

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Писарев Сергей Станиславович
Должность: Ректор
Дата подписания: 08.09.2025 13:34:26
Уникальный программный ключ:
b9d7463b91f434da3d4dc1afa9a0cf32d3c58650

**Государственное образовательное учреждение высшего образования
«Школа управления СКОЛКОВО»**

Утверждено
ректор С.С. Писарев
ШКОЛА УПРАВЛЕНИЯ
СКОЛКОВО
“21” августа 2025 г.
ГОРН 1237700448413
МОСКВА * БИРЮЛОВО

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Визуализация и моделирование**

Направление подготовки	38.03.02 Менеджмент
Квалификация выпускника	Бакалавр
Образовательная программа	Управление и предпринимательство
Форма обучения	Очная
Рабочая программа дисциплины разработана	

Трудоемкость		Контактная работа		Самостоятельная работа	Форма контроля	Семестр
з.е.	часы	лекции	семинарские занятия			
2	74	24	24	24	Экзамен	3

**Москва
2025**

1. АННОТАЦИЯ ДИСЦИПЛИНЫ

Дисциплина предполагает формирование у студентов компетенций в области анализа, визуализации и интерпретации данных с использованием современных методов статистики, машинного обучения и инструментов визуализации (включая Python). Дисциплина направлена на развитие навыков критического анализа данных, построения гипотез и эффективного представления результатов.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В случае успешного освоения курса студенты будут:

знать

- основы постановки задач отбора данных и постановки задач статистического исследования;
- основные методы визуализации данных и статистического анализа;
- основные этапы анализа данных;
- основы математической статистики;
- основные классы задач машинного обучения, их достоинства и недостатки;
- основы постановки задач отбора данных и постановки задач экспериментального исследования;
- основные модули языка Python;
- онлайн-ресурсы и прикладные программы для визуализации данных;
- основные методы и принципы проведения исследования и представления его результатов;
- роли и этапы командной работы в аналитических проектах;

уметь

- поставить задачу исследования, собрать необходимые данные, провести предобработку и анализ данных в рамках исследования;
- подобрать способ визуализации и анализа данных с учетом характеристик данных;
- подобрать средство визуализации данных с учетом поставленной задачи и имеющихся ресурсов;
- собрать набор данных для проведения исследования;
- сформулировать гипотезу исследования, обосновать теоретическую и практическую значимость результатов анализа данных;
- планировать этапы выполнения проекта с учетом ограниченных сроков;
- распределять задачи в команде и координировать совместную работу;

владеть

- навыком применения языка Python для предобработки данных, визуализации результатов исследования;
- навыком применения языка Python для подготовки данных к проведению исследования;
- навыком визуализации данных и критической оценки выбранной модели визуализации и анализа данных;
- навыком визуализации данных при помощи современных программных решений;
- навыками самоорганизации при выполнении многоэтапных заданий.

Дисциплина направлена на развитие следующих компетенций и их индикаторов:

Код компетенции	Формулировка компетенции и/или ее индикатора (ов)
УК-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения

	поставленных задач
УК-1-1.	Анализирует задачу, осуществляет ее декомпозицию, определяет приоритетность и этапность действий, направленных на решение задачи
УК-1-2.	Находит и критически анализирует информацию, необходимую для решения поставленной задачи
УК-1-3.	Выбирает оптимальный вариант решения задачи, аргументируя свой выбор
УК-2.	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений
УК-2-1.	Ставит задачи, необходимые для достижения цели с учетом правовых норм
УК-2-2.	Рассматривает возможные, в том числе нестандартные решения задач, оценивает достоинства и риски возможных решений, выбирает оптимальные решения с учетом ресурсов и ограничений
УК-3.	Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде
УК-3-1.	Знает принципы эффективной командной работы; участвует в распределении ролей в команде, взаимодействует с членами команды в соответствии со своей ролью
УК-3-2.	Участвует в обмене информацией, знаниями и опытом, в презентации результатов работы команды
УК-6.	Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни
УК-6-1.	Планирует и решает перспективные задачи собственной деятельности с учетом условий, средств, личностных возможностей, этапов карьерного роста, временной перспективы развития деятельности и требований рынка труда
УК-6-2.	Владеет навыками управления своим временем
УК-6-3.	Проявляет интерес к образованию и использует предоставляемые возможности для приобретения новых знаний и навыков

3. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Название раздела/темы	Всего часов	Трудоемкость (час.) по видам учебных занятий			Самостоятельная работа
		Контактная работа			
		Всего	Лекции	семинары	

Тема 1. Процесс анализа данных	16	12	6	6	4
Тема 2. Описательная статистика	18	12	6	6	6
Тема 3. Методы визуализации данных	18	12	6	6	6
Тема 4. Элементы машинного обучения	20	12	6	6	8
Итого	72	48	24	24	24

Тема 1. Процесс анализа данных

Методология CRISP-DM. Этапы анализа данных. Место визуализации в процессе анализа данных.

Тема 2. Описательная статистика

Распределения данных. Визуализация данных разных шкал. Статистические методы. Построение гипотез.

Тема 3. Методы визуализации данных

Основные виды графиков и диаграмм. Частые ошибки визуализации данных. Правила визуализации данных. Дата-сторителлинг.

Тема 4. Элементы машинного обучения

Применение визуализации для отбора признаков. Модели машинного обучения с учителем и без учителя. Применение визуализации для оценки модели.

4. ОЦЕНОЧНЫЕ СРЕДСТВА И ПРИМЕРЫ ЗАДАНИЙ ДЛЯ ОЦЕНКИ РЕЗУЛЬТАТОВ ОСВОЕНИЯ ДИСЦИПЛИНЫ

4.1 Текущий контроль

Оценка за курс складывается из следующих видов заданий текущего контроля:

1. Практическая работа «Подбор и оценка данных для визуализации, манипуляции с табличными данными» (максимум 16 баллов).
2. Практическая работа «Применения методов визуализации данных» (максимум 16 баллов).
3. Практическая работа «Построение гипотез в визуальном анализе данных» (максимум 16 баллов).
4. Практическая работа «Визуализация текстовых данных» (максимум 16 баллов).
5. Практическая работа «Методы визуализации данных в дата-сторителлинге» (максимум 16 баллов).

Практическая работа

Курс содержит 5 практических работ, раскрывающих содержание разделов дисциплины. Каждая практическая работа состоит из 2-4 заданий, выполняемых в формате interactive python notebook (.ipynb). На выполнение каждой практической работы отводится две недели. Работы должны быть сданы в период прочтения курса. Сдача работы представляет собой публикацию ссылки на свою работу в LMS. В случае необходимости студент дает устные ответы на уточняющие вопросы по отдельным задачам или делает дополнительные задачи по теме практической работы.

Требования к выполнению практических работ:

В случае если оформление кода решения и доклад обучающегося во время защиты соответствуют указанным требованиям, обучающийся получает максимальное количество баллов.

Основаниями для снижения количества баллов являются:

- несоответствие программного кода PEP8;
- небрежное выполнение работы;
- неполные или неточные ответы на вопросы.

Отчет не может быть принят и подлежит доработке в случае:

- отсутствия решения для обязательных частей практической работы.

Критерий	Минимальное количество баллов	Максимальное количество баллов
Выполнены все задания, обучающийся четко и без ошибок ответил на все контрольные вопросы	12	16
Выполнены все задания; обучающийся ответил на все контрольные вопросы с замечаниями	9	11
Выполнены все задания; обучающийся ответил на все контрольные вопросы с замечаниями	5	8
Обучающийся не выполнил или выполнил неправильно задания	0	4

4.2 Промежуточная аттестация

Промежуточная аттестация проводится в формате защиты группового проекта.

Групповой проект выполняется обучающимися в группах по 2-3 человека, тема выбирается обучающимися самостоятельно. Отчет по проекту должен содержать формулировку задачи и подробное описание путей ее реализации с соответствующим графическим материалом. Защита проекта проводится в устной форме в формате доклада с презентацией.

Структура отчета по проекту:

- 1) Тема проекта.
- 2) Цель и задачи проекта.
- 3) Описание входных данных.
- 4) Описание этапов выполнения проекта.
- 5) Результаты.
- 6) Выводы.

Требования к выполнению группового проекта:

Требования	Минимальное количество баллов	Максимальное количество баллов
I. Качество проекта		
1. Выбор данных произведен корректно и обоснован	0	2
2. Порядок анализа данных соответствует методологии CRISPDM	0	2
3. Наличие и корректное применение необходимых графических материалов	0	4
4. Построение и оценка модели при анализе данных	0	4
5. Обоснованность выводов	0	3
Общая оценка за выполнение проекта	0	15
II. Качество доклада		
1. Соответствие содержания доклада содержанию проекта	0	3
2. Качество изложения материала	0	2

Общая оценка за доклад	0	5
Итоговый балл	0	20

4.3 Примеры заданий

Пример практической работы

В этой практической работе предстоит провести визуальный анализ данных об особенностях жанров музыки. Датасет содержит 41700 уникальных композиций.

Данные содержат 18 столбцов:

- instance_id - id строки датасета
- artist_name - исполнитель
- track_name - название трека
- popularity - индекс популярности трека
- acousticness - акустичность
- danceability - танцевальность (насколько трек подходит для танцев)
- duration_ms - длительность трека
- energy - энергичность трека
- instrumentalness - инструментальность
- key - регистр (высота тона)
- liveness - живость
- loudness - громкость
- mode - тональность
- speechiness - насколько много текста
- tempo - музыкальный темп
- obtained_date - дата загрузки трека в датасет
- valence - настроение
- music_genre - жанр

Для выполнения задания возьмите 30 тысяч случайных строк из датасета.

1. Покажите, какие характеристики музыки коррелируют между собой. Сделайте выводы о корреляциях.
2. Покажите на диаграмме разброса три пары признаков, имеющих наибольшую отрицательную корреляцию.
3. Покажите на горизонтальной столбиковой диаграмме распределение количества треков по жанрам. Выделите контрастным цветом жанр, в котором больше всего песен.
4. Для топ-3 жанров покажите на круговой диаграмме топ-3 самых популярных исполнителей. Диаграмма должна быть одна.
5. На вертикальной столбиковой диаграмме для самого популярного и самого непопулярного жанров покажите средние значения по характеристикам. На одной диаграмме покажите 5 характеристик, в которых разница средних значений максимальна. На второй – в которых разница средних значений минимальна.
6. Добавьте в данные категориальный столбец «длительность трека», в котором будут значения: «короткая», «средняя», «длинная». Значения для каждой категории определите, исследовав данные в столбце duration_ms. Покажите график плотности распределения данных в столбце loudness, сгруппированных по категориям длительности.
7. Покажите на ящичковой диаграмме отношение популярности треков к жанру. Какие жанры похожи между собой?
8. На круговой диаграмме вида «пончик» покажите численное соотношение треков разной длительности (для длительности используйте новый категориальный столбец, сформированный в задании 6).

Все задания выполняются только при помощи средств визуализации Python: matplotlib, seaborn, pandas, plotly. По каждому заданию должен быть сделан вывод. Для всех визуализаций используйте стиль seaborn. Обязательно добавляйте подпись графика и осей.

Примерная тематика групповых проектов

Обучающиеся самостоятельно выбирают набор данных для анализа, обосновывают свой выбор, производят анализ данных по методологии CRISP-DM, сопровождающийся визуализацией промежуточных и итоговых результатов.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Литература

1. Анализ данных : учебник для вузов / под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2025. — 448 с. — (Высшее образование). — ISBN 978-5-534-19964-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/560311> (дата обращения: 21.07.2025).

2. Чернышев, С. А. Основы программирования на Python : учебник для вузов / С. А. Чернышев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 349 с. — (Высшее образование). — ISBN 978-5-534-17139-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/567821> (дата обращения: 21.07.2025).

5.2 Электронные образовательные ресурсы

Материалы дисциплины размещены в LMS: <https://l.skolkovo.ru/login/index.php>

5.3 Профессиональные базы данных и информационные справочные системы (при наличии)

Нет

6. ЛИЦЕНЗИОННОЕ И СВОБОДНО РАСПРОСТРАНЯЕМОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Операционная система Simple Linux, браузер Yandex браузер, антивирусное ПО Calmantivirus;

Свободно распространяемое ПО, в том числе отечественного производства:

Офисный пакет Libre Office, Okular PDF Reader, 7-Zip Архиватор, GIMP Редактирования фотографий, Inkscape Векторная графика, Blender 3D графика, Kdenlive Видеоредактор, Audacity Аудиоредактор, VLC Медиаплеер, Thunderbird Почтовый клиент, Flameshot Создание скриншотов

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебная аудитория для проведения занятий лекционного типа, оснащенная мультимедийным оборудованием, учебной мебелью, доской или со стенами с маркерным покрытием.

Учебная аудитория для проведения занятий семинарского типа, оснащенная мультимедийным оборудованием, учебной мебелью, доской или со стенами с маркерным покрытием.

Аудитория (коворкинг) для самостоятельной работы, оснащенная учебной мебелью, ноутбуками.

Материально-техническое обеспечение аудиторий представлено на официальном сайте <https://bbask.ru/sveden/objects/>.