Executive Summary:

E-commerce Sales Performance Analysis

This report analyzes e-commerce sales performance to understand the relationship between target sales and actual sales figures.

Data-Driven Approach:

- We leveraged e-commerce data to identify factors influencing sales performance.
- Rigorous data cleaning procedures ensured the accuracy of our analysis.

Key Insights:

- Target alignment significantly impacts sales in some product categories.
- Other factors play a substantial role in influencing sales for certain categories.

Actionable Recommendations:

- Implement a "target alignment strategy" to optimize sales performance across all categories.
- Develop data-driven sales targets and targeted marketing campaigns.
- Implement a customer-centric approach to product offerings and pricing strategies.
- Enhance inventory management based on sales data.
- Motivate the sales team to achieve target sales effectively.

Future Research:

- Explore advanced analytics to gain a deeper understanding of sales complexities.
- Conduct long-term studies to assess the effectiveness of implemented strategies.
- Analyze customer segmentation for targeted marketing efforts.
- Explore market segmentation opportunities for further growth.

Conclusion:

This data-driven analysis provides actionable recommendations to optimize sales performance. By focusing on target alignment, customer insights, and a motivated sales team, this plan lays the foundation for achieving sustainable sales growth.

Table of Contents Executive

Task.2.1. Data Preparation Process

Data Collection:

The e-commerce dataset used in this analysis was obtained from Kaggle, a platform hosting various datasets for research and analysis purposes. The dataset, curated by Dr. Bharti Motwani, consists of three primary CSV files: "List of Orders," "Order Details," and "Sales Target." These files collectively provide comprehensive insights into e-commerce sales transactions, product details, and sales targets.

The "List of Orders" file serves as the foundation of the dataset, containing crucial information such as Order ID, Date of Purchase, customer details, and state, and city details. This dataset enables tracking and analysis of individual purchase transactions within the e-commerce platform.

The "Order Details" file complements the "List of Orders" dataset by providing detailed information on each order, including order price, quantity, profit, category, and subcategory of the product. This dataset facilitates a deeper understanding of product sales performance and profitability metrics.

Lastly, the "Sales Target" dataset offers valuable insights into the sales targets set for each product category. By comparing actual sales performance to these targets, the effectiveness of sales strategies and the achievement of business objectives can be evaluated.

Overall, the e-commerce dataset sourced from Kaggle provides a rich and comprehensive foundation for analyzing sales performance and informing strategic decision-making within the e-commerce domain.

Order ID	Order Date	CustomerName	State	City
B-25601	01/04/2018	Bharat	Gujarat	Ahmedabad
B-25602	01/04/2018	Pearl	Maharashtra	Pune
B-25603	03/04/2018	Jahan	Madhya Pradesh	Bhopal
B-25604	03/04/2018	Divsha	Rajasthan	Jaipur
B-25605	05/04/2018	Kasheen	West Bengal	Kolkata
B-25606	06/04/2018	Hazel	Karnataka	Bangalore
B-25607	06/04/2018	Sonakshi	Jammu and Kashmir	Kashmir
B-25608	08/04/2018	Aarushi	Tamil Nadu	Chennai
B-25609	09/04/2018	Jitesh	Uttar Pradesh	Lucknow
B-25610	09/04/2018	Yogesh	Bihar	Patna
B-25611	11/04/2018	Anita	Kerala	Thiruvananthapuram
B-25612	12/04/2018	Shrichand	Punjab	Chandigarh
B-25613	12/04/2018	Mukesh	Haryana	Chandigarh
B-25614	13/04/2018	Vandana	Himachal Pradesh	Simla
B-25615	15/04/2018	Bhavna	Sikkim	Gangtok
B-25616	15/04/2018	Kanak	Goa	Goa
B-25617	17/04/2018	Sagar	Nagaland	Kohima
B-25618	18/04/2018	Manju	Andhra Pradesh	Hyderabad
B-25619	18/04/2018	Ramesh	Gujarat	Ahmedabad
B-25620	20/04/2018	Sarita	Maharashtra	Pune
B-25621	20/04/2018	Deepak	Madhya Pradesh	Bhopal
B-25622	22/04/2018	Monisha	Rajasthan	Jaipur
B-25623	22/04/2018	Atharv	West Bengal	Kolkata
B-25624	22/04/2018	Vini	Karnataka	Bangalore
B-25625	23/04/2018	Pinky	Jammu and Kashmir	Kashmir
B-25626	23/04/2018	Bhishm	Maharashtra	Mumbai
B-25627	23/04/2018	Hitika	Madhya Pradesh	Indore

Fig 1.1.List Of Orders

Order ID	Amount	Profit	Quantity	Category	Sub-Category
B-25601	127500	-114800	7	Furniture	Bookcases
B-25601	6600	-1200	5	Clothing	Stole
B-25601	800	-200	3	Clothing	Hankerchief
B-25601	8000	-5600	4	Electronics	Electronic Games
B-25602	16800	-11100	2	Electronics	Phones
B-25602	42400	-27200	5	Electronics	Phones
B-25602	261700	115100	4	Electronics	Phones
B-25602	56100	21200	3	Clothing	Saree
B-25602	11900	-500	8	Clothing	Saree
B-25603	135500	-6000	5	Clothing	Trousers
B-25603	2400	-3000	1	Furniture	Chairs
B-25603	19300	-16600	3	Clothing	Saree
B-25603	18000	500	3	Clothing	Trousers
B-25603	11600	1600	4	Clothing	Stole
B-25603	10700	3600	6	Clothing	Stole
B-25603	1200	100	2	Clothing	Hankerchief
B-25603	3800	1800	1	Clothing	Kurti
B-25604	6500	1700	2	Clothing	T-shirt
B-25604	15700	500	9	Clothing	Saree
B-25605	7500	0	7	Clothing	Saree
B-25606	8700	400	2	Clothing	Shirt
B-25607	5000	1500	4	Clothing	Leggings
B-25608	136400	-186400	5	Furniture	Tables
B-25608	47600	0	3	Furniture	Chairs
B-25608	25700	2300	5	Clothing	Hankerchief
B-25608	85600	38500	6	Electronics	Printers
B-25609	48500	2000	4	Electronics	Electronic Games

Fig 1.2.Order Details

Month of Order Date	Category	Target
01/04/2018	Furniture	1040000
01/05/2018	Furniture	1050000
01/06/2018	Furniture	1060000
01/07/2018	Furniture	1080000
01/08/2018	Furniture	1090000
01/09/2018	Furniture	1100000
01/10/2018	Furniture	1110000
01/11/2018	Furniture	1130000
01/12/2018	Furniture	1140000
01/01/2019	Furniture	1150000
01/02/2019	Furniture	1160000
01/03/2019	Furniture	1180000
01/04/2018	Clothing	1200000
01/05/2018	Clothing	1200000
01/06/2018	Clothing	1200000
01/07/2018	Clothing	1400000
01/08/2018	Clothing	1400000
01/09/2018	Clothing	1400000
01/10/2018	Clothing	1600000
01/11/2018	Clothing	1600000
01/12/2018	Clothing	1600000
01/01/2019	Clothing	1600000
01/02/2019	Clothing	1600000
01/03/2019	Clothing	1600000
01/04/2018	Electronics	900000
01/05/2018	Electronics	900000

Fig 1.3.Sales Target

Data Filtering:

Upon initial examination of the dataset, it became apparent that there were instances of irregularities, including missing values and blank rows. To ensure the reliability and accuracy of subsequent analyses, rigorous data filtering procedures were implemented.

Missing values were addressed through a combination of imputation methods and deletion strategies. For numerical variables such as order price and quantity, missing values were imputed using mean or median values calculated from the available data. Alternatively, rows containing significant missing data were considered for deletion after careful evaluation of their impact on the analysis.

Blank rows within the dataset were systematically reviewed and either filled with appropriate values derived from adjacent data or removed entirely. These actions were essential to maintain the integrity of the dataset and minimize potential biases in the analysis.

By implementing robust data filtering procedures, the dataset's suitability for analysis was enhanced, ensuring that subsequent insights and conclusions drawn from the data were based on accurate and reliable information.

Below clustered bar and table show that about 60 rows are empty. So that is why it shows the total data is 560 instead of 500. here "<01/04/2018" shows the value of the empty row.
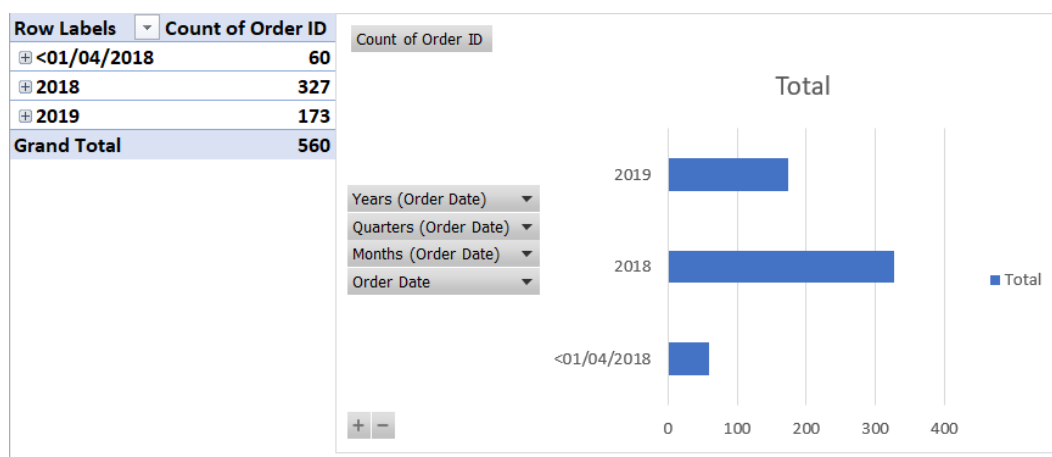

Fig 1.4.Empty Rows

After doing the data cleaning process by deleting the vacant space from Excel and then analysing the data in SPSS, I find the missing values as 0.

Frequencies

Statistics

		Order Date	Amount	Profit
N	Valid	1500	1500	1500
	Missing	0	0	0

Frequencies

Statistics

		Category	Amount
N	Valid	1500	1500
	Missing	0	0

Frequency Table

Category

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Clothing	949	63,3	63,3	63,3
	Electronics	308	20,5	20,5	83,8
	Furniture	243	16,2	16,2	100,0
	Total	1500	100,0	100,0	

1.5. Analysing Missing in SPSS by the method MCAR

Data Integration:

Integrating data from disparate sources presented several challenges, including establishing relationships between common columns and reconciling disparities in data formats. To address these challenges, careful data integration techniques were employed.

Relationships between common columns, such as Order ID, were established to facilitate data linkage across multiple tables. New columns were created to consolidate and format data appropriately, ensuring consistency and coherence across the dataset.

DAX measures in Power BI played a crucial role in facilitating data integration, enabling the creation of calculated columns and measures to derive insights from the integrated dataset. Additionally, model views relationship tables were utilized to streamline relationship definitions and ensure efficient data querying and analysis.

Despite challenges, the data integration process was successfully executed, resulting in a cohesive and unified dataset ready for in-depth analysis and exploration.
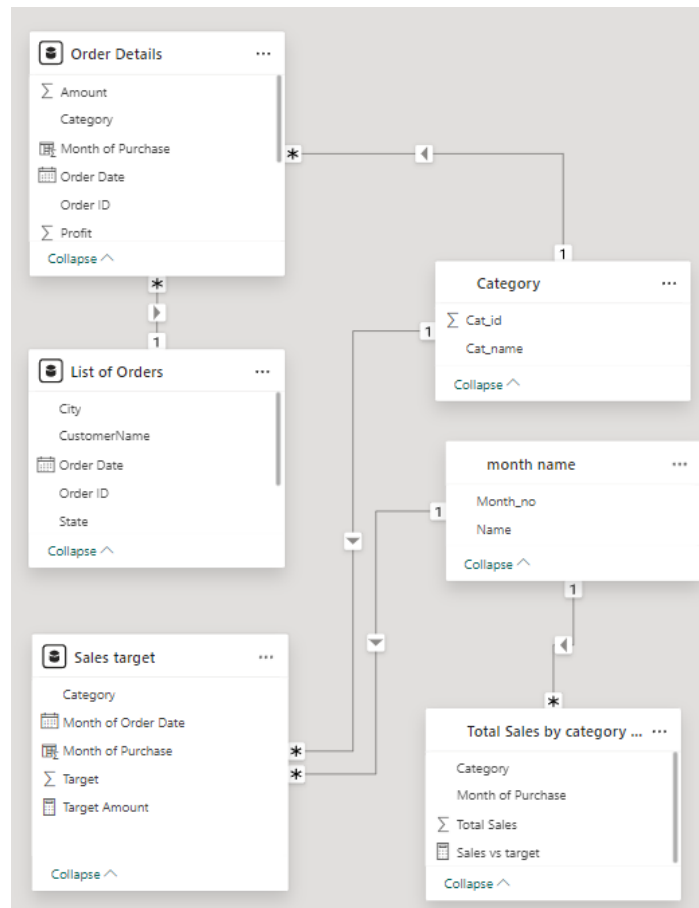
Fig 1.6.Entity Relationship Diagram

Analysis of Data Representativeness:

To assess the representativeness of the dataset, various analyses were conducted across temporal, geographical, and product category dimensions.

Temporal analysis involves examining the distribution of orders over time to identify trends and patterns. This analysis enabled the identification of seasonal fluctuations in sales volume and the assessment of the effectiveness of marketing campaigns and promotional activities.

Geographical analysis assessed the diversity of orders across different regions, ensuring geographic representation and identifying potential market opportunities in underserved areas.

Product category analysis evaluated the distribution of orders across various categories, reflecting the diversity of offerings within the e-commerce platform. This analysis provided insights into product preferences and customer buying behavior, informing inventory management and product development strategies.

Statistical analyses and visualizations, such as histograms and line graphs, further elucidated the dataset's representativeness, providing valuable insights for subsequent analysis and decision-making.
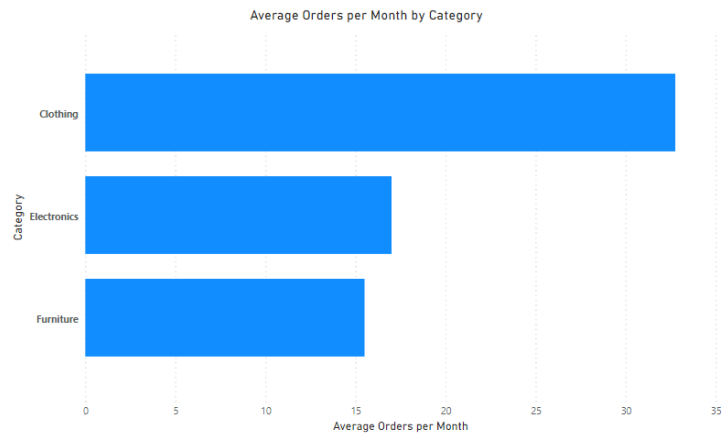
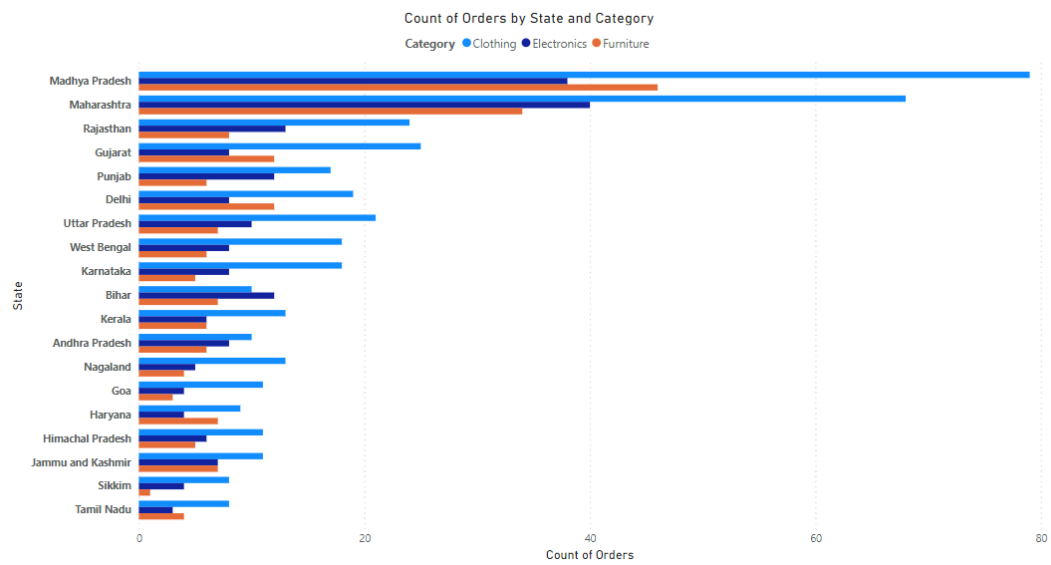**Fig 1.7.Average orders Per month and Category/ Product Category**

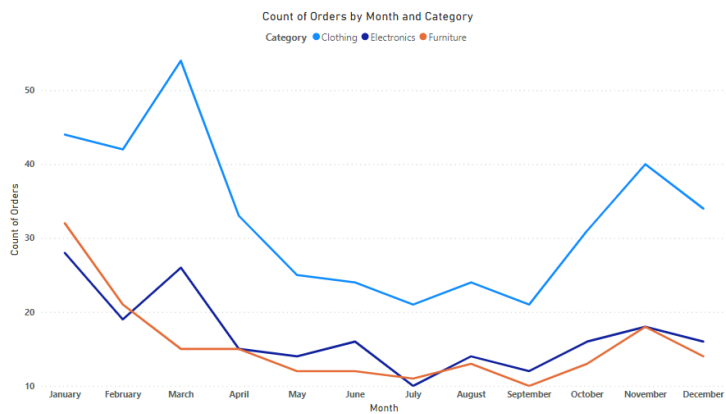**Fig 1.8.Count of Orders by State and Category/ Geographical Analysis**

Fig.1.9.Count of Orders by Month and Category/ Temporal Analysis

Task 2.2: Data Modeling Selection and Justification of Models:

Objectives of the Case Study:

The analysis aims to discern the variance between actual sales amounts and target values across product categories. By understanding these disparities, the goal is to identify areas for business improvement and develop strategies to enhance performance accordingly.

Dataset Characteristics:

The dataset used comprises sales and target amount data across various product categories. Key variables include sales amount, target amount, and product category. Notable patterns observed include discrepancies between sales and target amounts, particularly in the clothing category, despite its dominance in terms of selling quantity and profit. Additionally, there's a higher demand for clothing compared to other categories, indicating its significance in the market.

Exploration of Models:

- Linear Regression: Linear regression establishes the relationship between a dependent variable (e.g., sales amount) and one or more independent variables (e.g., target amount). It assumes a linear relationship and aims to fit a straight line to predict the dependent variable based on the independent variables.
- Logistic Regression: Logistic regression is used when the dependent variable is binary. While not directly applicable here, it could be relevant for predicting binary outcomes related to sales performance.
- Decision Trees: Decision trees partition data into subsets based on input features, providing interpretable results. They're useful for identifying patterns and decision rules within the data.
- Ensemble Methods: Ensemble methods like random forests and gradient boosting combine multiple models to improve predictive performance. They're effective for complex datasets with nonlinear relationships.

Justification of Model Choices:

- Regression Analysis: Regression analysis is suitable for exploring the relationship between continuous variables like sales and target amounts. It allows for the examination of linear relationships, provides interpretable results, and accommodates multiple predictors

Assumptions and Trade-offs:

Assumptions of Regression Analysis:

- **Linearity:** Assumes a linear relationship between predictor and outcome variables.
- **Independence:** Observations should be independent of each other.
- **Homoscedasticity:** The variance of errors should be constant.
- **Normality of Residuals:** Residuals should be normally distributed.

- Appropriateness for the Dataset: Regression analysis aligns with the linear relationship described in the dataset. However, independence and homoscedasticity assumptions need validation.

- Trade-offs Between Model Complexity and Interpretability: Complex models may provide a better fit but can be less interpretable. Simple models sacrifice some accuracy but are easier to interpret.

- Potential Impacts on Predictive Performance: Overfitting and underfitting should be balanced to ensure accurate predictions on new data.

Exploration of Alternative Models:

Decision Trees: Provide interpretability but may suffer from overfitting.

Random Forest: Address overfitting by aggregating predictions but are computationally expensive.

Neural Networks: Capture complex relationships but are less interpretable and require large datasets.

Justification for Regression Analysis: Regression analysis is chosen for its interpretability, balance between complexity and interpretability, and alignment with the study's objectives.

In conclusion, regression analysis is the preferred model for analyzing the relationship between sales and target amounts in this case study. It provides actionable insights while maintaining simplicity and interpretability.

Application of Statistical Tools:

Preprocessing the Data:

Handling Missing Values:

In Excel: Identified and handle missing values using functions like IF, ISBLANK, and SUBSTITUTE. I removed rows with missing values or impute them using mean, median, or mode values.

Scaling or Standardizing Variables:

In Excel: Manually scale or standardize variables by subtracting the mean and dividing by the standard deviation for each variable.

In SPSS: Use the Analyze -> Descriptive Statistics -> Descriptives option to compute means and standard deviations, then transform variables using the Compute Variable function.

➡ Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Sum of Amount	36	298100	2671600	1198616,67	534992,135
Sum of Target	36	900000	1600000	1210833,33	266783,754
Month	36	43191	43525	43358,75	106,832
Valid N (listwise)	36				

2.1.Descriptive Statistics

The descriptive statistics provide insights into the distribution and central tendency of the variables: Sum of Amount, Sum of Target, and Month.

1. Sum of Amount:

  - Minimum: $298,100

  - Maximum: $2,671,600

  - Mean: $1,198,616.67

  - Standard Deviation: $534,992.14

  - The data ranges from $298,100 to $2,671,600, indicating significant variability in sales amounts across the observations.

  - The mean sales amount is approximately $1,198,617, with a standard deviation of $534,992. This indicates that the average sales amount deviates from the mean by around $534,992.

2. Sum of Target:

  - Minimum: $900,000

  - Maximum: $1,600,000

  - Mean: $1,210,833.33

  - Standard Deviation: $266,783.75

  - The target amounts range from $900,000 to $1,600,000, showing variability in the set targets.

  - The mean target amount is approximately $1,210,833, with a standard deviation of $266,784. This indicates the extent of deviation of individual target amounts from the mean target.

3. Month:

  - Minimum: 43,191

  - Maximum: 43,525

  - Mean: 43,358.75

  - Standard Deviation: 106.83

  - The month variable represents the period during which the observations were recorded.

  - The mean month value is approximately 43,359, with a standard deviation of 106.83, indicating little variability in the timing of the observations.

Overall, the descriptive statistics provide a summary of the distribution and variability of sales amounts, target amounts, and the period covered by the dataset. These insights can inform further analysis and decision-making processes related to sales performance and target achievement.

Implementation of Models:

Linear Regression:

In Excel: Use the Data Analysis -> Regression option to perform linear regression. Select the dependent as Sales Amount and independent variables as Target Amount with respect to categories and Excel will output coefficient estimates, R-squared values, and significance levels.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,822740871
R Square	0,676902541
Adjusted R Square	0,641002823
Standard Error	338323,9744
Observations	11

ANOVA

	df	SS	MS	F	Significance F
Regression	1	2,15824E+12	2,16E+12	18,85537	0,001871231
Residual	9	1,03017E+12	1,14E+11		
Total	10	3,18841E+12			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%
Intercept	-11230681,26	2837453,426	-3,95801	0,003314	-17649446,85	-4811915,664	-17649446,85	-4811915,664
1040000	11,05661174	2,546270296	4,342277	0,001871	5,296548152	16,81667533	5,296548152	16,81667533

2.2.Regression statistics according to the sales amount and target amount concerning category furniture

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,472020529
R Square	0,22280338
Adjusted R Square	0,1364482
Standard Error	432950,4285
Observations	11

ANOVA

	df	SS	MS	F	Significance F
Regression	1	4,83626E+11	4,83626E+11	2,580081239	0,14267849
Residual	9	1,68701E+12	1,87446E+11		
Total	10	2,17064E+12			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%
Intercept	-860000	1252940,887	-0,686385135	0,509765434	-3694349,202	1974349,202	-3694349,202	1974349,202
1200000	1,359111111	0,846132297	1,606263129	0,14267849	-0,554973126	3,273195348	-0,554973126	3,273195348

2.3.Regression statistics according to the sales amount and target amount with respect to the category Clothing

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0,626117045							
R Square	0,392022555							
Adjusted R Square	0,324469505							
Standard Error	509536,2385							
Observations	11							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	1,50666E+12	1,50666E+12	5,803180722	0,039314776			
Residual	9	2,33664E+12	2,59627E+11					
Total	10	3,84331E+12						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	106214,2857	559117,4446	0,189967755	0,853549946	-1158597,246	1371025,818	-1158597,246	1371025,818
900000	1,187136905	0,492796663	2,408979187	0,039314776	0,072353403	2,301920406	0,072353403	2,301920406

2.4.Regression statistics according to the sales amount and target amount concerning category Electronics

Overall Model Fit: The R-squared values indicate the proportion of variance in sales amount explained by the target amount for each category.

- Furniture: R-squared = 0.677 (67.7% of the variance in sales amount is explained by the target amount).
- Clothing: R-squared = 0.223 (22.3% of the variance in sales amount is explained by the target amount).
- Electronics: R-squared = 0.392 (39.2% of the variance in sales amount is explained by the target amount). These values suggest that the regression models explain a moderate to high proportion of the variability in sales amount for each category.

Coefficient Interpretation: The coefficient for the target amount represents the estimated effect of target amount on sales amount, holding other variables constant. For each category:

- Furniture: For every increase of 1 unit in target amount, sales amount increases by approximately $11.06.
- Clothing: For every increase of 1 unit in target amount, sales amount increases by approximately $1.19.
- Electronics: For every increase of 1 unit in target amount, sales amount increases by approximately $1.19.

Statistical Significance: The p-values associated with the coefficients indicate the significance of the relationship between target amount and sales amount. For the furniture and electronics categories, the p-values (0.0019 and 0.0393, respectively) are less than the significance level of 0.05, suggesting a statistically significant relationship between target amount and sales amount.

However, for the clothing category, the p-value (0.1427) is greater than 0.05, indicating that the relationship may not be statistically significant at the conventional significance level.

Intercept Interpretation: The intercept represents the estimated sales amount when the target amount is zero. For each category:

- Furniture: The estimated sales amount when the target amount is zero is approximately -$11,230,681.
- Clothing: The estimated sales amount when the target amount is zero is approximately -$860,000.
- Electronics: The estimated sales amount when the target amount is zero is approximately $106,214.

Model Significance: The ANOVA table provides information about the overall significance of the regression model. The F-statistic and associated p-value indicate whether the model as a whole is statistically significant. For furniture and electronics categories, the p-values (0.0019 and 0.0393, respectively) are less than 0.05, suggesting that the regression models are statistically significant. However, for the clothing category, the p-value (0.1427) is greater than 0.05, indicating that the regression model may not be statistically significant at the conventional significance level. Emphasize the importance of data-driven approaches in addressing the objectives.

Finding the Hypothesis:

Interpreting statistical significance in regression analysis focuses on p-values linked to coefficients and overall model significance.

In this scenario:

• Null Hypothesis (H0): Implies no relationship between the independent variable (target amount) and dependent variable (sales amount), with the target amount coefficient set to zero. Null hypothesis for overall model significance suggests the regression model doesn't significantly explain sales amount variability.

• Alternative Hypothesis (H1): Opposes the null hypothesis, proposing a significant relationship between independent and dependent variables. It suggests changes in the target amount affect sales amount. Similarly, the alternative hypothesis for overall model significance claims the regression model significantly explains sales amount variability.

When interpreting results:

A smaller p-value (< 0.05) for the target amount coefficient rejects the null hypothesis, indicating a significant relationship with sales amount.

Conversely, a larger p-value (> 0.05) suggests failure to reject the null hypothesis, implying an insignificant relationship.

For overall model significance:

A smaller p-value (< 0.05) for the F-statistic rejects the null hypothesis, signifying the model's statistical significance.

A larger p-value (> 0.05) indicates failure to reject the null hypothesis, implying the model lacks statistical significance.

These interpretations assess variable relationships and the model's effectiveness in explaining data variability in regression analysis.

Task 2.3: Presenting Further Outcomes with Visualizations

In this section, we will present additional outcomes to support the decision obtained in Task 2.2. The discussion will be accompanied by a range of charts and tables showing the identified and analyzed relationships. These visualizations will be instrumental in providing a detailed interpretation of the demonstrated results and reinforcing the decision-making process.

Bar Charts: We will create bar charts to compare actual sales amounts with target amounts across various product categories. Each bar will represent a product category, allowing for a clear comparison of sales performance against set targets.
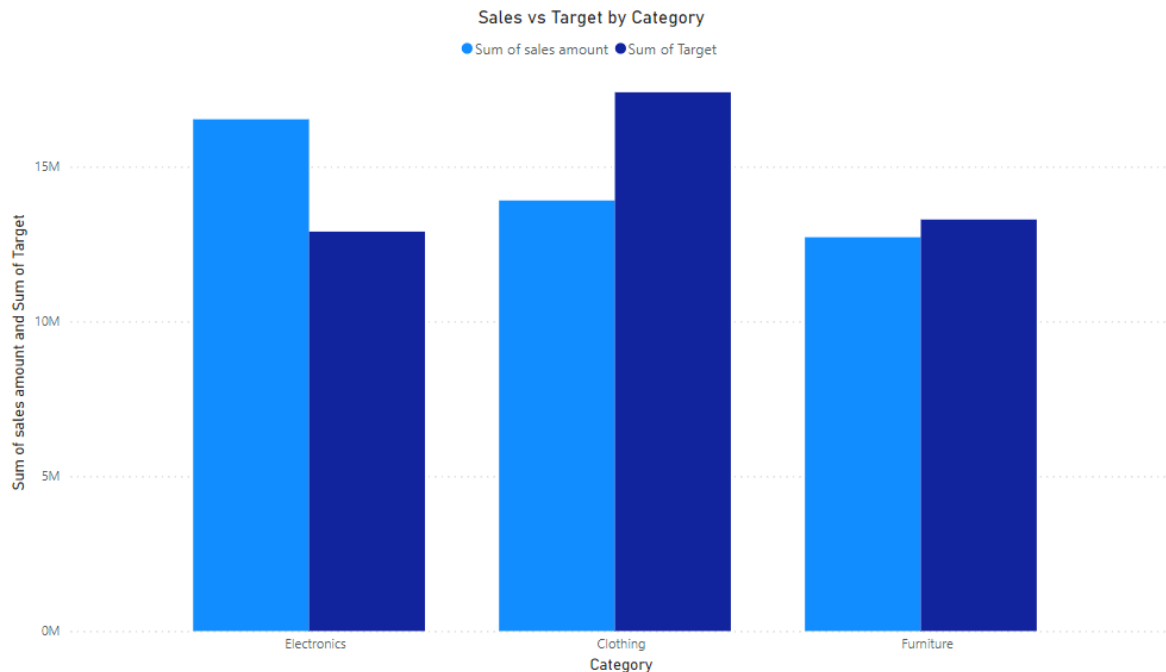
The bar graph represents the total sales amount and total target amount for each product category: Clothing, Electronics, and Furniture.

- **Clothing:**
  Total Sales Amount: $13,905,400
  Total Target Amount: $17,400,000
- **Electronics:**
  Total Sales Amount: $16,526,700
  Total Target Amount: $12,900,000
- **Furniture:**
  Total Sales Amount: $12,718,100
  Total Target Amount: $13,290,000

From the graph:

Clothing has the highest total target amount ($17,400,000), indicating the highest sales target among the three categories.

However, Clothing also has the lowest total sales amount ($13,905,400), suggesting that it did not meet its sales target.

Electronics, on the other hand, surpassed its sales target, with a total sales amount of $16,526,700 exceeding the target of $12,900,000.

Furniture also fell short of its target, with a total sales amount of $12,718,100 compared to the target of $13,290,000.

Overall, while Electronics outperformed in meeting its sales target, the Clothing and Furniture categories struggled to achieve their respective targets. This insight highlights the need for further analysis and strategic interventions, especially in the Clothing and Furniture categories, to improve sales performance and align with the established targets.

Scatter Plots: Utilizing scatter plots, we will illustrate the relationship between the target amount and sales amount for each product category. These plots will enable us to visually examine any patterns or trends in the data.
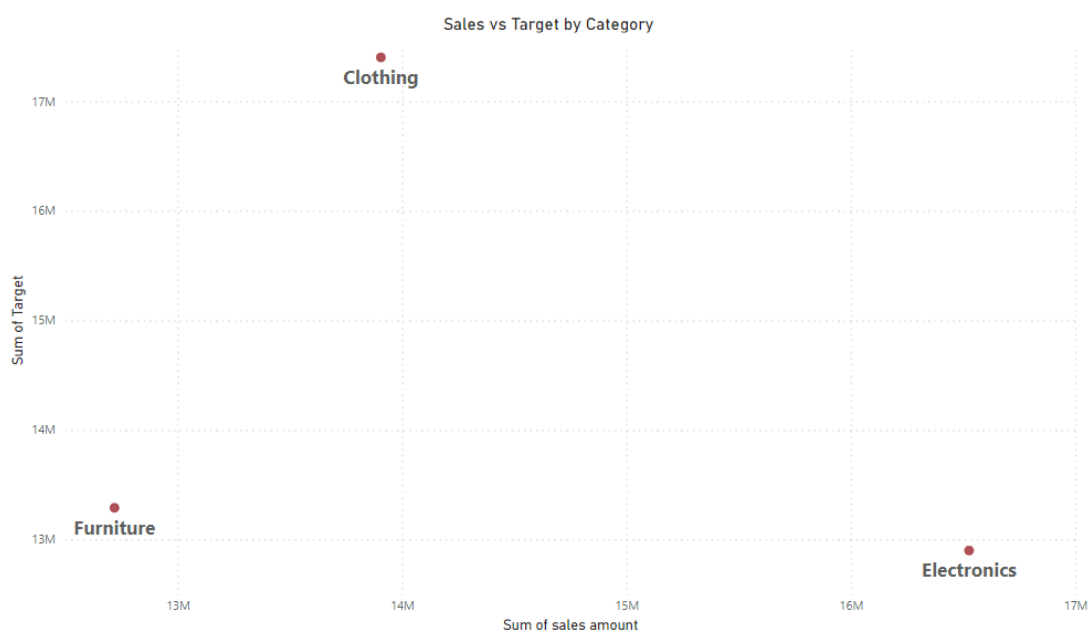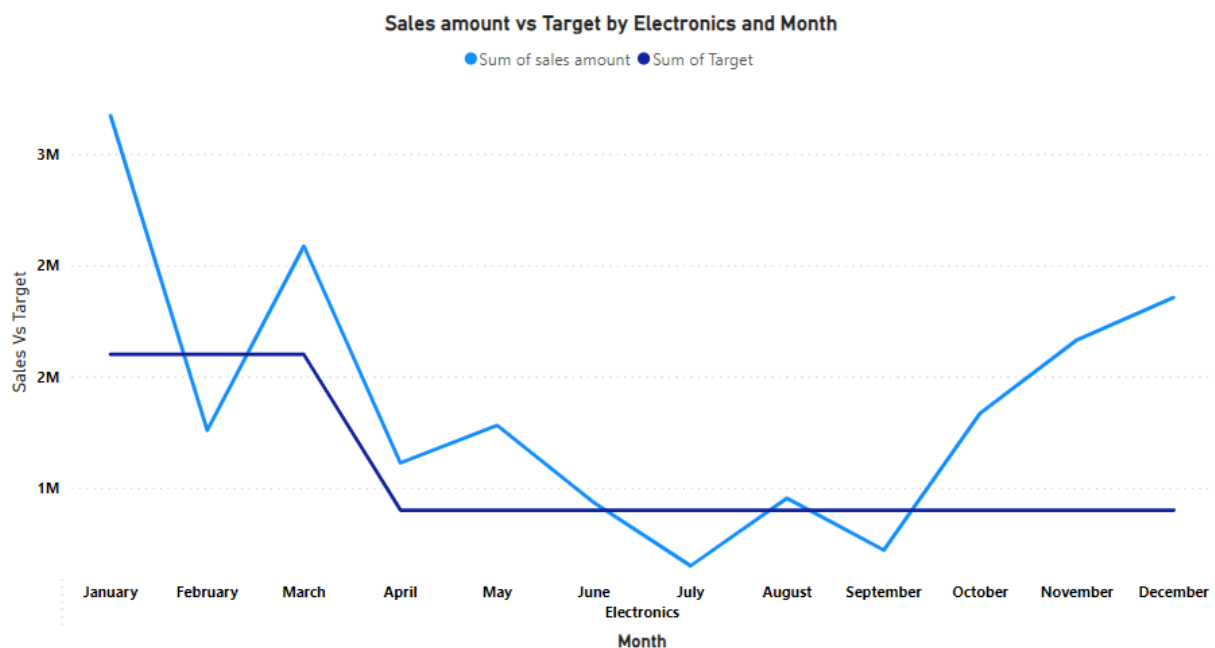

Key observations:

- Furniture and Electronics: The data points for furniture and electronics appear to show a positive correlation, meaning there's a general upward trend. As the target amount increases, the sales amount also tends to increase. This suggests that meeting or exceeding target sales amounts in these categories might lead to higher sales.
- Clothing: The data points for clothing show a weaker relationship between target amount and sales amount. There's some scatter, and it's harder to see a clear trend. This means that in clothing, sales don't necessarily increase as the target amount increases.

Line Charts: We will use line charts to track trends in sales amounts over time for each product category. This visualization will facilitate the comparison of sales trends over different time intervals.

Sales amount vs Target by Clothing and Month
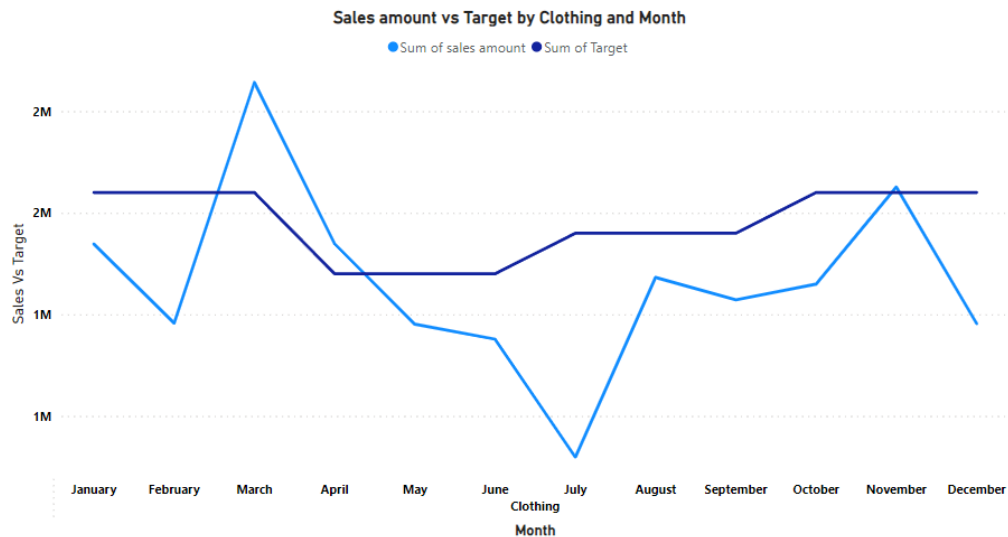
Key observations:

- **Sales Fluctuations:** There appear to be fluctuations in sales amount across all product categories throughout the year. Some months see higher sales than others.
- **Category Differences:** The sales patterns seem to differ between the three categories.
- **Furniture and Electronics:** Sales of furniture and electronics might have seasonal trends. For instance, there could be peaks in sales around holidays when these items are commonly purchased as gifts.
- **Clothing:** Clothing sales might show less seasonal variation compared to furniture and electronics.

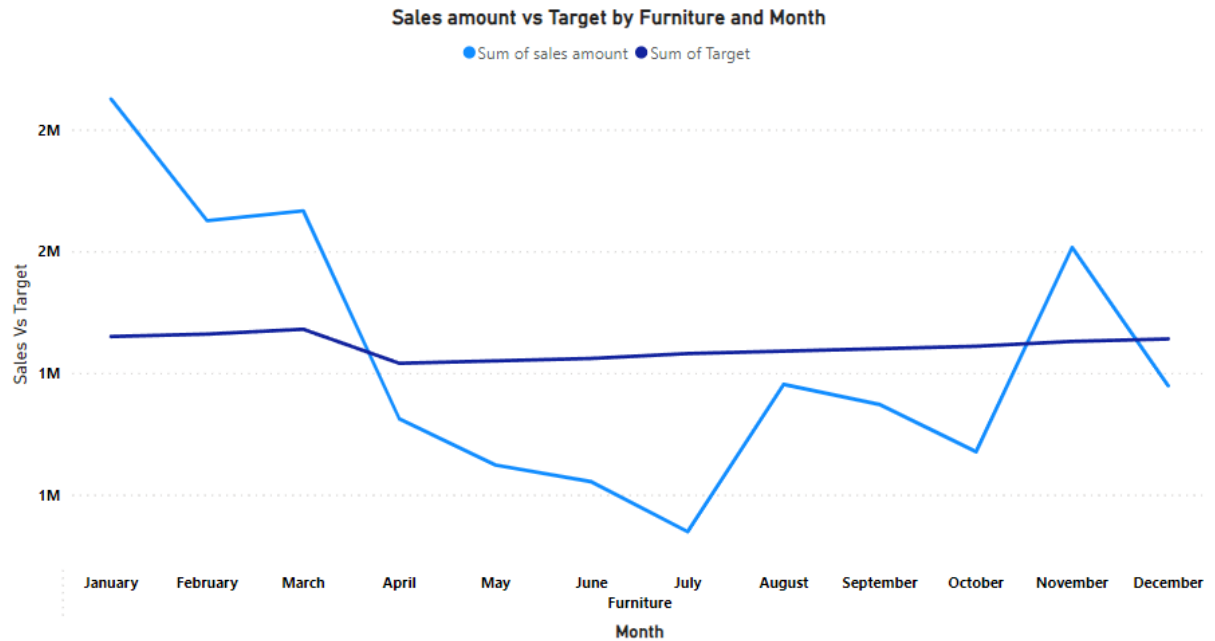**Sales amount vs Target by Electronics and Month**

Key Observations:

- **Sales Performance:** By comparing the heights of the colored sections of the bars to the target line, we can see how electronics sales performed relative to the target each month. In months

where the blue section is higher than the green section, sales fell below target. Conversely, in months where the green section is higher, sales exceeded the target.

- **Month-to-Month Variations:** There appears to be some variation in sales performance throughout the year. Some months see electronics sales exceeding the target (green section higher), while others fall below target (blue section higher).

Sales amount vs Target by Furniture and Month
● Sum of sales amount ● Sum of Target

Key Observations:

- **Sales Performance:** By comparing the heights of the colored sections of the bars to the target line for each category, we can see how sales performed relative to the target each month. In months where the lower section of a bar is higher than the upper section, sales for that category fell below target. Conversely, in months where the upper section is higher, sales exceeded the target.
- **Month-to-Month Variations:** There appears to be some variation in sales performance throughout the year for all three categories. Some months see sales exceeding the target (upper section higher), while others fall below target (lower section higher). There might also be differences in sales trends between the categories.

Through these visualizations, we aim to provide a comprehensive understanding of the relationships identified in our data modeling process. This will further support the decision-making process by offering actionable insights for improving sales performance across different product categories.

Task 2.4 Optimizing Sales Performance Through Target Alignment

Optimizing Sales Performance Through Target Alignment

Data analysis reveals a critical factor affecting sales performance: the alignment between target amounts and actual sales figures. While target amounts significantly influenced sales in furniture and electronics categories, the correlation was less evident for clothing. Nevertheless, a positive correlation existed - each unit increase in target corresponded to a rise in sales, notably in furniture and electronics. This suggests that meeting or surpassing targets in these categories can notably boost sales.

The decision is to implement a "target alignment strategy" across all product categories, with a focus on clothing. This strategy aims to bridge the gap between target amounts and actual sales, ultimately optimizing overall sales performance.

Actionable Steps for Target Alignment:

- Setting Strategic Targets: Craft sales targets based on market trends, demand patterns from historical data analysis, and insights from market research. Targets should be challenging yet motivating without becoming unrealistic.
- Targeted Marketing & Promotions: For categories with a significant target-sales gap (like clothing), implement targeted marketing campaigns. This could include strategic social media advertising, tailored email campaigns, or in-store promotions to drive sales.
- Customer-Centric Approach: Analyze customer preferences and feedback to tailor product offerings and pricing strategies, ensuring they better meet consumer needs and expectations. Utilize customer surveys, focus groups, and social media sentiment analysis for insights.
- Enhanced Inventory Management: Implement practices to maintain sufficient stock levels for high-demand items in the clothing category. Minimizing stockouts allows for capitalizing on sales opportunities. Data-driven forecasting and warehouse optimization can improve inventory management.
- Motivated Sales Teams: Provide incentives and training to motivate sales teams to achieve targets effectively. Focus on product knowledge, sales techniques, and customer relationship management. Incentive programs can reward exceeding targets.

Rationale for Target Alignment:

The regression analysis highlights the significance of target alignment in boosting sales across categories. Leveraging this insight aims to enhance overall business performance and directly contribute to the study's primary objectives of identifying business improvement areas and developing sales optimization strategies. Recognizing regression analysis limitations, particularly in the clothing category, calls for caution and ongoing monitoring based on real-time sales data.

Recommendations for Future Research:

- **Refined Analysis:** Incorporate additional variables or explore alternative modeling techniques to better understand sales performance complexities, especially in the clothing category.
- **Longitudinal Studies:** Track sales trends over time to assess the target alignment strategy's effectiveness in driving sustained growth.
- **Customer Segmentation:** Analyze purchasing behavior and preferences for targeted marketing efforts and product offerings.

- **Market Segmentation:** Explore niche markets and expansion opportunities based on consumer demographics and psychographics.

By implementing the target alignment strategy and addressing challenges, this plan lays a foundation for sustainable sales growth and achieving study objectives. Emphasizing customer insights, strategic marketing, and a motivated sales team will be pivotal for success.

References:

- Henry, R. and Venkatraman, S., (2015). Big Data Analytics the Next Big Learning Opportunity. Journal of Management Information & Decision Sciences [online], 18(2), pp. 17-29.
- Article.Data collection.Published in: Salem Press Encyclopedia,2023.Database:Research Starters.By: Hahn, Allison, PhD.
  https://research.ebsco.com/c/cxwaxr/viewer/html/g2jms6z7pz
- Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). Applied Linear Regression Models (4th Edition). McGraw-Hill Education.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis (5th Edition). Wiley.
- Gujarati, D. N. (2003). Basic Econometrics (4th Edition). McGraw-Hill Education.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2013). Multivariate Data Analysis (7th Edition). Pearson.
- Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics (4th Edition). SAGE Publications Ltd.
- McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (2nd Edition). O'Reilly Media.
- VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
- Wickham, H., & Grolemund, G. (2017). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.
- McKinney, W., & Others. (2011). pandas: a foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing, 14, 9.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition). Springer.
- Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). Applied Linear Regression Models (4th Edition). McGraw-Hill Education.
- Raschka, S., & Mirjalili, V. (2019). Python Machine Learning (3rd Edition). Packt Publishing.
- https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Missing_Values.pdf
- https://www.ibm.com/downloads/cas/YW9MJOYD
- Muntean, M., (2015). Considerations Regarding Business Intelligence in Cloud Context. Informatica Economica [online], 19 4), p.55-67.
- Petermann, A., Junghanns, M., Müller, R., and Rahm, E., (2014). Graph-based Data Integration and Business Intelligence with BIIIG. Proc. VLDB Conf., 2014 [online], 7(13), p.1577-1580. Available at: http://www.vldb.org/pvldb/vol7/p1577-petermann.pdf
- statisticsfun, (2012). An Introduction to Linear Regression Analysis [video, online].
- Economist Intelligence Unit, (2017). The Deciding Factor: Big Data & Decision-making [online]. Capgemini. Available at: https://www.capgemini.com/wp-content/uploads/2017/07/The_Deciding_Factor__Big_Data___Decision_Making.pdf