

SUMMARY OF FINDINGS, INSIGHTS AND RECOMMENDATION BY ATINUKE ODESANMI

INTRODUCTION

The dataset was loaded into an excel sheet and the relevant columns were extracted which was used to analyze the dataset and only the first one thousand (1000) dataset was used.

The following were the findings and insights of the Pakistani largest E-commerce dataset:

DATA EXPLORATION

- The dataset's structure was examined and it was found out that some values on the grand total column was not properly computed so computation was done using the formula ($= E2*EF$) and the correct result is in column H. As for the missing values, the formula $=IF(COUNTIF(E2:F1000,H2),"OK","NOT OK")$ was used to determine the missing values as its seen in column L.
- A summary statistic was created for PRICE, QTY_ORDERED AND GRAND TOTAL using the summary statistics on the data analysis command button and the average was found to be 2458.3, 1.168 and 3680.58 respectively.

DESCRIPTIVE STATISTICS

- The mean, median, standard deviation and quartiles were found for price and qty_ordered. For PRICE, the formula used for mean was =average, for median was =median, for mode was = mode, for standard deviation was =stdeva, and for quartiles it was =quartile as seen at the bottom of the visualization table.
- Visualization was done for price, qty_ordered and grand total and the average, median, 1st quartile, 3rd quartile, range, minimum and maximum values were calculated and used to plot the charts. The min values for the three variables were found to be 1 while the maximum values were found to be 96499, 34 and 1155966 respectively. These were represented graphically as seen in the visualization page.
- the product categories and the payment method were categorical data and were change to numerical data(frequency table) using the countif

formula and both were calculated to find the trend forecast using the =trend formula. It was found out that the products categories that are most popular are: beauty & grooming having a value of 302, soghaat having a value of 290 and Men's fashion having a value of 114. The payment methods that are most commonly used is COD (cash on delivery) having a value of 824. It shows that most orders are actually completed successfully using COD. A graphical representation of the trendline of the products categories and payment method is shown on the visualization sheets.

INFERENTIAL STATISTICS

- Hypothesis were formulated to find out if there's a difference between the order status and the product categories and if certain products have a highly likelihood to be cancelled or refunded. And different statistical test such as t-test, anova and chi-squared test were used to test this hypothesis and using the status, payment method and product category frequency table it was found out that:

For the t-test which has a value of 0.2806 greater than the table value of 0.05. the chi-squared test which has a value of 2.299, and the ANOVA test which has a value greater than 0.05, indicates that there's no significant difference between the order status and the product categories. And certain products are not likely to be cancelled or refunded but they are likely to be completed.

NOTE: The summary and insights of this findings are for the first one thousand (1000) dataset of the Pakistani largest e-commerce dataset.

RECOMMENDATIONS

Mobile & tablet product should be added as it appears there's a good probability of pakistina customers buying it. Also, More studies need to be carried out on the large dataset to get more insights.

