Для прогнозування продажів сформуємо собі такий флоу:

1. Завантаження та огляд даних
2. Попередня обробка даних
3. EDA (Exploratory Data Analysis)
4. Підготовка даних для моделі
5. Побудова моделі
6. Прогнозування на основі df\_future.csv

Notes:

Видалити колонку Unnamed: 0, оскільки вона не потрібна.

Перетворити колонку date на формат дати.

EDA щоб зрозуміти розподіл і взаємозв'язки між даними.

**3.EDA**

3.1 Розподіл продажів по днях.

3.2 Розподіл продажів по категоріях товарів.

* 1. Розподіл продажів по товарах (SKU).

3.4 Аналіз середніх цін.

**4. Підготовка даних для моделі. Feature selection**

4.1 Підготовка даних для моделювання.

4.2 Створення ознак (features) для моделі.

4.3 Розбиття даних на тренувальний та тестовий набори.

**5. Побудова моделі прогнозування:**

5.1 Кодування категоріальних даних (OnehotEncoding). В нас немає ординальних даних.

**sparse\_output=False**, вихідні дані у вигляді звичайного щільного масиву NumPy (формату numpy.ndarray)

5.2 Вибір алгоритму моделювання. Random Forest та LinearReg.

5.3 Тренування та валідація моделі.

5.4 Прогнозування на майбутні дати з df\_future.csv.