

Data collection:

البيانات المختارة هي بيانات لمعاملات تجارية لمتجر الكتروني مأخوذة من Kaggle قمت بقراءتها باستخدام Pandas تحتوي على 514909 سجل و تشمل الأعمدة التالية:

InvoiceNo
StockCode
Description
Quantity
InvoiceDate

Data cleaning:

- 1- إزالة الأسطر التي بها قيم فارغة من العمودين **CustomerID** و **Description**.
 - 2- إزالة القيم المكررة.
 - 3- معالجة **Outliers**.
- بعد التنظيف، أصبح حجم البيانات **319806** سطرا و **9** أعمدة.

Exploratory Data Analysis

- إجراء تحليل استكشافي من أجل فهم واكتشاف الأنماط في سلوك العملاء المنتجات والدول :
- 1 - تتبع شراء المنتجات على مدى الأشهر :
اتضح وجود تصاعد إجمالي المبيعات لسنة 2011.
سجل شهر نوفمبر أعلى قدر من المبيعات مما يدل هذا زيادة طلبات موسمية أو كانت عروض ترويجية ناجحة في هذا الشهر .
 - 2- تحليل المبيعات حسب الدول:
أكبر عدد مبيعات سجل في المملكة المتحدة بلغ حوالي 3016377 هذا يشير إلى أن العملاء البريطانيين يمثلون النسبة الأكبر من الإيرادات مقارنة بالدول الأخرى.
 - 3- تحليل أعلى 10 عملاء من حيث إجمالي الإنفاق:
لاحظنا أن العميل صاحب المعرف 14911 ينفق أكثر من ضعف أقرب منافسيه بينما باقي العملاء يظهرون بفروق تدريجية , وهذا أشار الى وجود عدد محدود من العملاء ذوي القيمة العالية .
 - 4- تحليل أكثر 10 منتجات مبيعا:
المنتجات الأكثر مبيعا ASSORTED PACK OF 72 RATROSPOT CAKE CASES بعده COLOUR BIRD ORNAMENT و LUNCH BAG RED RETROSPOT , مما يعكس تفضيل العملاء للمنتجات والهدايا البسيطة.

Advanced Analysis

استخدمت تحليل linear regression لدراسة العلاقة بين المتغيرات والسعر الكلي , حقق النموذج $R^2=0.57$ وهذا يشير إلى أن النموذج قادر على تفسير 57% من التغير في المبيعات بالاعتماد على المتغيرات (اسم المنتج , الكمية , سعر الوحدة).

Dashboard

من خلال Dashboard لاحظنا أن المبيعات ترتفع بشكل كبير مع نهاية السنة خاصة في شهري نوفمبر وديسمبر, مما يعكس تأثير العروض الموسمية على المعاملات التجارية, كما نجد أن المملكة المتحدة تشغل السوق الرئيسي للشركة وهذا يمكن أن يكون عائد الى موقع الشركة في المملكة او شي من هذا القبيل, بينما تبقى مساهمة باقي الدول محدودة مقارنة بها وهذا يشير امكانية العمل على توسع نحو أسواق جديدة أو العمل أكثر مع الدول الأقل تعاملًا , كما تظهر البيانات أن هناك منتجات تحقق مبيعات مرتفعة مقارنة بغيرها وهذا يستدعي التركيز على تسويقها وتوفيرها , كذلك نلاحظ اختلاف في متوسط قيمة السلة الشرائية بين الدول مما يدل على تباين سلوك العملاء في الأخير يساعد تحليل العملاء حسب الانفاق في تقسيم العملاء الى فئات ومعرفة العملاء الأكثر ولاء و تقديم عروض لعزیز ولأنهم .

Automation and Scheduling

لجعل هذه العملية الية تم أتمتة جميع المراحل (جمع البيانات, تنظيفها , تحليلها..) داخل دالة واحدة باستخدام لغة Python, كذلك استخدمت مكتبة **schedule** من أجل جدولة تنفيذ العملية يوميا بهدف تحديث النتائج تلقائيا عند توفر بيانات جديدة .

توصيات :

- العمل أكثر على توجيه الجهود التسويقية نحو الدول التي تشهد مبيعات ضعيفة .
- العمل على تأمين مخزون كافي من اجل المنتجات الأكثر مبيعا والأكثر طلبا .
- استغلال المواسم التي يكثر فيها الشراء من أجل تحقيق أقصى قيمة من الإيرادات .