

CA1 – Business Analytics

Author: Alice Corry

Student Number: N00211635

Module: Research & Analytics

Year 4 2024-25

DL836 BSc (Hons) in Creative Computing

Table of Contents

Business Problem 1 - Interpretation of Findings	4
1. Cleaning Up the Data	4
2. Top 3 Salespeople by City	4
3. Sale Amount Calculation	4
4. Total Sales by City	4
5. Top 5 Sales by Category and Bike	4
6. Sum and Average of Price by Segment and Category (Pivot Table)	5
7. Cruisers Sold in Chicago Over \$2000 in 2024	5
8. Salespeople Selling Hybrid Bikes to 50-60-Year-Olds in 2023	5
9. Commission Calculation	5
10. Top 5 Salespeople by Total Commission	5
11. Sales Visualisation	6
Conclusion	6
Business Problem 2 - Interpretation of Findings	7
1. Cleaning the Data	7
2. Reviewing Data Types	7
3. Identifying Relevant Data	7
4. Correlation Analysis	7
5. Regression Analysis: Monthly Purchases vs. CLV	7
6. Regression Analysis: Monthly Purchases vs. Ad Spend	8
7. Building a Predictive Model for CLV	8
8. Predicting CLV for New Customers	8
9. Visualising Actual vs. Predicted CLV	8
Conclusion	8
Business Problem 3 - Interpretation of Findings	10
1. Clean the Data	10
2. Review Data Types and Focus on Relevant Columns	10
3. Find Users Who Liked the Same Items	10
4. Calculate How Much All Users Liked Each Product	10
5. Compare Similar Users and All Users' Ratings	11

6. Create a Recommendation Score (Affinity Score)	11
7. Merge Data with Product Catalogue	11
8. Create a Ranked Item List for Recommendations	11
9. Create a List Showing the Best Recommended Item for Each Product	11
10. Visualise the Data	12
Conclusion	12
References	13

Business Problem 1 - Interpretation of Findings

Summary: Alexa Ltd. is an online and offline bike retailer looking to use its sales data to make better business decisions. This report focuses on understanding sales trends, identifying top salespeople, calculating commissions, and breaking down sales by city and category. The insights from this analysis will help Alexa improve sales strategies, reward top performers, and better allocate resources.

1. Cleaning Up the Data

The first step in the analysis was to handle any missing data. Some sales records had missing customer information, price, or quantity values. We decided to replace these missing values with zeros to avoid gaps in the data. This ensures that every sale is included in the analysis, giving a more complete picture of the business.

Business Impact:

- Clean data is essential for making accurate decisions. By filling in missing values, Alexa can rely on the analysis to identify trends and opportunities without bias from incomplete records.

2. Top 3 Salespeople by City

We identified the top 3 salespeople in each city based on their total sales. These individuals contributed the most to sales in their locations. Understanding who these salespeople are can help Alexa recognise high performers and learn from their success.

Business Impact:

- Knowing the top performers in each city helps Alexa reward successful salespeople, understand what strategies are working, and replicate their methods across other teams to boost overall sales.

3. Sale Amount Calculation

We created a new column called **Sale Amount**, which is the result of multiplying the price of a bike by the quantity sold. This shows the total value of each sale and is an important metric for understanding the business's revenue.

Business Impact:

- Knowing the total value of each sale helps Alexa track revenue more accurately, allowing the company to see which products are making the most money and adjust marketing strategies accordingly.

4. Total Sales by City

We calculated the total **sale amount** for each city, ordered from the highest to the lowest. This shows which cities are generating the most revenue for Amazon.

Business Impact:

- Understanding which cities are performing best allows Alexa to focus its resources and marketing efforts on high-revenue areas. It also helps identify underperforming cities where additional support might be needed.

5. Top 5 Sales by Category and Bike

We grouped the sales by bike category and model to find out which bikes are the most popular in terms of sales volume. This helps Alexa understand customer preferences and which products are driving the most sales.

Business Impact:

- Identifying the top-selling bikes helps Alexa make better inventory and marketing decisions, ensuring that popular bikes are well-stocked and promoted, while less popular models can be re-evaluated.

6. Sum and Average of Price by Segment and Category (Pivot Table)

Using a pivot table, we analysed the sum and average of bike prices by sales channel (online vs. store) and bike category (e.g., electric, hybrid, off-road). This gives a clearer picture of how different types of bikes are priced across different sales channels.

Business Impact:

- This analysis helps Alexa determine pricing strategies and identify where certain categories of bikes are more profitable. It also supports decisions about pricing adjustments across different sales channels.

7. Cruisers Sold in Chicago Over \$2000 in 2024

We filtered the data to find all cruisers sold in Chicago in 2024 that were priced above \$2000 and sold by a specific salesperson (ID 782). This helps Alexa focus on high-value sales and understand trends in high-ticket items.

Business Impact:

- By focusing on high-value sales, Alexa can identify opportunities to promote premium bikes, tailor marketing campaigns for high-end products, and reward salespeople who are successful in this segment.

8. Salespeople Selling Hybrid Bikes to 50-60 Year Olds in 2023

We identified which salespeople sold hybrid bikes to customers aged 50-60 with a discount in 2023. This helps Alexa understand which salespeople perform well in specific customer segments and which discounts work best.

Business Impact:

- This analysis can guide targeted marketing and sales efforts towards specific customer demographics, such as older customers who may be more interested in hybrid bikes. It also helps in evaluating the effectiveness of discount strategies.

9. Commission Calculation

We calculated the commission for each salesperson based on the quantity of bikes sold. A 5% commission rate was used to determine the amount each salesperson earns from their sales.

Business Impact:

- Commission data is essential for motivating and rewarding salespeople. Alexa can optimise compensation strategies by understanding how commissions are earned, and aligning them with business goals.

10. Top 5 Salespeople by Total Commission

We listed the top 5 salespeople based on their total commission. This provides insight into which individuals are generating the most revenue for Alexa.

Business Impact:

- Knowing the top earners in terms of commission helps Alexa reward high performers and identify successful sales strategies. It also enables management to make better sales team performance and resource allocation decisions.

11. Sales Visualisation

We created a bar chart showing total sales by city and bike category. This visual representation makes it easier for Alexa's management to see how different cities and bike categories are performing, highlighting areas for improvement or further investment.

Business Impact:

- Visualisations simplify complex data, helping Alexa's management quickly understand key sales trends. This supports strategic decision-making, from inventory management to targeted marketing.

Conclusion

The insights gained from this analysis provide Alexa Ltd. with a comprehensive view of its sales performance, identifying top-performing salespeople, high-value products, and profitable cities. By leveraging these insights, Alexa can optimise its sales strategies, improve commission structures, and make data-driven decisions to drive growth in both online and offline channels.

Business Problem 2 - Interpretation of Findings

Summary: Alexa Ltd. aims to predict the **Customer Lifetime Value (CLV)** for new customers using their monthly purchase data. By understanding how customers' spending over time relates to their total lifetime value, Alexa can improve marketing efforts, predict future revenue, and enhance customer retention. This report outlines the steps taken to clean the data, explore key relationships, build a predictive model, and visualise the results.

1. Cleaning the Data

The first step was to clean the dataset by removing any rows with missing values. After cleaning, the dataset contains **100 records** and **13 variables**, including monthly purchases for 10 months, CLV (customer lifetime value), and ad spend.

Business Impact:

- Cleaning the data ensures the predictions are based on accurate and complete information. Missing data can lead to unreliable results, so removing or handling it appropriately is important.

2. Reviewing Data Types

We checked each column's data types to ensure they were suitable for analysis. The columns for monthly purchases, CLV, and ad spend were all numeric (float64), while **CUST_ID** was an integer, but it wasn't needed for the analysis. We dropped it.

Business Impact:

- Using the correct data types is essential for running the model smoothly. Removing unnecessary columns like **CUST_ID** helps keep the analysis focused on the important data.

3. Identifying Relevant Data

We removed the **CUST_ID** column as it wasn't needed to predict CLV. The remaining columns include **MONTH_1** to **MONTH_10** (monthly purchases), **CLV** (customer lifetime value), and **Ad Spend** (advertising spend).

Business Impact:

- By focusing on the relevant columns, we ensure that the analysis is efficient and accurate, leading to better predictions of customer value.

4. Correlation Analysis

We analysed how monthly purchases relate to **CLV** and **ad spend**. The results showed that **MTH_7**, **MTH_9**, and **MTH_10** had the strongest correlations with CLV. Ad Spend also showed a positive correlation with both CLV and monthly purchases.

Business Impact:

- Identifying these key relationships helps Alexa understand which months are most important for driving CLV. It also shows that increasing ad spend can boost customer value, helping Alexa allocate its marketing budget more effectively.

5. Regression Analysis: Monthly Purchases vs. CLV

We created regression plots to see how **monthly purchases** relate to **CLV**. The plots show that higher spending in certain months leads to a higher lifetime value.

Business Impact:

- By understanding how customer spending impacts CLV, Alexa can prioritise high-value customers and create targeted strategies to increase their lifetime value.

6. Regression Analysis: Monthly Purchases vs. Ad Spend

We also examined how **monthly purchases** are related to **ad spend**. The analysis suggests that higher ad spend is linked to higher customer spending.

Business Impact:

- This insight allows Alexa to fine-tune its ad budget. By focusing on high-purchase customers, Alexa can increase sales and maximise return on its advertising investment.

7. Building a Predictive Model for CLV

We built a **linear regression** model to predict CLV based on the customer's monthly purchases. We split the data into **training** and **testing** sets, and the model showed an **R-squared** score of **0.74** for the training data and **0.62** for the testing data.

Business Impact:

- The model's good R-squared value means it can reasonably predict CLV. Alexa can use this model to identify high-potential customers and plan marketing strategies accordingly.

8. Predicting CLV for New Customers

We used the model to predict the CLV of a new customer based on their monthly purchases. The model predicted a CLV of **4938.64** for this new customer.

Business Impact:

- By predicting CLV for new customers, Alexa can forecast future revenue. The company can target high-value customers early to maximise retention and sales.

9. Visualising Actual vs. Predicted CLV

We compared **actual** vs. **predicted CLV** by creating a scatter plot. This helps evaluate the model's performance and identify any discrepancies.

Business Impact:

- This visualisation helps Alexa see how accurate the model's predictions are. If significant differences exist, Alexa can adjust its strategy to improve prediction accuracy and meet customer needs.

Conclusion

By analysing the relationship between **monthly purchases** and **CLV**, Alexa Ltd. can predict **customer lifetime value** more accurately. The model helps Alexa identify high-value customers, tailor marketing efforts, and predict future revenue. These insights can improve customer retention and drive business growth.

Next Steps:

- Alexa should integrate this model into its customer relationship management (CRM) system to predict CLV for all new customers.
- The model should be updated regularly with new data to improve accuracy.

- Alexa can also use **customer segmentation** to target high-value customers with personalised marketing, improving retention and sales

Business Problem 3 - Interpretation of Findings

Summary: Alexa wants to recommend more "tech products" to customers based on their past ratings and orders. The goal is to analyse the `bike_reviews.csv` dataset and use the findings to improve Alexa's sales by recommending the best products to users.

1. Clean the Data

The first step is to remove any rows with missing values to ensure the data is complete and accurate.

Business Impact:

- **Reliable Insights:** Clean data leads to more accurate recommendations, helping Alexa suggest the right products.
- **Improved Efficiency:** Removing incomplete records makes the analysis faster and more focused.

2. Review Data Types and Focus on Relevant Columns

We need to check the data types and keep only the important columns for the analysis. For example, we can remove unnecessary details like product model names that are unnecessary for this recommendation task.

Business Impact:

- **Better Data Management:** Focusing on key columns helps avoid confusion and simplifies analysis.
- **Streamlined Recommendations:** By narrowing down the relevant data, Alexa can focus on the most helpful customer preferences.

3. Find Users Who Liked the Same Items

Identify users who gave high ratings (above 3) to the same products. This helps find "similar users" who liked the same items.

Business Impact:

- **Customer Segmentation:** Understanding which users like the same products allows Alexa to target similar customers with tailored recommendations.
- **Personalised Recommendations:** The more we know about what similar users like, the better Alexa can suggest products that are highly likely to be appreciated.

4. Calculate How Much All Users Liked Each Product

Calculate the percentage of users who rated each product positively (above 3). This gives a clear picture of overall customer satisfaction with each product.

Business Impact:

- **Product Popularity:** This helps identify which products are generally liked by customers, giving Alexa a ranking of popular items to recommend.
- **Customer Satisfaction:** Focusing on products with high ratings ensures that Alexa promotes items that customers are likely to enjoy.

5. Compare Similar Users and All Users' Ratings

Combine the information from similar users and overall user ratings to get a clearer picture of product preference. This comparison helps Alexa identify the most highly recommended products from both perspectives.

Business Impact:

- **Cross-Validation:** By looking at both individual and general customer preferences, Alexa can refine recommendations to be more precise and reliable.
- **Targeted Marketing:** This helps Alexa recommend products with a proven track record of being liked, increasing the likelihood of sales.

6. Create a Recommendation Score (Affinity Score)

Calculate an "affinity score" for each product, which combines the number of similar users with how much those users like the product. This score helps rank products by how much similar users like them.

Business Impact:

- **Effective Prioritisation:** By calculating a score, Alexa can prioritise products with the most substantial potential for recommendation, driving better sales.
- **Better Insights:** The affinity score highlights which products have the most loyal customer base, making them ideal for promotion.

7. Merge Data with the Product Catalogue

Merge the affinity score with product details from the `bike_catalogue.csv` to fully view each product's characteristics, such as price, category, and availability.

Business Impact:

- **Holistic View:** Combining product information with customer preferences helps Alexa recommend products that fit customer needs and stock availability.
- **Comprehensive Recommendations:** Having all product details ensures that Alexa can suggest products with the correct pricing and stock levels, increasing customer satisfaction.

8. Create a Ranked Item List for Recommendations

Sort the products by their affinity score, showing customers the items most likely to be of interest based on their preferences and similar users' likes.

Business Impact:

- **Optimised Recommendations:** By sorting products based on their affinity score, Alexa can ensure users see the products most likely to appeal to them first, improving conversion rates.
- **Targeted Suggestions:** Customers are more likely to purchase when they see products that match their preferences and those of others like them.

9. Create a List Showing the Best Recommended Item for Each Product

For each product, show only the highest recommended item, displaying its name and price. This allows customers to find the top recommended items in each category quickly.

Business Impact:

- **Simplicity:** A focused list of the top recommended products makes it easier for customers to decide, leading to faster purchasing decisions.
- **Clear Value:** By showing the best items, Alexa highlights products most likely to meet customer needs, enhancing their shopping experience.

10. Visualise the Data

Plot graphs to visualise the affinity scores or the distribution of ratings. This can help Alexa's management better understand customer preferences and identify areas for improvement in product offerings.

Business Impact:

- **Better Decision-Making:** Visualisations can provide clear insights into trends and patterns, helping Alexa's management make informed decisions about which products to push for recommendation.
- **Data-Driven Strategy:** Graphs and charts allow for a clearer understanding of the data, making it easier to identify successful products and optimise the recommendation engine.

Conclusion

- **Increased Sales:** Alexa can increase conversion rates and boost sales by recommending products that are most likely to be liked based on customer preferences.
- **Personalised Experience:** Recommendations create a more personalised shopping experience, increasing customer satisfaction and loyalty.
- **Improved Customer Retention:** When customers find products they like quickly, they are more likely to return, ensuring long-term customer relationships.
- **Optimised Marketing:** With clear insights into customer preferences, Alexa can market products more effectively, targeting users with the items they are most likely to buy.

References

The following references offer key concepts on handling missing data, linear regression, recommender systems, and statistical analysis, which support the methods used in this work.

- Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). Wiley.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.
- Aggarwal, C. C. (2016). *Recommender systems: The textbook*. Springer.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage.