Analysis of Superstore Dataset:

Data Extraction, Cleaning, Loading and Transformation:

The Superstore dataset includes information about the products, customers, and sales associated with a retail store. The dataset contains 21 columns and multiple rows with information on order ID, order date, ship date, ship mode, customer ID, customer name, segment, postal code, region, product ID, category, sub-category, product name, quantity, discount, buying price, and selling price.

In order to prepare the dataset for efficient use, we need to clean and transform it. We can begin by removing empty rows and columns from the dataset. Then we can change the data type of certain columns as required. For instance, the order date and ship date columns should be converted to a date format, while the discount column should be converted to a decimal format.

The values in the Ship Mode column should be standardized. For instance, we can replace FC with First Class. The address column can be split into City, State, Country, and Pin code columns for better analysis.

Data Modelling:

Once the data has been cleaned and transformed, we can create a STAR schema for better performance of the analysis. The schema shall have a central fact table, Orders, and three-dimension tables, Order details, Customer, and Product. We need to remove duplicate rows from the newly created dimension tables and ensure there are no empty rows.

After creating the tables, we need to ensure one-to-many relationships are created between dimensions and the fact table. This will allow us to analyze the data more effectively and draw insights from it.

Insights:

We can use Power BI to create visuals that will help us draw insights from the data. For instance, we can create a scatter plot to show the relationship between buying price and selling price. This can help us identify if there are any products that are not selling as expected. We can also create a bar chart to show the sales performance of different regions, categories, and sub-categories.

From the data, we can draw insights on the sales performance of different regions, categories, and sub-categories. We can identify the best-selling products and use this information to optimize the supply chain. We can also identify customer segments that are most profitable and develop targeted marketing campaigns for them.

Calculation Purpose and Method:

To calculate the profitability of a product, we can subtract the buying price from the selling price and divide the result by the selling price. This will give us the profit margin as a percentage.

Profit Margin = (Selling Price - Buying Price) / Selling Price

We can use this calculation to identify which products have the highest profit margins and optimize the supply chain accordingly.

Conclusion:

In conclusion, the Superstore dataset contains valuable information that can be used to optimize the supply chain and develop targeted marketing campaigns. By cleaning and transforming the data and creating a STAR schema, we can draw insights from the data more effectively. We can use Power BI to create visuals that will help us identify trends and patterns in the data. With this information, we can make data-driven decisions that will improve the profitability of the retail store.