**Project Overview**

This project focuses on analyzing historical sales data from a sample business dataset. The primary goal is to derive actionable insights by analyzing orders, products, customers, and employee performance. The project is designed to showcase data cleaning, transformation, aggregation, and visualization techniques using tools like Python, Power BI, or SQL.

**Datasets**

1. **Categories**
   * Contains product categories and descriptions.
   * Fields: CategoryID, CategoryName, Description, Picture.
2. **Customers**
   * Contains customer information such as contact details, address, and company.
   * Fields: CustomerID, CompanyName, ContactName, Address, City, Region, PostalCode, Country, Phone, Fax.
3. **Employees**
   * Contains employee records, including job titles, birth dates, and reporting structure.
   * Fields: EmployeeID, LastName, FirstName, Title, TitleOfCourtesy, BirthDate, HireDate, Address, City, Region, PostalCode, Country, HomePhone, Extension, Photo, Notes, ReportsTo.
4. **Order Details**
   * Contains detailed line items for each order, including product, price, and quantity.
   * Fields: OrderID, ProductID, UnitPrice, Quantity, Discount.
5. **Orders**
   * Contains order records, customer info, shipping details, and freight charges.
   * Fields: OrderID, CustomerID, EmployeeID, OrderDate, RequiredDate, ShippedDate, ShipVia, Freight, ShipName, ShipAddress, ShipCity, ShipRegion, ShipPostalCode, ShipCountry.
6. **Products**
   * Contains product details including supplier, pricing, and stock information.
   * Fields: ProductID, ProductName, SupplierID, CategoryID, QuantityPerUnit, UnitPrice, UnitsInStock, UnitsOnOrder, ReorderLevel, Discontinued.
7. **Shippers**

* Contains shippers details including company and phone\_no
* Fields:ShipperID, CompanyName, Phone

**Key Metrics and Insights**

**1. Sales Performance**

* Total revenue by month, quarter, and year.
* Top-selling products and categories.
* Revenue contribution by region, city, and customer.
* Customer segmentation by purchase frequency and revenue.

**2. Customer Analysis**

* Identify high-value and repeat customers.
* Customer geographic distribution.
* Seasonal demand patterns by customer location.

**3. Product Insights**

* Inventory levels vs. sales trends.
* Products nearing discontinuation or reorder thresholds.
* Price sensitivity and discount patterns.

**4. Employee Performance**

* Orders handled per employee.
* Employee sales contribution over time.
* Relationship between hire dates and performance.

**5. Operational Insights**

* Shipping delays and their effect on customer satisfaction.
* Freight cost analysis.
* Order fulfillment trends.

**Data Preparation Steps**

1. **Data Cleaning**
   * Handle missing values (e.g., missing regions in customers and orders).
   * Normalize date formats.
   * Remove duplicates and irrelevant columns (like pictures and notes where unnecessary).
2. **Data Transformation**
   * Merge datasets using appropriate keys:
     + Orders → Customers on CustomerID.
     + Orders → Employees on EmployeeID.
     + Order Details → Orders on OrderID.
     + Order Details → Products on ProductID.
     + Products → Categories on CategoryID.
3. **Feature Engineering**
   * Create TotalPrice = UnitPrice \* Quantity \* (1 - Discount).
   * Derive OrderAge = ShippedDate - OrderDate.
   * Segment customers by order count or spending.

**Visualizations**

1. **Total Sales Over Time** – Line chart with monthly granularity.
2. **Top 10 Products by Revenue** – Bar chart.
3. **Customer Distribution by Country** – Map visualization.
4. **Inventory vs. Sales** – Scatter plot or heatmap.
5. **Employee Performance Dashboard** – Table with KPIs like orders handled, revenue contribution, and average fulfillment time.