

**Федеральное государственное бюджетное учреждение науки
Институт машиноведения им. А.А. Благонравова
Российской академии наук
(ИМАШ РАН)**

УТВЕРЖДАЮ
Заместитель директора по вопросам
подготовки кадров высшей
квалификации ИМАШ РАН
« 01 » _____ 2026 года
к.ф.-м.н. Жданов Э.Р.
М.П.



**Комплект оценочных материалов
по дисциплине «Применение языка Python для сбора и анализа данных»**

Направление подготовки:	27.04.07 Научно-технические технологии и экономика инноваций
Профиль подготовки:	Инновационное проектирование и разработка технологий
Квалификация:	Магистр
Форма обучения	Очная

Москва

Содержание

1. Общие положения.....	3
2. Результаты освоения дисциплины, подлежащие проверке.....	3
3. Контроль и оценка освоения учебной дисциплины.....	3
4. Оценочные материалы для проведения текущего контроля успеваемости.....	4
5. Оценочные материалы для проведения промежуточной аттестации.....	6

1. Общие положения

Комплект оценочных материалов (далее - ОМ) предназначен для контроля и оценки образовательных достижений обучающихся, освоивших программу дисциплины «Применение языка Python для сбора и анализа данных».

ОМ включают контрольные материалы для проведения текущего контроля успеваемости и промежуточной аттестации в форме зачета.

ОМ разработаны на основании:

- основной профессиональной образовательной программы по направлению подготовки 27.04.07 Научноёмкие технологии и экономика инноваций;
- рабочей программы дисциплины «Применение языка Python для сбора и анализа данных».

2. Результаты освоения дисциплины, подлежащие проверке

В результате освоения дисциплины обучающийся должен обладать следующими компетенциями

ПК-2	Способен организовывать и координировать сбор, обработку и анализ информации по результатам деятельности машиностроительной организации
ПК-2.1	Способен контролировать и анализировать основные показатели деятельности организации

Код индикатора компетенции	Планируемые результаты обучения по дисциплине
ПК-2.1.	<i>Знать:</i> - методы автоматизации сбора и очистки данных для анализа; - принципы визуализации данных для контроля показателей. <i>Уметь:</i> - писать скрипты на Python для сбора, расчёта и анализа KPI. - находить и интерпретировать отклонения фактических показателей от плановых; - создавать наглядные отчёты и дашборды для мониторинга деятельности. <i>Иметь практический опыт:</i> - разработки на Python системы автоматизированного мониторинга KPI для проекта/подразделения; - проведения анализа отклонений экономических показателей с помощью Python; - создания автоматизированного отчёта для оценки эффективности внедрённых изменений.

3. Контроль и оценка освоения учебной дисциплины

Элемент учебной дисциплины	Текущий контроль		Промежуточная аттестация	
	Форма контроля	Проверяемые идк	Форма контроля	Проверяемые идк
Тема 1. Основы Python и работа с табличными данными	Отчет по лабораторному практикуму	ПК-2.1.		
Тема 1. Основы Python и работа с табличными данными	Отчет по практикуму по решению задач	ПК-2.1.		
Тема 2. Визуализация данных и базовый анализ	Отчет по лабораторному практикуму	ПК-2.1.		
Тема 2. Визуализация данных и базовый анализ	Отчет по практикуму по решению задач	ПК-2.1.		

Тема 3. Автоматизация сбора данных	Отчет по лабораторному практикуму	ПК-2.1.		
Тема 3. Автоматизация сбора данных	Отчет по практикуму по решению задач	ПК-2.1.		
Тема 4. Анализ КРП и оценка эффективности изменений	Отчет по лабораторному практикуму	ПК-2.1.		
Тема 4. Анализ КРП и оценка эффективности изменений	Отчет по практикуму по решению задач	ПК-2.1.		
			Зачет	ПК-2.1.

4. Оценочные материалы для проведения текущего контроля успеваемости

Показатели и критерии оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Практикум по решению задач
10-8 – работа и отчет выполнены в срок, самостоятельно, правильно определены соответствующие спецификации, использована требуемая информация, правильно выполнены требуемые расчеты, сделаны необходимые выводы, хорошо аргументированы, даны исчерпывающие ответы на все поставленные вопросы; 7-5 – работа и отчет выполнены в срок, самостоятельно, правильно определены соответствующие спецификации, использована требуемая информация, правильно выполнены требуемые расчеты, необходимые выводы сделаны частично, хорошо аргументированы, даны ответы на все поставленные вопросы; 4-1 – работа и отчет выполнены в срок, в основном самостоятельно, выводы сделаны частично, слабо аргументированы, даны ответы не на все вопросы; 0 – обучающийся подготовил работу и отчет несамостоятельно и/или не завершил в срок, описание содержит ошибки, выводы и ответы на вопросы отсутствуют.
Лабораторный практикум
15-12 – работа выполнена в срок, самостоятельно, правильно поняты и использованы соответствующие формулы, правильно определены соответствующие спецификации, использована требуемая информация, правильно выполнены требуемые расчеты, правильно выбраны совместимые комплектующие, сделаны необходимые выводы, хорошо аргументированы, даны исчерпывающие ответы на все поставленные вопросы; 11-8 – работа выполнена в срок, самостоятельно, правильно поняты и использованы соответствующие формулы, правильно определены соответствующие спецификации, использована требуемая информация, правильно выполнены требуемые расчеты, правильно выбраны совместимые комплектующие, необходимые выводы сделаны частично, хорошо аргументированы, даны ответы на все поставленные вопросы; 7-4 – работа выполнена в срок, в основном самостоятельно, использованы соответствующие формулы; определены соответствующие спецификации, имеются ошибки в расчетах; выбраны совместимые комплектующие необходимые, выводы сделаны частично, слабо аргументированы, даны ответы не на все вопросы; 3-1 – обучающийся подготовил работу несамостоятельно или не завершил в срок, описание спецификации содержит незначительные ошибки, выводы и ответы на вопросы отсутствуют.

Порядок начисления баллов

Если студент набрал достаточное количество баллов ТКУ по дисциплине (более 70 баллов) в течение семестра, он может отказаться от сдачи зачета, и результат будет выставлен автоматически по его согласию.

Типовые задания для лабораторных практикумов

Тема 1. Основы Python и работа с табличными данными

Лабораторный практикум № 1. Работа в Jupyter Notebook

Задание: «Очистка производственного отчета»

Студенту предоставляется «грязный» Excel-файл с данными о работе цеха за месяц. В нем есть пустые ячейки, дубликаты строк, опечатки в названиях материалов, смешанный формат дат.

Что нужно сделать:

Загрузить данные в Pandas DataFrame.

Найти и удалить дубликаты.

Обработать пропуски (заполнить или удалить).

Привести столбец «Дата» к единому формату.

Переименовать столбцы для удобства.

Сохранить очищенный датасет в новый файл.

Тема 2. Визуализация данных и базовый анализ

Лабораторный практикум № 2. Работа с библиотеками Matplotlib/Seaborn

Задание: «Построение панели мониторинга (Dashboard)»

Дан чистый датасет с данными о продажах разных продуктов по месяцам.

Что нужно сделать:

Построить линейный график динамики общих продаж по месяцам.

Построить столбчатую диаграмму сравнения продаж разных категорий товаров за последний квартал.

Создать диаграмму рассеяния для анализа связи между ценой товара и объемом его продаж.

Тема 3. Автоматизация сбора данных

Лабораторный практикум № 3. Работа с библиотеками Requests/BeautifulSoup

Задание: «Сбор цен конкурентов»

Дана ссылка на страницу интернет-магазина конкурента, где представлен список товаров с ценами.

Что нужно сделать:

С помощью библиотеки requests получить HTML-код страницы.

Используя BeautifulSoup, найти в коде названия товаров и их цены.

Сохранить собранные данные (название, цена) в CSV-файл или DataFrame.

Тема 4. Анализ KPI и оценка эффективности изменений

Лабораторный практикум № 4. Комплексная работа

Задание: «Анализ эффективности внедрения нового ПО»

Компания внедрила новую программу для менеджеров. Есть данные за месяц до внедрения и за месяц после. Данные содержат время обработки одной заявки.

Что нужно сделать:

Загрузить оба датасета.

Рассчитать ключевой показатель: среднее время обработки заявки до и после изменений.

Визуализировать результат (сравнение «до/после»).

Рассчитать экономический эффект (например, сколько заявок обработано дополнительно за сэкономленное время).

Типовые задания для практикумов по решению задач

Тема 1. Основы Python и работа с табличными данными

«Подготовка данных для анализа эффективности»

Перед отделом стоит задача оценить эффективность работы двух смен. Даны два разных отчета: один в CSV (данные по сменам), другой в Excel (данные по выработке).

Задача 1. Объединить данные из двух источников в одну таблицу.

Задача 2. Сгруппировать данные по сменам.

Задача 3. Рассчитать суммарную выработку и среднее время простоя для каждой смены.

Задача 4. Сделать вывод, какая смена работает эффективнее, на основе подготовленных данных.

Тема 2. Визуализация данных и базовый анализ

«Анализ причин брака на производстве»

Даны данные о количестве бракованных изделий по дням и по типам дефектов (трещины, сколы, несоответствие размеров).

Задача 1. Визуализировать распределение дефектов по типам (какой тип встречается чаще всего).

Задача 2. Построить график, показывающий, есть ли рост брака в определенные дни недели.

Задача 3. На основе графиков сформулировать гипотезу о возможных причинах брака (например, проблема с конкретным станком или сырьем).

Тема 3. Автоматизация сбора данных

«Мониторинг курса валют для закупки сырья»

Компании необходимо закупать комплектующие за рубежом. Цена контракта зависит от курса доллара к рублю. Нужно автоматизировать сбор курса.

Задача 1. Найти публичный API (например, банка или биржи), предоставляющий курс валют в формате JSON.

Задача 2. Написать скрипт, который обращается к API и получает текущий курс.

Задача 3. (Расширенное задание) Настроить скрипт так, чтобы он сохранял курс в файл каждый день, формируя временной ряд для дальнейшего анализа колебаний.

Тема 4. Анализ KPI и оценка эффективности изменений

«Разработка системы контроля KPI инновационного проекта».

Студент выступает в роли аналитика проекта по разработке нового продукта. Ему нужно предложить систему контроля для руководителя проекта.

Задача 1. Определить ключевые показатели (KPI) для контроля: соблюдение сроков, соблюдение бюджета, достижение технических параметров прототипа.

Задача 2. Предложить источники данных для этих KPI (из каких отчетов или систем их брать).

Задача 3. Описать процесс: как часто собирать данные, как визуализировать отклонения от плана и какие действия предпринимать при их возникновении.

5. Оценочные материалы для проведения промежуточной аттестации

Промежуточная аттестация по дисциплине проводится в форме зачета.

Зачет представляет собой выполнение обучающимся заданий билета, включающего в себя:

Задание №1 – теоретический вопрос на знание базовых понятий предметной области дисциплины, а также позволяющий оценить степень владения обучающегося принципами предметной области дисциплины, понимание их особенностей и взаимосвязи между ними.

Задание №2 – задание на анализ ситуации из предметной области дисциплины и

выявление способности обучающегося выбирать и применять соответствующие принципы и методы решения практических проблем, близких к профессиональной деятельности.

Задание №3 – задания на проверку умений и навыков, полученных в результате освоения дисциплины.

Выполнение обучающимся заданий билета оценивается по следующей балльной шкале:

Задание 1: 0-30 баллов

Задание 2: 0-30 баллов

Задание 3: 0-40 баллов

«Зачтено»

– 90-100 – ответ правильный, логически выстроен, использована профессиональная терминология. Задания решены правильно. Обучающийся правильно интерпретирует полученный результат.

– 70 -89 – ответ в целом правильный, логически выстроен, использована профессиональная терминология. Ход решения заданий правильный, ответ неверный. Обучающийся в целом правильно интерпретирует полученный результат.

– 50 - 69 – ответ в основном правильный, логически выстроен, использована профессиональная терминология. Задание решено частично.

«Не зачтено»

– менее 50 – ответы на теоретическую часть неправильные или неполные. Задания не решены.

Типовые задания для проведения промежуточной аттестации обучающихся

Задания 1 типа

1. Что такое библиотека Pandas и для чего она используется в анализе данных?
2. В чем главное различие между структурой данных list (список) и DataFrame в Pandas?
3. Что такое Jupyter Notebook и какие преимущества он дает аналитику?
4. Какая библиотека в Python считается основной для научных вычислений и работы с массивами?
5. Что такое API и зачем он нужен при сборе данных?
6. Назовите две основные библиотеки Python для построения графиков.
7. Для чего используется библиотека Requests?
8. Что такое «веб-скрапинг» (web scraping)?
9. Какая библиотека чаще всего используется для парсинга HTML-кода веб-страниц?
10. Что такое NaN в данных и почему с ними нужно работать?
11. Какой метод в Pandas позволяет прочитать данные из файла Excel?
12. Какой метод в Pandas позволяет сохранить данные из DataFrame в файл CSV?
13. Что делает функция head() в Pandas?
14. Что такое «векторизация» и почему она работает быстрее обычных циклов for?
15. Какой тип графика лучше всего подходит для сравнения категорий (например, продаж по регионам)?
16. Какой тип графика лучше всего подходит для отображения изменения показателя во времени (например, курс валют)?
17. Что такое NumPy-массив (ndarray) и чем он отличается от обычного списка?
18. Для чего в Python используются виртуальные окружения (virtualenv) или менеджер пакетов Anaconda?
19. Что такое «коммит» (commit) в контексте работы с кодом в Jupyter Notebook или Git?
20. Что означает аббревиатура CSV и как этот файл хранит данные?

Задания 2 типа

1. В отчете о продажах нового продукта обнаружены пропуски в колонке «Сумма сделки» и отрицательные значения в колонке «Количество». Как бы вы подготовили эти данные к анализу с помощью Pandas перед расчетом выручки?
2. Вам нужно сравнить эффективность работы двух производственных линий за год. Какую визуализацию (Matplotlib/Seaborn) вы выберете для сравнения распределений дневной выработки и почему?
3. Анализ показал сильную положительную корреляцию между затратами на маркетинг и объемом продаж. Означает ли это, что маркетинг — единственная причина роста продаж? Какие еще факторы следует проверить, используя методы анализа данных?
4. Руководство утверждает, что внедрение новой технологии снизило себестоимость единицы продукции. У вас есть данные о себестоимости за 6 месяцев до и 6 месяцев после внедрения. Опишите пошагово, как вы докажете или опровергнете это утверждение с помощью Python.
5. При анализе данных о работе оборудования вы заметили несколько аномально высоких значений времени простоя. Как бы вы использовали методы описательной статистики и визуализации для выявления и анализа этих выбросов?
6. Вам нужно оценить, как изменился средний чек клиента после запуска программы лояльности. У вас есть два файла Excel: «До» и «После». Опишите процесс объединения этих данных и расчета итогового показателя в Pandas.
7. Финансовый директор просит спрогнозировать расходы на следующий квартал на основе данных за последние 3 года. Какой тип анализа (регрессионный анализ или анализ временных рядов) здесь уместен и почему?
8. В компании внедрена новая система мотивации. Вам нужно проверить, повлияла ли она на производительность труда. Какие KPI вы бы рассчитали и какие графики построили для презентации руководству?
9. Вам необходимо ежедневно отслеживать цены на ключевое сырье у трех основных поставщиков, чтобы выбрать лучшее предложение. Опишите, какой метод (API или веб-скрапинг) вы бы выбрали для автоматизации этого процесса и почему.
10. Вы решили использовать веб-скрапинг для сбора отзывов о продукте конкурента с маркетплейса. В процессе парсинга сайт выдает ошибку доступа. Опишите возможные причины и способы решения этой проблемы.
11. Вы получили данные от поставщика в формате JSON через API. Структура данных вложенная (есть списки и словари внутри). Опишите, как бы вы преобразовали эти данные в плоский DataFrame Pandas для дальнейшего анализа.
12. Компания хочет отслеживать упоминания своего бренда в новостях. Предложите план автоматизации сбора таких данных с использованием Python (укажите библиотеки).
13. Анализ показал, что процент брака резко вырос в последнюю неделю месяца. У вас есть данные о сменах, станках и операторах. Как бы вы использовали методы группировки (groupby) в Pandas для выявления первопричины проблемы?
14. Вам нужно представить отчет о динамике KPI за год генеральному директору, который не любит таблицы. Какие типы графиков (линейный график, тепловая карта) вы выберете и почему?
15. Проект по разработке прототипа выходит за рамки бюджета на 15%. У вас есть дорожная карта проекта и фактические данные о затратах по неделям. Как бы вы визуализировали это отклонение для отчета?
16. Отдел кадров утверждает, что текучесть кадров не связана с уровнем зарплаты. У вас есть анонимизированные данные по сотрудникам (стаж, зарплата, факт увольнения). Как бы вы проверили эту гипотезу?
17. Вам нужно автоматизировать еженедельную рассылку отчета о продажах руководителям отделов. Опишите процесс: от извлечения данных из базы до формирования графика и отправки письма.
18. В DataFrame с данными о поставках названия одних и тех же поставщиков

написаны с опечатками («ООО Ромашка», «Ромашка ООО», «ООО Ромашкка»). Как бы вы стандартизировали эти данные для корректного подсчета общей суммы закупок у каждого поставщика?

19. Компания планирует выход на новый рынок. Вам поручено собрать данные о ценах конкурентов и их доле рынка для первичного анализа. Опишите ваш план действий по сбору и первичной обработке этих данных.

20. Вы провели анализ эффективности отдела после внедрения нового ПО. Данные говорят об улучшении одних показателей (скорость), но ухудшении других (количество ошибок). Как бы вы представили эти противоречивые результаты руководству, чтобы они могли принять взвешенное решение?

Задания 3 типа

Задание 1. Вам предоставлен файл `purchases.xlsx` с данными о закупках сырья. Данные собраны из разных отделов, поэтому в них есть ошибки.

Задача:

Загрузить данные в `DataFrame`.

Найти и удалить дубликаты.

В столбце Цена заменить запятые на точки и привести тип данных к `float`.

Заполнить пропуски в столбце Поставщик значением `Неизвестно`.

Сохранить очищенный файл как `clean_purchases.xlsx`.

Задание 2. у вас есть `DataFrame sales` с колонками: Дата, Менеджер, Сумма_сделки.

Задача:

Сгруппировать данные по менеджерам.

Рассчитать общую сумму продаж и количество сделок для каждого.

Вычислить средний чек ($\text{Средний_чек} = \text{Сумма_сделки} / \text{Кол-во}$).

Отсортировать менеджеров по общей сумме продаж (от большего к меньшему) и вывести топ-3.

Задание 3. Дан `DataFrame defects` с данными о браке за неделю: Дата, Тип_дефекта, Количество.

Задача:

Построить линейный график, показывающий общую динамику брака по дням.

Построить столбчатую диаграмму, показывающую общее количество дефектов по типам за всю неделю.

Задание 4. Вам нужно получить текущий курс доллара к рублю с помощью публичного API (например, от ЦБ РФ или через RapidAPI).

Задача:

Используя библиотеку `requests`, отправить GET-запрос к API.

Извлечь значение курса из полученного JSON-ответа.

Вывести на экран сообщение: «Текущий курс USD: <значение> руб.».

Задание 5. На странице сайта (предоставляется URL) есть таблица с рейтингом инновационных компаний.

Задача:

Получить HTML-код страницы с помощью `requests`.

Найти таблицу в коде с помощью `BeautifulSoup`.

Извлечь названия первых 5 компаний из таблицы и вывести их списком.

Задание 6. У вас есть `DataFrame production` с данными о работе цеха: Количество_рабочих, Выработка_тонн, Время_простоя_мин.

Задача:

Рассчитать коэффициент корреляции Пирсона между Количеством_рабочих и Выработкой_тонн.

Интерпретировать результат (сильная/слабая/нет связи, прямая/обратная).

Задание 7. Дан `DataFrame clients` со списком клиентов банка: Баланс, Возраст, Статус.

Задача:

Отфильтровать клиентов, у которых баланс больше 100 000 руб. ИЛИ возраст больше 45 лет.

Из отфильтрованных выбрать только тех, у кого статус «VIP».

Вывести количество таких клиентов.

Задание 8. DataFrame projects содержит данные о R&D проектах: Отдел, Бюджет, Статус.

Задача:

Сгруппировать проекты по отделам.

Для каждого отдела посчитать количество проектов, общую сумму бюджетов и найти средний бюджет одного проекта.

Задание 9. В DataFrame orders столбец Дата_заказа имеет строковый тип в формате «25.10.2025».

Задача:

Преобразовать этот столбец в формат даты (datetime).

Создать новый столбец Месяц, содержащий только месяц и год заказа (например, «Октябрь-2025»).

Посчитать количество заказов в месяц «Апрель-2025».

Задание 10. Вы выполнили анализ данных в Jupyter Notebook. Вам нужно сохранить результаты для руководителя, который не умеет программировать.

Задача:

Взять DataFrame из любого предыдущего задания (например, топ-менеджеров из Задания 2).

Сохранить этот DataFrame в файл Excel (report.xlsx) на отдельный лист с названием «Топ-менеджеры».

Сохранить график из Задания 3 в файл PNG (defects_chart.png).