

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО
Заместитель директора

А.И. Рыбакова

	Рабочая программа дисциплины (модуля)
по дисциплине:	Прикладная аналитика данных
по направлению:	Прикладные математика и физика
профиль подготовки:	Суперкомпьютерное моделирование в прикладной физике Физтех-школа природоподобных, плазменных и ядерных технологий им. И.В. Курчатова центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
курс:	1
квалификация:	магистр

Семестры, формы промежуточной аттестации:

- 1 (осенний) - Зачет
- 2 (весенний) - Зачет

Аудиторных часов: 90 всего, в том числе:

- лекции: 30 час.
- семинары: 60 час.
- лабораторные занятия: 0 час.

Самостоятельная работа: 180 час.

Всего часов: 270, всего зач. ед.: 6

Количество контрольных работ, заданий: 2

Программу составили:

Н.А. Салина, преподаватель
Н.А. Долгополов, преподаватель
А.А. Бузулуцкая, методист
Ж.И. Зубцова, канд. физ.-мат. наук, эксперт

Программа обсуждена на заседании центра дополнительного, дополнительного профессионального и онлайн-образования "Пуск" 28.01.2025

Аннотация

Дисциплина состоит из 7 модулей:

Модуль 1. Введение в продуктовую аналитику. Анализ продуктовых метрик и Unit экономика

Модуль 2. Организация и проведение исследований

Модуль 3. Web и app аналитика

Модуль 4. SQL для анализа данных

Модуль 5. Введение в Python. Python для автоматизации анализа данных

Модуль 6. A/B-тестирование

Модуль 7. Визуализация данных

1. Цели и задачи

Цель дисциплины

- формирование/совершенствование компетенций в области сбора, обработки, анализа и визуализации данных.

Задачи дисциплины

- понимать роли аналитика в команде и его инструментов;
- научиться работать в команде и с подрядчиками;
- сформировать умение презентовать результаты;
- научиться работать с основными типами бизнес-метрик;
- сформировать навык по построению метрик;
- сформировать умение расчета Unit экономики;
- понимать общую организацию исследований, сбора и оценки данных для исследования;
- сформировать умение анализа рынка digital-продуктов на открытых данных;
- научиться проводить конкурентный анализ;
- сформировать умение работы с Google Analytics и Яндекс Метрикой;
- научиться составлять ТЗ/карты событий;
- сформировать умение работы с Firebase и атрибуцией;
- научиться писать типовые запросы для выборки различных данных;
- сформировать умение создавать корректную структуру базы данных;
- сформировать знание основ программирования на Python;
- научиться применять синтаксис Python для написания простых программ;
- сформировать знание основных инструментов Python для анализа данных;
- научиться применять Python для сбора и обработки данных;
- сформировать умение применять Python для визуализации данных;
- научиться решать практические задачи анализа данных с помощью Python;
- сформировать умение организовывать и проводить A/B-тестирование;
- научиться делать выводы по результатам A/B-тестирования;
- сформировать умение применять A/B-тестирование для решения задач анализа данных;
- сформировать знание основных принципов визуализации данных;
- научиться наглядно представлять результаты анализа данных;
- сформировать умение использовать инструменты визуализации.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-3 Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной задачи	УК-3.1 Организует и координирует работу участников проекта, способствует конструктивному преодолению возникающих разногласий и конфликтов
	УК-3.4 Способен планировать командную работу, распределять поручения членам команды, организовать обсуждение разных идей и мнений

ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.1 Способен применять знания и навыки по использованию информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов
--	---

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны
знать:

- инструменты оценки эффективности работы аналитика;
- основные понятия веб-аналитики;
- используемые метрики;
- когнитивные законы и принципы восприятия информации человеком;
- отличия и ограничения типов данных SQL;
- отличия неявного и явного преобразования типов данных в SQL;
- роль коллекций и функций в программировании на Python;
- особенности работы с функциями в Python (множественное присваивание, упаковка и распаковка значений, аргументы по умолчанию);
- виды данных, источники данных, способы хранения данных (csv, tsv-файлы и другие);
- назначение A/B-тестирования и инструменты Python, требуемые для его выполнения;
- свойства нормального распределения;
- возможности и ограничения параметрических и непараметрических критериев;
- варианты проведения A/B-тестов;
- понятие A-A теста;
- назначение таблиц, строк и столбцов при проектировании БД;
- назначение ключевых полей в структуре таблицы;
- необходимые теоретические сведения для понимания агрегации данных;
- назначение функций в языках программирования;
- принципы работы с базой данных SQLite с помощью библиотек Python;
- основные понятия математической статистики, используемые для анализа данных;
- случаи применения Т-теста Стьюдента и критерия Манна-Уитни для проверки статистических гипотез;
- принципы формулирования, отбора и формирования гипотез;
- способы приоритизации гипотез;
- типы данных SQL;
- терминологию БД;
- виды связей между таблицами;
- назначение ключевых полей для создания структуры базы данных;
- различия в способах импорта модулей;
- структуры данных и инструменты, предоставляемые библиотекой Pandas для работы с данными;
- функциональные возможности одного инструмента для визуализации Yandex DataLens;
- обзорно основные инструменты для A/B-тестирования;
- понятие команды продукта;
- роль аналитика в команде;
- рынок вакансий;
- цели и задачи продуктовой аналитики;
- основные способы оценки эффективности работы аналитика;
- понятие продукта, как объекта исследования;
- понятие конкурентной группы и открытых источников конкурентного анализа;
- способы изменения продукта, исходя из его результатов;
- принципы работы счетчиков;
- элементы интерфейса Яндекс Метрики, их назначение;
- функционал сервиса «Яндекс Метрика»;
- последовательность действий для создания и настройки цели в Яндекс Метрике;
- функционал Яндекс Метрики для настройки отчетов;
- основные принципы работы с отчетами в Метрике;
- как использовать Вебвизор и карты для повышения конверсий сайта;
- как работать с рекламными кампаниями и контентом в сервисе «Яндекс Метрика»;
- как подключить электронную коммерцию для анализа данных;
- подключаемые внешние интеграции для Яндекс Метрики;
- как учитывать офлайн-данные в сервисе;
- для чего используются метрики и KPI;
- классификацию метрик;
- понятие Unit-экономика;
- классификации Unit-экономик;
- основные фреймворки декомпозиции метрик, зачем они нужны;
- основные этапы и методы исследований;
- методы и инструменты анализа;
- как настроить сервис AppMetrica;
- отличия сервиса AppMetrica от Яндекс Метрики, особенности работы;
- как выполняется сбор данных и его настройка в сервисе;

уметь:

- анализировать данные сервиса AppMetrica, делать выводы и предлагать рекомендации;
- выбрать и обосновать выбор вида диаграммы, наилучшим образом визуализирующую заданные метрики;
- выбрать тип данных в зависимости от контекста задачи;
- составить запрос, используя математические и логические функции в вычисляемых столбцах;
- выделять в задаче на естественном языке необходимость применения базовых конструкций языка;
- выполнять выгрузку данных с использованием библиотеки Pandas;
- решать системы линейных уравнений в Python матричным методом, решать задачи с помощью системы линейных уравнений;
- использовать функциональные возможности нескольких библиотек в одном проекте;
- выбрать способ оценки статистических данных;
- применять методы математической статистики для проверки гипотез;
- использовать Т-тест Стьюдента для проверки гипотезы;
- объяснить назначение А-А теста;
- сопоставить схему БД с предметной областью;
- составить запрос для выбора уникальных элементов столбцов;
- составить сложный запрос с использованием подзапросов;
- сопоставить и выбрать необходимую структуру данных для конкретной практической задачи;
- создать программу по описанию задачи на естественном языке с использованием пользовательских функций;
- работать с пропущенными значениями и дубликатами, содержащимися в DataFrame;
- получать информацию о DataFrame, вычислить описательные статистики для числовых данных, обратиться к элементам DataFrame по индексу и порядковому номеру, изменить индекс;
- создавать интерактивные графики с animation(), FuncAnimation(), camera.animate();
- вычислять основные статистические показатели;
- использовать свойства нормального распределения для оценки набора данных в условиях конкретной практической задачи;
- различить параметры нормального и t-распределения;
- правильно и однозначно сформулировать гипотезу для заданной предметной области;
- определить приоритетность гипотезы;
- написать и запустить запрос в эмуляторе SQL;
- составить запрос, задающий новое название (псевдоним) для столбцов;
- составить запрос, выполняющий вычисление по каждой строке таблицы с помощью вычисляемого столбца;
- написать комментарий в коде SQL;
- соотнести запрос на естественном языке с ключевыми словами SQL;
- написать программу на Python с применением модулей стандартной библиотеки для конкретной практической задачи;
- использовать коллекции и функции для написания программ на Python;
- написать собственную функцию на языке Python;
- подключаться к базе данных и выполнять запросы с использованием Python;
- визуализировать данные с помощью различных видов графиков;
- применять функции для построения основных видов графиков;
- спроектировать и реализовать свой дашборд для визуализации набора данных для конкретной практической задачи;
- найти ошибки в проектировании дашбордов и предложить их улучшения;
- выбирать метрики;
- самостоятельно выбирать инструменты для тестирования;
- спланировать тестирование: рассчитать размер выборки и время проведения тестирования, разделить пользователей на группы;
- оценивать эффективность работы аналитика;
- создавать аккаунт и устанавливать счетчик на готовый сайт;
- анализировать данные и делать выводы с использованием функционала Яндекс Метрики (отчеты, Вебвизор);
- подбирать ключевые продуктовые метрики;
- считать UNIT и рассчитывать Unit-экономику;
- применять основные фреймворки декомпозиции метрик;
- собирать и оценивать данные;
- проводить внутренний и сравнительный анализ;
- анализировать данные личного кабинета Яндекс Метрики (рекламные кампании, контент);
- разработать запрос исходя из требований бизнес-задачи и оформить его на языке SQL;
- проводить предварительную обработку данных (получать информацию о DataFrame, работать

владеть:

- методами исследования и анализа рынка;
- инструментами web и app аналитики;
- python для решения задач анализа данных;
- postgresSQL;
- yandex DataLens для визуализации данных.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Введение в продуктовую аналитику. Анализ продуктовых метрик и unit-экономика	5	10		25
2	Организация и проведение исследований	5	10		25
3	Web и app аналитика	5	10		40
4	SQL для анализа данных	3	6		22
5	Введение в Python. Python для автоматизации анализа данных	4	8		22
6	А/Б-тестирование	4	8		23
7	Визуализация данных	4	8		23
Итого часов		30	60		180
Подготовка к экзамену		0 час.			
Общая трудоёмкость		270 час., 6 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 1 (Осенний)

1. Введение в продуктовую аналитику. Анализ продуктовых метрик и unit-экономика

1.1. Введение в продуктовую аналитику. Работа в команде аналитики:

Лекция

Составляющие команды продукта

Какие бывают аналитики

Обзорная экскурсия в продуктовую аналитику

Место аналитика в команде и зоны ответственности

Практическая работа

Введение в продуктовую аналитику. Работа в команде аналитики

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов.

1.2. Основные типы бизнес метрик

Лекция

Что такое метрика

Какие бывают метрики

Как искать метрики для своего продукта

Практическая работа

Основные типы бизнес-метрик

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов.

1.3. Unit-экономика

Лекция

Что такое Unit-экономика

Зачем считать Unit-экономику

Какие бывают Unit-экономики

Считаем Unit-экономику

Практическая работа

Unit-экономика

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов:

Методы приоритизации гипотез

1.4. Декомпозиция метрик и основные фреймворки работы с метриками

Лекция

Декомпозиция метрик

Пирамида метрик

Иерархия метрик

Другие фреймворки

Практическая работа

Декомпозиция метрик и основные фреймворки работы с метриками

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов.

2. Организация и проведение исследований

2.1. Продукт. Организация и проведение исследований

Лекция

Введение

Продукт как объект исследования

Методы продуктовых исследований

Продуктовый анализ данных

Практическая работа

Продукт. Организация и проведение исследований

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов.

2.2. Анализ рынка и конкурентной среды

Лекция

Определение конкурентной среды

Открытые источники информации

Методы конкурентного анализа

Инструментарий анализа и обработка результатов

Практическая работа

Анализ рынка и конкурентной среды

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов.

2.3. Клиенты. Исследование целевой аудитории и маркетинговых кампаний

Лекция

Продуктовая матрица и CJM

Целевая аудитория

Маркетинговые кампании
Анализ данных в клиентской аналитике
Практическая работа
Клиенты. Исследование целевой аудитории и маркетинговых кампаний
Самостоятельная работа
Тест и задания для самопроверки, изучение дополнительных материалов.

3. Web и app аналитика

3.1. Введение в web-аналитику. Знакомство с системой Яндекс Метрика

Лекция
Введение в web-аналитику
Работа со счетчиком аналитики и Яндекс Метрика
Начало работы со счетчиком Яндекс Метрика
Работа в интерфейсе Яндекс Метрики
Практическая работа
Введение в web-аналитику. Знакомство с системой Яндекс Метрика
Самостоятельная работа
Тест и задания для самопроверки, изучение дополнительных материалов.

3.2. Работа с отчетами в Яндекс Метрике

Лекция
Настройка и применение целей
Принципы работы и настройка отчетов
Настройка отчетов. Роботность и семплирование
Настройка отчетов. Сегментация трафика
Применение аналитики Вебвизора для повышения конверсии
Практическая работа
Работа с отчетами в Яндекс Метрике
Самостоятельная работа
Тест и задания для самопроверки, изучение дополнительных материалов.

3.3. Расширенные возможности Яндекс Метрики для стратегического развития продукта

Лекция
Анализ рекламных кампаний Директа и других систем.mpr4
Анализ контента
Электронная коммерция
Интеграции под основные задачи
Передача офлайн-данных
Практическая работа
Расширенные возможности Яндекс Метрики для стратегического развития продукта
Самостоятельная работа
Тест и задания для самопроверки, изучение дополнительных материалов.

3.4. Основы аналитики мобильного приложения на базе AppMetrica

Лекция
Начало работы с AppMetrica
Сбор статистики с использованием AppMetrica SDK
Настройка сбора данных в AppMetrica
Трекинг мобильных приложений
Работа с отчетами в AppMetrica
Практическая работа
Основы аналитики мобильного приложения на базе AppMetrica
Самостоятельная работа
Тест и задания для самопроверки, изучение дополнительных материалов.

4. SQL для анализа данных

4.1. Основные понятия баз данных. Типы данных и синтаксис SQL

Лекция

Цели и план занятия

Терминология реляционных баз данных

Понятие о графическом представлении схемы базы данных

Типы данных. Особенности, ограничения типов данных

Конвертация типов данных

Структура запроса SELECT, обязательные блоки в запросе

Получение данных, удовлетворяющих определенным условиям

Выборка данных в определённом порядке

Ограничение выборки

Итоги занятия

Практическая работа

Основные понятия баз данных. Типы данных и синтаксис SQL

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов:

Работа с тренировочной базой данных

4.2. Агрегирование и группировка таблиц

Лекция

Цели и план занятия

Использование алиасов

Комментарии в SQL

Правила написания кода SQL, читабельность кода.

Теория агрегации данных

Использование вычисляемых столбцов

Выборка уникальных элементов столбцов.

Группировка и агрегатные функции

Использование оператора HAVING

Пример агрегации данных

Итоги занятия

Практическая работа

Агрегирование и группировка таблиц

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов

4.3. Соединение нескольких таблиц в запросе

Лекция

Цели и план занятия

Использование ключевых полей для создания связей между таблицами. Виды связей.

Объединение нескольких таблиц в запросе с помощью JOIN.

Пример использования оператора JOIN.

Использование ключевого слова UNION.

Итоги занятия

Практическая работа

Соединение нескольких таблиц в запросе

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов

4.4. Исследование и сбор данных с помощью SQL

Лекция

Цели и план занятия

Подзапросы

Примеры использования SQL в работе аналитика.
Формирование выборки для маркетинговой коммуникации.
Когортный анализ.
Поиск инсайтов в данных.
Ошибки в запросах. Стоит ли их бояться.
Итоги занятия
Практическая работа
Исследование и сбор данных с помощью SQL
Самостоятельная работа
Тест и задания для самопроверки, изучение дополнительных материалов:
Работа с тренировочной базой данных

5. Введение в Python. Python для автоматизации анализа данных

5.1. Основы программирования на Python

Лекция
О языке Python
Введение в Python
Ввод и вывод данных
Примеры простейших программ
Инструкция import
Модули стандартной библиотеки
Установка внешних библиотек Python
Правила записи кода PEP 8
Практическая работа
Основы программирования на Python
Самостоятельная работа
Тест и задания для самопроверки, изучение дополнительных материалов:
Листинги программ
Установка интерпретатора Python на Windows
Установка интерпретатора Python на Linux
Установка интерпретатора Python на MacOS
Работа в IDE PyCharm. Первая программа
Создание виртуального окружения

5.2. Типы данных и управляющие конструкции языка

Лекция
Числовые типы и операции над ними
Строковый тип данных
Логический тип данных
Условный оператор
Цикл while
Цикл for
Операторы continue и break
Практическая работа
Типы данных и управляющие конструкции языка
Самостоятельная работа
Тест и задания для самопроверки, изучение дополнительных материалов:
Листинги программ

5.3. Встроенные структуры данных Python

Лекция
Встроенные структуры данных
Множества. Операции с множествами
Строки. Индексация строк
Списки

Методы split() и join(). Списочные выражения

Кортежи

Словари

Практическая работа

Встроенные структуры данных Python

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов:

Листинги программ

Примеры работы с множествами

Примеры работы со строками

Примеры работы со списками

Примеры работы с методами split() и join()

Методы списков и строк

5.4. Функции в Python

Лекция

Именные функции, инструкция def

Возврат значений из функции

Множественное присваивание, упаковка и распаковка значений.

Аргументы по умолчанию и именованные аргументы.

Инструкция pass(). Согласованность аргументов

Функция как объект. Функции высших порядков.

Практическая работа

Функции в Python

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов:

Лямбда-функция

5.5. Получение и предобработка данных

Лекция

Виды и источники данных

Предобработка данных

Модуль sqlite3 языка Python

Подключение и работа с базами данных в Python

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов

5.6. Библиотека NumPy

Лекция

Обзор библиотек для анализа данных

Вычислительные функции библиотеки NumPy. Массивы.

Векторы. Решение линейных уравнений

Практическая работа

Получение и предобработка данных. Библиотека NumPy.

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов:

Примеры кода для работы с библиотекой NumPy

5.7. Библиотека Pandas

Лекция

Первичная работа с датафреймом

Введение в агрегирование и сводные таблицы

Базовые операции с DataFrame

Работа с пропусками и операции над данными

Работа с несколькими таблицами (Join)

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов:

Примеры кода для работы с библиотекой Pandas

5.8. Автоматизация в работе. Аналитика данных

Лекция

Обзор библиотек Python для работы с расписанием и автоматизации. Библиотека schedule

Обзор Apache Airflow для аналитика данных

Практическая работа

Библиотека Pandas. Автоматизация в работе аналитика данных

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов

5.9. Примеры использования библиотек Python в работе аналитика данных

Лекция

Пример анализа данных с помощью библиотеки Pandas

Пример анализа данных из нескольких источников

Работа с матрицей корреляции

Создание интерактивных графиков

Практическая работа

Примеры использования библиотек Python в работе аналитика данных

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов:

Примеры кода

6. А/Б-тестирование

6.1. Основные понятия математической статистики. Проверка статистических гипотез

Лекция

Цели и план занятия

А/В-тестирование

Генеральная совокупность и выборка

Основные термины, используемые в математической статистике

Виды распределений. Нормальное распределение

Параметрическое и непараметрическое тестирование

Проверка гипотезы с помощью Т-теста Стьюдента

Критерий Манна-Уитни

Варианты проведения А/В-тестов

Итоги занятия

Практическая работа

Основные понятия математической статистики. Проверка статистических гипотез

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов.

6.2. Организация А/В-тестирования

Лекция

Цели и план занятия

Отбор и формирование гипотез

Прогнозирование и планирование тестирования

Итоги занятия

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов.

6.3. Инструменты для А/В-тестирования

Лекция

Цели и план занятия

Обзор инструментов для А/В-тестирования

Калькуляторы А/В-тестов

Использование Python

Визуальные конструкторы

Прочие внешние сервисы

Выбор инструмента для A/B-тестирования

Итоги занятия

Практическая работа

Инструменты для A/B-тестирования

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов.

6.4. Примеры проведения A/B-тестов

Лекция

Подготовка к A/B-тестированию (поиск ключевой и смежных метрик)

Пример A/B-теста нового варианта лендинга

Дополнительные аспекты проведения A/B-тестов

Практическая работа

Примеры проведения A/B-тестов

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов.

7. Визуализация данных

7.1. Основные принципы визуализации данных

Лекция

Экскурс в историю и современное состояние

Какие бывают данные

Как мы воспринимаем информацию.

Практическая работа

Основные принципы визуализации данных

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов.

7.2. Разбор прикладных инструментов визуализации данных

Лекция

Yandex DataLens. Построение диаграмм без программирования

Использование библиотек Python для визуализации. Библиотека Altair

Использование библиотек Python для визуализации. Библиотека Seaborn

Практическая работа

Разбор прикладных инструментов визуализации данных

Самостоятельная работа

Задание для самопроверки, изучение дополнительных материалов.

7.3. Библиотека Matplotlib

Лекция

Введение в Matplotlib

Примеры построения графиков и их кастомизация

Кастомизация графиков

Практическая работа

Библиотека Matplotlib

Самостоятельная работа

Тест и задания для самопроверки, изучение дополнительных материалов:

Примеры кода для работы с библиотекой Matplotlib

7.4. Диаграммы в контексте: инфопанели и презентации

Лекция

Интерактивные средства и связанные представления

Презентации на основе диаграмм. Общие практики

Подготовка инфопанелей и презентаций с помощью Yandex DataLens

Подготовка инфопанелей и презентаций с помощью библиотеки Altair

Практическая работа

Диаграммы в контексте: инфопанели и презентации

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Система дистанционного обучения:

Обучающемуся необходимо наличие доступа в сеть интернет, компьютер.

Преподавателю курса необходимо наличие доступа администратора курса и оборудование для проведения дистанционных семинаров (вебинаров), качественный отказоустойчивый доступ в сеть интернет.

6.Перечень рекомендуемой литературы

Основная литература

Литература из средств кафедры:

1. Абдрахманов, М. И. Python. Визуализация данных. – М.: Matplotlib. Seaborn. Mayavi, 2020.
2. Болье, А. Изучаем SQL. Генерация, выборка и обработка данных. – М.: Диалектика, 2021.
3. Васильев, А. Н. Программирование на Python в примерах и задачах. – М.: Эксмо, 2021.
4. Доверительное А/В-тестирование / Кохави, Р., Тан, Д., Сюй, Я. – М.: ДМК Пресс, 2021.
5. Дули, Р. Нейромаркетинг. Как влиять на подсознание потребителя. – Минск: Попурри, 2021.
6. Любанович, Б. Простой Python. Современный стиль программирования. – М.: Питер, 2021.
7. Маккинни, У. Python и анализ данных. Третье издание. – М.: ДМК-Пресс, 2023.
8. Манцнер, Т. Визуализация данных. Полный и исчерпывающий курс для начинающих. – М.: Бомбора, 2023.
9. Молинаро, Э., де Грааф, Р. SQL. Сборник рецептов. 2-е изд. – СПб.: БХВ-Петербург, 2022.
10. Понкин, И. В., Лаптева, А. И. Методология научных исследований и прикладной аналитики. – М.: Буки Веди, 2023.
11. Ренвуазе, П. Тренинг по нейромаркетингу. Где находится кнопка «Купить» в сознании покупателя? – М.: Эксмо, 2019.
12. Савельев, В. Статистика и котики – М.: АСТ, 2021
13. Саммерфилд, М.; пер. с англ. А. Киселева. Программирование на Python 3: подробное руководство – М.: Символ-Плюс, 2020.
14. Сирокер, Д., Кумен, П. А/В тестирование. Лучший способ превратить посетителей сайта в покупателей. – М.: Wiley, 2020.
15. Танимура, К. SQL для анализа данных. – СПб.: БХВ-Петербург, 2024.
16. Федоров, Д. Ю. Программирование на Python: учебное пособие для вузов. – М.: Юрайт, 2020.
17. Шилдс, У. SQL: быстрое погружение. – М.: Питер, 2022.

Дополнительная литература

Дополнительная литература из средств кафедры:

1. Альварес С. Как создать продукт, который купят: Метод Lean Customer Development. — М.: Альпина Паблишер, 2018.
2. Андерсон К. Аналитическая культура. От сбора данных до результатов. — М.: Манн, Иванов и Фербер, 2017.
3. Джеффри М. Маркетинг, основанный на данных. — М.: Манн, Иванов и Фербер, 2013.
4. Дэвенпорт Т., Ким Джин Хо. О чем говорят цифры. — М.: Манн, Иванов и Фербер, 2014.
5. Курносов Ю. Алфавит аналитики. — М.: Концептуал, 2018.
6. Льюис Д. Нейромаркетинг в действии. Как проникнуть в мозг покупателя. — М.: Манн, Иванов и Фербер, 2015.
7. Портер М. Конкурентная стратегия: Методика анализа отраслей и конкурентов. — М.: Альпина Диджитал, 2016.
8. Юрьева Л. А. Основы стратегического конкурентного анализа. — Екатеринбург: Изд-во Уральского университета, 2014.
9. Введение в СУБД MySQL. — М.: Национальный открытый университет "ИНТУИТ", 2016.
10. Грабер М. Введение в SQL / пер. с англ. В. А. Ястребова. — СПб.: Лори, 1996.
11. Мартишин С. А., Симонов В. Л., Храпченко М. В. Базы данных. Практическое применение СУБД SQL и NoSQL-типа для проектирования информационных систем: учеб. пособие. — М.: ИД «ФОРУМ», ИНФРА-М, 2018.
12. Полякова Л. Н. Основы SQL: курс лекций: учеб. пособие для вузов. — М.: Интернет-Ун-т Информ. технологий, 2004.
13. Астахова И. Ф. и др. Язык SQL в примерах и задачах: учеб. пособие для вузов. — М.: Физматлит, 2009.
14. Уилтон П., Колби Дж. SQL для начинающих / пер. с англ. А. Г. Сивака. — М.: Вильямс, 2006.
15. Сузи Р. А. Язык программирования Python: учеб. пособие для вузов. — 2-е изд., испр. — М.: Интернет-Ун-т Информ. Технологий, БИНОМ. Лаб. знаний, 2007.
16. Вандер Плас Д. Python для сложных задач: наука о данных и машинное обучение: учеб. пособие для вузов / пер. с англ. И. Пальти. — СПб.: Питер, 2018.
17. Саммерфилд М. Python на практике: создание качественных программ с использованием параллелизма, библиотек и паттернов / пер. с англ. А. А. Слинкина. — Научное изд. — М.: ДМК Пресс, 2014.
18. Свейгарт Э. Автоматизация рутинных задач с помощью Python: практическое руководство для начинающих. — СПб.: Питер, 2017.
19. Федоров Д. Основы программирования на примере языка Python: учебное пособие. — СПб.: Питер, 2018.
20. Tufte E. R. The Visual Display of Quantitative Information. — 2nd ed. — Graphics Press, 2001.
21. Yau N. Visualize This: The FlowingData Guide to Design, Visualization, and Statistics. — Wiley, 2011.
22. Базалева О. И. Мастерство визуализации данных. — М.: Диалектика-Вильямс, 2018.
23. Железняы Д. Говори на языке диаграмм. Пособие по визуальным коммуникациям. — М.: Манн, Иванов и Фербер (МИФ), 2008.
24. Сирокер Д., Кумен П. А/Б-тестирование. — Wiley, 2012.

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. KPI и OKR: в чем разница - <https://www.wrike.com/ru/blog/kpi-i-okr-v-chem-raznitsa-i-pochemu-dlya-uspeshnogo-koordinirovaniya-proektov-nuzhno-i-to-i-drugoe/>
2. 12 методов приоритизации продуктовых целей - <https://vc.ru/marketing/274778-12-metodov-prioritizacii-produktovyh-celey-rice-wsjf-kano-i-prochie>
3. Гайд по продуктовым метрикам - <https://medium.com/@uxunicorn/гайд-по-продуктовым-метрикам-4ccfc7b0504b>
4. Как выбрать метрики продукта - https://skillbox.ru/media/management/kak_vybrat_metriki_dlya_produkta_razbiraem_na_primerakh/
5. Гайд по unit-экономике - <https://vc.ru/finance/48822-gayd-razobratsya-v-yunit-ekonomike-za-odin-den>
6. Unit-экономика за 30 слов - <https://gopractice.ru/product/unit-economics/>
7. Методы продуктовых исследований - <https://academy.yandex.ru/journal/11-issledovatel'skikh-metodov-dlya-prodakt-menedzherov>

8. Разработка продукта - https://portal.tpu.ru/SHARED/e/ERMUSHKO/ucheba/Tab5/05_Tema_Student.pdf
9. Метод анализа конкурентов ФАС - <https://docs.cntd.ru/document/902215421>
10. Что такое CJM - <https://roistat.com/rublog/customer-journey-map/>
11. Как определить целевую аудиторию продукта - <https://dasreda.ru/learn/blog/article/1214>
12. Атрибуция Яндекс.Метрики - <https://yandex.ru/support/metrica/reports/attribution-model.html>
13. Атрибуция MyTracker - <https://tracker.my.com/docs/tracking/attribution/about>
14. Постбэк MyTracker - <https://tracker.my.com/docs/tracking/postbacks/about>
15. 5 практических советов по разметке событий в приложении - <https://tracker.my.com/blog/210/5-prakticheskikh-sovetov-po-razmetke-custom-events-v-mobilnom-prilozhenii>
16. Интерактивный учебник по SQL - <http://www.sql-tutorial.ru/ru/content.html>
17. Официальный сайт Python - <https://www.python.org/>
18. Среда разработки IDE PyCharm - <https://www.jetbrains.com/pycharm/>
19. PEP 8 – Style Guide for Python Code - <https://peps.python.org/pep-0008/>
20. Начало работы с Power BI - <https://docs.microsoft.com/ru-ru/power-bi/fundamentals/desktop-getting-started>
21. <https://reader.lanbook.com/book/2101#1>
22. <https://e.lanbook.com/book/100713>
23. <http://znanium.com/catalog/product/926871>
24. Sample Size Calculator (Evan’s Awesome A/B Tools) - <https://www.evanmiller.org/ab-testing/sample-size.html>
25. Калькулятор достоверности АБ-тестирования - <https://mindbox.ru/academy/ab-test-calculator/>
26. <https://biconsult.ru/services/razvitie-bi-sistem-trendy-i-dvizhenie-v-storonu-abi-vzglyad-so-storony-vizualizacii>
27. <https://www.unisender.com/ru/glossary/business-intelligence/>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

Документация Postgres про сравнение строк - <https://postgrespro.ru/docs/postgrespro/9.5/functions-matching>

Документация Postgres про другие функции работы со строками - <https://postgrespro.ru/docs/postgrespro/9.5/functions-string>

Тестер регулярных выражений - <https://www.regextester.com>

Интерактивный учебник по SQL - <http://www.sql-tutorial.ru/ru/content.html>

Введение в анализ данных с помощью Pandas - <https://habr.com/ru/post/196980/>

Начало работы с Power BI - <https://docs.microsoft.com/ru-ru/power-bi/fundamentals/desktop-getting-started>

Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных – <http://www.machinelearning.ru/>

Информационная система «Единое окно доступа к образовательным ресурсам» (ИС «Единое окно») – <http://window.edu.ru/>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Самостоятельная работа подразделяется на аудиторную и внеаудиторную. Аудиторную самостоятельную работу составляют практические задания, которые выполняются слушателями во время учебных занятий, результаты ее выполнения проверяются и оцениваются преподавателем в учебном процессе.

Внеаудиторная самостоятельная работа включает формы: изучение дополнительной литературы, подготовка итоговых проектов по модулям, подготовка проекта.

Основными критериями качества организации самостоятельной работы служит наличие контроля результатов самостоятельной работы.

Основными современными формами организации самостоятельной работы являются творческие работы и работа с информационными компьютерными технологиями.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладные математика и физика
профиль подготовки:	Суперкомпьютерное моделирование в прикладной физике Физтех-школа природоподобных, плазменных и ядерных технологий им. И.В. Курчатова центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
курс:	<u>1</u>
квалификация:	магистр

Семестры, формы промежуточной аттестации:

- 1 (осенний) - Зачет
- 2 (весенний) - Зачет

Разработчики:

Н.А. Салина, преподаватель
Н.А. Долгополов, преподаватель
А.А. Бузулуцкая, методист
Ж.И. Зубцова, канд. физ.-мат. наук, эксперт

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-3 Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной задачи	УК-3.1 Организует и координирует работу участников проекта, способствует конструктивному преодолению возникающих разногласий и конфликтов
	УК-3.4 Способен планировать командную работу, распределять поручения членам команды, организовать обсуждение разных идей и мнений
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.1 Способен применять знания и навыки по использованию информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов

2. Показатели оценивания компетенций

В результате изучения дисциплины «Прикладная аналитика данных» обучающийся должен:

знать:

- инструменты оценки эффективности работы аналитика;
- основные понятия веб-аналитики;
- используемые метрики;
- когнитивные законы и принципы восприятия информации человеком;
- отличия и ограничения типов данных SQL;
- отличия неявного и явного преобразования типов данных в SQL;
- роль коллекций и функций в программировании на Python;
- особенности работы с функциями в Python (множественное присваивание, упаковка и распаковка значений, аргументы по умолчанию);
- виды данных, источники данных, способы хранения данных (csv, tsv-файлы и другие);
- назначение A/B-тестирования и инструменты Python, требуемые для его выполнения;
- свойства нормального распределения;
- возможности и ограничения параметрических и непараметрических критериев;
- варианты проведения A/B-тестов;
- понятие A-A теста;
- назначение таблиц, строк и столбцов при проектировании БД;
- назначение ключевых полей в структуре таблицы;
- необходимые теоретические сведения для понимания агрегации данных;
- назначение функций в языках программирования;
- принципы работы с базой данных SQLite с помощью библиотек Python;
- основные понятия математической статистики, используемые для анализа данных;
- случаи применения Т-теста Стьюдента и критерия Манна-Уитни для проверки статистических гипотез;
- принципы формулирования, отбора и формирования гипотез;
- способы приоритизации гипотез;
- типы данных SQL;
- терминологию БД;
- виды связей между таблицами;
- назначение ключевых полей для создания структуры базы данных;
- различия в способах импорта модулей;
- структуры данных и инструменты, предоставляемые библиотекой Pandas для работы с данными;
- функциональные возможности одного инструмента для визуализации Yandex DataLens;
- обзорно основные инструменты для A/B-тестирования;
- понятие команды продукта;
- роль аналитика в команде;
- рынок вакансий;
- цели и задачи продуктовой аналитики;
- основные способы оценки эффективности работы аналитика;
- понятие продукта, как объекта исследования;
- понятие конкурентной группы и открытых источников конкурентного анализа;
- способы изменения продукта, исходя из его результатов;
- принципы работы счетчиков;
- элементы интерфейса Яндекс Метрики, их назначение;
- функционал сервиса «Яндекс Метрика»;
- последовательность действий для создания и настройки цели в Яндекс Метрике;
- функционал Яндекс Метрики для настройки отчетов;
- основные принципы работы с отчетами в Метрике;
- как использовать Вебвизор и карты для повышения конверсий сайта;
- как работать с рекламными кампаниями и контентом в сервисе «Яндекс Метрика»;
- как подключить электронную коммерцию для анализа данных;
- подключаемые внешние интеграции для Яндекс Метрики;
- как учитывать офлайн-данные в сервисе;
- для чего используются метрики и KPI;
- классификацию метрик;
- понятие Unit-экономика;
- классификации Unit-экономик;
- основные фреймворки декомпозиции метрик, зачем они нужны;
- основные этапы и методы исследований;
- методы и инструменты анализа;
- как настроить сервис AppMetrica;
- отличия сервиса AppMetrica от Яндекс Метрики, особенности работы;
- как выполняется сбор данных и его настройка в сервисе;

уметь:

- анализировать данные сервиса AppMetrica, делать выводы и предлагать рекомендации;
- выбрать и обосновать выбор вида диаграммы, наилучшим образом визуализирующую заданные метрики;
- выбрать тип данных в зависимости от контекста задачи;
- составить запрос, используя математические и логические функции в вычисляемых столбцах;
- выделять в задаче на естественном языке необходимость применения базовых конструкций языка;
- выполнять выгрузку данных с использованием библиотеки Pandas;
- решать системы линейных уравнений в Python матричным методом, решать задачи с помощью системы линейных уравнений;
- использовать функциональные возможности нескольких библиотек в одном проекте;
- выбрать способ оценки статистических данных;
- применять методы математической статистики для проверки гипотез;
- использовать Т-тест Стьюдента для проверки гипотезы;
- объяснить назначение А-А теста;
- сопоставить схему БД с предметной областью;
- составить запрос для выбора уникальных элементов столбцов;
- составить сложный запрос с использованием подзапросов;
- сопоставить и выбрать необходимую структуру данных для конкретной практической задачи;
- создать программу по описанию задачи на естественном языке с использованием пользовательских функций;
- работать с пропущенными значениями и дубликатами, содержащимися в DataFrame;
- получать информацию о DataFrame, вычислить описательные статистики для числовых данных, обратиться к элементам DataFrame по индексу и порядковому номеру, изменить индекс;
- создавать интерактивные графики с animation(), FuncAnimation(), camera.animate();
- вычислять основные статистические показатели;
- использовать свойства нормального распределения для оценки набора данных в условиях конкретной практической задачи;
- различить параметры нормального и t-распределения;
- правильно и однозначно сформулировать гипотезу для заданной предметной области;
- определить приоритетность гипотезы;
- написать и запустить запрос в эмуляторе SQL;
- составить запрос, задающий новое название (псевдоним) для столбцов;
- составить запрос, выполняющий вычисление по каждой строке таблицы с помощью вычисляемого столбца;
- написать комментарий в коде SQL;
- соотнести запрос на естественном языке с ключевыми словами SQL;
- написать программу на Python с применением модулей стандартной библиотеки для конкретной практической задачи;
- использовать коллекции и функции для написания программ на Python;
- написать собственную функцию на языке Python;
- подключаться к базе данных и выполнять запросы с использованием Python;
- визуализировать данные с помощью различных видов графиков;
- применять функции для построения основных видов графиков;
- спроектировать и реализовать свой дашборд для визуализации набора данных для конкретной практической задачи;
- найти ошибки в проектировании дашбордов и предложить их улучшения;
- выбирать метрики;
- самостоятельно выбирать инструменты для тестирования;
- спланировать тестирование: рассчитать размер выборки и время проведения тестирования, разделить пользователей на группы;
- оценивать эффективность работы аналитика;
- создавать аккаунт и устанавливать счетчик на готовый сайт;
- анализировать данные и делать выводы с использованием функционала Яндекс Метрики (отчеты, Вебвизор);
- подбирать ключевые продуктовые метрики;
- считать UNIT и рассчитывать Unit-экономику;
- применять основные фреймворки декомпозиции метрик;
- собирать и оценивать данные;
- проводить внутренний и сравнительный анализ;
- анализировать данные личного кабинета Яндекс Метрики (рекламные кампании, контент);
- разработать запрос исходя из требований бизнес-задачи и оформить его на языке SQL;
- проводить предварительную обработку данных (получать информацию о DataFrame, работать

владеть:

- методами исследования и анализа рынка;
- инструментами web и app аналитики;
- python для решения задач анализа данных;
- postgresSQL;
- yandex DataLens для визуализации данных.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Модуль 1

Тестовые задания

1. Что из перечисленного НЕ является плюсом матричной структуры аналитики?
 - a) Аналитик глубоко погружен в конкретный продукт и знает все особенности
 - b) Аналитик имеет тесную связь со всей командой продукта
 - c) Аналитик выполняет задачи только конкретного продукта (части продукта) и глубоко погружен в специфику
 - d) Взаимозаменяемость членов команды
2. Что из перечисленного является плюсом (плюсами) горизонтальной структуры аналитики?
 - a) Обмен опытом, обучение и поддержка в команде аналитиков
 - b) Стандартизация процессов
 - c) Аналитик выполняет задачи только конкретного продукта (части продукта) и глубоко погружен в специфику
 - d) Легкая заменяемость участников команды
3. Что из перечисленного является недостатком матричной структуры аналитики?
 - a) Нет обмена опытом между командами аналитиков
 - b) Проблемы заменяемости аналитиков
 - c) Аналитики разных команд могут одновременно решать одну и ту же проблему, не зная об этом
 - d) Сложно обеспечить одинаковый уровень аналитиков во всех командах
4. Что из перечисленного является недостатком горизонтальной структуры аналитики?
 - a) Проблемы заменяемости аналитиков
 - b) Аналитик не чувствует себя полноправной частью команды продукта (бизнеса)
 - c) Аналитик может не знать деталей развития продукта/иметь неглубокое погружение в продукт
 - d) Нет обмена опытом между командами аналитиков
5. Что из перечисленного НЕ является метрикой эффективности?
 - a) daily active users
 - b) monthly active users
 - c) кол-во установок приложения
 - d) конверсия в покупку
6. Что из перечисленного является метрикой роста?
 - a) уникальные клики
 - b) monthly active users
 - c) ctr
 - d) конверсия в покупку
7. Что из перечисленного не может являться ключевыми метриками чисто подписочного продукта ?
 - a) клики на уника
 - b) monthly active users
 - c) ctr

d) рекламная выручка на юзера

8. Что из перечисленного может являться метриками продукта с рекламной монетизацией?

a) ctr карточек

b) MAU

c) выручка на 1000 сессий

d) конверсия в добавление товара в корзину

Модуль 2

1. Выберите верный вариант ответа.

Рассмотрим в качестве продукта упаковку с молоком. Что происходит с воспринимаемой ценностью молока для рядового потребителя при увеличении розничной цены в 3 раза?

a. Снижается

b. Не изменяется

c. Увеличивается

2. Выберите верные варианты ответа. Что относится к моделям создания нового продукта?

a. IDEO

b. VTEM

c. Stage-gate

d. BAH

e. SWOT

f. PEST

3. Выберите верные варианты ответа. Какие методы НЕ относятся к количественным методам исследования продукта?

a. Метод фокус-групп

b. UX-тестирование

c. Глубинные интервью

d. A/B-тестирование.

4. Выберите верные варианты ответа. Что из списка является основными группами критериев VTEM-анализа?

a. Экономические

b. Технологические

c. Производственные

d. Логистические

e. Коммуникационные

f. Политические

g. Маркетинговые

h. Ценностные

Модуль 3

1. Выберите задачи, которые можно решить с помощью веб-аналитики

Определить, какие материалы сайта больше нравятся посетителям

Узнать средний чек от продаж новым клиентам

Посчитать маржинальность бизнеса, включая затраты на труд

Выявить причины ухода с сайта

2. С помощью какого инструмента можно увидеть запись поведения пользователя на сайте в формате видео?

Карта скроллинга

Карта кликов
Вебвизор
Карта ссылок

3. Данные в отчетах Яндекс Метрики и Яндекс Директа будут отличаться, если
не указан счетчик в рекламных кампаниях Я.Директа
на сайте стоит циклический редирект
счетчик стоит не на всех страницах сайта
в отчетах есть данные по роботам

4. Первый обязательный этап при настройке отправки In-App покупок и Revenue это
Создание тестового приложения
Подключение валидации покупок
Создание объекта YMMutableRevenueInfo
Настройка отправки Ad Revenue на API Key

Модуль 4

1. Какой атрибут таблицы базы данных используют для однозначной идентификации записи?

- первичный ключ
- главный ключ
- ключевой идентификатор
- уникальный ключ

2. Сопоставьте типы данных SQL и принимаемые ими значения.

вопрос	ответ
integer	12012020
timestamp	2020-01-12 14:20:59.537
date	2020-01-12
time	14:20:59.537
varchar	'12 января'

3. Таблица 'auto', содержащая данные об автомобилях, имеет структуру:

Название поля	Назначение поля
auto_id (primary key)	уникальный идентификатор автомобиля
auto_brand	марка автомобиля
auto_color	цвет
auto_year	год выпуска
auto_number	госномер
auto_cost	стоимость

Как правильно написать SQL-запрос для вывода всей информации о десяти наиболее дорогих автомобилях, содержащихся в таблице?

- SELECT * FROM auto ORDER BY auto_cost DESC LIMIT 10
- SELECT * FROM auto WHERE auto_cost LIMIT 1, 10
- SELECT * FROM auto (1, 10) ORDER BY cost DESC
- SELECT auto_cost FROM auto ORDER BY 10 ASC

4. Таблица 'auto', содержащая данные об автомобилях, имеет структуру:

Название поля	Назначение поля
auto_id (primary key)	уникальный идентификатор автомобиля
auto_brand	марка автомобиля
auto_color	цвет

auto_year	год выпуска
auto_number	госномер
auto_cost	стоимость

Как правильно написать SQL-запрос для вывода списка госномеров автомобилей, выпущенных после 2020 года?

- SELECT auto_number FROM auto WHERE auto_year>2020
- SELECT * FROM auto WHERE auto_year<2020
- SELECT auto_number>2020 FROM auto
- SELECT auto_number FROM auto LIMIT auto_year>2020

Модуль 5

1. Какая версия языка Python является актуальной в настоящее время?

- Python 3.X
- Python 2.6.X
- Python 1.X

2. Что такое инструкция в языках программирования?

- команда для выполнения какого-либо действия
- документ с перечислением всех команд языка
- данные, которые пользователь вводит с консоли в программу
- спецсимволы, используемые для написания кода

3. Какие элементы используют для формирования структуры программы в Python? Выберите все верные варианты.

- пробелы
- двоеточие
- табуляция
- фигурные скобки
- команды языка
- знак #

4. Как выглядит структура программы на Python?

- Программа → Модуль → Инструкция
- Программа → Инструкция → Модуль
- Модуль → Программа → Инструкция
- Программа → Модуль → Выражение

5. Какой из нижеприведенных источников данных хорошо подходит для тренировки и отработки навыка анализ данных?

- Kaggle
- GitHub
- Всемирный Банк
- Росстат

6. Какое из приведенных расширений данных основано на другом языке программирования?

- json
- .sav
- .csv
- .tsv

7. Вы записали команду для получения информации о датасете df.info(). Какая информация НЕ содержится в методе info()?

вопрос с выбором одного варианта ответов

- количество пропущенных значений
- размер датасета
- имена колонок
- все это содержится в выдаче метода

8. Что будет получено на выходе при применении метода `duplicated()` к датафрейму?

I II

- a 1 4
- b 2 5
- c 3 6

- False, False, False
- False
- True
- False, False, False, False, False, False

Модуль 6

1. Что понимают под репрезентативностью данных?

- свойство выборки, которое означает, что выборка соответствует генеральной совокупности
- параметр, определяющий внешний вид нормального распределения
- свойство выборки, которое означает, что выборка имеет нормальное распределение
- свойство выборки, означающее, что ее элементы были сформированы абсолютно случайным образом

2. Выберите из списка описательные статистики, которые являются мерами изменчивости. Отметьте все верные варианты.

- дисперсия
- стандартное отклонение
- размах
- медиана
- среднее значение
- гистограмма частот

3. Какое распределение будет приобретать выборка при увеличении количества независимых случайных величин в ней?

- нормальное
- экспоненциальное
- равномерное
- биномиальное
- логонормальное

4. Вы выполняете сравнение двух выборок. В каких случаях можно использовать критерий Стьюдента для такой задачи? Выберите все верные варианты.

- Выборки сформированы из нормально распределенных генеральных совокупностей
- Сравнимые генеральные совокупности имеют одинаковую дисперсию
- Сравнимые генеральные совокупности имеют одинаковое среднее значение
- Сравнимые генеральные совокупности имеют одинаковое математическое ожидание
- Сравнимые выборки имеют дискретное распределение

Модуль 7

1. Сколько примерно тысячелетий насчитывает визуализация данных (не считая глиняных таблиц и совсем редких примеров)?

- половину тысячелетия

- одно тысячелетие
- полтора тысячелетия
- два тысячелетия
- два с половиной тысячелетия

2. Что демонстрирует пример Энскомба?

- пользу визуализации
- пользу численных расчётов

3. Какие два из трёх перечисленных значений слова «визуализация» релевантны цели визуализации данных?

- представление физического явления или процесса в форме, удобной для зрительного восприятия;
- методика направленного вызова образа;
- создание ментальной модели чего-либо.

4. К какой из областей визуализации данных ближе всего визуальная бизнес-аналитика?

- Математическая визуализация
- Научная визуализация данных
- Визуализация статистических данных (InfoViz)
- Инфографика, в т.ч. в СМИ (Infographics, Data journalism)
- Визуализация данных как искусство (Data Art)

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

1. Кейс: клиент совершил первую покупку: средний чек — 1600 рублей, стоимость привлечения клиента — 200 рублей, себестоимость продукции — 750 рублей. Затем совершил вторую покупку. Данные о второй покупке: средний чек — 300 рублей, стоимость удержания — 50 рублей, себестоимость материалов — 40 рублей. Посчитайте, какую прибыль принёс клиент, который совершил две покупки. Ответ запишите цифрой.

2. Задача: Интернет-магазин по продаже вещей для новорождённых; Компания в январе потратила 2 млн рублей на привлечение аудитории с помощью рекламы в «Google ads»; Валовая прибыль с рекламного канала в январе составила 700 тыс. рублей, в феврале — 300 тыс., в марте — 100 тыс., в апреле — 65 тыс., в мае — 15 тыс., а в июне — 0; За всё время через «Google ads» пришло 20 тыс. пользователей.

3. Просмотрите 10 записей в Вебвизоре в тестовом аккаунте Яндекс Метрики с наибольшей активностью (<https://metrika.yandex.ru/stat/visor>) и напишите 5 предложений для повышения эффективности сайта.

Ответ должен включать в себя не только предложения по повышению производительности сайта, но и обоснования - т.е. почему вы именно так считаете.

Примеры, в каком виде вы можете ответить:

- “стоит поправить верстку в первом блоке, чтобы на мобильном экране не уезжал заголовок, это видим в записи от 27.07.23 в 11.05”
- “форма на пятом блоке не кликается с первого раза, эта ошибка повторяется в 2х сеансах - от 24.09.23 в 19.09 и от 16.10.23 в 11.08, стоит проверить почему - при необходимости изменить механику работы”

4. У вас есть задача отправить из CRM объем продаж по клиентам. Напишите последовательность действий по шагам, как именно вы это сделаете

При ответе учитывайте данные из справки <https://yandex.ru/support/metrika/crm/about.html>

5. Создайте deeplink для приложения Ozon

—<https://play.google.com/store/apps/details?id=ru.ozon.app.android>

Для этого вам потребуется зарегистрироваться в Апп Метрика.

Далее - создать тестовый вариант приложения.

Ответ - созданный диплинк.

Для ответа на это задание, потребуется зарегистрироваться здесь: <https://appmetrica.yandex.ru/about> и нажать “Подключить”.

Далее ввести приложение. Помните, что вы создаете именно тестовый вариант, для того чтобы понять, как это работает, и без внесения изменений в сдк приложений данных по приложению вы не получите, но для создания диплинка нам этого варианта будет достаточно.

6. Месяц и время года

На вход программы поступает целое число — порядковый номер месяца. Напишите программу, которая выводит построчно название месяца и время года, к которому этот месяц относится. Название месяца и времени года должны выводиться с маленькой буквы.

Если вводимое число не попадает в диапазон [1; 12], программа выводит «ошибка».

Пример работы программы:

Ввод	Вывод
------	-------

7	июль лето
---	--------------

15	ошибка
----	--------

7. Уникальный список слов

Напишите программу, которая будет считывать список слов и выводить только уникальные слова в данном списке.

Каждое слово вводится в новой строке, список считывается до тех пор, пока не будет введено слово stop.

Порядок вывода слов должен соответствовать порядку ввода. Каждое слово выводится с новой строки.

Пример работы программы:

Ввод	Вывод
------	-------

apple	apple
-------	-------

cherry	cherry
--------	--------

apple	banana
-------	--------

banana	orange
--------	--------

orange	apricot
--------	---------

banana	
--------	--

banana	
--------	--

apricot	
---------	--

orange	
--------	--

stop	
------	--

8. Базовые операции с DataFrame

Примечание для редактора:

это задание с автопроверкой в ЛМС

суть задания — написать код на Python

Ваш коллега создал df — объект DataFrame и передал в работу вам:

	I	II	III
0	1.0	5.0	NaN
1	2.0	6.0	NaN
2	3.0	NaN	7.0
3	4.0	NaN	6.0

Выполните следующие операции применительно к данному объекту df:

Замените индексы строк на последовательность чисел от 1 до 4, используя соответствующий метод библиотеки Pandas.

Переименуйте названия колонок в последовательность букв А, В, С, используя соответствующий метод библиотеки Pandas.

Замените пропущенные значения числом 55.

Примечание: никакие результаты работы программы выводить на экран НЕ требуется.

Предварительная загрузка окна ответа:

```
import numpy as np
import pandas as pd
df = pd.DataFrame(np.nan, index=[0, 1, 2, 3], columns=['I', 'II', 'III'])
df.loc[0, "I"] = 1
df.loc[1, "I"] = 2
df.loc[2, "I"] = 3
df.loc[3, "I"] = 4
df.loc[0, "II"] = 5
df.loc[1, "II"] = 6
df.loc[2, "III"] = 7
df.loc[3, "III"] = 6
# далее запишите ваш код
```

9. Базовые операции с массивами NumPy

У вас есть 2 одномерных массива:

- `array = np.arange(10) ** 4` — массив платежей по играм из сервиса;
- `array_2 = np.arange(10) ** 3` — массив платежей по подпискам в сервисе.

Выполните последовательно следующие задания применительно к исходным массивам.

- 1) Определите общую сумму дохода сразу по 2-м потокам. Для этого сначала сложите полученные массивы, после чего получите сумму всех элементов. Результат сохраните в переменную `array_sum`.
 - 2) Вычислите, насколько больше денег принесла продажа игр за указанный период. Результат сохраните в переменную `array_difference`.
 - 3) Сохраните 2-й элемент массива по платежам игр в переменную `game_payments2`.
 - 4) Сохраните последний элемент массива по платежам подписок в переменную `subscription_last`.
- Обратите внимание: в заданиях НЕ требуется выполнять вывод полученных результатов на экран.

Предварительная загрузка окна ответа

```
import numpy as np
array = np.arange(10) ** 4
array_2 = np.arange(10) ** 3
# далее напишите ваше решение
```

Задание 1. Анализ различий в среднем времени, затрачиваемом студентами на учебу с использованием t-теста.

Было замечено, что студенты имеют низкую концентрацию внимания и отвлекаются во время работы с учебными материалами. Поэтому был предложен подход к изучению с использованием геймификации. Мы думаем, что таким образом студенты будут больше увлечены материалом и меньше отвлекаться. В результате сократится время работы с контентом курса.

Студенты поделены на две группы. Группа А состоит из 20 студентов, а группа В — из 24 студентов. Вам требуется проверить, есть ли статистически значимая разница в среднем количестве часов, которые две группы студентов проводят в неделю за учебой. Исходные данные о количестве затраченных часов в неделю для каждого студента были собраны и находятся в файле [Ссылка на файл Excel – `students.xlsx`] на листе «Расчет t-критерия».

Выполните сравнение результатов двух групп с использованием t-теста Стьюдента.

Для этого предлагаем вам проделать следующие шаги:

Сформулировать нулевую и альтернативную гипотезы.

Рассчитать среднее значение и стандартное отклонение для каждой выборки.

Рассчитать t-критерий.

Определить критическое значение p-value.

Сделать вывод о принятии или отвержении нулевой гипотезы, приняв уровень значимости равным 0,05.

Предложенные данные имеют нормальное распределение и равные дисперсии, поэтому вы спокойно можете использовать t-тест. В качестве решения укажите в полях ввода результаты, полученные на каждом из шагов выполнения задания.

Поля для ввода ответа

Все дробные значения, за исключением критического p-value, округлите до двух знаков после запятой. В качестве разделителя целой и дробной части используйте символ запятой «,».

	Группа А	Группа В
Среднее значение	_____	_____
Среднее квадратичное отклонение	_____	_____

Значение t-критерия _____

Критическое значение p-value (округлите до трех знаков после запятой) _____

Есть ли статистически значимое отличие между группами А и В? Запишите ответ одним словом «да» или «нет».

Рекомендации к выполнению задания:

попробуйте часть шагов задания выполнять не вручную, а с использованием Python.

Дополнительное необязательное задание:

выполните сравнение результатов двух групп, используя $p\text{-value} = 0,01$.

Задание 2. Проверка гипотезы.

По историческим данным конверсия сайта составляет 20%. При проведении акций конверсия увеличивается на 12% (до 22,4%). Каждый день на сайт приходит около 800 новых посетителей.

Продакт-менеджер предложил добавить в продукт «фишку» и ожидает, что конверсия повысится до уровня конверсии во время акций. Для проверки гипотезы команда решила воспользоваться А/В-тестированием.

Ваша задача — определить размер выборки одной группы, общий размер выборки и количество дней (целое число) для получения статистически значимого результата.

Подсказка:

для выполнения задания воспользуйтесь одним из калькуляторов для А/В-тестов;

при вычислении количества дней результат необходимо округлить в большую сторону.

Поля для ввода ответа

Объем выборки одной группы, чел. _____

Общий объем выборки, чел. _____

Количество дней эксперимента _____

Задание 3. t-тест для парных выборок в Python.

В вашем распоряжении имеются данные по результатам длительности пребывания пользователей на сайте до и после внедрения нового функционала. Данные записаны в файле data.csv в формате:

- id — ID пользователя;
- before — средняя длительность сессии до эксперимента, с;
- after — средняя длительность сессии после эксперимента, с.

id	before	after
_____	_____	_____

Используя функцию `ttest_rel()` из библиотеки `scipy.stats` выполните парный t-тест. Выведете на экран результат выполнения этой функции, например, как:

Ttest_relResult(statistic=-2.5289026942943655, pvalue=0.02802807458682508)

Примечания: никакие другие результаты работы программы выводить на экран НЕ требуется.

Задание 4. А/В-тестирование.

Выполните А/В-тест, используя датасет.

При выполнении заданий рекомендуем:

- скачать датасет и изучить представленные в нем данные;
- все задания выполнять в самостоятельно созданном Python-ноутбуке, а в LMS прикрепить ваш ноутбук с результатами.

Задание 5.

Набор данных — информация о пассажирах Титаника.

Оригинал набора данных: <https://www.kaggle.com/c/titanic/data>

Файл CSV: https://drive.google.com/drive/folders/1-gUtrFmEl5K-yJh_XuFpcs2ndPxREYso

Вопрос 1. Классифицируйте атрибуты данных в этом наборе данных. Какого типа атрибут PassengerId?

- номинальный
- ординальный
- количественный (interval)
- количественный (ratio)

Вопрос 2. Классифицируйте атрибуты данных в этом наборе данных. Какого типа атрибут Pclass?

- номинальный
- ординальный
- количественный (interval)
- количественный (ratio)

Вопрос 3. Классифицируйте атрибуты данных в этом наборе данных. Какого типа атрибут Age?

- номинальный
- ординальный
- количественный (interval)
- количественный (ratio)

Вопрос 4. Какие два вида графиков из перечисленных ниже лучше подходят для визуализации распределения процента выживших в зависимости от возраста (при условии, что хочется показать возраст непрерывной величиной или ввести около 15 разных возрастных групп)?

- scatter plot
- pie chart
- bar chart (histogram)
- line graph

Вопрос 5. Какой вид графика подходит для визуализации класса билетов выживших пассажиров?

- scatter plot
- pie chart
- line graph

Критерии оценивания

Максимальная сумма, которую можно набрать, успешно выполнив все контрольные мероприятия, составляет 100 баллов. Для получения положительной оценки «зачтено» необходимо набрать не менее 30 баллов.

Оценка «зачтено» выставляется студенту, если он показал всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка «не зачтено» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

Составляющие процесса обучения, которые оцениваются в ходе обучения, и их вклад в итоговый зачет представлены ниже:

Модуль	Вклад в зачет, %
Оценка за модуль 1	10
Оценка за модуль 2	10
Оценка за модуль 3	10
Оценка за модуль 4	10
Оценка за модуль 5	10
Оценка за модуль 6	10
Оценка за модуль 7	10
Зачет	30

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Во время проведения зачета обучающиеся могут пользоваться программой дисциплины.