



HULT
INTERNATIONAL
BUSINESS SCHOOL

A3: Retail Product Sales Data Analysis and Reporting

943 Words

Aleksas Slavinskas

Visualizing & Analyzing Data with R: Methods & Tools - DAT-5323 - LMBAN1

Professor: Dr. Martin Boeckle

March 16th, 2025

Video link: <https://youtu.be/1Dp6rWldcno>

Contents

Introduction.....	3
Data Loading and Exploration	3
Data cleaning, handling missing values, transformation and feature engineering	3
Grouping and Aggregation.....	3
Data Visualization	4
Predicting Future Sales with Regression	4
Recommendations	5

Introduction

This report will aim to analyse a dataset from a global datastore to provide insight into business metrics, and to provide future recommendations.

Data Loading and Exploration

The dataset is in Excel CSV format, and is quite straight forward: it has Row Id for each unique row, going as integers from 1 to 9800, Order Id for each purchase, Order and Ship Dates (two columns, both in UK date format) as well as shipping type, customer information such as Customer Id and their full name, their address including region, (however no street name column). All of these have a non-numerical character in them, meaning that such data can be seen as “string”, meaning that calculations will not be possible, except for dates and postal code, with the latter being a 4 or 5-digit numerical code. Lastly, there are columns regarding the product bought, such as Product Id, Category, Sub-category, Product Name, and lastly the price, which can be seen as a “float” a number with fractional values.

Data Cleaning, Handling Missing Values, Transformation And Feature Engineering

After the initial exploration, the dataset was cleaned from duplicate rows, for instances where Order Date is after Ship Date, and where Sale Price is negative. Lastly, it was noticed that all orders are from USA, where the standard postal ZIP code is always 5-digit numerical (United States Postal Service, 2010), however, there were 4-digit postal codes in the dataset, which means that they are incorrect. After gaining insight that all states that did have 4-digit ZIP codes where those which start with a “0”, they were fixed. Also, it was checked if the number of states was 50 (U.S. Embassy & Consulates in the United Kingdom, 2017). Also, it should be noted that dates were converted into dd/mm/yy format, since while the format in the dataset was the same, in some cases months such as July were displayed as “06” while other times as “6”, which could have caused errors later.

Grouping and Aggregation

Once this was done, some data was transformed, or features were added. Order Month was added as a new column, and as a number, delivery times were calculated for each order, each sale was categorized into three categories (low (<100), medium (100-500), and high (>500)),

and finally, a binary column was created to indicate whether express shipping was selected. Such features will allow to improve the analysis in the future and to draw better insights, while current wants allow to easily find make insights into logistic performance, and sales values per order.

Data Visualization

Following this, sales were aggregated by product and region, as this will allow to understand which products generate the most revenue in each region. In three regions, Central, East, and West, “Canon imageCLASS 2200 Advanced Copier” was the most popular product. It should be remembered that this reflects revenues brought to the business, however, not profits.

Further, two bar charts were created, to visualize sales total sales per category and sub-category, as such insight will help to understand which product categories and sub-categories perform the best (in terms of revenues), and which might require a change in sales strategy. It was found that Technology products generate the most revenue, however, Furniture and office Supplies are not too far behind. In terms of sub-categories, Phones and Chairs (interestingly) bring the most revenue by a very big margin, while fasteners generate the least, also by quite a margin, compared to the 2nd least popular product.

Afterwards, a line-chart graph was created to display year and monthly total sales. Such insight allows to better understand seasonal demand changes, thus allowing to optimize logistics and marketing strategies, as well as, indicating whether the business is growing or not. While from the graph it can be seen that the business is indeed growing, there are major fluctuations, with demand reaching over 90’000 in Christmas period on a monthly basis, after which it falls over three times in post-Christmas period.

Predicting Future Sales with Regression

Following this, two pivot tables were created to summarize product sales by category and segment, as well as by region and month. Such tables provide insight into how much different customer segments are bring revenue, as well as how much there are purchases in each region, thus allowing to better understand seasonality and demand changes.

The last table includes aggregated total sales on a monthly basis of the company, as well as 3-month prediction, which was done via simple linear regression. While the graph, as mentioned

before, shows that the sales are highly fluctuating, the 3-month prediction is displayed as levelled-off and having a slow increase, starting at slightly above 60'000 a month. As it can be suspected, this is unlikely to be true, and the Mean Absolute Error of ~17'400, with Root Mean Squared Error of 21435, thus it would be more accurate to take this as a 3-month average, rather than an accurate prediction.

Recommendations

Lastly, a few recommendations should be mentioned for the business. First, while unclear about current Christmas season preparations, the business should focus on stocking up for the Christmas, and plan it accordingly, so that little stock is left afterwards, for the upcoming few months. This way storing prices could be decreased, while inventory turnover increased. Further, as phones are the most purchased product, however there are no accessories offered, introducing smartphone accessories such as phone cases, headphones, chargers, etc., is very likely to improve the revenues as such products are often sold together. Furthermore, if phones are a popular purchase, moving more towards consumer electronics, such as laptops, might be profitable as well. On the other hand, it should be reconsidered if working with Fasteners is still worth it.

References

U.S. Embassy & Consulates in the United Kingdom. (2017, September 1). Facts about the 50 States of the Union. U.S. Embassy & Consulates in the United Kingdom.
<https://uk.usembassy.gov/states-of-the-union-states-of-the-u-s/>

United States Postal Service. (2010). U.S. ZIP Codes: Free ZIP code map and zip code lookup. UnitedStatesZipCodes. <https://www.unitedstateszipcodes.org/>