

Перечень сведений о рабочей программе модуля	Учетные данные
Образовательная программа Инженерия искусственного интеллекта	Код ОП 09.04.01
Направление подготовки Информатика и вычислительная техника	Код направления и уровня подготовки 09.04.01

Области образования, в рамках которых реализуется модуль образовательной программы по СУОС УрФУ:

№ п/п	Перечень областей образования, для которых разработан СУОС УрФУ	Уровень подготовки
1.	Инженерное дело, технологии и технические науки	магистратура

Программа модуля составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

Рекомендовано учебно-методическим советом института радиоэлектроники и информационных технологий - РТФ

Протокол № 7 от 11 октября 2021 г.

1. ОБЩАЯ ХАРАКТЕРИСТИКА МОДУЛЯ Инжиниринг данных

1.1. Аннотация содержания модуля

Модуль «Инжиниринг данных» состоит из одноименной дисциплины, посвященной подготовке данных для моделей машинного обучения. Рассматриваются особенности работы с данными в различных форматах на языке Python. Уделяется внимание инструментам и технологиям загрузки данных из интернета и социальных сетей. Подробно изучаются методы очистки данных и соответствующие библиотеки на Python.

1.2. Структура и объем модуля

Таблица 1

№ п/п	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах и часах
1.	Инжиниринг данных	3 з.е. / 108 ч.
ИТОГО по модулю:		3 з.е. / 108 ч.

1.3. Последовательность освоения модуля в образовательной программе

Пререквизиты модуля	<i>отсутствуют</i>
Постреквизиты и корреквизиты модуля	<i>отсутствуют</i>

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

Таблица 2

Перечень дисциплин модуля	Код и наименование компетенции	Планируемые результаты обучения (индикаторы)
1	2	3
Инжиниринг данных	ОПК-3. Способен планировать и проводить комплексные исследования и изыскания для решения инженерных задач, относящихся к профессиональной деятельности, включая проведение измерений, планирование и постановку экспериментов, интерпретацию полученных результатов.	ОПК-3. З-3. Сделать обзор основных методов статистической обработки и анализа результатов измерений. ОПК-3. У-2. Обоснованно выбрать необходимую аппаратуру и метод исследования для решения инженерных задач, относящихся к профессиональной деятельности ОПК-3. П-1. Выполнять в рамках поставленного задания экспериментальные комплексные научно-технические исследования и изыскания для решения инженерных задач в области профессиональной деятельности, включая обработку, интерпретацию и оформление

		результатов ОПК-3. Д-1. Проявлять умение видеть детали, упорство, аналитические умения.
--	--	--

1.5. Форма обучения

Обучение по дисциплине модуля может осуществляться в очной форме.

2. СОДЕРЖАНИЕ И ОБЕСПЕЧЕНИЕ РЕАЛИЗАЦИИ ДИСЦИПЛИН МОДУЛЯ

**ПРОГРАММА МОДУЛЯ
ИНЖИНИРИНГ ДАННЫХ**

**РАЗДЕЛ 2. СОДЕРЖАНИЕ И ОБЕСПЕЧЕНИЕ РЕАЛИЗАЦИИ ДИСЦИПЛИН
МОДУЛЯ**

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
ИНЖИНИРИНГ ДАННЫХ**

Рабочая программа дисциплины составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ- РТФ, УрФУ

Рекомендовано учебно-методическим советом института радиоэлектроники и информационных технологий - РТФ

Протокол № 7 от 11 октября 2021 г.

2. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ

Инжиниринг данных

2.1. Технологии обучения, используемые при изучении дисциплины модуля

- Традиционная (репродуктивная) технология;
- Смешанная модель обучения с использованием онлайн-курса;
- Исключительно электронного обучения с использованием онлайн-курса;

2.2. Содержание дисциплины

Таблица 1.3

Код раздела, темы	Раздел, тема дисциплины	Содержание
1	Работа с данными в Python.	Библиотеки для работы с данными в различных форматах в Python: файлы CSV, JSON, HTML. Работа с базами данных в Python. Работа с изображениями, видео и звуковыми файлами. Форматы хранения больших данных и работа с ними: Parquet, Avro. Графы знаний.
2	Подготовка данных для систем машинного обучения.	Сбор данных и формирование набора данных для систем машинного обучения. Загрузка данных из интернет и социальных сетей. Методы очистки и подготовки данных. Очистка и подготовка данных на Python. Разметка данных. Общедоступные платформы для хранения данных. Подход Data-Centric AI.
3	Параллельная и распределенная обработка данных.	Архитектура центров обработки данных, кластеры для параллельных и распределенных вычислений. Экосистема для распределенного хранения и обработки больших объемов данных: Apache Hadoop, HDFS. Распределенная обработка данных в Apache Spark. Архитектура Apache Spark: Resilient Distributed Dataset (RDD), действия трансформации. Работа с данными с использованием Spark DataFrame. Источники данных для Spark DataFrame. Обработка данных в Spark DataFrame. Использование SQL в Spark DataFrame.

2.3. Программа дисциплины реализуется на государственном языке Российской Федерации.

2.4. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ Инжиниринг данных

Электронные ресурсы (издания)

1. Онлайн-курс “Прикладное программирование на языке Python”. URL: <https://openedu.ru/course/urfu/PYAP/> (дата обращения: 05.10.2021).
2. Real Python Tutorials. URL: <https://realpython.com/> (дата обращения: 05.10.2021).
3. Jake VanderPlas. Python Data Science Handbook. URL: <https://jakevdp.github.io/PythonDataScienceHandbook/> (дата обращения: 05.10.2021).
4. Jeroen Janssens. Data Science at the Command Line. URL: <https://www.datascienceatthecommandline.com/> (дата обращения: 05.10.2021).

5. Andrew Ng. MLOps: From Model-centric to Data-centric AI. URL: <https://www.deeplearning.ai/wp-content/uploads/2021/06/MLOps-From-Model-centric-to-Data-centric-AI.pdf> (дата обращения: 05.10.2021).
6. Spark SQL, DataFrames and Datasets Guide. URL: <https://spark.apache.org/docs/latest/sql-programming-guide.html> (дата обращения: 05.10.2021).

Профессиональные базы данных, информационно-справочные системы

1. Applied Science & Technology Source. EBSCO publishing <http://search.ebscohost.com>
2. Wiley Online Library <http://onlinelibrary.wiley.com/>
3. Гугл Академия <https://scholar.google.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. Academic Search Ultimate EBSCO publishing – <http://search.ebscohost.com>
2. eBook Collections Springer Nature – <https://link.springer.com/>
3. Гугл Академия – <https://scholar.google.ru/>
4. Электронный научный архив УрФУ <https://elar.urfu.ru/>
5. Зональная научная библиотека (УрФУ) - <http://lib.urfu.ru/>
6. Портал информационно-образовательных ресурсов УрФУ <https://study.urfu.ru/>
7. Электронно-библиотечная система «Лань» – <https://e.lanbook.com/>
8. Университетская библиотека ONLINE – <https://biblioclub.ru/>
9. Электронно-библиотечная система "Библиокомплектатор" (IPRbooks) <http://www.bibliocomplectator.ru/available>
10. Электронные информационные ресурсы Российской государственной библиотеки <https://www.rsl.ru/>
11. Научная электронная библиотека «КиберЛенинка» <https://cyberleninka.ru/>

2.5. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Инжиниринг 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. Реквизиты подтверждающего документа
1	Лекции; Практические занятия.	Компьютерный класс. Мультимедийный проектор с экраном.; Сетевое оборудование.; Локальная сеть с выходом в глобальную сеть Internet.	В курсе используется бесплатно распространяемое программное обеспечение: 1. Python – https://www.python.org/ 2. Веб - среда разработки

			для языка программирования Python: google colab - https://colab.research.google.com/ 3. Anaconda solution - https://www.anaconda.com/ 4. Система распределенной обработки данных Apache Spark – https://spark.apache.org/
--	--	--	--

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

Код модуля
М.1.7

Модуль
Инжиниринг данных

Екатеринбург, 2021

Оценочные материалы по модулю составлены авторами:

№ п/п	Фамилия, имя, отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

1. СТРУКТУРА И ОБЪЕМ МОДУЛЯ ИНЖИНИРИНГ ДАННЫХ

№ п/п	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах и часах	Форма итоговой промежуточной аттестации по дисциплинам модуля и в целом по модулю
1.	Инжиниринг данных	3 з.е. /108 час.	Зачет
ИТОГО по модулю:		3 з.е. /108 час.	

2. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО МОДУЛЮ

Не предусмотрено

**Раздел 3. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ
ИНЖИНИРИНГ ДАННЫХ**

Модуль М 1.7 Инжиниринг данных

Оценочные материалы составлены автором(ами):

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Созыкин Андрей Владимирович	кандидат технических наук, нет	доцент	Кафедра информационных технологий и систем управления, ИРИТ-РТФ, УрФУ

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ ИНЖИНИРИНГ ДАННЫХ

Таблица 1.1

Код и наименование компетенций, формируемые с участием дисциплины	Планируемые результаты обучения	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3
<p>ОПК-3. Способен планировать и проводить комплексные исследования и изыскания для решения инженерных задач, относящихся к профессиональной деятельности, включая проведение измерений, планирование и постановку экспериментов, интерпретацию полученных результатов.</p>	<p>ОПК-3. З-3. Сделать обзор основных методов статистической обработки и анализа результатов измерений.</p> <p>ОПК-3. У-2. Обоснованно выбрать необходимую аппаратуру и метод исследования для решения инженерных задач, относящихся к профессиональной деятельности</p> <p>ОПК-3. П-1. Выполнять в рамках поставленного задания экспериментальные комплексные научно-технические исследования и изыскания для решения инженерных задач в области профессиональной деятельности, включая обработку, интерпретацию и оформление результатов</p> <p>ОПК-3. Д-1. Проявлять умение видеть детали, упорство, аналитические умения.</p>	<p>1) Контрольная работа 2) Домашние работы 3) Выполнение практических работ 4) Зачет</p>

2. ВИДЫ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ, ВКЛЮЧАЯ МЕРОПРИЯТИЯ ТЕКУЩЕЙ АТТЕСТАЦИИ

2.1. Распределение объема времени по видам учебной работы

Таблица 2

№ п/ п	Наименование дисциплины модуля Инжиниринг данных	Объем времени, отведенный на освоение дисциплины модуля								
		Аудиторные занятия, час.				Промежуточная аттестация (форма итогового контроля)	Контактная работа (час.)	Самостоятельная работа студента, включая текущую аттестацию (час.)	Всего по дисциплине	
		Занятия лекцион ного типа	Практиче ские работы	Лаборатор ные работы	Всего				Час.	Зач. ед.
1	2	3	4	5	6	7	8	9	10	11
1.	Инжиниринг данных	18	18	0	36	Зачет	41.65	66.35	108	3

2.2. Виды СРС, количество и объем времени на контрольно-оценочные мероприятия СРС по дисциплине

Контрольно-оценочные мероприятия СРС включают самостоятельное изучение материала, подготовку к аудиторным занятиям и мероприятиям текущего контроля, выполнение и оформление внеаудиторных мероприятий текущего контроля и подготовку к мероприятиям промежуточного контроля.

Таблица 3

№ п/п	Вид самостоятельной работы студента по дисциплине модуля	Количество контрольно-оценочных мероприятий СРС	Объем контрольно-оценочных мероприятий СРС (час.)
1.	Подготовка к аудиторным занятиям и мероприятиям текущего контроля: лекционным, практическим занятиям.		13,5 час.
2	Выполнение и оформление мероприятий текущего контроля:		
2.1	Контрольная работа	1	5 час.
2.2	Домашняя работа	2	10 час.
3.	Подготовка к зачету	зачет	12 час.
4.	Самостоятельное изучение материала		25,85 час.
Итого на СРС по дисциплине:			66,35 час.

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

Процедуры текущей и промежуточной аттестации по дисциплине

1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Контрольная работа</i>	1 сем.	70
<i>Самостоятельное изучение материала</i>	1 сем.	30
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5		
Промежуточная аттестация по лекциям – Зачет		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5		
2. Практические/семинарские занятия: коэффициент значимости совокупных		

результатов практических/семинарских занятий – 0.5		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Выполнение и оформление практических работ</i>	1 сем.	50
<i>Домашняя работа №1</i>	1 сем.	25
<i>Домашняя работа №2</i>	1 сем.	25
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – 1		
Промежуточная аттестация по практическим/семинарским занятиям – не предусмотрена		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – 0		
3. Лабораторные занятия: Не предусмотрены		
коэффициент значимости совокупных результатов лабораторных занятий – 0		

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.
Личностные качества	Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов. Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения. Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.

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	Удовлетворительн о (40-59 баллов)		Пороговый (П)
4.	Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка	Неудовлетворител ьно (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

Задания по контрольно-оценочным мероприятиям в рамках текущей и промежуточной аттестации должны обеспечивать освоение и достижение результатов обучения (индикаторов) и предметного содержания дисциплины на соответствующем уровне.

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

5.1.1. Практические занятия

Номер занятия	Примерный перечень тем практических занятий
1	Библиотеки для работы с данными в Python: numpy, pandas.
2	Работа с текстовыми файлами разных форматов в Python: CSV, JSON, HTML.
3	Работа с базами данных в Python.
4	Работа с изображениями, видео и звуковыми файлами в Python.
5	Работа с файлами для хранения больших данных в Python.
6	Работа с графами знаний в Python.

7	Создание собственных наборов данных в Python. Очистка и подготовка данных.
8	Работа с данными в Apache Spark.
9	Использование SQL в Apache Spark.

5.1.2. Лабораторные занятия

Не предусмотрено

5.1.3. Курсовая работа / Курсовой проект

Не предусмотрено

5.1.4. Контрольная работа

Примерная тематика контрольных работ:

1. Работа с данными в различных форматах.

Примерные задания в составе контрольных работ:

Задание 1. Формат данных CSV может быть использован как замена:

- **реляционным СУБД**
- **нереляционным СУБД**
- **документно-ориентированным СУБД**

Задание 2. Что обычно выступает разделителем столбцов в формате CSV:

- **Запятая**
- **Точка с запятой**
- **Перенос строки**

Задание 3. Какие элементы не используются в формате JSON в качестве структурных элементов:

- **Теги**
- **Кавычки**
- **Двоеточие**

Задание 5. Для какого языка программирования впервые использовался формат JSON:

- **JavaScript**
- **Java**
- **Python**

Задание 6. В каком виде хранятся данные в MongoDB:

- **BSON**
- JSON
- XML

Задание 7. Выберите наиболее подходящие характеристики MongoDB:

- **Ключ-значение, неструктурированность данных, нереляционные свойства**
- Ключ-значение, неструктурированность данных, реляционные свойства
- SQL, неструктурированность данных, реляционные свойства

Задание 8. Преобразуйте следующий код в формат JSON:

```
<companies>
  <company>
    <company-id>7707040070</company-id>
    <name>Якорьбанк</name>
    <shortname>Якорьбанк</shortname>
    <name-other>Якорьбанк, платёжное устройство</name-other>
    <address>Россия, Республика Татарстан, Зеленодольский район, село
Нурлаты, улица Гагарина, 46</address>
    <phone>
      <type>phone</type>
      <number>+7 (800) 999-99-90</number>
    </phone>
    <url>http://www.yakorbank.ru/</url>
    <working-time>будни 8:30-18:00, сб 9:00-14:30</working-time>
    <rubric-id>184106974</rubric-id>
    <actualization-date>23.09.2019</actualization-date>
    <coordinates>
      <lat>55.616051</lat>
      <lon>48.295532</lon>
    </coordinates>
  </company>
</companies>
```

5.1.5. Домашняя работа

Примерная тематика домашних работ:

Домашняя работа №1:

Создание собственного набора данных.

Домашняя работа №2:

Обработка данных в Apache Spark.

Примерные задания в составе домашних работ:

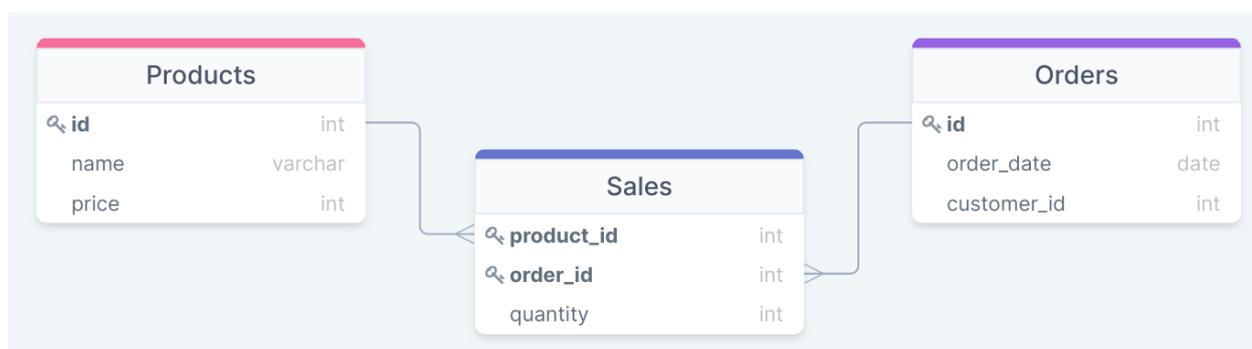
1. Подготовить собственный набор данных. Выберите задачу в одном из направлений создания системы искусственного интеллекту (компьютерное зрение, обработка естественного языка) и подготовьте для этой задачи набор данных для обучения с учителем. Соберите и очистите данные, проведите разметку. Готовый набор данных разместите на одной из общедоступных платформ для хранения данных по своему выбору. Подготовьте документацию к созданному набору данных.

Перечень задач, для которых рекомендуется подготовить набор данных:

- Классификация объектов на изображениях.
- Определение положения объектов на изображениях.
- Определение положения объектов в видео.
- Классификация текста на русском языке.
- Определение эмоциональной окраски текста на русском языке.

2. Создайте набор данных в Apache Spark и проведите его исследование с помощью Spark DataFrame API.

Схема данных выглядит следующим образом:



Онлайн-школа продает образовательные продукты: онлайн-курсы, книги, семинары и т.п. Описание и стоимость продуктов содержится в таблице **Products**. Когда клиент что-то покупает, создается заказ, который заносится в таблицу **Orders**. Заказ может содержать несколько продуктов, перечень продуктов в заказах содержится в таблице **Sales**.

Таблица **Products** - продукты онлайн-школы:

- **id** - идентификатор продукта
- **name** - название продукта
- **price** - стоимость продукта

Таблица **Orders** - заказы:

- **id** - идентификатор заказа
- **order_date** - дата заказа
- **customer_id** - идентификатор заказчика (таблица с заказчиками не создается для упрощения примера)

Таблица **Sales** - продажи:

- **product_id** - идентификатор продукта, ссылка на таблицу Products, поле id
- **order_id** - идентификатор заказа, ссылка на таблицу Orders, поле id
- **quantity** - количество продуктов в заказе

Ноутбук в облачной платформе Colab с заготовкой кода для домашней работы – https://colab.research.google.com/drive/1MLiHIZ2CcBbCp_U7wKs_CAPmnx2F7Q6?usp=sharing

Задания для анализа:

- Выведите список продуктов, которые не были проданы ни разу
- Определите, сколько продуктов любого типа было продано по дням.
- Определить, какая выручка от продуктов любого типа была получена по дням.

5.1.6. Расчетная работа / Расчетно-графическая работа

Не предусмотрено

5.1.7. Реферат / эссе / творческая работа

Не предусмотрено

5.1.8. Проектная работа

Не предусмотрено

5.1.9. Деловая (ролевая) игра / Дебаты / Дискуссия / Круглый стол

Не предусмотрено

5.1.10. Кейс-анализ

Не предусмотрено

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

5.2.1. Зачет в форме независимого тестового контроля (НТК).

НТК по дисциплине модуля не проводится.

5.2.2. Зачет в традиционной форме (устные / письменные ответы на вопросы)

Список примерных тем для зачёта:

1. Библиотека pandas в Python.
2. Работа с данными в формате CSV в Python.
3. Работа с данными в формате JSON в Python.
4. Работа с данными в формате HTML в Python.
5. Работа с изображениями в Python.
6. Работа с видео в Python.
7. Работа с аудио в Python.
8. Работа с Parquet в Python.
9. Работа с графами знаний в Python.
10. Этапы и инструменты создания наборов данных для машинного обучения.
11. Загрузка данных с Web-сайтов.
12. Загрузка данных из социальных сетей.
13. Методы и инструменты подготовки данных.
14. Методы и инструменты очистки данных.
15. Разметка данных.
16. Общедоступные платформы для хранения данных.
17. Архитектура центров обработки данных.
18. Кластеры для параллельных и распределенных вычислений.
19. Экосистема для распределенного хранения и обработки больших объемов данных: Apache Hadoop.
20. Распределенная файловая система HDFS.
21. Распределенная обработка данных в Apache Spark.

22. Работа с данными с использованием Apache Spark DataFrame.
23. Источники данных для Apache Spark DataFrame.
24. Обработка данных в Apache Spark DataFrame.
25. Использование SQL в Apache Spark DataFrame.