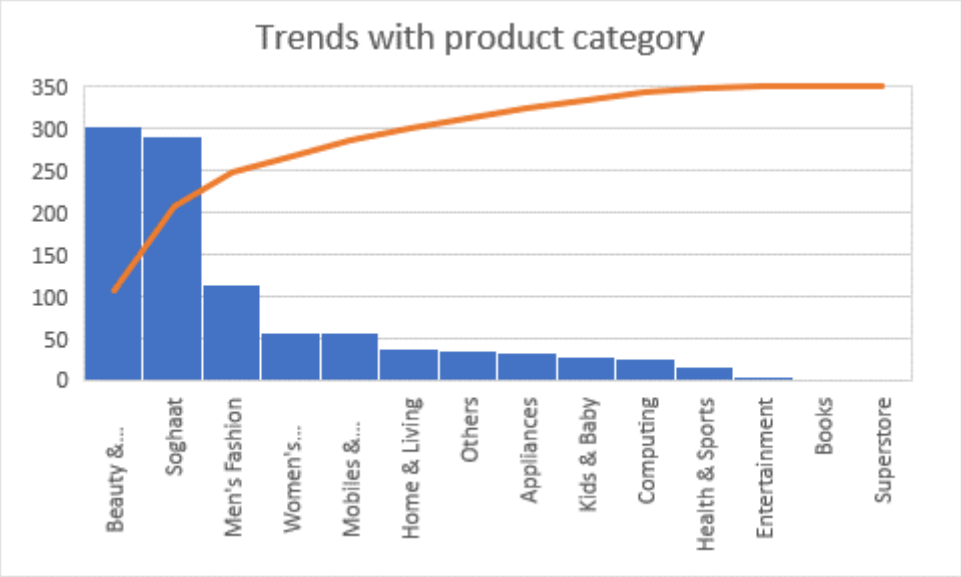
PROJECT TASK FINDINGS

1. DATA EXPLORATION
   - The dataset was loaded into an Excel sheet
   - The structure of the dataset was examined and it has about 4 datatypes; float, integer, string and date.
   - The missing values were checked with the Countblank function but there were no missing values.
   - The data analysis function was used to create summary statistics for price, quantity and grand total variables.

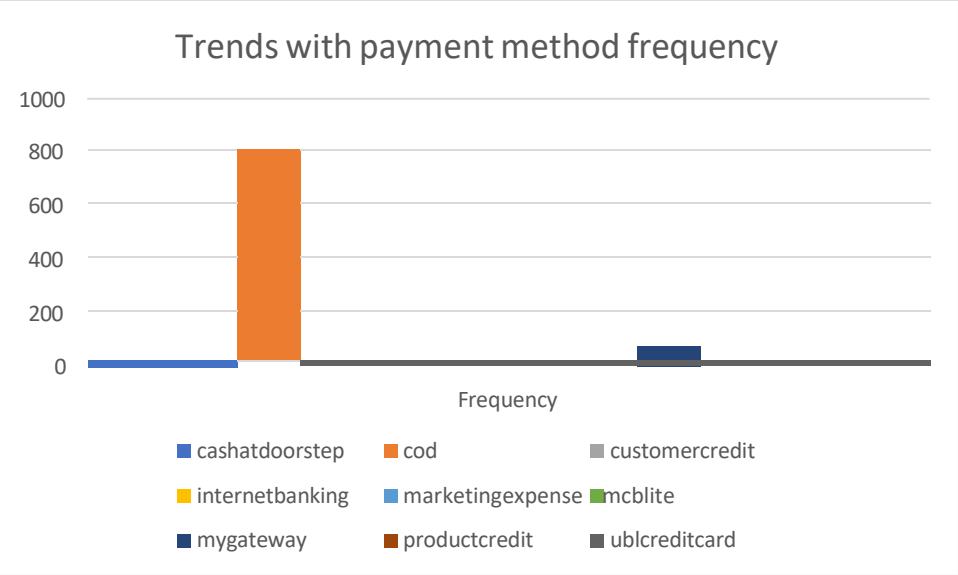
Summary Statistics for key variables (Price, Quantity and Grandtotal)					
price		*qty_ordered*		*grand_total*	
Mean	2456.34415	Mean	1.168	Mean	4828.16811
Standard Error	290.9716432	Standard Error	0.038978139	Standard Error	1193.26342
Median	360	Median	1	Median	639.5
Mode	360	Mode	1	Mode	360
Standard Deviation	9201.33127	Standard Deviation	1.232596972	Standard Deviation	37734.30255
Sample Variance	84664497.14	Sample Variance	1.519295295	Sample Variance	1423877589
Kurtosis	55.36675732	Kurtosis	512.6827926	Kurtosis	869.3557741
Skewness	6.963865882	Skewness	20.33475516	Skewness	28.56813012
Range	96498	Range	33	Range	1155966
Minimum	1	Minimum	1	Minimum	0
Maximum	96499	Maximum	34	Maximum	1155966
Sum	2456344.15	Sum	1168	Sum	4828168.11
Count	1000	Count	1000	Count	1000

2. **DESCRIPTIVE STATISTICS**
   - The descriptive statistics (mean, median, standard deviation and quartiles) were calculated for the price, quantity, grand_total, discount and MV variables.
     I used the 'AVERAGE' function for mean, 'MEDIAN' function for median, 'STDEV' for standard deviation and 'QUARTILES' function for Q1, Q2 and Q3.
   - For the Visualizations, scatterplot and line chart were used for each of these variables to gain a visual understanding of the distribution and trends.
   - To identify the trends in the data, to know which product categories are more popular;
     A countIF function was used to count the product categories. Then a bar chart was created for visualization.
     I could identify from the trends that the most popular products are:
     - Beauty & Grooming
     - Soghaat



Also, from the trends for payment methods, the payment methods which are commonly used is:

- Cod



3. Inferential hypothesis;
   - Is there a significant difference in the average order total between different payment methods?

     Null hypothesis (H0): There is no significant difference in the average order total between different payment methods.
     Alternative hypothesis (Ha): There is a significant difference in the average order total between different payment methods.

Inferential statistics using Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
cashatdoorstep	24	14015	583.9583333	2489.085145
cod	825	1926600.95	2335.273879	49258752.49
customercredit	1	0	0	#DIV/0!
internetbanking	8	34478	4309.75	42378954.5
marketingexpense	1	1330	1330	#DIV/0!
mcblite	10	62389	6238.9	217157946.5
mygateway	78	2245955.16	28794.29692	16876138859
productcredit	1	0	0	#DIV/0!
ublcreditcard	36	543400	15094.44444	729769650.4

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	54229693787	8	6778711723	4.831866227	7.45022E-06	1.94788321
Within Groups	1.36784E+12	975	1402917921			
Total	1.42207E+12	983				

P-value = 0.0000074502199657382 which is less than 0.05
That means the 'Null hypothesis (H0)' is rejected and the 'Alternative hypothesis' is upheld.

- Do certain product categories have a higher likelihood of being canceled or refunded?

  Null Hypothesis (H0): There is no significant difference in the likelihood of cancellations or refunds across different product categories.
  Alternative Hypothesis (H1): There is a significant difference in the likelihood of cancellations or refunds across different product categories.

  Degree of freedom = 52
  significance level = 0.05
  chi-squared statistic = 227.29
  Critical Value = 69.83

Chi-Squared Statistics is greater than the critical value which means that there is a higher chance of some product categories to be cancelled or refunded.

The Alternative hypothesis is upheld while the Null hypothesis (H0) is cancelled.

RECOMMEDATIONS:

I will recommend that the e-commerce businesses operating in Pakistan should look at the product categories that may be cancelled or refunded and conduct a further analysis to understand it root cause which might be due to the payment methods, quality issues or customer satisfaction.
They can also get feedback from customers who have refunded or cancelled some products and use their insights to provide valuable information.