# CRISP-DM

## Business Understanding

### Determine business objectives:

Determine business objectives: You should first “thoroughly understand, from a business perspective, what the customer really wants to accomplish.” (CRISP-DM Guide) and then define business success criteria.

### Assess situation:

Determine resources availability, project requirements, assess risks and contingencies, and conduct a cost-benefit analysis.

### Determine data mining goals

In addition to defining the business objectives, you should also define what success looks like from a technical data mining perspective.

### Produce project plan

Select technologies and tools and define detailed plans for each project phase.

## Data Understanding

* Collect initial data: Acquire the necessary data and (if necessary) load it into your analysis tool.

<https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?select=product_category_name_translation.csv>

This is a Brazilian ecommerce public dataset of orders made at [Olist Store](http://www.olist.com/" \t "_blank). The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. We also released a geolocation dataset that relates Brazilian zip codes to lat/lng coordinates.

This is real commercial data, it has been anonymised, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses.

* Describe data:
* Customer Data
* Geolocation Data
* Order Items Data
* Order Payment Data
* Order Review Data
* Orders Data
* Products Data
* Seller data
* Product Category Name Trandlatin

Nvarchar data types were converted to

* Explore data: Dig deeper into the data. Query it, visualize it, and identify relationships among the data.
* Verify data quality: How clean/dirty is the data? Document any quality issues.

Data Preparation

Modeling

Evaluation

Deployment

orders:

- order\_id

- customer\_id

- order\_status

- order\_purchase\_timestamp

- order-approved\_at

- order\_delivered\_carrier\_date

- order\_delivered\_customer\_date

- order\_estimated\_delivery\_date

order\_items:

- order\_id

- order\_item\_id

- product\_id

- seller\_id

- shipping\_limit\_date

- price

- freight\_value

order\_payment:

- order\_id

- paymenr\_sequential

- payment\_type

- payment\_installments

- payment value

order\_reviews:

- review\_id

- order\_id

- review\_score

- review\_comment\_title

- review\_comment\_message

- review\_creation\_date

- review\_answer\_timestamp

products:

- product\_id

- product\_category\_name

- product\_name\_length

- product\_desription\_length

- product\_photos\_qty

- product\_weight\_g

- product\_height\_cm

- product\_width\_cm

sellers:

- seller\_id

- seller\_zip\_code\_prefix

- seller\_city

- seller\_state

customers:

- cutsomer\_id

- cutsomer\_unique\_id

- customer\_zip\_code\_prefix

- customer\_city

- customer state

geolocation:

- geolocation\_zip\_code\_prefix

- geolocation\_lat

- geolocation\_lng

- geolocation\_city

- geolocation\_state

With the datasets you have, you can create a comprehensive dashboard for Sales Performance Analysis. Here’s a breakdown of what you can include in your dashboard, organized by different areas of analysis:

### \*\*1. Sales Overview\*\*

- \*\*Total Sales\*\*:

- Metric displaying total sales revenue over a specific period.

- Visualization: KPI card.

- \*\*Sales Trend\*\*:

- Line chart showing sales revenue over time (daily, weekly, monthly).

- Data Source: `order\_purchase\_timestamp`, `order\_payment`.

- \*\*Sales by Product Category\*\*:

- Bar or pie chart showing total sales revenue by product category.

- Data Source: `order\_items`, `products`, `order\_payment`.

- \*\*Top Selling Products\*\*:

- List or bar chart of top-selling products by revenue or quantity sold.

- Data Source: `order\_items`, `products`.

### \*\*2. Customer Analysis\*\*

- \*\*Customer Segmentation\*\*:

- Pie chart or bar chart showing the distribution of customers based on purchase frequency, order value, or other metrics.

- Data Source: `orders`, `order\_payment`, `customers`.

- \*\*Customer Lifetime Value (CLV)\*\*:

- Visualization of CLV by customer segment.

- Data Source: `orders`, `order\_payment`.

- \*\*Customer Acquisition Trends\*\*:

- Line chart showing the number of new customers acquired over time.

- Data Source: `orders`, `customers`.

- \*\*Repeat vs. New Customers\*\*:

- Comparison of sales or order volume from repeat customers vs. new customers.

- Data Source: `orders`, `customers`.

### \*\*3. Order Insights\*\*

- \*\*Order Volume\*\*:

- Trend chart showing the number of orders over time.

- Data Source: `orders`.

- \*\*Order Status Breakdown\*\*:

- Pie chart showing the percentage of orders in different statuses (e.g., delivered, pending, canceled).

- Data Source: `orders`.

- \*\*Order Fulfillment Time\*\*:

- Average time from order purchase to delivery, displayed as a line chart or bar chart.

- Data Source: `orders`.

### \*\*4. Product Analysis\*\*

- \*\*Product Performance\*\*:

- Visualization of revenue and quantity sold by product.

- Data Source: `order\_items`, `products`, `order\_payment`.

- \*\*Product Categories Performance\*\*:

- Bar chart showing sales and number of orders by product category.

- Data Source: `order\_items`, `products`, `order\_payment`.

- \*\*Inventory Analysis\*\*:

- Visualization of product stock levels, if applicable.

- Data Source: Product data might need to be combined with inventory data if available.

### \*\*5. Seller Performance\*\*

- \*\*Top Performing Sellers\*\*:

- List or bar chart of sellers ranked by total sales revenue or number of orders.

- Data Source: `order\_items`, `sellers`, `order\_payment`.

- \*\*Seller Location Analysis\*\*:

- Map visualization showing seller locations and performance by region.

- Data Source: `sellers`, `geolocation`.

### \*\*6. Payment Analysis\*\*

- \*\*Payment Methods\*\*:

- Pie chart showing the distribution of payment methods used (e.g., credit card, PayPal).

- Data Source: `order\_payment`.

- \*\*Payment Success Rate\*\*:

- Metric showing the percentage of successful payments vs. failed transactions.

- Data Source: `order\_payment`.

- \*\*Average Payment Value by Method\*\*:

- Comparison of average payment values by different payment types.

- Data Source: `order\_payment`.

### \*\*7. Geolocation Insights\*\*

- \*\*Sales by Region\*\*:

- Heat map or bar chart showing sales performance by region or city.

- Data Source: `orders`, `customers`, `geolocation`.

- \*\*Customer Distribution\*\*:

- Map visualization showing the concentration of customers by city or state.

- Data Source: `customers`, `geolocation`.

- \*\*Shipping vs. Delivery Times\*\*:

- Visualization comparing estimated delivery times vs. actual delivery times by region.

- Data Source: `orders`, `geolocation`.

### \*\*8. Review Analysis\*\*

- \*\*Average Review Score\*\*:

- Line chart or bar chart showing the average review score over time.

- Data Source: `order\_reviews`.

- \*\*Review Sentiment\*\*:

- Visualization of sentiment analysis of review comments (positive, neutral, negative).

- Data Source: `order\_reviews`.

- \*\*Top Reviewed Products\*\*:

- List or chart showing products with the highest and lowest review scores.

- Data Source: `order\_reviews`, `products`.

### \*\*9. KPI Tracking\*\*

- \*\*Sales Growth Rate\*\*:

- Metric showing the percentage growth in sales over specified periods.

- Data Source: `order\_payment`.

- \*\*Customer Churn Rate\*\*:

- Metric displaying the rate at which customers stop purchasing.

- Data Source: `orders`, `customers`.

- \*\*Profit Margins\*\*:

- Visualization showing overall and product-specific profit margins.

- Data Source: `order\_items`, `order\_payment`.

### \*\*Dashboard Layout Suggestions\*\*

- \*\*Top Row\*\*: Summary KPIs (Total Sales, Total Orders, Customer Growth).

- \*\*Left Column\*\*: Sales overview and order insights.

- \*\*Center\*\*: Customer analysis and product performance.

- \*\*Right Column\*\*: Seller performance, payment analysis, and geolocation insights.

- \*\*Bottom Section\*\*: Review analysis and KPI tracking.

By including these elements in your dashboard, you'll provide a comprehensive view of sales performance that supports better decision-making and strategic planning.