



Программа курса «Junior data analyst: младший аналитик данных»

Номер	Название темы	Количество часов	Описание темы
1	Введение в Big Data и Data science. Data science в различных секторах экономики	1	Основные термины. Типы задач, которые решаются с помощью Data Science. Обоснованность применения.
2	Данные и источники, характеристики, корреляция. Структурированные, полуструктурированные и неструктурированные данные	2	Данные и источники, характеристики, корреляция. Структурированные, полуструктурированные и неструктурированные данные.
3	Введение в статистику. Закон больших чисел. Нормальное распределение. Проверка гипотез	1,5	Определение статистической науки. Закон больших чисел, нормальное распределение и его роль в описании и анализе данных. Понятие нулевой гипотезы и проверка нулевой гипотезы, ошибки первого и второго рода.
4	Типы данных	1	Основные типы данных, которые применяются в data science, программировании и базах данных: целочисленные, логические, строки, числа с плавающей точкой, дата и время, timestamp.
5	Введение в Python, среды исполнения (IDE). Типы данных в Python	2	Основные среды и продукты: PyCharm, Anaconda, Google Colaboratory. Язык программирования Python как стандарт для работы с большими данными: основные функции, типы данных в Python.
6	Базовые конструкции и структуры	2	Основные конструкции, списки, словари, множества, кортежи, стеки, очереди, логические условия, арифметические операции.
7	Циклы и условия. Функции и классы	3	Синтаксис построения циклов, условных операторов, пользовательских функций, классов, наследование и полиморфизм.
8	Библиотеки NumPy, SciPy для научных вычислений	3	Библиотека NumPy, типы данных, массивы и операции с ними, Матричные операции, семплирование, чтение файлов. Библиотека SciPy. Научные вычисления. Практическое применение. Работа с данными. Тестирование данных.
9	Библиотека Pandas, как стандарт исследования данных	5	Библиотека Pandas, загрузка и запись данных, срезы данных, мульти индексация, group by, datetime, статистические функции, стандартная визуализация. Преимущества и недостатки библиотеки Pandas.

10	Визуализация в Python и срезы данных. Библиотеки Matplotlib, seaborn, plotly	4	Библиотека Matplotlib, варианты отрисовки графиков и изображений. Библиотека Seaborn, визуализация парных взаимосвязей, heatmap, диаграммы, тепловая карта корреляции. Библиотека Plotly. Продвинутое визуализация. Динамические графики.
11	Поиск бизнес решений и анализ датасетов	4	Общий подход к описательному анализу данных. Выявление скрытых закономерностей и корреляций для принятия бизнес-решений.
12	Системы хранения данных. Базы данных	1,5	Основные подходы к хранению данных: горизонтально и вертикально масштабируемые системы хранения данных, реляционные и NoSQL. Key-Value хранилища.
13	Реляционные базы данных и их моделирование	2,5	Реляционные базы данных для хранения структурированных данных. Основные определения и признаки реляционных баз данных. Понятие и создание моделей (схем) баз данных и установление связей один-ко-многим (one2many), многие-ко-многим (many2many), многие-к-одному (many2one).
14	Функционал PGAdmin для PostgreSQL	2	Процесс установки реляционных баз данных PostgreSQL. Интерфейс оболочки PGAdmin, основной функционал PGAdmin для удобной навигации по реляционным базам данных. Восстановление и развертывание базы данных из файла.
15	Язык структурированных запросов SQL. Синтаксис	4,5	Язык структурированных запросов SQL. Общий подход. Базовая концепция, фильтрация, функции. Функционал PgAdmin. Подзапросы, Join, Табличные операции.
16	Обращение к базе данных PostgreSQL с помощью языка SQL	3	Задачи создания новых таблиц, внешних и внутренних ключей их связей. Формирование аналитических запросов в реляционной базе данных PostgreSQL.
17	Итоговая аттестация	2	Тестирование.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ

Первый проректор –

проректор по учебной работе

МГТУ им. Н.Э. Баумана

Б.В. Падалкин

2021 г.



Дополнительное профессиональное образование

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПОВЫШЕНИЯ КВАЛИФИКАЦИИ
«Junior data analyst: младший аналитик данных»

Регистрац. № 05.22.21.12.1

Москва, 2021

АВТОР ПРОГРАММЫ:

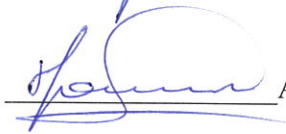
Заведующий лабораторией
Центра НТИ «Цифровое материаловедение:
новые материалы и вещества»



М.А. Орлов

СОГЛАСОВАНО:

Руководитель стратегического проекта «Bauman U2U»



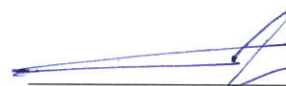
А.С. Комшин

Начальник УСП
МГТУ им. Н.Э. Баумана



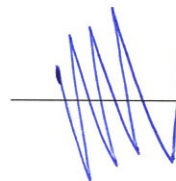
Т.А. Гузева

Директор ИСОТ
МГТУ им. Н.Э. Баумана



В.Г. Брекалов

Заместитель директора
Центра НТИ «Цифровое материаловедение:
новые материалы и вещества»



М.В. Стоянова

Оглавление

1. Общая характеристика дополнительной профессиональной программы (ДПП)	4
1.1. Цель ДПП	4
1.2. Планируемые результаты обучения	4
1.3. Дополнительные характеристики ДПП.....	4
1.4. Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения	5
1.5. Соответствие видов деятельности профессиональным компетенциям и их составляющих	5
2. Учебный план ДПП.....	6
2.1. Категория слушателей ДПП	6
2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа	6
2.3. Форма обучения.....	6
2.4. Учебный план	6
3. Календарный учебный график.....	7
4. Рабочая программа ДПП	10
5. Условия реализации ДПП	26
5.1. Организационные условия реализации ДПП.....	26
5.2. Педагогические условия реализации ДПП	27
5.3. Учебно-методическое обеспечение ДПП.....	27
5.4. Методические рекомендации	28
6. Формы итоговой аттестации ДПП.....	28
7. Оценочные материалы итоговой аттестации.....	29
7.1. Паспорт комплекта оценочных средств	29
7.2. Комплект оценочных средств	29

1. Общая характеристика дополнительной профессиональной программы (ДПП)

Программа подготовлена на основе:

- Федерального закона от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации»;
- приказа Минобрнауки России от 23 августа 2017 года № 816 «Об утверждении Порядка применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»;
- требований Приказа Минобрнауки России от 01.07.2013 года № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;
- методических рекомендаций-разъяснений Минобрнауки России по разработке дополнительных профессиональных программ на основе профессиональных стандартов от 22 апреля 2015 года № ВК-1030/06.

Реализация программы дополнительного профессионального образования направлена на получение новой(-ых) компетенции(-ий), необходимой(-ых) для профессиональной деятельности, в рамках реализации программы стратегического академического лидерства «Приоритет-2030».

1.1. Цель ДПП

Сформировать у обучающихся компетенции в области создания информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг.

1.2. Планируемые результаты обучения

Планируемые результаты обучения по ДПП:

- освоение профессиональных компетенций в процессе изучения перечисленных тем в учебном плане;
- успешное освоение программы повышения квалификации;
- успешное прохождение итоговой аттестации (зачет).

Обучающимся, успешно прошедшим обучение, выполнившим текущие контрольные задания и выдержавшим предусмотренное учебным планом итоговую аттестацию, выдается удостоверение о повышении квалификации по ДПП «Junior data analyst: младший аналитик данных».

1.3. Дополнительные характеристики ДПП

Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения, определены Приказом Министерства труда и социальной защиты Российской Федерации от 06.07.2020 №405н об утверждении профессионального стандарта «Специалист по большим данным».

Вид профессиональной деятельности:

- Создание и применение технологий больших данных (Код 06.042).

Обобщенные трудовые функции:

- Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры (ОТФ 06.042_А).

Трудовые функции:

- Выявление, формирование и согласование требований к результатам аналитических работ с применением технологий больших данных (ТФ 06.042_А/01.6).

1.4. Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения

Данная программа направлена на совершенствование и (или) получение новой(-ых) компетенции, необходимой для профессиональной деятельности, и (или) повышение профессионального уровня в рамках имеющейся квалификации.

При определении профессиональных компетенций на основе профессиональных стандартов, МГТУ им. Н.Э.Баумана осуществляет выбор профессиональных стандартов, соответствующих профессиональной деятельности выпускников, из числа указанных в приложении к ФГОС ВО и (или) иных профессиональных стандартов, соответствующих профессиональной деятельности выпускников.

Реализуемые компетенции базируются на основании Приказа Министерства образования и науки РФ от 19 сентября 2017 г. N 929 "Об утверждении федерального государственного образовательного стандарта высшего образования - бакалавриат по направлению подготовки 09.03.01 Информатика и вычислительная техника" (с изменениями и дополнениями) Редакция с изменениями N 1456 от 26.11.2020.

Перечень компетенций:

ПК-1. Способен участвовать в исследовательских и опытно-конструкторских разработках в области создания и совершенствования ИТ-систем.

ОПК-3. Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности.

1.5. Соответствие видов деятельности профессиональным компетенциям и их составляющих

Профессиональные компетенции	Практический опыт	Умения	Знания
Выявление, формирование и согласование требований к результатам аналитических работ с применением технологий больших данных (А/01.6)			
ПК-1. Способен участвовать в исследовательских и опытно-конструкторских разработках в области создания и совершенствования ИТ-систем	Выявление требований заказчика к результатам анализа, определение возможностей применения анализа больших данных в предметной области и конкретных задачах заказчика	Проводить сравнительный анализ методов и инструментальных средств анализа больших данных	Стандарты проведения анализа данных
ОПК-3. Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности			

2. Учебный план ДПП

2.1. Категория слушателей ДПП

Имеющаяся квалификация (требования к обучающимся) – к освоению ДПП допускаются лица, соответствующего профессиональному стандарту уровню образования или получающие соответствующий уровень образования (бакалавриат).

2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа

Общая трудоёмкость программы 44 академических часа, из них 23 часа аудиторной работы, 19 часов самостоятельной работы и 2 часа итоговой аттестации.

2.3. Форма обучения

Форма обучения по ДПП – очная, с применением электронного обучения и дистанционных образовательных технологий.

2.4. Учебный план

ДПП «Junior data analyst: младший аналитик данных» реализуется одним модулем.

№ п/п	Наименование темы, модуля	Форма контроля	Всего, акад. час*	В том числе		
				Лекции	Практические занятия	Самостоятельная работа
1.	Введение в Big Data и Data science. Data science в различных секторах экономики.	устный опрос	1	1	-	1
2.	Данные и источники, характеристики, корреляция. Структурированные, полуструктурированные и неструктурированные данные.	тест	2	1	-	1
3.	Введение в статистику. Закон больших чисел. Нормальное распределение. Проверка гипотез.	тест	1,5	1	-	0,5
4.	Типы данных.	тест	1	0,5	-	0,5
5.	Введение в Python, среды исполнения (IDE). Типы данных в Python.	устный опрос	2	1,5	-	0,5
6.	Базовые конструкции и структуры.	домашнее задание	2	-	1	1

7.	Циклы и условия. Функции и классы	домашнее задание	3	-	1	1
8.	Библиотеки Numpy, Scipy для научных вычислений.	домашнее задание	3	-	1	2
9.	Библиотека Pandas, как стандарт исследования данных	домашнее задание	5	-	3	2
10.	Визуализация в Python и срезы данных. Библиотеки Matplotlib, seaborn, plotly.	домашнее задание	4	-	2	2
11.	Поиск бизнес решений и анализ датасетов.	домашнее задание	4	1	1	2
12.	Системы хранения данных. Базы данных.	тест	1,5	1	-	0,5
13.	Реляционные базы данных и их моделирование.	тест	2,5	0,5	1	1
14.	Функционал PGAdmin для PostgreSQL	тест	2	0,5	0,5	1
15.	Язык структурированных запросов SQL. Синтаксис	тест	4,5	0,5	3	1
16.	Обращение к базе данных PostgreSQL с помощью языка SQL	тест	3	0,5	0,5	2
17.	Итоговая аттестация	зачет	2	-	-	-
	ИТОГО	-	44	9	14	19

*академический час составляет 45 минут

3. Календарный учебный график

№ п/п	Наименование темы, модуля	1 день	2 день	3 день	4 день	5 день	6 день
1.	Введение в Big Data и Data science. Data science в различных секторах экономики.						
2.	Данные и источники,						

	характеристики, корреляция. Структурированные, полуструктурированные и неструктурированные данные.						
3.	Введение в статистику. Закон больших чисел. Нормальное распределение. Проверка гипотез.						
4.	Типы данных.						
5.	Введение в Python, среды исполнения (IDE). Типы данных в Python.						
6.	Базовые конструкции и структуры.						
7.	Циклы и условия. Функции и классы						
8.	Библиотеки Numpy, Scipy для научных вычислений.						
9.	Библиотека Pandas, как стандарт исследования данных						
10.	Визуализация в Python и срезы данных. Библиотеки Matplotlib, seaborn, plotly.						
11.	Поиск бизнес решений и анализ датасетов.						
12.	Системы хранения данных. Базы данных.						
13.	Реляционные базы данных и их моделирование.						
14.	Функционал PGAdmin для PostgreSQL						
15.	Язык структурированных запросов SQL. Синтаксис						
16.	Обращение к базе данных PostgreSQL с помощью языка SQL						
17.	Итоговая аттестация						

№ п/п	Наименование темы, модуля	7 день	8 день	9 день	10 день	11 день
1.	Введение в Big Data и Data science. Data science в различных секторах экономики.					
2.	Данные и источники,					

	характеристики, корреляция. Структурированные, полуструктурированные и неструктурированные данные.					
3.	Введение в статистику. Закон больших чисел. Нормальное распределение. Проверка гипотез.					
4.	Типы данных.					
5.	Введение в Python, среды исполнения (IDE). Типы данных в Python.					
6.	Базовые конструкции и структуры.					
7.	Циклы и условия. Функции и классы					
8.	Библиотеки Numpy, Scipy для научных вычислений.					
9.	Библиотека Pandas, как стандарт исследования данных					
10.	Визуализация в Python и срезы данных. Библиотеки Matplotlib, seaborn, plotly.					
11.	Поиск бизнес решений и анализ датасетов.					
12.	Системы хранения данных. Базы данных.					
13.	Реляционные базы данных и их моделирование.					
14.	Функционал PGAdmin для PostgreSQL					
15.	Язык структурированных запросов SQL. Синтаксис					
16.	Обращение к базе данных PostgreSQL с помощью языка SQL					
17.	Итоговая аттестация					Зачет

Рекомендуемый срок освоения ДПП — 11 дней.

Фактическое расписание занятий утверждается при заключении договора с обучающимися или при формировании группы.

4. Рабочая программа ДПП

4.1 Рабочая программа модуля

4.1.1. Цель изучения модуля: сформировать у обучающихся компетенции в области создания информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг.

4.1.2. Задача изучения модуля: изучить средства анализа больших данных.

4.1.3. Планируемые результаты обучения

Процесс изучения раздела направлен на формирование следующих компетенций

Код/ наименование компетенции	Перечень планируемых результатов обучения по модулю	Методы и формы обучения, способствующие формированию и развитию компетенции
ПК-1	Знать: Методы поиска научно-технической информации по теме исследования; методы проведения экспериментов и наблюдений, обобщения и обработки их результатов. Уметь: Выполнять поиск и анализ научно-технической информации по теме исследования, проводить эксперименты, обобщать и обрабатывать полученные результаты. Владеть: Практическими навыками поиска научно-технической информации по теме исследования.	Методы обучения: Активные, пассивные, интерактивные. Формы обучения: лекция; практическое занятие; самостоятельная работа.
ОПК-3	Знать: Стандарты проведения анализа данных. Уметь: Проводить сравнительный анализ методов и инструментальных средств анализа больших данных. Владеть: Выявление требований заказчика к результатам анализа, определение возможностей применения анализа больших данных в предметной области и конкретных задачах заказчика.	Методы обучения: Активные, пассивные, интерактивные. Формы обучения: лекция; практическое занятие; самостоятельная работа.

4.1.4 Содержание курса

Тема 1. Введение в Big Data и Data science. Data science в различных секторах экономики (1 час)

Лекции (1 час). В рамках данной темы будут рассмотрены основные термины, типы задач, которые решаются с помощью Data Science. Обоснованность применения.

Самостоятельная работа (1 час).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Введение в Big Data и Data science. Data science в различных секторах экономики	Большие данные	Проработка дополнительных источников информации	Круз Р.Л. Структуры данных и проектирование программ: [учеб. пособие] / Круз Р.Л.; пер. 3-го англ. изд. Финогенов К.Г. – М.: БИНОМ. Лаборатория знаний, 2017. – 765 с. (https://library.bmstu.ru/Catalog/Details/476469)	тест

Тема 2. Данные и источники, характеристики, корреляция. Структурированные, полуструктурированные и неструктурированные данные (2 часа)

Лекции (1 час). Будет рассмотрено понятие данные, основные открытые и закрытые источники данных. Характеристики данных, основные различия по типу структурированности, способах обработки и хранению.

Самостоятельная работа (1 час).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Данные и источники, характеристики, корреляция. Структурированные, полуструктурированные и неструктурированные данные	Абстракция данных	Проработка дополнительных источников информации	Круз Р.Л. Структуры данных и проектирование программ: [учеб. пособие] / Круз Р.Л.; пер. 3-го англ. изд. Финогенов К.Г. – М.: БИНОМ. Лаборатория знаний, 2017. – 765 с. (https://library.bmstu.ru/Catalog/Details/476469)	тест

Тема 3. Введение в статистику. Закон больших чисел. Нормальное распределение. Проверка гипотез (1,5 часа)

Лекции (1 час). В рамках данной темы будут рассмотрены определения статистической науки, законы больших чисел, нормальное распределение и его роль в описании и анализе данных. Будет рассмотрено понятие нулевая гипотеза и проверка нулевой гипотезы, ошибки первого и второго рода.

Самостоятельная работа (0,5 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Введение в статистику. Закон больших чисел. Нормальное распределение. Проверка гипотез	Большие числа, нормальное распределение	Проработка дополнительных источников информации	Меженная Н.М. Основы теории вероятностей и математической статистики: курс лекций / Меженная Н.М.; МГТУ им. Н.Э. Баумана. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2016. – 108 с. (https://library.bmstu.ru/Catalog/Details/465260)	тест

Тема 4. Типы данных (1 час)

Лекции (0,5 часа). В рамках данной темы будут рассмотрены основные типы данных, которые применяются в data science, программировании и базах данных. Целочисленные, логические, строки, числа с плавающей точкой, дата и время, timestamp.

Самостоятельная работа (0,5 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Типы данных	Структурирование, большие массивы данных	Проработка дополнительных источников информации	Антонова В.А., Антонова В.М. Введение в анализ больших информационных массивов / Антонова В.А., Антонова В.М. – М.: Изд-во МГТУ им. Н. Э. Баумана, 2021. – 52 с. (https://library.bmstu.ru/Catalog/Details/553766)	тест

Тема 5. Введение в Python, среды исполнения (IDE). Типы данных в Python (2 часа)

Лекции (1,5 часа). в рамках данной темы будут рассмотрены основные среды и продукты PyCharm, Anaconda, Google Colaboratory. Язык программирования - Python, как стандарт для работы с большими данными основные функции, типы данных в Python.

Самостоятельная работа (0,5 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Введение в Python, среды исполнения (IDE). Типы данных в Python	Большие данные	Проработка дополнительных источников информации	Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных / Силен Д., Мейсман А., Али М.; пер. с англ. Матвеев Е. – СПб.: Питер, 2020. – 334 с. (https://library.bmstu.ru/Catalog/Details/544371)	устный опрос

Тема 6. Базовые конструкции и структуры (2 часа)

Практические занятия (1 час). Основные конструкции, списки, словари, множества, кортежи, стеки, очереди, логические условия, арифметические операции.

Самостоятельная работа (1 час).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Базовые конструкции и структуры	Грани данных	Проработка дополнительных источников информации	Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных / Силен Д., Мейсман А., Али М.; пер. с англ. Матвеев Е. – СПб.: Питер, 2020. – 334 с. (https://library.bmstu.ru/Catalog/Details/544371)	домашнее задание

Тема 7. Циклы и условия. Функции и классы (3 часа)

Практические занятия (1 час). Синтаксис построения циклов, условных операторов, пользовательских функций, классов, наследование и полиморфизм.

Самостоятельная работа (1 час).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Циклы и условия. Функции и классы	Циклы в Python	Проработка дополнительных источников информации	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение / Плас Дж. Вандер; пер. с англ. Пальти И. – СПб.: Питер, 2020. – 572 с. (https://library.bmstu.ru/Catalog/Details/550326)	домашнее задание

Тема 8. Библиотеки Numpy, Scipy для научных вычислений (3 часа)

Практические занятия (1 час). Библиотека NumPy, типы данных, массивы и операции с ними, Матричные операции, семплирование, чтение файлов. Библиотека SciPy. Научные вычисления. Практическое применение. Работа с данными. Тестирование данных.

Самостоятельная работа (2 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Библиотеки Numpy, Scipy для научных вычислений	Библиотека Numpy	Проработка дополнительных источников информации	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение / Плас Дж. Вандер; пер. с англ. Пальти И. – СПб.: Питер, 2020. – 572 с. (https://library.bmstu.ru/Catalog/Details/550326)	домашнее задание

Тема 9. Библиотека Pandas, как стандарт исследования данных (5 часов)

Практические занятия (3 часа). Библиотека Pandas, загрузка и запись данных, срезы данных, мульти индексация, group by, datetime, статистические функции, стандартная визуализация. Преимущества и недостатки библиотеки Pandas.

Самостоятельная работа (2 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Библиотека Pandas, как стандарт исследования данных	Библиотека Pandas	Проработка дополнительных источников информации	Лутц М. Python. Карманный справочник / Лутц М.; Пер. с англ. и ред. пер. Берштейн И.В. – 5-е изд. – М.: Издат. дом «Вильямс», 2015. – 318 с. (https://library.bmstu.ru/Catalog/Details/475602)	домашнее задание

Тема 10. Визуализация в Python и срезы данных. Библиотеки Matplotlib, seaborn, plotly (4 часа)

Практические занятия (2 часа). Библиотека Matplotlib, варианты отрисовки графиков и изображений. Библиотека Seaborn, визуализация парных взаимосвязей, heatmap, диаграммы, тепловая карта корреляции. Библиотека Plotly. Продвинутая визуализация. Динамические графики.

Самостоятельная работа (2 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Визуализация в Python и срезы данных. Библиотеки Matplotlib, seaborn, plotly	Визуализация в Python	Проработка дополнительных источников информации	Сузи Р. Python. В подлиннике: Наиболее полное руководство / Сузи Р. – СПб.: БХВ-Петербург, 2002. – 747 с. (https://library.bmstu.ru/Catalog/Details/84709)	домашнее задание

Тема 11. Поиск бизнес решений и анализ датасетов (4 часа)

Лекции (1 час). В рамках данной темы будет рассмотрен и описан общий подход к описательному анализу данных, для выявления скрытых закономерностей и корреляций, для принятия бизнес решений.

Практические занятия (1 час). Сравнительный анализ датасетов.

Самостоятельная работа (2 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Поиск бизнес решений и анализ датасетов	Управленческие решения, математическая модель	Проработка дополнительных источников информации	Методы и модели принятия решений в задачах оценки качества и технического уровня сложных технических систем: монография / Семёнов С.С., Воронов Е.М., Полтавский А.В., Крянев А.В.; ред. Рубинович Е.Я.; Ин-т проблем управления им. В.А. Трапезникова РАН. – 3-е изд. – М.: URSS: Ленанд, 2020. – 517 с. (https://library.bmstu.ru/Catalog/Details/534649)	домашнее задание

Тема 12. Системы хранения данных. Базы данных (1,5 часа)

Лекции (1 час). В рамках данной темы будут рассмотрены основные подходы к хранению данных, горизонтально и вертикально масштабируемые системы хранения данных, реляционные и NoSQL. Key-Value хранилища.

Самостоятельная работа (0,5 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Системы хранения данных. Базы данных	NoSQL	Проработка дополнительных источников информации	Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных / Силен Д., Мейсман А., Али М.; пер. с англ. Матвеев Е. – СПб.: Питер, 2020. – 334 с. (https://library.bmstu.ru/Catalog/Details/544371)	тест

Тема 13. Реляционные базы данных и их моделирование (2,5 часа)

Лекции (0,5 часа). Будут рассмотрены реляционные базы данных, для хранения структурированных данных. Основные определения и признаки реляционных баз данных. Понятие и создание моделей (схем) баз данных и установление связей один-ко-многим (one2many), многие-ко-многим (many2many), многие-к-одному (many2one).

Практические занятия (1 час). Создание моделей баз данных.

Самостоятельная работа (1 час).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Реляционные базы данных и их моделирование	Реляционный подход, организация баз данных	Проработка дополнительных источников информации	Гордеев С.И., Волошина В.Н. Организация баз данных: учебник для вузов: в 2 ч. / Гордеев С.И., Волошина В.Н. – 2-е изд., испр. и доп. – М.: Юрайт, – 2021. (https://library.bmstu.ru/Catalog/Details/555437)	тест

Тема 14. Функционал PGAdmin для PostgreSQL (2 часа)

Лекции (0,5 часа). В рамках данной темы будет рассмотрен процесс установки реляционной базы данных PostgreSQL, и интерфейс - оболочки PGAdmin, а также разобран основной функционал PGAdmin, для удобной навигации по реляционным базам данных. Будет разобран процесс восстановления и развертывания базы данных из файла.

Практические занятия (0,5 часа). Установка баз данных PostgreSQL и PGAdmin. Развертка баз данных.

Самостоятельная работа (1 час).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Функционал PGAdmin для PostgreSQL	PGAdmin, PostgreSQL	Проработка дополнительных источников информации	Лутц М. Python. Карманный справочник / Лутц М.; Пер. с англ. и ред. пер. Берштейн И.В. – 5-е изд. – М.: Издат. дом «Вильямс», 2015. – 318 с. (https://library.bmstu.ru/Catalog/Details/475602)	тест

Тема 15. Язык структурированных запросов SQL. Синтаксис (4,5 часа)

Лекции (0,5 часа). В рамках данной темы будет рассмотрен SQL - язык структурированных запросов. Общий подход. Базовая концепция, фильтрация, функции. Функционал PgAdmin. Подзапросы, Join, Табличные операции.

Практические занятия (3 часа). Изучение функционала PgAdmin.

Самостоятельная работа (1 час).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Язык структурированных запросов SQL. Синтаксис	Синтаксис SQL	Проработка дополнительных источников информации	Сузи Р. Python. В подлиннике: Наиболее полное руководство / Сузи Р. – СПб.: БХВ-Петербург, 2002. – 747 с. (https://library.bmstu.ru/Catalog/Details/84709)	тест

Тема 16. Обращение к базе данных PostgreSQL с помощью языка SQL (3 часа)

Лекции (0,5 часа). В рамках данной темы будут рассмотрены задачи создания новых таблиц, внешних и внутренних ключей их связей. Формирование аналитических запросов в реляционной базе данных PostgreSQL.

Практические занятия (0,5 часа). Формирование аналитических запросов к базам данных.

Самостоятельная работа (2 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Обращение к базе данных PostgreSQL с помощью языка SQL	Обращение к базам данных в SQL	Проработка дополнительных источников информации	Сузи Р. Python. В подлиннике: Наиболее полное руководство / Сузи Р. – СПб.: БХВ-Петербург, 2002. – 747 с. (https://library.bmstu.ru/Catalog/Details/84709)	тест

4.1.5. Оценочное средство для текущего контроля (темы для подготовки к тесту и устному опросу, формулировка условий домашних заданий):

Тема 1. Устный опрос: термины, задачи и определения в data science.

Тема 2. Тест: типы данных, способы хранения, обработки. Оптимальная структура хранения данных (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Что такое данные?
 - А. Это структуры представления фактов или понятий, или инструкций в любой свободной, даже непонятной форме
 - В. Представление фактов, понятий или инструкций в форме, приемлемой для интерпретации или обработки человеком или машиной
 - С. Это факты, понятия или инструкции, не обработанные до понятной формы, которую может обработать человек или машина

2. Что такое алгоритмическая сложность
 - А. Способ оценки стиля написания кода. Необходим для понимания насколько код является читаемым для других людей
 - В. Способ оценки технической наполненности кода. Необходим для оценки сложности используемых функций и способов их применения
 - С. Способ оценки производительности алгоритма. Необходим для проверки того, как долго будет выполняться алгоритм для входных данных определённого размера
 - Д. Все вышеперечисленное

3. Выберите описание соответствующие структурированным данным
 - А. Данные хранятся в фиксированном поле внутри записи в таблицах, в базах данных или файлах.
 - В. Отсутствует жесткая модель данных и типизация, значение определяется контекстом и требуются специальные методы извлечения таких данных.
 - С. Отсутствует жесткая модель данных и типизация, данные хранятся в таблицах, в базах данных или файлах.

4. Что такое неструктурированные данные?
 - А. Данные хранятся в фиксированном поле внутри записи в таблицах, в базах данных или файлах.
 - В. Отсутствует жесткая модель данных и типизация, значение определяется контекстом и требуются специальные методы извлечения таких данных.
 - С. Отсутствует жесткая модель данных и типизация, данные хранятся в таблицах, в базах данных или файлах.
 - Д. Данные хранятся в фиксированном поле внутри записи в таблицах, в базах данных или файлах. Такие данные определяются контекстом и требуются специальные методы для их извлечения

5. Какие уровни работы с данными бывают? (Несколько вариантов ответа)
 - А. Слой доступа к данным
 - В. Слой изучения
 - С. Корневой слой
 - Д. Слой хранения
 - Е. Железо

Тема 3. Тест: определения и термины статистики. Виды и типы нормального распределения. Нулевая гипотеза, ошибки первого и второго рода (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Что такое статистика?
 - А. Отрасль программирования, отвечающая на вопросы взаимосвязей между разными данными
 - В. Отрасль знаний, изучающая данные в интернете и законы которым они подчиняются

- С. Наука, занимающаяся сбором, анализом, мониторингом и изучением данных, а также их сравнением и поиском взаимосвязей между ними
- Д. Сфера, занимающаяся контролем и управлением потоками данных во всех возможных сферах их происхождения
2. Что такое корреляция?
- А. Закон, описывающий как сделать значения зависимыми к друг другу
- В. Правило, помогающее находить лучшие варианты исхода экспериментов
- С. Подход по разработке гипотез, помогающих решать регрессионные задачи
- Д. Мера, отображающая зависимость между двумя взятыми переменными
3. Что такое нормальное распределение?
- А. Это закон, позволяющий сказать в какой области с наибольшей вероятностью мы можем встретить измеряемые значения
- В. Это правило, которому подчиняются только данные, созданные без участия человека (скорость ветра)
- С. Это закон, позволяющий сказать в какой временной промежуток может произойти то или иное явление
- Д. Это правило создания новых данных, согласно которому все данные должны принадлежать определённым
4. Что такое статистическая гипотеза?
- А. Это предположение о возможной величине взаимосвязи между значениями
- В. Это предположение о возможных причинах, спровоцировавших реакцию, создавшую событие
- С. Это предположение о возможных причинах провала или успеха предсказания того или иного события
- Д. Это предположение о виде распределения и свойствах случайной величины, которое можно подтвердить с помощью статистических методов, применённых к данным
5. Может ли быть корреляция отрицательной?
- А. Да, и это важно учитывать
- В. Да, но обычно это является редким и незначительным явлением
- С. Нет, это невозможно исходя из определения корреляции

Тема 4. Тест: различные типы данных (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Для обработки какого вида данных требуется специальное ПО?
- А. Полу-Структурированные
- В. Структурированные
- С. Не структурированные
- Д. Для структурированных и полу структурированных
- Е. Для всех вышеперечисленных
2. Что такое многомерные массивы?
- А. Это массивы внутри массива
- В. Это Массивы по размеру сопоставимые с big data
- С. Это массивы, которые человек не способен обработать
- Д. Это массивы, которые машина не способна обработать

3. Есть ли способ достать из стека не только верхний элемент? Порядок стека или нарушать его функции нельзя.
 - A. Да, это входит в функционал работы стека
 - B. Да, но для этого нужно использовать специальную математическую функцию
 - C. Нет, это нельзя сделать
4. Что такое связанные списки?
 - A. Это списки со списками внутри
 - B. Это списки, которые только выполняют роль ссылки на другой список
 - C. Это списки, хранящиеся внутри массивов
 - D. Это списки, хранящие в себе как собственные данные, так и ссылку на следующий или предыдущий узел списка
5. Что такое MAP?
 - A. Это словарь пар (ключ и значение), где все ключи уникальны
 - B. Это словарь пар (ключ и значение)
 - C. Это список, где можно хранить списки, к которым можно обращаться специальными индексами
 - D. Это функция языка Python, помогающая в редактировании данных без явного цикла

Тема 5. Устный опрос: основные функции, типы данных Python.

Тема 6. Домашнее задание: решить практические задачи на Python, используя базовые конструкции и структуры.

Тема 7. Домашнее задание: решить практические задачи на Python, используя циклы, условия, функции и классы.

Тема 8. Домашнее задание: решить практические задачи на Python, используя циклы, условия, функции и классы.

Тема 9. Домашнее задание: решить практические задачи на Python, используя библиотеки Pandas. Загрузить датасет, преобразовать, сделать срезы данных и визуализацию распределения.

Тема 10. Домашнее задание: решить практические задачи на Python, используя библиотеки Matplotlib, seaborn, plotly. Загрузить датасет, сделать визуализацию разных признаков, их распределения и корреляции их взаимосвязей.

Тема 11. Домашнее задание: провести описательный анализ датасета (Датасет описывает социально - демографические признаки и уровень дохода больше или меньше 50K \$), построить графики распределения, выполнить агрегирование признаков, выявить топ коррелируемых признаков, оценить взаимосвязи.

Тема 12. Тест: термины, задачи и признаки машинного обучения (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Что такое системы хранения данных
 - A. Это технические рекомендации по поводу того, как и где нужно хранить данные, как получать к ним доступ
 - B. Это программно-аппаратный комплекс. Основная задача которого - хранение и передача больших массивов информации и предоставление к ним доступа.
 - C. Это специальные программы, которые помогают организовать системы, по правилам которых будут храниться и использоваться все данные
2. В системе хранения DAS физические накопители памяти должны ли подключаться напрямую к ПК или серверу?
 - A. Да, это является основой этой системы
 - B. Да, но будет потеряна скорость доступа к данным

- C. Необязательно, можно и без подключения
 - D. Нет, это не предусмотрено в такой системе
3. NAS - это система...
- A. Система хранения, где накопитель присоединен к сети. При таком типе подключения используется только файловая передача данных
 - B. Система хранения, где накопитель присоединен к виртуальной машине визуализируемой на стационарных машинах
 - C. Система хранения, где накопитель присоединен напрямую к стационарной машине, но только к одной
 - D. Система хранения, где накопитель не присоединяется постоянно к какому-либо месту, по необходимости доступа производится физическое подключение к нужной машине
4. Выберите главную особенность системы хранения SAN.
- A. Прямое подключения накопителей к вычислительным машинам
 - B. Необязательно использовать сеть
 - C. Вертикальное масштабирование
 - D. Использует блочный тип хранения данных
5. Какие платформы проще и дешевле начать эксплуатировать с коммерческой и производственной точки зрения? (Несколько вариантов ответа)
- A. Распределительные платформы
 - B. Физические платформы
 - C. Облачные платформы
 - D. Локальные платформы
 - E. Серверные платформы

Тема 13. Тест: основные понятия реляционных моделей и связей в базах данных (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Что такое реляционные БД (базы данных)?
- A. база данных, использующая транзакции и с независимой от объемов обрабатываемых данных скоростью обработки
 - B. база данных, не использующая транзакции и с независимой от объемов обрабатываемых данных скоростью обработки
 - C. база данных, использующая транзакции и с зависимой от объемов обрабатываемых данных скоростью обработки
 - D. база данных, не использующая транзакции и с зависимой от объемов обрабатываемых данных скоростью обработки
2. Что в списке является свойствами данных (Несколько вариантов ответа)
- A. Атомарность
 - B. Изолированность
 - C. Надёжность
 - D. Консистентность
3. Что такое Entity Relationship Diagram?
- A. Это система, показывающая отсутствие какого-либо взаимоотношения между объектами
 - B. Это визуальное представление различных сущностей в системах и как они относятся друг к другу

- С. Это система визуальных представлений, рассматривающих объекты по отдельности
- Д. Это визуальное представление, отображающие все происходящие над данными вычисления
4. Что делает relationships?
- А. Вычисляет закономерности
- В. Описывает взаимодействие с помощью глагола, определяющего это взаимодействие
- С. Описывает сущность объекта с помощью прилагательного, определяющего сущность объекта
- Д. Создаёт искусственную связь между объектами, у которых этой связи не может быть
5. Можно ли соединять два relationships напрямую?
- А. Можно, это не противоречит логике
- В. Можно, но не во всех случаях
- С. Нельзя, это противоречит логике

Тема 14. Тест: способы установки и развертывания базы данных (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Что такое pgAdmin?
- А. Это платформа с закрытым исходным кодом для администрирования и разработки для PostgreSQL и не связанных с ней систем управления базами данных
- В. Это платформа с открытым исходным кодом для администрирования и разработки для PostgreSQL и не связанных с ней систем управления базами данных
- С. Это платформа с закрытым исходным кодом для администрирования и разработки для PostgreSQL и связанных с ней систем управления базами данных
- Д. Это платформа с открытым исходным кодом для администрирования и разработки для PostgreSQL и связанных с ней систем управления базами данных
2. На каком языке написана pgAdmin?
- А. на Python и jQuery
- В. на Python и JavaScript
- С. на C# и jQuery
- Д. на JavaScript и jQuery
3. Сможет ли ПК на Windows с 4 гб оперативной памяти работать с pgAdmin?
- А. Да, pgAdmin не является платформой с большими требованиями
- В. Да, но работа будет идти медленно
- С. Нет, нужно минимум 8 гб оперативной памяти
4. Сможет ли ПК на Linux с 6 гб оперативной памяти работать с pgAdmin?
- А. Да, такое количество оперативной памяти является достаточным
- В. Да, но работа будет идти медленно
- С. Нет, нужно минимум 8 гб оперативной памяти
5. Какая минимальная тактовая частота нужна для работы с pgAdmin?
- А. 4 GHz
- В. 2 GHz
- С. 3 GHz

D. 6 GHz

Тема 15. Тест: проверка знаний основного синтаксиса SQL запросов (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Какие типы запросов можно делать в SQL? (Несколько вариантов ответа)
 - A. Запросы на редактирование и создание новых данных
 - B. Запросы на получение данных
 - C. Запросы на добавление данных
 - D. Запрос на создание данных с нуля
 - E. Запросы на удаление данных
 - F. Запрос на обращение СУБД
2. Какие из операторов описывают данные? (Несколько вариантов ответа)
 - A. CREATE
 - B. DROP
 - C. LOCK
 - D. ALTER
 - E. GRANT
3. Что такое DDL?
 - A. Язык определения данных
 - B. Язык манипулирования данными
 - C. Язык управления данными
 - D. Язык управления транзакциями
4. Что такое TCL?
 - A. Язык определения данных
 - B. Язык манипулирования данными
 - C. Язык управления данными
 - D. Язык управления транзакциями
5. Что такое DCL?
 - A. Язык определения данных
 - B. Язык манипулирования данными
 - C. Язык управления данными
 - D. Язык управления транзакциями

Тема 16. Тест: ключевые особенности при формировании запросов к базе данных PostgreSQL (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Найдите номер модели и скорость для всех ПК стоимостью более 200 дол.
 - A. SELECT PC
FROM model, speed
WHERE price > 200
 - B. FROM PC
SELECT model, speed
WHERE price > 200
 - C. FROM model, speed

```
SELECT PC
WHERE price > 200
```

- D.

```
SELECT model, speed
FROM PC
WHERE price > 200
```
2. Найдите размеры жестких дисков, совпадающих у двух и более PC. Вывести:
HD
- A.

```
SELECT hd
FROM PC
GROUP BY hd
HAVING COUNT(hd) > 1
```
- B.

```
SELECT hd
FROM PC
GROUP BY hd
HAVING COUNT(hd) > 2
```
- C.

```
SELECT hd
FROM PC
HAVING COUNT(hd) > 1
```
- D.

```
SELECT hd
FROM PC
HAVING COUNT(hd) > 2
```
3. Найдите размеры жестких дисков, совпадающих у двух и более PC. Вывести:
HD
- A.

```
SELECT hd
FROM PC
GROUP BY hd
HAVING COUNT(hd) > 1
```
- B.

```
SELECT hd
FROM PC
GROUP BY hd
HAVING COUNT(hd) > 2
```
- C.

```
SELECT hd
FROM PC
HAVING COUNT(hd) > 1
```
- D.

```
SELECT hd
FROM PC
HAVING COUNT(hd) > 2
```
4. Найдите номер модели, скорость и размер жесткого диска ПК, имеющих 4х или 12х CD и цену менее 450 дол.
- A.

```
SELECT model, speed, hd
FROM PC
```

WHERE (cd = '4x' , '12x') AND price < 450

- B. SELECT model, speed, hd
FROM PC
WHERE cd = ('4x', '12x') AND price < 450
- C. SELECT model, speed, hd
FROM PC
WHERE (cd = '4x' OR cd = '12x') AND price < 450
- D. SELECT model, speed, hd
FROM PC
WHERE cd = '4x' OR cd = '12x' AND price < 450
5. Для каждого производителя, выпускающего ПК-блокноты с объёмом жесткого диска не менее 6 Гбайт, найти скорости таких ПК-блокнотов.
- A. SELECT DISTINCT Product.maker, Laptop.speed
FROM Product INNER JOIN
Laptop ON Laptop.model = Product.model
WHERE Laptop.hd >= 6
ORDER BY maker ASC, speed ASC
- B. SELECT DISTINCT maker, speed
FROM Product INNER JOIN
Laptop ON model = model
WHERE hd >= 6
ORDER BY maker ASC, speed ASC
- C. SELECT DISTINCT Product.maker, Laptop.speed
FROM Product INNER JOIN
Laptop ON model = model
WHERE hd >= 6
ORDER BY maker ASC, speed ASC
- D. SELECT DISTINCT Product.maker, Laptop.speed
FROM Product INNER JOIN
WHERE Laptop.hd >= 6
ORDER BY maker ASC, speed ASC

5. Условия реализации ДПП

5.1. Организационные условия реализации ДПП

Наименование аудитории	Вид занятия	Наименование оборудования, программного обеспечения
Компьютерный класс	Лекции	Материальное обеспечение: компьютер, мультимедийный проектор, экран, доска, пишущий инструмент, Программное обеспечение: Anaconda
Компьютерный	Практические занятия	Материальное обеспечение: компьютер, мультимедийный проектор, экран,

класс		доска, пишущий инструмент, Программное обеспечение: Anaconda
Компьютерный класс	Самостоятельная работа	Материальное обеспечение: компьютер, мультимедийный проектор, экран, доска, пишущий инструмент, Программное обеспечение: Anaconda
Компьютерный класс	Итоговая аттестация	Материальное обеспечение: компьютер, мультимедийный проектор, экран, доска, пишущий инструмент, Программное обеспечение: Anaconda

5.2. Педагогические условия реализации ДПП

Реализация программы обеспечивается преподавательским составом, удовлетворяющим следующим условиям:

- наличие высшего образования, соответствующее профилю программы, из числа штатных преподавателей, или привлеченных на условиях почасовой оплаты труда;
- опыт практической деятельности в соответствующей сфере из числа штатных преподавателей или привлеченных на условиях почасовой оплаты труда.

5.3. Учебно-методическое обеспечение ДПП

1. Круз Р.Л. Структуры данных и проектирование программ: [учеб. пособие] / Круз Р.Л.; пер. 3-го англ. изд. Финогенов К.Г. – М.: БИНОМ. Лаборатория знаний, 2017. – 765 с. (<https://library.bmstu.ru/Catalog/Details/476469>)
2. Меженная Н.М. Основы теории вероятностей и математической статистики: курс лекций / Меженная Н.М.; МГТУ им. Н.Э. Баумана. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2016. – 108 с. (<https://library.bmstu.ru/Catalog/Details/465260>)
3. Антонова В.А., Антонова В.М. Введение в анализ больших информационных массивов / Антонова В.А., Антонова В.М. – М.: Изд-во МГТУ им. Н. Э. Баумана, 2021. – 52 с. (<https://library.bmstu.ru/Catalog/Details/553766>)
4. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных / Силен Д., Мейсман А., Али М.; пер. с англ. Матвеев Е. – СПб.: Питер, 2020. – 334 с. (<https://library.bmstu.ru/Catalog/Details/544371>)
5. Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение / Плас Дж. Вандер; пер. с англ. Пальти И. – СПб.: Питер, 2020. – 572 с. (<https://library.bmstu.ru/Catalog/Details/550326>)
6. Лутц М. Python. Карманный справочник / Лутц М.; Пер. с англ. и ред. пер. Берштейн И.В. – 5-е изд. – М.: Издат. дом «Вильямс», 2015. – 318 с. (<https://library.bmstu.ru/Catalog/Details/475602>)
7. Сузи Р. Python. В подлиннике: Наиболее полное руководство / Сузи Р. – СПб.: БХВ-Петербург, 2002. – 747 с. (<https://library.bmstu.ru/Catalog/Details/84709>)
8. Методы и модели принятия решений в задачах оценки качества и технического уровня сложных технических систем: монография / Семёнов С.С., Воронов Е.М., Полтавский А.В., Крянев А.В.; ред. Рубинович Е.Я.; Ин-т проблем управления им. В.А. Трапезникова РАН. – 3-е изд. – М.: URSS: Ленанд, 2020. – 517 с. (<https://library.bmstu.ru/Catalog/Details/534649>)
9. Гордеев С.И., Волошина В.Н. Организация баз данных: учебник для вузов: в 2 ч. / Гордеев С.И., Волошина В.Н. – 2-е изд., испр. и доп. – М.: Юрайт, – 2021. (<https://library.bmstu.ru/Catalog/Details/555437>)

5.4. Методические рекомендации

ДПП построена по тематическому принципу, каждый раздел представляет собой логически заверченный материал.

Преподавание программы основано на личностно-ориентированной технологии образования, сочетающей два равноправных аспекта этого процесса: обучение и учение. Личностно-ориентированный подход развивается при участии слушателей в активной работе на практических занятиях. Личностно-ориентированный подход направлен, в первую очередь, на развитие индивидуальных способностей обучающихся, создание условий для развития творческой активности слушателя и разработке инновационных идей, а также на развитие самостоятельности мышления при решении учебных задач разными способами, нахождение рационального варианта решения, сравнения и оценки нескольких вариантов их решения и т.п. Это способствует формированию приемов умственной деятельности по восприятию новой информации, ее запоминанию и осознанию, созданию образов для сложных понятий и процессов, приобретению навыков поиска решений в условиях неопределенности.

Лекции проводятся для приобретения навыков реализации знаний в предметной области, с использованием активных методов обучения.

Практические занятия проводятся для приобретения навыков решения практических задач в предметной области. Задания, выполняемые на практических занятиях, выполняются с использованием активных и интерактивных методов обучения.

Самостоятельная работа слушателей предназначена для проработки дополнительных источников информации. При изучении ДПП предусмотрены следующие методы организации и осуществления учебно-познавательной деятельности:

- объяснительно-иллюстративный метод;
- репродуктивный метод;
- частично-поисковый метод.

При изучении ДПП предусмотрены активные формы проведения занятий:

- управляемая дискуссия;
- разбор конкретных ситуаций.

6. Формы итоговой аттестации ДПП

Итоговая аттестация проводится в форме зачета для проверки сформированности компетенций, полученных в рамках ДПП.

Зачет проводится в формате тестирования. Результатом зачета служат правильные ответы на вопросы билета.

По результатам итоговой аттестации обучающемуся выставляется оценка «ЗАЧТЕНО/НЕ ЗАЧТЕНО»:

Оценка «ЗАЧТЕНО» выставляется обучающемуся, который:

- ответил на 8 из 12 вопросов теста;
- продемонстрировал необходимые систематизированные знания и достаточную степень владения принципами предметной области программы, понимание их особенностей и взаимосвязь между ними в течение всего срока обучения по ДПП.

Оценка «НЕ ЗАЧТЕНО» ставится обучающемуся, который:

- ответил менее, чем на 8 из 12 вопросов теста;
- имеет крайне слабые теоретические и практические знания, обнаруживает неспособность к построению самостоятельных заключений.

7. Оценочные материалы итоговой аттестации

7.1. Паспорт комплекта оценочных средств

Предметы оценивания	Объекты оценивания	Показатели оценки
ПК-1. Способен участвовать в исследовательских и опытно-конструкторских разработках в области создания и совершенствования ИТ-систем	Ответы на вопросы	Количество правильных ответов (60% правильных ответов)
ОПК-3. Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	Ответы на вопросы	Количество правильных ответов (60% правильных ответов)

7.2. Комплект оценочных средств

7.2.1. Темы для подготовки к зачету:

1. Данные и источники.
2. Статистика и закон больших чисел.
3. Типы данных.
4. Циклы в Python.
5. Условия в Python.
6. Библиотеки в Python.
7. Визуализация в Python.
8. Системы хранения данных.
9. Моделирование баз данных.
10. Обращение к базам данных.

7.2.2. Пример билета:

1 – Как называются массивы данных большого объема для обработки которых нужны специальные инструменты?

A – Closed data

B – Big closed data

C – Big data

D – Slow date

2 – Зачем Data Science бизнесу?

A – Он нужен потому что неструктурированная информация содержит очень важные для компании знания, способные повлиять на результаты бизнеса

В – Он нужен потому что бизнес не является возможным по определению при отсутствии Data science

С – Он нужен потому что данные засоряют ‘Обзор’ аналитиков и нужны механизмы их отвода от бизнеса

3 – Что такое выборка данных, систематизированных по определенным критериям, в которой критерии выборки задаются с помощью запроса к базе данных?

А – Источники данных

В – Данные сети интернет

С – Таблицы

Д – Полуструктурированные данные

4 – Если в переменной находятся такие данные: «K1p1lx2 д0м@», какой тип данных будет у переменной?

А – Int

В – Str

С – Float

Д – Object

5 – Чем является среда в которой будет писаться код?

А – GST

В – ARW

С – IDE

Д – KIoStr

6 – Что из вариантов ниже связано с термином функция больше всего?

А – for

В – elif

С – drop

Д – def

7 – Что из вариантов ниже связано с термином цикл больше всего?

А – for

В – elif

С – drop

Д – def

8 – Какая библиотека позволяет работать с таблицами в Python?

А – Numpy

В – Pandas

С – Scipy

9 – Какие из вариантов библиотек ниже позволяют визуализировать данные? (Несколько вариантов ответа)

А – Matplotlib

В – Matplotlib

С – Seaborn

Д – Seaborn

Е – Plotly

Ф – VisionPlot

10 – Если аналитику нужно будет сделать запрос к Big Data на получение данных, кокой язык он будет использовать?

A – Python

B – Keras

C – SQL

D – PGAdmin

11 – Система хранения структурированных данных, основная цель которой - хранить информацию и предоставлять ее по требованию это...

A – СКЛ

B – СВВ

C – БВВ

D – СБД

12 – ‘план/инструкция’, из которого создаются объекты в Python это...

A – Класс

B – Функция

C – Комментарий

D – Генератор