

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«24» июня 2025 г.
Протокол № 2

**Рабочая программа дисциплины (модуля)
«Python для анализа данных»**

Направление подготовки: 38.04.05 Бизнес-информатика

Направленность (профиль) подготовки: Продуктовый менеджмент

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Срок освоения программы: 2 года

Год набора: 2025

**Москва
2025**

Содержание

| | |
|---|----|
| 1. Краткая характеристика дисциплины (модуля) | 3 |
| 2. Перечень планируемых результатов обучения..... | 5 |
| 3. Тематический план..... | 7 |
| 4. Содержание дисциплины (модуля)..... | 7 |
| 5. Учебно-методическое обеспечение | 8 |
| 6. Материально-техническое обеспечение | 8 |
| 7. Методические и оценочные материалы | 10 |

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Python для анализа данных» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – магистратура по специальности 38.04.05 Бизнес-информатика, профиль Продуктовый менеджмент, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 990 от 12.08.2020 года.

Изучение дисциплины (модуля) «Python для анализа данных» имеет огромное значение в современном мире, где данные стали ключевым ресурсом для принятия решений в бизнесе, науке и технологиях, позволяя специалистам эффективно обрабатывать большие объемы информации и извлекать из них ценную информацию. Освоение Python и его библиотек, таких как pandas, numpy и matplotlib, открывает широкие возможности для карьерного роста в областях data science, машинного обучения и аналитики, делая выпускников востребованными на рынке труда.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки магистратуры по направлению 38.04.05 Бизнес-информатика, профиль Продуктовый менеджмент и входит в обязательную часть Блока 1.

Дисциплина (модуль) изучается на 1 курсе в 1 семестре, доступна для прохождения при условии успешного завершения дисциплины (модуля) «Основы Python».

Цель изучения дисциплины (модуля): формирование знаний и навыков в применении языка Python и его специализированных библиотек для эффективной обработки, анализа и визуализации данных в профессиональной деятельности.

Задачи изучения дисциплины (модуля):

- освоить основные объекты и методы библиотек pandas и numpy для выполнения операций с датафреймами и многомерными массивами;
- научиться применять инструменты matplotlib, seaborn и plotly для создания разнообразных типов графиков и диаграмм;
- изучить способы взаимодействия с внешними источниками данных через API, системы контроля версий Git и базы данных;
- практиковать базовые техники анализа данных, включая факторный анализ, когортный анализ и методы прогнозирования;
- развить навыки использования сред разработки, таких как Jupyter Notebook, JupyterHub и Google Colab, для автоматизации рассылки метрик и решения прикладных задач.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- объекты и методы pandas для работы с датафреймами;
- объекты и методы Numpy для работы с многомерными массивами;
- объекты и методы matplotlib, seaborn и plotly для визуализации;
- основные методы и библиотеки для работы с API, Git и базами данных;
- базовые методы анализа данных: факторный анализ, когортный анализ, методы прогнозирования.

уметь:

- обрабатывать данные в pandas и numpy;
- проводить вычисления из линейной алгебры с помощью многомерных объектов numpy;
- строить различные виды графиков с помощью библиотек matplotlib, seaborn и plotly;
- подключаться к API и базам данных, а также работать с Git;

- настраивать автоматическую рассылку метрик через телеграм-бота;
- использовать основные средства разработки: jupyter notebook, jupyterhub, Google Colab.

владеть:

- навыками решения прикладных задач с использованием python;
- навыками анализа данных с использованием python;
- навыками подготовки и обработки данных с помощью python;
- навыками выбора корректного типа визуализации и построения визуализаций в python;
- навыками настройки подключения к внешним источникам.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

| Компетенция | Содержание компетенции | Индикатор компетенции | Перечень планируемых результатов обучения по дисциплине (модулю) |
|-------------|--|-----------------------|---|
| ОПК-3. | Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта | ОПК-3.1. | Знает современные методы анализа данных, прогнозирования и поддержки принятия решений, включая технологии машинного обучения и искусственного интеллекта |
| | | ОПК-3.2. | Умеет применять программные инструменты и интеллектуальные системы для сбора, обработки и интерпретации данных в целях стратегического планирования и управления продуктом |
| | | ОПК-3.3. | Имеет практический опыт применения аналитических инструментов и технологий искусственного интеллекта для принятия обоснованных стратегических решений и прогнозирования развития бизнес-продуктов |
| ОПК-5. | Способен проводить исследования, организовывать самостоятельную и коллективную научно-исследовательскую, проектную и учебно-профессиональную деятельность для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий | ОПК-5.1. | Знает методы научных исследований и проектной деятельности, применяемые для анализа и разработки инновационных решений в сфере ИКТ |
| | | ОПК-5.2. | Умеет планировать и организовывать исследовательскую и проектную работу как индивидуально, так и в составе команды, с учётом целей и задач цифровой трансформации |
| | | ОПК-5.3. | Имеет практический опыт участия в научно-исследовательских и прикладных проектах, направленных на создание и внедрение новых ИКТ-продуктов и сервисов, включая презентацию результатов |

| | | | |
|-------|--|---------|--|
| ПК-1. | Способен выявлять и формулировать потребности целевых пользователей на основе качественных и количественных методов исследования | ПК-1.1. | Знает основные методы и инструменты проведения пользовательских исследований, включая интервью, опросы, наблюдение и анализ пользовательских данных |
| | | ПК-1.2. | Умеет планировать и проводить исследования для выявления потребностей и болей целевой аудитории, а также интерпретировать полученные данные |
| | | ПК-1.3. | Имеет практический опыт формулирования пользовательских инсайтов и требований к продукту на основе результатов качественного и количественного анализа |
| ПК-2. | Способен разрабатывать и внедрять цифровые продукты с учетом принципов устойчивого развития и цифровой инклюзивности | ПК-2.1. | Знает принципы устойчивого развития и цифровой инклюзивности, а также их применение в процессе создания цифровых продуктов |
| | | ПК-2.2. | Умеет проектировать цифровые решения, учитывающие экологические, социальные и этические аспекты, включая доступность для различных категорий пользователей |
| | | ПК-2.3. | Имеет практический опыт внедрения цифровых продуктов с соблюдением стандартов инклюзивного дизайна и оценки их воздействия на устойчивое развитие |

3. Тематический план

| № п/п | Наименование раздела дисциплины (модуля) | Трудоемкость, академические часы | | | | ТКУ (текущий контроль успеваемости) |
|----------|---|----------------------------------|-----------|----------|-------------------------------|--|
| | | <i>Очная форма</i> | | | | |
| | | Аудиторная работа | | Контроль | Самостояте льная работа | |
| Лекции | Семинары (практичес кие занятия) | | | | | |
| 1 | Введение и настройка среды разработки | 2 | 2 | | 10 | Домашние задания |
| 2 | Pandas | 10 | 10 | | 50 | Домашние задания |
| 3 | NumPy | 6 | 6 | | 30 | Домашние задания |
| 4 | Визуализация | 4 | 4 | | 22 | Домашние задания |
| 5 | API and Virtual Machine | 4 | 4 | | 22 | Домашние задания |
| | <i>Зачет с оценкой</i> | | | 4 | | Защита проекта |
| | Итого: | 26 | 26 | 4 | 134 | |
| | <i>Объем дисциплины (модуля) (в ак. ч.)</i> | 190 | | | | |
| | <i>Объем дисциплины (модуля) (в зач. ед.)</i> | 5 | | | | |

4. Содержание дисциплины (модуля)

| №п/п | Наименование раздела дисциплины (модуля) | Содержание дисциплины (модуля) по темам |
|------|---|--|
| 1 | Введение и настройка среды разработки | Онбординг |
| 2 | Pandas | Pandas Факторный анализ Маппинг и лямбда-функции Когортный анализ |
| 3 | NumPy | NumPy Прогнозирование с NumPy |
| 4 | Визуализация | Matplotlib Seaborn & Plotly |
| 5 | API and Virtual Machine | API |

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Плас, Дж. В. Python для сложных задач: наука о данных и машинное обучение : практическое руководство / Дж. В. Плас. - Санкт-Петербург : Питер, 2021. - 576 с. - (Серия «Бестселлеры O'Reilly»). - ISBN 978-5-4461-0914-2. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/1739601>.

2. Масис, С. Интерпретируемое машинное обучение на Python : практическое руководство / С. Масис. - Санкт-Петербург : БХВ-Петербург, 2023. - 640 с. - ISBN 978-5-9775-1735-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2123391>.

3. Нисчал, Н. Python — это просто. Пошаговое руководство по программированию и анализу данных : практическое руководство / Н. Нисчал. - Санкт-Петербург : БХВ-Петербург, 2022. - 416 с. - ISBN 978-5-9775-6849-4. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2139457>.

Дополнительная литература:

1. Васильев, Ю. Python для data science : практическое руководство / Ю. Васильев. - Санкт-Петербург : Питер, 2023. - 272 с. - (Библиотека программиста). - ISBN 978-5-4461-2392-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2123381>.

2. Ричардс, Т. Streamlit для Data Science. Создаем интерактивные приложения в Python : практическое руководство / Т. Ричардс, А. Груздев ; пер. с англ. А. В. Груздева. – Москва : ДМК Пресс, 2024. - 356 с. – ISBN 978-5-93700-275-4. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2205064>.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;

— специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

| № | Наименование портала (издания, курса, документа) | Ссылка |
|----|--|---|
| 1. | Научная электронная библиотека elibrary.ru библиотека | https://elibrary.ru/defaultx.asp |
| 2. | База данных для IT-специалистов | https://habr.com |
| 3. | База данных ScienceDirect | https://www.sciencedirect.com |
| 4. | Официальный сайт Министерства науки и высшего образования Российской Федерации | https://minobrnauki.gov.ru/ |
| 5. | Федеральный портал «Российское образование» | https://www.edu.ru/ |
| 6. | Информационная система "Единое окно доступа к образовательным ресурсам" | http://window.edu.ru/ |
| 7. | Единая коллекция цифровых образовательных ресурсов | http://school-collection.edu.ru/ |
| 8. | Федеральный центр информационно - образовательных ресурсов | http://fcior.edu.ru/ |

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

| Наименование ПО | Производство | Лицензионное / свободно распространяемое |
|--|---------------|--|
| Операционные системы: | | |
| Microsoft Imagine (Windows Client, Server) | зарубежное | лицензионное |
| Браузеры: | | |
| Яндекс.Браузер | отечественное | свободно распространяемое |
| Google Chrome | зарубежное | свободно распространяемое |
| Офисные приложения: | | |
| Microsoft Imagine (Visio, OneNote) | зарубежное | лицензионное |
| TeXstudio | зарубежное | свободно распространяемое |
| Adobe Acrobat Reader | зарубежное | свободно распространяемое |
| Программное обеспечение для планирования и учета времени: | | |
| Toggle app | зарубежное | свободно распространяемое |
| Системы управления проектами: | | |
| Microsoft Imagine (Project) | зарубежное | лицензионное |
| Системы управления базами данных: | | |
| Microsoft Imagine (SQL Server) | зарубежное | лицензионное |
| Системы резервного копирования (backup): | | |
| Acronis Backup Advanced for HyperV | зарубежное | лицензионное |
| Справочно-правовые системы: | | |

| | | |
|---|---------------|---------------------------|
| КонсультантПлюс: справочно-правовая система | отечественное | лицензионное |
| Средства антивирусной защиты: | | |
| Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition | отечественное | лицензионное |
| Среды разработки: | | |
| Visual Studio Code | зарубежное | свободно распространяемое |
| Bash (Unix shell) | зарубежное | свободно распространяемое |
| Anaconda | зарубежное | свободно распространяемое |
| Robotic Operating System | зарубежное | свободно распространяемое |
| CopelliaSim | зарубежное | свободно распространяемое |
| Google Colaboratory | зарубежное | свободно распространяемое |
| Пакеты программных средств и библиотек: | | |
| AutoPsy | зарубежное | свободно распространяемое |
| Interactive Disassembler (IDA) | зарубежное | свободно распространяемое |
| Системы управления библиографической информацией: | | |
| Zotero | зарубежное | свободно распространяемое |
| Сервисы и службы: | | |
| Bind | зарубежное | свободно распространяемое |
| Docker | зарубежное | свободно распространяемое |

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Python для анализа данных» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, домашние задания, проект, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар — это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники

информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Проект – исследовательская работа по дисциплине (модулю) и презентация результатов.

Для успешной подготовки к проекту рекомендуется: четко определить цели и задачи проекта; составить план работы, разбив проект на этапы с указанием сроков выполнения каждого из них; использовать разнообразные источники информации и инструменты для исследования темы; регулярно проверять прогресс и вносить коррективы в план, если это необходимо.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Python для анализа данных»

Оценивание уровня учебных достижений обучающихся по дисциплине (модулю) осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *зачета с оценкой*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

| Десятибалльная оценка | Пятибалльная оценка | Оценка за зачет | Общая характеристика результата обучения по дисциплине (модулю) |
|------------------------------|----------------------------|------------------------|---|
| 10 | Отлично | Зачтено | Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для |
| 9 | Отлично | Зачтено | |
| 8 | Отлично | Зачтено | |

| Десятибалльная оценка | Пятибалльная оценка | Оценка за зачет | Общая характеристика результата обучения по дисциплине (модулю) |
|-----------------------|---------------------|-----------------|--|
| | | | практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами. |
| 7 | Хорошо | Зачтено | Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами. |
| 6 | Хорошо | Зачтено | |
| 5 | Удовлетворительно | Зачтено | Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования. |
| 4 | Удовлетворительно | Зачтено | |
| 3 | Не сдан | Не зачтено | Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы. |
| 2 | Не сдан | Не зачтено | |
| 1 | Не сдан | Не зачтено | |

Дисциплина (модуль) «Python для анализа данных» оценивается следующим образом:

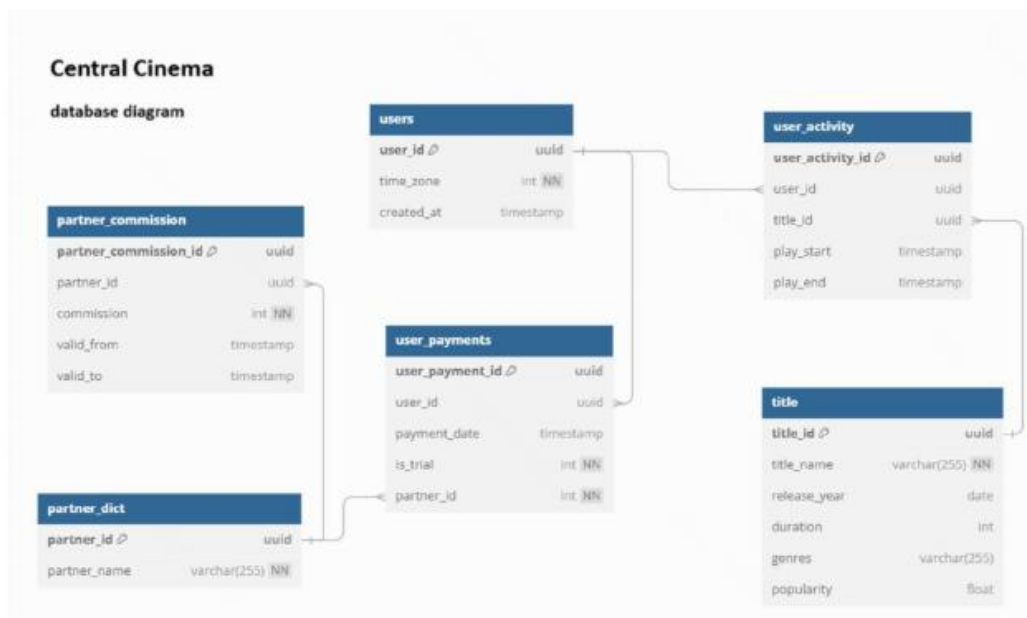
| Активность | Вес | Количество | Описание |
|------------------|-----|------------|--|
| Домашние задания | 80% | 7 | Набор задач по темам недели |
| Зачет с оценкой | 20% | 1 | Защита проекта – исследовательская работа по дисциплине (модулю) и презентация результатов |

Формула расчёта итоговой оценки по дисциплине (модулю) «Python для анализа данных»: « $0,8 \times$ среднее за домашние задания + $0,2 \times$ зачет с оценкой».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

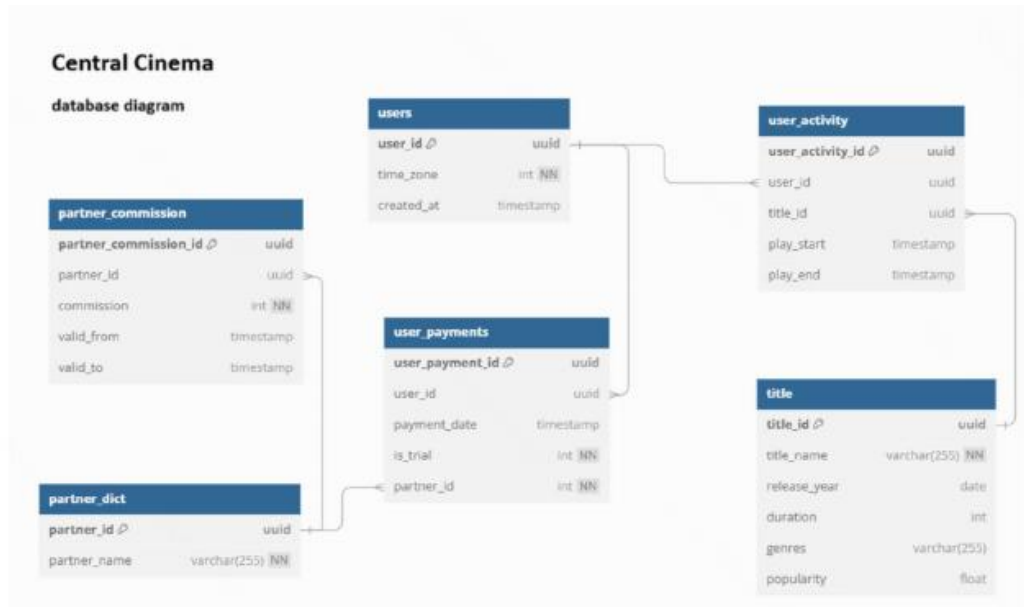
Домашнее задание 1



Описание таблиц

| | |
|---------------------------|--|
| users | справочник Пользователи |
| user_id | уникальный идентификатор клиента |
| time_zone | тайм-зона клиента |
| created_at | дата и время регистрации |
| user_payments | сведения об оплатах |
| user_payment_id | уникальный идентификатор платежа |
| user_id | идентификатор клиента |
| payment_date | дата оплаты |
| is_trial | признак триальности (подписка в подарок) |
| partner_id | идентификатор партнера, который продал подписку |
| user_activity | сведения о просмотрах фильмов |
| user_activity_id | уникальный идентификатор просмотра |
| user_id | идентификатор клиента |
| title_id | идентификатор фильма |
| play_start | дата и время начала просмотра |
| play_end | дата и время окончания просмотра |
| partner_dict | справочник Партнеры (кто продал подписку) |
| partner_id | уникальный идентификатор Партнера |
| partner_name | наименование |
| partner_commission | справочник Стоимость подписки |
| partner_commission_id | уникальный идентификатор |
| partner_id | идентификатор Партнера |
| commission | стоимость подписки в месяц в рублях |
| valid_from | дата действия цены С |
| valid_to | дата действия цены ПО |
| title | справочник Фильмы |
| title_id | уникальный идентификатор Фильма |
| title_name | наименование |
| release_year | дата выхода |
| duration | продолжительность в минутах |
| genres | жанры |
| popularity | популярность |

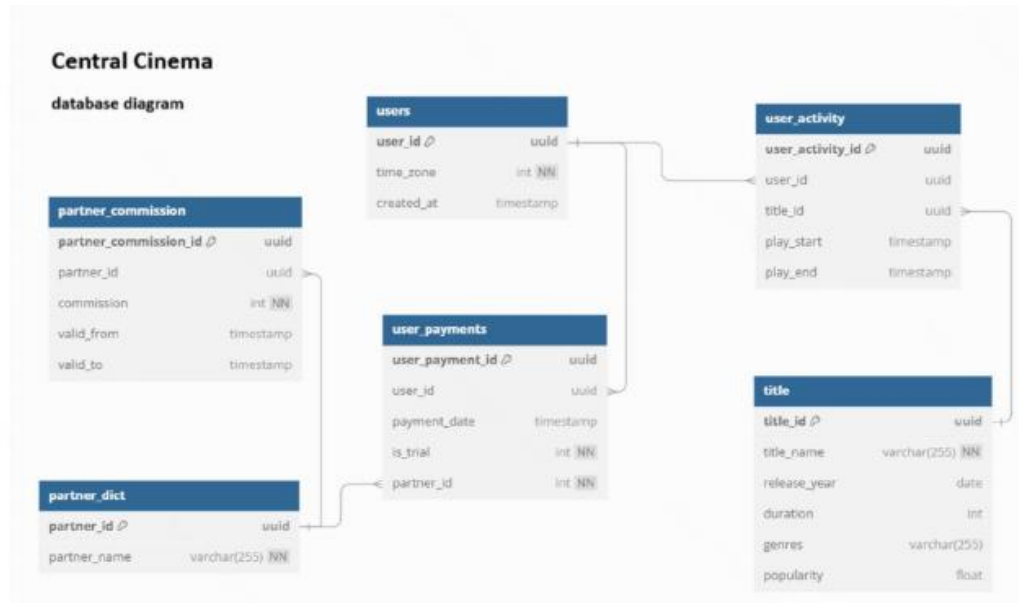
Домашнее задание 2



Описание таблиц

| | |
|------------------------------|--|
| users | справочник Пользователи |
| user_id | уникальный идентификатор клиента |
| time_zone | тайм-зона клиента |
| created_at | дата и время регистрации |
| user_payments | сведения об оплатах |
| user_payment_id | уникальный идентификатор платежа |
| user_id | идентификатор клиента |
| payment_date | дата оплаты |
| is_trial | признак триальности (подписка в подарок) |
| partner_id | идентификатор партнера, который продал подписку |
| user_activity | сведения о просмотрах фильмов |
| user_activity_id | уникальный идентификатор просмотра |
| user_id | идентификатор клиента |
| title_id | идентификатор фильма |
| play_start | дата и время начала просмотра |
| play_end | дата и время окончания просмотра |
| partner_dict | справочник Партнеры (кто продал подписку) |
| partner_id | уникальный идентификатор Партнера |
| partner_name | наименование |
| partner_commission | справочник Стоимость подписки |
| partner_commission_id | уникальный идентификатор |
| partner_id | идентификатор Партнера |
| commission | стоимость подписки в месяц в рублях |
| valid_from | дата действия цены С |
| valid_to | дата действия цены ПО |
| title | справочник Фильмы |
| title_id | уникальный идентификатор Фильма |
| title_name | наименование |
| release_year | дата выхода |
| duration | продолжительность в минутах |
| genres | жанры |
| popularity | популярность |

Домашнее задание 3



Описание таблиц

| | |
|---------------------------|--|
| users | справочник Пользователи |
| user_id | уникальный идентификатор клиента |
| time_zone | тайм-зона клиента |
| created_at | дата и время регистрации |
| user_payments | сведения об оплатах |
| user_payment_id | уникальный идентификатор платежа |
| user_id | идентификатор клиента |
| payment_date | дата оплаты |
| is_trial | признак триальности (подписка в подарок) |
| partner_id | идентификатор партнера, который продал подписку |
| user_activity | сведения о просмотрах фильмов |
| user_activity_id | уникальный идентификатор просмотра |
| user_id | идентификатор клиента |
| title_id | идентификатор фильма |
| play_start | дата и время начала просмотра |
| play_end | дата и время окончания просмотра |
| partner_dict | справочник Партнеры (кто продал подписку) |
| partner_id | уникальный идентификатор Партнера |
| partner_name | наименование |
| partner_commission | справочник Стоимость подписки |
| partner_commission_id | уникальный идентификатор |
| partner_id | идентификатор Партнера |
| commission | стоимость подписки в месяц в рублях |
| valid_from | дата действия цены С |
| valid_to | дата действия цены ПО |
| title | справочник Фильмы |
| title_id | уникальный идентификатор Фильма |
| title_name | наименование |
| release_year | дата выхода |
| duration | продолжительность в минутах |
| genres | жанры |
| popularity | популярность |

Примерные задания по проекту

Задание 1. Расширение отчётности

Текущую отчетность необходимо обогатить новыми метриками, пользуясь доступными данными.

- Придумай минимум две дополнительные метрики, которые можно рассчитать на имеющихся данных, и предоставь код на Python с вычислением данных метрик на имеющихся таблицах.

- Добавь эти метрики в текущую версию репортинга с телеграм-ботом.

В результате

- Предоставь `irunb` файл с описанием и реализацией новых метрик.
- Обнови репортинг через телеграм таким образом, чтобы расчёт и визуализация новых метрик там появились.

Критерии

1. Описано (словами в маркдаунах) минимум две метрики, которые имеют отношение к бизнесу онлайн-кинотеатра (1 балл за каждую метрику, максимум 2)

2. Для данных метрик представлены примеры кода на Python для их расчёта в динамике (1 балл за каждую метрику, максимум 2)

1. Сравнительные значения данных метрик, а также графики рассылаются телеграм-ботом (1 балл)

Задание 2. Презентация результатов

Необходимо презентовать результаты вашей работы

- Сделай слайды `pptx` с демонстрацией результатов твоей работы, а также с демонстрацией

финансовых и продуктовых результатов онлайн-кинотеатра (не забудь использовать материалы из предыдущего задания).

- Сними видео (OBS/Loom/etc.) со скринкастом слайдов и устным объяснением содержимого на 5-7 минут.

В результате

- Предоставь слайды в формате `pptx`
- Предоставь видео с объяснением слайдов

Критерии

1. Длина видео незначительно отличается от установленного коридора 5-7 мин. Ваше лицо не обязано быть на видео, но с ним, конечно же, лучше. (1.5 балла)

2. Слайды оформлены с использованием визуализаций и должны объяснять выполненную работу и бизнес онлайн-кинотеатра. (1.5 балла)

3. Первое задание, наравне с другими метриками, учтено в презентации и в видео (0.5 балла)

Задание 3. Техническое задание для DWH

Общая рекомендация по сдаче задания

Необходимо обратиться к дата инженерам и запросить доп. данные

Придумай метрики, которые невозможно рассчитать на имеющихся таблицах, но которые были бы полезны для мониторинга

В результате

Предоставь текст (не более 2000 знаков) с описанием новых метрик и данных (витрин), которые нужны для их расчета

Критерии

1. В тексте указаны необходимые таблицы (0.75 балла)

- Набор колонок
- Тип данных
- Смысловое наполнение колонок
- Временной промежутков, который хранится в таблице
- Частоту обновления таблицы

2. В тексте названы и описаны метрики, которые будут строиться на новых данных, а также описан их бизнес-смысл и интерпретация (0.75 балла)

Задания для промежуточной аттестации по дисциплине (модулю)

| № п/п | Задание | Ответ | Компетенция |
|-------|---|-----------------|-------------|
| 1. | <p>Рассмотрите приведенный ниже код:</p> <pre>import numpy as np arr = np.array([10, 15, 20, 25]) res = np.where(arr > 18, 1, 0) print(res.sum())</pre> <p>Что выведет данный код? (целым числом)</p> | 2/2.0/2,0 | ОПК-3 |
| 2. | <p>Рассмотрите приведенный ниже код:</p> <pre>import pandas as pd data = { 'City': ['A', 'B', 'A', 'B'], 'Sales': [100, 200, 50, 50] } df = pd.DataFrame(data) res = df.groupby('City')['Sales'].sum() print(res.max())</pre> <p>Какое число будет выведено в результате выполнения кода? (целым числом)</p> | 250/250.0/250,0 | ОПК-3 |
| 3. | <p>Рассмотрите приведенный ниже код:</p> <pre>import pandas as pd s = pd.Series([10, 20, 30]) s = s.apply(lambda x: x + 5 if x < 25 else x) print(s[1])</pre> <p>Что выведет данный код? (целым числом)</p> | 25/25.0/25,0 | ОПК-3 |
| 4. | <p>Даны итоговые показатели рекламной кампании:</p> <ul style="list-style-type: none"> - Marketing Cost (Расходы) = 200 - Revenue (Выручка) = 450 - Visitors (Посетители) = 1000 - Buyers (Покупатели) = 130 <p>Чему равны ROI (в долях, по прибыли) и конверсия в покупку (в долях)?</p> <p><i>В качестве ответа укажите сумму этих двух значений. Ответ округлите до двух знаков после запятой.</i></p> | 1,38/1.38 | ОПК-3 |
| 5. | <p>Какой тип визуализации в библиотеке Seaborn наиболее эффективно используется для принятия решений при факторном анализе (например, для отображения корреляционной матрицы)</p> <p>а) Круговая диаграмма (Pie chart) б) Тепловая карта (Heatmap) в) Столбчатая диаграмма (Bar plot) г) Линейный график (Line plot)</p> | б/б)/Б/Б) | ОПК-3 |

| | | | |
|-----|--|---|-------|
| 6. | <p>Установите соответствие между этапом работы с данными и программным инструментом (библиотекой), который для этого используется.</p> <p>Этап: А) Сбор данных из интернета (парсинг) Б) Математическое моделирование и прогноз В) Работа с таблицами и группировка Г) Визуализация результатов для отчета</p> <p>Инструмент: 1) Pandas 2) Plotly 3) BeautifulSoup 4) NumPy</p> <p><i>Ответ запишите в виде сочетания букв и цифр (например: АЗБ1В4Г2)</i></p> | АЗБ4В1Г2/а3б4в1г2 | ОПК-3 |
| 7. | <p>Напишите название популярной библиотеки Python, которая является стандартом для построения двумерных графиков и визуализации данных в ходе исследования. В коде её часто импортируют с псевдонимом plt. <i>(ответ на английском языке строчными буквами)</i></p> | Matplotlib/matplotlib | ОПК-5 |
| 8. | <p>Рассмотрите приведенный ниже код:</p> <pre>import pandas as pd df = pd.DataFrame({ 'A': [1, 2], 'B': [3, 4], 'C': [5, 6] }) print(df.shape)</pre> <p>Результатом выполнения является кортеж (2, 3). Что означает число 3 в данном выводе? <i>(запишите одно слово в именительном падеже)</i></p> | столбцы/столбцов/колонок/колонок/атрибутов/атрибутов/атрибутов/свойств/свойства | ОПК-5 |
| 9. | <p>Рассмотрите приведенный ниже код:</p> <pre>import pandas as pd df = pd.DataFrame({ 'val': [100, 200, None] }) print(int(df['val'].isna().sum()))</pre> <p>Что выведет данный код, направленный на поиск пропущенных значений? <i>(ответ целым числом)</i></p> | 1/1.0/1,0 | ОПК-5 |
| 10. | <p>Рассмотрите приведенный ниже код:</p> | 9/9.0/9,0 | ОПК-5 |

| | | | |
|-----|--|--|-------|
| | <pre>import pandas as pd data = { 'X': [1, 2, 3], 'Y': [4, 5, 6] } df = pd.DataFrame(data) df['Z'] = df['X'] + df['Y'] print(df['Z'].max())</pre> <p>Какое число будет выведено в результате выполнения кода? (ответ целым числом)</p> | | |
| 11. | <p>Какой формат файла (расширение) преимущественно использует интерактивная среда Jupyter Notebook, применяемая для организации исследовательской работы и написания кода на Python?</p> <p>а) .py б) .ipynb в) .txt г) .exe</p> | б/б)/Б/Б) | ОПК-5 |
| 12. | <p>Установите логически верную последовательность действий при организации процесса анализа данных:</p> <p>1) Импорт необходимых библиотек (pandas, numpy) 2) Загрузка данных из источника (csv, excel, sql) 3) Очистка и предобработка данных (удаление пропусков, дубликатов) 4) Построение графиков и интерпретация результатов</p> <p>Ответ запишите последовательностью цифр (например: 4321)</p> | 1234/1)2)3)4) | ОПК-5 |
| 13. | <p>Рассмотрите приведенный ниже код, используемый для сбора данных через API:</p> <pre>import requests response = requests.get('https://api.example.com/data') print(response.status_code)</pre> <p>Какой числовой код ответа (status code) мы ожидаем получить, если запрос прошел успешно и данные получены корректно? (ответ целым числом)</p> | 200 | ПК-1 |
| 14. | <p>В рамках анализа продукта рассчитывается метрика LTV (Lifetime Value. Известно: ARPU (средний доход на пользователя в месяц) = 500 руб. Lifetime (среднее время жизни пользователя) = 6 месяцев. Чему равен LTV? (ответ целым числом)</p> | 3000/3000,0/3000.0 | ПК-1 |
| 15. | <p>Какой HTTP-метод обычно используется для получения (запроса) данных с сервера при работе с API без изменения состояния сервера?</p> | Get/get/GET/.get()/get()/requests.get()/requests.get | ПК-1 |

| | | | |
|-----|---|--|------|
| | <i>(ответ запишите одним словом на английском языке)</i> | | |
| 16. | <p>Рассмотрите код для анализа пропущенных значений в данных:</p> <pre>import pandas as pd import numpy as np df = pd.DataFrame({ 'id': [1, 2, 3, 4], 'score': [10, np.nan, 30, np.nan] }) print(df['score'].isna().sum())</pre> <p>Какое число будет выведено? <i>(ответ целым числом)</i></p> | 2/2.0/2,0 | ПК-1 |
| 17. | <p>Какой метод в библиотеке Pandas позволяет быстро получить сводную статистику (количество, среднее, стандартное отклонение, минимум, максимум) по числовым столбцам датафрейма для первичного количественного анализа?</p> <p>а) df.info() б) df.describe() в) df.head() г) df.count()</p> | б/б)/Б/Б) | ПК-1 |
| 18. | <p>Установите правильную последовательность действий при проведении когортного анализа с использованием Python:</p> <ol style="list-style-type: none"> 1) Расчет целевых метрик (Retention, LTV) внутри групп 2) Загрузка и первичная предобработка данных 3) Интерпретация результатов и визуализация 4) Группировка пользователей в когорты (например, по месяцу первой покупки) <p><i>Ответ запишите последовательностью цифр (например: 1234)</i></p> | 2413 | ПК-1 |
| 19. | <p>В цифровом продукте необходимо реализовать функционал отправки данных формы на сервер (регистрация пользователя). Какой HTTP-метод запроса следует использовать в библиотеке requests для этой задачи?</p> <p><i>Ответ запишите одним словом на английском языке.</i></p> | Post/post/POST/ пост/Пост/.post ().post/post()/re quests.post()/req uests.post | ПК-2 |
| 20. | <p>Рассмотрите приведенный ниже код, который объединяет данные о пользователях и их покупках:</p> <pre>import pandas as pd users = pd.DataFrame({ 'id': [1, 2], 'name': ['Ian', 'Oleg'] }) buys = pd.DataFrame({ 'id': [1, 3], 'item': ['Book', 'Pen'] }) res = pd.merge(users, buys, on='id', how='inner') print(len(res))</pre> | 1/1.0/1,0 | ПК-2 |

| | | | |
|-----|---|------------------------|------|
| | Сколько строк останется в итоговом датафрейме после выполнения данного объединения? (<i>ответ целым числом</i>) | | |
| 21. | <p>При прогнозировании нагрузки на сервис используется работа с временными рядами в NumPy. Дан массив нагрузок по часам: <code>load = np.array([10, 20, 15, 30, 40])</code></p> <p>Чему равно среднее значение для среза <code>load[1:3]</code> ?</p> <p><i>Ответ укажите числом (при необходимости округлите до десятых, разделитель — точка или запятая).</i></p> | 17.5/17,5 | ПК-2 |
| 22. | <p>Для оценки эффективности цифрового продукта рассчитывается конверсия в покупку (CR). Посетители сайта: 2000 человек. Совершили покупку: 100 человек.</p> <p>Чему равен CR(в процентах) ? (<i>в ответе укажите только число</i>)</p> | 5/5.0/5,0/0.05/0,05/5% | ПК-2 |
| 23. | <p>Что означает продуктовая метрика LTV, рассчитываемая в рамках когортного анализа?</p> <p>а) Средний чек пользователя б) Стоимость привлечения одного клиента в) Прибыль компании за один день г) Совокупная прибыль, которую приносит пользователь за все время использования продукта</p> | г/г)/Г/Г) | ПК-2 |
| 24. | <p>Установите соответствие между названием библиотеки Python и её основной сферой применения в анализе данных.</p> <p>Библиотеки: А) Pandas Б) Requests В) NumPy Г) Matplotlib</p> <p>Назначение: 1) Выполнение HTTP-запросов к API 2) Работа с матрицами и линейной алгеброй 3) Обработка и анализ структурированных (табличных) данных 4) Базовая визуализация и построение графиков</p> <p><i>Ответ запишите в виде сочетания букв и цифр (например: АЗБ1В2Г4)</i></p> | АЗБ1В2Г4/а3б1в2г4 | ПК-2 |