**Kmart E-commerce Data Analysis**

**Dataset:** The dataset for this project contains 12 csv files, each file holds the sales data of individual months of the year 2019.

* One Csv file named ‘Combined\_Sales\_Data’ is created to combine all these sales data of the year 2019. This file has 1,85,688 rows and 6 columns in it.

The dataset has the following columns.

* + - * Order ID: Represents the unique number associated with each number for the order identification.
* Product: Product name
* Order Quantity: Indicated the quantity of product being ordered.
* Price Each: Price of each unit
* Order Date: Contains date and time information of order placed.
* Purchase Address: Address of the customer. It contains street, city, state and pin code information in it.

The following columns can be derived from the existing ones.

* Generated Revenue: It indicates the revenue generated by Kmart on each order, which is Order Quantity multiplied by the Price Each.
* Order Date: Date of the order.
* Order Day: Day of a month on which an order is placed.
* Order Month: Month part of the Order Date.
* Order Time: Time of the day on which the order is placed.
* Street Address: Street Address element of the customer address.
* City: City of the customer address.
* State: State of the customer address.
* Pin code: Pin code of the customer address.

**Business Problem Statements:** Observations on the following are made.

1. Which day of the month recorded the highest sales?
2. During the month with the highest sales, which specific day recorded the highest sales?
3. Which specific day of month recorded the highest sales annually?
4. During which time of the day do people prefer to purchase products?
5. Which state recorded the highest number of sales, and within that state, which city ranked at the top?
6. What was the highest-value order placed in San Francisco, California?
7. What was the most modest purchase made at Kmart?
8. What is the most affordable and the most expensive item that Kmart offered for sale?
9. Which items received the most and least orders at Kmart?
10. Best time to display advertising to maximize sales?

**Issues:**

1. The Order Date column contains data in multiple formats.

E.g. YYYY-MM-DD, DD-MM-YYYY, DD-MM-YY

So, for this reason, formatting into single format is not possible in SQL. Need to solve this task in excel using formulas, formatting function and also data literacy.

This column contains data in below mentioned two formats.

* 02/17/19 18:29 = MM/DD/YY HH:MM
* 2002-07-19 14:33 = In this format, first 4-digit part represents the month of purchase out of which first two digits are unwanted, next two digits represents the day of purchase, and last two digits represents the year of purchase. The above-mentioned conclusion is derived by analysing the sales data of individual months.

1. The Purchase\_Time\_Slot has null values in it due to incorrect logic building.

The records in which Order Date contains single digit hour information has null values in Purchase\_Time\_Slot column.

This is taken care of by deleting the whole column and re-building it with the new logic. All these steps are commented off in main script.

**Tools Used:**

* Microsoft SQL Server Management Studio
* MSSQL (For Data Cleaning & Transformation)
* MICROSOFT EXCEL
* POWER BI (For Data Visualization)

**Insights:**

* In the evening, specifically between 7 and 8 PM, customers tend to place high-value orders, significantly contributing to Kmart's annual revenue. Conversely, the time period from 2 to 6 in the morning was the least preferred for customers to place orders.
* December emerged as the most profitable month, generating around $4,608,295.70 in revenue, making it a standout month along with October in terms of financial performance.
* The revenue generated on the 9th of each month was notably higher, totaling $1,169,049.21, compared to other days. The top two days annually for total revenue were the 9th and 10th, while the 30th and 31st were the least favorable.
* California (CA) leads in terms of revenue generation, with San Francisco and Los Angeles topping the list, generating $8,254,743.55 and $5,448,304.28, respectively. On the contrary, Portland city in the state of Maine (ME) contributed the least, generating only $449,321.38 to Kmart's overall revenue growth.
* The largest order in terms of revenue in San Francisco, California, was the purchase of 2 MacBook Pro Laptops on April 27, 2019, at 21:01 PM from 668 Park St, San Francisco, CA, amounting to $3,400.
* The least ordered products on Kmart were the LG Dryer and LG Washing Machine, with 646 and 666 orders, respectively, throughout the year 2019.
* The cheapest item on Kmart's list was a 4-pack of AAA batteries, costing $2.99, while the most expensive product was the MacBook Pro Laptop, priced at $1700.
* The USB-C Charging Cable topped the list as the most ordered product on Kmart.
* January and November had the lowest monthly orders, 9699 and 11603, respectively, contrary to expectations for peak orders. This suggests an opportunity for strategic advertising and promotions to attract more customers and boost revenue during these slower periods.

**Visualizations:**

A screenshot of a graph

Description automatically generated

Figure 1. Sales by time.

A screenshot of a pie chart

Description automatically generatedFigure 2. Sales by demographics.

A screenshot of a graph

Description automatically generatedFigure 3. Sales by Products