

# تقرير

الجزء الأول: مقدمة المشروع

❖ **عنوان المشروع :** تحليل بيانات مبيعات لفهم الأنماط السلوكية واتخاذ قرارات تجارية مدروسة

❖ **الهدف من المشروع:** يهدف هذا المشروع إلى تحليل بيانات المبيعات من أجل:

- تحديد الفترات الزمنية التي تشهد أعلى المبيعات.
- معرفة أكثر المنتجات مبيعًا.
- دراسة علاقة السعر بالطلب.
- تقديم توصيات لتحسين الأداء التجاري.

❖ **بيئة العمل:**

تم تنفيذ المشروع باستخدام Google Colab لما يوفره من:

- إمكانية التعامل مع بيانات كبيرة.
- سهولة استخدام مكتبات تحليل البيانات.
- دعم التكامل مع Google Drive لتخزين الملفات.

❖ **نوع البيانات المستخدمة:**

تم استخدام بيانات مبيعات Sales Dataset تتضمن معلومات مثل:

- رقم الطلب Order ID
- اسم المنتج Product Name
- الفئة Category
- السعر Price
- الكمية المباعة Quantity
- تاريخ البيع Date
- المنطقة Region

## ❖ الهدف التحليلي:

من خلال تحليل هذه البيانات سنقوم بـ:

1. استكشاف البيانات Data Exploration
2. تنظيف البيانات Data Cleaning
3. التحليل الإحصائي والوصفي Descriptive Analysis
4. رسم المخططات البيانية لفهم الاتجاهات Visualization
5. استخلاص التوصيات النهائية Insights & Recommendations

## ➤ الخطوة 1: جمع البيانات Data Collection

في هذه المرحلة قمنا بتحميل بيانات المبيعات من ملف CSV يدويًا باستخدام Python في Google Colab البيانات تتضمن معلومات حول:

رقم الطلب Order ID

المنتج Product

الكمية والسعر Quantity Ordered , Price Each

تاريخ الشراء Order Date

عنوان المشتري Purchase Address

الشهر، المدينة، والساعة Month, City, Hour

تم تخزين البيانات في DataFrame باسم df بنجاح، وهي جاهزة لمرحلة التنظيف والتحليل في الخطوة القادمة.

## ➤ الخطوة 2: تنظيف البيانات Data Cleaning & Preprocessing

في هذه المرحلة نعمل على جعل البيانات جاهزة للتحليل من خلال:

- حذف القيم المفقودة Missing Values.
  - إزالة التكرارات Duplicates.
  - تحويل الأعمدة إلى الصيغ المناسبة مثل التواريخ.
  - التأكد من القيم الرقمية.
  - معالجة القيم غير المنطقية أو الشاذة Outliers.
- هذه الخطوة تضمن أن التحليل اللاحق سيكون دقيقًا وموثوقًا.

## ➤ الخطوة 3: التحليل الاستكشافي للبيانات Exploratory Data Analysis - EDA

في هذه المرحلة نستكشف البيانات لفهمها بشكل أعمق، وتشمل:

- الإحصاءات الوصفية (المتوسط، الانحراف المعياري، إلخ).
- تحليل المبيعات حسب الوقت (الشهور، الأيام، الساعات).
- تحديد أفضل المنتجات مبيعًا.
- تحديد المدن الأكثر مبيعات.
- اكتشاف الاتجاهات الموسمية في الطلب.

يتم ذلك من خلال الجداول والرسوم البيانية لفهم الأنماط والعلاقات داخل البيانات.

## ➤ الخطوة 4: التحليل المتقدم Advanced Analysis

في هذه المرحلة نستخدم خوارزميات من التعلم الآلي Machine Learning لاستخراج أنماط خفية من البيانات. سنطبق تقنية التجميع Clustering لتقسيم العملاء أو المدن أو المنتجات إلى مجموعات متشابهة بناءً على سلوك المبيعات. على سبيل المثال:

تحديد أنواع الزبائن (كبار المشتريين، المشتريين العاديين...).

تحديد المنتجات التي تُشترى معًا.

مقارنة المدن حسب حجم المبيعات وعدد الطلبات.

سنستخدم خوارزمية K-Means Clustering لتصنيف المدن حسب الأداء التجاري.

## ➤ الاستنتاج العام:

- من خلال تحليل بيانات المبيعات و مختلف التمثيلات البيانية في كل من colab و powerBI يمكن استخلاص مجموعة من الملاحظات المهمة حول أداء المبيعات خلال الفترة المدروسة :
- أظهرت نتائج الدائرة النسبية ان المنتج MacBook Laptop كان الأعلى مبيعا ثم يليه iphone في حين جاءت باقي مبيعات بنسب اقل نسبيا.
- عند تمثيل اجمالي المبيعات حسب السنة ،لاحظنا انه في سنة 2019 كانت الفترة التي حققت فيها الشركة اعلى مبيعات، لتبدأ بعدها في الانخفاض التدريجي خلال سنة 2020.
- كما بين التمثيل المنحنى الافقي حسب المدن ان اكبر حجم مبيعات سجل في مدينة san francisco تليها los Angeles مما يشير الى ان النشاط التجاري للشركة اكثر تركزا في المناطق الكبرى .
- يبين الجدول ان اجمالي المبيعات خلال 2019 بلغ 34.449.203.579 وهو رقم قياسي يعكس أداء مالي قوي للمؤسسة كما بلغ اجمالي عدد الوحدات المباعة الى 209.079 وحدة مما يشير الى حجم الطلب المرتفع على المنتجات وهذا يدل على ان النشاط التجاري يسير بوتيرة إيجابية وان هنالك منتجات تحقق أرباحا كبيرة سواءا من حيث السعر او حجم المبيعات