

Федеральное государственное образовательное бюджетное учреждение
высшего образования
**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»**
(Финансовый университет)

Кафедра бизнес-информатики
Факультет информационных технологий и анализа больших данных

Документ подписан усиленной неквалифицированной электронной подписью
Организация: Финансовый университет при Правительстве РФ
Утверждено: Проректор по учебной и методической работе Е.А. Каменева
Сертификат: 36yXdkaKPNRjCmTS3nckrDtvsE9H/ic2
Дата: 08.10.2025 г.

А.Ю. Мишин

Методы визуализации данных

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки:

38.03.01 - Экономика,

Образовательная программа «Бизнес-архитектура и аналитика»

Рекомендовано

*Факультет информационных технологий и анализа больших данных
(протокол № 03 от 16.12.2025 г.)*

Одобрено

*Кафедра бизнес-информатики
(протокол № 4 от 20.11.2025 г.)*

© Москва 2026

СОДЕРЖАНИЕ

1.	Наименование дисциплины	3
2.	Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	3
3.	Место дисциплины в структуре образовательной программы	7
4.	Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	7
5.	Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	8
5.1.	Содержание дисциплины	8
5.2.	Учебно-тематический план	11
5.3.	Содержание семинаров	12
6.	Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине	14
6.1.	Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	14
6.2.	Перечень вопросов, заданий, тем для подготовки к текущему контролю	16
7.	Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	20
8.	Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	41
9.	Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	41
10.	Методические указания для обучающихся по освоению дисциплины	43
11.	Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	43
12.	Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	44

1. Наименование дисциплины

«Методы визуализации данных».

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКН-3	Способность осуществлять сбор, обработку и статистический анализ данных, применять математические методы для решения стандартных профессиональных финансово-экономических задач, интерпретировать полученные результаты	1. Проводит сбор, обработку и статистический анализ данных для решения финансово-экономических задач	знать: 1. Способы сбора данных, классификацию данных и методы обеспечения их качества в целях исследовательского анализа данных;2. Высокоуровневый алгоритм анализа данных и этапы Knowledge Discovery in Databases (KDD) для системного подхода к решению задач;3. Принципы и инструменты исследовательского анализа данных (EDA), включая статистические методы для первичного анализа. уметь: 1. Осуществлять первичную обработку данных: проверять их качество, идентифицировать и устранять типовые проблемы качества с использованием инструментов очистки данных;2. Строить и интерпретировать EDA-диаграммы (гистограммы, блочные диаграммы и т.д.) с помощью библиотек визуализации;3. Подготавливать данные для анализа, используя возможности BI-инструментов.
		2. Формулирует математические постановки финансово-экономических задач, переходит от экономических постановок задач к математическим моделям	знать: 1. Типовые классы финансово-экономических задач, решаемых методами Data Mining и Process Mining;2. Знать основные визуальные представления;3. Взаимосвязь бизнес-сущностей и данных - показателей и атрибутов. уметь: 1. Декомпозировать аналитическую задачу в набор

			измеримых показателей, привязанных к модели данных;2. Определять состав данных, необходимых для построения математической модели;3. Выбирать адекватный исследовательской задаче тип визуализации.
		3. Системно подходит к выбору математических методов и информационных технологий для решения конкретных финансово-экономических задач в профессиональной области	знать: 1. Континуум BI-решений и функционал популярных BI-систем;2. Правила бизнес-визуализации;3. Концептуальную архитектуру BI-решений и их взаимосвязь с задачами анализа. уметь: 1. Осуществлять рациональный выбор между использованием BI-платформы и python-библиотеками для решения аналитической задачи;2. Выбирать конкретный тип BI-чарта в зависимости от цели анализа и природы данных;3. Применять различные технологии доступа и обработки данных (SQL-запросы, подключения к OLAP-кубам, работа с плоскими файлами) в BI-платформе
		4. Анализирует результаты исследования математических моделей финансово-экономических задач и делает на их основании количественные и качественные выводы и рекомендации по принятию финансово-экономических решений	знать: 1. Правила и принципы эффективной визуализации данных, позволяющие избежать неверной интерпретации и донести суть результатов до стейкхолдеров; 2. Принципы интерактивного сторителлинга и построения дашбордов, направленные на выделение ключевых метрик и упрощение принятия решений;3. Типовые бизнес-эффекты от визуализации данных. уметь: 1. Читать и интерпретировать сложные визуализации, извлекая из них количественные оценки и качественные закономерности;2. Формулировать содержательные выводы и практические рекомендации для принятия финансово-экономических решений на основе анализа визуализированных данных;3.

			Создавать комплексные отчеты и дашборды в BI-платформах
УК-4	Способность использовать прикладное программное обеспечение при решении профессиональных задач	Использует основные методы и средства получения, представления, хранения и обработки данных.	знать: 1. Архитектурные компоненты современных BI-решений;2. Принципы работы, преимущества и ограничения различных типов СУБД и файловых систем для задач хранения данных;3. Стандарты и протоколы безопасного доступа к данным в корпоративной среде. уметь: 1. Настраивать подключения к различным источникам данных в BI-платформе, корректно выбирая параметры;2. Проектировать и реализовывать процесс загрузки и трансформации данных (ETL) в рамках BI-инструмента для построения аналитических датасетов;3. Управлять инфраструктурой данных в облачной платформе: создавать иерархии папок/воркбуков, распределять права доступа к данным и отчетам.
		Демонстрирует владение профессиональными пакетами прикладных программ.	знать: 1. Полный жизненный цикл объекта в BI-платформе от создания подключения и датасета до версионирования и публикации дашборда;2. Архитектуру и модульную структуру библиотек визуализации, принципы построения графиков и их настройки;3. Встроенные высокоуровневые интерфейсы библиотеки Seaborn для быстрого построения сложных статистических графиков. уметь: 1. Реализовывать в BI-платформе сложные интерактивные элементы - иерархии, QL-чарты, связанные дашборды;2. Программировать сложные, кастомизированные визуализации в Python, комбинируя методы Matplotlib и интерфейсы Seaborn;3. Использовать встроенные языки разметки и запросов (Markdown, QL) в BI-платформах для оформления отчетов.

		Выбирает необходимое прикладное программное обеспечение в зависимости от решаемой задачи.	знать: 1. Рыночный ландшафт BI-решений и ключевые критерии выбора BI-решений;2. Технические ограничения различных BI-инструментов;3. Конкретные ситуации и сферы эффективного применения различных BI-технологий. уметь: 1. Проводить сравнительный анализ функционала BI-решений для целей конкретного проекта;2. Определять состав необходимого сопутствующего программного обеспечения для построения целостного аналитического решения;3. Выбирать BI-инструмент исходя из системных требований.
		Использует прикладное программное обеспечение для решения конкретных прикладных задач.	знать: 1. Концептуальную архитектуру типового BI-решения;2. Встроенные механизмы оптимизации в BI-решениях;3. Принципы проектирования пользовательского интерфейса дашбордов (UI/UX) и интерактивного сторителлинга. уметь: 1. Разворачивать и настраивать рабочее BI-окружение в облачной платформе (Yandex.Cloud);2. Разрабатывать и внедрять единые стили оформления визуализаций в Matplotlib/Seaborn;3. Оценивать производительность созданного BI-функционала и применять инструментальные методы для его оптимизации.

3. Место дисциплины в структуре образовательной программы

Дисциплина «Методы визуализации данных» относится к «Цикл математики и информатики».

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Очная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 5 (в часах)
Общая трудоёмкость дисциплины	4/144	144
Контактная работа- Аудиторные занятия	50	50
Лекции	16	16
Семинары, практические занятия	34	34
Самостоятельная работа	94	94
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Зачет	Зачет

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Области применения визуализации данных

Понятие визуализации данных. Бизнес-эффекты от визуализации данных. Достоинства и недостатки визуализации данных. Понятие Process Mining. Задачи, решаемые в рамках Process Mining. Виды данных по степени структурированности. Способы сбора данных. Высокоуровневый алгоритм анализа данных. Высокоуровневый алгоритм моделирования. Роли в аналитической команде. Понятие Data Mining. Основные задачи Data Mining. Исследовательский анализ данных (EDA, Exploratory data analysis). Способы, методы и инструменты EDA.

Тема 2. Описание инструментов, технологий и инфраструктуры бизнес-визуализации

Определения Business Intelligence. Этапы Knowledge Discovery in Databases (KDD). Континуум BI-решений. Рынок BI-решений. Сравнение функционала популярных BI-систем. Понятие adhoc-анализа. Понятие Extract, Transform, Load (ETL). Понятие качества данных. Дерево проблем качества данных. Методы очистки данных. Инструменты и технологии очистки данных. Понятие Online Transaction Processing (OLTP). Понятие Online Analytical Processing (OLAP). Java DataBase Connectivity (JDBC). Open Database Connectivity (ODBC). Понятие системы хранения данных (Data Warehouse, DWH). Подходы к разработке DWH. Hadoop Distributed File System (HDFS). Плоские файлы. Понятие обогащения данных. Виды обогащения данных.

Тема 3. Формы и правила визуализации данных

Способы представления и визуализации данных. Виды диаграмм и гистограмм. Виды карт и картограмм. Кросс-таблицы и кросс-диаграммы. Интерактивный сторителлинг. Отчёты. Приборные панели (dashboards, дашборды). Правила визуализации данных.

Тема 4. Визуализация данных в Yandex DataLens

Архитектура данных в Yandex.Cloud. Концептуальная архитектура BI-решения на базе Yandex DataLens. Основные сущности Yandex DataLens: источник, подключение, датасет, чарт, дашборд. Виды источников данных в Yandex DataLens. Варианты развертывания. Общий алгоритм работы Yandex DataLens. Материализованные

представления. Типы данных в Yandex DataLens. Виды чартов в Yandex DataLens. Древовидные иерархии в Yandex DataLens. Drill-down через древовидные структуры. Язык разметки Markdown. СУБД ClickHouse. Механизм работы колоночной СУБД. Процедуры аутентификации, идентификации и авторизации. Федеративное управление идентификацией (FIM). Технология однократного входа (Single sign-on, SSO). Технология единого входа (Same Sign-On). Стандарты безопасности Yandex Cloud. Yandex Cloud Organization. Папки, воркбуки и каталоги в Yandex DataLens. Критерии распределения прав. Обогащение датасета. Yandex Cloud Marketplace. QL-чарты. Ограничения на чарты на основе связанных датасетов в DataLens. Версионирование чартов в Yandex DataLens. Виды версий чартов в DataLens. Создание чарта с иерархией. Общие настройки чарта. Настройки измерений. Настройки показателей. Особенности настройка различных видов чартов. Особенности QL-чартов. Настройка QL-чарта.

Тема 5. Библиотека визуализации Matplotlib

Двумерные диаграммы в Matplotlib: линейная (plot), столбчатая (bar), рассеяния (scatter), стеблевая (stem), заливки областей (fill_between), с накоплением областей (stackplot), ступенчатая (stairs) - синтаксис, аргументы, ограничения, примеры использования.

Статистические диаграммы в Matplotlib: гистограмма (hist), блочная ("ящик с усами", boxplot), с погрешностями (errorbar), скрипичная (violinplot), событий (eventplot), двумерная тепловая карта (hist2d), гексагональный график с бинами (hexbin), круговая (pie), эмпирическая кумулятивная функция распределения (ecdf) - синтаксис, аргументы, ограничения, примеры использования.

Диаграммы для данных с координатной привязкой в Matplotlib: двумерная регулярная цветовая сетка (imshow), псевдоцветовая нерегулярная прямоугольная сетка (pcolormesh), контурные линии (contour), контурные области (contourf), ветровых стрелок (barbs), поля градиентов (quiver), поточная (streamplot) - синтаксис, аргументы, ограничения, примеры использования.

Диаграммы для нерегулярно распределенных данных в Matplotlib: контурные линии на неструктурированной треугольной сетке (tricontour), контурные области на неструктурированной треугольной сетке (tricontourf), псевдоцветовая неструктурированной треугольной сетки (tripcolor), неструктурированной треугольной сетки (tripplot) - синтаксис, аргументы, ограничения, примеры использования.

Диаграммы для 3D и объемных данных в Matplotlib: трехмерная столбчатая (bar3d), заливки трехмерных областей (fill_between), трёхмерная линейная (plot), трёхмерная поля градиентов (quiver), трёхмерная рассеяния (scatter), трехмерная стеблевая (stem), трехмерной поверхности (plot_surface), трехмерной поверхности на треугольной сетке (plot_trisurf), вокселей (voxels), трёхмерная каркасная (plot_wireframe).

Matplotlib API. Модули Matplotlib.

Тема 6. Библиотека визуализации Seaborn

Настройка осей в Seaborn. Настройка палитр в Seaborn. Маркеры в Seaborn. Implot. scatterplot. lineplot. displot. relplot. catplot. boxplot. histplot. boxenplot. kdeplot. displot. residplot. Seaborn API. pairgrid. heatmap. stripplot. JointGrid. FacetGrid. jointplot. violinplot. relplot. swarmplot. clustermap.

5.2. Учебно-тематический план

№ п/п	Наименование тем(разделов) дисциплины	Трудоемкость в часах				
		Всего	Контактная работа - Аудиторная работа			Самостоя тельная работа
			Общая, в т.ч.:	Лекции	Семинары, практическ ие занятия	
1	Области применения визуализации данных	12	4	2	2	8
2	Описание инструментов, технологий и инфраструктуры бизнес- визуализации	14	6	2	4	8
3	Формы и правила визуализации данных	16	4	2	2	12
4	Визуализация данных в Yandex DataLens	40	16	4	12	24
5	Библиотека визуализации Matplotlib	36	12	4	8	24
6	Библиотека визуализации Seaborn	26	8	2	6	18
	Итого	144	50	16	34	94

5.3. Содержание семинаров

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарах, практических занятиях	Формы проведения занятий
Области применения визуализации данных	Понятие визуализации данных. Бизнес-эффекты от визуализации данных. Достоинства и недостатки визуализации данных. . Способы сбора данных. Исследовательский анализ данных (EDA, Exploratory data analysis). Способы, методы и инструменты EDA.	Дискуссия, обсуждение.Выполнение индивидуальных заданий.
Описание инструментов, технологий и инфраструктуры бизнес-визуализации	Определения Business Intelligence. Этапы Knowledge Discovery in Databases (KDD). Континуум BI-решений. Рынок BI-решений. Сравнение функционала популярных BI-систем. Понятие ад-хок-анализа. Понятие Extract, Transform, Load (ETL). Понятие качества данных. Дерево проблем качества данных. Методы очистки данных. Инструменты и технологии очистки данных. Понятие обогащения данных. Виды обогащения данных.	Дискуссия, обсуждение.Выполнение индивидуальных заданий.
Формы и правила визуализации данных	Способы представления и визуализации данных. Виды диаграмм и гистограмм. Виды карт и картограмм. Кросс-таблицы и кросс-диаграммы. Интерактивный сторителлинг.	Дискуссия, обсуждение.Выполнение индивидуальных заданий.
Визуализация данных в Yandex DataLens	Архитектура данных в Yandex. Cloud. Концептуальная архитектура BI-решения на базе Yandex DataLens. Основные сущности Yandex DataLens: источник, подключение, датасет, чарт, дашборд. Виды источников данных в Yandex DataLens. Варианты развертывания. Общий алгоритм работы Yandex DataLens. Материализованные представления. Типы данных в Yandex DataLens. Виды чартов в Yandex DataLens. Древовидные иерархии в Yandex DataLens. Drill-down через древовидные структуры. Механизм работы колоночной СУБД. Папки, воркбуки и каталоги в Yandex DataLens. Критерии распределения прав. Обогащение датасета. Yandex Cloud Marketplace. QL-чарты. Ограничения на чарты на основе связанных датасетов в DataLens. Версионирование чартов в Yandex DataLens. Создание чарта с иерархией. Общие настройки чарта. Настройки измерений. Настройки показателей. Особенности настройка различных видов чартов. Настройка QL-чарта.	Дискуссия, обсуждение.Выполнение индивидуальных заданий.

Библиотека визуализации Matplotlib	Двумерные диаграммы в Matplotlib: линейная (plot), столбчатая (bar), рассеяния (scatter), стеблевая (stem), заливки областей (fill_between), с накоплением областей (stackplot), ступенчатая (stairs) - синтаксис, аргументы, ограничения, примеры использования. Статистические диаграммы в Matplotlib: гистограмма (hist), блочная ("ящик с усами", boxplot), с погрешностями (errorbar), скрипичная (violinplot), событий (eventplot), двумерная тепловая карта (hist2d), гексагональный график с бинами (hexbin), круговая (pie), эмпирическая кумулятивная функция распределения (ecdf) - синтаксис, аргументы, ограничения, примеры использования.	Дискуссия, обсуждение.Выполнение индивидуальных заданий.
Библиотека визуализации Seaborn	Настройка осей в Seaborn. Настройка палитр в Seaborn. Маркеры в Seaborn. lmplo. scatterplot. lineplot. displot. relplot. catplot. boxplot. histplot. boxenplot. kdeplot. displot. residplot	Дискуссия, обсуждение.Выполнение индивидуальных заданий.

6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Области применения визуализации данных	Виды данных по степени структурированности. Высокоуровневый алгоритм анализа данных. Высокоуровневый алгоритм моделирования. Роли в аналитической команде. Понятие Data Mining. Основные задачи Data Mining. Понятие Process Mining. Задачи, решаемые в рамках Process Mining	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников
Описание инструментов, технологий и инфраструктуры бизнес-визуализации	Понятие Online Transaction Processing (OLTP). Понятие Online Analytical Processing (OLAP). Java DataBase Connectivity (JDBC). Open Database Connectivity (ODBC). Понятие системы хранения данных (Data Warehouse, DWH). Подходы к разработке DWH. Hadoop Distributed File System (HDFS). Плоские файлы.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников
Формы и правила визуализации данных	Отчёты. Приборные панели (dashboards, дашборды). Правила визуализации данных.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников
Визуализация данных в Yandex DataLens	Язык разметки Markdown. СУБД ClickHouse. Процедуры аутентификации, идентификации и авторизации. Федеративное управление идентификацией (FIM). Технология однократного входа (Single sign-on, SSO). Технология единого входа (Same Sign-On). Стандарты безопасности Yandex Cloud. Yandex Cloud Organization. Виды версий чартов в DataLens. Особенности QL-чартов.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников

Библиотека визуализации Matplotlib	Диаграммы для данных с координатной привязкой в Matplotlib: двумерная регулярная цветовая сетка (imshow), псевдоцветовая нерегулярная прямоугольная сетка (pcolormesh), контурные линии (contour), контурные области (contourf), ветровых стрелок (barbs), поля градиентов (quiver), поточная (streamplot) - синтаксис, аргументы, ограничения, примеры использования. Диаграммы для нерегулярно распределенных данных в Matplotlib: контурные линии на неструктурированной треугольной сетке (tricontour), контурные области на неструктурированной треугольной сетке (tricontourf), псевдоцветовая неструктурированной треугольной сетки (tripcolor), неструктурированной треугольной сетки (tripplot) - синтаксис, аргументы, ограничения, примеры использования. Диаграммы для 3D и объемных данных в Matplotlib: трехмерная столбчатая (bar3d), заливки трехмерных областей (fill_between), трёхмерная линейная (plot), трёхмерная поля градиентов (quiver), трёхмерная рассеяния (scatter), трехмерная стеблевая (stem), трехмерной поверхности (plot_surface), трехмерной поверхности на треугольной сетке (plot_trisurf), вокселей (voxels), трёхмерная каркасная (plot_wireframe). Matplotlib API. Модули Matplotlib.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников
Библиотека визуализации Seaborn	Seaborn API. pairgrid. heatmap. stripplot. JointGrid. FacetGrid. jointplot. violinplot. relplot. swarmplot. clustermap.	Изучение методических материалов по теме в электронном виде и рекомендуемых разделов основной и дополнительной литературы, интернет – источников

6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерные темы контрольной работы:

1. Разработка бизнес-визуализации для сетевого ретейлера: анализ продаж с целью исследовательского анализа данных (EDA) для выявления региональных и сезонных паттернов.
2. Разработка бизнес-визуализации для коммерческого банка: процесс кредитного скоринга с целью проверки качества данных и выявления аномалий.
3. Разработка бизнес-визуализации для логистической компании: процесс управления цепочками поставок с целью Process Mining для оптимизации маршрутов.
4. Разработка бизнес-визуализации для телеком-оператора: функция удержания клиентов (Churn Rate) с целью бизнес-анализа ключевых причин оттока.
5. Разработка бизнес-визуализации для медицинского учреждения: процесс обработки пациентов с целью исследовательского анализа данных для оценки загруженности отделений.
6. Разработка бизнес-визуализации для производственного предприятия: функция контроля качества с целью мониторинга дефектов в режиме реального времени.
7. Разработка бизнес-визуализации для финтех-стартапа: процесс транзакций с целью анализа поведения пользователей (User Behavior Analytics).
8. Разработка бизнес-визуализации для авиакомпании: процесс бронирования и регистрации с целью Process Mining для выявления узких мест.
9. Разработка бизнес-визуализации для агрохолдинга: функция прогнозирования урожайности с целью исследовательского анализа данных на основе погодных и почвенных данных.
10. Разработка бизнес-визуализации для IT-компании: процесс разработки ПО с целью анализа производительности команд (Team Performance Analysis).
11. Разработка бизнес-визуализации для страховой компании: процесс урегулирования убытков с целью бизнес-анализа временных циклов и стоимости операций.

12. Разработка бизнес-визуализации для энергетической компании: функция мониторинга потребления с целью проверки качества данных с интеллектуальных счетчиков.
13. Разработка бизнес-визуализации для ритейлера электроники: анализ эффективности маркетинговых кампаний с целью оценки возврата на инвестиции (ROI).
14. Разработка бизнес-визуализации для консалтинговой фирмы: управление проектами с целью визуализации использования ресурсов и сроков.
15. Разработка бизнес-визуализации для рекламного агентства: процесс проведения медиа-кампаний с целью анализа ключевых показателей эффективности (CPC, CTR, ROI).
16. Разработка бизнес-визуализации для гостиничной сети: функция динамического ценообразования с целью исследовательского анализа данных о спросе.
17. Разработка бизнес-визуализации для фармацевтической компании: процесс клинических испытаний с целью мониторинга ключевых показателей эффективности (KPI).
18. Разработка бизнес-визуализации для онлайн-кинотеатра: анализ поведения зрителей с целью построения рекомендательной системы.
19. Разработка бизнес-визуализации для строительной корпорации: управление портфелем проектов с целью анализа рисков и бюджетов.
20. Разработка бизнес-визуализации для интернет-магазина: процесс выполнения заказов с целью Process Mining для сокращения времени доставки.
21. Разработка бизнес-визуализации для университета: функция приема абитуриентов с целью бизнес-анализа эффективности рекрутинговых каналов.
22. Разработка бизнес-визуализации для транспортно-логистической компании: управление автопарком с целью мониторинга расхода топлива и пробега.
23. Разработка бизнес-визуализации для инвестиционного фонда: анализ портфеля ценных бумаг с целью исследовательского анализа данных для выявления трендов.

24. Разработка бизнес-визуализации для call-центра: процесс обработки обращений с целью анализа производительности операторов.
25. Разработка бизнес-визуализации для нефтегазовой компании: процесс добычи сырья с целью проверки качества данных с датчиков оборудования.
26. Разработка бизнес-визуализации для SaaS-компании: анализ метрик клиентского успеха (Customer Success) с целью снижения оттока.
27. Разработка бизнес-визуализации для государственного учреждения: процесс оказания государственных услуг с целью бизнес-анализа прозрачности и сроков.
28. Разработка бизнес-визуализации для автомобильного дилера: функция управления запасами запчастей с целью прогнозного анализа спроса.
29. Разработка бизнес-визуализации для игровой студии: анализ игровых событий (Game Analytics) с целью исследовательского анализа данных о поведении игроков.
30. Разработка бизнес-визуализации для HR-департамента крупной корпорации: функция подбора персонала с целью Process Mining для оптимизации воронки рекрутинга.

Предусмотрено выполнение контрольной работы, заключающейся в разработке комплексного BI-решения, сочетающего как безкодovou часть (дашборд в BI-платформе, так и низкоуровневую часть, разработанную студентов на python с использованием библиотек matplotlib/seaborn)

Содержание контрольной работы:

1. Формирование студентом и согласование учебного кейса с преподавателем;
2. Разработка логической модели данных, подлежащих визуализации;
3. Генерация студентом синтетических данных для наполнения источников данных;

4. Разработка чартов и дашборда на BI-платформе, визуализирующих сгенерированные данные
5. Разработка низкоуровневых визуализаций с использованием библиотек matplotlib/seaborn
6. Интеграция низкоуровневой и безкодвой части BI-решения;
7. Настройка ролевой модели BI-решения
8. Разработка пояснительной записки

Дополнительная информация

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях Кафедры бизнес-информатики.

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. *Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине.*

Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

ПКН-3 Способность осуществлять сбор, обработку и статистический анализ данных, применять математические методы для решения стандартных профессиональных финансово-экономических задач, интерпретировать полученные результаты

1) 1. Проводит сбор, обработку и статистический анализ данных для решения финансово-экономических задач

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: 1. Способы сбора данных, классификацию данных и методы обеспечения их качества в целях исследовательского анализа данных; 2. Высокоуровневый алгоритм анализа данных и этапы Knowledge Discovery in Databases (KDD) для системного подхода к решению задач; 3. Принципы и инструменты исследовательского анализа данных (EDA), включая статистические методы для первичного анализа.

Уметь: 1. Осуществлять первичную обработку данных: проверять их качество, идентифицировать и устранять типовые проблемы качества с использованием инструментов очистки данных; 2. Строить и интерпретировать EDA-диаграммы (гистограммы, блочные диаграммы и т.д.) с помощью библиотек визуализации; 3. Подготавливать данные для анализа, используя возможности BI-инструментов.

Типовые контрольные задания

Задание 1

Предоставлен датасет с ежемесячными продажами по филиалам компании за последние 3 года. Данные содержат пропуски, дубликаты и аномальные значения (например, отрицательные суммы продаж). Также в данных присутствуют следующие поля: "Дата", "Филиал", "Менеджер", "Категория товара", "Сумма продажи", "Количество".

Выполните следующее:

1. Проведите аудит и очистку данных: устранили дубликаты, заполните пропуски и обработайте аномалии, обосновав выбранные методы;
2. Проведите исследовательский анализ (EDA), используя библиотеки Matplotlib/Seaborn, построив и интерпретировав: гистограмму распределения суммы продаж, блочную диаграмму для выявления выбросов по сумме продаж по филиалам, график динамики общего объема продаж по месяцам.

Задание 2

В Yandex DataLens загружены два CSV-файла: sales.csv (информация о продажах) и products.csv (справочник товаров с категориями и закупочной ценой). В данных есть несоответствия в названиях товаров, а для расчета маржинальности необходимо объединить таблицы.

Выполните следующее:

1. Создайте подключения и датасеты для обоих файлов в Yandex DataLens;
2. Создайте связь между датасетами по полю "ID_товара".
3. В датасете продаж создайте вычисляемое поле "Маржа", которое рассчитывается как "[Сумма продажи] - [Закупочная цена]".
4. Проверьте данные на наличие некорректных значений после объединения (например, продажи товаров, которых нет в справочнике).

2) 2. Формулирует математические постановки финансово-экономических задач, переходит от экономических постановок задач к математическим моделям

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: 1. Типовые классы финансово-экономических задач, решаемых методами Data Mining и Process Mining; 2. Знать основные визуальные представления; 3. Взаимосвязь бизнес-сущностей и данных - показателей и атрибутов.

Уметь: 1. Декомпозировать аналитическую задачу в набор измеримых показателей, привязанных к модели данных; 2. Определять состав данных, необходимых для построения математической модели; 3. Выбирать адекватный исследовательской задаче тип визуализации.

Типовые контрольные задания

Задание 1

Руководство банка поставило бизнес-задачу: «Снизить отток клиентов из премиум-сегмента».

Дана логическая модель данных банка в виде графической БД-схемы.

Выполните следующее:

1. Декомпозируйте бизнес-задачу в набор конкретных, измеримых показателей, которые можно оценить с помощью анализа данных;
2. Для каждого показателя определите, какие именно данные (показатели и атрибуты) необходимы для построения математической модели или анализа.
3. Составьте список требуемых полей из базы данных.

Задание 2

Дан набор данных. Выдвинута исследовательская гипотеза: «Существует зависимость между возрастом клиента и средним чеком в интернет-магазине электроники».

Выполните следующее:

1. Сформулируйте, какую математическую модель или статистический метод можно использовать для проверки этой гипотезы;
2. Выберите и обоснуйте два наиболее подходящих типа диаграмм для визуальной проверки этой гипотезы.

3) 3. Системно подходит к выбору математических методов и информационных технологий для решения конкретных финансово-экономических задач в профессиональной области

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: 1. Континуум BI-решений и функционал популярных BI-систем; 2. Правила бизнес-визуализации; 3. Концептуальную архитектуру BI-решений и их взаимосвязь с задачами анализа.

Уметь: 1. Осуществлять рациональный выбор между использованием BI-платформы и python-библиотеками для решения аналитической задачи; 2. Выбирать конкретный тип BI-чарта в зависимости от цели анализа и природы данных; 3. Применять различные технологии доступа и обработки данных (SQL-запросы, подключения к OLAP-кубам, работа с плоскими файлами) в BI-платформе

Типовые контрольные задания

Задание 1

Вам необходимо выполнить два проекта:

Проект А: Регулярно (еженедельно) готовить для отдела маркетинга дашборд с ключевыми метриками (LTV, САС, конверсия по каналам). Данные хранятся в корпоративной DWH.

Проект Б: Разработать и визуализировать нестандартную модель кластеризации клиентов на основе их транзакционного поведения для отдела аналитики.

Выполните следующее:

1. Для каждого проекта обоснуйте выбор инструментария (BI-платформа или python-библиотеки), опираясь на такие критерии, как регулярность, интерактивность, сложность анализа и потребности конечных пользователей.

Задание 2

В организации есть данные по продажам товаров. Данные хранятся в SQL-базе.

Необходимо проанализировать выполнение плана продаж по регионам и товарным категориям за текущий квартал. Данные хранятся в SQL-базе. Необходимо предоставить возможность детализации (drill-down) от региона к конкретному менеджеру.

Выполните следующее:

1. Настройте подключение к данным вы будете использовать;
2. Разработайте чарт и реализуйте в нём drill-down с помощью иерархий так, чтобы детализировать данные от региона к конкретному менеджеру;
3. Создайте ещё три чарта разных типов, наилучшим образом отображающие план и факт, структуру продаж и динамику по регионам и товарным категориям;
4. Создайте дашборд в Yandex DataLens, разместив на нем все созданные чарты и расшарьте его преподавателю.

4) 4. Анализирует результаты исследования математических моделей финансово-экономических задач и делает на их основании количественные и качественные выводы и рекомендации по принятию финансово-экономических решений

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: 1. Правила и принципы эффективной визуализации данных, позволяющие избежать неверной интерпретации и донести суть результатов до стейкхолдеров; 2. Принципы интерактивного сторителлинга и построения дашбордов, направленные на выделение ключевых метрик и упрощение принятия решений; 3. Типовые бизнес-эффекты от визуализации данных.

Уметь: 1. Читать и интерпретировать сложные визуализации, извлекая из них количественные оценки и качественные закономерности; 2. Формулировать содержательные выводы и практические рекомендации для принятия финансово-экономических решений на основе анализа визуализированных данных; 3. Создавать комплексные отчеты и дашборды в BI-платформах

Типовые контрольные задания

Задание 1

Представлен разработанный сторонним аналитиком дашборд. На нем представлены:

- 1) Heatmap (тепловая карта) корреляции между различными факторами (доход, возраст, количество кредитов) и вероятностью одобрения кредита;
- 2) Диаграмма рассеяния с двумя ключевыми факторами из heatmap.

Выполните следующее:

1. Дайте количественную и качественную интерпретацию результатов, отображенным на heatmap;
2. Сформулируйте две конкретные рекомендации для кредитного отдела на основе анализа этих визуализаций.

Задание 2

Перед запуском новой рекламной кампании необходимо проанализировать эффективность предыдущих кампаний по различным каналам (контекстная реклама, социальные сети, email-рассылки). Ключевые метрики: ROI, CTR, Cost per Lead.

Указанные данные есть в датасете.

Выполните следующее:

1. Создайте в Yandex DataLens интерактивный дашборд, позволяющий наглядно сравнивать эффективность каналов и показывать динамику ключевых метрик;
2. Настройте сторителлинг: выделите цветом лучший и худший канал, добавьте краткие текстовые выводы с помощью Markdown, обеспечьте возможность фильтрации по периоду.
3. Напишите 3-4 кратких вывода под дашбордом с рекомендациями по распределению бюджета на следующую кампанию.

УК-4 Способность использовать прикладное программное обеспечение при решении профессиональных задач

1) Использует основные методы и средства получения, представления, хранения и обработки данных.

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: 1. Архитектурные компоненты современных BI-решений; 2. Принципы работы, преимущества и ограничения различных типов СУБД и файловых систем для задач хранения данных; 3. Стандарты и протоколы безопасного доступа к данным в корпоративной среде.

Уметь: 1. Настраивать подключения к различным источникам данных в BI-платформе, корректно выбирая параметры; 2. Проектировать и реализовывать процесс загрузки и трансформации данных (ETL) в рамках BI-инструмента для построения аналитических датасетов; 3. Управлять инфраструктурой данных в облачной платформе: создавать иерархии папок/воркбуков, распределять права доступа к данным и отчетам.

Типовые контрольные задания

Задание 1

Необходимо построить дашборд в Yandex DataLens на основе данных из двух источников: операционная БД (PostgreSQL) и файл с данными о регионах (CSV в облачном хранилище). Данные в БД обновляются ежечасно.

Выполните следующее:

1. Создайте в Yandex DataLens два подключения: к PostgreSQL и к облачному хранилищу. Для подключения к БД укажите корректные параметры (хост, порт, база данных, тип подключения);
2. Объедините данные из двух источников в одном датасете, создав связь по полю region_id;
3. Настройте материализованное представление для датасета с суточной частотой обновления.

Задание 2

В Yandex Cloud Organization создана структура: каталог "Финансы" с папками "Бюджетирование" и "Отчетность". Необходимо настроить права доступа для двух групп пользователей: "Аналитики" (полный доступ в "Бюджетирование", только просмотр в "Отчетность") и "Менеджеры" (только просмотр готовых дашбордов в "Отчетность").

Выполните следующее:

1. В интерфейсе Yandex Cloud Organization создайте указанную структуру папок;
2. Назначьте соответствующие права (роли datalens.creator и datalens.viewer) на нужные папки для групп пользователей, следуя принципу наименьших привилегий.

2) Демонстрирует владение профессиональными пакетами прикладных программ.

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: 1. Полный жизненный цикл объекта в BI-платформе от создания подключения и датасета до версионирования и публикации дашборда; 2. Архитектуру и модульную структуру библиотек визуализации, принципы построения графиков и их настройки; 3. Встроенные высокоуровневые интерфейсы библиотеки Seaborn для быстрого построения сложных статистических графиков.

Уметь: 1. Реализовывать в BI-платформе сложные интерактивные элементы - иерархии, QL-чарты, связанные дашборды; 2. Программировать сложные, кастомизированные визуализации в Python, комбинируя методы Matplotlib и интерфейсы Seaborn; 3. Использовать встроенные языки разметки и запросов (Markdown, QL) в BI-платформах для оформления отчетов.

Типовые контрольные задания

Задание 1

На основе датасета с продажами (Регион -> Город -> Менеджер, Продажи) необходимо создать дашборд.

Выполните следующее:

1. Создайте в датасете Yandex DataLens древовидную иерархию "География": Регион -> Город -> Менеджер;
2. Постройте чарт (сводную таблицу), реализующий механизм drill-down по созданной иерархии;
3. Добавьте на дашборд селектор для фильтрации по периоду и свяжите его с чартом.

Задание 2

Имеется DataFrame df с полями date, product_category, sales, profit.

Выполните следующее:

1. Используя связку Matplotlib и Seaborn, постройте два графика на одной фигуре (subplots): линейный график динамики продаж по месяцам (верхний) и heatmap корреляции между продажами и прибылью по товарным категориям (нижний);
2. Настройте графическое оформление: для линейного графика задайте цвет и стиль линии, для heatmap — цветовую палитру Seaborn;
3. Добавьте заголовки и подписи осей.

3) Выбирает необходимое прикладное программное обеспечение в зависимости от решаемой задачи.

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: 1. Рыночный ландшафт BI-решений и ключевые критерии выбора BI-решений; 2. Технические ограничения различных BI-инструментов; 3. Конкретные ситуации и сферы эффективного применения различных BI-технологий.

Уметь: 1. Проводить сравнительный анализ функционала BI-решений для целей конкретного проекта; 2. Определять состав необходимого сопутствующего программного обеспечения для построения целостного аналитического решения; 3. Выбирать BI-инструмент исходя из системных требований.

Типовые контрольные задания

Задание 1

Стартап с ограниченным бюджетом и командой из 5 аналитиков и 20 менеджеров ищет BI-решение. Ключевые требования: низкая стоимость, возможность быстрого создания дашбордов без написания кода, интеграция с Google BigQuery, удобный механизм расшаривания отчетов.

Выполните следующее:

1. Проведите сравнительный анализ 3 популярных BI-решений по заданным критериям;
2. Выберите наиболее подходящий инструмент и обоснуйте свой выбор.

Задание 2

Для построения системы мониторинга логистики в масштабах холдинга сформированы требования:

- данные из 15 различных ERP-систем,
- хранение истории за 5 лет,
- построение 30+ дашбордов с интерактивными отчетами для 100+ пользователей,
- ежедневное обновление данных.

Выполните следующее:

1. Определите и обоснуйте состав необходимого программного обеспечения для каждого этапа: сбор и консолидация данных (ETL), хранение (DWH), визуализация и аналитика (BI).
2. Объясните, почему выбранный стек является оптимальным для данной задачи.

4) Использует прикладное программное обеспечение для решения конкретных прикладных задач.

Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции

Знать: 1. Концептуальную архитектуру типового BI-решения; 2. Встроенные механизмы оптимизации в BI-решениях; 3. Принципы проектирования пользовательского интерфейса дашбордов (UI/UX) и интерактивного сторителлинга.

Уметь: 1. Разворачивать и настраивать рабочее BI-окружение в облачной платформе (Yandex.Cloud); 2. Разрабатывать и внедрять единые стили оформления визуализаций в Matplotlib/Seaborn; 3. Оценивать производительность созданного BI-функционала и применять инструментальные методы для его оптимизации.

Типовые контрольные задания

Задание 1

Необходимо организовать новое рабочее пространство для отдела маркетинга в Yandex DataLens и обеспечить единый стиль для всех будущих дашбордов.

Выполните следующее:

1. Разверните новое окружение: создайте каталог "Маркетинг", внутри него — папки "В работе" и "Одобренные отчеты";
2. Создайте шаблонный дашборд, который будет включать:
 - а) стандартный заголовок с названием отдела, оформленный с помощью Markdown,
 - б) цветовую схему, соответствующую брендбуку компании (3 определенных цвета),
 - в) типовую легенду.

Задание 2

Дашборд в Yandex DataLens, построенный на основе сложного датасета с 10+ соединениями таблиц, стал загружаться очень медленно (более 30 секунд).

Выполните следующее:

1. Предложите 3 конкретных действия по оптимизации производительности этого дашборда, основанные на встроенных механизмах BI-платформы;
2. Обоснуйте, как каждое предложенное действие повлияет на скорость загрузки.

Примеры практико-ориентированных заданий

Задание 1

Предоставлен датасет с ежемесячными продажами по филиалам компании за последние 3 года. Данные содержат пропуски, дубликаты и аномальные значения (например, отрицательные суммы продаж). Также в данных присутствуют следующие поля: "Дата", "Филиал", "Менеджер", "Категория товара", "Сумма продажи", "Количество".

Выполните следующее:

1. Проведите аудит и очистку данных: устранили дубликаты, заполните пропуски и обработайте аномалии, обосновав выбранные методы;
2. Проведите исследовательский анализ (EDA), используя библиотеки Matplotlib/Seaborn, построив и интерпретировав: гистограмму распределения суммы продаж, блочную диаграмму для выявления выбросов по сумме продаж по филиалам, график динамики общего объема продаж по месяцам.

Задание 2

В Yandex DataLens загружены два CSV-файла: sales.csv (информация о продажах) и products.csv (справочник товаров с категориями и закупочной ценой). В данных есть несоответствия в названиях товаров, а для расчета маржинальности необходимо объединить таблицы.

Выполните следующее:

1. Создайте подключения и датасеты для обоих файлов в Yandex DataLens;
2. Создайте связь между датасетами по полю "ID_товара".
3. В датасете продаж создайте вычисляемое поле "Маржа", которое рассчитывается как "[Сумма продажи] - [Закупочная цена]".
4. Проверьте данные на наличие некорректных значений после объединения (например, продажи товаров, которых нет в справочнике).

Задание 3

Руководство банка поставило бизнес-задачу: «Снизить отток клиентов из премиум-сегмента».

Дана логическая модель данных банка в виде графической БД-схемы.

Выполните следующее:

1. Декомпозируйте бизнес-задачу в набор конкретных, измеримых показателей, которые можно оценить с помощью анализа данных;
2. Для каждого показателя определите, какие именно данные (показатели и атрибуты) необходимы для построения математической модели или анализа. 3. Составьте список требуемых полей из базы данных.

Задание 4

Дан набор данных. Выдвинута исследовательская гипотеза: «Существует зависимость между возрастом клиента и средним чеком в интернет-магазине электроники».

Выполните следующее:

1. Сформулируйте, какую математическую модель или статистический метод можно использовать для проверки этой гипотезы;
2. Выберите и обоснуйте два наиболее подходящих типа диаграмм для визуальной проверки этой гипотезы.

Задание 5

Вам необходимо выполнить два проекта:

Проект А: Регулярно (еженедельно) готовить для отдела маркетинга дашборд с ключевыми метриками (LTV, САС, конверсия по каналам). Данные хранятся в корпоративной DWH.

Проект Б: Разработать и визуализировать нестандартную модель кластеризации клиентов на основе их транзакционного поведения для отдела аналитики.

Выполните следующее:

1. Для каждого проекта обоснуйте выбор инструментария (BI-платформа или python-библиотеки), опираясь на такие критерии, как регулярность, интерактивность, сложность анализа и потребности конечных пользователей.

Задание 6

В организации есть данные по продажам товаров. Данные хранятся в SQL-базе.

Необходимо проанализировать выполнение плана продаж по регионам и товарным категориям за текущий квартал. Данные хранятся в SQL-базе. Необходимо предоставить возможность детализации (drill-down) от региона к конкретному менеджеру.

Выполните следующее:

1. Настройте подключение к данным вы будете использовать;
2. Разработайте чарт и реализуйте в нём drill-down с помощью иерархий так, чтобы детализировать данные от региона к конкретному менеджеру;
3. Создайте ещё три чарта разных типов, наилучшим образом отображающие план и факт, структуру продаж и динамику по регионам и товарным категориям;
4. Создайте дашборд в Yandex DataLens, разместив на нем все созданные чарты и расшарьте его преподавателю.

Задание 7

Представлен разработанный сторонним аналитиком дашборд. На нем представлены:

- 1) Heatmap (тепловая карта) корреляции между различными факторами (доход, возраст, количество кредитов) и вероятностью одобрения кредита;
- 2) Диаграмма рассеяния с двумя ключевыми факторами из heatmap.

Выполните следующее:

1. Дайте количественную и качественную интерпретацию результатов, отображенным на heatmap;
2. Сформулируйте две конкретные рекомендации для кредитного отдела на основе анализа этих визуализаций.

Задание 8

Перед запуском новой рекламной кампании необходимо проанализировать эффективность предыдущих кампаний по различным каналам (контекстная реклама, социальные сети, email-рассылки). Ключевые метрики: ROI, CTR, Cost per Lead.

Указанные данные есть в датасете.

Выполните следующее:

1. Создайте в Yandex DataLens интерактивный дашборд, позволяющий наглядно сравнивать эффективность каналов и показывать динамику ключевых метрик;
2. Настройте сторителлинг: выделите цветом лучший и худший канал, добавьте краткие текстовые выводы с помощью Markdown, обеспечьте возможность фильтрации по периоду.
3. Напишите 3-4 кратких вывода под дашбордом с рекомендациями по распределению бюджета на следующую кампанию.

Задание 9

Необходимо построить дашборд в Yandex DataLens на основе данных из двух источников: операционная БД (PostgreSQL) и файл с данными о регионах (CSV в облачном хранилище). Данные в БД обновляются ежечасно.

Выполните следующее:

1. Создайте в Yandex DataLens два подключения: к PostgreSQL и к облачному хранилищу. Для подключения к БД укажите корректные параметры (хост, порт, база данных, тип подключения);
2. Объедините данные из двух источников в одном датасете, создав связь по полю `region_id`;
3. Настройте материализованное представление для датасета с суточной частотой обновления.

Задание 10

В Yandex Cloud Organization создана структура: каталог "Финансы" с папками "Бюджетирование" и "Отчетность". Необходимо настроить права доступа для двух групп

пользователей: "Аналитики" (полный доступ в "Бюджетирование", только просмотр в "Отчетность") и "Менеджеры" (только просмотр готовых дашбордов в "Отчетность").

Выполните следующее:

1. В интерфейсе Yandex Cloud Organization создайте указанную структуру папок;
2. Назначьте соответствующие права (роли `datalens.creator` и `datalens.viewer`) на нужные папки для групп пользователей, следуя принципу наименьших привилегий.

Задание 11

На основе датасета с продажами (Регион -> Город -> Менеджер, Продажи) необходимо создать дашборд.

Выполните следующее:

1. Создайте в датасете Yandex DataLens древовидную иерархию "География": Регион -> Город -> Менеджер;
2. Постройте чарт (сводную таблицу), реализующий механизм drill-down по созданной иерархии;
3. Добавьте на дашборд селектор для фильтрации по периоду и свяжите его с чартом.

Задание 12

Имеется DataFrame `df` с полями `date`, `product_category`, `sales`, `profit`.

Выполните следующее:

1. Используя связку Matplotlib и Seaborn, постройте два графика на одной фигуре (subplots): линейный график динамики продаж по месяцам (верхний) и heatmap корреляции между продажами и прибылью по товарным категориям (нижний);
2. Настройте графическое оформление: для линейного графика задайте цвет и стиль линии, для heatmap — цветовую палитру Seaborn;
3. Добавьте заголовки и подписи осей.

Задание 13

Стартап с ограниченным бюджетом и командой из 5 аналитиков и 20 менеджеров ищет BI-решение. Ключевые требования: низкая стоимость, возможность быстрого создания дашбордов без написания кода, интеграция с Google BigQuery, удобный механизм расшаривания отчетов.

Выполните следующее:

1. Проведите сравнительный анализ 3 популярных BI-решений по заданным критериям;
2. Выберите наиболее подходящий инструмент и обоснуйте свой выбор.

Задание 14

Для построения системы мониторинга логистики в масштабах холдинга сформированы требования:

- данные из 15 различных ERP-систем,
- хранение истории за 5 лет,
- построение 30+ дашбордов с интерактивными отчетами для 100+ пользователей,
- ежедневное обновление данных.

Выполните следующее:

1. Определите и обоснуйте состав необходимого программного обеспечения для каждого этапа: сбор и консолидация данных (ETL), хранение (DWH), визуализация и аналитика (BI).
2. Объясните, почему выбранный стек является оптимальным для данной задачи.

Задание 15

Необходимо организовать новое рабочее пространство для отдела маркетинга в Yandex DataLens и обеспечить единый стиль для всех будущих дашбордов.

Выполните следующее:

1. Разверните новое окружение: создайте каталог "Маркетинг", внутри него — папки "В работе" и "Одобренные отчеты";
2. Создайте шаблонный дашборд, который будет включать:

- а) стандартный заголовок с названием отдела, оформленный с помощью Markdown,
- б) цветовую схему, соответствующую брендбуку компании (3 определенных цвета),
- в) типовую легенду.

Задание 16

Дашборд в Yandex DataLens, построенный на основе сложного датасета с 10+ соединениями таблиц, стал загружаться очень медленно (более 30 секунд).

Выполните следующее:

1. Предложите 3 конкретных действия по оптимизации производительности этого дашборда, основанные на встроенных механизмах BI-платформы;
2. Обоснуйте, как каждое предложенное действие повлияет на скорость загрузки.

Примерные вопросы для подготовки к зачету

Примерные вопросы к зачёту:

1. Дайте определение понятию "визуализация данных";
2. Перечислите основные бизнес-эффекты, достигаемые с помощью визуализации данных;
3. Приведите определение понятия "Process Mining";
4. Охарактеризуйте виды данных по степени структурированности. Приведите примеры для каждого вида.
5. Перечислите основные способы сбора данных.
6. Опишите высокоуровневый алгоритм анализа данных.
7. Опишите высокоуровневый алгоритм моделирования.
8. Приведите состав ключевых ролей аналитической команды data-проекта;

9. Дайте определение понятию "Data Mining";
10. Перечислите основные задачи Data Mining;
11. Дайте определение понятию "Исследовательский анализ данных" (EDA);
12. Перечислите способы проведения EDA;
13. Перечислите конкретные визуализации, применяемые для исследовательского анализа данных;
14. Приведите определение понятия "Business Intelligence" (BI);
15. Перечислите и охарактеризуйте этапы Knowledge Discovery in Databases (KDD);
16. Перечислите и кратко охарактеризуйте составляющие континуума BI-решений;
17. Охарактеризуйте ключевые тенденции рынка BI-решений;
18. Охарактеризуйте ключевые функциональные критерии выбора BI-системы;
19. Охарактеризуйте понятие adhoc-анализа;
20. Охарактеризуйте понятие "Extract, Transform, Load" (ETL);
21. Охарактеризуйте понятие "качество данных";
22. Охарактеризуйте понятие "дерево проблем качества данных";
23. Перечислите и кратко охарактеризуйте основные методы очистки данных;
24. Перечислите и кратко охарактеризуйте инструменты и технологии, используемые для очистки данных;
25. Охарактеризуйте понятие "Online Transaction Processing" (OLTP);
26. Охарактеризуйте понятие "Online Analytical Processing" (OLAP);
27. Охарактеризуйте понятие JDBC и его предназначение;
28. Охарактеризуйте понятие ODBC и его предназначение;
29. Дайте определение понятию хранилищу данных (Data Warehouse, DWH);
30. Кратко охарактеризуйте основные подходы к разработке DWH;
31. Охарактеризуйте Hadoop Distributed File System (HDFS) и её роль в анализе данных?

32. Охарактеризуйте понятие "обогащение данных". Какие существуют виды обогащения данных?
33. Перечислите и охарактеризуйте основные способы представления и визуализации данных;
34. Перечислите и охарактеризуйте виды диаграмм и гистограмм;
35. Перечислите и охарактеризуйте виды карт и картограмм, используемых в визуализации;
36. Перечислите и охарактеризуйте кросс-таблицы и кросс-диаграммы;
37. Охарактеризуйте понятие "интерактивный сторителлинг"? Каковы его принципы?
38. Охарактеризуйте понятие "дашборд". Чем он отличается от отчета?
39. Перечислите основные правила эффективной визуализации данных;
40. Опишите общую архитектуру данных в Yandex.Cloud;
41. Охарактеризуйте концептуальную архитектуру BI-решения на базе Yandex DataLens;
42. Перечислите и охарактеризуйте ключевые сущности Yandex DataLens;
43. Перечислите и охарактеризуйте виды источников данных, поддерживаемых Yandex DataLens;
44. Перечислите и охарактеризуйте варианты развертывания BI-решений на Yandex DataLens;
45. Опишите общий алгоритм работы Yandex DataLens;
46. Дайте понятие определению "материализованные представления". Для чего они используются?
47. Перечислите типы данных, поддерживаемые в Yandex DataLens;
48. Перечислите основные виды чартов, доступные в Yandex DataLens;
49. Охарактеризуйте понятие "древовидная иерархия" и охарактеризуйте его предназначение в Yandex DataLens;
50. Опишите процесс настройки Drill-down через древовидные структуры в Yandex DataLens;

51. Язык разметки Markdown: охарактеризуйте его предназначение и приведите примеры практического использования;
52. Охарактеризуйте особенности СУБД ClickHouse и опишите ключевые кейсы ее применения;
53. Охарактеризуйте механизм работы колоночной СУБД, сравните его со строковой СУБД;
54. Приведите определение понятий аутентификации, идентификации и авторизации;
55. Приведите определение понятия "федеративное управление идентификацией" (FIM);
56. Охарактеризуйте разницу между технологиями Single Sign-On и Same Sign-On;
57. Перечислите стандарты безопасности, поддерживаемые Yandex Cloud;
58. Охарактеризуйте понятие "Yandex Cloud Organization";
59. Опишите, как организована структура папок, воркбуков и каталогов в Yandex DataLens;
60. Перечислите критерии распределения прав доступа в Yandex DataLens;
61. Охарактеризуйте процесс обогащения датасета в Yandex DataLens;
62. Охарактеризуйте Yandex Cloud Marketplace, его предназначение и текущие возможности;
63. Охарактеризуйте понятие "QL-чарт" и его отличие от обычных чартов;
64. Перечислите ограничения, накладываемые на чарты на основе связанных датасетов в Yandex DataLens?
65. Охарактеризуйте процессы версионирования чартов в Yandex DataLens. Какие виды версий чартов существуют?
66. Перечислите общие настройки чарта в Yandex DataLens?
67. Охарактеризуйте процесс настройки измерений чарта в Yandex DataLens;
68. Охарактеризуйте процесс настройки показателей чарта в Yandex DataLens;
69. Охарактеризуйте особенности настройки различных видов чартов (линейного графика, столбчатой диаграммы, карты) в Yandex DataLens;
70. Охарактеризуйте синтаксис и аргументы функции plot() библиотеки matplotlib;

71. Охарактеризуйте синтаксис и аргументы функции `bar()` библиотеки `matplotlib`;
72. Охарактеризуйте синтаксис и аргументы функции `scatter()` библиотеки `matplotlib`. В чем её отличие от линейного графика?
73. Охарактеризуйте синтаксис и аргументы функции `stem()` библиотеки `matplotlib`. Приведите примеры ситуаций ее применения;
74. Охарактеризуйте синтаксис и аргументы функций `fill_between()` и `stackplot()` библиотеки `matplotlib`;
75. Охарактеризуйте синтаксис и аргументы функции `hist()` библиотеки `matplotlib`;
76. Охарактеризуйте синтаксис и аргументы функции `boxplot()` библиотеки `matplotlib`. Охарактеризуйте правила интерпретации блочной диаграммы;
77. Охарактеризуйте синтаксис и аргументы функции `pie()` библиотеки `matplotlib`. Приведите примеры ситуаций ее применения;
78. Охарактеризуйте синтаксис и аргументы функции `ecdf()` библиотеки `matplotlib`. В чем статистический смысл эмпирической кумулятивной функции распределения?
79. Перечислите диаграммы `matplotlib`, используемые для данных с координатной привязкой. Охарактеризуйте синтаксис и аргументы одной из таких функций
80. Охарактеризуйте разницу между функциями `pcolormesh()` и `imshow()`;
81. Охарактеризуйте синтаксис и аргументы функций `quiver()` и `streamplot()` библиотеки `matplotlib`;
82. Перечислите диаграммы `matplotlib`, используемые для визуализации данных на нерегулярных сетках. Охарактеризуйте синтаксис и аргументы одной из таких функций;
83. Перечислите трехмерные диаграммы `matplotlib`. Охарактеризуйте синтаксис и аргументы одной из таких функций;
84. Охарактеризуйте разницу между функциями `plot_surface()` и `plot_wireframe()`;
85. Охарактеризуйте ключевые модули `matplotlib`;
86. Охарактеризуйте `Matplotlib API`. Опишите его основные функции и приведите примеры их реализации;
87. Охарактеризуйте ключевые различия между `Seaborn` и `Matplotlib`;
88. Охарактеризуйте процесс настройки осей и стилей графиков в `Seaborn`;

89. Охарактеризуйте возможности Seaborn для работы с палитрами;
90. Охарактеризуйте синтаксис и аргументы функции `lmpplot()` библиотеки Seaborn;
91. Охарактеризуйте синтаксис и аргументы функции `relplot()` библиотеки Seaborn;
92. Охарактеризуйте синтаксис и аргументы функции `scatterplot()` библиотеки Seaborn;
93. Охарактеризуйте синтаксис и аргументы функции `lineplot()` библиотеки Seaborn;
94. Охарактеризуйте синтаксис и аргументы функции `displot()` библиотеки Seaborn;
95. Охарактеризуйте синтаксис и аргументы функции `histplot()` библиотеки Seaborn;
96. Охарактеризуйте синтаксис и аргументы функции `kdeplot()` библиотеки Seaborn;
97. Охарактеризуйте синтаксис и аргументы функций `FacetGrid()` и `PairGrid()` библиотеки Seaborn;
98. Охарактеризуйте синтаксис и аргументы функции `boxplot()` библиотеки Seaborn;
99. Охарактеризуйте синтаксис и аргументы функций `violinplot()` и `swarmplot()` библиотеки Seaborn;
100. Охарактеризуйте синтаксис и аргументы функции `boxplot()` библиотеки Seaborn;
101. Охарактеризуйте результат построения `clustermap()` и его отличие от `heatmap()` библиотеки Seaborn?

8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Нормативные правовые акты:

1. ГОСТ Р 56907-2016 «Бережливое производство. Визуализация»
2. ГОСТ Р 70846.6-2023: «Визуализация пространственных данных. Основные положения».

Основная литература:

1. Разведочный анализ данных. Python / Демидова Л. А. Ч. 1 : Разведочный анализ данных. Python. Часть 1 . Ч. 1 / Демидова Л. А. Москва : РТУ МИРЭА, 2022 107 с. Книга из коллекции РТУ МИРЭА - Информатика <https://e.lanbook.com/book/310970>. [БИК ID: RU-LAN-BOOK-310970]
2. Титов, А. Н. Визуализация данных в Python. Работа с библиотекой Matplotlib [Электронный ресурс] : учебно-методическое пособие / Титов А. Н.,Тазиева Р. Ф. Казань : КНИТУ, 2022 92 с. Книга из коллекции КНИТУ - Информатика <https://e.lanbook.com/book/331025> ISBN 978-5-7882-3176-1. [БИК ID: RU-LAN-BOOK-331025]
3. Разведочный анализ данных. Python / Демидова Л. А. Ч. 2 : Разведочный анализ данных. Python. Часть 2 . Ч. 2 / Демидова Л. А. Москва : РТУ МИРЭА, 2023 92 с. Книга из коллекции РТУ МИРЭА - Информатика <https://e.lanbook.com/book/382691> ISBN 978-5-7339-1933-1. [БИК ID: RU-LAN-BOOK-382691]

Дополнительная литература:

1. Глебов, В. И. Практикум по математической статистике: проверка гипотез с использованием Excel, MatCalc, R и Python : учебное пособие / В. И. Глебов, С. Я. Криволапов ; Финансовый университет при Правительстве Российской Федерации Москва : Прометей, 2019 87 с. : ил. Библиогр. в кн Режим доступа: электронная библиотечная система «Университетская библиотека ONLINE», требуется авторизация <https://biblioclub.ru/index.php?page=book&id=576035> ISBN 978-5-907100-66-4. [БИК ID: BIBLIOCLUB\0000576035]

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Блог о визуализации данных и информационном дизайне -<http://www.vmethods.ru>
2. Сервис визуализации и анализа данных от Yandex Cloud. URL: <https://cloud.yandex.ru/docs/dataLens/>
3. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf>)
4. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
5. Электронно-библиотечная система "Университетская библиотека ОНЛАЙН" <http://biblioclub.ru/>
6. Образовательная платформа "ЮРАЙТ" <https://urait.ru/>
7. Электронно-библиотечная система издательства "Лань" <https://e.lanbook.com/>

10. Методические указания для обучающихся по освоению дисциплины

Студентам необходимо руководствоваться «Методическими рекомендациями по планированию и организации внеаудиторной самостоятельной работы по образовательным программам бакалавриата и магистратуры в Финансовом университете» (Приказ ректора № 1040_о от 11.05.2021) и данной рабочей программой дисциплины.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

Комплект лицензионного программного обеспечения:

1. Jupyter Notebook (Windows 10)
2. Python 3.8
3. Пакет офисных программ
4. Реляционная СУБД
5. Anaconda
6. Онлайн-приложение для бизнес-аналитики Yandex DataLens
7. Kaspersky
8. Astra Linux
9. Libre Office

Современные профессиональные базы данных и информационные справочные системы:

1. Язык программирования Python 3. <https://pythonworld.ru/>
2. Информационно-правовая система «Гарант»

Сертифицированные программные и аппаратные средства защиты информации:

1. не предусмотрены

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

1. **Учебная аудитория** для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная оборудованием и техническими средствами обучения: мебель аудиторная (столы, стулья, доска аудиторная), набор демонстрационного оборудования (проектор, экран)

2. **Помещение для самостоятельной работы** обучающихся оснащенное компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа к электронной информационно-образовательной среде Университета.

3. **Компьютерный класс** для проведения учебных занятий, предусмотренных программой, в том числе групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенный оборудованием и техническими средствами обучения: мебель аудиторная (столы, стулья, доска аудиторная), персональные компьютеры, набор демонстрационного оборудования (проектор, экран)

4. Библиотека, читальный зал