Retail sales analytics

Power BI

by Olena Petrova, Data Analyst

Abstract

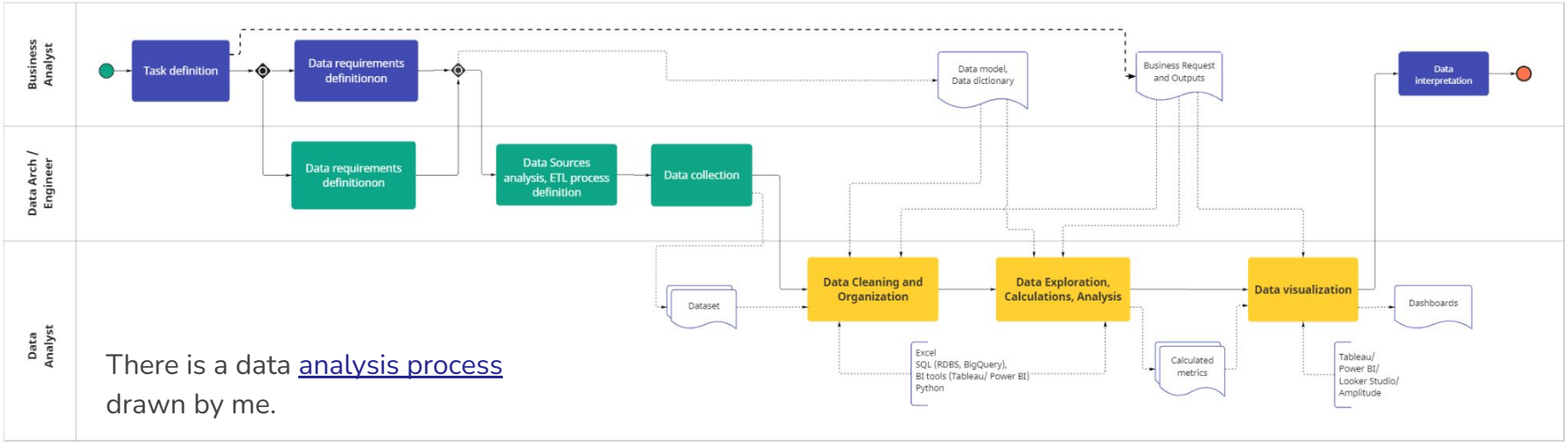
The case study is based on an artificial dataset from the internet.

In this project I performed the roles of all the stakeholders, including the business request provider.

The Business Request is formed in a way to not really get answers to business questions (but that doesn’t mean it does not), but also to demonstrate technical skills.

Business and Requirements analysis skills were applied and helped a lot to build dashboards.

Scope of the Case



There is a data analysis process drawn by me.

The current case study includes tasks for the following steps:

Task definition(p.4), Outputs requirements (p.5-7), Data preparation and organization (p.8), Data Exploration and Calculation (p.9), Data visualization and Interpretation (p.10-12).

Business request

Based on data from the order tracking system, we need:

- to analyse **Sales vs. Profit results** of business operations, evaluating Products sub-categories, Customer Segments and Regions
- to evaluate **orders to shipment processing** for specified Order Priorities,
- to get insight into **usage of shipment services**.

Data period: January - July, 2015.
Region: USA.

Questions to answer:

1. If there is any way to optimize sales to get more profit?
2. If there is a need to optimize order-to-shipment process?
    a. There should be a way to see cost in USD and UAH

Data Requirements and Output definition

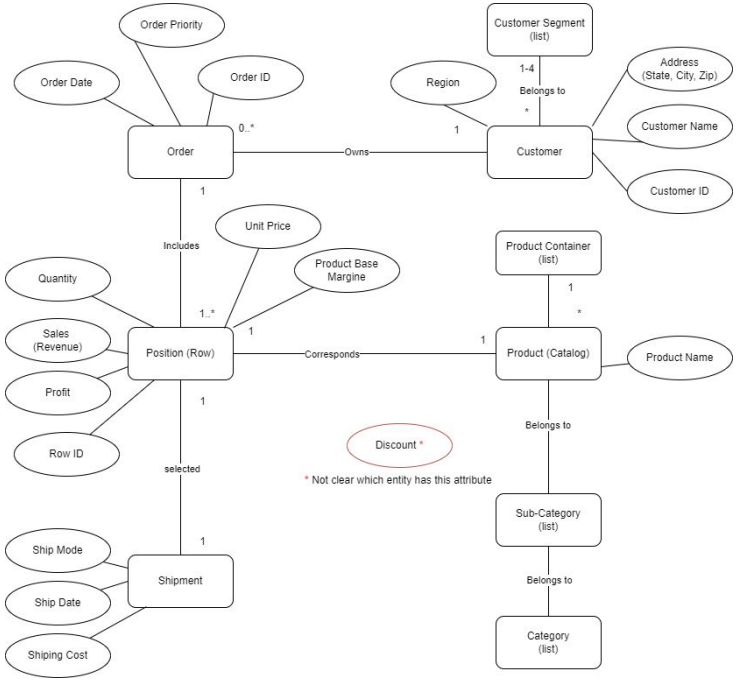
The raw data are in .xlsx format. There are 3 denormalized tables, exported from order tracking system.

Expected Output: **An analytical dashboard in Power BI**.

To enable accurate data pre-processing and analytics the following operations have been made:

- A conceptual **data model** was created (Entity Relation Diagram) to identify main entities and their attributes.
- **Data dictionary** was created to get the Business Stakeholders, Data Analysts and Engineers on the same page about terms and data formats.
- The **data aspects** for dashboard to show was created * (**requirements to the output**).

* In real situation it would be a result of interviews with the report users and subject matter experts. Here are suggestions that allow to demonstrate not just business case, but technical skills as well



Data Requirements and Output definition

Data aspects to show: Sales and Profit

Sales and Profit in the datasource table are specified **for each Position** in Order (“Row ID” in the table). Position specifies **quantity of the particular Product Name** in the Order. Each Product Name has corresponding Category and Sub-Category.

1. To demonstrate Sales (Revenue) and Profit of sales in the following dimensions (see result here):
    a. By Product Categories and Sub-Cutegories to find **the most and the least profitable ones**, and to see their part in generated revenue (%). Profitability should be provided in terms of Profit Margin. Net income (Profit) should also be present.
    b. Overall sums of Sales and Profit day by day to **identify periods with outliers and possible patterns and correlations**.
    c. Percent of Sales by Customer Segments and by Regions in Sales grand total to compare with Profits they give (see the d. requirement).
    d. Profit by Regions with clustering by Customer Segments **to identify they share in profits and visually compare with sales**.
2. Additional dashboard should be created to get insight from analysis of Profit by region and segment :
    a. **Drill down to States** when analysing Sales and Profit.
    b. Possibility **to investigate negative Profit values** should be in place.
    c. **Filter** by Customer Segments and Product Category should be in place.

Data Requirements and Output definition

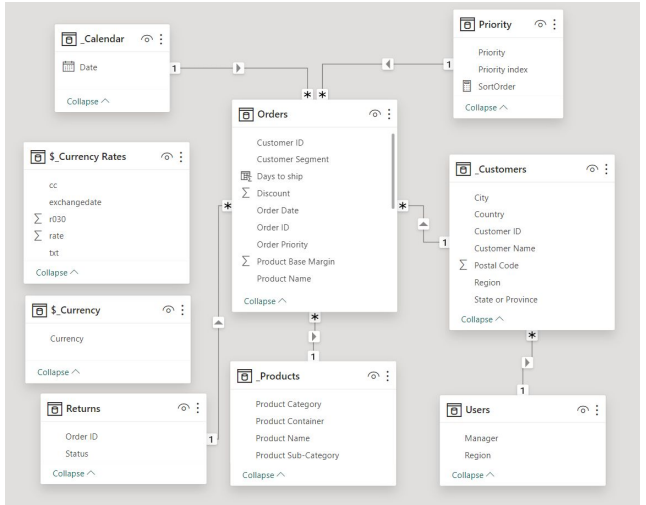
Data aspects to show: Shipment analysis

"Shipment" is the **state of Position** in Order, when it is shipped with particular Shipping Mode. Position specifies **quantity of the particular Product Name** in the Order.
Each Product Name has corresponding Product Container.
Each Shipment **inherits priority from the Order Priority** attribute.

1. To analyse whether there is a need to optimize order-to-shipment process, the data should be presented in the following dimensions:
    a. Average days to ship for each Priority with clustering by Shipping Mode - to see **if there is a significant difference in processing speed** depending on the Shipping Mode.
    b. Distribution of Shipments by days to ship with clustering by Priority - to see **if the process is stable**.
    c. It should be clear which Product Sub-categories leads in shipments - **to further analyze Shipping Mode is Cost** based on relevant Product Container
    d. Each Shipment Mode share in the total shipments, and Product Containers their are used for.
    e. **The distribution of Shipping Cost sums** by Container in each Shipping Mode with specification of the range, in which lies the biggest number of values.
    f. **Filter** by Priorities should be in place
2. To see the Shipping costs in USD and UAH, the currency selection should be in place with specification of current rate.

Data Preparation and Organization

1. Analysis of data model and **dataset preparation** (Excel):
    a. Sheets with list of unique values and their attributes were created: Products and Customers.
    b. Attributes data were deleted from the original sheet Orders.
2. Dataset **import to Power BI** and **data transformation with Power Query** :
    a. Got data from prepared Excel worksheet
    b. Headers and data formats were checked and updated updated
    c. Got data of the currency rate via WEB connector (API provided by the national bank of Ukraine)
    d. A new table with currency names was created
3. Data **organization in Power BI Model view**:
    a. Relations among tables were checked and updated
    b. Additional tables were created to sort Order Priorities and to keep dates separately



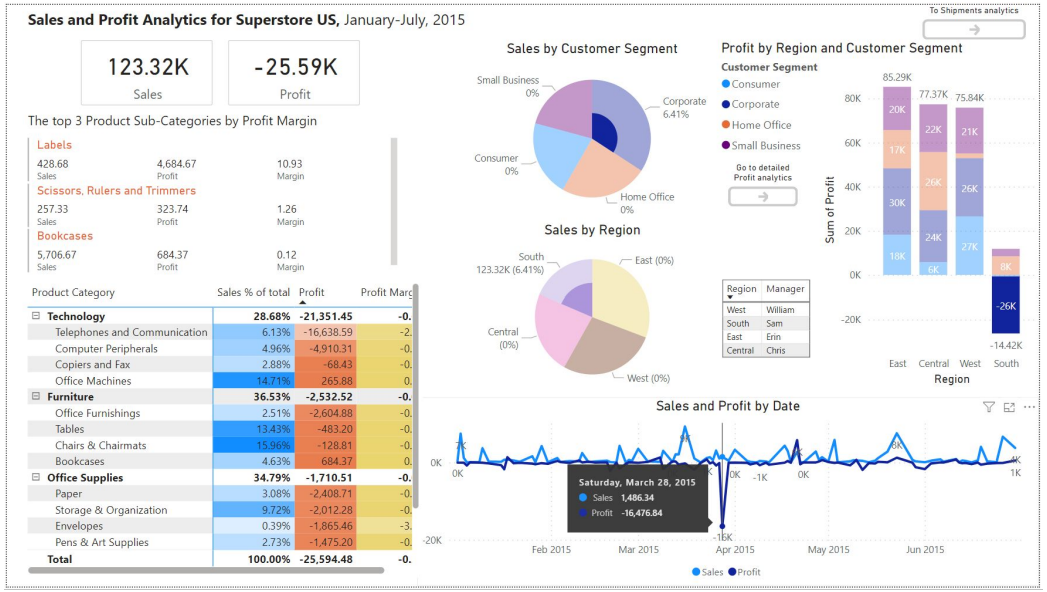
Data Exploration and Calculation

1. Data **calculation** in Power BI **Table view** and **DAX** usage:
   a. New column in the Orders table created to calculate Days to ship
   b. Calendar table was filled with data based on Order date and Ship date data
   c. Priority table was filled with data
2. New **measures** in Power BI **Report view** and **DAX** usage**:**
   a. Profit Margin measure (according to requirements to the dashboard)
   b. Shipping Cost change depending on selected Currency
   c. Sort Order for Priority

```
$_ShippingCost = IF(
    MAX('$_Currency'[Currency]) = "UAH",
    SUM('Orders'[Shipping Cost]) * SUM('$_Currency Rates'[rate]),
    SUM('Orders'[Shipping Cost])
)
```

Data visualization and Interpretation:

Sales and Profit Analytics

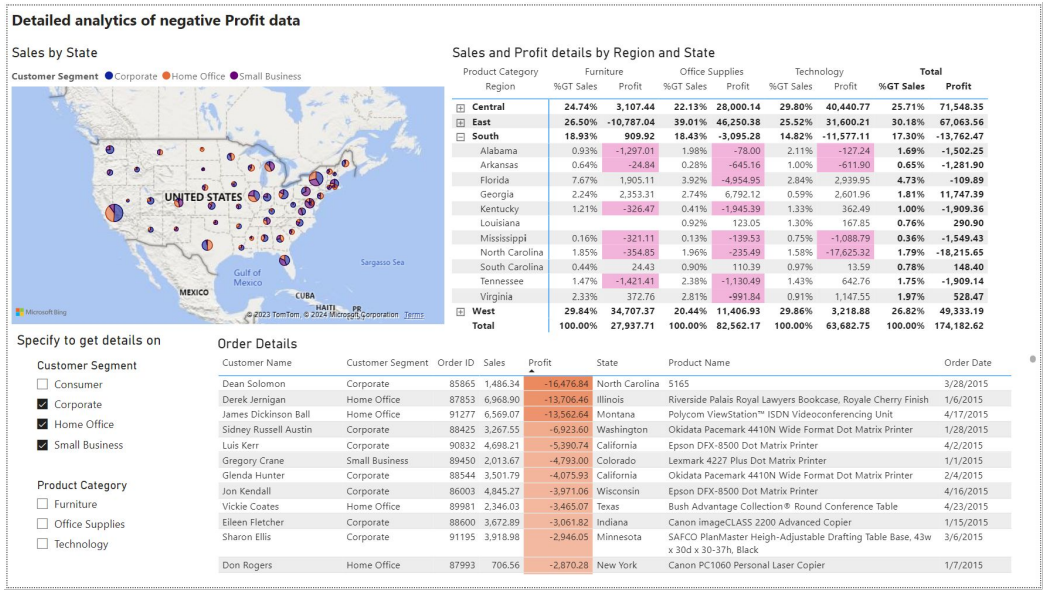


(see requirements here)

- Conditional formatting was used for a table to focus on leaders
- Time Series visualize **extreme values** for Sales and Profit. Interactivity allows to **identify details of the day transactions** in terms of
  - Customer Segment and Region,
  - Product Sub-categories
- **Insight example**:
  - For the entire period there is a significant negative profit in in South region in Corporate segment (-26K)
  - This negative Profit is mostly due to transaction(s) made on March 28 for the Telephones and Communication products (-16,5K). And that is accountability of manager Sam.
  - Now we can go to the page Profit Details to find relevant orders.

Data visualization and Interpretation:

Detailed Profit Analytics

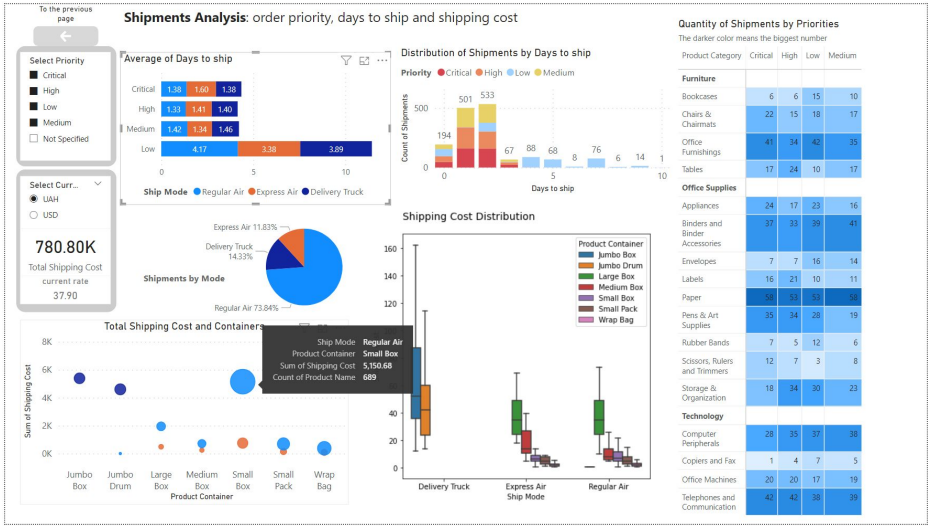


(see requirements here)

- Map shows Sales by States, clustered by Customer Segments. Map focuses on the region, selected in the table with number by state.
- Conditional formatting was used for a table to focus on negative Profit
- Slicers are applied to focus on Customer Segment and Product Category
- A list of Shipments which corresponds all the selections is provided. There is a possibility to focus on the biggest values of negative Profit for further investigation.

Data visualization and Interpretation:

Shipment Analytics



(see requirements here)

- Slicer to focus on Shipment Priority is provided.
- Average days to ship are clustered by Shipping Mode (Stacked Bar Chart).
- **Distribution** of Shipments among Days to ship **shows** that
  - For Critical, High and Medium priority distribution is normal, that means that the process is stable.
- Shipments by Mode **pie chart**, scatter chart with **sized bubbles**, the **table** with Sub-Categories and **box plot show** that
  - the majority of shipments are made via Regular Air, a half of them are in Small Box containers (Binders and Paper), and the Shipment cost is about 5-10.
  - Jumbo Box and Jumbo Drum containers (Furniture and Office Machines) are shipped via Delivery Tracks only and have the biggest price with the wide distribution of values (30-90).
- **Python script** was used to create box plot diagram.
- A container for **switching Currency** is added.