

Документ подписан простой электронной подписью
 Информация о владельце:
 ФИО: Косенок Сергей Михайлович
 Должность: ректор
 Дата подписания: 23.07.2025 08:37:05
 Уникальный программный ключ:
 e3a68f3eaa1e62b7485449909a5abba1c836

Оценочные материалы для текущего контроля и промежуточной аттестации по дисциплине «Технологии программирования», 4 семестр

Код направления подготовки	38.03.05 Бизнес-информатика
Направленность (профиль)	Экономика предприятий и управление бизнес-процессами
Форма обучения	Очная
Кафедра-разработчик	Менеджмента и бизнеса
Выпускающая кафедра	Менеджмента и бизнеса

Тематика контрольных работ

1. Работа с Pandas
2. Предобработка данных
3. Анализ данных
4. Изменение типов данных
5. Графики и выводы
6. Изучение срезов данных
7. Работа с несколькими источниками данных

Примерные задания к экзамену

Здание 1. Вам предоставлены два набора данных:

1. **sales.csv** : Содержит информацию о продажах товаров за год.
 - Колонки: **date, product_id, quantity, price.**
2. **products.csv** : Содержит информацию о товарах.
 - Колонки: **product_id, category, brand.**

Необходимо выполнить следующие шаги:

Часть 1: Загрузка и объединение данных

1. Загрузите оба файла в DataFrame с помощью Pandas.
2. Объедините данные из двух файлов по ключу **product_id**. Убедитесь, что данные корректно соединены.
3. Проверьте наличие пропущенных значений. Если они есть, обработайте их (например, удалите строки или заполните значения).

Часть 2: Предобработка данных

1. Преобразуйте колонку **date** в формат даты (**datetime**).
2. Создайте новую колонку **total_revenue**, которая будет содержать выручку от продажи каждого товара (формула: **quantity * price**).
3. Измените тип данных для колонок, если это необходимо (например, преобразуйте числовые колонки в соответствующие типы).

Часть 3: Анализ данных

1. Определите общую выручку за год.
2. Найдите топ-5 самых популярных категорий товаров по количеству проданных единиц.

3. Вычислите среднюю цену товаров для каждой категории.
4. Найдите месяц с максимальной выручкой.
5. Постройте график, показывающий динамику выручки по месяцам.

Часть 4: Работа с несколькими источниками данных

1. Создайте новый DataFrame, который содержит только данные о товарах категории "Electronics".
2. Для этой категории найдите бренд с наибольшей общей выручкой.
3. Постройте круговую диаграмму, показывающую долю выручки каждого бренда в категории "Electronics".

Часть 5: Выводы

1. На основе проведенного анализа сделайте выводы:
 - Какая категория товаров является самой популярной?
 - Какой месяц был самым прибыльным?
 - Какой бренд в категории "Electronics" лидирует по выручке?

Задание 2. Вам предоставлены два набора данных:

1. `trips.csv` : Содержит информацию о поездках такси за год.
 - Колонки: **trip_id, pickup_datetime, dropoff_datetime, distance, fare_amount, tip_amount**.
2. `drivers.csv` : Содержит информацию о водителях такси.
 - Колонки: **driver_id, name, rating, total_trips**.

Необходимо выполнить следующие шаги:

Часть 1: Загрузка и объединение данных (5 баллов)

1. Загрузка данных (1 балл):
 - Загрузите оба файла в DataFrame с помощью Pandas.
 - Убедитесь, что данные загружены корректно.
2. Объединение данных (2 балла):
 - Объедините данные из двух файлов по ключу **driver_id** (предположим, что в `trips.csv` есть колонка **driver_id**).
 - Убедитесь, что данные корректно соединены.
3. Проверка и обработка пропущенных значений (2 балла):
 - Проверьте наличие пропущенных значений.
 - Обработайте пропущенные значения (например, удалите строки или заполните их).

Часть 2: Предобработка данных (6 баллов)

1. Преобразование временных данных (2 балла):
 - Преобразуйте колонки **pickup_datetime** и **dropoff_datetime** в формат даты (**datetime**).
 - Создайте новую колонку **trip_duration**, которая будет содержать длительность поездки в минутах.
2. Создание новых колонок (2 балла):
 - Создайте колонку **total_amount**, которая будет содержать общую сумму поездки (**fare_amount + tip_amount**).
 - Создайте колонку **average_speed**, которая будет содержать среднюю скорость поездки (в км/ч). Формула: $average_speed = \frac{trip_duration \text{ (в часах)} \times distance}{trip_duration}$

3. Изменение типов данных (2 балла):
 - Преобразуйте числовые колонки (например, **distance**, **fare_amount**, **tip_amount**) в соответствующие типы данных.
 - Убедитесь, что все преобразования выполнены корректно.

Часть 3: Анализ данных (8 баллов)

1. Общая выручка за год (1 балл):
 - Вычислите общую сумму всех поездок (**total_amount**) за год.
2. Топ-5 самых популярных водителей (2 балла):
 - Определите топ-5 водителей по количеству поездок.
 - Отсортируйте результаты по убыванию.
3. Средний рейтинг водителей (2 балла):
 - Вычислите средний рейтинг (**rating**) для всех водителей.
 - Найдите водителя с самым высоким рейтингом.
4. Месяц с максимальной выручкой (2 балла):
 - Агрегируйте выручку по месяцам.
 - Определите месяц с максимальной выручкой.
5. График динамики выручки по месяцам (1 балл):
 - Постройте график, показывающий динамику выручки по месяцам.
 - Убедитесь, что график читаемый и оси подписаны.

Часть 4: Работа с несколькими источниками данных (7 баллов)

1. Создание DataFrame для водителей с рейтингом выше 4.5 (2 балла):
 - Создайте новый DataFrame, содержащий только данные о водителях с рейтингом выше 4.5.
 - Убедитесь, что фильтрация выполнена корректно.
2. Водитель с наибольшей общей выручкой (2 балла):
 - Для водителей с рейтингом выше 4.5 найдите водителя с наибольшей общей выручкой.
 - Используйте группировку и агрегацию для вычислений.
3. Круговая диаграмма доли выручки водителей (3 балла):
 - Постройте круговую диаграмму, показывающую долю выручки каждого из топ-5 водителей с рейтингом выше 4.5.
 - Убедитесь, что диаграмма правильно отображает данные, подписи и легенда добавлены.

Часть 5: Выводы (4 балла)

1. Выводы по анализу (4 балла):
 - Ответьте на следующие вопросы:
 - Какой водитель выполнил больше всего поездок? (1 балл)
 - Какой месяц был самым прибыльным? (1 балл)
 - Какой водитель с рейтингом выше 4.5 лидирует по выручке? (1 балл)
 - Выводы логичны, основаны на результатах анализа, четко сформулированы (1 балл).

Критерии оценивания

Часть	Критерии	Макс.	Общая шкала оценивания
-------	----------	-------	------------------------

		баллы	
Часть 1: Загрузка и объединение	Загрузка данных	1	<ul style="list-style-type: none"> • 30–27 баллов: Отлично. Все задания выполнены полностью, код написан аккуратно, выводы сделаны логично. • 26–21 балл: Хорошо. Большинство заданий выполнено корректно, но есть незначительные ошибки или недочеты. • 20–15 баллов: Удовлетворительно. Задания выполнены частично, есть существенные ошибки или пропуски. • Менее 15 баллов: Неудовлетворительно. Задания выполнены некорректно или не выполнены вообще.
	Объединение данных	2	
	Обработка пропущенных значений	2	
Часть 2: Предобработка данных	Преобразование колонки	2	
	Создание колонки	2	
	Изменение типов данных	2	
Часть 3: Анализ данных	Анализ 1. Подсчет общего	1	
	Анализ 2. Топ 5	2	
	Анализ 3. Поиск среднего	2	
	Анализ 4. Поиск максимального	2	
	График	1	
Часть 4: Работа с данными	Новый DataFrame	2	
	Поиск наибольшего	2	
	Круговая диаграмма	3	
Часть 5: Выводы	Выводы по анализу	4	
Итого		30	

