# Data visualization audience/persona

The audience of the Data visualization is the executive manager and business unit of the Rockin Robin, the famous creamery. In fact, and in the annual meeting of the managers, it’s important to summarize the financial activity of the creamery and analyze the trend, the performance of the product and other insights that could lead us to eventual improvements in the next financial planning.

So, the dashboard will mainly analyze:

* The trend and seasonality of the profit, net margins and sales
* The performance of the different brand’s product
* The geographic distribution of the brand products

Presenting this data visualization will help us extract insights to direct actions:

* Market targeting and segmentation function of the area codes;
* Better management of the costs
* Optimizing and maximizing of the profits per product.

# Data elements

First, we will categorize the variables function of their types:

## Continuous variables

We dispose of 11 continuous variables related to financial and performance indicators:

1. Profit: Profit generated measured (in $,000) = Margin - Total Expenses
2. Margin: Net sales revenue (in $,000) = Sales - Cost of Goods
3. Sales: Operating revenue (in $,000) earned from selling IceCream/Beverages
4. COGS: Cost of Goods (in $,000) that is the direct costs attributable to the production of the goods sold by the company
5. Total: Expenses Total costs in (in $,000) associated with managing and operating the business (COG not included)
6. Marketing: Marketing costs (in $,000)
7. Inventory: Total value of products and goods (in $,000) that are ready or for sale.
8. Budget Profit: Projected profit (in $,000)
9. Budget COGS: Projected cost of goods (in $,000)
10. Budget Margin: Projected margin (in $,000)
11. Budget Sales: Projected sales (in $,000)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Profit* | *Margin* | *Sales* | *COGS* | *Total Expenses* | *Marketing* |
| Mean | 61.097693 | 104.293315 | 192.987524 | 84.433145 | 54.0635593 | 31.1850282 |
| Median | 40 | 76 | 138 | 60 | 46 | 22 |
| Mode | 47 | 43 | 43 | 52 | 45 | 14 |
| Standard Deviation | 101.708546 | 94.3425225 | 151.133127 | 67.249769 | 32.3525975 | 27.0232638 |
| Variance | 10344.6283 | 8900.51154 | 22841.2221 | 4522.53144 | 1046.69057 | 730.256785 |
| Minimum | -638 | -302 | 17 | 0 | 10 | 0 |
| Maximum | 778 | 613 | 912 | 364 | 190 | 156 |
| Count | 4248 | 4248 | 4248 | 4248 | 4248 | 4248 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Inventory* | *Budget Sales* | *Budget Profit* | *Budget COGS* | *Budget Margin* |
| Mean | 749.381356 | 175.649718 | 60.913371 | 74.8305085 | 100.819209 |
| Median | 619 | 130 | 40 | 50 | 70 |
| Mode | 777 | 100 | 30 | 40 | 60 |
| Standard Deviation | 661.031896 | 148.891522 | 79.5461226 | 66.2381451 | 92.6027251 |
| Variance | 436963.167 | 22168.6854 | 6327.58562 | 4387.4a9187 | 8575.26469 |
| Minimum | -3534 | 0 | -320 | 0 | -210 |
| Maximum | 8252 | 1140 | 560 | 450 | 690 |
| Count | 4248 | 4248 | 4248 | 4248 | 4248 |

In order to describe the continuous variables, we will use some descriptive statistics and measures of central tendency. From the table above:

* The mean sales equals 193,000$ which is superior to the budget sales. The minimum of sales is 17000$ and the maximum 912,000 with a standard deviation that equals 15,000$, that is synonym of the fluctuations of sales and the presence of daily and monthly seasonality.
* The mean profit equals 61,000$ which is superior to the budget profit. The minimum of profit is –638,000$ and the maximum 778,000 with a standard deviation that equals 101,000$, that is synonym of the fluctuations of profit and the presence of daily and monthly seasonality.
* The mean margin equals 104,000$ which is superior to the budget margin. The minimum of margin is –302000$ and the maximum 613,000 with a standard deviation that equals 94,000$, that is synonym of the fluctuations of margin and the presence of daily and monthly seasonality.

## Geographic data

Location: GPS localization of the US States

Area Code: The 3 digits code of the US states

State: a Nominal geographic variable of US State

## Date data

Date: a date that present the day, month and year .

## Categorical variables

Market Market region (East, West, Central, South)

Market Size Market size as a factor of population and demand (Major vs. Small)

Product

Product ID

Product Type Type of ice cream and beverages sold

Product : Product sub-category

The Market variable is well distributed for Rockin Robin creamery. In fact, 29% of creameries are located in the East, 26% in the central, 24% in the West and 21% in tbe south

Using what you have learnt in the lectures and tutorials, first please describe each column in the dataset (note: you should describe all columns regardless of whether you will ultimately use them in your visualisation or not). Then classify the data (example, Categorical, Numerical: Discrete, Continuous / Scales of measurements: Nominal, Ordinal, Interval and Ratio / Time/Date, Location).

# Types of charts

Based on the data types you described earlier in this section, along with the objectives of the persona you described, we will describe what charts and visualisations that we will use to display our dashboard:

* Symbol map chart: We will use this chart to display the geographic distribution of the Rockin Robin creamery per state and use the Market Size via the layer size to differentiate between the small and major market.
* Line plot: We will use the line plot to visualize the evolution of the financial indicators per day and extract the characteristics of trend and seasonality of the financial indicator
* Barplot : The bar plot is used to visualize the distribution of categorical variables, we will use it to visualize the top products in each state.
* Scatterplot: In order order extract the type of relationship between continuous variables, we will use the scatter plot to see what kind link exist between the financial indicators.
* Area plot: Area plot is quiet similar to the line plot and could be used to visualize time series along with Date. It becomes more useful from time series that follow certain hierarchy and allow to sum graphically the trend of the categories to get the global evolution.
* Histogram is used for continuous variables and allows showing its distribution by defining a variable of bins used
* Pie chart is somehow similar to the barplot and present the repartition of categorical variables in terms of count or percent.
* Tree map is a map that is based on characters and the frequency of items to visualize it using squares which size change function of the frequency of the item.

# Data ethics

The ethical aspect in data analytics is very important and mostly not being taken seriously by Business Intelligence engineers and Data analysts. The respect of ethics principles is indispensable to build a great dashboard. In our dashboard, I’ve managed to respect the following ethics criteria:

* **Reliability**: Sharing reliable information is primordial to make reasonable actions without any bias. Furthermore, it’s mandatory to present reliable information to clients, customers and public audience because everyone deserves truth.
* **Clarity**: A dashboard is a powerful tool that must be clear and concise and must as possible. In our dashboard, I’ve followed all the requirements to present clear information using appropriate visualization and being sure to choose the right title and to display the adequate layers and information on the graph.
* **Impact**: A dashboard requires having an impact and not showing any useless information. Every single component of dashboard should bring some insights and lead to actions.

# Dashboard

## Color layer

I’ve used color to highlight the distribution of the products type per state, this layer has helped show that the product type sold depends significantly on the geographic location. In fact, one important insight is that the Cup and Cones are a special product highly sold in the south of US.

Also, I’ve used the color layer to differentiate between the products in the Top products and give each product each own color, that would be automatically recognized by stakeholders.

## Hierarchy

* Symbol map chart: We will use this chart to display the geographic distribution of the Rockin Robin creamery per state and use the Market Size via the layer size to differentiate between the small and major market.
* Line plot: We will use the line plot to visualize the evolution of the financial indicators per day and extract the characteristics of trend and seasonality of the financial indicator
* Barplot : The bar plot is used to visualize the distribution of categorical variables, we will use it to visualize the top products in each state.
* Scatterplot: In order order extract the type of relationship between continuous variables, we will use the scatter plot to see what kind link exist between the financial indicators.
* Tables: Tables are the most classical way to present information and probably the most used ones. Being sample to apprehend, tables are used by many BI engineers and data analysts to present information that could be understood by everyone.

## Interactive components

The Rockin Robin dataset contains many referential data that could be used as interactive component of the dashboards:

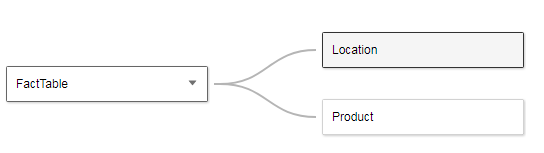
* The State filter is an important component that would be used to make a zoom or focus on a specific US state.
* The Financial indicators are also displayed as interactive component in order to make the dashboard show the zoom based on the each chosen financial indicator.
* We could also add interactive components linked to the product, the product type and the area code to make the dashboard more interactive.

## Data sources and data manipulation

The source of data is Rockin Robin creamery. It contains information on great level of details showing the daily financial indicators of the product shown per each single US area and state.

Once connecting the Tableau Desktop to our database, it’s important to build the relationship between the tables to get one tidy and clean one that contains all the relevant information is structured way to better build graphics.

For this purpose, we use the union statement in Tableau in order to link the FactTable that contains the financial indicators along the product and area codes with the reference tables that display the labels of the products and the areas hierarchically with the US state.



Finally, we obtain a structured table that contains all the needed information that we will use the build the data visualizations, dashboard and story.

We have performed many types of calculations in the dataset:

* We have a created a parameter to display the top 3 products in each state and for each financial indicator.
* We have created a calculated field to display the % of marketing cost in total expenditures.

## Other layers

* Many other layers types could be used in a graphic to enhance the number of dimension to visualize:
* Size : the size layers allows showing different mark with different size, each range represents a category or an interval of information.
* Label: Labels allows combining the graphical visualization with text information to maximize the impact on the audience.
* Tooltip: Tooltips are efficient ways to display control charts on the graph and allow the audience to follow the story that we want to convey.

# Story

In tableau, stories are an efficient way to display visualization to public audience and simultaneously provide interpretation and text analyses that everyone could understand to maximize the apprehension and the reactivity.