UNCOVERING THE STRATEGIES OF BOOSTING THE SALES OF ZEDEF CORPORATION

1. Problem Definition:

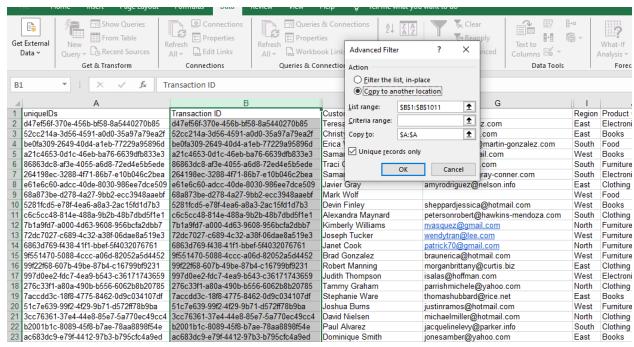
What should the management do to boost the sales of ZEDEF Corporation?

2. Data Collection:

An unstructured extensive sales data was provided. The dataset contains 1010 rows of data and 11 columns. The columns are: Transaction ID, Customer Name, Customer Email, Region, Product Category, Product, Quantity, Unit Price, Total Sales, Transaction Date, and Payment Method, Quantity, Unit Price and Total Sales are numeric columns. Transaction date is a date column while the rest of the columns are text columns.

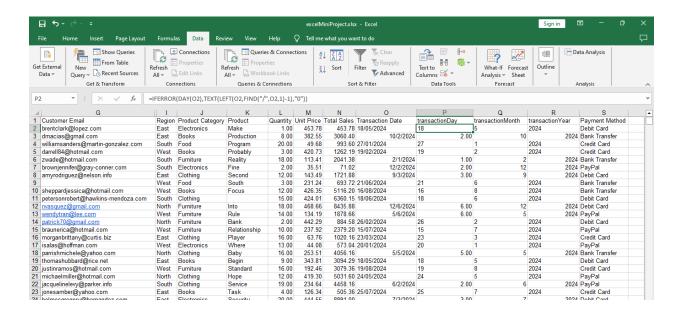
3. Data Cleaning:

- 1. Generic Use of VLOOKUP: VLOOKUP was used on all the columns to bring their values from the raw data to the new sheet where they will be cleaned. The lookup values are the unique transaction IDs that were copied from the raw data sheet.
- 2. Removal of Duplicates in the Transaction ID Column: Data > Advanced Filter was used to select the unique Transaction IDs and copy them to a newly created column called uniqueIDs in column A. The IDs in the uniqueIDs were copied to the cleanedDataSet sheet where further cleaning processes were carried out selecting only the unique transactions by using these unique IDs as lookup values.

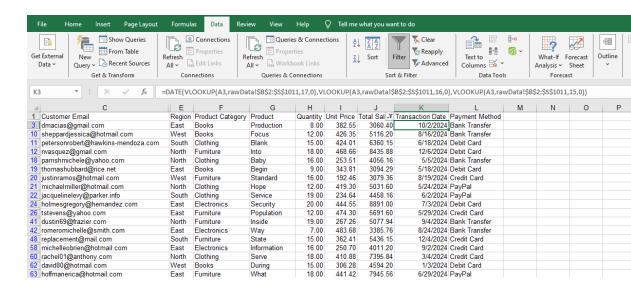


3. Conversion of Transaction Date Column to Consistent Date Data Type

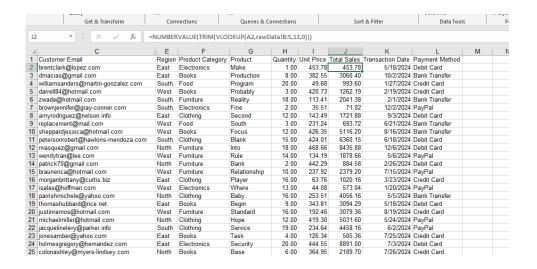
 a. Creation of three helper columns – transactionDay, transactionMonth, and transactionYear – using IFERROR, TEXT, FIND, LEFT, DAY, MONTH, AND YEAR functions



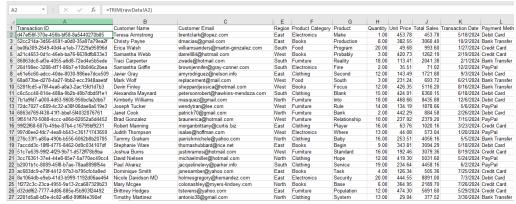
b. Generation of consistent date values based on the initial inconsistent date values using VLOOKUP and DATE functions.



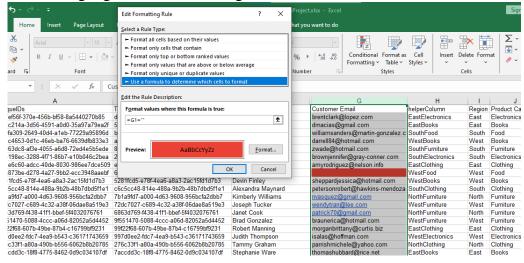
4. Removal of extra spaces in the three number columns – Quantity, Unit Price, and Total Sales - using TRIM, NUMBERVALUE, and VLOOKUP.



5. Elimination of extra spaces in the text columns using TRIM

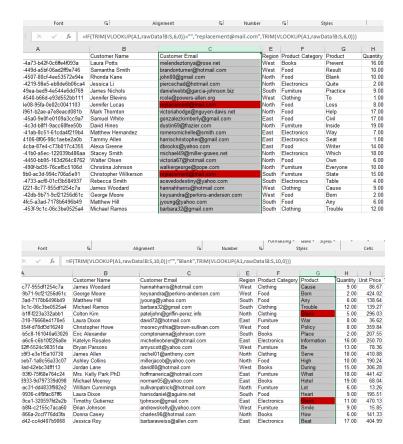


6. Identification of Missing values in the email column: Home > Conditional Formatting was used to highlight cells with missing values.



7. Replacement of Missing Values: replacement@mail.com was used to replace all missing mail values and 'Blank' was used to replace all missing product values. This decision was made because an invalid mail or product is not totally relevant to the

question of this analysis but not having an email or product value at all for a confirmed transaction reduces the completeness of the data set.

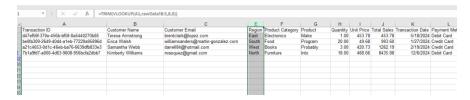


4. Data Exploration:

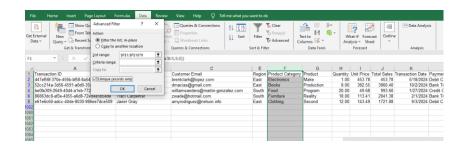
Sorting and Filtering

All sorting was done with Data > Sort and all filtering was done with Data > Filter

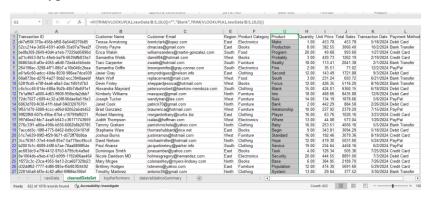
1. Regions: 4



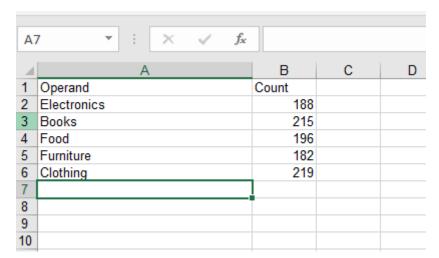
2. Product Categories: 5



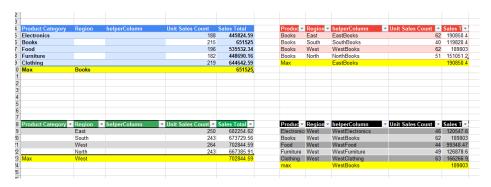
3. Product Type Count: 622



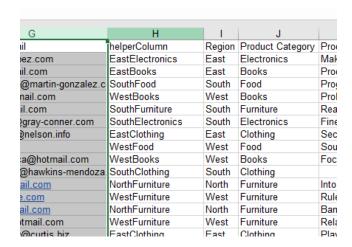
4. Product Category Count Summary



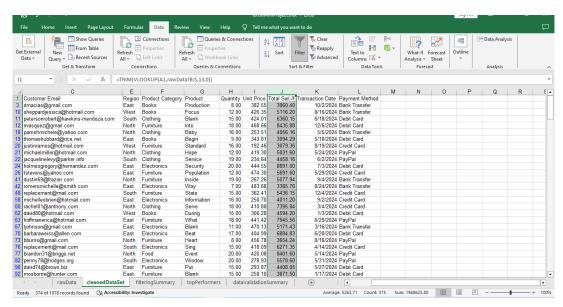
5. Summary of Top Performing Products and Regions



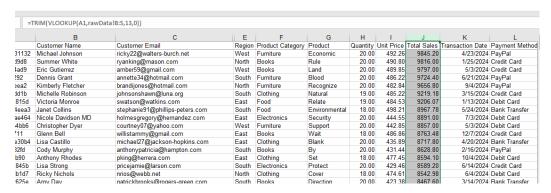
6. Use of helperColumn: In calculating the top performers, helperColumn was used to evaluate the products that have highest sales in a particular region or the region that performed best with a particular product in the Summary of the top-performing products and regions.



7. 375 transactions worth more than 3000 total sales threshold.

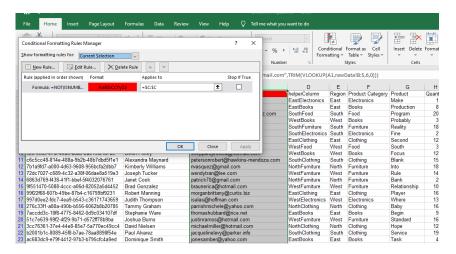


8. Sorting the data in descending order of the sales amount.

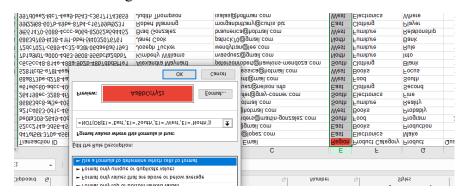


5. Data Exploration:

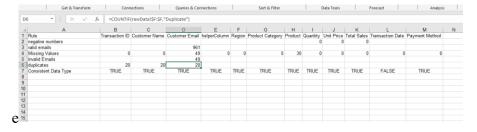
1. Identification of invalid emails



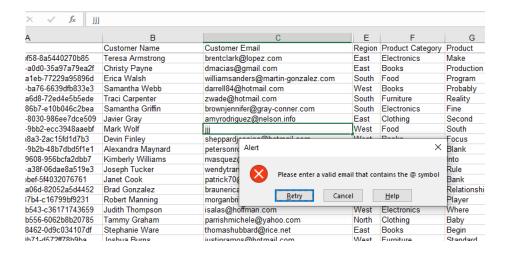
2. Detection of invalid regions



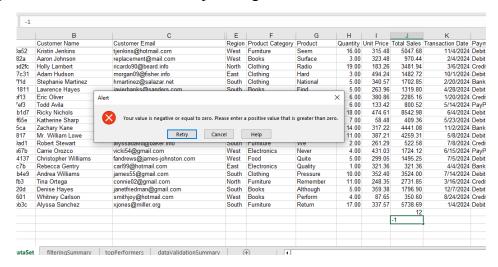
3. Summary of data impurities that require cleaning



4. Application of data validation on email column to allow only values that contain the @ symbol of an email



5. Application of data validation on the total sales column to allow only nonzero positive numbers. It therefore rejects negative numbers and zero.

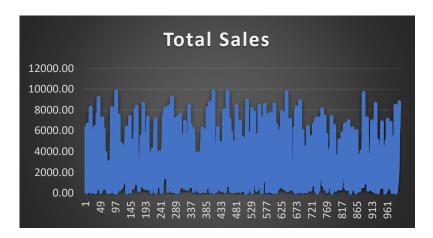


6. Data Analysis:

Descriptive Data Analysis

| 4 | A | В |
|---|-------------------------|-------------|
| | Total Sales | |
| | | |
| | Mean | 2726.21468 |
| | Standard Error | 73.02932117 |
| | Median | 2078.28 |
| | Mode | #N/A |
| | Standard Deviation | 2309.389909 |
| | Sample Variance | 5333281.751 |
| | Kurtosis | -0.01498451 |
|) | Skewness | 0.921875351 |
| 1 | Range | 9825.67 |
| 2 | Minimum | 19.53 |
| 3 | Maximum | 9845.2 |
| 1 | Sum | 2726214.68 |
| 5 | Count | 1000 |
| 3 | Largest(1) | 9845.2 |
| 7 | Smallest(1) | 19.53 |
| 3 | Confidence Level(95.0%) | 143.3084648 |

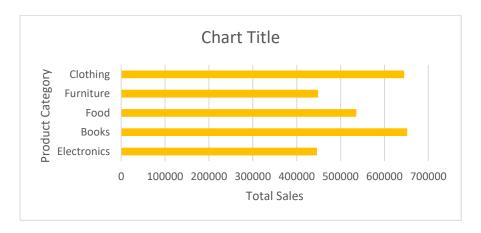
There is a total of 1000 sales captured in the dataset after cleaning. The high value of standard deviation implies that there is high variability in the data, and this can be verified by the first chart below.







The *Total Sales Per Month* column chart above shows that throughout the whole year, highest number of sales was recorded in the sixth month, which is June.



Even though clothing and Books are very close, the total Sales recorded on books is higher than the one recorded on Clothing.



This bar chart shows that the West region records the highest number of sales throughout the period under study.

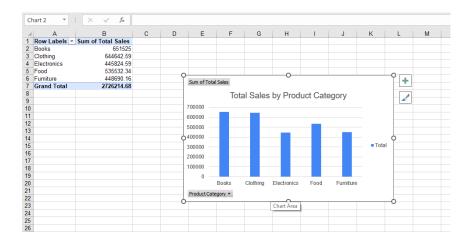


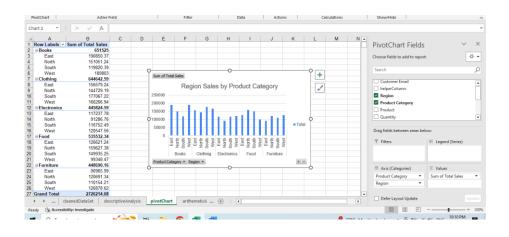
This bar chart shows that the books, which was discovered to have the highest sales throughout the period under study, has its highest sales in the East. This means that people love to read books in the East more than other regions.



The chart above shows that Books are the most sold product in the West region which recorded the highest among all regions.

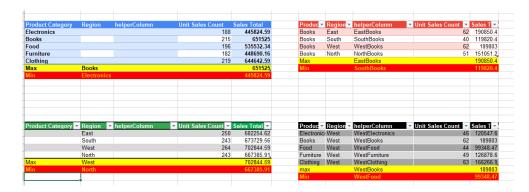
Using Pivot Charts







Underperforming Products and Regions



Electronics is the product with the least sales. For Books product that have the highest sales, they have the least sales in the South Region. This implies that people in the South do not like reading Books.

Also, the region with the least sales is the North region. And in the West region where we have the maximum sales, the product with the least sales is Food. This implies that people in the West do not like eating Food!

It is also important to note that in all categories, Clothing has the second highest record of sales and can therefore be considered alongside Books.

7. Interpretation of Results:



The above illustration shows the products and regions where the management can fund further and the regions they should start withdrawing fund from. The regions and products in the Green Area are the regions and products where the management can invest to increase their sales. While the regions and products in red are the regions where the company should not for once consider investing in them due to their bad sales.

So the following are the recommendations for the management:

- 1. Invest in Books in the East region (Recommended)
- 2. Invest in Books in the West region
- 3. Invest in Clothing generally
- 4. Invest in Books in the West region. (Recommended)
- 5. Invest in Clothing in the West region

And the investments that the management should watch out for are the following:

- 1. Do not invest in Electronics product
- 2. Do not invest in Books in the South (Recommended)
- 3. Do not invest in the North
- 4. Do not invest in Foods in the West.