**أولاً: جمع البيانات:**

1. .مصادر جمع البيانات

البنوك المشاركة : الحصول على البيانات من الأنظمة الداخلية للبنوك.

1. طرق جمع البيانات **:**

**المصادر الثانوية هي الأنسب بسبب**:

* تتعامل مع بيانات موجودة مسبقًا
* تتعلق بالقروض التي تم منحها خلال 10 سنوات سابقة
* تستخرج البيانات من أنظمة داخلية للبنك(مثل نظام القروض، الرواتب، إدارة العملاء)
* **- اسم الملف الذي يحتوي البيانات :** **bank\_loans\_data.csv**

**ثانياً معالجة وتحليل البياناات :**

كود يشمل عملية المعالجة :  
التنظيف(ازالة القيم المفقود، ازالة التكرارات)

التقطيع

التحويل(باستخدام الترميز) .

التحليل المستخدم(كافتراض):  
 1.عدد العملاء حسب نوع التعاقد

2. متوسط القرض لكل فئة عمرية

3. نسبة التخلف عن السداد حسب المدينة

4.حفظ البيانات المعالجة في ملف خاص

**import pandas as pd**

**import numpy as np**

**# =================== 1. قراءة البيانات ===================**

**df = pd.read\_csv('bank\_loans\_data.csv')**

**print(" تم قراءة البيانات.")**

**print(f"عدد السجلات قبل التنظيف: {len(df)}")**

**# =================== 2. تنظيف البيانات ===================**

**# تحويل العمر إلى رقم**

**df['age'] = pd.to\_numeric(df['age'], errors='coerce')**

**# معالجة القيم الناقصة**

**df.fillna({**

**'salary': df['salary'].median(),**

**'loan\_amount': df['loan\_amount'].median(),**

**'gender': 'Unknown',**

**'job\_title': 'Unknown',**

**'contract\_type': 'Unknown',**

**'repayment\_status': 'Unknown'**

**}, inplace=True)**

**# إزالة التكرارات**

**df.drop\_duplicates(subset=['client\_id'], keep='first', inplace=True)**

**# معالجة القيم الشاذة**

**Q1 = df['salary'].quantile(0.25)**

**Q3 = df['salary'].quantile(0.75)**

**IQR = Q3 - Q1**

**df = df[~((df['salary'] < (Q1 - 1.5 \* IQR)) | (df['salary'] > (Q3 + 1.5 \* IQR)))]**

**print(" 2. تم تنظيف البيانات.")**

**print(f"عدد السجلات بعد التنظيف: {len(df)}")**

**# =================== 3. تصنيف الفئات العمرية ===================**

**bins = [20, 29, 39, 49, 59, 70]**

**labels = ['20-29', '30-39', '40-49', '50-59', '60-70']**

**df['age\_group'] = pd.cut(df['age'], bins=bins, labels=labels, right=False)**

**print(" 3. تم إضافة الفئة العمرية.")**

**# =================== 4. تحليل أولي ===================**

**# عدد العملاء حسب نوع التعاقد**

**contract\_counts = df['contract\_type'].value\_counts()**

**print("\n عدد العملاء حسب نوع التعاقد:")**

**print(contract\_counts)**

**# متوسط القرض لكل فئة عمرية**

**avg\_loan\_by\_age = df.groupby('age\_group')['loan\_amount'].mean()**

**print("\n متوسط مبلغ القرض حسب الفئة العمرية:")**

**print(avg\_loan\_by\_age)**

**# نسبة التخلف عن السداد حسب المدينة**

**default\_by\_city = df.groupby('city')['repayment\_status'].apply(lambda x: ((x == 'Default') | (x == 'Late')).mean() \* 100)**

**print("\نسبة التخلف عن السداد حسب المدينة (%):")**

**print(default\_by\_city)**

**# =================== 5. حفظ البيانات المعالجة ===================**

**df.to\_csv('cleaned\_bank\_loans\_data.csv', index=False)**

**print("\n 5. تم حفظ البيانات المعالجة في ملف: cleaned\_bank\_loans\_data.csv")**

**جميع الملفات التابعة للتكليف مرفقة في نفس الملف.**