

E-commerce Analytics: Understanding Pakistani Market

The e-commerce industry continues to experience rapid growth, driven by evolving consumer preferences, technological advancements, and global market expansion. This report provides an analysis of the current state of e-commerce, key trends, challenges, and future prospects. It aims to offer insights for businesses and stakeholders in the e-commerce sector.

This report contains the breakdown of the analysis performed on the Pakistan E-commerce dataset using a spreadsheet tool. The dataset contains an initial size of 584525 rows and 21 columns, in which a sample of 5000 rows was extracted for analysis using python, due to the size and complexity of the dataset, for easy analysis.

1. Data Exploration: The dataset structure was examined, where missing values were dropped due to the low percentage of missing values, attributes were set to the correct data type. Thereafter, to understand key variables of the dataset, summary statistics was carried out; which include *Price*, *Quantity Ordered*, *Grand Total*, *MV*.

Summary Statistics									
price		qty_ordered		grand_total		discount_amount		MV	
Mean	3007.21327	Mean	1.1434	Mean	4249.84921	Mean	11.568198	Mean	3437.3802
Standard Error	157.120262	Standard Error	0.013507175	Standard Error	296.5262823	Standard Error	1.391802927	Standard Error	292.6021735
Median	425	Median	1	Median	780	Median	0	Median	460
Mode	360	Mode	1	Mode	360	Mode	0	Mode	360
Standard Deviation	11110.08027	Standard Deviation	0.955101504	Standard Deviation	20967.5745	Standard Deviation	98.41532876	Standard Deviation	20690.09811
Sample Variance	123433883.6	Sample Variance	0.912218884	Sample Variance	439639180.5	Sample Variance	9685.576934	Sample Variance	428080159.6
Kurtosis	63.18039353	Kurtosis	457.0947007	Kurtosis	1843.605969	Kurtosis	1361.136026	Kurtosis	1950.261086
Skewness	7.134099695	Skewness	17.75606026	Skewness	35.49617994	Skewness	29.77749858	Skewness	36.99457136
Range	190724.85	Range	33	Range	1155966	Range	5000	Range	1155966
Minimum	0.15	Minimum	1	Minimum	0	Minimum	0	Minimum	0
Maximum	190725	Maximum	34	Maximum	1155966	Maximum	5000	Maximum	1155966
1st Quartile	240	1st Quartile	1	1st Quartile	320	1st Quartile		1st Quartile	246.5
3rd Quartile	1101.5	3rd Quartile	1	3rd Quartile	1998	3rd Quartile		3rd Quartile	1200
Sum	Rs 15,036,066.35	Sum	5717	Sum	Rs 21,249,246.05	Sum	Rs 57,840.99	Sum	Rs 17,186,901.00
Count	5000	Count	5000	Count	5000	Count	5000	Count	5000

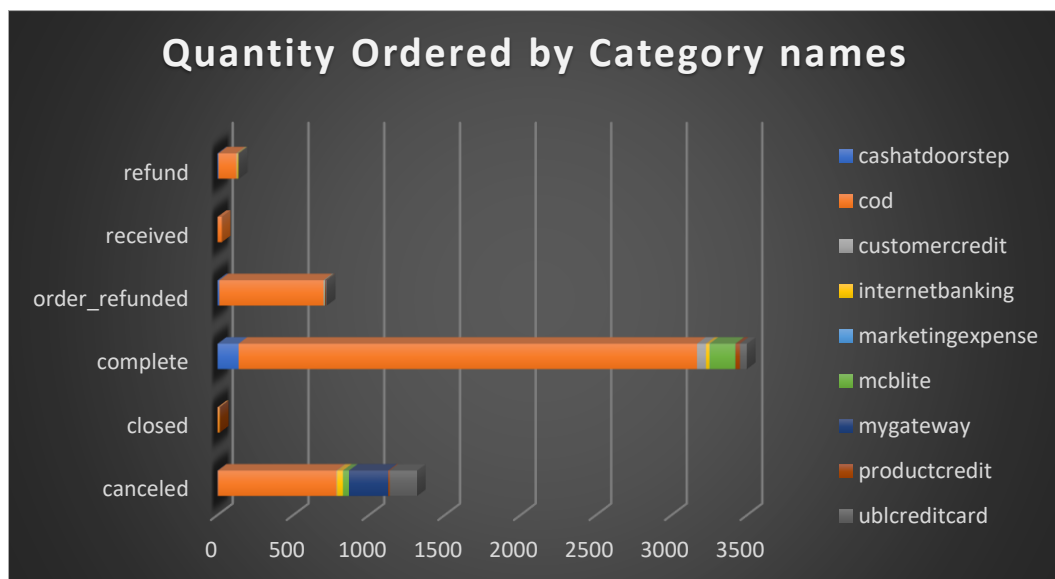
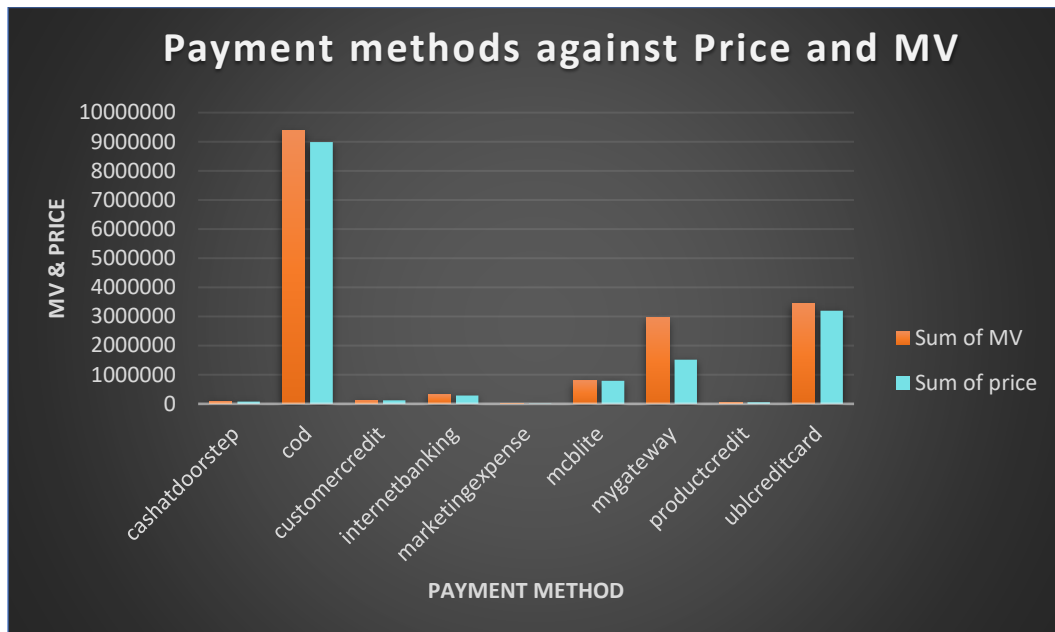
2. Descriptive Statistics: Here, a description on categories was carried out displaying each total unit for each unique category. Trends in the dataset were found.

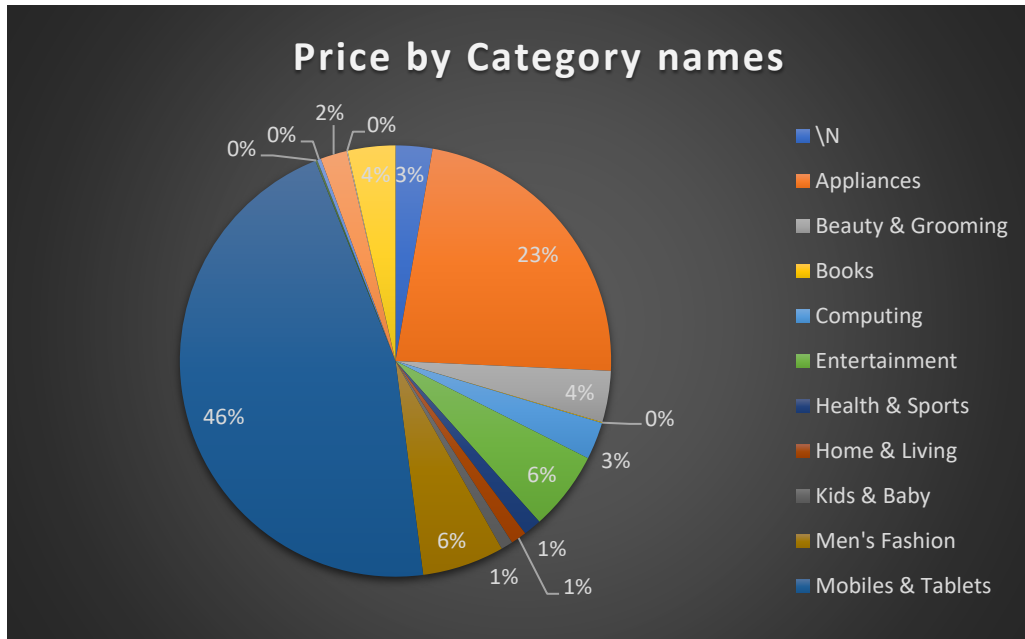
Descriptive Statistics								
product_category	Units	payment_method	Number	status	Number	Total Orders	5717	
Women's Fashion	261	cod	4061	complete	3133	Average Orders	1.1434	
Beauty & Grooming	1313	ublcrcard	194	canceled	1030	Payment Variance	1734085.778	
Soghaat	1066	mygateway	211	order_refunded	674	Max Product_category	1313	Beauty & Grooming
Mobiles & Tablets	398	customercredit	70	received	25	Min Product_category	11	Superstore
Appliances	230	cashatdoorstep	145	refund	123	Popular Payment Method	4061	cod
Home & Living	225	mclite	212	closed	15	Unpopular payment method	1	marketingexpense
Men's Fashion	576	internetbanking	60					
Kids & Baby	98	marketingexpense	1					
\N	418	productcredit	46					
Others	20							
Entertainment	32							
Computing	151							
Superstore	11							
Health & Sports	120							
Books	11							
School & Education	70							

3. Inferential Statistics: In this section, hypothesis testing was carried out. The z-test, t-test was carried out to check the association of the variables.

Inferential Statistics					
z-Test: Two Sample for Means			t-Test: Two-Sample Assuming Unequal Variances		
	Orders	Payment Methods		Orders	Payment Methods
Mean	1.1434	555.5555556	Mean	1.1434	555.5555556
Known Variance	0.912218884	1734085.778	Variance	0.912218884	1734085.778
Observations	5000	9	Observations	5000	9
Hypothesized Mean	0		Hypothesized Mean	0	
z	-1.263044669		df	8	
P(Z<=z) one-tail	0.103286564		t Stat	-1.263044669	
z Critical one-tail	1.644853627		P(T<=t) one-tail	0.12107007	
P(Z<=z) two-tail	0.206573127		t Critical one-tail	1.859548038	
z Critical two-tail	1.959963985		P(T<=t) two-tail	0.24214014	
			t Critical two-tail	2.306004135	

4. Visualisation: This section consists of plots carried out in the analysis.





5. Conclusion: The e-commerce industry is poised for continued growth, driven by technological innovation and evolving consumer behaviours. This analysis serves as a snapshot of the e-commerce landscape as of July, 2016, and the industry is expected to undergo further transformation in the years to come. Staying informed about market dynamics and consumer preferences will be essential for success in this dynamic sector.

It is observed that consumers are more comfortable with the Cash-on-delivery (cod) method. From the hypothesis carried out, it shows that there is an association between Orders and Payment methods.