



Supply Chain Analysis (Sales, Profit and Impact Summaries)

A case study of Dataco Supply chain dataset

Sadakhya Narnur

DATA 230 - Data Visualization for Data Analytics

Department of Applied Data Science

San Jose State University

California

Introduction

Often customers and stakeholders demand for better visibility into the last mile of the supply chain, there is a need to have a consistently good idea on the way things work throughout. Beginning from the order of a product by a customer, the acquisition of the product, stocking of the products, managing the inventory, managing the logistics, keeping track of the profits and losses, and so on.

Objectives

The aim of this project is to draw insights on the supply chain dataset picked and give the reader a clear idea of the Sales, Profit and Impact they have to be prepared for with the data provided. The dataset “ DataCo Supply Chain” is available on Kaggle. It is the dataset built for analysis on the company DataCo Global. It consists of about 53 columns and 180519 rows representing customer data, orders data, products, location and year wise data.

This project is an attempt to put into practice all the concepts learned throughout the course. A clear description of principles employed are discussed in later sections.

Steps to develop

- Data Collection : The dataset is downloaded from Kaggle and stored as csv in drive as DataCoSupplyChainDataset.

Link:

<https://www.kaggle.com/datasets/shashwatwork/dataco-smart-supply-chain-for-big-data-analysis?select=DataCoSupplyChainDataset.csv>

- Data cleaning : The dataset is read into a Colab Notebook and checked for all the columns.

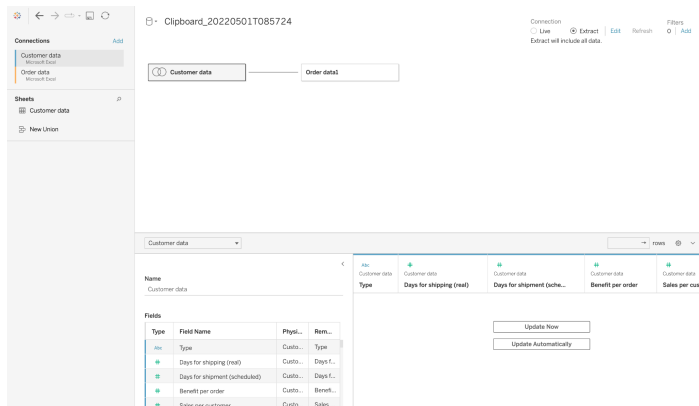
```
df.columns
Index(['Type', 'Days for shipping (real)', 'Days for shipment (scheduled)',
      'Benefit per order', 'Sales per customer', 'Delivery Status',
      'Late_delivery_risk', 'Category Id', 'Category Name', 'Customer City',
      'Customer Country', 'Customer Email', 'Customer Fname', 'Customer Id',
      'Customer Lname', 'Customer Password', 'Customer Segment',
      'Customer State', 'Customer Street', 'Customer Zipcode',
      'Department Id', 'Department Name', 'Latitude', 'Longitude', 'Market',
      'Order City', 'Order Country', 'Order Customer Id',
      'Order date (DateOrders)', 'Order Id', 'Order Item Cardprod Id',
      'Order Item Discount', 'Order Item Discount Rate', 'Order Item Id',
      'Order Item Product Price', 'Order Item Profit Ratio',
      'Order Item Quantity', 'Sales', 'Order Item Total',
      'Order Profit Per Order', 'Order Region', 'Order State', 'Order Status',
      'Order Zipcode', 'Product Card Id', 'Product Category Id',
      'Product Description', 'Product Image', 'Product Name', 'Product Price',
      'Product Status', 'shipping date (DateOrders)', 'Shipping Mode'],
      dtype='object')
```

Next the dataset is checked for NULL values and it was observed that Order Zip Code and Product Description had the maximum NULL values and are not really of interest for our visualization hence dropped the two columns.

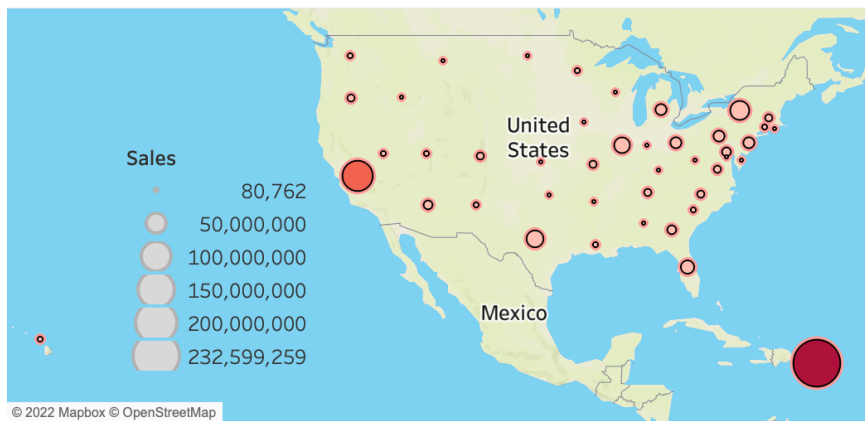
```
Sales                                0
Order Item Total                     0
Order Profit Per Order               0
Order Region                        0
Order State                         0
Order Status                        0
Order Zipcode                       155679
Product Card Id                     0
Product Category Id                 0
Product Description                  180519
Product Image                       0
Product Name                        0
Product Price                       0
```

Inorder to maintain the anonymity and secure the privacy of customers the columns regarding any personal data like Customer Email, Customer Fname, Customer Lname and Customer Password are removed.

- The dataset is too huge and for convenience it is split to Order and customer datasets to separate the Customers related data from Orders data with still Order Id as common key in both.
- Importing Data: The two csv files Customer and Order are imported and exported to Tableau for further analysis.

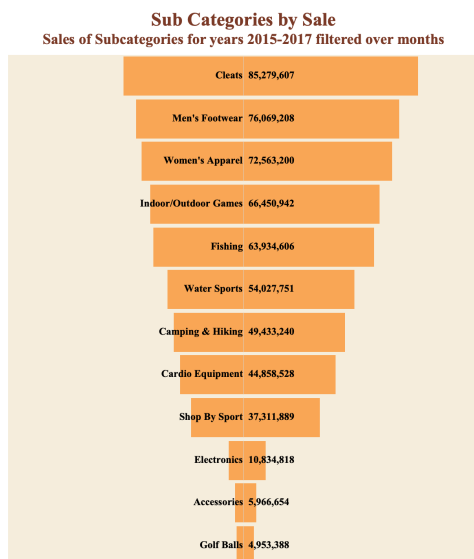


- Sales for Country: The Geospatial chart is visualized as a point map for the orders data showing the orders over the states in the US based on Latitude and Longitude columns.



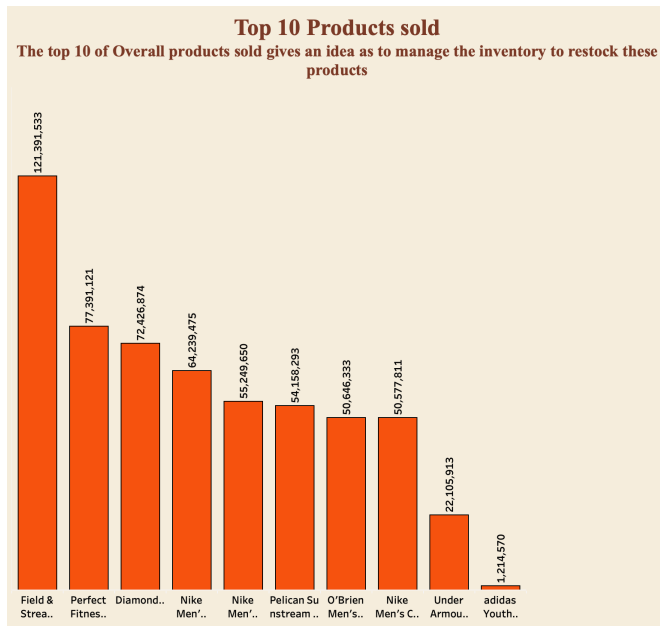
The sizes of the points shows the order's intensity in each region of the country.

- Sub Categories by sale: Based on Sales and Categories columns the graph is visualized.



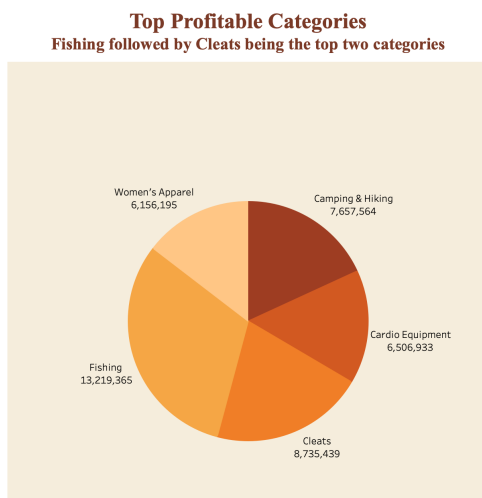
The graph gives a good understanding of the sub categories sales in a descending order showing the products of focus in years 2015 to 2017.

- Sales by product: Based on Products and Sum of Sales bar graph is plotted.



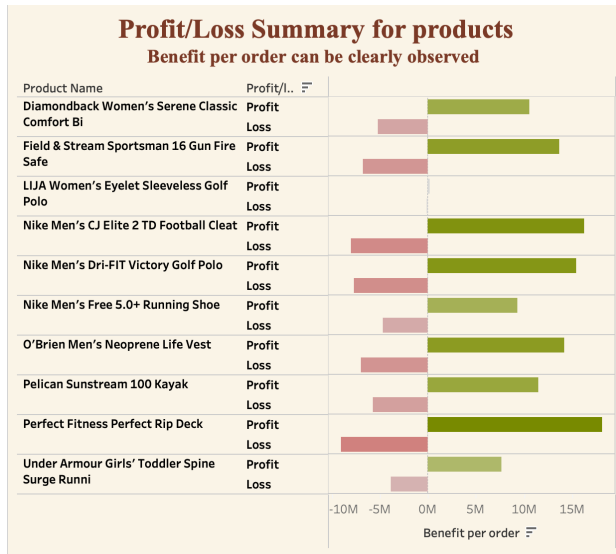
This graph shows the top products ordered by customers which gives a picture of what are the products that could be focused on.

- Top Profitable Categories: Top 5 profitable categories are shown based on Category and Benefit per Order columns.



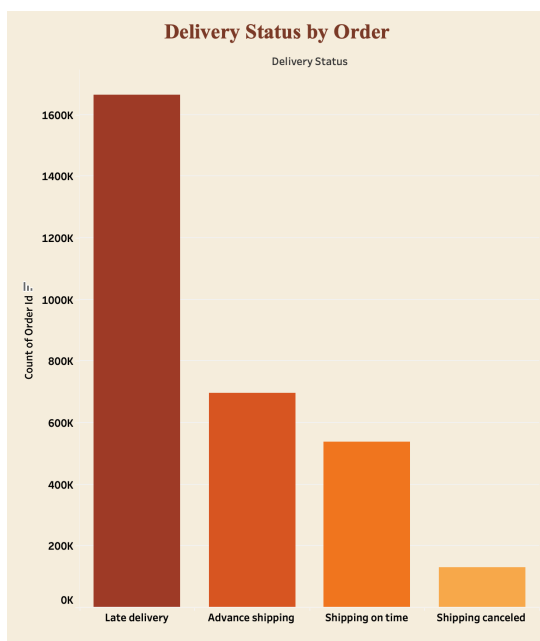
It summarizes the top 5 categories that have been profitable for the company sales.

- Profit/loss summary for products: A tabular summary of Profit/Loss of products is shown as a tabular format with benefit shown as bar graph.



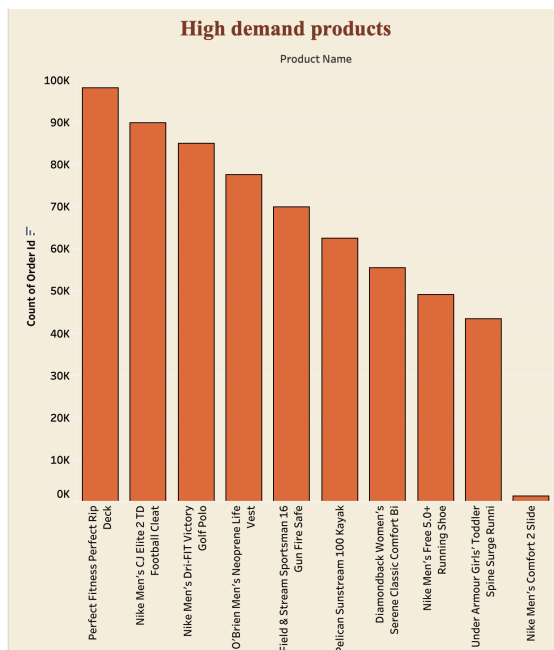
It gives a good idea of the profits and loss obtained per order of the products.

- Delivery status by order: A bar chart is visualized based on Delivery status and Order id columns.



It shows that most of the orders are under late delivery which could be a concern to the company.

- High demand products based on delivery status: A bar chart to visualize the top 10 highly demanded products for different delivery status like Advance shipping, late delivery, shipping canceled and shipped on time. It is built on Product name, Order date, category and delivery status.



The filter for the graph available gives a good picture of the high demand products for different delivery status.

- Once all the individual worksheets are prepared on Tableau a dashboard is created with all these sheets. The graphs are organized into 3 containers for Sales, Profit and Impact to meaningfully convey the story of the insights obtained.

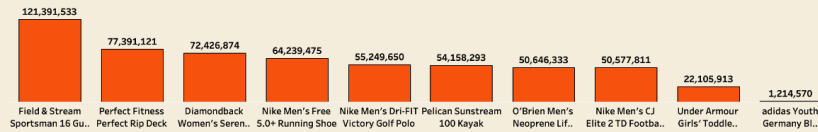
The Dashboard named “Supply Chain Dashboard” can be seen below:

Supply Chain Analysis (Sales, Profit and Impact Summaries)

A case study of Dataco Supply chain dataset

Top 10 Products sold

The top 10 of Overall products sold gives an idea as to manage the inventory to restock these products



Sales Summary

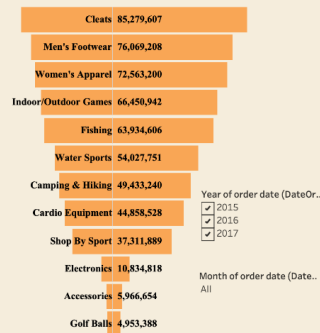
Sales: 603,784,093
Profit: 65,074,682
Quantity: 6,591,463

Sales by States in United States



Sub Categories by Sale

Sales of Subcategories for years 2015-2017 filtered over months

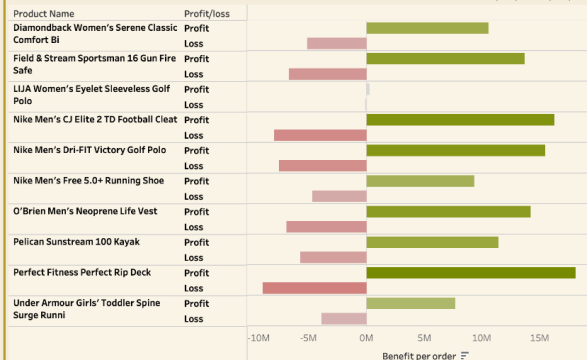


Sales Summary:
Overall sales yearly and categorized gives a good overview on the sales so far

Profits Summary

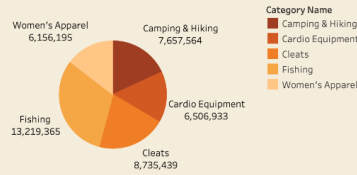
Profit/Loss Summary for products

Benefit per order can be clearly observed



Top Profitable Categories

Fishing followed by Cleats being the top two categories



Profit Summary:
Fishing products are more profitable and depending on the profit/loss for each category strategic decisions could be taken.

Insights that could Impact the supply chain

Delivery Status by Order



Insight:
It is observed that most of the orders are in Late delivery status which could greatly impact the supply chain

High demand products



Delivery Status:
● Advance shipping
○ Late delivery
○ Shipping canceled
○ Shipping on time
Limit:
Top 10 by COUNT([Order id])

Insight:
Depending on the filters based on delivery status the high demand products are mainly Perfect Fitness Perfect Rip Deck.

Key principles employed

● Interactions

1. Tooltips / Hover Help: All the graphs in the Dashboard are enabled with tooltips and hover help which gives a brief description of data points.
2. Selection: The bar graphs all the chart components can be selected and deselected.
3. Zooming and Scaling : The point map provides zooming and all charts can accommodate scaling based on the filters applied.

4. Storytelling: The entire dashboard includes Headings and descriptions indicating the takeaways from the graph with components separated horizontally for clear storytelling.
 5. Sorting: The tabular representation has an option to sort alphabetically the product names.
 6. Filters: A yearly and monthly filter is available for Sub categories by sale chart, selectable highlighter for pie chart and bars. A delivery status filter is available for the High demand products bar chart.
- Graphic Design Principles
 1. Text and Typography : The text levels show the main heading in larger font. The next main features we want to address are the chart headings in second level and the descriptions in smaller font size.
 2. Color Theory and Color Use: Safe color palette - The usage of colors is based on the set standards for safe colors.
 3. Layout and White Space: The graphs and dashboard is entirely organized into containers and negative spaces are given to ensure the proper layout.
 4. Consistency and Design Vocabulary : An overall consistency in font face as well as vocabulary is maintained throughout the dashboard. Negative spacing is used where necessary to show a proper alignment of components.
 5. Visual Encoding: Besides the colors the graphs are drawn such that the data is visualized without solely relying on the color. The points in the point map have varying sizes indicating the density of orders rather than just the intensity of the

color. The graphs also allow selecting specific data components which shows a highlighted distinction from other points which is one form of encoding.

- Cognitive Models and Gestalt Principles

1. Gestalt Principles - A number of principles are addressed in this:

Similarity - The pink and green bars in the Profit/Loss Summary for Products clearly shows that pink represents Loss and green represents Profit. The brain perceives items that are of similar colors, shapes, sizes, or orientations as belonging to the same groups.

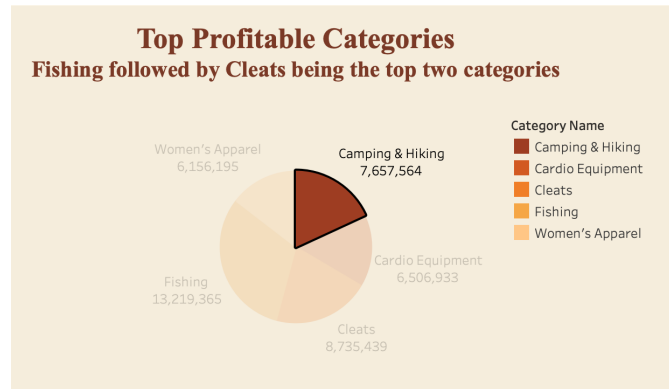
Enclosure - This principle states the fact that enclosing a group of objects brings our attention to them, they are processed together, and our mind perceives them as connected. Separating the Sales, Profit and Impact containers have made them clearly distinguishable.

Figure and Ground - The selection of Lighter background and Graphs colors, Text popping in darker shades makes it distinguishable.

Continuity - Our tendency is to see shapes as continuous to the greatest degree possible. The human eye follows lines, curves or a sequence of shapes to create pathways. The bar graphs are ordered in decreasing order to give a visual cue as of which have highest sales compared to others.



Focal Point - This states that the elements that pop out the most are the entities that people mostly focus on. When using filters one selects a particular component it pops out while others fade giving a focal effect.



2. Infographic: The dashboard follows most of the principles of infographic with some elements taken from all the various types.
 3. Top-down Processing : A number of visual encodings are mentioned above aids the viewer towards understanding our motive.
 4. Supporting Parallel Processing: Charts are selected that are generally used for visualizing that type of data. Bar, Pie and tabular representations are selected for simplicity.
 5. Serial processing: The compartmentalizing of the charts are done to enable forced serial processing to ensure it makes it clear about a view.
 6. Pre-Attentive Visual Processing : The parallel processing support and top-down processing provided helps in pre attentive visual processing. It is easy to read while filling the data we want to convey.
- User Interface Principles : All the following principles are kept in mind while designing the dashboard.

Giving Control to the User at every possible graph.

Any action can be reversible like selection and highlighting.

Kept it Simple.

Employed Graphic Design Principles.

Employed Cognitive Principles to Organize Elements.

Relied on Recognition, not Recall.

Have been Consistent throughout.