



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Магнитогорский государственный технический университет им. Г.И.
Носова»



УТВЕРЖДАЮ:
Директор ИЭиАС
С.И. Лукьянов
26.02.2020 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

АНАЛИЗ БОЛЬШИХ ДАННЫХ

Направление подготовки (специальность)
38.03.05 Бизнес-информатика

Уровень высшего образования – бакалавриат

Программа подготовки – академический бакалавриат

Форма обучения
очная

Институт
Кафедра
Курс
Семестр

Институт энергетики и автоматизированных систем
Бизнес-информатики и информационных технологий
2
4


Магнитогорск
2020 г.

Рабочая программа составлена на основе ФГОС ВО по направлению подготовки 38.03.05 Бизнес-информатика (уровень бакалавриата) утвержденного приказом МОиН РФ от 11.08.2016 г. № 1002.

Рабочая программа рассмотрена и одобрена на заседании кафедры Бизнес-информатики и информационных технологий 11.02.2020 г., протокол № 6.

Зав. кафедрой  / Г.Н. Чусавитина /

Рабочая программа одобрена методической комиссией Института энергетики и автоматизированных систем 26.02.2020 г., протокол № 5.

Председатель  / С.И. Лукьянов /


Рабочая программа составлена:

доцент каф. БИиИТ, к.п.н, доцент

 / Е.В. Карманова /

Рецензент:

Руководитель группы анализа
ИТ-проектов ЗАО "Консом СКС"

 / В.А. Ошурков /

Лист актуализации рабочей программы

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2021 - 2022 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2022 - 2023 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2023 - 2024 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2024 - 2025 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2025 - 2026 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

1 Цели освоения дисциплины (модуля)

Целями освоения дисциплины «Анализ больших данных» являются: является теоретическая и практическая подготовка студентов к работе с большими данными. Знания, полученные в результате освоения дисциплины, помогут при сборе и анализе огромных объемов структурированной или неструктурированной информации, при разработке моделей данных и получении новых знаний. Все это необходимо выпускнику, освоившему программу бакалавриата, для решения различных задач практической и научно-исследовательской деятельности.

2 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина Анализ больших данных входит в базовую часть учебного плана образовательной программы.

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения дисциплин/ практик:

- Теория вероятностей и математическая статистика для анализа больших данных
- Программирование на Python
- Управление данными
- Математические модели и методы цифровой экономики

Знания (умения, владения), полученные при изучении данной дисциплины будут необходимы для изучения дисциплин/практик:

- Основы машинного обучения
- Инструменты продвинутой бизнес-аналитики
- Нейронные сети

3 Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля) и планируемые результаты обучения

В результате освоения дисциплины (модуля) «Анализ больших данных» обучающийся должен обладать следующими компетенциями:

Структурный элемент компетенции	Планируемые результаты обучения
ОПК-1 - способностью решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	
Знать	<ul style="list-style-type: none">• методы решения задач предобработки, обработки и анализа больших данных;• методы визуализации больших данных;• возможности высокопроизводительных вычислительных систем;• технологии распределенных вычислений, методы и модели Data Mining.
Уметь	<ul style="list-style-type: none">• разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных;• оценивать время и необходимые аппаратные ресурсы для решения задач анализа и обработки данных;• создавать алгоритмы анализа и обработки большого объема данных с применением моделей Data Mining.
Владеть	<ul style="list-style-type: none">• навыками применения программных систем, предназначенных для анализа больших данных.
ПК-2 - проведение исследования и анализа рынка информационных систем и информационно-коммуникативных технологий	
Знать	<ul style="list-style-type: none">• методы анализа больших данных;• сервисы хранения больших данных;
Уметь	<ul style="list-style-type: none">• осуществлять поиск больших данных в сети Интернет;• проводить исследования на основе анализа больших данных
Владеть	<ul style="list-style-type: none">• навыками поиска и проведения исследований на основе больших данных

ПК-4 - проведение анализа инноваций в экономике, управлении и информационно-коммуникативных технологиях	
Знать	<ul style="list-style-type: none">• основные и специальные (передовые) методы сбора и обработки больших данных для проведения углубленного анализа инноваций в экономике
Уметь	<ul style="list-style-type: none">• использовать основные и специальные (передовые) методы сбора и обработки больших данных для проведения углубленного анализа инноваций в экономике
Владеть	<ul style="list-style-type: none">• основными и специальными (передовыми) методами сбора больших данных для проведения углубленного анализа инноваций в экономике

4 Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 4 зачетных единиц 144 акад. часов, в том числе:

- контактная работа – 89 акад. часов:
 - аудиторная – 85 акад. часов;
 - внеаудиторная – 4 акад. часов
- самостоятельная работа – 19,3 акад. часов;
- подготовка к экзамену – 35,7 акад. часа

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа (в акад. часах)	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код и структурный элемент компетенции
		лекции	лаборат. занятия	практич. занятия				
1. Введение в анализ больших данных. Предобработка данных								
1.1. Жизненный цикл аналитики данных. Методы анализа больших данных. Обзор программных средств для анализа больших данных	4	4	4/2И		2	Выполнение лабораторных заданий	Отчет по лабораторной работе	ОПК-1 – зув ПК-2-зув ПК-3-зув
1.2. Методы поиска и сбора больших данных. Web-scraping. Технологии хранения больших данных	4	4	6/4И		2	Выполнение лабораторных заданий	Отчет по лабораторной работе	ОПК-1 – зув ПК-2-зув ПК-3-зув
1.3. Методы обработки пропусков и дубликатов данных. Приведение типов данных. Технологии парсинга данных.	4	4	8/4И		2	Выполнение лабораторных заданий	Отчет по лабораторной работе	ОПК-1 – зув ПК-2-зув ПК-3-зув
1.4. Категоризация данных. Лемматизация, стемминг данных	4	2	6/2И		2	Выполнение лабораторных заданий	Отчет по лабораторной работе	ОПК-1 – зув ПК-2-зув ПК-3-зув
Итого по разделу		14	24/12И		8		Тестирование	
2. Визуальный анализ данных								
2.1. Методы объединения данных.	4	4	2/2И		2	Выполнение лабораторных заданий, изучение литературы	Отчет по лабораторной работе	ОПК-1 -зув

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа (в акад. часах)	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код и структурный элемент компетенции
		лекции	лаборат. занятия	практич. занятия				
2.2. Визуализация данных. Инструменты построения графиков. Гистограммы, диаграммы (с областями, столбчатые, точечные, пузырьковые, лепестковые и др), «ящики с усами». Библиотеки Seaborn, Bokeh	4	2	10/2И		0,3	Выполнение лабораторных заданий, изучение литературы	Отчет по лабораторной работе	ОПК-1 –зுவ ПК-2-зுவ ПК-3-зுவ
2.3. Анализ и визуализация геоданных. Библиотека folium, интерактивная визуализация геоданных.	4	6	6/2И		2	Выполнение лабораторных заданий, изучение литературы	Отчет по лабораторной работе	ОПК-1 –зув ПК-2-зув ПК-3-зув
2.4. Визуализация графов и сетевых структур. Библиотека networkx. Gephi.	4	4	3/2И		3	Выполнение лабораторных заданий, изучение литературы	Отчет по лабораторной работе	ОПК-1 –зув ПК-2-зув ПК-3-зув
Итого по разделу		16	21/10И		11,3		Тестирование	
3. Анализ больших данных социальных сетей								
3.1. Анализ сетей. Степенной закон. Меры сосредоточенности и ранжирование по сети	4	2	3			Выполнение лабораторных заданий, изучение литературы	Отчет по лабораторной работе	ОПК-1 –зув ПК-2-зув ПК-3-зув
3.2. Анализ графов сообществ социальной сети		2	3			Выполнение лабораторных заданий, изучение литературы	Отчет по лабораторной работе	ОПК-1 –зув ПК-2-зув ПК-3-зув
Итого по разделу		4	6				Тестирование	
Итого за семестр		34	51/22И		19,3		Экзамен	
Итого по дисциплине		34	51/22И		19,3		Экзамен	

5 Образовательные технологии

В ходе проведения занятий предусматриваются следующие образовательные технологии:

1. Информационная лекция – последовательное изложение материала в дисциплинарной логике, осуществляемое преимущественно вербальными средствами (монолог преподавателя).
2. Лабораторная работа – организация учебной работы с реальными материальными и информационными объектами, экспериментальная работа с аналоговыми моделями реальных объектов.
3. Технологии проблемного обучения – организация образовательного процесса, которая предполагает постановку проблемных вопросов, создание учебных проблемных ситуаций для стимулирования активной познавательной деятельности студентов.
4. Лабораторное занятие на основе кейс-метода – обучение в контексте моделируемой ситуации, воспроизводящей реальные условия научной, производственной, общественной деятельности. Обучающиеся должны проанализировать ситуацию, разобраться в сути проблем, предложить возможные решения и выбрать лучшее из них. Кейсы базируются на реальном фактическом материале или же приближены к реальной ситуации.
5. Информационно-коммуникационные образовательные технологии – организация образовательного процесса, основанная на применении специализированных программных сред и технических средств работы с информацией.
6. Формы учебных занятий с использованием информационно-коммуникационных технологий:
7. Лекция-визуализация – изложение содержания сопровождается презентацией (демонстрацией учебных материалов, представленных в различных знаковых системах, в т.ч. иллюстративных, графических, аудио- и видеоматериалов).
8. Лабораторное занятие в форме презентации – представление результатов проектной или исследовательской деятельности с использованием специализированных программных сред.

В рамках практических занятий предусматривается использование средств вычислительной техники при выполнении индивидуальных заданий. Используется существующий образовательный портал университета (newlms.magt.u.ru) для размещения ЭУМК по дисциплине. Текущий, промежуточный и рубежный контроль проводится на образовательном портале университета.

6 Учебно-методическое обеспечение самостоятельной работы обучающихся

Представлено в приложении 1.

7 Оценочные средства для проведения промежуточной аттестации

Представлены в приложении 2.

8 Учебно-методическое и информационное обеспечение дисциплины (модуля)

а) Основная литература:

1. Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для вузов / Д. Ю. Федоров. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2020. — 161 с. — (Высшее образование). — ISBN 978-5-534-10971-9. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/454100>
2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2020. — 174 с. — (Высшее образование). — ISBN 978-5-9916-5009-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450262>

б) Дополнительная литература:

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2020. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450166> (дата обращения: 11.05.2020).
2. Журнал «Программные продукты и системы» [Электронный ресурс]. Научно-исследовательский институт «Центрпрограммсистем» — Режим доступа: https://e.lanbook.com/journal/2276#journal_name — Загл. с экрана.

в) Методические указания:

Методические указания по выполнению индивидуальных домашних заданий представлены в приложении 3

г) Программное обеспечение и Интернет-ресурсы:

Программное обеспечение

Наименование ПО	№ договора	Срок действия лицензии
MS Windows 7 Professional(для классов)	Д-1227-18 от 08.10.2018	11.10.2021
MS Office 2007 Professional	№ 135 от 17.09.2007	бессрочно
7Zip	свободно распространяемое ПО	бессрочно
Anaconda Python, Jupyter Notebook (anaconda)	свободно распространяемое ПО	бессрочно
Google Chrome	свободно распространяемое ПО	бессрочно
Браузер Mozilla Firefox	свободно распространяемое ПО	бессрочно

Профессиональные базы данных и информационные справочные системы

Название курса	Ссылка
Поисковая система Академия Google (Google Scholar)	URL: https://scholar.google.ru/
Информационная система - Единое окно доступа к информационным ресурсам	URL: http://window.edu.ru/

9 Материально-техническое обеспечение дисциплины (модуля)

Материально-техническое обеспечение дисциплины включает:

Тип и название аудитории	Оснащение аудитории
Учебные аудитории для проведения занятий лекционного типа	Персональный компьютер (или ноутбук) с пакетом MS Office, с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Доска, мультимедийный проектор, экран. Мультимедийные презентации к лекциям, учебно-наглядные пособия
Учебные аудитории для проведения лабораторных занятий	Персональные компьютеры с пакетом MS Office; Anaconda Python, Jupyter Notebook (anaconda), с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Браузер Google Chrome.
Аудитории для самостоятельной работы: компьютерные классы; читальные залы библиотеки	Персональные компьютеры с пакетом MS Office; Anaconda Python, Jupyter Notebook (anaconda), с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Браузер Google Chrome.
Аудитории для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации	Персональные компьютеры с пакетом MS Office и выходом в Интернет и с доступом в электронную информационно-образовательную среду университета. Anaconda Python, Jupyter Notebook (anaconda); Браузер Google Chrome.
Аудитория для хранения и профилактического обслуживания учебного оборудования	Мебель для хранения и обслуживания оборудования (шкафы, столы), учебно-методические материалы, компьютеры, ноутбуки, принтеры.

Учебно-методическое обеспечение самостоятельной работы обучающихся

По дисциплине «Программирование на Python» предусмотрена аудиторная и внеаудиторная самостоятельная работа обучающихся.

Аудиторная самостоятельная работа студентов предполагает выполнение заданий по лабораторным работам, подготовка к тестированию.

Для выполнения лабораторных работ вам будут полезны следующие ресурсы:

- Большие данные и всё о них - <https://habr.com/ru/hub/bigdata/>
- Сайт компании Basegroup Labs- <https://basegroup.ru/>
- Портал открытых данных РФ - <https://data.gov.ru/>

Предметные области для определения темы проекта:

- Работа с твитами – эмодзи
- Данные с росстата
- Данные с центробанка
- Данные по экологии
- Данные по стоимости недвижимости
- Др.

Примечание. При выборе предметной области, продумайте возможность использования геоданных, поскольку во второй половине курса потребуются их обработка.

Лабораторная работа №1

Тема «Сбор и хранение данных»

Задания:

1. Выберите предметную область для реализации дальнейшего проекта:
 - Работа с твитами – эмодзи
 - Данные с росстата
 - Данные с центробанка
 - Данные по экологии
 - Данные по стоимости недвижимости
 - Др.
2. изучите строение HTML-страниц, правила работы GET-запросов;
3. составьте несколько запросов к сайтам и соберите данные.
4. примените методы автоматизированного веб-скрейпинга при сборе данных.

Лабораторная работа №2

Тема «Предобработка данных»

Задания:

1. По выбранной предметной области необходимо подготовить датафрейм.
2. Полученный датафрейм обработать на предмет пропусков, привести типы данных в соответствие, найти и убрать дубликаты данных; привести к одному виду названия столбцов.
3. Для анализа текстовых данных примените методы стемминга и лемматизации.

Лабораторная работа №3

Тема «Визуализация»

Задания:

1. Изучите методы объединения данных в Python. Реализуйте объединение данных в вашем датафрейме для решения конкретной задачи.
2. По полученным данным датафрейма постройте графики, используя методы: `hist()`, `boxplot()`, `plot()`.
3. *Проанализируйте полученные результаты, какие выводы можно сделать по построенным графикам?*

Лабораторная работа №4

Тема «Числовые показатели»

Задания:

1. Найдите среднее, медиану, количество наблюдений в выборке и разброс их значений.
2. Какие выводы можно сделать о данных в датафрейме по полученным показателям?

Лабораторная работа №5

Тема «Геоданные»

Задания:

1. Рассмотрите проблему обработки и визуализации геоданных.
2. Получите геоданные по одному ваших столбцов датафрейма.
3. Проведите группировку данных, решая некоторую задачу.
4. Постойте интерактивную карту, используя библиотеку folium.

Лабораторная работа №6

Тема «Визуализация графов»

Задания:

1. Создайте полносвязный граф. Полносвязный граф — граф, где каждая вершина соединена с каждой другой.
2. С использованием библиотеки networkx постройте визуализацию данного графа.

Лабораторная работа № 7

Тема «Анализ данных социальных сетей»

Задания:

1. Подключитесь к социальной сети Вконтакте и проанализируйте свой социальный граф

7 Оценочные средства для проведения промежуточной аттестации

а) Планируемые результаты обучения и оценочные средства для проведения промежуточной аттестации:

Структурный элемент компетенции	Планируемые результаты обучения	Оценочные средства
ОПК-1 - способностью решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности		
Знать	<ul style="list-style-type: none"> • методы решения задач предобработки, обработки и анализа больших данных; • методы визуализации больших данных; • возможности высокопроизводительных вычислительных систем; • технологии распределенных вычислений, методы и модели Data Mining. 	<p>Вопросы к экзамену:</p> <ol style="list-style-type: none"> 1. Методы анализа больших данных. 2. Обзор программных средств для анализа больших данных. 3. Технологии хранения больших данных. 4. Методы и подходы к предобработке больших данных 5. Технология парсинга больших данных. 6. Технология веб-скрейпинга. 7. Методы визуализации данных.
Уметь	<ul style="list-style-type: none"> • разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных; • оценивать время и необходимые аппаратные ресурсы для решения задач анализа и обработки данных; • создавать алгоритмы анализа и обработки большого объема данных с применением моделей Data Mining. 	<p><i>Практические задания:</i></p> <ol style="list-style-type: none"> 1. Получите датафрейм, используя API любой социальной сети. 2. Получите статистические данные по датафрейму. 3. Обработайте пропуски и дубликаты в датафрейме 4. Постройте диаграмму – ящик с усами по отдельным столбцам датафрейма

Структурный элемент компетенции	Планируемые результаты обучения	Оценочные средства
Владеть	<ul style="list-style-type: none"> • навыками применения программных систем, предназначенных для анализа больших данных. 	<p><i>Проектное задание:</i> Получите данные по стоимости недвижимости отдельного города за несколько лет (не менее 3 лет). На основе полученных данных ответьте на следующие вопросы:</p> <ol style="list-style-type: none"> 1. Какова тенденция изменения стоимости недвижимости? 2. Как стоимость недвижимости отличается внутри города, от каких показателей зависит, есть ли корреляция между отдельными показателями? 3. Какими данными необходимо дополнить датафрейм? Визуализируйте полученные данные.
ПК-2 - проведение исследования и анализа рынка информационных систем и информационно-коммуникативных технологий		
Знать	<ul style="list-style-type: none"> • методы анализа больших данных; • сервисы хранения больших данных; 	<p>Вопросы к экзамену:</p> <ol style="list-style-type: none"> 1. Методики анализа больших данных. 2. Дайте характеристику Big Data на мировом рынке. 3. Охарактеризуйте Big Data в России. 4. Определите понятие Data Mining. 5. Методы поиска и сбора больших данных. 6. Технологии хранения больших данных
Уметь	<ul style="list-style-type: none"> • осуществлять поиск больших данных в сети Интернет; • проводить исследования на основе анализа больших данных 	<p><i>Практическое задание:</i></p> <ol style="list-style-type: none"> 1. Реализуйте чтение данных из таблиц Excel 2. Реализуйте чтение данных из файла csv 3. Получите данные с помощью GET-запроса. 4. Проанализируйте данные с помощью библиотеки Beautiful Soup

Структурный элемент компетенции	Планируемые результаты обучения	Оценочные средства
Владеть	<ul style="list-style-type: none"> • навыками поиска и проведения исследований на основе больших данных 	<p><i>Проектное задание:</i> Получите данные по стоимости недвижимости отдельного города за несколько лет (не менее 3 лет). На основе полученных данных ответьте на следующие вопросы:</p> <ol style="list-style-type: none"> 1. Какова тенденция изменения стоимости недвижимости? 2. Как стоимость недвижимости отличается внутри города, от каких показателей зависит, есть ли корреляция между отдельными показателями? 3. Какими данными необходимо дополнить датафрейм? 4. Визуализируйте полученные данные.
ПК-4 - проведение анализа инноваций в экономике, управлении и информационно-коммуникативных технологиях		
Знать	<ul style="list-style-type: none"> • основные и специальные (передовые) методы сбора и обработки больших данных для проведения углубленного анализа инноваций в экономике 	<p>Вопросы к экзамену:</p> <ol style="list-style-type: none"> 1. Методы поиска и сбора больших данных. Web-scraping 2. Технологии хранения больших данных 3. Сервисы для получения больших данных. Сервис открытые данные РФ. 4. Зарубежные сервисы больших данных. 5. Средства получения и визуализации геоданных. 6. Данные социальных сетей. Возможности для бизнеса.
Уметь	<ul style="list-style-type: none"> • использовать основные и специальные (передовые) методы сбора и обработки больших данных для проведения углубленного анализа инноваций в экономике 	<p><i>Практическое задание:</i> Составьте обзор сервисов, хранящих большие данные (не менее 10 сервисов). Заполните таблицу: название сервиса, URL, описание данных, форматы хранения, методы доступа к данным.</p>

Структурный элемент компетенции	Планируемые результаты обучения	Оценочные средства
Владеть	<ul style="list-style-type: none">• основными и специальными (передовыми) методами сбора больших данных для проведения углубленного анализа инноваций в экономике	<p><i>Проектное задание:</i> Получите данные по стоимости недвижимости отдельного города за несколько лет (не менее 3 лет). На основе полученных данных постройте карту с визуализацией дорогих и дешевых городов с точки зрения стоимости в них недвижимости.</p>

б) Порядок проведения промежуточной аттестации, показатели и критерии оценивания:

Промежуточная аттестация по дисциплине «Анализ больших данных» включает теоретические вопросы, позволяющие оценить уровень усвоения обучающимися знаний, и практические задания, выявляющие степень сформированности умений и владений, проводится в форме экзамена.

Экзамен по данной дисциплине проводится в устной форме по экзаменационным билетам, каждый из которых включает 2 теоретических вопроса и одно практическое задание.

Показатели и критерии оценивания экзамена:

– на оценку **«отлично»** (5 баллов) – обучающийся демонстрирует высокий уровень сформированности компетенций, всестороннее, систематическое и глубокое знание учебного материала, свободно выполняет практические задания, свободно оперирует знаниями, умениями, применяет их в ситуациях повышенной сложности.

– на оценку **«хорошо»** (4 балла) – обучающийся демонстрирует средний уровень сформированности компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.

– на оценку **«удовлетворительно»** (3 балла) – обучающийся демонстрирует пороговый уровень сформированности компетенций: в ходе контрольных мероприятий допускаются ошибки, проявляется отсутствие отдельных знаний, умений, навыков, обучающийся испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.

– на оценку **«неудовлетворительно»** (2 балла) – обучающийся демонстрирует знания не более 20% теоретического материала, допускает существенные ошибки, не может показать интеллектуальные навыки решения простых задач.

– на оценку **«неудовлетворительно»** (1 балл) – обучающийся не может показать знания на уровне воспроизведения и объяснения информации, не может показать интеллектуальные навыки решения простых задач.

Методические по выполнению индивидуальных домашних заданий

Цель методических рекомендаций - обеспечить студенту оптимальную организацию процесса изучения дисциплины, а также выполнения различных форм самостоятельной работы.

Студентам необходимо ознакомиться: с содержанием рабочей программы дисциплины (далее - РПД), с целями и задачами дисциплины, ее связями с другими дисциплинами образовательной программы, методическими разработками по данной дисциплине, имеющимся на образовательном портале и сайте кафедры, с графиком консультаций преподавателей кафедры.

Инструкции по работе с Anaconda

Anaconda — это дистрибутивы Python и R. Он предоставляет все необходимое для решения задач по анализу и обработке данных (с применимостью к Python).

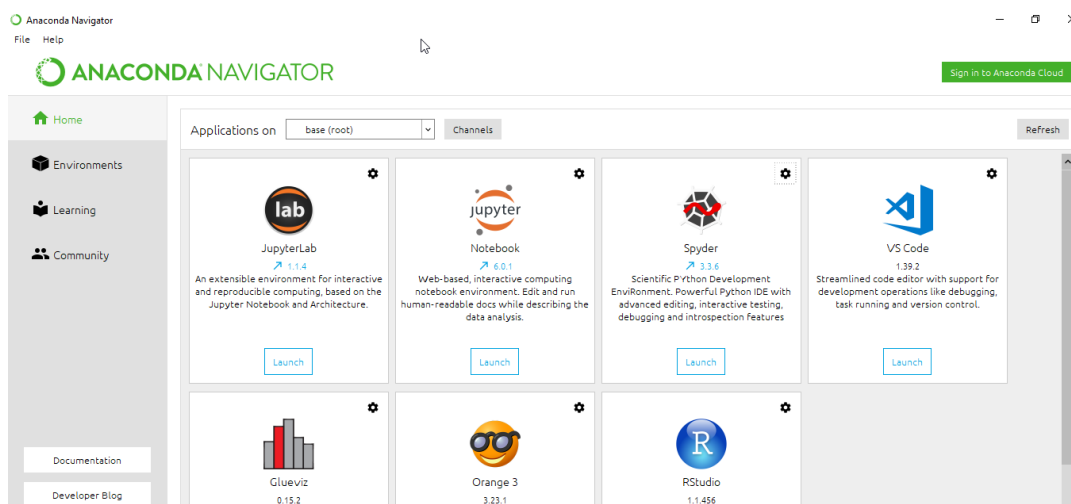
Anaconda — это набор бинарных систем, включающий в себя Scipy, Numpy, Pandas и их зависимости.

Scipy — это пакет статистического анализа.

Numpy — это пакет числовых вычислений.

Pandas — уровень абстракции данных для объединения и преобразования данных.

Anaconda Navigator — это графический интерфейс пользователя на рабочем столе (GUI), включенный в дистрибутив Anaconda, который позволяет запускать приложения и легко управлять пакетами, средами и каналами conda без использования команд командной строки. Навигатор может искать пакеты в Anaconda Cloud или в локальном репозитории Anaconda. Он доступен для Windows, MacOS и Linux.



JupyterLab — это интерактивная среда разработки для работы с блокнотами, кодом и данными.

Jupyter Notebook — удобный инструмент для создания красивых аналитических отчетов, позволяет хранить вместе код, изображения, комментарии, формулы и графики. Работа ведется в браузере.

Spyder — интерактивной IDE для научных расчетов на языке Python. Данная IDE позволяет писать, редактировать и тестировать код. Spyder предлагает просмотр и редактирование переменных с помощью GUI, динамическую интроспекцию кода, нахождение ошибок на лету и многое другое. Также, по необходимости, можно интегрировать Anaconda с другими Python IDE, включая PyCharm и Atom.

VS Code — это оптимизированный редактор кода с поддержкой таких операций разработки, как отладка, запуск задач и контроль версий.

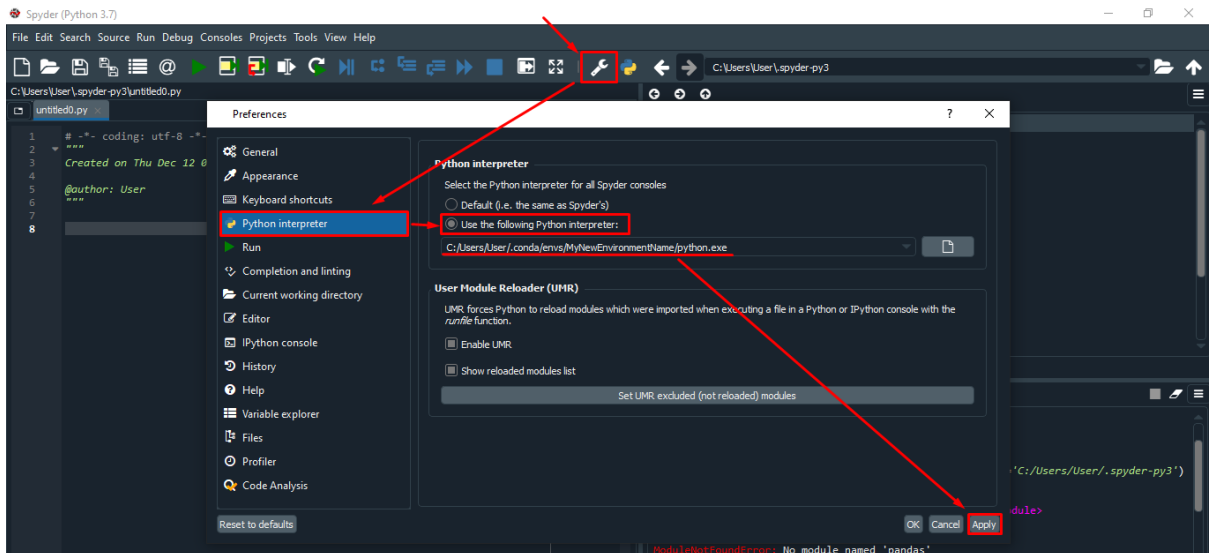
Glueviz — используется для визуализации многомерных данных в файлах. Он исследует отношения внутри и между связанными наборами данных.

Orange 3 — это основанная на компонентах структура интеллектуального анализа данных. Это может быть использовано для визуализации данных и анализа данных. Рабочие процессы в Orange 3 очень интерактивны и предоставляют большой набор инструментов.

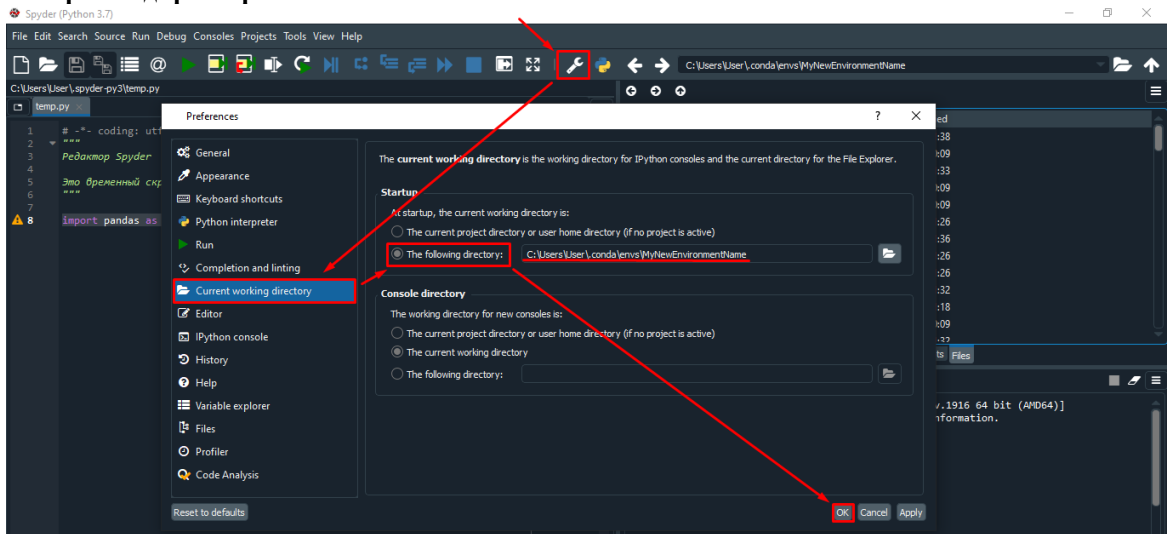
RStudio — это набор интегрированных инструментов, предназначенных для повышения продуктивности работы с R. Он включает в себя основы R и Notebooks.

Настройка среды для Spyder

1. Настройка интерпретатора

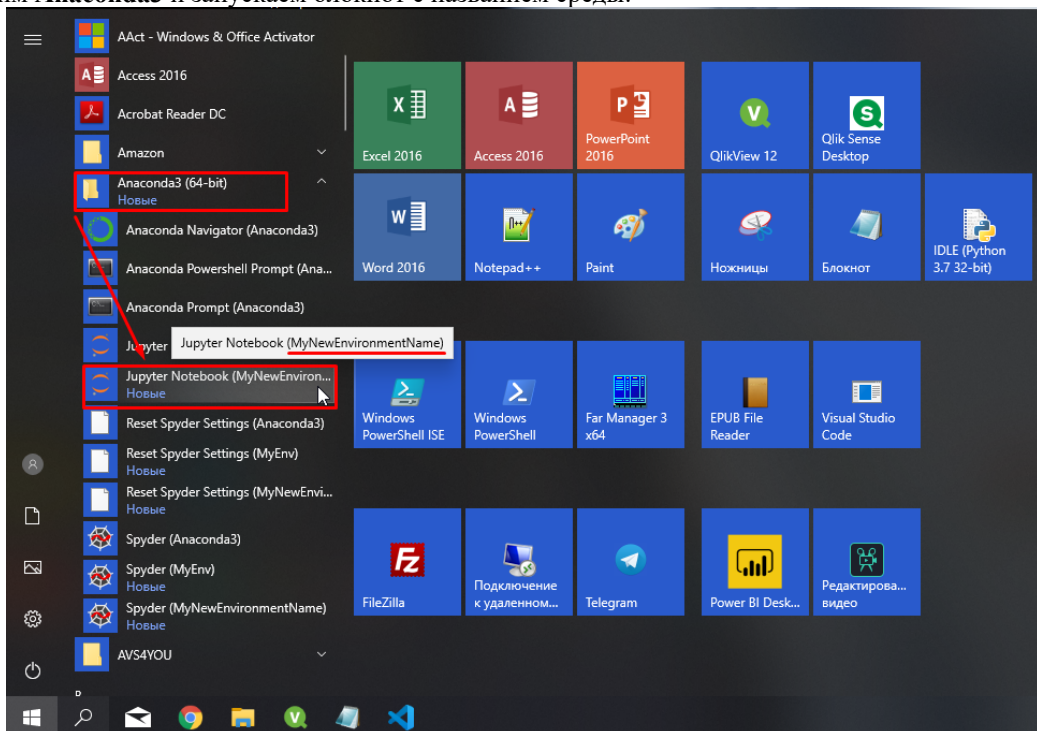


Настройка директории



Как открыть Jupyter Notebook в новой среде MyNewEnvironmentName

Для того, чтобы запустить **Jupyter Notebook** в созданной среде **MyNewEnvironmentName**, в поиске находим **Anaconda3** и запускаем блокнот с названием среды:

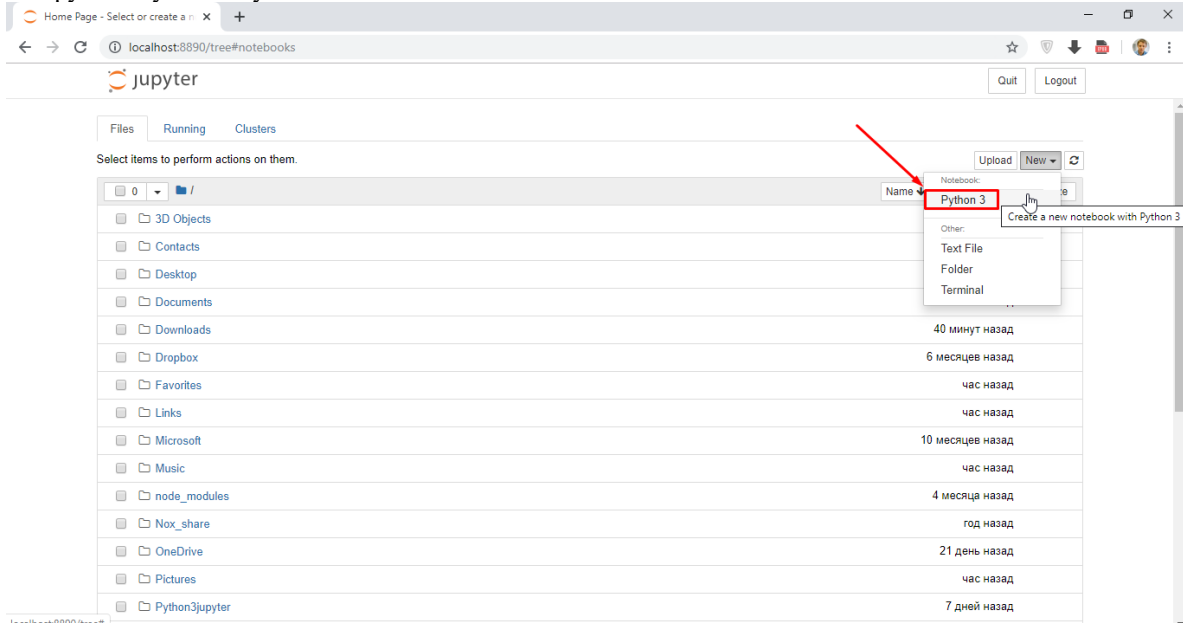


Появится консольное окошко — это движок Jupyter Notebook, который работает в фоновом режиме:

```
Jupyter Notebook (MyNewEnvironmentName)
[I 23:57:47.710 NotebookApp] The port 8888 is already in use, trying another port.
[I 23:57:47.710 NotebookApp] The port 8889 is already in use, trying another port.
[I 23:57:47.757 NotebookApp] Serving notebooks from local directory: C:\Users\User
[I 23:57:47.757 NotebookApp] The Jupyter Notebook is running at:
[I 23:57:47.757 NotebookApp] http://localhost:8890/?token=57cb326d88828682f9384226df11d58fff936a7547108cc6
[I 23:57:47.757 NotebookApp] or http://127.0.0.1:8890/?token=57cb326d88828682f9384226df11d58fff936a7547108cc6
[I 23:57:47.757 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 23:57:47.969 NotebookApp]

To access the notebook, open this file in a browser:
file:///C:/Users/User/AppData/Roaming/jupyter/runtime/nbserver-9088-open.html
Or copy and paste one of these URLs:
http://localhost:8890/?token=57cb326d88828682f9384226df11d58fff936a7547108cc6
or http://127.0.0.1:8890/?token=57cb326d88828682f9384226df11d58fff936a7547108cc6
```

В Jupyter запускаем Python 3:



Для того, чтобы убедиться в какой среде мы работаем, можно вбить ряд команд (ниже приведен текст этих команд для Python 3):

```
Jupyter MyNewEnvironmentName_Python_Example Last Checkpoint: несколько секунд назад (unsaved changes)
File Edit View Insert Cell Kernel Help Trusted Python 3

In [15]: import sys
         print(sys.version)
         print(sys.base_prefix)

3.7.5 (default, Oct 31 2019, 15:18:51) [MSC v.1916 64 bit (AMD64)]
C:\Users\User\.conda\envs\MyNewEnvironmentName

In [16]: print('\n'.join(sys.modules.keys()))

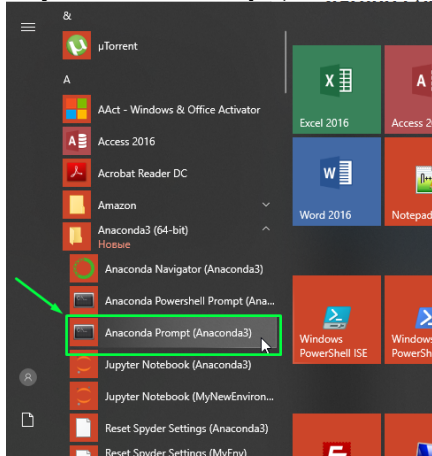
sys
builtins
_frozen_importlib
_imp
_thread
_warnings
_weakref
zipimport
_frozen_importlib_external
_io
marshal
nt
winreg
encodings
codecs
codecs
encodings.aliases
encodings.utf_8
signal
```

Узнать среду, в которой работает Jupyter Notebook:

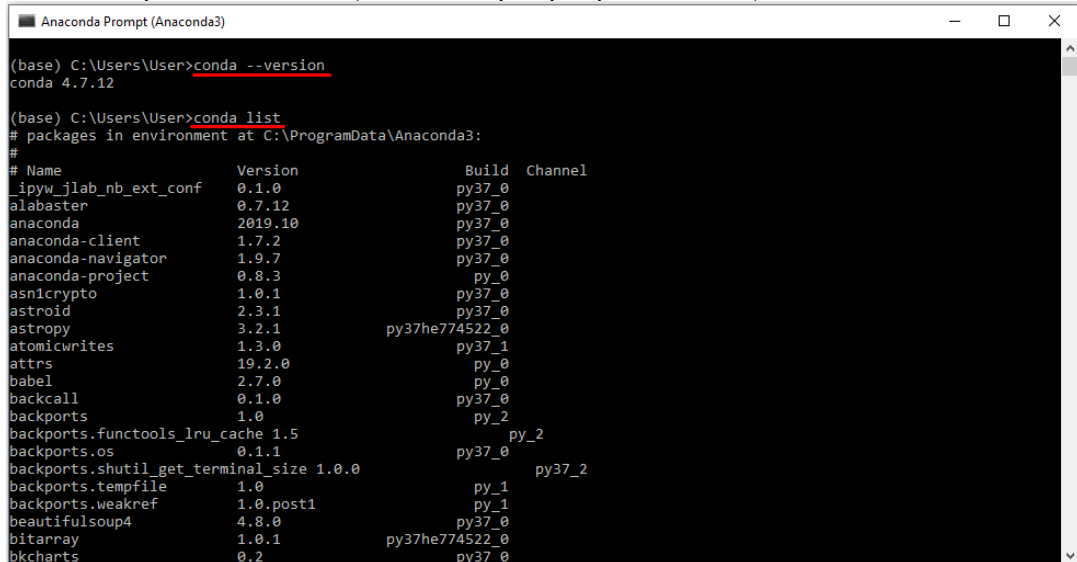
```
import sys
print(sys.version)
print(sys.base_prefix)
Получить список модулей, доступных в Env:
print('\n'.join(sys.modules.keys()))
```

Anaconda3 Prompt cmd Conda Command — Запуск команд через консоль

Для того, чтобы использовать команды conda через командную строку (cmd), необходимо запустить программу **Anaconda Prompt (Anaconda3)**

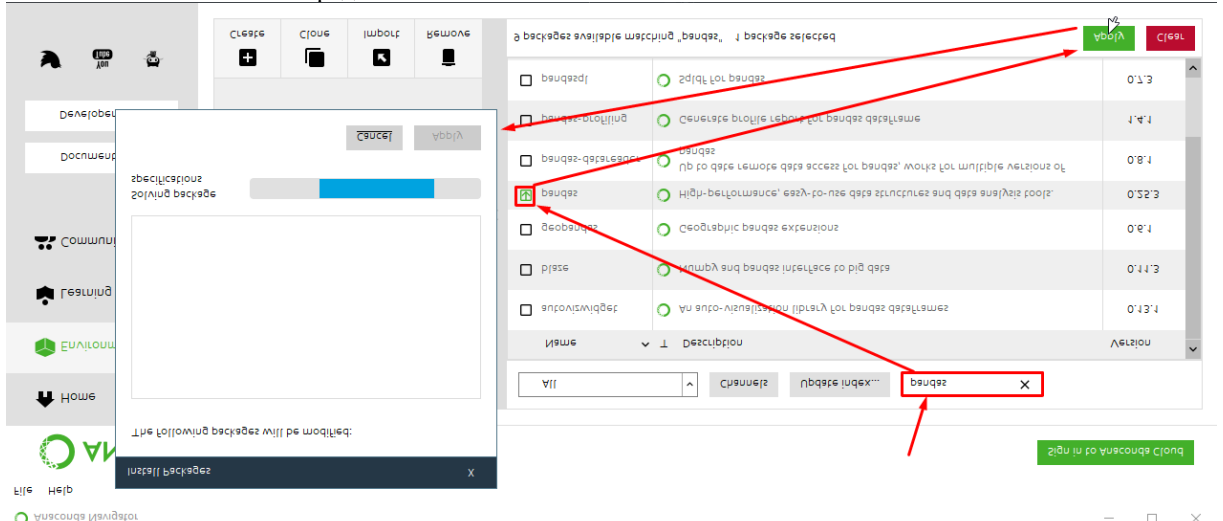


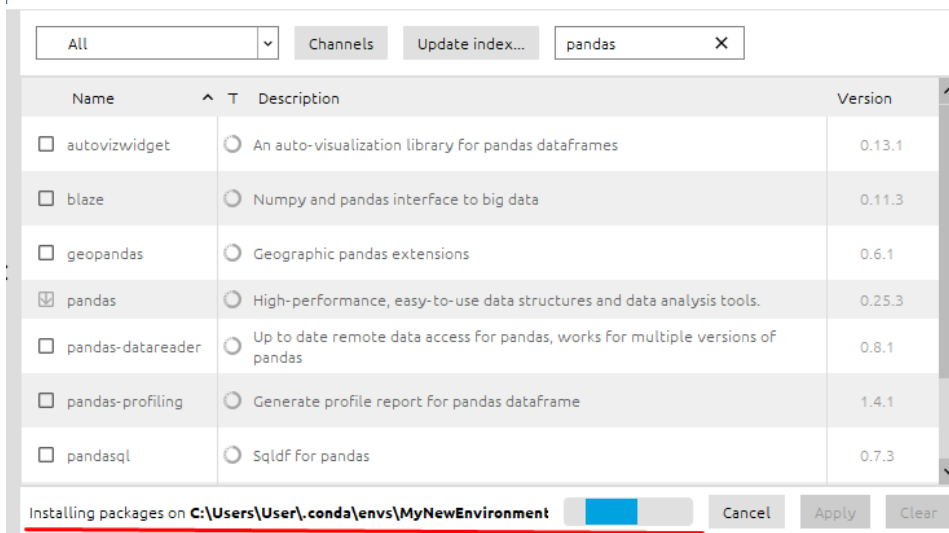
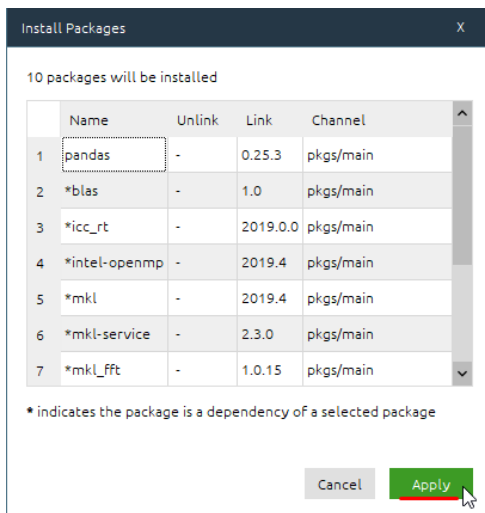
Можете набрать две команды (в качестве проверки работы conda):



Установка новой библиотеки (пакета) в среду

Пакеты управляются отдельно для каждой среды. Изменения, которые вы вносите в пакеты, применяются только к активной среде.





Исполняемые файлы в среде Conda

python.exe — исполняемый файл Python для приложений командной строки. Так, например, если вы находитесь в каталоге Example App, вы можете выполнить его: `python.exe exampleapp.py`

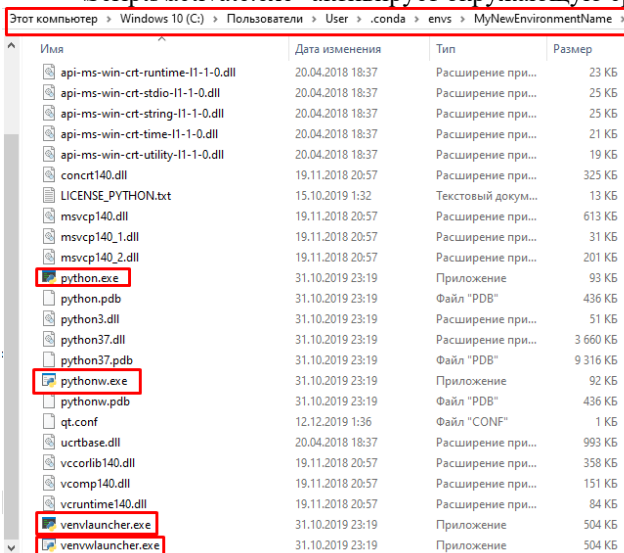
pythonw.exe — исполняемый файл Python для приложений с графическим интерфейсом или приложений без интерфейса пользователя

venvlauncher.exe —

venvwlauncher.exe —

`\Scripts` — исполняемые файлы, являющиеся частью установленных пакетов. После активации среды этот каталог добавляется в системный путь, поэтому исполняемые файлы становятся доступными без их полного пути.

`\Scripts\activate.exe` - активирует окружающую среду



Источники:

- <https://protostar.space/why-you-need-python-environments-and-how-to-manage-them-with-conda>
- https://kapeli.com/cheat_sheets/Conda.docset/Contents/Resources/Documents/index