1. **Анализ поведения участников аукционов**

*Описание:*

Датасет " Shill Bidding Dataset" содержит информацию о посвященный изучению поведения участников аукционов, в частности, шилл-биддеров (shill bidders), которые используют мошеннические методы для манипуляции аукционами. Вот краткое описание переменных:

1. Record ID: Уникальный идентификатор записи.

2. Auction ID: Уникальный идентификатор аукциона.

3. Bidder ID: Уникальный идентификатор участника.

4. Bidder Tendency: Описание поведения шилл-биддера, который предпочитает участвовать в аукционах у немногих определенных продавцов.

5. Bidding Ratio: Частота участия шилл-биддера в торгах для повышения цен.

6. Successive Outbidding: Процесс, при котором шилл-биддер последовательно повышает свои собственные ставки для увеличения цены.

7. Last Bidding: Поведение шилл-биддера в конце аукциона, когда он становится неактивным, чтобы избежать выигрыша.

8. Auction Bids: Общее количество ставок в аукционе с участием шилл-биддеров, которое обычно выше среднего.

9. Auction Starting Price: Начальная цена, которая часто низкая, чтобы привлечь законных участников.

10. Early Bidding: Активация шилл-биддера на ранних стадиях аукциона для привлечения внимания пользователей.

11. Winning Ratio: Соотношение, показывающее, что шилл-биддер участвует в большом количестве аукционов, но редко побеждает.

12. Auction Duration: Длительность аукциона.

13. Class: Классификация поведения — 0 для обычного участия в торгах, 1 для шилл-биддеров.

1. **Предобработка данных:**  
   • Загрузите датасет и проверьте наличие пропущенных значений. Обработайте их  
   • Преобразуйте категориальные переменные, если они есть

2. **Выбор переменных для кластеризации:**  
  
3. **Анализ данных:**  
   • Постройте визуализации (гистограммы, диаграммы рассеяния) для понимания распределения выбранных переменных.  
   • Рассчитайте основные статистики (средние значения, медианы) для количественных переменных  
  
4. **Нормализация данных:**  
   • Примените нормализацию или стандартизацию к количественным переменным чтобы избежать влияния масштаба на кластеризацию.  
  
5. **Определение числа кластеров:**  
   • Используйте метод локтя или силуэтный метод для определения оптимального количества кластеров для K-means или другого алгоритма кластеризации.  
  
6. **Кластеризация:**  
   • Примените алгоритм K-means или иерархическую кластеризацию к подготовленным данным.

7. **Визуализация кластеров:**  
   • Постройте визуализации для представления результатов кластеризации (например, графики с центрами кластеров и их распределением по различным переменным).

8. Удалить из датасета признак (Class) и повторить кластеризацию. Сравнить результаты