

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»

УТВЕРЖДАЮ
Директор по образовательной
деятельности

М.И.И. С.Т. Князев
«7» сентября 2023



РАБОЧАЯ ПРОГРАММА МОДУЛЯ

Код модуля	Модуль
1164099	Аналитика и визуализация данных

Екатеринбург

Перечень сведений о рабочей программе модуля	Учетные данные
Образовательная программа Прикладной искусственный интеллект	Код ОП 09.03.03
Направление подготовки Прикладная информатика	Код направления и уровня подготовки 09.03.03

Программа модуля составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Медведева Марина Александровна	кандидат физико- математических наук, доцент	Доцент	Базовая кафедра «Аналитика больших данных и методы видеоанализа»
2	Балунгу Даниель		ассистент	Базовая кафедра «Аналитика больших данных и методы видеоанализа»

1. ОБЩАЯ ХАРАКТЕРИСТИКА МОДУЛЯ Аналитика и визуализация данных

1.1. Аннотация содержания модуля

Модуль «Аналитика и визуализация данных» состоит из одноименной дисциплины. Собрать и обработать данные - лишь часть работы дата-сайентиста. Для того чтобы их правильно проинтерпретировать, нужно их визуализировать. На этом курсе студенты освоят использование функций визуализации при помощи python-библиотек, например, pandas для создания простейших графиков, познакомятся с более продвинутыми инструментами: matplotlib, seaborn и plotly, научатся использовать различные типы графиков: гистограммы, линейные графики, древовидные диаграммы, визуализации сезонности, pie-chart, heatmap, а также создавать интерактивные графики и отчеты (дэшборды).

1.2. Структура и объем модуля

Таблица 1

№	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах
1	Аналитика и визуализация данных	3
ИТОГО по модулю:		3

1.3. Последовательность освоения модуля в образовательной программе

Пререквизиты модуля	Не предусмотрено
Постреквизиты и кореквизиты модуля	Не предусмотрено

1.4. Распределение компетенций по дисциплинам модуля, планируемые результаты обучения (индикаторы) по модулю

Таблица 2

Перечень дисциплин модуля	Код и наименование компетенции	Индикаторы достижения компетенции	Планируемые результаты обучения
1	2	3	4
Аналитика и визуализация данных	ПК-4. Способен использовать инструментальные средства для решения задач машинного обучения	ПК-4.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи	ПК-4.1. 3-1. Знает возможности современных инструментальных средств и систем программирования для решения задач анализа данных и машинного обучения ПК-4.1. У-1. Умеет проводить сравнительный анализ и осуществлять выбор

			инструментальных средств для решения задач машинного обучения
--	--	--	---

1.5. Форма обучения

Обучение по дисциплинам модуля может осуществляться в очной форме.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Аналитика и визуализация данных

Рабочая программа дисциплины составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Медведева Марина Александровна	кандидат физико- математических наук, доцент	Доцент	Базовая кафедра «Аналитика больших данных и методы видеоанализа»
2	Балунгу Даниель		ассистент	Базовая кафедра «Аналитика больших данных и методы видеоанализа»

1. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ

Авторы:

- Медведева Марина Александровна, доцент, Базовая кафедра «Аналитика больших данных и методы видеоанализа»
- Балунгу Даниель, ассистент, Базовая кафедра «Аналитика больших данных и методы видеоанализа»

1.1. Технологии реализации, используемые при изучении дисциплины модуля

- Традиционная (репродуктивная) технология
- Разноуровневое (дифференцированное) обучение
 - Базовый уровень

**Базовый I уровень – сохраняет логику самой науки и позволяет получить упрощенное, но верное и полное представление о предмете дисциплины, требует знание системы понятий, умение решать проблемные ситуации. Освоение данного уровня результатов обучения должно обеспечить формирование запланированных компетенций и позволит обучающемуся на минимальном уровне самостоятельности и ответственности выполнять задания; Продвинутый II уровень – углубляет и обогащает базовый уровень как по содержанию, так и по глубине проработки материала дисциплины. Это происходит за счет включения дополнительной информации. Данный уровень требует умения решать проблемы в рамках курса и смежных курсов посредством самостоятельной постановки цели и выбора программы действий. Освоение данного уровня результатов обучения позволит обучающемуся повысить уровень самостоятельности и ответственности до творческого применения знаний и умений.*

1.2. Содержание дисциплины

Таблица 1.1

Код раздела, темы	Раздел, тема дисциплины*	Содержание
1	Введение в анализ данных с помощью Python	Обзор процесса анализа данных. Введение в концепции программирования на Python. Установка и настройка Python и необходимых библиотек
2	Импорт и очистка данных	Импорт и экспорт данных на Python. Обработка недостающих данных. Методы очистки и предварительной обработки данных
3	Предварительный анализ данных (EDA)	Распределение данных. Одномерный и многомерный анализ. Обобщение данных и описательная статистика
4	Библиотека Pandas	Введение в библиотеку Pandas. Сортировка, фильтрация и агрегирование данных. Объединение наборов данных
5	Статистический анализ	Основы проверки гипотез. Параметрические и непараметрические тесты. Корреляционный и регрессионный анализ
6	Визуализация данных с помощью Matplotlib	Введение в визуализацию данных. Линейные графики, точечные графики и столбчатые диаграммы. Настройка визуализаций с помощью Matplotlib
7	Интерактивная визуализация с помощью Plotly	Введение в библиотеку Plotly. Создание интерактивных графиков и информационных панелей. Встраивание визуальных элементов в записные книжки Jupyter

8	Библиотека Seaborn	Библиотека Seaborn для статистической визуализации. Построение графиков категориальных данных. Тепловые карты и визуализация географических данных
9	Анализ временных рядов	Введение в данные временных рядов. Декомпозиция временных рядов и прогнозирование. Визуализация данных временных рядов
10	Очистка веб-страниц и API-интерфейсы	Извлечение данных с веб-сайтов с использованием Python. Работа с API-интерфейсами для извлечения данных
11	Инструменты для бизнес-аналитики	Power BI, Google Data Studio

1.3. Направление, виды воспитательной деятельности и используемые технологии

Таблица 1.2

Направление воспитательной деятельности	Вид воспитательной деятельности	Технология воспитательной деятельности	Компетенция	Результаты обучения
Профессиональное воспитание	профориентационная деятельность	Технология формирования уверенности и готовности к самостоятельной успешной профессиональной деятельности Технология самостоятельной работы	ПК-4.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи	ПК-4.1. У-1. Умеет проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения

1.4. Программа дисциплины реализуется на государственном языке Российской Федерации.

2. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Аналитика и визуализация данных

Электронные ресурсы (издания)

1. Формирование и объединение данных в Power BI Desktop. <https://docs.microsoft.com/ru-ru/power-bi/desktop-shape-and-combine-data>
2. Что такое Power BI Desktop? Структура и схема работы с программой <https://biprosto.ru/power-bi/desktop.html>
3. Как работать с Microsoft Power BI — подробное руководство. <https://netpeak.net/ru/blog/kak-rabotat-s-microsoft-power-bi-podrobnoe-rukovodstvo/>
4. [Представление отчетов в Power BI Desktop](https://docs.microsoft.com/ru-ru/power-bi/desktop-report-view) <https://docs.microsoft.com/ru-ru/power-bi/desktop-report-view>

5. Основные сведения о DAX в Power BI Desktop <https://docs.microsoft.com/ru-ru/power-bi/desktop-quickstart-learn-dax-basics>
6. Колоколов А.С. Power BI для бизнес-пользователей. Основные советы по разработке аналитической панели мониторинга в Power BI. Открытый онлайн-курс Udemу
7. Massive. URL-адрес: <https://www.udemy.com/course/power-bi-for-бизнес-пользователи/?instructorPreviewMode=гость>

Печатные издания

1. Лимановская, О. В., Обабков, И. Н.; Основы машинного обучения : учебное пособие для студентов вуза, обучающихся по направлениям подготовки: 09.03.04 - Программная инженерия; 09.03.03 - Прикладная информатика; 02.04.02 - Фундаментальная информатика и информационные технологии; 09.04.02 - Информационные системы и технологии.; Издательство Уральского университета, Екатеринбург; 2020
2. Жуков, Р.А. Язык программирования Python. Практикум: учебное пособие/ Р.А. Жуков. – Москва: ИНФРА-М, 2022-216 с.
3. Уэс, Маккинли. Python и анализ данных. Электронный ресурс / Маккинли Уэс ; пер. А. А. Слинкин. - Python и анализ данных - Саратов : Профобразование, 2017. - 482 с.
4. Сузи, Р.А. Язык программирования Python Электронный ресурс : учебное пособие / Р.А. Сузи. - Язык программирования Python - Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), 2016. - 350 с.
5. Стенли, Липпман. Язык программирования С++ Электронный ресурс:
6. Полное руководство / Липпман Стенли, Лажойе Жози ; пер. А. Слинкин. - Язык программирования С++ - Саратов : Профобразование, 2017.- 1104 с.

Профессиональные базы данных, информационно-справочные системы

- 1) Единое окно доступа к образовательным ресурсам. Раздел Информатика и информационные технологии - http://window.edu.ru/catalog/p_rubr=2.2.75.6
- 2) Зональная научная библиотека УрФУ <http://lib.urfu.ru>
- 3) Научная электронная библиотека Elibrary.ru <https://www.elibrary.ru/>
- 4) Электронная библиотечная сеть "Лань" <http://e.lanbook.com/>
- 5) Портал информационно-образовательных ресурсов УрФУ <http://study.urfu.ru/>

Материалы для лиц с ОВЗ

Весь контент ЭБС представлен в виде файлов специального формата для воспроизведения синтезатором речи, а также в тестовом виде, пригодном для прочтения с использованием экранной лупы и настройкой контрастности.

Базы данных, информационно-справочные и поисковые системы

- 1) Государственная публичная научно-техническая библиотека <http://www.gpntb.ru>
- 2) Список библиотек, доступных в Интернет и входящих в проект «Либнет» <http://www.valley.ru/-nicr/listrum.htm>
- 3) Российская национальная библиотека <http://www.rsl.ru>
- 4) Свободная энциклопедия Википедия <https://ru.wikipedia.org/>

5) Портал национального общества имитационного моделирования <http://simulation.su/>

6) Портал российской ассоциации искусственного интеллекта <https://raai.org/>

3. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Аналитика и визуализация данных

Сведения об оснащённости дисциплины специализированным и лабораторным оборудованием и программным обеспечением

Таблица 3.1

№ п/п	Виды занятий	Оснащённость специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения
1	Лекции	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Доска аудиторная Подключение к сети Интернет	Office 365 EDUA3 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmc Браузер Chrome Google Colab
2	Практические занятия	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Доска аудиторная Персональные компьютеры по количеству обучающихся Подключение к сети Интернет	Office 365 EDUA3 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmc Браузер Chrome Google Colab Visual studio code Power BI (десктопное приложение)
3	Самостоятельная работа студентов	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Персональные компьютеры по количеству обучающихся Подключение к сети Интернет	Office 365 EDUA3 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmc Браузер Chrome

			<p>Google Colab</p> <p>Visual studio code</p> <p>Power BI (desktopное приложение)</p>
4	Текущий контроль и промежуточная аттестация	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Персональные компьютеры по количеству обучающихся</p> <p>Подключение к сети Интернет</p>	<p>Office 365 EDUA3 ShrdSvr ALNG SubsVL MVL PerUsr B Faculty EES</p> <p>Microsoft Windows 8.1 Pro 64-bit RUS OLP NL Acdmс</p> <p>Браузер Chrome</p> <p>Google Colab</p> <p>Visual studio code</p> <p>Power BI (desktopное приложение)</p>

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Аналитика и визуализация данных

Код модуля
1164099(1)

Модуль
Аналитика и визуализация данных

Екатеринбург

Оценочные материалы составлены автором(ами):

№ п/п	Фамилия, имя, отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Медведева Марина Александровна	кандидат физико-математических наук, доцент	Доцент	Базовая кафедра «Аналитика больших данных и методы видеоанализа»
2	Балунгу Даниель	-	ассистент	Базовая кафедра «Аналитика больших данных и методы видеоанализа»

Авторы:

- Медведева Марина Александровна, доцент, Базовая кафедра «Аналитика больших данных и методы видеоанализа»
- Балунгу Даниель, ассистент, Базовая кафедра «Аналитика больших данных и методы видеоанализа»

1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ Аналитика и визуализация данных

1.	Объем дисциплины в зачетных единицах	3	
2.	Виды аудиторных занятий	Лекции Практические занятия	
3.	Промежуточная аттестация	зачет	
4.	Текущая аттестация	Контрольная работа	1
		Домашняя работа	1

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ Аналитика и визуализация данных

Индикатор – это признак / сигнал/ маркер, который показывает, на каком уровне обучающийся должен освоить результаты обучения и их предъявление должно подтвердить факт освоения предметного содержания данной дисциплины, указанного в табл. 1.3 РПМ-РПД.

Таблица 1

Код и наименование компетенции	Индикаторы достижения компетенции	Планируемые результаты обучения	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3	4
ПК-4. Способен использовать инструментальные средства для решения задач машинного обучения	ПК-4.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи	ПК-4.1. З-1. Знает возможности современных инструментальных средств и систем программирования для решения задач анализа данных и машинного обучения	Лекции Практические занятия Домашняя работа Контрольная работа Зачет

		ПК-4.1. У-1. Умеет проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения	
--	--	--	--

3. ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)

3.1. Процедуры текущей и промежуточной аттестации по дисциплине

1) Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0,4		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>домашняя работа</i>	6,14	50
<i>контрольная работа</i>	6,7	50
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0,6		
Промежуточная аттестация по лекциям – зачет		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0,4		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – 0,6		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>выполнение и защита практических работ</i>	6,16	100
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – 1		
Промежуточная аттестация по практическим/семинарским занятиям – нет		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – не предусмотрено		
3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий – не предусмотрено		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям – не предусмотрено		
Промежуточная аттестация по лабораторным занятиям – нет		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – не предусмотрено		

4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий –не предусмотрено		
Текущая аттестация на онлайн-занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям -не предусмотрено		
Промежуточная аттестация по онлайн-занятиям –нет		
Весовой коэффициент значимости результатов промежуточной аттестации по онлайн-занятиям – не предусмотрено		

3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено		
Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено		

4. КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ

4.1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

Таблица 4

Критерии оценивания учебных достижений обучающихся

Результаты обучения	Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам
Знания	Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Умения	Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Опыт /владение	Студент демонстрирует опыт в области изучения на уровне указанных индикаторов.
Другие результаты	Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов.

	<p>Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения.</p> <p>Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.</p>
--	---

4.2 Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

Таблица 5

Шкала оценивания достижения результатов обучения (индикаторов) по уровням

Характеристика уровней достижения результатов обучения (индикаторов)				
№ п/п	Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание)	Шкала оценивания		
		Традиционная характеристика уровня		Качественная характеристика уровня
1.	Результаты обучения (индикаторы) достигнуты в полном объеме, замечаний нет	Отлично (80-100 баллов)	Зачтено	Высокий (В)
2.	Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения	Хорошо (60-79 баллов)		Средний (С)
3.	Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания	Удовлетворительно (40-59 баллов)		Пороговый (П)
4.	Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка	Неудовлетворительно но (менее 40 баллов)	Не зачтено	Недостаточный (Н)
5.	Результат обучения не достигнут, задание не выполнено	Недостаточно свидетельств для оценивания		Нет результата

5. СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ

5.1. Описание аудиторных контрольно-оценочных мероприятий по дисциплине модуля

5.1.1. Лекции

Изучение теоретического материала по темам/разделам лекций в соответствии с содержанием дисциплины (п. 1.2. РПД).

5.1.2. Практические занятия

Примерный перечень тем

1. Установка программной среды Python, пакетов NUMPY, PANDAS, IPYTHON
2. Методы очистки и обработки данных
3. Предварительный анализ данных
4. Визуализация данных с Pandas
5. Статистический анализ данных
6. Настройка визуализаций с помощью Matplotlib
7. Seaborn для статистической визуализации
8. Парсинг и анализ данных из интернета
9. Визуализация данных и отчеты Power BI
11. Реализация проекта анализа данных

LMS-платформа – не предусмотрена

5.2. Описание внеаудиторных контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля

Разноуровневое (дифференцированное) обучение.

Базовый

5.2.1. Контрольная работа

Примерный перечень тем

1. Методы обработки и анализа данных
2. Методы и инструменты визуализации данных

Примерные задания

1. Для чего используется визуализация данных?
 1. Изучение заданного набора данных.
 2. Обмен объективным представлением данных.
 3. Поддержка рекомендаций для различных заинтересованных сторон.
 4. Все вышеперечисленное.
2. Что из перечисленного ниже является слоем в архитектуре Matplotlib?
 1. Рисунок Слой
 2. FigureCanvas Слой
 3. Backend_Bases Слой
 4. Серверный уровень
3. Какой слой позволяет полностью контролировать и тонко настраивать Matplotlib — контейнер верхнего уровня для всех элементов графика?
 1. Сценарий уровня
 2. Серверный уровень
 3. Уровень событий
 4. Слой художника
4. Какой из приведенных ниже кодов создаст график с накопленной областью данных в pandas датафрейм, area_df, со значением прозрачности 0,75?
 - 1).

```
import matplotlib.pyplot as plt
transparency = 1 - 0.75
area_df.plot(kind='area', alpha=transparency, stacked=False, figsize=(20, 10))
plt.title('Plot Title')
plt.ylabel('Метка вертикальной оси')
plt.xlabel('Метка горизонтальной оси')
```

```
plt.show()
```

```
2) прозрачность = 0,75  
ax = area_df.plot(kind='площадь', alpha=прозрачность, stacked=False, figsize=(20, 10))  
ax.set_title('Название графика')  
ax.set_ylabel('Метка вертикальной оси')  
ax.set_xlabel('Метка горизонтальной оси')
```

```
3) import matplotlib.pyplot as plt  
прозрачность = 0.35  
area_df.plot(kind='область', alpha=прозрачность, figsize=(20, 10))  
plt.title('Заголовок графика')  
plt.ylabel('Метка вертикальной оси')  
plt.xlabel('Метка горизонтальной оси')  
plt.show()
```

```
4) import matplotlib.pyplot as plt  
transparency = 0.75  
area_df.plot(kind='area', alpha=transparency, figsize=(20, 10))  
plt.title('Plot Title')  
plt.ylabel('Vertical Axis Label')  
plt.xlabel('Horizontal Axis Label')  
plt.show()
```

5. Каков способ статистического представления распределения данных по пяти основным измерениям?

1. Гистограмма
2. Коробка
3. Точечная диаграмма
4. Линейный график

6. Что является разновидностью точечной диаграммы, отображающей три измерения данных?

1. Ничего из вышеперечисленного
2. Пузырьковый сюжет
3. Точечная карта
4. Тепловая карта

7. Что из перечисленного НЕ ВЕРНО в отношении облака слов?

1. Облако слов может быть сгенерировано на Python с помощью пакета `word_cloud`, разработанного Андреасом Мюллером.
2. Облако слов — это изображение частоты различных слов в некоторых текстовых данных.
3. Ничего из вышеперечисленного.
4. Облако слов — это изображение частоты стоп-слов.

9. Что из перечисленного НЕ является стилем плиток карт Folium?

1. Тычиночная местность
2. Тычиночный тонер
3. Река Прибрежная
4. OpenStreetMap

10. Что из перечисленного НЕ соответствует стилю плитки Stamen Terrain для карт Folium?

1. Идеально подходит для гибридных приложений данных и исследования речных меандров и прибрежных зон
2. Демонстрирует расширенную маркировку и обобщение линий дорог с двусторонним движением

3. Особенности естественных цветов растительности
4. Особенности затенения холмов
11. Визуализации в виде графика не могут быть отображены каким из следующих способов:
 1. Отображается в записной книжке Jupyter
 2. Сохранено в HTML-файлы
 3. Служит в качестве чистого python-приложения для сборки с использованием Dash
 4. Ничего из вышеперечисленного
12. Площадные графики по умолчанию не складываются
 1. Правда
 2. Неверно
13. Следующий код создаст гистограмму ряда панд, `series_data`, и выровняет края ячейки с горизонтальными галочками.


```
count, bin_edges = np.histogram(series_data)
series_data.plot(kind='hist', xticks = count, bin_edges)
```

 - 1). Правда.
 - 2). Неверно.
14. Что создаст горизонтальную гистограмму рассматриваемых данных?
 - 1). `question.plot(type='bar', rot=90)`
 - 2). `question.plot(kind='bar', orientation='horizontal')`
 - 3). `question.plot(kind='barh')`
 - 4). `question.plot(kind='bar')`
 - 5). `question.plot(kind='bar', type='horizontal')`

Визуализация данных с помощью Python

Задание 1. Был проведен опрос для оценки интереса аудитории к различным темам науки о данных, а именно: Большие данные (Spark/Hadoop) Анализ данных / Статистика / Визуализация данных / Глубокое обучение / Машинное обучение/ У участников было три варианта оценки по каждой теме: «Очень интересно», «В некоторой степени заинтересовано» и «Не заинтересовано». В опросе приняли участие 2 233 респондента. Результаты опроса были сохранены в CSV-файле, и к ним можно получить доступ по этой ссылке: https://cocl.us/datascience_survey_data. Если вы изучите CSV-файл, вы обнаружите, что в первом столбце представлены разделы для обработки и анализа данных, а в первой строке представлены варианты для каждого раздела. Используйте метод `pandas read_csv` для чтения CSV-файла в кадр данных `pandas`. Чтобы считывать данные во фрейм данных, одним из способов сделать это является использование параметра `index_col` для загрузки первого столбца в качестве индекса кадра данных. Вот документация по методу `read_csv` панд: https://pandas.pydata.org/pandas-docs/stable/generated/pandas.read_csv.html После того, как вам удалось создать вышеуказанный кадр данных, загрузите снимок экрана вашего кадра данных с фактическими числами.

Задание 2. Используйте слой художника `Matplotlib`, чтобы воспроизвести приведенную ниже гистограмму, чтобы визуализировать процент интереса респондентов к различным опрошенным темам науки о данных. Чтобы создать эту гистограмму, вы можете выполнить следующие шаги: Отсортируйте кадр данных в порядке убывания - Очень интересно. Переведите цифры в проценты от общего числа респондентов. Напомним, что в опросе приняли участие 2 233 респондента. Округлите проценты до 2 знаков после запятой. Что касается диаграммы: используйте размер рисунка (20, 8), ширину полосы 0,8, используйте цветной `#5cb85c` для очень интересующих столбцов, цветной `#5bc0de` для несколько интересующих столбцов и цветной `#d9534f` для неинтересующих столбцов, используйте размер шрифта 14 для подписей столбцов, процентов и легенды, используйте размер шрифта 16 для заголовка, и отобразите проценты над полосами, как показано выше, и удалите левую, верхнюю и правую границы. Как только вы будете удовлетворены своей диаграммой, загрузите скриншот вашего графика.

Задание 3. В этом вопросе вам необходимо создать карту Choropleth для визуализации преступности в Сан-Франциско. Прежде чем вы будете готовы приступить к построению карты, давайте реструктурируем данные так, чтобы они были в правильном формате для карты Choropleth. По сути, вам нужно будет создать кадр данных, в котором будет указан каждый район Сан-Франциско вместе с соответствующим общим количеством преступлений. Основываясь на наборе данных о преступности в Сан-Франциско, вы обнаружите, что Сан-Франциско состоит из 10 основных районов, а именно: Центральный, Южный, Бэйвью, Миссия, Парк, Ричмонд, Инглсайд, Таравал, Северный и Тендерлойн. Преобразуйте набор данных Сан-Франциско, который вы также можете найти здесь, https://cocl.us/sanfran_crime_dataset, в кадр данных pandas, который представляет общее количество преступлений в каждом районе. Как только вы будете довольны своим фреймом данных, загрузите скриншот фрейма данных pandas.

Задание 4. Теперь вы должны быть готовы приступить к созданию карты Choropleth. Как вы узнали в лаборатории карт Choropleth, вам понадобится файл GeoJSON, который отмечает границы различных районов Сан-Франциско. Нужный файл доступен по этой ссылке: https://cocl.us/sanfran_geojson. Для карты убедитесь, что: она сосредоточена вокруг Сан-Франциско, вы используете уровень масштабирования 12, вы используете `fill_color = 'YlOrRd'`, вы определяете `fill_opacity = 0.7`, вы определяете `line_opacity = 0.2`, а также вы определяете легенду и используете пороговую шкалу по умолчанию. Когда вы будете готовы отправить свою карту, загрузите скриншот вашей карты Choropleth.

LMS-платформа – не предусмотрена

5.2.2. Домашняя работа

Примерный перечень тем

1. Первичный анализ данных
2. Исследование данных сервиса «Яндекс.Музыка» — сравнение пользователей двух городов
3. Исследование надёжности заёмщиков — анализ банковских данных
4. Продажа квартир в Санкт-Петербурге — анализ рынка недвижимости
5. Определение выгодного тарифа для телека компании
6. Изучение закономерностей, определяющих успешность игр
7. Изучение базы данных интернет-сервиса для чтения книг
8. Анализ убытков приложения ProcrastinatePRO+
9. Проверка гипотез по увеличению выручки в интернет-магазине
10. Рынок заведений общественного питания Москвы
11. Анализ пользовательского поведения в мобильном приложении
12. Создание дашборда по пользовательским событиям для агрегатора новостей
13. Прогнозирование вероятности оттока пользователей для фитнес-центров

Примерные задания

1. На реальных данных Яндекс.Музыки с помощью библиотеки Pandas и её возможностей проверить данные и сравнить поведение и предпочтения пользователей двух столиц — Москвы и Санкт-Петербурга.

2. На основе статистики о платёжеспособности клиентов исследовать влияет ли семейное положение и количество детей клиента на факт возврата кредита в срок.

3. Используя данные сервиса Яндекс.Недвижимость, определить рыночную стоимость объектов недвижимости и типичные параметры квартир.

4. На основе данных клиентов оператора сотовой связи проанализировать поведение клиентов и поиск оптимального тарифа.

5. Используя исторические данные о продажах компьютерных игр, оценки пользователей и экспертов, жанры и платформы, выявить закономерности, определяющие успешность игры.

6. Проект по анализу базы данных сервиса для чтения книг по подписке.

7. Задача для маркетингового аналитика развлекательного приложения Procrastinate Pro+. Несмотря на огромные вложения в рекламу, последние несколько месяцев компания терпит убытки. Задача — разобраться в причинах и помочь компании выйти в плюс.

8. Используя данные интернет-магазина приоритезировать гипотезы, произвести оценку результатов A/B-тестирования различными методами.

9. Исследование рынка общественного питания на основе открытых данных, подготовка презентации для инвесторов.

10. На основе данных использования мобильного приложения для продажи продуктов питания проанализировать воронку продаж, а также оценить результаты A/A/B-тестирования.

11. Используя данные Яндекс.Дзена построить дашборд с метриками взаимодействия пользователей с карточками статей.

12. На основе данных о посетителях сети фитнес-центров спрогнозировать вероятность оттока для каждого клиента в следующем месяце, сформировать с помощью кластеризации портреты пользователей.

LMS-платформа – не предусмотрена

5.3. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля

5.3.1. Зачет

1. Какие инструментальные средства используются для организации рабочего места специалиста Data Science?
2. Какие библиотеки Python используются для работы в области машинного обучения? Дайте краткую характеристику каждой библиотеке.
3. Почему при реализации систем машинного обучения широкое распространение получили библиотеки Python?
4. Перечислите функции Python для визуализации данных.
5. Какая библиотека python предназначена для управления наборами данных: numpy, pandas, sklearn, opencv, matplotlib?
6. Какая стратегия является нежелательной при обработке пропусков в данных?
 - а) замена пропущенных значений в столбце медианным значением по данному столбцу;
 - б) удаление строк, содержащих пропуски в данных;
 - в) замена пропущенных значений в столбце средним арифметическим значением по данному столбцу;
 - г) замена пропущенных значений в столбце наиболее часто встречающимся значением по данному столбцу;
7. Обоснуйте ответ на следующую проблему предварительной обработки данных: имеется независимая категориальная переменная y , которая представляет собой категориальный признак, определенный на домене {C#, Java, Python, R}. Нужно ли применять к данному целевому признаку OneHotEncoder?
8. Поясните принцип разбиения набора данных на обучающую и тестовую выборку. Какое соотношение «тестовая:обучающая» наиболее оптимально: 20:80, 50:50, 25:75, 5:95, 40:30?
9. Какой код лучше использовать при загрузке данных из csv-файла?

- а) dataset = read_csv("data.csv")
- б) dataset = import("data.csv")
- в) dataset = read.csv("data.csv")
- г) dataset = import.csv("data.csv")
- д) dataset = read_xls("data.csv")

5.4 Содержание контрольно-оценочных мероприятий по направлениям воспитательной деятельности

Направление воспитательной деятельности	Вид воспитательной деятельности	Технология воспитательной деятельности	Компетенция	Результаты обучения	Контрольно-оценочные мероприятия
Профессиональное воспитание	Профориентационная деятельность	Технология формирования уверенности и готовности к самостоятельной успешной профессиональной деятельности Технология самостоятельной работы	ПК-4.1	ПК-4.1. У-1	Лекции Зачет Домашняя работа