наличие сердечного заболевания чаще встречается у мужчин, которые ранее курили или бросили – 1 место, курит – 2 место, никогда не курили – 3 место, и неизвестно – 4 место. У женщин точно также распределились места, но наличие сердечного заболевания, встречается намного реже.

11. Индекс массы тела влияет на состояние здоровья, а также может влиять на развитие появления инсульта. Выяснили, что 520 человек имеют низкий риск развития инсульта из-за дефицита массы тела (ИМТ <18,5); 1060 человек имеют обычный риск из-за нормальной массы тела (ИМТ 18,5-24,9.), 1409 человек имеют повышенный риск из-за избыточной массы тела (ИМТ 25,0-29,9.), 1000 человек имеют высокий риск из-за ожирения I степени (ИМТ 30,0-34,9), 506 человек имеют очень высокий риск из-за ожирения II степени (ИМТ 35,0-39,9), 409 человек имеет чрезвычайно высокий риск из-за ожирения III степени (ИМТ >40).

12. Определили возрастную категорию людей: от 50 до 55 лет, которые являются пациентами.

13. Определили, как зависит тип проживания от ИМТ и типа работы, к примеру, средняя масса тела у неработающих людей в сельской местности – 28,4 (избыточная масса), а в городской местности – 22,9 (нормальная масса тела), у городских жителей вес в норме, в отличие от сельских.

14. Определили, как зависит пол пациента от ИМТ и типа работы, к примеру, рассмотрим работу с детьми. У мужчин средняя масса тела – 18,8 (дефицит массы тела), а у женщин – 18,9 (дефицит массы тела). Но избыточный вес преобладает у женщин, чем у мужчин.

15. Сердечное заболевание встречается чаще всего у людей, которых повышен индекс массы тела.

16. Сердечное заболевание встречается чаще всего у людей в возрасте от 50 до 90.

3. Создав матричную модель, можно выделить, например, три группы, которые входят в зону риска возникновения инсульта: возраст пациента от 50 до 90 лет и индекс массы тела которого равен от избыточной массы тела до ожирения III степени; возраст пациента от 50 до 90 лет и наличие гипертонии; возраст пациента от 50 до 90 лет и наличие сердечных заболеваний.

Промежуточный контроль

Выполнение проекта «Создание синквейнов и визуальной карты знаний по машинному обучению»

Задания проекта нацелены на контроль усвоения основных терминов и понятий по курсу

Требование к синквейну

1. Содержание синквейна должно соответствовать одной из 6 тем курса (понятие машинного обучения, линейная регрессия, логистическая регрессия, деревья решений, случайный лес, задача кластеризации). Тема курса для составления синквейна выбирается случайным образом или назначается преподавателем.

2. Синквейн должен раскрывать содержание одной из 6 тем курса.

3. Синквейн состоит из пяти строк, каждая из которых имеет свое содержание:

- Первая строка — тема синквейна, включает в себе одно-два слова, которые обозначают объект или понятие, о котором пойдет речь.

- Вторая строка — два слова (чаще всего прилагательные или причастия), которые описывают признаки и свойства выбранного в синквейне предмета или объекта.

- Третья строка — образована тремя глаголами или деепричастиями, описывающими характерные действия объекта.

- Четвертая строка — фраза из четырёх слов, выражающая личное отношение автора синквейна к описываемому предмету или объекту.

- Пятая строка — одно слово, характеризующее суть предмета или объекта

Критерии оценивания. Задание считается выполненным, если разработанный синквейн соответствует требованиям, а результаты представлены в ходе публичной презентации синквейна.

Шкала итоговой оценки

Шкала	Критерии
зачтено	Структура и содержание синквейна соответствуют требованиям. Отсутствуют содержательные ошибки. Имеется не более трех незначительных неточностей. В ходе презентации синквейна обучающийся правильно ответил на уточняющие вопросы
не зачтено	Структура и содержание синквейна не соответствуют требованиям. Имеются содержательные ошибки. В ходе презентации синквейна обучающийся отвечал с ошибками/не отвечал на уточняющие вопросы

Требования к визуальной карте знаний

1. Содержание визуальной карты знаний должно соответствовать содержанию курса.

2. Должны быть выделены ключевые понятия курса.

3. Должна быть установлена взаимосвязь между ключевыми понятиями курса с помощью стрелок.

4. Используемая система знаков или визуальных образов должна отражать сущность ключевого понятия курса.

Критерии оценивания. Задание считается выполненным, если разработанная визуальная карта знаний соответствует требованиям, а результаты представлены в ходе ее публичной презентации.

Шкала итоговой оценки

Шкала	Критерии
зачтено	Структура и содержание визуальной карты знаний соответствуют требованиям. Корректно выделены ключевые понятия курса, а также установлены взаимосвязи между ними. Выбранная система знаков и/визуальных образов отражает сущность ключевых понятий. Отсутствуют содержательные ошибки. Имеется не более трех незначительных неточностей. В ходе презентации визуальной карты знаний обучающийся правильно ответил на уточняющие вопросы
не зачтено	Структура и содержание визуальной карты знаний не соответствуют требованиям. Выделенные ключевые понятия не соответствуют содержанию курса, имеются неверно установленные взаимосвязи между этими понятиями или отсутствуют взаимосвязи между ключевыми понятиями. Система знаков и/визуальных образов выбрана случайным образом и не отражает сущность ключевых понятий. Имеются содержательные ошибки. В ходе презентации синквейна обучающийся отвечал с ошибками/не отвечал на уточняющие вопросы

Рубежный контроль

Выполнение проекта «Разработка модели машинного обучения для решения задачи классификации»

Постановка задачи: исследование файла с данными о грибах, который содержит более 8000 строк описания параметров различных грибов, включающее описание формы и цвета шляпки, формы ножки, цвет спор и т.д. Задача машинного обучения – научиться отличать съедобные грибы от ядовитых.

Содержание самостоятельной работы обучающихся (План работы над проектом)

1. Изучение, анализ и преобразование данных.
2. Выбор подходящей модели, ее импорт и обучение.
3. Оценка качества работы модели.
4. Рефлексия.
5. Разработка презентации.
6. Выступление.
7. Самооценка и взаимооценка

Требования к результату проектной деятельности

Разработка модели

- Выполнены все этапы разработки модели.
- Все команды выполняются корректно, отсутствуют ошибки при запуске кода.

- Отсутствуют фактические и логические ошибки в коде программы.
- Точность работы программы на тестовом датасете составляет не менее 97%

Оформление блокнота

- Код сопровождается текстовыми комментариями
- Вывод результатов оценки модели оформлен с текстовыми пояснениями

Презентация и выступление

Презентация включает:

- Этапы работы над проектом:
- анализ данных и принцип их отбора и обработки;
- аргументация выбора модели и процесс ее создания;
- аргументация выбора метрики для оценки модели, результаты оценки;
- шаги по доработке модели после оценки (если они были необходимы);
- Рефлексия: с какими проблемами столкнулись участники команды во время разработки модели, как они их преодолели.
- Оформление презентации соответствует общим требованиям к дизайну визуальных продуктов.
- Выступление содержательно, при этом выдержан регламент выступления (5-7 минут).

Шкала итоговой оценки

Шкала	Критерии
зачтено	Структура и содержание выполненного проекта полностью соответствуют требованиям: все команды выполняются корректно, отсутствуют ошибки при запуске кода, а также отсутствуют фактические и логические ошибки в коде программы, точность работы программы на тестовом датасете составляет не менее 97%; код сопровождается текстовыми комментариями, вывод результатов оценки модели оформлен с текстовыми пояснениями. Презентация соответствует требованиям, выступление содержательно, при этом выдержан регламент выступления
не зачтено	Структура и содержание выполненного проекта не соответствуют требованиям: команды выполняются недостаточно корректно, присутствуют ошибки при запуске кода, а также имеются фактические и логические ошибки в коде программы, точность работы программы на тестовом датасете составляет менее 97%; код не сопровождается текстовыми комментариями, вывод результатов оценки модели не оформлен с текстовыми пояснениями. Презентация не соответствует требованиям, выступление содержательно, но при этом не выдержан регламент выступления