



Data Analysis in e-Commerce

Maria-Cristina ENACHE*

ARTICLE INFO

Article history:

Accepted March 2023

Available online March 2023

JEL Classification

K22, M21

Keywords:

e-commerce, data analysis,
methods

ABSTRACT

The race for which businesses of various types are in today, they need every advantage. These companies that want to stay in business, competitive and profitable can improve their chances of success by making smart choices, such as the strategy of collecting as much useful information as possible, analyzing them to make the best decisions, but especially on time. That's why we can no longer discuss successful businesses in 2023 without implementing data analysis methods. In this article we will present some such methods.

© 2023 EAI. All rights reserved.

1. Introduction

Data analysis is a process of collecting and especially analyzing large amounts of information from various fields. In e-commerce and businesses developed through it, data analysis primarily involves collecting data about buyer habits and website performance to improve the online shopping experience and increase sales.

There are many ways to analyze data, depending on the type of data you have and your analysis goals. Here are some examples of data analysis methods:

1. Statistical Analysis: This involves the use of various statistical methods to extract information from the data. This can include regression analysis, analysis of variance, principal factor analysis, and more.

2. Exploratory data analysis: This involves exploring the data through various graphs and visual methods to identify patterns and trends in the data. This can include creating graphs, box plots, scatter plots, and more.

3. Social Network Analysis: This involves analyzing social data to identify relationships and patterns between individuals and groups. This can include social media analysis, social influence analysis and more.

4. Text Analysis: This involves analyzing written text to identify patterns and trends in text content. This can include sentiment analysis, thematic analysis, sentiment analysis and more.

5. Simulation Analysis: This involves the use of mathematical and statistical models to simulate the behavior and results of data in a hypothetical scenario. This can include event simulation, resistance simulation, and more.

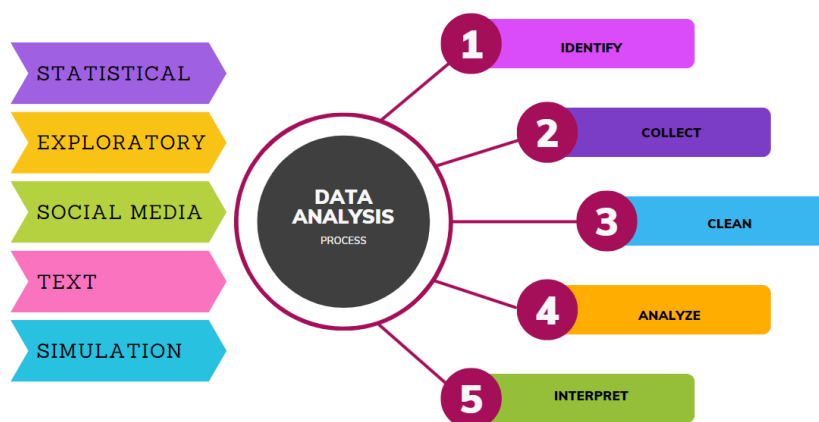


Figure 1. Data analysis process

2. Data analysis methods

Regardless of the data analysis method you choose, it's important to have a good understanding of the data you have and your analysis goals. This will help you select the best methods and analysis tools to get valuable and meaningful insights from your data.

We will detail how sales data analysis could be used concretely using the 5 methods mentioned:

- a) **Statistical analysis:** We may use regression analysis to identify factors that influence our sales and estimate the impact of each factor on them. We may also use analysis of variance to identify significant differences between groups of customers or products. Regression analysis is a method of statistical analysis used to identify the relationship between a dependent variable (in our case, sales) and one or more independent variables (or factors) that might influence our sales. Regression analysis can be used to identify the factors that have the greatest impact on our sales, as well as to estimate the magnitude of that impact.

In order to perform a regression analysis on our sales, we need to have data on sales and the factors that might influence them. For example, if we are a food company, our factors could be our prices, our advertising, the season, or weather conditions.

To start the analysis, we can use statistical analysis software such as R or Python, or we can use analysis tools such as Microsoft Excel or Power BI. Next, we follow the following steps:

- **Data Collection:** We collect data about our sales and our factors. The data should be as detailed as possible and cover a significant period.
- **Identifying the type of regression:** We identify the type of regression we want to use. In our case, we can use simple linear regression, which identifies a direct relationship between a single independent variable and the dependent variable.
- **Data analysis:** We analyze our data and identify patterns and trends in it. This can be done by using graphs and charts, such as scatter plots or box-and-whisker plots.
- **Regression calculation:** We calculate our regression using the software of your choice. This step involves using a mathematical formula to estimate the value of the dependent variable based on the values of the independent variables.
- **Interpretation of results:** We interpret our results and identify the factors that have the greatest impact on our sales. This can be done by examining the regression coefficients, which represent the magnitude of the impact of each factor on our sales.

In conclusion, regression analysis is a powerful method of statistical analysis that can help identify the factors that influence our sales. By using this method, we can develop more effective strategies to grow our business and improve our performance over time.

- b) **Exploratory Data Analysis:** We may use various graphs, such as line graphs, bar graphs, and boxplots, to identify patterns and trends in our sales data. For example, we can use a line graph to visualize the evolution of our sales over time, or a bar graph to compare the sales of our products in different regions. Exploratory data analysis is a method used to examine data sets and identify patterns, trends, and relationships between variables. In our case, we can use exploratory data analysis to identify the factors that influence our sales. This method allows us to explore our data and identify relationships between variables before performing more advanced statistical analysis such as regression.

Here are some steps we can take to perform exploratory data analysis for our sales:

- We start by examining our data and understanding the structure of the data set. We check if the data set is complete, if there is any missing data or if there are anomalies in the data.
- We identify the variables in the data set and analyze each variable separately. For each variable, we should examine the minimum and maximum values, the mean and median values, the standard deviation, and the distribution of the values. We can also use charts to visualize the distribution of values, such as histograms or box-and-whisker charts.
- We analyze the relationships between the variables. We identify variables that could influence our sales and analyze the relationship between these variables and sales. We can use scatterplots to visualize the relationship between two variables and calculate the Pearson correlation coefficient to measure the strength of the relationship.
- We identify atypical values (outliers) and analyze how they affect the data set and the relationships between the variables. We can use box-and-whisker plots to identify outliers and decide whether they should be removed from the data set.
- We identify patterns and trends in the data set. We identify variables that could influence our sales and analyze how these variables change over time. We can use line charts to visualize the evolution of variables over time and analyze data trends and seasonality.

In conclusion, exploratory data analysis is an important method of data analysis that can help us identify the factors that influence our sales. This method allows us to visualize and understand our

data before performing more advanced statistical analysis and can help develop more effective strategies to grow our business and improve our performance over time.

- c) **Social Media Analytics:** We may use social media analytics to identify our most influential customers and understand how they influence our sales. We may also use social influence analysis to identify the individuals or groups that have the greatest influence on our customers' purchasing decisions. [6]

Social network analysis is a method used to identify and analyze the interactions and relationships between individuals, groups, or organizations on different social media platforms. In our case, we can use social media analysis to identify the factors that influence our sales.

Here are some steps we can take to conduct a social media analysis for our sales:

- We identify social media platforms relevant to our business and collect available data. We may use social media analytics tools to collect data about user interactions with our business and our competitors.
- We analyze the types of interactions and user influences. We identify influential users and groups of users who are discussing our and our competitors' products or services. We may also analyze the sentiment of conversations about our business and our products.
- We identify the factors that influence our sales. We analyze data to identify patterns and trends in conversations about our business and our products. We can identify keywords and themes that appear frequently in conversations and use text analysis to identify the sentiment of conversations.
- We analyze interactions between users and our business. We identify positive and negative interactions between users and our business and can identify when users need additional support or guidance.
- We use data to develop more effective strategies. We use the collected data to develop more effective marketing strategies and to improve users' experience with our business.

In conclusion, social media analysis is an important data analysis method that can help us identify the factors that influence our sales. This method allows us to visualize and understand interactions between users and the business.

- d) **Text Analysis:** We may use sentiment analysis to understand how our customers perceive our products and services. We may also use topic analysis to identify the most common words and topics mentioned by our customers in their comments. [6]

Text analytics is a data analysis method that can be used to identify factors that influence sales. This method involves analyzing a textual data set to identify relevant patterns, trends, and themes.

Here are some steps we can take to perform a text analysis to identify the factors that influence our sales:

- We collect relevant textual data. We may collect comments, reviews, feedback, social media messages and other relevant textual data about our products and services.
- We clean and process the textual data. In this step, we remove redundant data, remove stop words, and apply natural language processing (NLP) techniques to improve data quality.
- We identify key themes and patterns in textual data. We use text analysis techniques such as word frequency analysis, clustering, and sentiment analysis to identify key themes and patterns that appear in our data.
- We analyze the sentiment of feedbacks and comments. We identify positive and negative feedback and analyze their sentiment to understand how users perceive our products and services.
- We identify key words and phrases that appear frequently in feedback and comments. These words and phrases can help us identify factors that influence our sales, such as product quality, prices, service, or other relevant factors.
- We use data to develop more effective strategies. We use the collected and analyzed data to develop more effective marketing strategies and to improve users' experience with our business.

In conclusion, text analysis is an important data analysis method that can help us identify the factors that influence our sales. This method allows us to identify key themes and patterns in our textual data and can help develop more effective strategies to grow our business and improve our performance over time.

- e) **Simulation analysis:** We may use event simulation to estimate the impact of unforeseen events, such as pandemics or economic crises, on our sales. We may also use stress simulation to identify the most vulnerable aspects of our business and develop risk management strategies.

Simulation analysis is a data analysis method that allows us to assess the impact of factors on our business performance. This method involves using mathematical or statistical models to simulate different scenarios and identify the key factors that influence our business performance. Here are some steps we can take to perform a simulation analysis to identify the factors influencing our sales:

- We identify the key variables. First, we need to identify the key variables that could influence our sales, such as prices, stock levels, advertising, and other relevant factors.
- We develop a simulation model. To simulate different scenarios, we develop a mathematical or statistical model that considers the key variables identified in the first step. This model can be built in Excel or other analysis tools.
- We establish the input parameters of the model. In this step, we set the input parameter values for each key variable, such as price, stock level, and advertising spend.
- We run the simulation. We run the simulation model with the specified input parameters and analyze the results. This can help us identify the factors that most influence our sales and determine which of the key variables should be adjusted to improve our business performance.
- We interpret the results. We analyze the results of the simulation model to understand how different variables influence our sales. For example, we may identify that a lower price for a product could increase sales, while an increase in advertising spending would not have a significant impact on sales.
- We develop more effective strategies. We use the collected and analyzed data to develop more effective marketing strategies and to improve users' experience with our business.

In conclusion, simulation analysis is an important method of data analysis that can help us identify the factors that influence our sales. This method allows us to simulate different scenarios and identify the key variables that influence our business performance, allowing us to develop more effective strategies to grow our business and improve our performance over time.

3. Exploratory analysis of data

We will exemplify in what follows the generation of a set of sales data and we will perform an exploratory analysis of the data. To generate a sales dataset, we'll use the Python Pandas library to create a dummy dataset with sales information, including date of sale, place of sale, value of sale, and category of products sold.

```
import pandas as pd
import numpy as np
import random

# Generate sales data
date = pd.date_range(start='2020-01-01', end='2022-12-31', freq='D')
locations = ['Bucuresti', 'Cluj-Napoca', 'Iasi', 'Constanta', 'Timisoara', 'Brasov']
categories = ['Telefoane', 'Electrocasnice mari', "Electrocasnice mici", 'Carti', 'Asigurari']
sales_data = pd.DataFrame(columns=['Sale Date', 'Location', 'Value', 'Category'])
for i in range(10000):
    date = random.choice(date)
    location = random.choice(locations)
    value = round(random.uniform(10, 10000), 2)
    category = random.choice(categories)
    sales_data.loc[i] = [date, location, value, category]
# View sales data
print(sales_data.head())
```

This code generates a set of sales data for 10,000 transactions between January 1, 2020, and December 31, 2022. The next step is to perform an exploratory analysis of this data to identify factors that influence sales.

We used Seaborn which is a Python library for data visualization based on Matplotlib. It provides a simplified interface for creating attractive and informative graphs, with a set of styles and preconfigured default settings options for different types of data visualizations. Seaborn is often used in conjunction with Pandas to analyze and visualize data in an efficient and easy-to-understand way.

```
import seaborn as sns
import matplotlib.pyplot as plt
# Visualization of the distribution of sales values
sns.histplot(sales_data['Valoare'], kde=True)
```

```

plt.title('Distribuția valorilor de vânzare')
plt.xlabel('Valoare')
plt.show()

# Visualization of sales distribution by product category
sns.histplot(data=sales_data, x='Valoare', hue='Categorie', kde=True)
plt.title('Distribuția valorilor de vânzare pe categorii de produse')
plt.xlabel('Valoare')
plt.show()

# Visualization of sales distribution by locations
sns.histplot(data=sales_data, x='Valoare', hue='Locatie', kde=True)
plt.title('Distribuția valorilor de vânzare pe locații')
plt.xlabel('Valoare')
plt.show()

# Visualization of the evolution of sales over time
sales_by_date = sales_data.groupby('Data Vanzare')['Valoare'].sum().reset_index()
sns.lineplot(x='Data Vanzare', y='Valoare', data=sales_by_date)
plt.title('Evoluția vânzărilor în timp')
plt.xlabel('Data Vanzare')
plt.ylabel('Valoare')
plt.show()

```

4. Conclusions

In this article, we have reviewed some methods of data analysis, used successfully especially in business and especially in e-commerce. Data analysis methods must result in improved business intelligence. By using these five methods of data analysis, we can gain a deeper and more complete understanding of our sales and develop more effective strategies to grow our business.

We exemplified an exploratory data analysis by generating a sales data set for an e-commerce platform that sells several product categories.

We used the Python Pandas library as well as Seaborn to create a dataset with sales information, including date of sale, place of sale, value of sale, and category of products sold.

References

1. Hao Wang - *Application of Data Analysis in E-commerce, Highlights in Business, Economics and Management, Volume 1, 2022, p. 204.*
2. El Flah Zineb , Rafalia Najat , Abouchabaka Jaafar - *An Intelligent Approach for Data Analysis and Decision Making in Big Data; International Journal of Advanced Computer Science and Applications, Vol. 12, No. 7, p. 702, 2021*
3. Hossein H., Beneki C., Unger S - *Text Mining in Big Data Analytics, Big Data and Cognitive Computing., January 2020*
4. Enache M.C - *AI for Advertising, Economics and Applied Informatics Journal, Nr.1/2020, Universitatea „Dunărea de Jos” Galați, ISSN 1584-0409, p. 28.*
5. Elgendy, N., & Elragal, A. - *Big Data Analytics in Support of the Decision-Making Process, Procedia Computer Science, 100, p. 1071*
6. Enache M.C - *Machine Learning for dynamic pricing in e-commerce, Economics and Applied Informatics Journal, Nr.3/2021, Universitatea „Dunărea de Jos” Galați, ISSN 1584-0409*
7. <https://towardsdatascience.com/data-science-for-e-commerce-with-python-a0a97dd7721d>
8. <https://github.com/marcogdepinto/Python-for-Data-Analysis-and-Machine-Learning/blob/master/Ecommerce%20Purchases%20Exercise%20.ipynb>
9. <https://www.simplilearn.com/tutorials/data-analytics-tutorial/data-analytics-with-python>
10. <https://www.datapine.com/blog/data-analysis-methods-and-techniques/>
11. <https://www.bigcommerce.com/blog/ecommerce-analytics/>
12. <https://www.indeed.com/career-advice/career-development/data-analytics-in-ecommerce>
13. <https://scandiweb.com/blog/how-to-use-data-in-ecommerce-analytics/>
14. <https://supermetrics.com/blog/ecommerce-analytics>
15. <https://www.hotjar.com/blog/ecommerce-website-analysis/>