

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Косенок Сергей Михайлович
Должность: ректор
Дата подписания: 16.06.2026 12:02:04
Уникальный программный ключ:
e3a68f3eaa1e62674b54f4998099d3d6bfdcf836

Оценочные материалы для промежуточной аттестации по дисциплине

«Анализ данных на языке Python»

Квалификация выпускника	бакалавр
Направление подготовки	38.03.01 <i>шифр</i> Экономика <i>наименование</i>
Направленность (профиль)	Учет, налогообложение, финансы и аудит <i>наименование</i>
Форма обучения	очная <i>наименование</i>
Кафедра- разработчик Выпускающая кафедра	Прикладная математика <i>наименование</i> Экономика, учет и финансы

Типовое задание для контрольной работы

Тема: «Исследовательский анализ экономических показателей региона»

Контекст: Вы работаете аналитиком в региональном департаменте экономики. Вам предоставлен датасет `regional_economy.csv` с данными по муниципальным образованиям: численность населения, средний доход, уровень безработицы, количество предприятий, объём инвестиций, расходы бюджета.

Задание:

№ 1. Загрузка и первичная обработка данных (Pandas)

1. Загрузите датасет, выведите информацию о структуре данных (`.info()`, `.describe()`).
2. Найдите и обработайте пропущенные значения: обоснуйте выбор стратегии (удаление/замена медианой/средним) с экономической точки зрения.
3. Отфильтруйте данные: оставьте только муниципалитеты с населением $> 10\,000$ чел.

№ 2. Описательная статистика и визуализация

4. Рассчитайте корреляцию между уровнем безработицы и средним доходом. Постройте scatter-plot с линией тренда (`matplotlib/seaborn`).
5. Постройте гистограмму распределения объёма инвестиций. Есть ли аномалии? Как они могут влиять на экономические выводы?
6. Сгруппируйте данные по уровням урбанизации (создайте признак: «город» / «посёлок» / «сельская местность» по пороговому значению численности). Рассчитайте средние показатели по группам.

№ 3. Интерпретация (экономический смысл)

7. Сформулируйте 2–3 исследовательских гипотезы на основе визуального анализа (например: «В городских муниципалитетах выше концентрация инвестиций на душу населения»).
8. напишите краткий аналитический вывод (5–7 предложений): какие закономерности вы обнаружили и как их можно использовать при планировании бюджетной политики?

Типовые вопросы и практические задания к экзамену

Сформулируйте развернутые ответы на следующие теоретические вопросы :

1. Аналитика и программирование. Python в аналитике.
2. Основы синтаксиса Python. Определение данных и их типов. Условные конструкции. Циклы и их организация.
3. Пользовательские методы и встроенные функции. Создание функций.
4. Работа с данными в Python с использованием Pandas: создание, заполнение и выгрузка датафрейма, фильтрация данных, изменение данных в соответствии с условием задачи.
5. Агрегация данных, расчет по группам, сводные таблицы, сортировка данных. Обработка пропусков, выбросов, аномалий, дубликатов в данных.
6. Визуализация данных: библиотеки matplotlib, seaborn, plotly.
7. Исследовательский анализ данных. Анализ аномалий и категоризация данных.
8. Одномерные и двумерные распределения. Корреляция данных.
9. Анализ агрегированных данных, группировок, сводных таблиц, графиков. Проверка исследовательских гипотез.
10. Статистический анализ: повторение математической статистики, статистические тесты, A/B-тестирование.
11. Машинное обучение в аналитике.
12. Обучение без учителя, кластеризация k-mean.
13. Обучение с учителем. Матрица ошибок, дерево решений, регрессия.
14. E-commerce метрики.
15. Анализ временных рядов.
16. Сегменты и когорты. RFM-сегментация клиентов.

Задание:

Тема: «RFM-сегментация клиентов интернет-магазина для оптимизации маркетингового бюджета»

Контекст: Вы — data scientist в e-commerce компании. Руководство хочет персонализировать рассылки и оптимизировать рекламные расходы. Вам

предоставлен датасет `transactions.csv`: ID клиента, дата заказа, сумма чека, категория товара, канал привлечения.

Этапы выполнения (полный цикл анализа):

№ 1. Подготовка данных и расчёт RFM-метрик

1. Загрузите данные, приведите даты к формату `datetime`.
2. Рассчитайте для каждого клиента три метрики (на дату отсечки — последний день в датасете):
 - **Recency**: сколько дней прошло с последнего заказа;
 - **Frequency**: общее количество заказов;
 - **Monetary**: суммарная выручка от клиента.
3. Проведите категоризацию: разбейте каждую метрику на квантили (1–5), сформируйте итоговый RFM-скор (например, "543").

№ 2. Кластеризация и профилирование сегментов

4. 4. Примените k-means кластеризацию к нормированным значениям R, F, M. Обоснуйте выбор числа кластеров (метод локтя, силуэтный коэффициент).
5. 5. Опишите каждый кластер экономически: например, «Лояльные крупные покупатели», «Уходящие клиенты», «Новички с низким чеком».
6. 6. Визуализируйте кластеры в 2D (PCA или `pairplot`), подпишите группы.

№ 3. Принятие решений и валидация

7. Предложите маркетинговую стратегию для 2 выбранных сегментов: какие каналы коммуникации, тип предложений, ожидаемый ROI?