# **1. OVERVIEW**

The sales table is a critical part of any business, as it records all of the transactions that have taken place within a given period of time. This table is usually used to track the revenue generated by the business, as well as to identify any trends or patterns that may be emerging in the market.

However, it's important to note that almost half of the sales table is often empty. This can be due to a variety of factors, such as a slow period for the business or a lack of sufficient data. Regardless of the reason, it's important for businesses to be aware of this issue and to work towards filling in these empty values.

In addition to the sales table, there are several other tables that are commonly used in businesses to track various types of data. For example, the returns list table is used to track all of the products that have been returned by customers. Similarly, the product reviews table is used to track customer feedback and ratings on various products.

Another table used by the firm is the Amazon Campaigns table, which is used to track the effectiveness of marketing campaigns on Amazon. Finally, the other promotions table is used to track any other promotional activities that the business may be engaged in, such as discounts or special offers.

It's worth noting that many rows in these tables may also contain null values. This can be due to a variety of reasons, such as missing data or incomplete information. As with the empty values in the sales table, it's important for businesses to be aware of these null values and to work towards filling them in as much as possible. This can help to ensure that the data in these tables is accurate and up to date, which can be critical for making informed business decisions.Formun Üstü

**Table 1:Our Methodology by feature engineering**

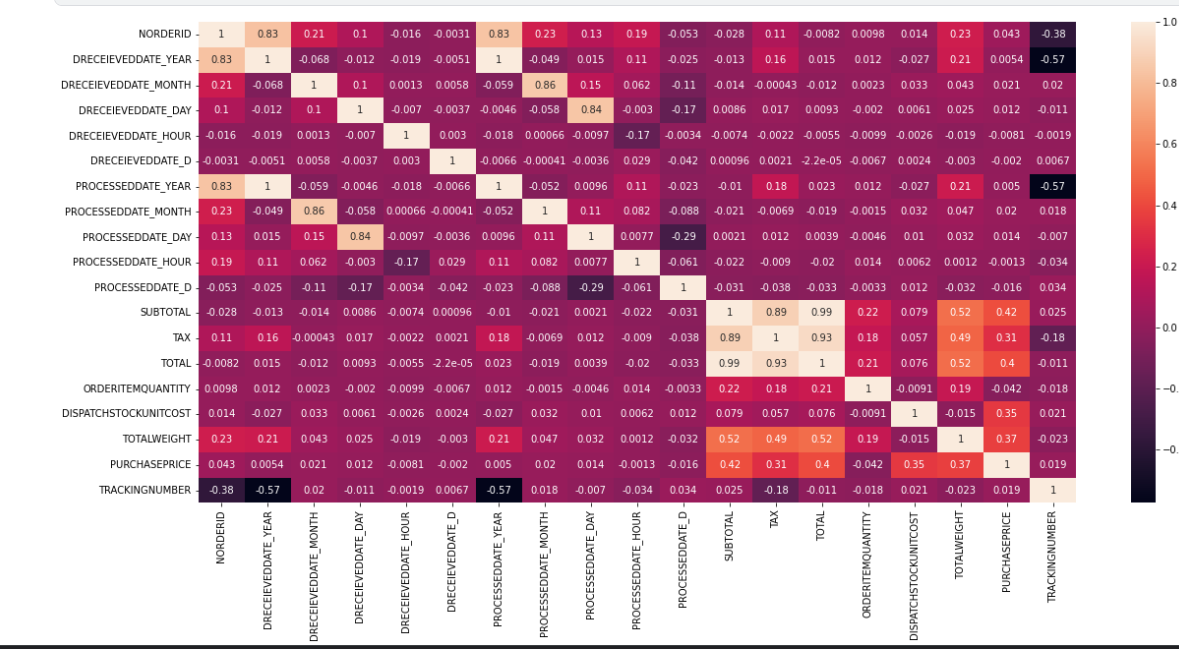
The methodology we followed during this period examining the datas row by row and at the end get an insight or pre-look about our feature analysis. At this point we marked the features from which we have taken the following actions on the table. If you do not see any table below the explanation, that means we left the table with its main existing columns.

|  |  |
| --- | --- |
| **FEATRUE (COLUMN)** | **ACTIONS TAKEN** |
|  | * Yellow filled cells mean that we dropped the related feature or column. |
|  | * Purple filled cells mean that we added the related feature as a new column. |

## **SALES TABLE**

When we analyzed the sales table, the raw data we received contained 190 thousand rows. Almost more than half of these lines consisted of blank lines. As a result of the analysis, all of these blank lines were removed from the data set. Then, data in the form of day, month and year were obtained from the date columns. And values ​​like TotalWeight,DispatchUnitCost were dropped because they include excessively blank or zero rows.

When we analyzed the sales table, the raw data we received contained 190 thousand rows. Almost more than half of these lines consisted of blank lines. As a result of the analysis, all of these blank lines were removed from the data set. Then, data in the form of day, month and year were obtained from the date columns. And values ​​like TotalWeight,DispatchUnitCost were dropped because they didn't give us an insight. After these processes, new features such as TP\_per\_Quantity,ST\_per\_Quantity,UK\_CPI,Turkey\_CPI,Net\_Profit,Day\_Time have been added as required.



**Table 2: Correlation Map of Sales Table**

|  |  |
| --- | --- |
| **FEATURE (COLUMN)** | **THE FUNCTION OF COLUMN** |
| nOrderId | * The uniqiue ID of each order taken by the company. |
| Company | * The company which shops or buys items. |
| Customer ID | * The ID number of customer |
| dReceievedDate | * The Date on which orders taken |
| Country | * Country of the customer |
| status | * Status of Orders (paid or unpaid) |
| Processed | * processing of the order if it is true the order is delivered to customer. |
| ProcessedDate | * The date of the delivery process begins |
| Source | * The source of orders taken(Amazon,Ebay etc.) |
| Currency | * The currency on that an order transacted |
| Subtotal | * The Price without tax |
| Tax | * The amount of tax a customer paid on an order |
| Total | * The Total price a customer paid on an order |
| OrderItemSKU | * Unique item number which is on sale |
| OrderItemTitle | * The ItemTitle which a customer sees on sale platform |
| ItemCategory | * Brands which items on sale |
| DispatchStockUnitCost | * Delivery Cost of an item |
| OrderItemQuantity | * Quantity of items order per nOrderID |
| TotalWeight | * The weight of dipatched goods. |
| PurchasePrice | * Purchase price of items sold |
| TrackingNumber | * Delivery tracking number of items sold |
| PostalService | * The postal service over it delivery sent |

**Table 3 Raw Features of Sales Table**

|  |  |
| --- | --- |
| **FEATURE (COLUMN)** | **ACTIONS TAKEN** |
| dReceievedDate | The column is splitted up in to the following columns.To get insight about sales date based on time. The main column also stays.  •DRECEIEVEDDATE\_YEAR •DRECEIEVEDDATE\_MONTH •DRECEIEVEDDATE\_DAY •DRECEIEVEDDATE\_HOUR •DRECEIEVEDDATE\_WEEKDAY |
| ProcessedDate | It is also one splitted up following columns.The main ProcessedDateColumn also stays. •PROCESSEDDATE\_MONTH •PROCESSEDDATE\_YEAR •PROCESSEDDATE\_DAY  •PROCESSEDDATE\_HOUR  •PROCESSEDDATE\_WEEKDAY |
| DispatchStockUnitCost | * There was lots of zero values so we decided to drop it. Also it didn’t give so much insight. |
| TotalWeight | * This column did not give us the expected research perspective. |
| TrackingNumber | * It was determined that there was a delivery tracking number, but the process of obtaining the desired information did not take place. |
| Processed | * After dropping false values then the column was dropped. |
| TP\_Per\_Quantity | * We would like to know the price that a customer paid per quantity with tax. |
| ST\_Per\_Quantity | * We would like to know the price that a customer paid per quantity with without tax. |
| UK\_CPI | * To use it on analysis as on the project instructions |
| Turkey\_CPI | * To use it on analysis as on the project instructions |
| Net\_Profit | * To know the net profit of the seller per sale (Subtotal-PurchasePrice) |
| Day\_Time | * 24 hours of a day splitted up 4 parts like 'AfterNoon','Night','Early Morning' and 'Morning' |
| status | * There was no different values as 'PAID' because of that we have dropped this. |

**Table 4: Actions Taken on Sales Table**

### **1.1.1 Recommendations for Sales Table**

When the analysis phase of the sales table was over, it was seen that most of the rows were filled with blank data. Although the exact reason for this is not known, it is thought that if it is a technical malfunction, the cause of the malfunction must be found. In addition, it has been determined that some zero values on data such as on columns DispatchUnitCost, TotalWeight recommended to be recorded completely. In addition, it is recommended to the company that the process date data with the missing date is required to be filled in correctly.

Finally, unidentified characters were seen in a few product titles in the table, and it was determined that these characters belonged to the Swedish language. In this regard, it can be eliminated with this character with a renewal to be made within the system.

## **RETURNS TABLE**

The raw form of the data in the returns table contained approximately 15 thousand rows of data, and it was seen that more than half of this data was filled with null raws, as in the Sales table, in this context, these rows were removed from the data set.

The remaining columns in this table were found to contain important data and we did not need to add any new columns.

|  |  |
| --- | --- |
| **FEATRUE (COLUMN)** | **THE FUNCTION OF COLUMN** |
| Type | * The reason of return |
| nOrderId | * The Id of Order returned |
| cPostCode | * Postcode of customer who makes item return |
| Customer ID | * The ID of customer |
| ItemNumber | * The number of item Returned |
| ItemTitle | * The title of item which customer returned |
| dReceievedDate | * The receviedate of return request |
| cCountry | * Country of the customer |
| cCountryCode | * The code of country from which rreturn made |
| cCurrency | * The currency of returned order |
| source | * The platform return made (amazon,ebay,etc) |
| subsource | * The sub platform returns made |
| Return Date | * The date on that return of item realized |
| ReturnQty | * The quantity of item returned to company by customer |
| Category | * The reason of return category |
| ResendOrExchangeQty | * Resended or Exchanged Item Quantity from the company to customer |
| RMA Actioned | * The status of the return process |
| Refund Amount | * The amount money which refunded to customer |
| Return Reason | * Explaination from a customer about the return |

**Table 5: Raw Features of Returns Table**



**Table 6: Correlation Map of Returns Table**

### **Recommendations For Returns Table**

Again, as in the previous sales table, in this returns data set, it is recommended to the company that these lines, which mostly consist of blank rows, are likely to be filled if it is possible. Otherwise, these data cells will not be used for analysis.

It was seen that the returns made in terms of the Return Date column have dates, most of which show the return date of products recalled from customers. However, the same is not valid for orders shown as Refund, therefore it was thought that for the company it is proper to create a separate RefundDate column for Refund orders and add it to the sales system.

The Return Reason and Category columns are considered to be the reason for similar customer returns. But these columns contain half empty values. In order for this data to be evaluated better, the customer's reasons for return is recommended to be recorded fully and correctly. For this, necessary arrangements must be made on the platform where the sale is made.

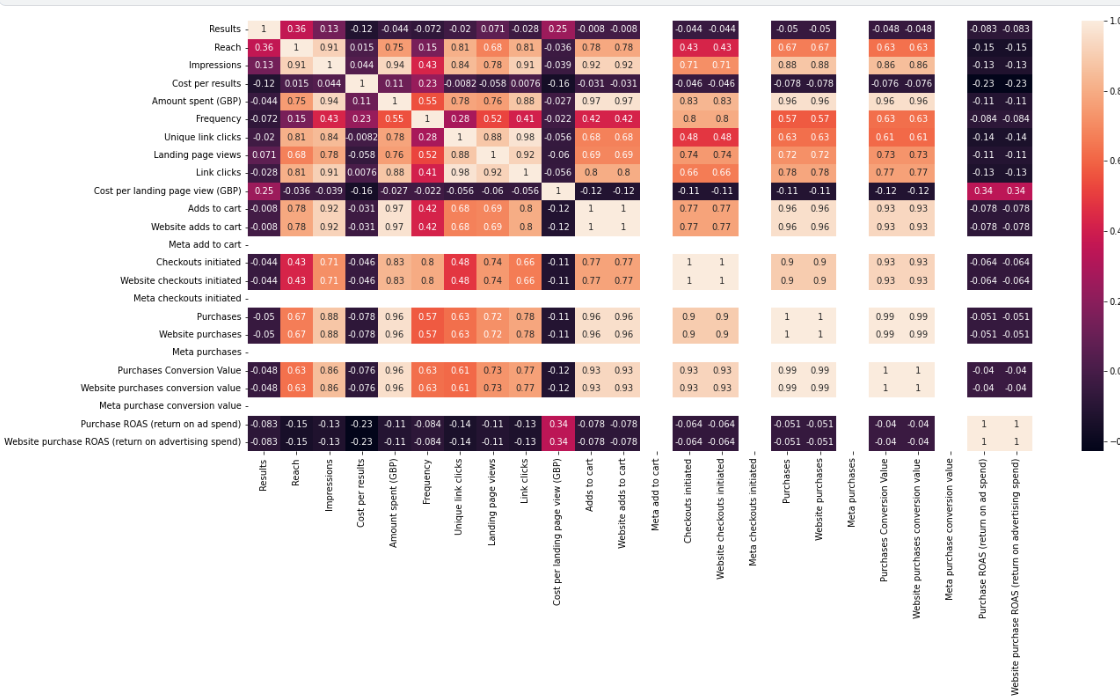
## **OTHER CAMPAIGNS**

In the examinations made in the other campaigns table, it was seen that many of the examined columns contained exactly the same values. These are: Website with Adds to cart, Adds to cart, Checkouts initiated with Website checkouts initiated, Website purchases with Purchases, Purchases Conversion Value with Website purchases conversion value, are dropped.

The completely empty Meta Add to Cart, Meta Purchases, Meta Purchase Conversion Value columns were removed from the dataset.

|  |  |
| --- | --- |
| **FEATURE (COLUMN)** | **THE FUNCTION OF COLUMN** |
| Reporting starts | * That column is mentioning about the date on which reporting starts. |
| Reporting ends | * That column contains the date on which the reporting ends. |
| Campaign name | * The campaign name which made |
| Campaign delivery | * Mentions the sitiation of campaign |
| Ad set budget | * This tells us the amount of the money that the seller spent on budget type basis. |
| Ad set budget type | * In this column it is mentioned that on which basis the seller spent the ad budget. (Daily or Lifetime) |
| Attribution setting | * The purpose of this column is telling the type of the ads. (28-day click or 1-day view,7-day click or 1-day view) |
| Results | * Number of Results which was aimed in Result indicator. |
| Result indicator | * The purpose of add set |
| Reach | * The number the ad reached |
| Impressions | * Total impression the ad taken(click,view etc.) |
| Cost per results | * Shows the cost per result |
| Amount spent (GBP) | * Result \* Cost Per Result |
| Ends | * Shows the ende date of campaign |
| Frequency | * the average number of times users see the ad. |
| Unique link clicks | * the number of people who clicked |
| Landing page views | * people landing on ad's destination URL |
| Link clicks | * the number of clicks on links within the ad that led to destinations |
| Cost per landing page view (GBP) | * the total amount spent divided by the amount of landing page views. |
| Adds to cart | * allows customers to choose items to purchase without actually completing the payment. |
| Website adds to cart | * allows customers to choose items to purchase without actually completing the payment. |
| Meta add to cart | * allows customers to choose items to purchase without actually completing the payment on Facebook or Instagram |
| Checkouts initiated | * The number of purchase launch events tracked by the pixel or Conversions API on the website and attributed to the ads. |
| Website checkouts initiated | * The number of purchase launch events tracked by the pixel or Conversions API on the website and attributed to the ads. |
| Meta checkouts initiated | * The number of initiate checkout events attributed to the ads |
| Purchases | * The number of purchases made within Meta technologies (such as Pages or Messenger) and attributed to the ads |
| Website purchases | * The number of total purchases made within the website. |
| Meta purchases | * The number of total purchases made within Meta Technologies. |
| Purchases Conversion Value | * tracks the total value of purchases made from your advertising efforts |
| Website purchases conversion value | * tracks the total value of purchases made from your advertising efforts |
| Meta purchase conversion value | * The total value of website purchases conversions. |
| Purchase ROAS (return on ad spend) | * the total revenue generated from your Facebook ads (your return) divided by your total ad spend. |
| Website purchase ROAS (return on advertising spend) | * the total revenue generated from your Facebook ads (your return) divided by your total ad spend |

**Table 7: Raw Features of Other Campaigns**



**Table 8:Correlation Map of Other Campaigns**

|  |  |
| --- | --- |
| **FEATURE (COLUMN)** | **ACTIONS TAKEN** |
| Adds to cart | * It contains same values with WebSite Adds to cart. It is duplicated so it is dropped. |
| Meta add to cart | * It is completely null.So it is dropped. |
| Checkouts initiated | * It contains same values withWebsite checkouts initiated. It is duplicated so it is dropped. |
| Meta checkouts initiated | * It is completely null.So it is dropped. |
| Purchases | * It contains same values withWebsite purchases. It is duplicated so it is dropped. |
| Meta purchases | * It is completely null.So it is dropped. |
| Purchases Conversion Value | * It contains same values withWebsite purchases conversion value. It is duplicated so it is dropped. |
| Meta purchase conversion value | * It is completely null.So it is dropped. |
| Purchase ROAS (return on ad spend) | * It contains same values withWebsite purchase ROAS (return on advertising spend). It is duplicated so it is dropped. |

**Table 9:Actions Taken on Other Campaigns**

### **Recommendations For Other Campaigns**

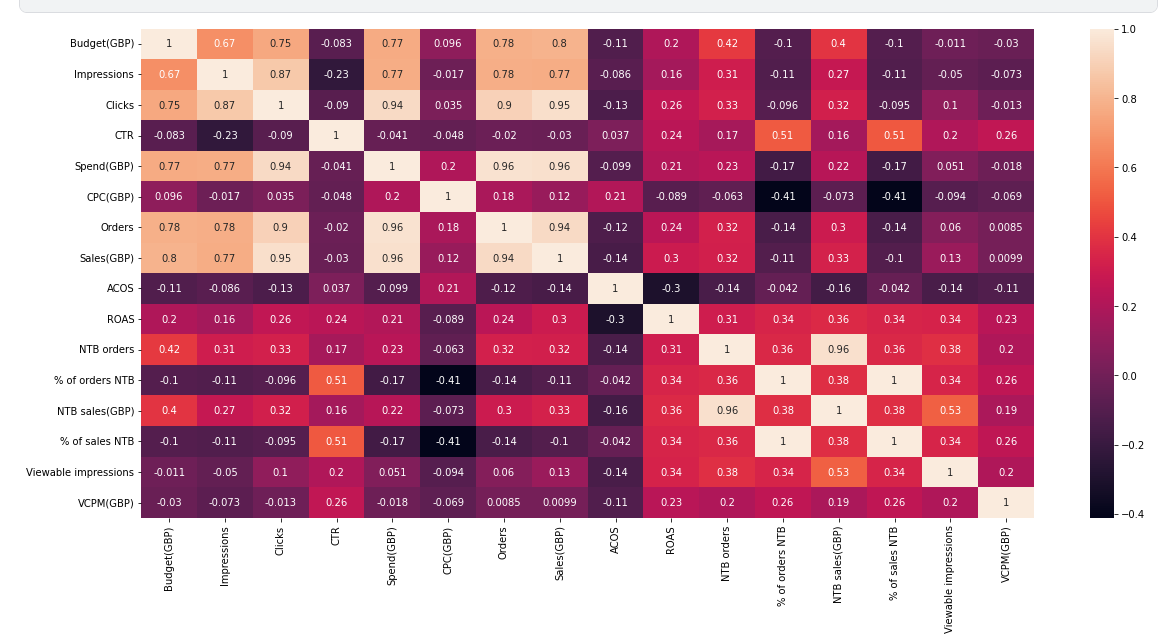
After examining the Other Campaigns table, it is thought that if columns such as Meta Add to Cart, Meta Checksout Initiated, Meta Purchases are completely blank, this data behooves to be filled completely.

## **AMAZON CAMPAIGNS**

When the table was examined, it was seen that the data rows were empty in the same way and these rows were dropped. The other columns have been including significant values. Besides that to get an idea about columns as Type,Campaign Binding Strategy, Portfolio,ROAS discussion goes further. At the and we have taken the necessary answers. Consequently we have decided,it is recomended that the dataset stay the same with existing columns.

|  |  |
| --- | --- |
| **FEATURE (COLUMN)** | **THE FUNCTION OF COLUMN** |
| State | * That column shows the status of the campanígn |
| Campaigns | * This is the name of the campaign |
| Status | * The actual status of the campaign it includes valuable information. |
| Type | * Sponsored Products (SP), Sponsored Brands (SB),  Sponsored Brand Video (SBV), Sponsored Display (SD), |
| Targeting | * choosing the specific keywords and products you wish to target and set bids accordingly. |
| Campaign bidding strategy | * When an Amazon customer performs a search for a product, the sellers with the highest bids on relevant keywords win the auction, and their product ads get listed in their chosen placement |
| Start date | * Start date of campaign |
| End date | * End date of campaign |
| Portfolio | * If you have an existing Portfolio in your account, you can optionally associate the campaign being created to a particular portfolio |
| Budget(GBP) | * A daily budget |
| Top-of-search IS | * the percentage of top-of-search impressions your campaign received out of the total top-of-search impressions it was eligible to serve on |
| Cost type | * CPC is the cost per click that an ad receives. |
| Impressions | * measure the number of times Amazon shows shoppers your Ad, regardless of whether they clicked on it or not. |
| Clicks | * Total number of clicks on ad |
| CTR | * the ratio between how many people have clicked on your Ad and the number of people who have seen it: |
| Spend(GBP) | * Total Spend for campaign |
| CPC(GBP) | * Cost per Click |
| Orders | * The number of orders taken |
| Sales(GBP) | * The amount of sales made through campaign |
| ACOS | * It compares the amount spent on PPC campaigns to the amount earned, and it helps determine if your brand generated campaigns that were cost-efficient. |
| ROAS | * ROAS (Return on advertising spend) is a metric that allows sellers to calculate the amount of income (or loss) from each invested dollar and evaluate the productivity of a particular ad campaign or even a keyword. |
| NTB orders | * The number of first-time orders for products on Amazon within the brand over a one-year lookback window |
| % of orders NTB | * The percent of first-time orders for products on Amazon within the brand over a one-year lookback window |
| NTB sales(GBP) | * The total amount of NTB sales |
| % of sales NTB | * The percentage of NTB sales |
| Viewable impressions | * This means that almost the number of measurable impressions. |
| VCPM(GBP) | * Viewable CPM represents the cost to serve only viewable impressions, which can be compared directly with the CPM that you have paid to serve all of your ad impressions. |

**Table 10: Raw Data of Amazon Campaigns**



**Table 11: Correlation Map of Amazon Campaigns**

* + 1. **Recommendatins For Amazon Campaigns**

After examining the Amazon Campaigns table, it was seen that a large part of it consisted of empty rows. It was suggested to the company that it is recommended to be filled in a more appropriate way in the same way.

## **PRODUCT REVIEWS**

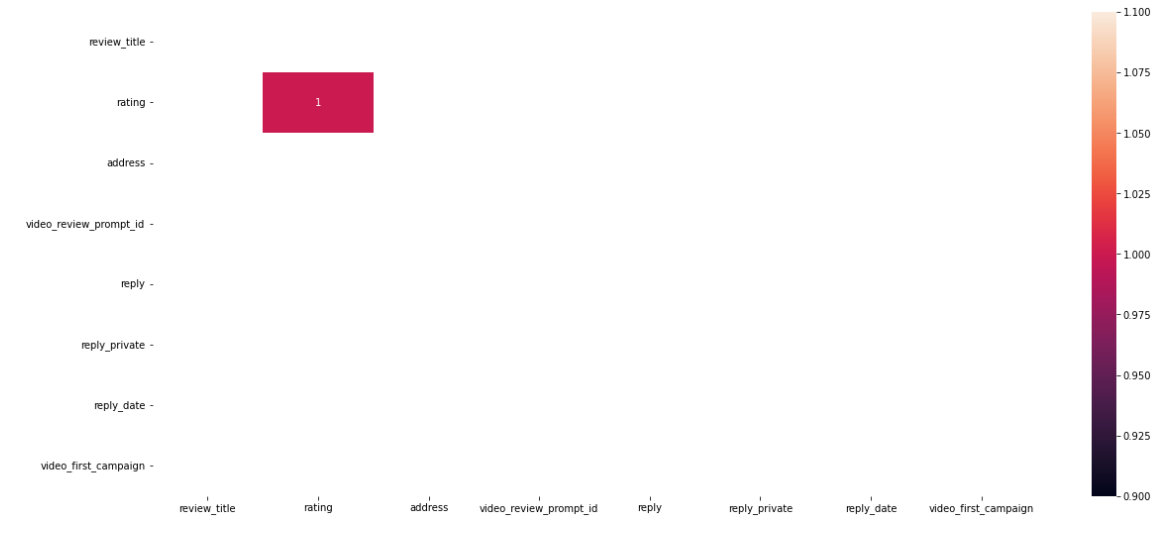
As seen in the data set received in the product comments table, empty data lines were seen as in the other tables. These were removed from the dataset.

|  |  |
| --- | --- |
| **FEATURE (COLUMN)** | **THE FUNCTION OF COLUMN** |
| order\_id | * The order id of the comment made |
| review\_title | * The review title(empty) |
| comments | * The comments made by customers |
| rating | * The rating made by customers 1 to 5. (1:lowest-5:greatest) |
| status | * The status of review active or inactive. |
| date\_created | * The date review created |
| sku | * The unique number of item |
| Customer ID | * The ID of customer who made the review |
| address | * Adress of customer (empty) |
| product\_sku | * The unique number of item |
| product\_name | * The name of product |
| product\_link | * The link of review made |
| video\_review\_prompt\_id | * Full of 199 |
| tags | * Tags of reviews |
| reply | * Replies to review made |
| reply\_private | * Private reply made |
| reply\_date | * The date of reply |
| published\_images | * The images published by making review |
| unpublished\_images | * Unpublished images by making review |
| published\_videos | * Published videos by making review |
| unpublished\_videos | * Unpublished videos by making review |
| source | * The source in that the comment made |
| location | * Location of customer who made the review |
| timeago | * The time indicator for how long time ago the comment or review made |
| video\_first\_campaign | * It is almost impossible to get insight from this column |

**Table 12:Raw data of product reviews**

|  |  |
| --- | --- |
| **FEATURE (COLUMN)** | **ACTIONS TAKEN** |
| sku | * We dropped the column because it has same values with 'product\_sku'. |
| address | * It is totally an empty column we can drop it |
| tags | * There are few values but we can drop it. |
| reply | * We can drop it. There are lots of null values. |
| reply\_private | * It is empty. |
| published\_images | * There are few values but it won't give us insight.The links on this column are not available. |
| unpublished\_images | * There are few values but it won't give us insight.The links on this column are not available. |
| published\_videos | * There are few values but it won't give us insight.The links on this column are not available. |
| unpublished\_videos | * There are few values but it won't give us insight.The links on this column are not available. |
| timeago | * There are few values but it won't give us insight. Because it has a lot of null rows. |
| video\_first\_campaign | * There are few values but it won't give us insight. Because it has a lot of null rows. |
|  |  |

**Table 13:Actions Taken on Product Reviews**



**Table 14: Correlation Map of Reviews Table**

### **Recommendations For Product Reviews Table**

The opinions obtained as a result of the investigations are as follows:

Columns mentioned or dropped above must be taken into account to be filled if the sales system allows it to fill. Additonally, the links of videos and images those provided on the Reviews Table, it is recommended to make on work. The relations between tables could’t be provided lack of table contents.

Formun Altı