

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

УТВЕРЖДЕНО

**Директор по цифровизации
образования**

Д.И. Гриц

	Рабочая программа дисциплины (модуля)
по дисциплине:	Прикладная аналитика данных
по направлению:	Бизнес-информатика
профиль подготовки:	Финансовые технологии и аналитика центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
курс:	1
квалификация:	магистр

Семестры, формы промежуточной аттестации:

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

Аудиторных часов: 135 всего, в том числе:

лекции: 60 час.
семинары: 75 час.
лабораторные занятия: 0 час.

Самостоятельная работа: 135 час.

Всего часов: 270, всего зач. ед.: 6

Программу составили:

Р.С. Кулиев, старший преподаватель
А.О. Бугрий, старший методист

Программа обсуждена на заседании центра дополнительного, дополнительного профессионального и
онлайн-образования "Пуск" 30.06.2022

Аннотация

Дисциплина состоит из 9 модулей:

Модуль 1. Введение в продуктовую аналитику

Модуль 2. Анализ продуктовых метрик и Unit экономика

Модуль 3. Организация и проведение исследований

Модуль 4. Web и app аналитика

Модуль 5. SQL для анализа данных

Модуль 6. Введение в Python

Модуль 7. Python для автоматизации анализа данных

Модуль 8. А/Б-тестирование

Модуль 9. Визуализация данных.

По итогам освоения программы обучающийся будет способен формализовать и алгоритмизировать поставленную задачу, написать программный код с использованием языков программирования, оформить код в соответствии с установленными требованиями.

1. Цели и задачи

Цель дисциплины

Целью реализации дисциплины «Прикладная аналитика данных» является формирование/совершенствование компетенций в области сбора, обработки, анализа и визуализации данных.

Задачи дисциплины

- Сформировать понимание роли аналитика в команде и его инструментов;
- сформировать умение работать в команде и с подрядчиками;
- сформировать умение презентовать результаты;
- сформировать умение работы с основными типами бизнес-метрик;
- сформировать навык по построению метрик;
- сформировать умение расчета Unit экономики;
- сформировать понимание общей организации исследований, сбора и оценки данных для исследования;
- сформировать умение анализа рынка digital-продуктов на открытых данных;
- сформировать умение проведение конкурентного анализа;
- сформировать умение работы с Google Analytics и Яндекс Метрикой;
- сформировать умение составления ТЗ/карты событий;
- сформировать умение работы с Firebase и атрибуцией;
- сформировать умение писать типовые запросы для выборки различных данных;
- сформировать умение создавать корректную структуру базы данных;
- сформировать знание основ программирования на Python;
- сформировать умение применять синтаксис Python для написания простых программ;
- сформировать знание основных инструментов Python для анализа данных;
- сформировать умение применять Python для сбора и обработки данных;
- сформировать умение применять Python для визуализации данных;
- сформировать умение решать практические задачи анализа данных с помощью Python;
- сформировать умение организовывать и проводить А/Б-тестирование;
- сформировать умение делать выводы по результатам А/Б-тестирования;
- сформировать умение применять А/Б-тестирование для решения задач анализа данных;
- сформировать знание основных принципов визуализации данных;
- сформировать умение 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
--------------------------------	-----------------------------------

ОПК-1 Способен разрабатывать стратегию развития информационных технологий, инфраструктуры предприятия и управлять её реализацией	ОПК-1.1 Применяет на практике методики оценки качества ресурсов информационных технологий, управления активами и конфигурации информационных технологий, методики определения потребностей в уровне качества ресурсов ИТ
--	--

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- Модель Lean Canvas;
- HADI-циклы;
- основные бизнес-метрики (анализ продуктовых метрик);
- матрицы BCG, ABC, XYZ (организация и проведение исследований);
- SWOT-анализ, матрица McKinsey, PESTELI-анализ, ситуационный анализ (организация и проведение исследований);
- инструмент Google Analytics;
- инструмент Yandex Metrika;
- инструмент Google Tag Manager;
- математические термины и понятия, используемые для анализа данных;
- методы статистического анализа;
- синтаксис языка запросов SQL;
- команды модификации;
- принципы работы представлений, хранимых процедур, триггеров;
- принципы работы оконных функций;
- методы оптимизации SQL-запросов;
- возможности языка Python и его особенности;
- синтаксис Python;
- базовые конструкции языка Python;
- основные библиотеки для работы с данными;
- способы визуализации данных;
- методы сбора обработки данных;
- практики проверки гипотез;
- способы проведения А/Б-теста;
- основные принципы визуализации данных;
- инструменты для визуализации данных.

уметь:

- Работать в команде и с подрядчиками;
- презентовать результаты;
- выбирать и рассчитывать продуктовые метрики и бизнес-метрики при реализации проектов (анализ продуктовых метрик);
- рассчитывать unit-экономику (анализ продуктовых метрик);
- работать в команде;
- строить модели и формулировать гипотезы для улучшения продукта и регулирования процессов анализа продуктовых метрик;
- применять алгоритмы создания запросов в SQL;
- загружать данные из БД с помощью SQL;
- выбрать наиболее оптимальный способ написания SQL-запросов для извлечения данных;
- использовать среду программирования на Python;
- писать простые программы на Python;
- выбирать инструменты для работы с данными в зависимости от условий задачи;
- решать задачи анализа данных с помощью Python;
- наглядно представлять результаты анализа данных;
- анализировать результаты А/Б-теста;
- применять А/Б-тестирование для решения типовых задач;
- выбирать способ визуализации в зависимости от условий задачи;
- наглядно представлять результаты анализа данных.

владеть:

- Методами исследования и анализа рынка;
- инструментами web и app аналитики;
- python для решения задач анализа данных;
- postgresQL;
- 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	7	7		19
2	Анализ продуктовых метрик и Unit экономика	8	8		19
3	Организация и проведение исследований	7	7		19
4	Web и app аналитика	8	8		18
5	SQL для анализа данных	6	9		12
6	Введение в Python	6	9		12
7	Python для автоматизации анализа данных	6	9		12
8	А/Б-тестирование	6	9		12
9	Визуализация данных	6	9		12
Итого часов		60	75		135
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

Роль и место аналитика в продуктовой команде. Основные инструменты аналитика. Работа с командой и подрядчиками. Решение бизнес-задач в команде. Презентация результатов команде.

2. Анализ продуктовых метрик и Unit экономика

Основные типы бизнес-метрик. Навыки построения метрик. Unit-экономика. Декомпозиция метрик и факторный анализ.

3. Организация и проведение исследований

Введение в организацию исследований. Сбор и оценка данных. Анализ рынка digital-продуктов на открытых данных. Сравнение с конкурентами. Способы анализа продукта и продуктовых матриц. Инструменты комплексного анализа рынка. Оценка емкости рынка. Конкурентный анализ. Особенности проведения исследований клиентов.

4. Web и app аналитика

Введение в веб-аналитику/инструменты. Google Analytics и Яндекс Метрика. Введение в app-аналитику/инструменты. Составление ТЗ/карта событий. Firebase и атрибуция.

Семестр: 2 (Весенний)

5. SQL для анализа данных

Введение в SQL. Работа с командами. Функции фильтрации и вычисляемые поля. Функции аналитики. Подзапросы и объединение таблиц. Команды модификации языка DML. Создание и модификации таблиц. Представления и хранимые процедуры. Переменные. Триггеры. Расширенные возможности SQL и основные ограничения. Аналитические функции. Основные особенности PostgreSQL. Оконные функции.

6. Введение в Python

Введение. Типы данных. Условия. Циклы. Модули и пакеты. Коллекции: множества, строки, списки, кортежи. Функции. Словари.

7. Python для автоматизации анализа данных

Библиотека Pandas для работы с данными. Библиотека NumPy. Обработка и визуализация с Matplotlib и Seaborn. Библиотека SciPy.

8. А/Б-тестирование

Введение в А/Б-тесты. Математическая статистика. Параметрические критерии. Непараметрические критерии. А/Б-тестирование для решения типовых задач.

9. Визуализация данных

Основы визуализации. Принципы визуализации. Диаграммы. Прикладные инструменты визуализации.

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

1. Как создать продукт, который купят: Метод Lean Customer Development, Электрон. версия печ. публикации / С. Альварес. — Москва, Альпина Паблишер, 2016
2. Прикладная математическая статистика [Текст] : для инженеров и научных работников, / А. И. Кобзарь, М., Физматлит, 2012
3. Математическая статистика, Электронная версия печатной публикации / А. А. Боровков. — Санкт-Петербург, Лань, 2021
4. Введение в реляционные базы данных и язык SQL [Текст] / Т. М. Дадашев [и др.] , Долгопрудный, МФТИ, 2002
5. СУБД : Язык SQL в примерах и задачах [Текст] / И. Ф. Астахова [и др.], М., Физматлит, 2009

6. Программирование на Python 3 : Подробное руководство [Текст] = Programming in Python 3 : [учеб. пособие для вузов] / М. Саммерфилд; пер. с англ. А. Киселева .— СПб : Символ-Плюс, 2015 .— 608 с.

7. Программирование на языке высокого уровня Python , учебное пособие для вузов/Д. Ю. Федоров,-Москва, Юрайт, 2020

Дополнительная литература

1. Прикладная статистика [Текст] / И. А. Палий - М.Высшая школа,2004
2. Введение в SQL [Текст] = Understanding SQL / М. Грабер ; пер. с англ. В. А. Ястребова .— М. : Лори, 1996 .— 375 с.
3. СУБД: язык SQL в примерах и задачах, Электронная версия печатной публикации / И. Ф. Астахова, И. Ф. Астахова, В. М. Мельников, В. В. Фертников. — Москва, Физматлит, 2009
4. Основы SQL [Текст] : курс лекций : учеб. пособие для вузов / Л. Н. Полякова .— М. : Интернет-Ун-т Информ. технологий, 2004 .— 368 с.
5. SQL для начинающих [Текст] / П. Уилтон, Дж. Колби ; пер. с англ. А. Г. Сивака .— М. : Вильямс, 2006 .— 496 с.
6. Язык программирования PYTHON [Текст] : учеб. пособие для вузов / Р. А. Сузи .— 2 изд., испр. — М. : Интернет-Ун-т Информ. Технологий : БИНОМ. Лаб. знаний, 2007 .— 326 с.
7. Python на практике [Текст], создание качественных программ с использованием параллелизма, библиотек и паттернов/М. Саммерфилд, -М, ДМК Пресс, 2014

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

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>

<https://reader.lanbook.com/book/2101#1>

<https://e.lanbook.com/book/100713>

<http://znanium.com/catalog/product/926871>

<https://biconsult.ru/services/razvitie-bi-sistem-trendy-i-dvizhenie-v-storonu-abi-vzglyad-so-storony-vizualizacii>

<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/>

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:	Бизнес-информатика		
профиль подготовки:	Финансовые технологии и аналитика	▲	▲
	онлайн-образования "Пуск"	▲	▲
	онлайн-образования "Пуск"		
курс:	1		
квалификация:	магистр		

Семестры, формы промежуточной аттестации:

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

Разработчики:

Р.С. Кулиев, старший преподаватель
А.О. Бугрий, старший методист

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-1 Способен разрабатывать стратегию развития информационных технологий, инфраструктуры предприятия и управлять её реализацией	ОПК-1.1 Применяет на практике методики оценки качества ресурсов информационных технологий, управления активами и конфигурации информационных технологий, методики определения потребностей в уровне качества ресурсов ИТ

2. Показатели оценивания компетенций

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

знать:

- Модель Lean Canvas;
- HADI-циклы;
- основные бизнес-метрики (анализ продуктовых метрик);
- матрицы BCG, ABC, XYZ (организация и проведение исследований);
- SWOT-анализ, матрица McKinsey, PESTELI-анализ, ситуационный анализ (организация и проведение исследований);
- инструмент Google Analytics;
- инструмент Yandex Metrica;
- инструмент Google Tag Manager;
- математические термины и понятия, используемые для анализа данных;
- методы статистического анализа;
- синтаксис языка запросов SQL;
- команды модификации;
- принципы работы представлений, хранимых процедур, триггеров;
- принципы работы оконных функций;
- методы оптимизации SQL-запросов;
- возможности языка Python и его особенности;
- синтаксис Python;
- базовые конструкции языка Python;
- основные библиотеки для работы с данными;
- способы визуализации данных;
- методы сбора обработки данных;
- практики проверки гипотез;
- способы проведения А/Б-теста;
- основные принципы визуализации данных;
- инструменты для визуализации данных.

уметь:

- Работать в команде и с подрядчиками;
- презентовать результаты;
- выбирать и рассчитывать продуктовые метрики и бизнес-метрики при реализации проектов (анализ продуктовых метрик);
- рассчитывать unit-экономику (анализ продуктовых метрик);
- работать в команде;
- строить модели и формулировать гипотезы для улучшения продукта и регулирования процессов анализа продуктовых метрик;
- применять алгоритмы создания запросов в SQL;
- загружать данные из БД с помощью SQL;
- выбрать наиболее оптимальный способ написания SQL-запросов для извлечения данных;
- использовать среду программирования на Python;
- писать простые программы на Python;
- выбирать инструменты для работы с данными в зависимости от условий задачи;
- решать задачи анализа данных с помощью Python;
- наглядно представлять результаты анализа данных;
- анализировать результаты А/Б-теста;
- применять А/Б-тестирование для решения типовых задач;
- выбирать способ визуализации в зависимости от условий задачи;
- наглядно представлять результаты анализа данных.

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

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

Критерии оценивания

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.

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

Модуль 1

Тема 1. Роль и место аналитика в продуктовой команде

Что из перечисленного не относится к градации профессии аналитика?

- а) junior analyst
- б) chief data officer**
- в) middle analyst
- г) senior analyst

Найти 3–5 хороших вакансий продуктовых аналитиков. Выписать требования к знаниям и опыту. Сравнить с тем, что вы уже умеете и знаете и составить собственный план развития.

Тема 2. Основные инструменты аналитика

Выберите из перечисленного основные инструменты аналитика (несколько вариантов ответа)

- а) miro
- б) Microsoft Excel и его аналоги**
- в) Figma
- г) Google Analytics**

Тема 3. Работа с командой и подрядчиками

Выберите основные параметры характеристик, которые являются для продукта Must have (несколько вариантов ответа)

- А) никак не влияют на удовлетворенность потребителей
- Б) их отсутствие вызывает раздражение и непонимание**
- В) их присутствие очень раздражает пользователей
- Г) если они есть и хорошо реализованы, то пользователи этого почти не замечают**
- Д) чем лучше они реализованы, тем больше положительных эмоций вызывают

Тема 4. Решение бизнес-задач в команде

Выберите продукт (или придумайте свой) и опишите его по шаблону Lean Canvas, как будто только-только собираетесь выпустить его на рынок. Можно выполнять задание в Google Docs, Miro и т. д.

Модуль 2

Тема 1. Основные типы бизнес-метрик

Выберите любые 2–3 продукта. Выпишите все метрики продукта, за которыми будете следить;

Выделите среди всех метрик ключевые (3–5 штук). Почему вы выбрали именно их?

Тема 2. Навыки построения метрик

Для своего продукта постройте воронку, разнесите свои метрики по шагам воронки.

Определите проблемные шаги в своей воронке по цифрам конверсии, поясните, из-за чего такие проблемы могут быть. Если цифр нет, предположите, на каких шагах могут потенциально быть проблемы. Обоснуйте свой ответ.

Тема 3. Unit-экономика

Для своего продукта опишите, что является «единицей» («юнитом»).

Допустим, ваш бюджет на привлечение пользователей в текущем месяце составляет 500 тысяч рублей. Для своего продукта посчитайте unit-экономику, как в примере. Рассчитайте LTV, CPA и ROI. «Сходится» ли unit-экономика?

Тема 4. Декомпозиция метрик и факторный анализ

1. Вы делаете приложение для занятия спортом (например, бегом). К вам неожиданно прибегает CEO и говорит: ”У нас упал Retention!!! Почему так?!”

Что вы будете делать? Опишите пошаговый план – чек-лист, того что вы будете делать, что спрашивать, что проверять и т.д.

2. Продолжите работать по своему продукту. Представьте я ваш руководитель. Задайте мне 3 вопроса про развитие продукта.

Модуль 3

Тема 1. Введение в организацию исследований. Сбор и оценка данных

Представьте, что вы аналитик VK. Вам нужно провести исследование и выяснить причины роста оттока работодателей.

- Сформулируйте проблему исследования;
- Определите цели и задачи исследования;
- Определите объект исследования;
- Определите перечень показателей, которые будете анализировать, источники информации и затраты времени на анализ;
- Определите примерную стоимость исследования (если понадобятся внешние данные);
- Запишите несколько потенциально возможных выводов по вашей проблеме.

Тема 2. Анализ рынка digital-продуктов на открытых данных. Сравнение с конкурентами. Способы анализа продукта и продуктовых матриц

1. На примере вашего проекта попробуйте выделить товары / продукты по матрице BCG. Достаточно взять для примера любые 4 товара любых направлений и т.п. Относительную долю рынка можете определять примерно — если больше, чем у основного конкурента, то вы будете в квадранте с высокой долей, и наоборот;

2. Если у вас нет своего проекта, воспользуйтесь данными из прикрепленного файла;

3. По данным прикрепленного файла проведите ABC и XYZ-анализ по выручке.

Тема 3. Инструменты комплексного анализа рынка. Оценка емкости рынка

1. На примере вашего проекта проведите подробный SWOT-анализ: опишите факторы внешней и внутренней среды вашего продукта или компании, сильные и слабые стороны, угрозы и возможности;

2. Кратко опишите возможную стратегию работы с выявленными факторами: усиливаем слабые, используем сильные.

Если по какой-то причине вы не можете проводить анализ собственного проекта, в качестве основы можете взять VK либо сделайте персональный SWOT-анализ.

Тема 4. Конкурентный анализ. Особенности проведения исследований клиентов

Представьте, что вам нужно провести глубинное интервью по вашему продукту.

1. Определите основной вопрос, ответ на который вы хотели бы получить;
2. Опишите сегмент клиентов, которых вы бы хотели опросить в первую очередь;
3. Сформулируйте 5–7 главных вопросов, которые вы бы хотели задать респонденту при проведении глубинного интервью. Важно, что эти вопросы должны помочь вам достичь основной цели, поставленной в п.1.

Модуль 4

Тема 1. Введение в веб-аналитику/инструменты

Распределите метрики веб-аналитики на собственной доске в Miro в соответствии с типом сайта:

- e-commerce (М.Видео, например);
- новостной сайт (контентный проект);
- бренд-сайт с переходом на сайт партнёра.

Тема 2. Google Analytics и Яндекс Метрика

1. Создайте ресурс GA4 и тег GA4 в GTM с триггером All pages для тестового сайта.
2. Сделайте анализ удержания пользователей за последние 6 месяцев, скриншот репорта приложите к анализу.
3. Зайдите в раздел «Аудитории» и настройте аудиторию «Недавно проявлявшие активность», в названии укажите вашу фамилию.

Тема 3. Введение в app-аналитику/инструменты

1. Привести еще по 1 примеру систем аналитики Рекламные/in-App/OpenSource/Crash

Тема 4. Составление ТЗ/карта событий

1. Пройти регистрацию в сервисе Google Cloud Platform и добавить тестовые данные приложения Flood-it
2. Показать скриншотом результат выполнения запроса к базе

3. Подключиться к тестовым данным
4. Сделать DashBoard с данными

Тема 5. Firebase и атрибуция

Задание 1

Представьте, что вы аналитик приложения VK. Сделайте в карте ивентов описание событий по выкладке поста. Какие ивенты вы будете считать? На какие моменты стоит обратить внимание ?

Задание 2

Создайте воронку из ивентов для публикации поста. Покажите в какой последовательности идут ивенты.

Задание 3

Придумайте AB-test, который можно провести по изменению интерфейса Instagram и опишите все шаги для его создания.

Задание 4

Создайте этот AB-test в своем кабинете. Укажите название проекта.

Модуль 5

1. Чем различаются запросы с UNION и UNION ALL?
2. Какая агрегатная функция используется для расчета суммы?
3. Запрос, возвращающий все значения из таблицы «Countries», за исключением страны с ID=8, имеет вид:
 - SELECT * FROM Countries EXP ID=8
 - SELECT * FROM Countries WHERE ID !=8
 - SELECT ALL FROM Countries LIMIT 8
4. Напишите запрос для выборки данных из таблицы «Customers», где условием является проживание заказчика в городе Москва
5. Заполните пробелы в запросе «SELECT ___, Country FROM ___ », который возвращает имена заказчиков и страны, где они находятся, из таблицы «Customers».

Модуль 6

1. bool(0.000001) - True или False?
2. Что получится в результате выполнения "a"*3 ?

*А: Строка "aaa"

В: Упадет исключение
3. Какое расширение обычно дают файлам с кодом на Python?
4. pi = 3.1415926 pi_fmt = f"{pi:#0.2f}"

Чему будет равен pi_fmt?

*А: Строке "3.14"

В: Числу 3.14

5. Какая утилита позволяет ставить внешние Python пакеты в систему?

6. `def foo(*args, **kwargs): pass`

Что произойдет при вызове функции `foo`?

А: Синтаксическая ошибка

*В: Все позиционные аргументы запишутся в кортеж `args`

С: Все именованные аргументы запишутся в кортеж `kwargs`

*D: Все именованные аргументы запишутся в словарь `kwargs`

7. В каком случае можно вызвать функцию без параметров?

8. К чему приведет обращение к непустому списку по индексу «-1»?

9. В чем отличие стандартного метода списка `sort` и встроенной функции `sorted`?

Модуль 7

1. Что вернет следующий код: `np.array([[1, 2], [1, 2]]) + np.array([[1, 2]])` ?

А: `array([[2, 3], [3, 4]])`

В: `array([[2, 4], [1, 2]])`

*С: `array([[2, 4], [2, 4]])`

2. Как с помощью `numpy` можно посчитать произведение двух матриц $A \times B$?

А: $A * B$

*В: `np.matmul(A, B)`

С: `np.matmul(B, A)`

D: `np.cross(A, B)`

3. Можно ли сохранять таблицы со сложными иерархическими индексами в `.csv`?

4. В каком случае предпочтительнее заполнять медианой, а не арифметическим средним?

*А: Данные о зарплатах людей. Есть некоторое количество людей с зарплатами в несколько раз выше, чем остальные.

В: Данные о температуре. Распределены нормально, все значения укладываются в интервал от 15 до 25 градусов.

5. Как можно настраивать шрифт для подписей делений на оси X?

A: matplotlib не позволяет этого сделать

*B: передав параметр fontdict в функцию set_xticklabels

C: передав параметр fontdict в функцию set_xticks

6. За что отвечает параметр density в методе plt.hist?

7. Почему важно, чтобы при построении boxplot с помощью seaborn в столбце были только численные значения?

8. Почему обычно не рекомендуется использовать параметр inplace=True?

Модуль 8

1. В каких ситуациях лучше не использовать АБ-тестирование?

- Когда мы плохо понимаем, как и зачем пользователь взаимодействует с нашим продуктом, — в таком случае лучше провести фокус-группу.
- Когда мы проверяем эффективность какой-то совершенно новой функции нашего продукта, введение которой полностью меняет пользовательский сценарий.
- Когда мы проверяем эффективность небольшого изменения: вряд ли оно сильно влияет на пользователей, поэтому, чтобы заметить эффект, нужно очень много испытуемых. Делать на них АБ-тест дорого, дешевле провести опрос.
- Когда мы хотим исследовать взаимодействие пользователей с продуктом в максимально естественной среде: АБ-тест слишком искусственен, лучше провести опрос и явно выяснить у пользователей, нравится ли им изменение.

2. Как провести АБ-тест в случае, когда отсутствует возможность напрямую измерить КР! (ключевой бизнес показатель) метрику непосредственно во время эксперимента?

- Принять решение о том, что интересующее изменение не будет внедрено, так как нет возможности его предварительно оценить.
- В этой ситуации проведение АБ-теста невозможно, нужно выбрать другую методику тестирования изменения — например, провести опрос или фокус-группу.
- Использовать промежуточную прокси-метрику, которая согласована с КР!-метрикой и которую можно измерить непосредственно во время эксперимента.
- Принять решение о внедрении изменения по offline метрике качества.

3. В ходе исследования влияния лесозаготовки на биоразнообразие лесов острова Борнео собраны данные о количестве видов деревьев в 12 лесах, где вырубка не ведётся:

[22, 22, 15, 13, 19, 19, 18, 20, 21, 13, 13, 15,]

и в 9 лесах, где идёт вырубка:

[17, 18, 18, 15, 12, 4, 14, 15, 10.]

Проверьте гипотезу о равенстве среднего количества видов в двух типах лесов против односторонней альтернативы о снижении биоразнообразия в вырубаемых лесах. Используйте ранговый критерий. Чему равен достигаемый уровень значимости? Округлите до четырёх знаков после десятичной точки.

4. 28 января 1986 года космический шаттл "Челленджер" взорвался при взлёте. Семь астронавтов, находившихся на борту, погибли. В ходе расследования причин катастрофы основной версией была неполадка с резиновыми уплотнительными кольцами в соединении с ракетными ускорителями. Для 23 предшествовавших катастрофе полётов "Челленджера" известны температура воздуха и появление повреждений хотя бы у одного из уплотнительных колец.

С помощью бутстрепа постройте 95% доверительный интервал для разности средних температур воздуха при запусках, когда уплотнительные кольца повреждались, и запусках, когда повреждений не было. Чему равна его ближайшая к нулю граница? Округлите до четырёх знаков после запятой.

Чтобы получить в точности такой же доверительный интервал, как у нас:

- установите `random seed = 0` перед первым вызовом функции `get_bootstrap_samples`, один раз
- сделайте по 1000 псевдовыборок из каждой выборки.

Модуль 9

1. Отметьте оппонентные пары цветов:

- *Красный—зелёный
- *Синий—жёлтый
- Красный—синий
- Жёлтый—зелёный
- Чёрный—белый
- Жёлтый—красный

2. Какой из визуальных атрибутов наиболее пригоден для отображения точных количественных данных?

- *Длина двумерного объекта
- Угол
- Объём трёхмерной фигуры в плоской проекции
- Насыщенность цвета

3. Где чаще всего встречается семантическое масштабирование (semantic zooming)?

- *в средствах визуализации геопозиционных данных
- в средствах визуализации временных данных
- в средствах визуализации иерархических данных

4. Какой из вариантов столбчатых диаграмм предпочтителен?

- *С плоскими столбиками — меньше погрешность в «считывании» данных.
- С трёхмерными столбиками — человеку привычнее изображение 3D

объектов.

5. Как выравниваются данные в столбцах таблиц, содержащих короткие текстовые значения?

- по правому краю
- *по левому краю
- по центру

В каком случае НЕ следует визуализировать данные непрерывной линией?

- *Значения по горизонтальной шкале не допускают естественной интерполяции
- Значения по вертикальной шкале меняются в небольшом интервале
- Значения по горизонтальной шкале меняются в небольшом интервале
- Значение по вертикальной шкале имеют слишком большой разброс

Тест

1. `bool(0.000001)` - True или False?

2. Что получится в результате выполнения `"a"*3` ?

*A: Строка "aaa"

B: Упадет исключение

3. Что вернет следующий код: `np.array([[1, 2], [1, 2]]) + np.array([[1, 2]])` ?

A: `array([[2, 3], [3, 4]])`

B: `array([[2, 4], [1, 2]])`

*C: `array([[2, 4], [2, 4]])`

4. Как с помощью numpy можно посчитать произведение двух матриц A x B?

A: `A * B`

*B: `np.matmul(A, B)`

C: `np.matmul(B, A)`

D: `np.cross(A, B)`

5. Напишите запрос для выборки данных из таблицы «Customers», где условием является проживание заказчика в городе Москва

6. Заполните пробелы в запросе «SELECT ___, Country FROM ___ », который возвращает имена заказчиков и страны, где они находятся, из таблицы «Customers».

7. Сколькими способами можно определить последовательность 5 гостей за круглым столом?

- 24
- 10
- 120
- 12

8. Статистическая гипотеза — это:

а) любое предположение, используемое в статистическом исследовании;

б) предположение, которое можно проверить с использованием имеющейся статистической информации;

в) научное предположение, выдвигаемое для объяснения какого-либо явления и требующее проверки на опыте.

9. Какой принцип отвечает за то, что точки, лежащие на одной линии, воспринимаются как единая группа?

- Замкнутость/завершение (closure)
- Метастабильность (множественная стабильность)
- Продолжение (continuation)
- Близость (proximity)
- Принадлежность общей области (common region)

10. В каком случае задача выбора конфигурации подписей на диаграмме решается за гарантированно полиномиальное время (не является NP-трудной)?

- При не более чем двух вариантах на каждую подпись.
- При не более чем трёх вариантах на каждую подпись.
- При не более чем четырёх вариантах на каждую подпись.
- При не менее чем четырёх вариантах на каждую подпись.

11. Для своего продукта опишите, что является «единицей» («юнитом»).

12. Допустим, ваш бюджет на привлечение пользователей в текущем месяце составляет 500 тысяч рублей. Для своего продукта посчитайте unit-экономику, как в примере. Рассчитайте LTV, CPA и ROI. «Сходится» ли unit-экономика?

В зачетно-экзаменационную ведомость оценка выставляется в соответствии с нижеприведенной таблицей 1.

Таблица 1

Сумма баллов	Оценка
50-100	Зачтено
менее 50	Не зачтено

Составляющие процесса обучения, которые оцениваются в ходе обучения, и их вклад в итоговую оценку представлены в таблице 2.

Таблица 2

№ п/п	Основные показатели оценки	Вклад в итоговую оценку
1	Практические занятия	30%
2	Выполнение домашних заданий	40%
3	Промежуточная аттестация	30%

3.2. Оценочные материалы

Таблица 3

Наименование модуля	Формы и методы контроля и оценки	Вес задания
Основы программирования на Python	Практические задания по темам лекций Домашние задания в системе дистанционного обучения по темам лекций	14
Структуры данных и функции	Практические задания по темам лекций Домашние задания в системе дистанционного обучения по темам лекций	14
Объектно-ориентированное программирование	Практические задания по темам лекций Домашние задания в системе дистанционного обучения по темам лекций	14
Углубленный Python	Практические задания по темам лекций Домашние задания в системе дистанционного обучения по темам лекций	14
Многопоточное и асинхронное программирование	Практические задания по темам лекций Домашние задания в системе дистанционного обучения по темам лекций	14
Промежуточная аттестация	Итоговое тестирование	30