

Финансовый университет при Правительстве Российской Федерации

УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН

программа профессиональной переподготовки

«Аналитик данных»

Категория слушателей: – для отдельных категорий граждан в соответствии с Постановлением Правительства Российской Федерации от 7 марта 2025 № 291

Срок обучения: – 256 ак.ч, не менее 10 недель

Форма обучения: – очно-заочная с применением дистанционных образовательных технологий

№ п/п	Наименование дисциплин (модулей, курсов), разделов, тем	Трудоёмкость		Контактные часы, в т.ч. с применением ДОТ						СРС, ч	Формы контроля
				лекции		лабораторные работы		практические и семинарские занятия			
		Всего, ч.	Из них с ДОТ, ч	Всего, ч.	Из них с ДОТ, ч	Всего, ч.	Из них с ДОТ, ч	Всего, ч.	Из них с ДОТ, ч		
I	Дисциплина 1. Сбор данных	66	66	12	12	-	-	20	20	34	Зачет
	Тема 1.1. Аналитика данных в бизнесе и других сферах	10	10	2	2	-	-	2	2	6	
	Тема 1.2. Типы и источники данных	10	10	2	2	-	-	2	2	6	
	Тема 1.3. Сбор данных с различных источников	10	10	2	2	-	-	2	2	6	
	Тема 1.4. Основы работы с SQL	12	12	2	2	-	-	4	4	6	
	Тема 1.5. Загрузка и интеграция данных из различных форматов	10	10	2	2	-	-	4	4	4	
	Тема 1.6. Инструменты для сбора данных	12	12	2	2	-	-	4	4	6	
	Промежуточная аттестация	2	2	-	-	-	-	2	2	-	
II	Тема 2. Обработка данных	70	70	12	12	-	-	22	22	36	Зачет
	Тема 2.1. Основы Python для обработки данных	10	10	2	2	-	-	2	2	6	
	Тема 2.2. Очистка данных	10	10	2	2	-	-	2	2	6	
	Тема 2.3. Преобразование данных	12	12	2	2	-	-	4	4	6	

	Тема 2.4. Нормализация и стандартизация данных	12	12	2	2	-	-	4	4	6	
	Тема 2.5. Основы работы с R для обработки данных	12	12	2	2	-	-	4	4	6	
	Тема 2.6. Работа с большими данными	12	12	2	2	-	-	4	4	6	
	Промежуточная аттестация	2	2	-	-	-	-	2	2	-	
III	Тема 3. Интерпретация	68	68	12	12	-	-	20	20	36	Зачет
	Тема 3.1. Основы статистического анализа данных	10	10	2	2	-	-	2	2	6	
	Тема 3.2. Анализ взаимосвязей	10	10	2	2	-	-	2	2	6	
	Тема 3.3. Тестирование гипотез	10	10	2	2	-	-	2	2	6	
	Тема 3.4. Анализ временных рядов и прогнозирование	12	12	2	2	-	-	4	4	6	
	Тема 3.5. Методы кластеризации и классификации	12	12	2	2	-	-	4	4	6	
	Тема 3.6. Интерпретация и представление результатов анализа	12	12	2	2	-	-	4	4	6	
	Промежуточная аттестация	2	2	-	-	-	-	2	2	-	
IV	Тема 4. Построение отчетов	44	44	6	6	-	-	16	16	22	Зачет
	Тема 4.1. Первичная визуализация данных	10	10	2	2	-	-	2	2	6	
	Тема 4.2. Визуализация данных в Python и R	12	12	2	2	-	-	4	4	6	
	Тема 4.3. Визуализация данных в BI-инструментах	12	12	2	2	-	-	4	4	6	
	Тема 4.4. Презентация аналитических результатов для руководства и заинтересованных сторон	8	8	-	-	-	-	4	4	4	

	Промежуточная аттестация	2	2	-	-	-	-	2	2	-	
	Итоговая аттестация	8	8		-	-	-	8	8	-	Защита итогового практического о задания
	Итого	256	256 / 100%	42	42	-	-	86	86	128	

Директор Института развития
профессиональных компетенций и квалификаций



Т.А. Болтенко

Федеральное государственное образовательное бюджетное учреждение
высшего образования
«Финансовый университет при Правительстве Российской Федерации»
(Финансовый университет)

Институт развития профессиональных
компетенций и квалификаций

ОБСУЖДЕНО И ОДОБРЕНО

на Ученом совете институтов и школ
дополнительного профессионального
образования

Протокол от 14.08.2025 № 46

УТВЕРЖДАЮ



Проректор по дополнительному
профессиональному образованию

Е.А. Диденко

2025 г.

МП

ПРОГРАММА

профессиональной переподготовки

«Аналитик данных»

Москва – 2025

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ «Аналитик данных»

Общая характеристика программы

Цель программы: формирование и совершенствование компетенций, необходимых для выполнения нового вида профессиональной деятельности в сфере «Создание и применение технологий больших данных», совершенствование навыков по различным аспектам профессиональной деятельности, освоение слушателями новых способов решения профессиональных задач, развития кадрового потенциала для повышения эффективности деятельности организации с учетом профессиональных стандартов.

Перечень нормативных документов, определяющих квалификационные требования к выпускнику программы:

1. Профессиональный стандарт «Бизнес-аналитик», утвержденный приказом Министерства труда и социальной защиты Российской Федерации от 22.11.2023 № 821н, код 08.037;
2. Профессиональный стандарт «Специалист по большим данным», утвержденный приказом Министерства труда и социальной защиты Российской Федерации от 6 июля 2020 года N 405н, код 06.042;
3. Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 38.03.05 Бизнес-информатика, утвержденный приказом Министерства науки и высшего образования Российской Федерации от 29 июля 2020 г. N 838¹.

Характеристика нового вида профессиональной деятельности и/или трудовых функций:

Область профессиональной деятельности: Связь, информационные и коммуникативные технологии.

Вид профессиональной деятельности: Создание и применение технологий больших данных.

Обобщенные трудовые функции, на реализацию которых направлена программа, входящие в профессиональный стандарт «Бизнес-аналитик»:

¹ С изменениями и дополнениями от: 26 ноября 2020 г., 19 июля 2022 г., 27 февраля 2023 г.

Обобщенные трудовые функции			Трудовые функции		
код	наименование	уровень квалификации	наименование	код	уровень (подуровень) квалификации
C	Выявление бизнес-проблем или бизнес-возможностей	5	Сбор информации о бизнес-проблемах или бизнес-возможностях	C/001.5	5
D	Обоснование решений	6	Анализ, обоснование и выбор решения	D/02.6	6

Обобщенные трудовые функции, на реализацию которых направлена программа, входящие в профессиональный стандарт «Специалист по большим данным»:

Обобщенные трудовые функции			Трудовые функции		
код	наименование	уровень квалификации	наименование	код	уровень (подуровень) квалификации
A	Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры	6	Подготовка данных для проведения аналитических работ по исследованию больших данных	A/03.6	6
			Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика	A/04.6	6

Характеристика профессиональных компетенций, подлежащих совершенствованию, и/или перечень новых компетенций, формирующихся в процессе освоения программы

Слушатель, успешно прошедший обучение по программе профессиональной переподготовки «Аналитик данных», должен обладать

следующими компетенциями, необходимыми для выполнения нового вида профессиональной деятельности в сфере создание и применение технологий больших данных:

- способность к сбору информации о бизнес-проблемах или бизнес-возможностях (ТФ № С/001.5, ПС № 821н);
- способность к анализу, обоснованию и выбору решения (ТФ № D/02.6, ПС № 821н);
- способность готовить данные для проведения аналитических работ по исследованию больших данных (ТФ № А/03 .6, ПС № 405н);
- способность к проведению аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика (ТФ № А/04.6, ПС № 405н).

Планируемые результаты обучения по программе

По итогам освоения программы слушатели должны:

Знать:

- основные сферы применения аналитики данных в бизнесе и других отраслях;
- типы и форматы данных (структурированные, неструктурированные).
- особенности реляционных СУБД и основы языка SQL;
- общие подходы и инструменты для сбора и автоматизации выгрузки данных (ETL, API, парсинг);
- методы загрузки и интеграции данных из различных форматов (CSV, JSON, XML и др.);
- основные библиотеки Python для обработки данных;
- ключевые пакеты языка R и их функциональность;
- методы очистки, нормализации и стандартизации данных;
- типичные проблемы качества данных: пропуски, дубликаты, аномалии;
- основные концепции работы с большими данными;
- ключевые статистические понятия и методы (среднее, дисперсия, корреляция, регрессия);
- принципы постановки и проверки гипотез;
- основы анализа временных рядов и методы прогнозирования.
- базовые алгоритмы машинного обучения (регрессия, кластеризация, классификация) и иные нейросетевые возможности;
- методы оценки качества моделей (метрики, перекрёстная проверка);
- структуру и требования к интерпретации результатов;
- основные принципы и правила визуализации данных;

- возможности Python и R для визуализации;
- функциональные особенности BI-платформ;
- структуру и логику построения аналитических отчётов и презентаций;
- методы сторителлинга для донесения ключевых инсайтов;
- этические и юридические аспекты работы с данными.

Уметь:

- определять релевантные источники данных для решения конкретных бизнес-задач;
- формировать SQL-запросы средней сложности и оптимизировать их;
- собирать и консолидировать данные из нескольких источников в единый набор;
- использовать инструменты для извлечения и первичной фильтрации данных (ETL-платформы, скрипты);
- анализировать качество исходных данных и выбирать методы их корректной интеграции;
- загружать набор данных в Python или R и выполнять его первичную разведку;
- выполнять очистку и преобразование в соответствии с задачами исследования;
- применять методы нормализации/стандартизации и оценивать их влияние на дальнейший анализ;
- организовывать и обрабатывать большие объёмы данных в распределённых средах;
- создавать скрипты/ноутбуки для повторяемой предобработки данных;
- проводить статистический анализ данных и интерпретировать полученные статистические показатели;
- формулировать гипотезы для бизнес-задач и проверять их (A/B-тестирование и др.);
- применять методы прогнозирования на основе временных рядов (ARIMA, ETS и др.);
- выбирать соответствующие модели машинного обучения (кластеризация, классификация) и настраивать их гиперпараметры;
- оценивать качество обученных моделей и интерпретировать результаты;
- делать выводы на основе анализа и формировать рекомендации для дальнейших действий;
- выбирать соответствующий тип диаграммы/визуализации под конкретную аналитическую задачу;
- создавать интерактивные дашборды в BI-инструментах;

- использовать Python или R для формирования наглядных графиков и диаграмм с анимацией/взаимодействием;
- структурировать отчёт в соответствии с интересами и уровнем компетенции целевой аудитории;
- презентовать результаты анализа, формулируя чёткие выводы и рекомендации для руководства и стейкхолдеров.

Владеть:

- навыками планирования процесса сбора данных под конкретные аналитические цели;
- приёмами загрузки и объединения данных из разнородных систем;
- техниками базовой оптимизации SQL-запросов при работе с большими объёмами данных;
- практическими методами автоматизации процессов сбора (сценарии, планировщики задач);
- инструментами контроля версии для совместной работы над скриптами и SQL-запросами;
- приёмами эффективного использования Python и R для различного вида данных (числовые, текстовые);
- навыками быстрой диагностики и устранения проблем качества (пропуски, выбросы, несогласованные типы);
- методами группировки, агрегации и фильтрации данных для получения промежуточных результатов анализа;
- техниками оптимизации кода (векторизация, использование встроенных функций) при обработке больших наборов данных;
- базовой работой с инструментами Big Data и принципами их интеграции с Python/R;
- навыками анализа взаимосвязей между переменными и выявления причинно-следственных связей;
- приёмами применения библиотек машинного обучения;
- методами корректной визуализации статистических результатов и параметров моделей;
- техниками прогнозирования временных рядов и оценки точности предсказаний;
- практическими техниками разработки многопользовательских дашбордов и их интеграцией в корпоративную среду;
- приёмами дизайна визуализации (цветовые схемы, фокус на ключевых данных, упрощение графиков);

- навыками построения увлекательного и убедительного сторителлинга вокруг данных;
- средствами подготовки презентаций с данными, автоматически обновляемыми из BI;
- методами документирования и архивирования отчётов, включающими версии и историю изменений.

Федеральное государственное образовательное бюджетное
учреждение высшего образования
**«Финансовый университет при Правительстве Российской Федерации»
(Финансовый университет)**

Институт развития профессиональных
компетенций и квалификаций

ОБСУЖДЕНО И ОДОБРЕНО

на Ученом совете институтов и школ
дополнительного профессионального
образования

Протокол от 28.01.2025 № 46

УТВЕРЖДАЮ



МП

УЧЕБНЫЙ ПЛАН

Программа профессиональной переподготовки
«Аналитик данных»

Требования к уровню образования слушателей	лица, имеющие среднее профессиональное и (или) высшее образование, лица, получающие среднее профессиональное и (или) высшее образование
Категория слушателей	Руководители и специалисты: лица, желающие освоить новый вид профессиональной деятельности в области анализа данных и машинного обучения
Новый вид профессиональной деятельности/ новая присваиваемая квалификация	Создание и применение технологий больших данных
Срок освоения программы	256 часов, не менее 10 недель
Форма обучения	очная или очно-заочная, с применением дистанционных образовательных технологий и электронного обучения
Режим занятий	4 - 8 часов в день

№№ п/п	Наименование дисциплины, модуля	Трудоемкость		В том числе				Форма контроля
				Контактная работа ²			Самостоятельная работа*	
				Всего часов	из них			
		Лекции	Практические занятия					
1	2	3 В зачетных единицах	4 В часах	5	6	7	8	9
1	Дисциплина 1. Сбор данных	-	66	32	12	20	34	Зачет
2	Дисциплина 2. Обработка данных	-	70	34	12	22	36	Зачет
3	Дисциплина 3. Интерпретация	-	68	32	12	20	36	Зачет
4	Дисциплина 4. Построение отчетов	-	44	22	6	16	22	Зачет
5	Всего:	-	248	120	42	78	128	
6	Стажировка	-	-	-	-	-	-	-
7	Практика	-	-	-	-	-	-	-
8	Подготовка и проведение итоговой аттестации	-	8	8	-	8	-	Защита итогового практического задания
9	Общая трудоемкость программы:	-	256	128	42	86	128	

Разработчики программы:

Андрянов Никита Андреевич - кандидат технических наук, доцент кафедры искусственного интеллекта Факультета информационных технологий и анализа больших данных Финуниверситета.

Миронов Владимир Олегович - кандидат технических наук, старший преподаватель кафедры искусственного интеллекта Факультета информационных технологий и анализа больших данных Финуниверситета.

Сахнюк Павел Анатольевич - кандидат технических наук, ведущий аналитик департамента по развитию стратегических проектов "Ростелеком",

* С применением дистанционных образовательных технологий и электронного обучения (при наличии).

доцент кафедры бизнес-информатики Факультета информационных технологий и анализа больших данных (Внешнее совместительство).

Сорокин Артем Денисович - сотрудник Центра перспективных проектов Института развития профессиональных компетенций и квалификаций Финансового университета.

В реализации программы принимают участие преподаватели Финансового университета, приглашенные ведущие специалисты в профильной сфере.

Директор ИРПКиК



Т.А. Болтенко

«__» _____ 20__ г.



Учебно-тематический план

Программа профессиональной переподготовки «Аналитик данных»

Форма обучения – очная или очно-заочная, с применением
дистанционных образовательных технологий и электронного обучения

№№ п/п	Наименование дисциплины, модуля	Трудоемкость		В том числе				Форма контроля
				Контактная работа ³			Самостоятельная работа ⁴	
				Всего часов	из них			
		Лекции	Практические занятия					
1	2	В зачетных единицах	В часах	5	6	7	8	9
1	Дисциплина 1. Сбор данных	-	66	32	12	20	34	Зачет
2	Тема 1.1. Аналитика данных в бизнесе и других сферах	-	10	4	2	2	6	Тестирование
3	Тема 1.2. Типы и источники данных	-	10	4	2	2	6	Тестирование
4	Тема 1.3. Сбор данных с различных источников	-	10	4	2	2	6	Тестирование
5	Тема 1.4. Основы работы с SQL	-	12	6	2	4	6	Тестирование
6	Тема 1.5. Загрузка и интеграция данных из различных форматов	-	10	6	2	4	4	Тестирование
7	Тема 1.6. Инструменты для сбора данных	-	12	6	2	4	6	Тестирование
8	Промежуточная аттестация	-	2	2	-	2	-	Зачет
9	Дисциплина 2. Обработка данных	-	70	34	12	22	36	Зачет
10	Тема 2.1. Основы Python для обработки данных	-	10	4	2	2	6	Тестирование
11	Тема 2.2. Очистка данных	-	10	4	2	2	6	Тестирование
12	Тема 2.3. Преобразование данных	-	12	6	2	4	6	Тестирование

* С применением дистанционных образовательных технологий и электронного обучения (при наличии).

13	Тема 2.4. Нормализация и стандартизация данных	-	12	6	2	4	6	Тестирование
14	Тема 2.5. Основы работы с R для обработки данных	-	12	6	2	4	6	Тестирование
15	Тема 2.6. Работа с большими данными	-	12	6	2	4	6	Тестирование
16	Промежуточная аттестация	-	2	2	-	2	-	Зачет
17	Дисциплина 3. Интерпретация	-	68	32	12	20	36	Зачет
18	Тема 3.1. Основы статистического анализа данных	-	10	4	2	2	6	Тестирование
19	Тема 3.2. Анализ взаимосвязей	-	10	4	2	2	6	Тестирование
20	Тема 3.3. Тестирование гипотез	-	10	4	2	2	6	Тестирование
21	Тема 3.4. Анализ временных рядов и прогнозирование	-	12	6	2	4	6	Тестирование
22	Тема 3.5. Методы кластеризации и классификации	-	12	6	2	4	6	Тестирование
23	Тема 3.6. Интерпретация и представление результатов анализа	-	12	6	2	4	6	Тестирование
24	Промежуточная аттестация	-	2	2	-	2	-	Зачет
25	Дисциплина 4. Построение отчетов	-	44	22	6	16	22	Зачет
26	Тема 4.1. Первичная визуализация данных	-	10	4	2	2	6	Тестирование
27	Тема 4.2. Визуализация данных в Python и R	-	12	6	2	4	6	Тестирование
28	Тема 4.3. Визуализация данных в BI-инструментах	-	12	6	2	4	6	Тестирование
29	Тема 4.4. Презентация аналитических результатов для руководства и заинтересованных сторон	-	8	4	-	4	4	Тестирование
30	Промежуточная аттестация	-	2	2	-	2	-	Зачет
31	Всего:	-	248	120	42	78	128	-
32	Стажировка (при наличии)	-	-	-	-	-	-	-
33	Практика (при наличии)	-	-	-	-	-	-	-
34	Подготовка и проведение итоговой аттестации	-	8	8	-	8	-	Защита итогового практического задания
35	Общая трудоемкость программы:	-	256	128	42	86	128	

Федеральное государственное образовательное бюджетное
учреждение высшего образования
**«Финансовый университет при Правительстве
Российской Федерации»**

Институт развития профессиональных
компетенций и квалификаций

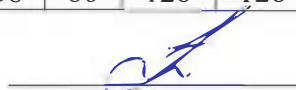
Календарный учебный график

Программа профессиональной переподготовки
«Аналитик данных»

Срок освоения программы 256 час.
Продолжительность
обучения 3 месяца(-ев)
Форма обучения - очная или очно-заочная, с применением
дистанционных образовательных технологий и
электронного обучения

№ п/п	Наименование дисциплины, модуля	1 месяц	2 месяц	3 месяц	КР	СР	П	С	ПА	ИА	Всего
1.	Дисциплина 1. Сбор данных	66			32	34			2		66
2.	Дисциплина 2. Обработка данных	22	48		34	36			2		70
3.	Дисциплина 3. Интерпретация		40	28	32	36			2		68
4.	Дисциплина 4. Построение отчетов			44	22	22			2		44
5.	Подготовка и проведение итоговой аттестации			8	8					8	8
6.	Общая трудоемкость программы	88	88	80	128	128			8	8	256

Директор ИРПКиК



Т.А. Болтенко

Условные обозначения	
ПА	Промежуточная аттестация
П	Практика
С	Стажировка
ИА	Итоговая аттестация
КР	Контактная работа

Программа профессиональной переподготовки
«Аналитик данных»

состоит из:

- Дисциплина 1. Сбор данных
- Дисциплина 2. Обработка данных
- Дисциплина 3. Интерпретация
- Дисциплина 4. Построение отчетов

Рабочая программа Дисциплины 1. Сбор данных

Цель освоения дисциплины - приобретения слушателями компетенций для комплексного сбора, консолидации и первичной интеграции данных из различных источников, включая реляционные СУБД и внешние форматы.

Компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения дисциплины:

- способность к сбору информации о бизнес-проблемах или бизнес-возможностях;
- способность к анализу, обоснованию и выбору решения;
- способность готовить данные для проведения аналитических работ по исследованию больших данных.

Планируемые результаты обучения

По итогам освоения программы слушатели должны:

Знать:

- основные сферы применения аналитики данных в бизнесе и других отраслях;
- типы и форматы данных (структурированные, неструктурированные).
- особенности реляционных СУБД и основы языка SQL;
- общие подходы и инструменты для сбора и автоматизации выгрузки данных (ETL, API, парсинг);
- методы загрузки и интеграции данных из различных форматов (CSV, JSON, XML и др.).

Уметь:

- определять релевантные источники данных для решения конкретных бизнес-задач;
- формировать SQL-запросы средней сложности и оптимизировать их;
- собирать и консолидировать данные из нескольких источников в единый набор;

- использовать инструменты для извлечения и первичной фильтрации данных (ETL-платформы, скрипты);
- анализировать качество исходных данных и выбирать методы их корректной интеграции.

Владеть:

- навыками планирования процесса сбора данных под конкретные аналитические цели;
- приёмами загрузки и объединения данных из разнородных систем;
- техниками базовой оптимизации SQL-запросов при работе с большими объёмами данных;
- практическими методами автоматизации процессов сбора (сценарии, планировщики задач);
- инструментами контроля версии для совместной работы над скриптами и SQL-запросами.

Учебно-тематический план дисциплины

№№ п/п	Наименование дисциплины, модуля	Трудоемкость		В том числе				Форма контроля
				Контактная работа ⁴			Самостоятельная работа*	
				Всего часов	из них			
		Лекции	Практические занятия					
1	2	В зачетных единицах	В часах	5	6	7	8	9
1	Дисциплина 1. Сбор данных	-	66	32	12	20	34	Зачет
2	Тема 1.1. Аналитика данных в бизнесе и других сферах	-	10	4	2	2	6	Тестирование
3	Тема 1.2. Типы и источники данных	-	10	4	2	2	6	Тестирование
4	Тема 1.3. Сбор данных с различных источников	-	10	4	2	2	6	Тестирование
5	Тема 1.4. Основы работы с	-	12	6	2	4	6	Тестирование

* С применением дистанционных образовательных технологий и электронного обучения (при наличии).

	SQL							
6	Тема 1.5. Загрузка и интеграция данных из различных форматов	-	10	6	2	4	4	Тестирование
7	Тема 1.6. Инструменты для сбора данных	-	12	6	2	4	6	Тестирование
8	Промежуточная аттестация	-	2	2	-	2	-	Зачет

Содержание дисциплины

Тема 1.1. Аналитика данных в бизнесе и других сферах

Роль данных в принятии решений: почему компании инвестируют в аналитику, примеры успешных кейсов. Основные направления аналитики: продуктовая, маркетинговая, финансовая, операционная. Требования к специалисту по аналитике данных: обзор компетенций, инструментов и методологий.

Тема 1.2. Типы и источники данных

Структурированные, полуструктурированные и неструктурированные данные: особенности, примеры использования. Внутренние и внешние источники: CRM, ERP-системы, веб-сайты, соцсети, открытые базы данных. Ключевые форматы данных: CSV, JSON, XML, SQL-таблицы, API-ответы.

Тема 1.3. Сбор данных с различных источников

ETL-процесс: Extract, Transform, Load — базовые понятия. Web-scraping и API: инструменты (Requests, BeautifulSoup, Postman) и нюансы юридической стороны. Парсинг файлов: чтение данных из файлов (CSV, Excel, JSON), первичная фильтрация.

Тема 1.4. Основы работы с SQL

Структура реляционных СУБД: таблицы, связи, ключи. Синтаксис базовых SQL-запросов: SELECT, FROM, WHERE, GROUP BY, JOIN. Оптимизация и индексация: понятия индексов, принцип их работы, простые приёмы оптимизации.

Тема 1.5. Загрузка и интеграция данных из различных форматов

Способы соединения и совмещения данных: объединение таблиц, импорт из CSV/JSON/XML. Обработка ошибок и конфликтов форматов: кодировки, несовместимые типы, пропущенные значения. Инструменты интеграции:

Python (pandas), SQL (LOAD DATA), ETL-платформы (например, Pentaho, Talend).

Тема 1.6. Инструменты для сбора данных

ETL-платформы: обзор Airflow, Luigi, SSIS, Talend. Сервисы и библиотеки для автоматизации сбора: PySpark, Python-скрипты, bash-скрипты. Организация расписаний и мониторинг; планировщики задач, логирование и алёрты.

Содержание практических занятий

№ темы	Наименование (содержание) темы, по которой предусмотрено практическое занятие	Формы и методы проведения
1	Тема 1.1. Аналитика данных в бизнесе и других сферах	Устный опрос, выполнение практических заданий, обмен опытом
2	Тема 1.2. Типы и источники данных	Устный опрос, выполнение практических заданий, обмен опытом
3	Тема 1.3. Сбор данных с различных источников	Устный опрос, выполнение практических заданий, анализ конкретных ситуаций, обмен опытом
4	Тема 1.4. Основы работы с SQL	Устный опрос, выполнение практических заданий, обмен опытом
5	Тема 1.5. Загрузка и интеграция данных из различных форматов	Устный опрос, выполнение практических заданий, обмен опытом
6	Тема 1.6. Инструменты для сбора данных	Устный опрос, выполнение практических заданий, обмен опытом

Содержание самостоятельной работы слушателей

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ п/п	Наименование (содержание) темы	Формы и методы проведения
1	Тема 1.1. Аналитика данных в бизнесе и других сферах	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
2	Тема 1.2. Типы и источники данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
3	Тема 1.3. Сбор данных с различных источников	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
4	Тема 1.4. Основы работы с SQL	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

5	Тема 1.5. Загрузка и интеграция данных из различных форматов	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
6	Тема 1.6. Инструменты для сбора данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

Список литературы:

Основная литература:

1. Бринк, Ричардс, Феверолф: Машинное обучение. – СПб.: Питер, 2023. – 336 с.
2. Зараменских, Е. П. Основы бизнес-информатики : учебник и практикум для вузов / Е. П. Зараменских. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 470 с. — (Высшее образование). — ISBN 978-5-534-15039-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511961>.
3. Data Science. Наука о данных с нуля. / Билл Фрэнкс.; пер. с англ. Евстигнеева И.В. – М.: Издательство «Альпина Паблишер». – 2022. – 320 с.
4. Набатова, Д. С. Математические и инструментальные методы поддержки принятия решений: учебник и практикум для вузов / Д. С. Набатова. — Москва: Издательство Юрайт, 2025. — 292 с. — (Высшее образование). — ISBN 978-5-534-02699-3.
5. Курносоев Ю.В. «Азбука аналитики», Издательство «Концептуал», 2018 -240 с.
6. Б. Марр «Ключевые инструменты бизнес-аналитики»/ пер с англ. Егоров В. Н., Издательство «Лаборатория знаний», 2021 – 339 с.
7. де Прадо М. «Машинное обучение: алгоритмы для бизнеса», Санкт Петербург: Издательский дом «Питер». – 2019. – 432 с.
8. Плас вандер Д. «Python для сложных задач: наука о данных и машинное обучение», Санкт Петербург: Издательский дом «Питер». – 2023. – 576с.
9. Лакшманан В., Тайджани Д. «Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении», Санкт Петербург: Издательский дом «Питер». – 2021. – 496с.

Дополнительная литература:

1. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени. / Натан Марц и Джеймс Уоррен.; пер. с англ. – М.: Вильямс. – 2020. – 336 с.
2. Глубокое обучение на Python. – СПб.: Питер, 2022. – 400 с.: ил. – (Серия «Библиотека программиста»).
3. Основы Data Science и Big Data. Python и наука о данных. / Силен Д., Мейсман А., Али М.; пер. с англ. – Санкт Петербург: Издательский дом «Питер». – 2018. – 336 с.

4. Паклин Н.Б. Орешков В.И. Бизнес-аналитика: от данных к знаниям, СПб: Питер 2013. – 706 с.
5. Радченко И.А, Николаев И.Н. Технологии и инфраструктура Big Data. – СПб: Университет ИТМО, 2020. – 52 с.
6. Data Science. Наука о данных с нуля. / Джоэл Грас.; пер. с англ. – 2-е изд., перераб. – Санкт Петербург: Издательство «БХВ-Петербург». – 2021. – 416 с.
7. Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении. – СПб.: Питер, 2021. – 496 с.: ил.
8. К. Андерсон «Аналитическая культура»/Издательство. Манн, Иванов, Фербер, 2023 – 332 с.
9. Д. Битти К. Вигерс «Разработка требований к программному обеспечению», Издательство BHV, 2020 -736 с.

Интернет-ресурсы, информационно-справочные и поисковые системы

1. <https://www.anaconda.com/> – страница загрузки Anaconda.
2. <https://powerbi.microsoft.com/ru-ru/desktop/> – страница загрузки платформы Power BI Desktop компании Microsoft.
3. <https://loginom.ru/download> – страница загрузки платформы Loginom компании BaseGroup Labs.
4. <https://www.knime.com/knime-analytics-platform> – страница загрузки платформы Knime Analytics Platform.
5. <https://community.cloud.databricks.com/login.html> – страница регистрации Databricks Community Edition
6. <https://rapidminer.com/get-started/> – страница загрузки платформы RapidMiner.
7. <http://h2o-release.s3.amazonaws.com/h2o/rel-zermelo/4/index.html> – страница загрузки платформы H2O.
8. <https://cloud.google.com/bigquery/docs/sandbox> – страница BigQuery sandbox.
9. <https://loginom.ru/> - аналитическая low-code платформа, которая позволяет проводить анализ данных любого уровня сложности без программирования
10. <https://cloud.yandex.ru/services/datалens> – сервис визуализации и анализа данных Яндекс.

Рабочая программа Дисциплины 2. Обработка данных

Цель освоения дисциплины – приобретения слушателями компетенций для эффективной очистки, преобразования и базовой аналитической подготовки данных с применением Python и R, включая работу с большими наборами данных.

Компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения дисциплины:

- способность к сбору информации о бизнес-проблемах или бизнес-возможностях;
- способность к анализу, обоснованию и выбору решения;
- способность готовить данные для проведения аналитических работ по исследованию больших данных;
- способность к проведению аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика.

Планируемые результаты обучения

По итогам освоения программы слушатели должны:

Знать:

- основные библиотеки Python для обработки данных;
- ключевые пакеты языка R и их функциональность;
- методы очистки, нормализации и стандартизации данных;
- типичные проблемы качества данных: пропуски, дубликаты, аномалии;
- основные концепции работы с большими данными.

Уметь:

- загружать набор данных в Python или R и выполнять его первичную разведку;
- выполнять очистку и преобразование в соответствии с задачами исследования;
- применять методы нормализации/стандартизации и оценивать их влияние на дальнейший анализ;
- организовывать и обрабатывать большие объёмы данных в распределённых средах;
- создавать скрипты/ноутбуки для повторяемой предобработки данных.

Владеть:

- приёмами эффективного использования Python и R для различного вида данных (числовые, текстовые);

- навыками быстрой диагностики и устранения проблем качества (пропуски, выбросы, несогласованные типы);
- методами группировки, агрегации и фильтрации данных для получения промежуточных результатов анализа;
- техниками оптимизации кода (векторизация, использование встроенных функций) при обработке больших наборов данных;
- базовой работой с инструментами Big Data и принципами их интеграции с Python/R.

Учебно-тематический план дисциплины

№№ п/п	Наименование дисциплины, модуля	Трудоемкость		В том числе				Форма контроля
				Контактная работа ⁵			Самостоятельная работа*	
				Всего часов	из них			
		Лекции	Практические занятия					
		В зачетных единицах	В часах					
1	Дисциплина 2. Обработка данных	-	70	34	12	22	36	Зачет
2	Тема 2.1. Основы Python для обработки данных	-	10	4	2	2	6	Тестирование
3	Тема 2.2. Очистка данных	-	10	4	2	2	6	Тестирование
4	Тема 2.3. Преобразование данных	-	12	6	2	4	6	Тестирование
5	Тема 2.4. Нормализация и стандартизация данных	-	12	6	2	4	6	Тестирование
6	Тема 2.5. Основы работы с R для обработки данных	-	12	6	2	4	6	Тестирование
7	Тема 2.6. Работа с большими данными	-	12	6	2	4	6	Тестирование
8	Промежуточная аттестация	-	2	2	-	2	-	Зачет

Содержание дисциплины

* С применением дистанционных образовательных технологий и электронного обучения (при наличии).

Тема 2.1. Основы Python для обработки данных

Структура языка Python: типы данных, управление потоком (if, for, while). Основные библиотеки: NumPy, pandas — их назначение и базовые операции (Series, DataFrame). Работа в Jupyter Notebook: организация кода, визуальный вывод, практические упражнения.

Тема 2.2. Очистка данных

Обнаружение и обработка пропущенных значений: методы (удаление, заполнение средними). Выбросы и аномалии: способы обнаружения (межквартильный размах, z-score). Устранение дубликатов и несоответствий: merge, drop_duplicates, валидация данных.

Тема 2.3. Преобразование данных

Изменение структуры набора данных: pivot, melt, stack/unstack, reshape. Объединение и агрегирование: join, merge, groupby для группового анализа. Выделение новых признаков: создание и преобразование колонок, кодирование категорий, генерация дат.

Тема 2.4. Нормализация и стандартизация данных

Нормализация: влияние масштабирования на методы анализа и обучения моделей. Способы нормализации: min-max, Z-преобразование, робастная масштабировка. Практические кейсы: когда и как правильно применять различные методы масштабирования.

Тема 2.5. Основы работы с R для обработки данных

Синтаксис R: особенности языка, базовые конструкции (векторы, факторы, фреймы данных). Популярные пакеты: dplyr, tidyr — операции фильтрации, группировки, преобразования. Интеграция с Python: обмен данными между Python и R, обзор RStudio, Jupyter R-kernel.

Тема 2.6. Работа с большими данными

Основные концепции Big Data: volume, velocity, variety, value (4V). Hadoop-экосистема: HDFS, MapReduce, YARN. Spark: резюме функционала Spark (RDD, DataFrame API), пример базовой обработки больших данных.

Содержание практических занятий

№ темы	Наименование (содержание) темы, по которой предусмотрено практическое занятие	Формы и методы проведения
1	Тема 2.1. Основы Python для обработки данных	Устный опрос, выполнение практических заданий, обмен опытом
2	Тема 2.2. Очистка данных	Устный опрос, выполнение практических заданий, обмен опытом
3	Тема 2.3. Преобразование данных	Устный опрос, выполнение практических заданий, анализ конкретных ситуаций, обмен опытом
4	Тема 2.4. Нормализация и стандартизация данных	Устный опрос, выполнение практических заданий, обмен опытом
5	Тема 2.5. Основы работы с R для обработки данных	Устный опрос, выполнение практических заданий, обмен опытом
6	Тема 2.6. Работа с большими данными	Устный опрос, выполнение практических заданий, обмен опытом

Содержание самостоятельной работы слушателей

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ п/п	Наименование (содержание) темы	Формы и методы проведения
1	Тема 2.1. Основы Python для обработки данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
2	Тема 2.2. Очистка данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
3	Тема 2.3. Преобразование данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
4	Тема 2.4. Нормализация и стандартизация данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
5	Тема 2.5. Основы работы с R для обработки данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
6	Тема 2.6. Работа с большими данными	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

Список литературы:

Основная литература:

1. Бринк, Ричардс, Феверолф: Машинное обучение. – СПб.: Питер, 2023. – 336 с.

2. Зараменских, Е. П. Основы бизнес-информатики : учебник и практикум для вузов / Е. П. Зараменских. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 470 с. — (Высшее образование). — ISBN 978-5-534-15039-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511961>.
3. Data Science. Наука о данных с нуля. / Билл Фрэнкс.; пер. с англ. Евстигнеева И.В. — М.: Издательство «Альпина Паблишер». — 2022. — 320 с.
4. Набатова, Д. С. Математические и инструментальные методы поддержки принятия решений: учебник и практикум для вузов / Д. С. Набатова. — Москва: Издательство Юрайт, 2025. — 292 с. — (Высшее образование). — ISBN 978-5-534-02699-3.
5. Курносоев Ю.В. «Азбука аналитики», Издательство «Концептуал», 2018 -240 с.
6. Б. Марр «Ключевые инструменты бизнес-аналитики»/ пер с англ. Егоров В. Н., Издательство «Лаборатория знаний», 2021 – 339 с.
7. де Прадо М. «Машинное обучение: алгоритмы для бизнеса», Санкт Петербург: Издательский дом «Питер». — 2019. — 432 с.
8. Плас вандер Д. «Python для сложных задач: наука о данных и машинное обучение», Санкт Петербург: Издательский дом «Питер». — 2023. — 576с.
9. Лакшманан В., Тайджани Д. «Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении», Санкт Петербург: Издательский дом «Питер». — 2021. — 496с.

Дополнительная литература:

1. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени. / Натан Марц и Джеймс Уоррен.; пер. с англ. — М.: Вильямс. — 2020. — 336 с.
2. Глубокое обучение на Python. — СПб.: Питер, 2022. — 400 с.: ил. — (Серия «Библиотека программиста»).
3. Основы Data Science и Big Data. Python и наука о данных. / Силен Д., Мейсман А., Али М.; пер. с англ. — Санкт Петербург: Издательский дом «Питер». — 2018. — 336 с.
4. Паклин Н.Б. Орешков В.И. Бизнес-аналитика: от данных к знаниям, СПб: Питер 2013. — 706 с.
5. Радченко И.А, Николаев И.Н. Технологии и инфраструктура Big Data. — СПб: Университет ИТМО, 2020. — 52 с.
6. Data Science. Наука о данных с нуля. / Джоэл Грас.; пер. с англ. — 2-е изд., перераб. — Санкт Петербург: Издательство «БХВ-Петербург», — 2021. — 416 с.
7. Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении. — СПб.: Питер, 2021. — 496 с.: ил.
8. К. Андерсон «Аналитическая культура»/Издательство. Манн, Иванов, Фербер, 2023 – 332 с.

9. Д. Битти К. Вигерс «Разработка требований к программному обеспечению», Издательство BHV, 2020 -736 с.

Интернет-ресурсы, информационно-справочные и поисковые системы

1. <https://www.anaconda.com/> – страница загрузки Anaconda.
2. <https://powerbi.microsoft.com/ru-ru/desktop/> – страница загрузки платформы Power BI Desktop компании Microsoft.
3. <https://loginom.ru/download> – страница загрузки платформы Loginom компании BaseGroup Labs.
4. <https://www.knime.com/knime-analytics-platform> – страница загрузки платформы Knime Analytics Platform.
5. <https://community.cloud.databricks.com/login.html> – страница регистрации Databricks Community Edition
6. <https://rapidminer.com/get-started/> – страница загрузки платформы RapidMiner.
7. <http://h2o-release.s3.amazonaws.com/h2o/rel-zermelo/4/index.html> – страница загрузки платформы H2O.
8. <https://cloud.google.com/bigquery/docs/sandbox> – страница BigQuery sandbox.
9. <https://loginom.ru/> - аналитическая low-code платформа, которая позволяет проводить анализ данных любого уровня сложности без программирования
10. <https://cloud.yandex.ru/services/datалens> – сервис визуализации и анализа данных Яндекс.

Рабочая программа Дисциплины 3. Интерпретация

Цель освоения дисциплины – приобретение слушателями компетенций для проведения статистического анализа, тестирования гипотез, а также применения методов машинного обучения (кластеризация, классификация, временные ряды) и корректной интерпретации результатов.

Компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения дисциплины:

- способность к сбору информации о бизнес-проблемах или бизнес-возможностях;
- способность к анализу, обоснованию и выбору решения;
- способность готовить данные для проведения аналитических работ по исследованию больших данных.

Планируемые результаты обучения

По итогам освоения программы слушатели должны:

Знать:

- ключевые статистические понятия и методы (среднее, дисперсия, корреляция, регрессия);
- принципы постановки и проверки гипотез;
- основы анализа временных рядов и методы прогнозирования.
- базовые алгоритмы машинного обучения (регрессия, кластеризация, классификация);
- методы оценки качества моделей (метрики, перекрёстная проверка);
- структуру и требования к интерпретации результатов.

Уметь:

- проводить статистический анализ данных и интерпретировать полученные статистические показатели;
- формулировать гипотезы для бизнес-задач и проверять их (A/B-тестирование и др.);
- применять методы прогнозирования на основе временных рядов (ARIMA, ETS и др.);
- выбирать соответствующие модели машинного обучения (кластеризация, классификация) и настраивать их гиперпараметры;
- оценивать качество обученных моделей и интерпретировать результаты;
- делать выводы на основе анализа и формировать рекомендации для дальнейших действий.

Владеть:

- навыками анализа взаимосвязей между переменными и выявления причинно-следственных связей;
- приёмами применения библиотек машинного обучения;
- методами корректной визуализации статистических результатов и параметров моделей;
- техниками прогнозирования временных рядов и оценки точности предсказаний;
- инструментами интерпретации результатов при работе с моделями ML;
- принципами описания ограничений моделей и рисков при принятии решений по результатам анализа.

Учебно-тематический план дисциплины

№№ п/п	Наименование дисциплины, модуля	Трудоемкость		В том числе				Форма контроля
				Контактная работа ⁶			Самостоятельная работа*	
				Всего часов	из них			
		Лекции	Практические занятия					
1	Дисциплина 3. Интерпретация	-	68	32	12	20	36	Зачет
2	Тема 3.1. Основы статистического анализа данных	-	10	4	2	2	6	Тестирование
3	Тема 3.2. Анализ взаимосвязей	-	10	4	2	2	6	Тестирование
4	Тема 3.3. Тестирование гипотез	-	10	4	2	2	6	Тестирование
5	Тема 3.4. Анализ временных рядов и прогнозирование	-	12	6	2	4	6	Тестирование
6	Тема 3.5. Методы кластеризации и классификации	-	12	6	2	4	6	Тестирование
7	Тема 3.6. Интерпретация и представление результатов	-	12	6	2	4	6	Тестирование

* С применением дистанционных образовательных технологий и электронного обучения (при наличии).

	анализа							
8	Промежуточная аттестация	-	2	2	-	2		Зачет

Содержание дисциплины

Тема 3.1. Основы статистического анализа данных

Типы статистических показателей: описательные (среднее, медиана, мода), дисперсионные (дисперсия, стандартное отклонение). Корреляционный анализ: коэффициенты (Пирсона, Спирмена), интерпретация результатов. Распределения: нормальное, биномиальное, Пуассона — в чём разница, когда применяются.

Тема 3.2. Анализ взаимосвязей

Парные и множественные взаимосвязи: ковариация, многомерный анализ. Методы выявления зависимостей: тепловые карты корреляций, парные диаграммы рассеяния. Качественные переменные: методы анализа (ANOVA, хи-квадрат).

Тема 3.3. Тестирование гипотез

Основные принципы: нулевая гипотеза, уровень значимости, мощность теста. А/В-тестирование: постановка, расчёт выборки, анализ результатов (p-value, доверительные интервалы). Ошибки I и II рода: как интерпретировать результаты тестов и что учитывать при принятии решений.

Тема 3.4. Анализ временных рядов и прогнозирование

Типовые компоненты временного ряда: тренд, сезонность, цикличность, шум. Модели прогнозирования: ARIMA, SARIMA, ETS — назначение, настройка параметров. Практическое применение: кейсы прогнозирования продаж, трафика, нагрузки на систему.

Тема 3.5. Методы кластеризации и классификации

Кластеризация: k-means, иерархическая кластеризация, DBSCAN — особенности и сферы применения. Классификация: логистическая регрессия, деревья решений, случайный лес, k-NN. Отбор признаков: влияние многообразия признаков на качество моделей, регуляризация. Нейросетевые возможности. Базовые концепции нейросетей (полносвязанные сети, функции активации, обучение и валидация).

Тема 3.6. Интерпретация и представление результатов анализа

Анализ результатов: как оценивать метрики (accuracy, precision, recall, F1, ROC-AUC). Формирование выводов: структура отчёта, формулировка гипотез, комментарии к цифрам. Применение результатов: рекомендации для бизнеса, ограничения анализа, возможные риски и допущения.

Содержание практических занятий

№ темы	Наименование (содержание) темы, по которой предусмотрено практическое занятие	Формы и методы проведения
1	Тема 3.1. Основы статистического анализа данных	Устный опрос, выполнение практических заданий, обмен опытом
2	Тема 3.2. Анализ взаимосвязей	Устный опрос, выполнение практических заданий, обмен опытом
3	Тема 3.3. Тестирование гипотез	Устный опрос, выполнение практических заданий, анализ конкретных ситуаций, обмен опытом
4	Тема 3.4. Анализ временных рядов и прогнозирование	Устный опрос, выполнение практических заданий, обмен опытом
5	Тема 3.5. Методы кластеризации и классификации	Устный опрос, выполнение практических заданий, обмен опытом
6	Тема 3.6. Интерпретация и представление результатов анализа	Устный опрос, выполнение практических заданий, обмен опытом

Содержание самостоятельной работы слушателей

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ п/п	Наименование (содержание) темы	Формы и методы проведения
1	Тема 3.1. Основы статистического анализа данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
2	Тема 3.2. Анализ взаимосвязей	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
3	Тема 3.3. Тестирование гипотез	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
4	Тема 3.4. Анализ временных рядов и прогнозирование	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
5	Тема 3.5. Методы кластеризации и классификации	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

6	Тема 3.6. Интерпретация и представление результатов анализа	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
---	---	---

Список литературы:

Основная литература:

1. Бринк, Ричардс, Феверолф: Машинное обучение. – СПб.: Питер, 2023. – 336 с.
2. Зараменских, Е. П. Основы бизнес-информатики : учебник и практикум для вузов / Е. П. Зараменских. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 470 с. — (Высшее образование). — ISBN 978-5-534-15039-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511961>.
3. Data Science. Наука о данных с нуля. / Билл Фрэнкс.; пер. с англ. Евстигнеева И.В. – М.: Издательство «Альпина Паблишер». – 2022. – 320 с.
4. Набатова, Д. С. Математические и инструментальные методы поддержки принятия решений: учебник и практикум для вузов / Д. С. Набатова. — Москва: Издательство Юрайт, 2025. — 292 с. — (Высшее образование). — ISBN 978-5-534-02699-3.
5. Курносоев Ю.В. «Азбука аналитики», Издательство «Концептуал», 2018 -240 с.
6. Б. Марр «Ключевые инструменты бизнес-аналитики»/ пер с англ. Егоров В. Н., Издательство «Лаборатория знаний», 2021 – 339 с.
7. де Прадо М. «Машинное обучение: алгоритмы для бизнеса», Санкт Петербург: Издательский дом «Питер». – 2019. – 432 с.
8. Плас вандер Д. «Python для сложных задач: наука о данных и машинное обучение», Санкт Петербург: Издательский дом «Питер». – 2023. – 576с.
9. Лакшманан В., Тайджани Д. «Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении», Санкт Петербург: Издательский дом «Питер». – 2021. – 496с.

Дополнительная литература:

1. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени. / Натан Марц и Джеймс Уоррен.; пер. с англ. – М.: Вильямс. – 2020. – 336 с.
2. Глубокое обучение на Python. – СПб.: Питер, 2022. – 400 с.: ил. – (Серия «Библиотека программиста»).
3. Основы Data Science и Big Data. Python и наука о данных. / Силен Д., Мейсман А., Али М.; пер. с англ. – Санкт Петербург: Издательский дом «Питер». – 2018. – 336 с.
4. Паклин Н.Б. Орешков В.И. Бизнес-аналитика: от данных к знаниям, СПб: Питер 2013. – 706 с.

5. Радченко И.А, Николаев И.Н. Технологии и инфраструктура Big Data. – СПб: Университет ИТМО, 2020. – 52 с.
6. Data Science. Наука о данных с нуля. / Джоэл Грас.; пер. с англ. – 2-е изд., перераб. – Санкт Петербург: Издательство «БХВ-Петербург». – 2021. – 416 с.
7. Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении. – СПб.: Питер, 2021. – 496 с.: ил.
8. К. Андерсон «Аналитическая культура»/Издательство. Манн, Иванов, Фербер, 2023 – 332 с.
9. Д. Битти К. Вигерс «Разработка требований к программному обеспечению», Издательство BHV, 2020 -736 с.

Интернет-ресурсы, информационно-справочные и поисковые системы

1. <https://www.anaconda.com/> – страница загрузки Anaconda.
2. <https://powerbi.microsoft.com/ru-ru/desktop/> – страница загрузки платформы Power BI Desktop компании Microsoft.
3. <https://loginom.ru/download> – страница загрузки платформы Loginom компании BaseGroup Labs.
4. <https://www.knime.com/knime-analytics-platform> – страница загрузки платформы Knime Analytics Platform.
5. <https://community.cloud.databricks.com/login.html> – страница регистрации Databricks Community Edition
6. <https://rapidminer.com/get-started/> – страница загрузки платформы RapidMiner.
7. <http://h2o-release.s3.amazonaws.com/h2o/rel-zermelo/4/index.html> – страница загрузки платформы H2O.
8. <https://cloud.google.com/bigquery/docs/sandbox> – страница BigQuery sandbox.
9. <https://loginom.ru/> - аналитическая low-code платформа, которая позволяет проводить анализ данных любого уровня сложности без программирования
10. <https://cloud.yandex.ru/services/datalens> – сервис визуализации и анализа данных Яндексе.

Рабочая программа Дисциплины 4. Построение отчетов

Цель освоения дисциплины – приобретения слушателями компетенций для наглядной визуализации и эффективной презентации результатов анализа данных с помощью инструментов Python, R и BI-платформ, а также навыков создания дашбордов и отчётов для разных заинтересованных сторон.

Компетенции, совершенствуемые и приобретаемые слушателями в процессе освоения дисциплины:

- способность к сбору информации о бизнес-проблемах или бизнес-возможностях;
- способность к анализу, обоснованию и выбору решения;
- способность готовить данные для проведения аналитических работ по исследованию больших данных;
- способность к проведению аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика.

Планируемые результаты обучения

По итогам освоения программы слушатели должны:

Знать:

- основные принципы и правила визуализации данных;
- возможности Python и R для визуализации;
- функциональные особенности BI-платформ;
- структуру и логику построения аналитических отчётов и презентаций;
- методы сторителлинга для донесения ключевых инсайтов;
- этические и юридические аспекты работы с данными.

Уметь:

- выбирать соответствующий тип диаграммы/визуализации под конкретную аналитическую задачу;
- создавать интерактивные дашборды в BI-инструментах;
- использовать Python или R для формирования наглядных графиков и диаграмм с анимацией/взаимодействием;
- структурировать отчёт в соответствии с интересами и уровнем компетенции целевой аудитории;
- презентовать результаты анализа, формулируя чёткие выводы и рекомендации для руководства и стейкхолдеров.

Владеть:

- практическими техниками разработки многопользовательских дашбордов и их интеграцией в корпоративную среду;
- приёмами дизайна визуализации (цветовые схемы, фокус на ключевых данных, упрощение графиков);
- навыками построения увлекательного и убедительного сторителлинга вокруг данных;
- средствами подготовки презентаций с данными, автоматически обновляемыми из BI;
- методами документирования и архивирования отчётов, включающими версии и историю изменений.

Учебно-тематический план дисциплины

№№ п/п	Наименование дисциплины, модуля	Трудоемкость		В том числе				Форма контроля
				Контактная работа ⁷			Самостоятельная работа ⁸	
				Всего часов	из них			
		Лекции	Практические занятия					
		В зачетных единицах	В часах					
1	Дисциплина 4. Построение отчетов	-	44	22	6	16	22	Зачет
2	Тема 4.1. Первичная визуализация данных	-	10	4	2	2	6	Тестирование
3	Тема 4.2. Визуализация данных в Python и R	-	12	6	2	4	6	Тестирование
4	Тема 4.3. Визуализация данных в BI-инструментах	-	12	6	2	4	6	Тестирование
5	Тема 4.4. Презентация аналитических результатов для руководства и заинтересованных сторон	-	8	4	-	4	4	Тестирование
6	Промежуточная аттестация	-	2	2		2		Зачет

Содержание дисциплины

⁷ С применением дистанционных образовательных технологий и электронного обучения (при наличии).

Тема 4.1. Первичная визуализация данных

Базовые типы графиков: столбчатые, линейные, круговые, диаграммы рассеяния. Основы визуальной грамотности: выбор цвета, масштаб, заголовки, подписи осей, легенда. Практика на простых примерах: как быстро визуализировать выборку в Excel/Google Sheets/Python.

Тема 4.2. Визуализация данных в Python и R

Инструменты Python: Matplotlib, Seaborn, Plotly — примеры построения интерактивных графиков. Графическая система R: ggplot2 — грамотная работа со слоями, темами и эстетиками. Сравнительный анализ: когда удобнее Python, когда R, примеры совместного использования.

Тема 4.3. Визуализация данных в BI-инструментах

Обзор популярных платформ: Loginom, Yandex DataLens, Power BI и другие. Подготовка и подключение данных: импорт из файлов, баз данных, облачных сервисов. Создание дашбордов: фильтры, слайсеры, интерактивные элементы, методы сторителлинга в BI.

Тема 4.4. Презентация аналитических результатов для руководства и заинтересованных сторон

Принципы эффективной презентации: структура доклада, акценты, простота восприятия. Фокус на бизнес-ценности: как объяснить технические детали «на языке» управленцев. Инструменты и форматы: слайды (PowerPoint, Google Slides), PDF-отчёты, интерактивные демо и т.д.

Содержание практических занятий

№ темы	Наименование (содержание) темы, по которой предусмотрено практическое занятие	Формы и методы проведения
1	Тема 4.1. Первичная визуализация данных	Устный опрос, выполнение практических заданий, обмен опытом
2	Тема 4.2. Визуализация данных в Python и R	Устный опрос, выполнение практических заданий, обмен опытом
3	Тема 4.3. Визуализация данных в BI-инструментах	Устный опрос, выполнение практических заданий, анализ коваретных ситуаций, обмен опытом
4	Тема 4.4. Презентация аналитических результатов для руководства и заинтересованных сторон	Устный опрос, выполнение практических заданий, обмен опытом

Содержание самостоятельной работы слушателей

Основная цель самостоятельной работы слушателей – закрепление знаний, полученных в ходе лекционных и практических занятий.

Индивидуальная консультационная работа преподавателей со слушателями осуществляется весь период обучения.

№ п/п	Наименование (содержание) темы	Формы и методы проведения
1	Тема 4.1. Первичная визуализация данных	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
2	Тема 4.2. Визуализация данных в Python и R	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
3	Тема 4.3. Визуализация данных в BI-инструментах	Изучение основной и дополнительной литературы по программе; выполнение практических заданий
4	Тема 4.4. Презентация аналитических результатов для руководства и заинтересованных сторон	Изучение основной и дополнительной литературы по программе; выполнение практических заданий

Список литературы:

Основная литература:

10. Бринк, Ричардс, Феверолф: Машинное обучение. – СПб.: Питер, 2023. – 336 с.
11. Зараменских, Е. П. Основы бизнес-информатики : учебник и практикум для вузов / Е. П. Зараменских. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 470 с. — (Высшее образование). — ISBN 978-5-534-15039-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511961>.
12. Data Science. Наука о данных с нуля. / Билл Фрэнкс.; пер. с англ. Евстигнеева И.В. – М.: Издательство «Альпина Паблишер». – 2022. – 320 с.
13. Набатова, Д. С. Математические и инструментальные методы поддержки принятия решений: учебник и практикум для вузов / Д. С. Набатова. — Москва: Издательство Юрайт, 2025. — 292 с. — (Высшее образование). — ISBN 978-5-534-02699-3.
14. Курносов Ю.В. «Азбука аналитики», Издательство «Концептуал», 2018 -240 с.
15. Б. Марр «Ключевые инструменты бизнес-аналитики»/ пер с англ. Егоров В. Н., Издательство «Лаборатория знаний», 2021 – 339 с.
16. де Прадо М. «Машинное обучение: алгоритмы для бизнеса», Санкт Петербург: Издательский дом «Питер». – 2019. – 432 с.
17. Плас вандер Д. «Python для сложных задач: наука о данных и машинное обучение», Санкт Петербург: Издательский дом «Питер». – 2023. – 576с.
18. Лакшманан В., Тайджани Д. «Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении», Санкт Петербург: Издательский дом «Питер». – 2021. – 496с.

Дополнительная литература:

1. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени. / Натан Марц и Джеймс Уоррен.; пер. с англ. – М.: Вильямс. – 2020. – 336 с.
2. Глубокое обучение на Python. – СПб.: Питер, 2022. – 400 с.: ил. – (Серия «Библиотека программиста»).
3. Основы Data Science и Big Data. Python и наука о данных. / Силен Д., Мейсман А., Али М.; пер. с англ. – Санкт Петербург: Издательский дом «Питер». – 2018. – 336 с.
4. Паклин Н.Б. Орешков В.И. Бизнес-аналитика: от данных к знаниям, СПб: Питер 2013. – 706 с.
5. Радченко И.А, Николаев И.Н. Технологии и инфраструктура Big Data. – СПб: Университет ИТМО, 2020. – 52 с.
6. Data Science. Наука о данных с нуля. / Джоэл Грас.; пер. с англ. – 2-е изд., перераб. – Санкт Петербург: Издательство «БХВ-Петербург». – 2021. – 416 с.
7. Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении. – СПб.: Питер, 2021. – 496 с.: ил.
8. К. Андерсон «Аналитическая культура»/Издательство. Манн, Иванов, Фербер, 2023 – 332 с.
9. Д. Битти К. Вигерс «Разработка требований к программному обеспечению», Издательство BHV, 2020 -736 с.

Интернет-ресурсы, информационно-справочные и поисковые системы

1. <https://www.anaconda.com/> – страница загрузки Anaconda.
2. <https://powerbi.microsoft.com/ru-ru/desktop/> – страница загрузки платформы Power BI Desktop компании Microsoft.
3. <https://loginom.ru/download> – страница загрузки платформы Loginom компании BaseGroup Labs.
4. <https://www.knime.com/knime-analytics-platform> – страница загрузки платформы Knime Analytics Platform.
5. <https://community.cloud.databricks.com/login.html> – страница регистрации Databricks Community Edition
6. <https://rapidminer.com/get-started/> – страница загрузки платформы RapidMiner.
7. <http://h2o-release.s3.amazonaws.com/h2o/rel-zermelo/4/index.html> – страница загрузки платформы H2O.
8. <https://cloud.google.com/bigquery/docs/sandbox> – страница BigQuery sandbox.
9. <https://loginom.ru/> - аналитическая low-code платформа, которая позволяет проводить анализ данных любого уровня сложности без программирования

10. <https://cloud.yandex.ru/services/datalens> – сервис визуализации и анализа данных Яндекса.

**Организационно-педагогические условия
реализации программы профессиональной переподготовки**

**1. Материально-технические условия, необходимые для
осуществления образовательного процесса**

Наименование специализированных учебных помещений	Вид занятий	Наименование оборудования, программного обеспечения
Учебный интерактивный класс	Лекции Практические занятия	Мультимедийное оборудование, компьютеры. Компьютер, подключенный к сети Интернет, интернет-браузер. Прикладные программы для просмотра текстовых и видеоматериалов.

Материально-технические условия соответствуют действующим санитарным и противопожарным правилам и нормам.

***Примечание.** В случае проведения учебных занятий с применением электронного обучения (ЭО) и дистанционных образовательных технологий (ДОТ) у слушателя должен быть персональный компьютер, оснащенный аудиоколонками, с доступом в сеть интернет и установленным видеоплеером, способным воспроизводить видеофайлы.*

2. Перечень информационных технологий и учебно-методическим условий, используемых при осуществлении образовательного процесса

При проведении занятий с применением ЭО и ДОТ проведение вебинаров для слушателей осуществляется в удаленном доступе. Преподавателями используются компьютерные презентации, работу в чате, индивидуальное консультирование слушателей.

Условия для функционирования электронной информационно-образовательной среды

Электронные информационные ресурсы	Вид занятий	Наименование оборудования, программного обеспечения
Система дистанционного обучения, система видеоконференцсвязи	Вебинар Итоговая аттестация	Компьютер, подключенный к сети Интернет; интернет-браузер; Прикладные программы для просмотра текстовых и видеоматериалов

3. Организация образовательного процесса

В образовательном процессе используются разнообразные формы работы со слушателями.

- лекция (видеолекция) с мультимедийным сопровождением по наиболее сложным вопросам программы;
- лекция-вебинар с использованием современных технических средств обучения;

- практические занятия и самостоятельная работа с использованием современных технических средств обучения;
 - кейс-стади (в том числе видео-кейсы)– изучение конкретных ситуаций из практики (casestudy), для выполнения данного вида заданий обучающимся должна быть представлена в письменной форме информация относительно реальной ситуации (профессиональной или жизненной) и поставлены конкретные задачи её изучения проблемы, обучающиеся анализируют различные аспекты проблемы и предлагают выработанные решения;
 - тестирование метод оценки знаний, умений, навыков обучающихся и др.
- Обучение проводится, в том числе с использованием ЭО и ДОТ, реализуемых посредством информационно-телекоммуникационных сетей при опосредованном взаимодействии слушателей и педагогических работников.

В процессе обучения слушатели обеспечиваются необходимыми для эффективного прохождения обучения учебно-методическими материалами и информационными ресурсами в объеме изучаемого курса, которые могут быть объединены в учебно-методический комплекс. Материалы учебно-методического комплекса доводятся до всех слушателей курса.

Итоговая аттестация проводится на образовательном портале Финансового университета посредством информационно-телекоммуникационных сетей.

4. Кадровое обеспечение образовательного процесса

Учебный процесс со слушателями обеспечивают квалифицированные сотрудники Финансового университета, а также приглашенные специалисты и действующие практики других организаций.

Описание системы оценки качества освоения программы

В систему оценки качества освоения программы «Аналитик данных» входят:

- текущий контроль;
- промежуточная аттестация по каждой дисциплине;
- итоговая аттестация.

1. Текущий контроль успеваемости реализуется в ходе проведения практических занятий в форме устного опроса, обмена опытом работы, выступлений слушателей по узловым вопросам программы, путем выполнения практических заданий, разбора конкретных ситуаций, тестирования.

2. Промежуточные аттестации – зачет в форме тестирования.

Примеры тестового вопроса для промежуточной аттестации:

Дисциплина 1. Сбор данных

1. Что из перечисленного относится к реляционным источникам данных?

- **Таблицы в базе данных, связанные ключами**
- JSON-файлы с данными
- Неструктурированные тексты статей
- Логи серверов приложений

2. Какое определение лучше всего описывает процесс ETL?

- **Extract, Transform, Load — извлечение, преобразование и загрузка данных из разных источников**

- Every Table Loaded — загрузка всех таблиц базы данных для анализа
- Extended Tabular Layout — расширенная табличная структура для визуализации
- Exclude, Transform, Link — исключение лишних данных, их трансформация и связывание

3. Какие данные относятся к неструктурированным?

- **Тексты сообщений в социальных сетях**
- Таблица Excel с продажами по датам
- Таблица MySQL с транзакциями
- CSV-файл с клиентской базой

4. Какой оператор SQL используется для выбора нужных строк по условию?

- **WHERE**
- GROUP BY
- ORDER BY
- JOIN

Дисциплина 2. Обработка данных

1. Какой из перечисленных инструментов не является библиотекой Python для обработки данных?

- **ggplot2**
- pandas
- NumPy
- SciPy

2. Что такое пропущенные значения (NaN)?

- **Отсутствующие или некорректные данные в наборе данных**
- Обязательная колонка для индексирования DataFrame
- Признак, который всегда равен нулю
- Специальный тип файла для обмена с сервером

3. В чём суть нормализации (MinMaxScaler) при работе с данными?

- **Приведение признаков к единому диапазону (обычно от 0 до 1)**
- Удаление всех пропущенных значений
- Оптимизация времени выборки в SQL-запросах
- Поиск выбросов в временном ряду

4. Какая платформа чаще всего упоминается в контексте работы с большими данными (Big Data)?

- **Hadoop**
- Excel
- PowerPoint
- Google Docs

Дисциплина 3. Интерпретация

1. Как называется мера тесноты линейной связи между двумя переменными?

- **Коэффициент корреляции**
- Коэффициент детерминации
- Медиана
- Мода

2. Что такое p-value в тестировании гипотез?

- **Вероятность получить результат не менее экстремальный, чем наблюдаемый, при условии справедливости нулевой гипотезы**
- Уровень доверия к исследованию (обычно 95%)
- Среднее отклонение результатов эксперимента
- Величина, показывающая силу эффекта при любых условиях

3. Что из перечисленного относится к методам кластеризации?

- **k-means**
- Линейная регрессия

- Логистическая регрессия
- Решающее дерево (Decision Tree)

4. Для чего используют модель ARIMA?

- **Прогнозирование временных рядов**
- Сегментацию клиентов на кластеры
- Анализ корреляции двух признаков
- Автоматическую визуализацию данных

Дисциплина 4. Построение отчётов

1. Какой принцип лучше всего отражает идею «Data Storytelling»?

- **Структурированная подача аналитических результатов в форме наглядного повествования**

- Обязательное использование 3D-графиков в отчётах
- Применение только цифровых метрик и статистических тестов без визуализации
- Скрытие промежуточных результатов ради лаконичности презентации

2. Какая библиотека в Python НЕ предназначена для визуализации данных?

- **requests**
- matplotlib
- seaborn
- plotly

3. Какое основное назначение BI-инструментов?

- **Создание интерактивных дашбордов и отчётов для принятия решений**

- Запуск SQL-запросов на сервере
- Тестирование статистических гипотез без кода
- Автоматическое написание скриптов на Python

4. При подготовке презентации аналитических результатов для руководства, что следует сделать в первую очередь?

- **Определить цели и целевую аудиторию, чтобы правильно подобрать язык и акценты**

- Использовать только максимально детальную информацию о каждом шаге анализа

- Приводить весь программный код, чтобы продемонстрировать навыки разработки
- Исключить все визуальные элементы и ограничиться текстовым отчётом

Порядок проведения: тестирование проводится с личного компьютера, 20 тестовых вопросов по каждой дисциплине, 60 мин, количество попыток – 3 по каждой дисциплине.

Для получения зачета по дисциплине необходимо правильно ответить не менее чем на 12 тестовых вопросов в любой попытке.

3. Итоговая аттестация

Итоговая аттестация представляет собой защиту итогового практического задания.

Итоговая аттестация имеет целью определить сформированность спланированных к освоению профессиональных компетенций.

Проведение итоговой аттестации. Итоговая аттестация проводится в форме защиты итогового практического задания.

Содержание итогового практического задания:

План итогового практического задания:

1. Описание проекта и набора данных

1.1. Цель и постановка задачи

- Кратко сформулируйте бизнес-задачу или исследовательскую цель, которую нужно решить с помощью анализа данных (например, прогноз продаж, сегментация клиентов, кластеризация товаров).

1.2. Описание набора данных

- Источник данных (внутренняя база, CSV-файл, API, открытые источники).
- Формат и объём данных (количество строк, столбцов, основные типы: числовые, категориальные, временные).
- Дополнительные данные (при наличии) из внешних источников для более богатого анализа (например, региональные статистические показатели).

1.3. Выделение целевой переменной (при наличии)

- Если задача предполагает прогноз (регрессия) или классификацию, укажите, какая переменная будет предсказываться.
- Если задача — кластеризация или анализ взаимосвязей, то целевая переменная может быть не нужна, но следует описать, какие результаты хотите получить (количество кластеров, выявление закономерностей и т.д.).

2. Сбор и интеграция данных

2.1. Источники и методы сбора

- Подключение к базе данных (SQL-запросы для выборки).
- Использование API (если применимо).
- Загрузка и парсинг CSV/JSON/XML.
- Краткое описание ETL-процесса (Extract, Transform, Load).

2.2. Интеграция из разных форматов

- Объединение данных из SQL-таблиц и файлов (CSV, Excel, JSON).
- Проблемы совместимости (разные схемы, форматы дат, кодировки) и способы их решения.

3. Предварительный анализ и подготовка данных

3.1. Разведочный анализ данных (EDA)

- Анализ структуры датасета: просмотр первых строк, типов столбцов, статистик (count, mean, std и т.д.).
- Визуальный обзор (гистограммы, диаграммы рассеяния, «ящики с усами» для выявления выбросов).
- Изучение категориальных переменных (частоты, распределения).

3.2. Очистка данных

- Работа с пропущенными значениями (удаление, заполнение средними, медианами).

- Обнаружение и корректировка выбросов.
- Проверка несбалансированных классов (если есть задача классификации) и возможные методы балансировки (Oversampling, Undersampling).

3.3. Преобразование данных

- Нормализация и стандартизация признаков (MinMaxScaler, StandardScaler).
- Кодирование категориальных признаков (One-Hot Encoding, Label Encoding).
- Слияние и группировка данных (JOIN, groupby).

4. Моделирование и анализ

4.1. Выбор подхода к анализу

- Классификация, регрессия, кластеризация, прогнозирование временных рядов и т.д. — в зависимости от цели проекта.
- Краткое обоснование выбора методов и алгоритмов (например, Random Forest, Gradient Boosting, k-means, ARIMA).

4.2. Проведение анализа / обучения моделей

- Подготовка обучающей и тестовой выборок (train/test split).
- Обучение и оценка нескольких моделей (Scikit-learn, statsmodels, R-библиотеки и т.д.).
- Если релевантно, проведение тестирования гипотез (A/B-тест, статистические проверки).

4.3. Сравнение результатов и интерпретация

- Подсчёт метрик (accuracy, precision/recall, RMSE, MAPE, silhouette score и др. в зависимости от задачи).
- Сравнение нескольких моделей / подходов, выбор лучшего решения.
- Интерпретация полученных закономерностей (какие факторы наиболее значимы, какие выводы можно сделать для бизнеса/проекта).

5. Визуализация и презентация

5.1. Подготовка отчётов

- Представление результатов анализа в виде таблиц, диаграмм, сюжетных графиков.
- Визуальное сравнение различных моделей и сценариев.

5.2. BI-инструменты и дашборды

- Использование Loginom, Yandex DataLens, Power BI и другие для построения интерактивных дашбордов.
- Настройка фильтров, срезов, интерактивных элементов (поиск, панель выбора дат).

5.3. Выводы и рекомендации

- Формирование ключевых инсайтов, понятных стейкхолдерам.
- Рекомендации по улучшению показателей (оптимизация маркетинговых кампаний, формирование сегментов клиентов, прогноз продаж).
- Ограничения анализа (недостаток данных, предположения при моделировании и т.д.).

Итоговое практическое задание состоит из выбора индивидуального кейса (набор данных) и выполнения заданий, которые охватывают все практические методы, подходы в соответствующих программных продуктах, применяемые в анализе больших данных и машинном обучении рассмотренные и изученные в соответствующих темах в рамках учебной программы.

В результирующую оценку по итоговому практическому заданию входит результат/продукт, полученный в ходе выполнения работы.

Итоговая аттестация проводится итоговой аттестационной комиссией, в состав которой могут входить преподаватели Финансового университета, приглашенные ведущие специалисты и представители работодателя в профильной сфере.

Порядок проведения итоговой аттестации: обучающиеся выполняют итоговое практическое задание на основе материалов по предложенному примеру. Выполненное итоговое практическое задание в соответствии с установленными сроками, загружается обучающимися на платформу и проверяется преподавателем. Преподаватели оценивают задания в системе и

предоставляют обратную связь по качеству выполнения задания, соответствию требованиям, дают рекомендации по улучшению работы и исправлению ошибок (при наличии).

Защита итогового практического задания является обязательной и может быть выполнена в одном из двух вариантов:

- синхронно (демонстрация презентации с комментариями, ответы на вопросы);
- асинхронно (прикрепление скринкаста, записи демонстрации презентации с комментариями, выполненную в любом редакторе, в электронную систему курса).

Вопросы и обратная связь от аттестационной комиссии может быть реализована также асинхронно.

Оценка выставляется по 4 балльной шкале: "неудовлетворительно"; "удовлетворительно", "хорошо", "отлично".

Итоговое практическое задание оформляется в текстовый файл Microsoft Word (.doc, .docx) или Adobe Acrobat (.pdf).

Критерии оценивания: для выставления оценки по итоговой аттестации необходимо пользоваться следующими критериями, приведенными в таблице.

№	Критерии оценки итоговой практической работы	Балл
1	Наличие таблиц (или датафреймов) «объекты-признаки». Строки соответствуют сущностям (объектам, примерам), столбцы — их атрибутам (признакам).	3
2	Наличие EDA (разведочного анализа данных). Выбор наиболее значимых переменных, обнаружение выбросов и аномалий, проверка основных гипотез, разработка первоначальных моделей.	3
3	Наличие визуализации данных в EDA. Диаграммы, графики, тепловые карты, соответствующие типу анализируемых данных; оценка качества исполнения визуализации.	3
4	Наличие моделей машинного обучения. Обучение как минимум одной модели (в любом аналитическом инструменте) с оценкой точности, сложности, интерпретируемости и времени предсказания.	4
5	Наличие сравнения моделей машинного обучения. Анализ результатов работы нескольких	5

	алгоритмов/библиотек, выбор более точной модели с обоснованием.	
6	Наличие методов предсказательной аналитики. Прогнозирование поведения объектов или процессов (например, временные ряды, классификация будущих событий), оценка качества прогноза.	2
7	Итого	20
Перевод баллов в оценку		≤ 10 баллов: «Неудовлетворительно» 10–11 баллов: «Удовлетворительно» 12–15 баллов: «Хорошо» 16–20 баллов: «Отлично»

Слушателям, которые успешно прошли итоговую аттестацию выдается диплом о переподготовке Финансового университета при Правительстве Российской Федерации.

Обсуждено и одобрено на заседании Научно-методического совета Института развития профессиональных компетенций и квалификаций, протокол № 28 от 20 декабря 2024 г.

	алгоритмов/библиотек, выбор более точной модели с обоснованием.	
6	Наличие методов предсказательной аналитики. Прогнозирование поведения объектов или процессов (например, временные ряды, классификация будущих событий), оценка качества прогноза.	2
7	Итого	20
Перевод баллов в оценку		≤ 9 баллов: «Неудовлетворительно» 10–11 баллов: «Удовлетворительно» 12–15 баллов: «Хорошо» 16–20 баллов: «Отлично»

Слушателям, которые успешно прошли итоговую аттестацию выдается диплом о переподготовке Финансового университета при Правительстве Российской Федерации.

Обсуждено и одобрено на заседании Научно-методического совета Института развития профессиональных компетенций и квалификаций, протокол № 28 от 20 декабря 2024 г.





Директор ИРПКиК



Т.А. Болтенко



**Лист информации о разработке и актуализации
программы профессиональной переподготовки
Аналитик данных, 256 ак.ч.**

№ п/п	Разработка/ актуализация программы	Дата и номер протокола рассмотрения ДПП на НМС СП/ совете департамента/ Ученом совете филиала	Дата и номер протокола рассмотрения ДПП на Экспертном совете по качеству ДО (при наличии)	Дата и номер протокола утверждения ДПП на Ученом совете институтков и школ ДПО	Подпись руководителя структурного подразделения, реализующего программу ДПО
1	2	3	4	5	6
1	Разработка	от 18.02.2021 № 8		от 15.03.2021 № 60	
2	Актуализация программы	от 01.06.2022 № 15		от 21.06.2022 № 14	
3	Актуализация программы	№ 16 от 01.03.2023 г.		от 21.03.2023 № 24	
4	Актуализация программы	№ 20 от 25.01.2024 г.		от 30.01.2024 № 33	
5	Актуализация программы	№ 20 от 28.12.2024 г.			