

# دراسة حالة: تنظيف وتحليل بيانات التجارة الإلكترونية (Data Amazon)

من البيانات الخام وغير المهيكلة إلى رؤى أعمال قابلة للتنفيذ — رحلة تحول بيانات أمازون إلى قرارات مؤثرة.

إعداد: يوسف — مهندس ومحلل بيانات. أدوات: Python (Pandas, Matplotlib,  
.Seaborn, WordCloud)



# التحدي: فوضى البيانات تعيق القرار

مشاكل جودة بارزة

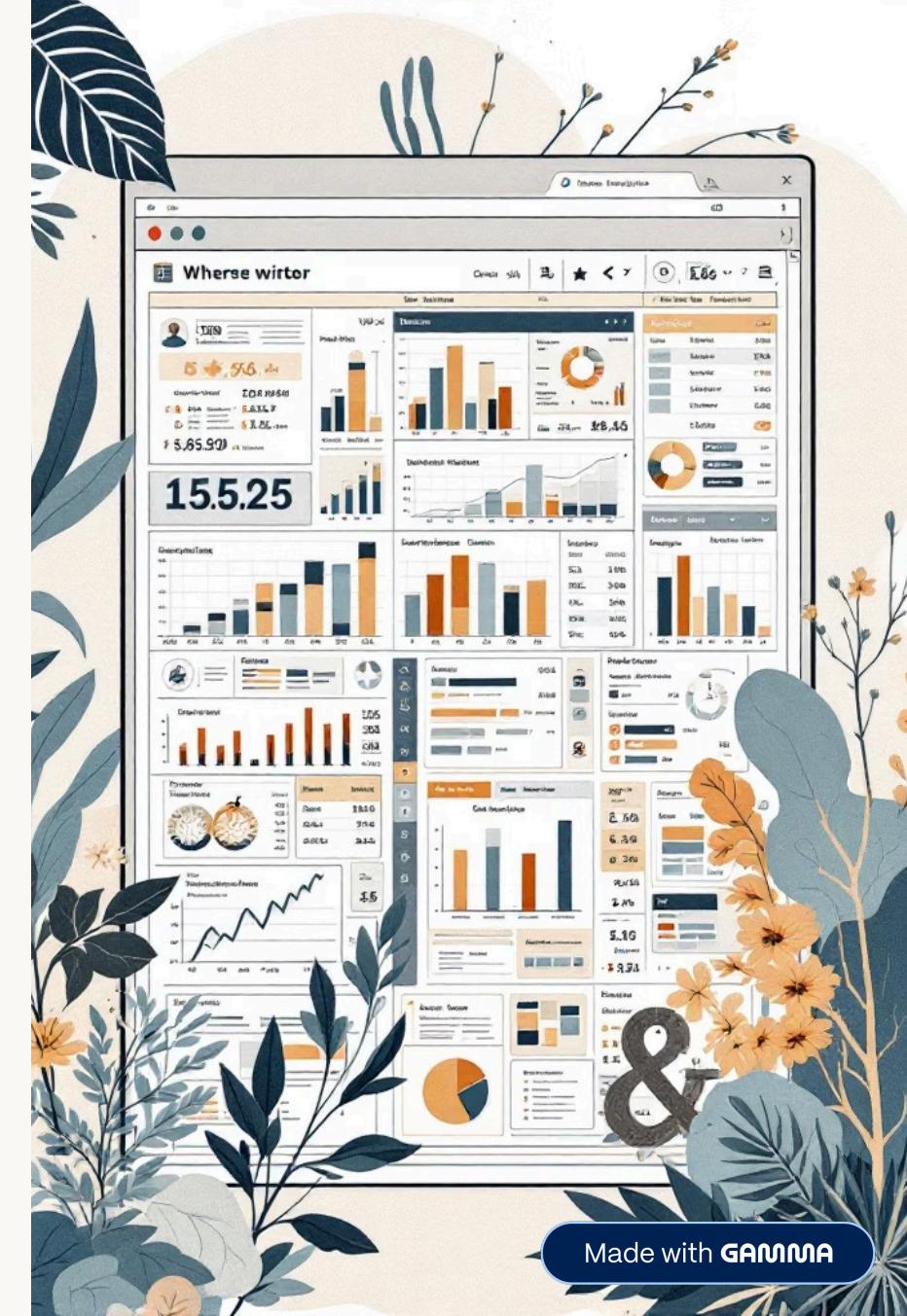
أسعار بنصوص ورموز عمليات، فئات متداخلة، وقيم مفقودة أو شاذة في التقييمات.

حجم ومصدر البيانات

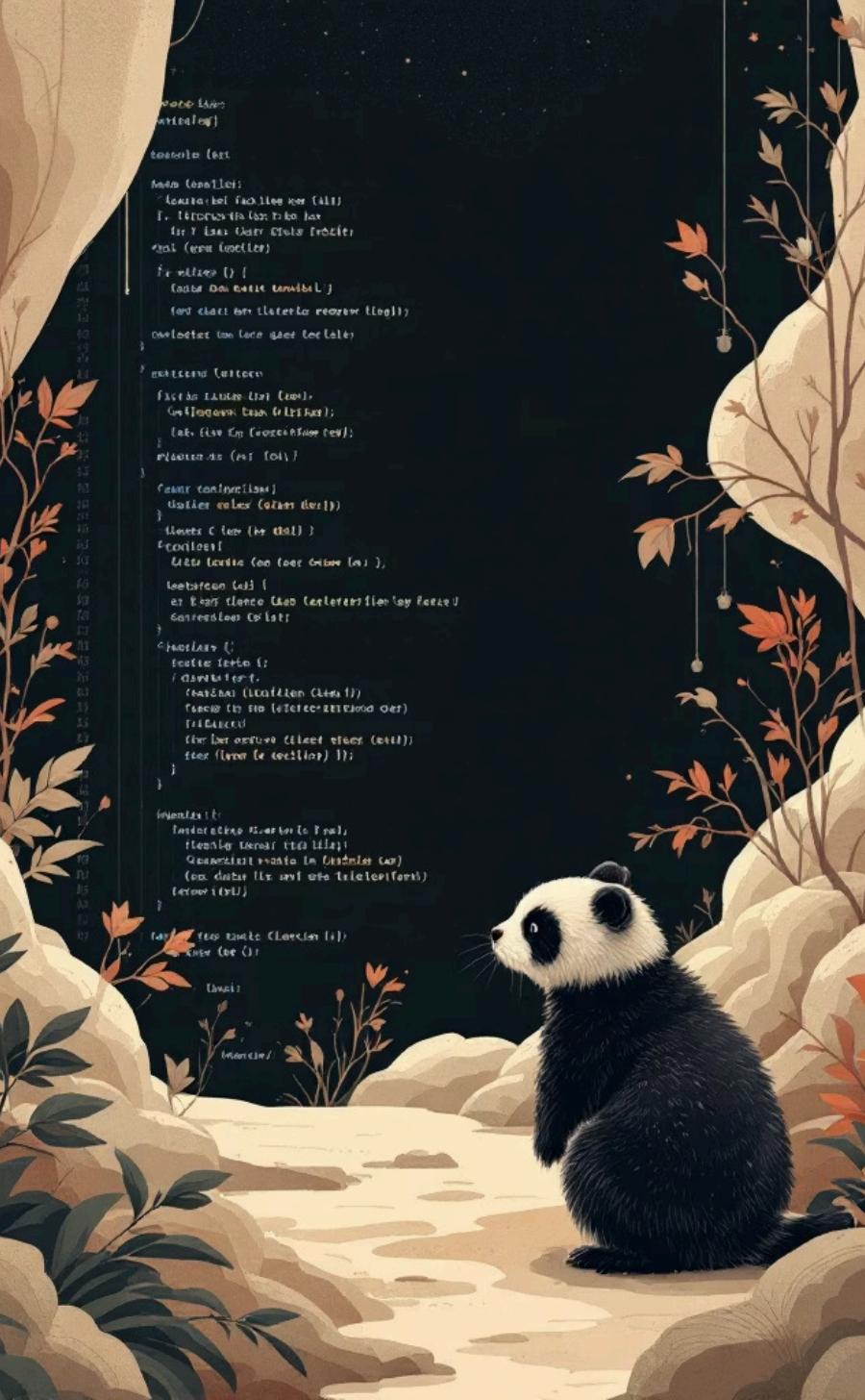
أكثر من 1000 منتج ومئاتآلاف المراجعات الحقيقية — بيانات غير موحدة من قوائم متعددة.

هدف العمل

فهم سلوك العملاء وتأثير التسويق على التقييمات لرفع المبيعات وتحسين الاستهداف التسويقي.



# الحل التقني: Pipeline متين لتنظيف البيانات



## تنقية الأسعار

إزالة رموز العملة وفواصل النص وتحويل القيم إلى Decimal دقيق للتحليل المالي والقياسات التجميعية.



## تصنيف الفئات

فصل الفئات المتداخلة واستخراج Main Category باستخدام قواعد نصية وقواعد شجيرية (hierarchical parsing).



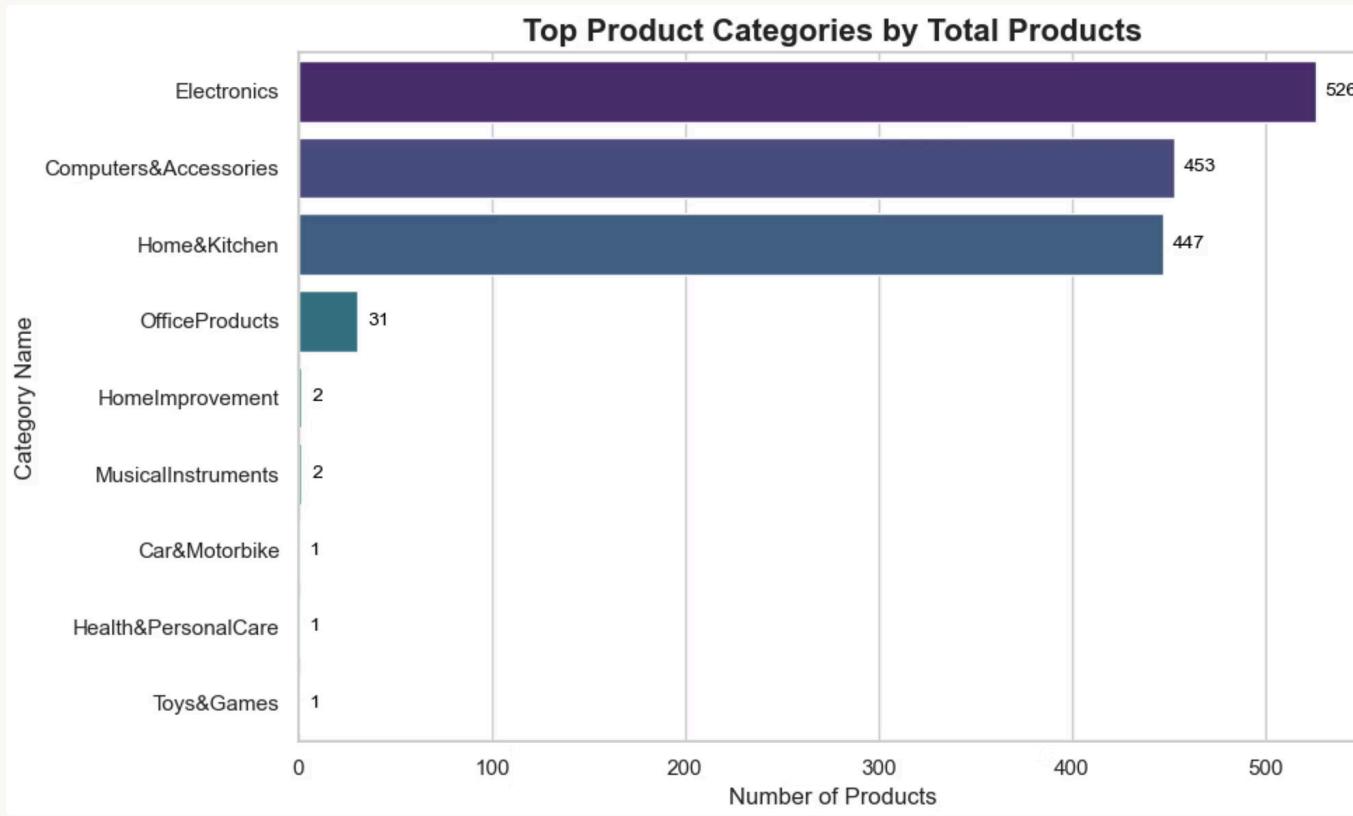
## معالجة الأخطاء والقيم الشاذة

تعويض القيم المفقودة، فلترة الشوادع، وتوثيق التحويلات لضمان استمرارية التجمعيات والإحصاءات.



النتيجة: Pipeline قابلة للتكرار مع سجلات (logs) وفحوصات جودة أوتوماتيكية وصلت إلى موثوقية تحليلات عالية.

# تحليلات استكشافية (EDA)



استخدمنا رسوم بيانية متنوعة لاستخراج مؤشرات تؤثر مباشرة في قرار التسويق والمخزون:

- تحديد الفئات الأعلى مبيعاً والأعلى تقييماً لتوجيه الميزانيات الإعلانية.
- تحليل علاقة نسبة الخصم بحجم التفاعل والمباعات عبر correlation وScatter plots analysis.
- تفريز الأداء حسب Main Category و weekday/hour للتقطاط فرص تحسين العرض.

المؤشرات أدت إلى توصيات عملية: إعادة تخصيص ميزانية إعلانية، تعديل استراتيجيات الخصم، وتحسين توقيت الحملات.

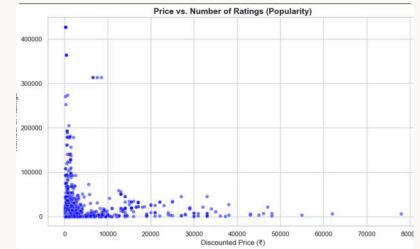
# القيمة المضافة: تحليل نصي وذكاء اصطناعي



تحويل الرؤى إلى إجراءات  
قائمة توصيات قصيرة: تحسين التغليف، تحديد  
سياسة ردود أسرع، وتجربة عروض تسعير  
A/B على فئات محددة.



سحابة كلمات  
استخراج الكلمات الأكثر تكراراً (شحن، تغليف،  
تأخير) لتوجيه تحسينات لوجستية وجودة المنتج.



## تحليل المشاعر (Sentiment Analysis)

تصنيف المراجعات لاكتشاف نقاط الألم — تركيز  
على تعليقات أقل من 3 نجوم لتمييز الشكاوى  
المتكررة.

# الختام: كيف نبدأ تطبيق الحل على عملك

المرحلة الثالثة —  
تسليم واستراتيجية  
لوحات تحكم،  
تقارير تشغيلية،  
وخطة تنفيذ توصيات  
تسويقية وعملية.

المرحلة الثانية —  
تنفيذ Pipeline  
بناء وتنفيذ أنابيب  
التنظيف والتحقق  
الآلي، واختبار EDA  
مبني خلال 3-4 أسابيع.

المرحلة الأولى —  
تقييم البيانات  
فحص سريع لمصادر  
البيانات وجودتها  
وتقدير حجم العمل  
خلال أسبوعين.

مستعد لتطبيق هذه الاستراتيجيات المتقدمة على بياناتكم لتحويلها إلى ميزة تنافسية.  
للتواصل: يوسف — مهندس ومحلل بيانات.

