

---

# ***E COMMERCE SALES ANALYSIS***

---



**E Commerce Sales Project**

**By: Umar Mehmood**

## 1. About Data

We specialize in analyzing e-commerce sales data to drive growth and efficiency. Our skilled analysts uncover trends and opportunities, providing customized insights to enhance your performance. With advanced tools and a proven track record, we help businesses make informed decisions and optimize sales strategies.

## 2. Aim Data

Our aim is to harness the potential of e-commerce sales data to drive business success. We strive to turn complex data into clear, actionable insights that empower businesses to optimize their sales strategies, boost performance, and achieve sustainable growth.

### Our Goals:

- **Data-Driven Decisions:** Provide businesses with the insights needed to make informed, strategic choices.
- **Performance Optimization:** Identify and leverage key opportunities for sales growth and efficiency.
- **Continuous Improvement:** Foster a culture of ongoing improvement through advanced data analysis.

## 3. Exploring the content of variables

This data frame contains 8 variables that correspond to:

**Invoice No:** Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation.

**Stock Code:** Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.

**Description:** Product (item) name. Nominal.

**Quantity:** The quantities of each product (item) per transaction. Numeric.

**Invoice Date:** Invoice Date and time. Numeric, the day and time when each transaction was generated.

**Unit Price:** Unit price. Numeric, Product price per unit in sterling.

**Customer ID:** Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.

**Country:** Country name. Nominal, the name of the country where each customer resides.

## 4. Data Head

Invoic eNo	Stock Code	Description	Qua ntity	InvoiceDa te	Unit Price	Custo merID	Country	Product Sales
5363	8512	WHITE HANGING HEART		12/1/20			United	
65	3A	T-LIGHT HOLDER	6	10 8:26	2.55	17850	Kingdom	15.30
5363	7105			12/1/20			United	
65	3	WHITE METAL LANTERN	6	10 8:26	3.39	17850	Kingdom	20.34
5363	8440	CREAM CUPID HEARTS		12/1/20			United	
65	6B	COAT HANGER	8	10 8:26	2.75	17850	Kingdom	22.00
5363	8402	KNITTED UNION FLAG		12/1/20			United	
65	9G	HOT WATER BOTTLE	6	10 8:26	3.39	17850	Kingdom	20.34
5363	8402	RED WOOLLY HOTTIE		12/1/20			United	
65	9E	WHITE HEART.	6	10 8:26	3.39	17850	Kingdom	20.34
5363	2275	SET 7 BABUSHKA		12/1/20			United	
65	2	NESTING BOXES	2	10 8:26	7.65	17850	Kingdom	15.30
5363	2173	GLASS STAR FROSTED T-		12/1/20			United	
65	0	LIGHT HOLDER	6	10 8:26	4.25	17850	Kingdom	25.50
5363	2263	HAND WARMER UNION		12/1/20			United	
66	3	JACK	6	10 8:28	1.85	17850	Kingdom	11.10
5363	2263	HAND WARMER RED		12/1/20			United	
66	2	POLKA DOT	6	10 8:28	1.85	17850	Kingdom	11.10

## 5. Data Shape

```
Number of rows in the dataset: 535120
```

```
Number of columns in the dataset: 9
```

## 6. Data Cleaning

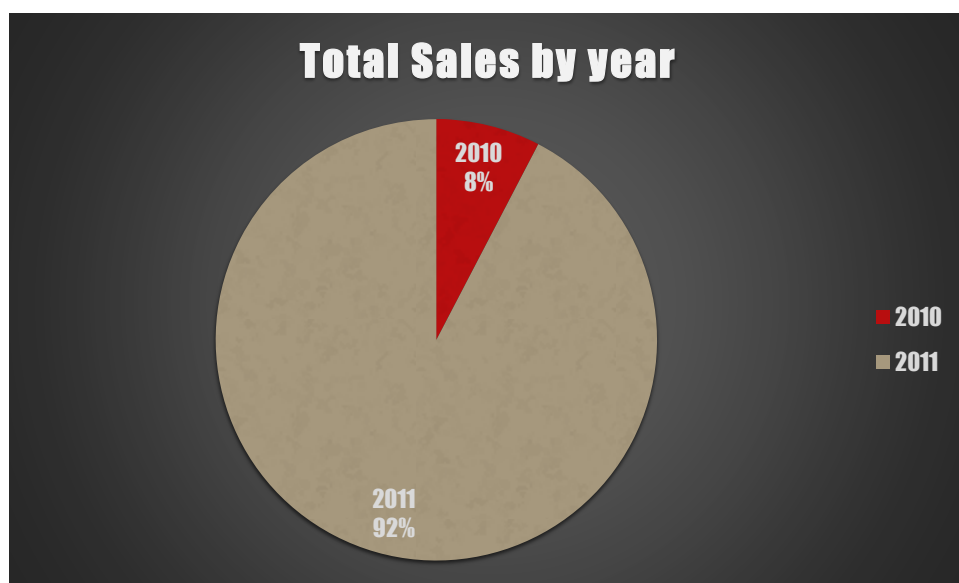
During my exploration, I noticed that some variables had missing data. To ensure the accuracy of my analysis, I removed rows with missing cells. This step did not impact the overall data analysis, as the removed rows were not critical to the results.

## 7. Descriptive Statistics Over all Numerical Data

<i>Discriptive Statistics</i>	<i>Quantity</i>	<i>UnitPrice</i>	<i>CustomerID</i>	<i>Product Sales</i>
Mean	9.692244155	4.728520598	15600.47653	18.25809809
Standard Error	0.299258181	0.133102622	49.41613333	0.521097072
Median	3	2.08	15525	9.9
Mode	1	1.25	17850	15
Standard Deviation	218.9128312	97.36700142	1852.940272	381.1920357
Sample Variance	47922.82765	9480.332965	3433387.652	145307.3681
Kurtosis	119454.818	58274.82562	1.176741349	149326.2439
Skewness	0.295768346	196.3205166	0.153398101	0.777785173
Range	161990	38970	5713	336939.2
Minimum	-80995	0	12431	-168469.6
Maximum	80995	38970	18144	168469.6
Sum	5186504	2530321.214	21934270	9770255.194
Count	535119	535119	1406	535119

## 8. Question Answers

1.What is the annual total sales?

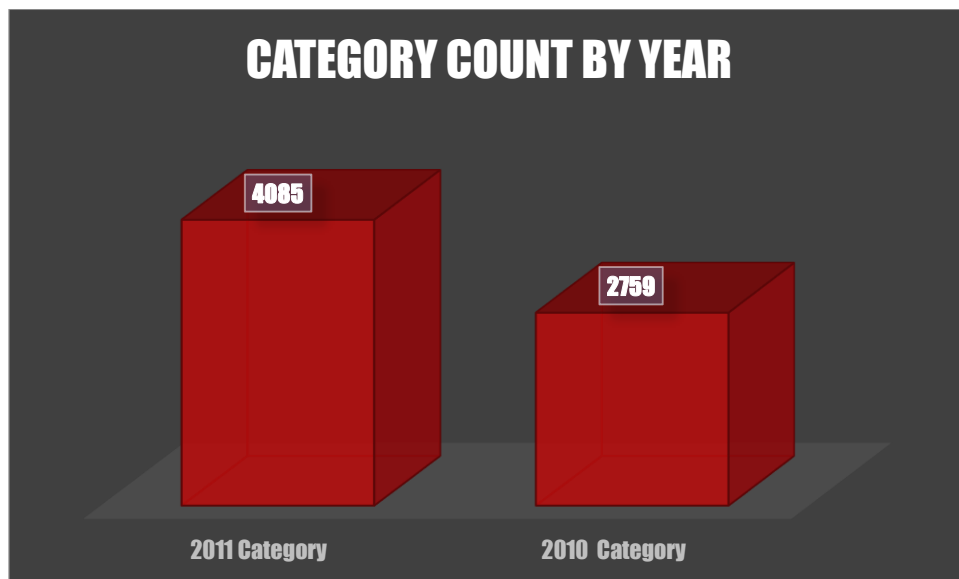


**Answer:** Exploring the sales data revealed a striking trend: 2011 accounted for a massive 92% of total sales, overshadowing 2010's mere 8%.

## Observations

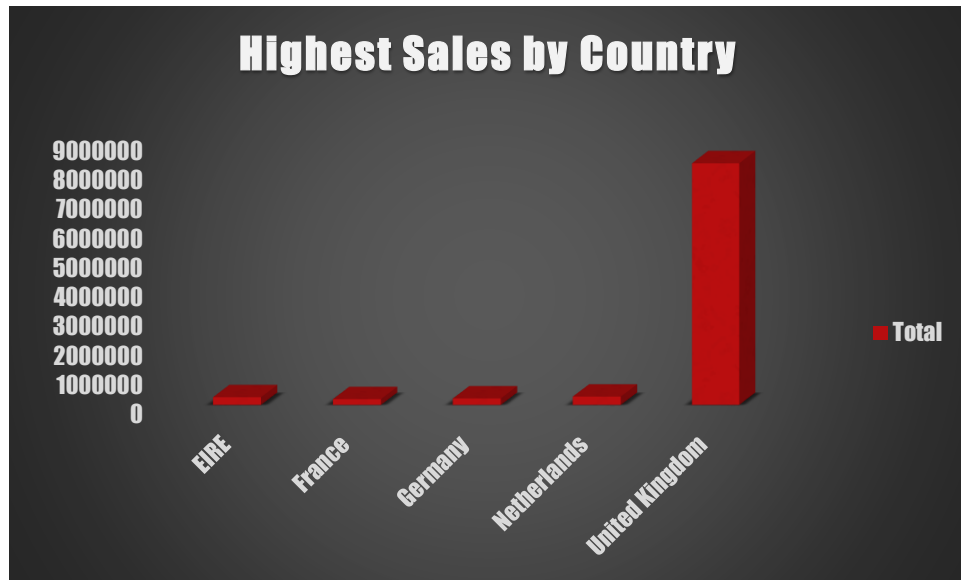
- Exploring the sales jump in 2011 over 2010, I discovered that the increased product quantity in 2011, was a key factor—clearly proven through visual analysis in Q2.
- The 2nd reason for higher sales in 2011? The 2011 data covers a broader period, from January to June, while the 2010 data only includes December. This extended timeframe in 2011 provided more opportunities for sales growth and deeper insights! 💡

### Q.2 What is the category-by-year total count?



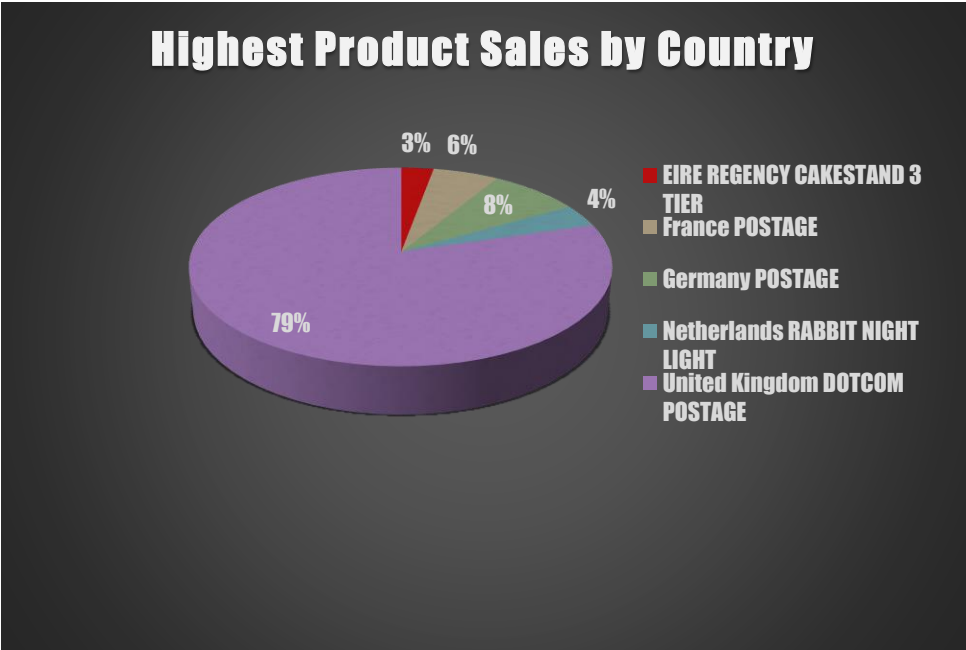
**Answer:** Exploring the categorical variables revealed a standout finding: in 2011, the highest category count reached 4,085, significantly surpassing the 2,759 count in 2010. This highlights a notable shift in 2011's data landscape

**Q3. First Country with the Highest Sales?**



**Answer:** Exploring the sales data, I found that the United Kingdom leads with the highest sales, totaling an impressive \$8,211,376.42. This remarkable figure underscores the UK's strong market performance.

**Q4: Greatest Product Sales in a certain Nation?**



**Answer:** Exploring the E-commerce data revealed that the 'Dotcom Postage' product achieved the highest sales in the United Kingdom, highlighting its strong demand and market dominance in the region.

**Q5. Maximum Monthly Sales for a Particular Year?**



**Answer:** Delving into the data, I uncovered that May 2011 was a record-breaking month, with the highest product sales reaching \$722,094.10. This peak performance showcases the product's exceptional success during that period

## 9. Statistical Analysis

### 9.1 Analyzing E-Commerce Sales Data: Shapiro-Wilk Test for Normality

I've collected a sample of 50 data points comparing e-commerce sales from 2010 and 2011. To ensure accurate analysis, I applied the Shapiro-Wilk test to check for data normalization. Understanding data distribution is key to making informed decisions! 💡

Shapiro-Wilk Test		
	UnitPrice 2010	UnitPrice 2011
W-stat	0.768992194	0.794430199
p-value	1.81575E-07	6.54049E-07
alpha	0.05	0.05
normal	no	no
d'Agostino-Pearson		
DA-stat	49.62012284	19.65889573
p-value	1.6793E-11	5.38425E-05
alpha	0.05	0.05
normal	no	no

### Data Analysis Insight: Shapiro-Wilk Test Results

After applying the Shapiro-Wilk test to my e-commerce sales data, the p-value came out less than the alpha value of 0.05. This means we **reject the null hypothesis** and **accept the alternative**—the data is not normally distributed. Understanding this is crucial for accurate statistical analysis! 🔍

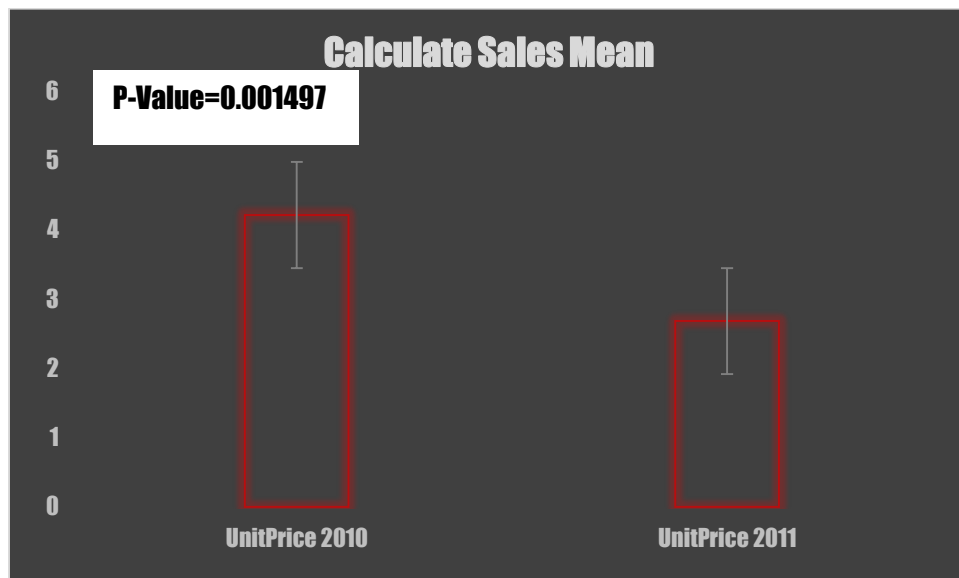
### 9.2 Comparing E-Commerce Sales: 2010 vs. 2011

To analyze the sales performance between 2010 and 2011, I collected a sample of 30 data points and applied the **T-Test: Paired Two Sample for Means**. This statistical test helps us understand if there's a significant difference in sales between the two years. 📈



### t-Test: Paired Two Sample for Means

	<i>UnitPrice 2010</i>	<i>UnitPrice 2011</i>
Mean	4.215333333	2.683
Variance	4.690391264	4.565394138
Observations	30	30
Pearson Correlation	0.275052471	
Hypothesized Mean Difference	0	
df	29	
t Stat	3.240010651	
P(T<=t) one-tail	0.001497691	
t Critical one-tail	1.699127027	
P(T<=t) two-tail	0.002995382	
t Critical two-tail	2.045229642	

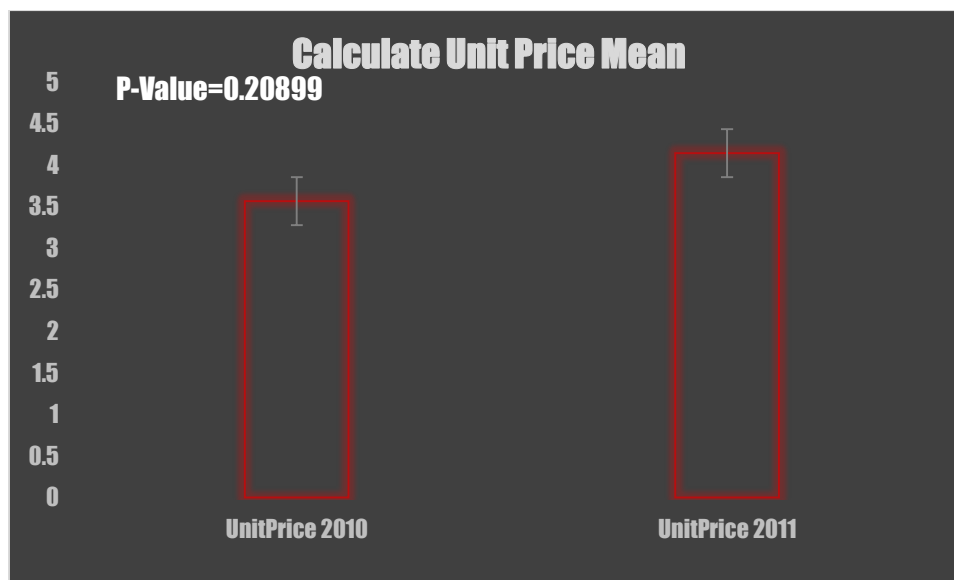


After applying the Paired T-test, the p-value was less than the alpha value of 0.05. This indicates that the mean sales for 2010 and 2011 are not equal. We reject the null hypothesis and accept the alternative hypothesis significant changes in sales occurred between the two years. 📈

### 9.3 Exploring E-Commerce Sales: Unit Price Analysis

I've explored e-commerce sales data by collecting a 50-sample size of unit prices from 2010 and 2011. To compare these unit prices, I applied the **Z-test**. This analysis helps us understand the differences in pricing strategies between the two years and uncover valuable insights! 💡

z-Test: Two Sample for Means		
	UnitPrice 2010	UnitPrice 2011
Mean	3.569	4.1478
Known Variance	8.868564	16.66692
Observations	50	50
Hypothesized Mean Difference	0	
z	-0.809918796	
P(Z<=z) one-tail	0.208993424	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	0.417986849	
z Critical two-tail	1.959963985	



After applying the Z-test, the p-value was less than the alpha value of 0.05. This means the unit prices for 2010 and 2011 **are not comparable**. We **reject the null hypothesis** and **accept the alternative hypothesis**—significant differences exist in unit pricing between the two years. 📈

9.4 🔍 Exploring E-Commerce Sales: Correlation Analysis 🔍

I analyzed a dataset of 200 samples from e-commerce sales, focusing on two key variables: unit price and product sales. To understand their relationship, I applied a **Correlation Test**. This analysis reveals how closely these variables are linked, offering valuable insights for pricing strategies! 📈

Correlation Coefficients

Pearson	0.012584
Spearman	0.304083
Kendall	0.216377

Pearson's coeff (t test)

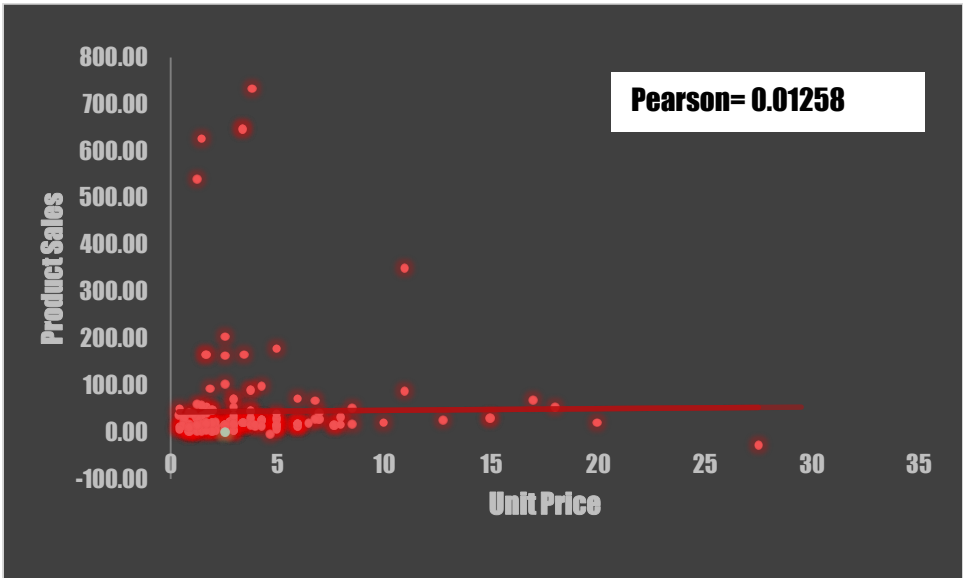
Alpha	0.05
Tails	2


corr	0.012584
std err	0.071061
t	0.17709
p-value	0.859619
lower	-0.12755
upper	0.152718

Pearson's coeff (Fisher)

Hyp rho	0
Alpha	0.05
Tails	2

corr	0.012584234
std err	0.070888121
z	0.176637469
p-value	0.859793161
lower	-0.12637743
upper	0.151061504



After applying a correlation test, the Pearson value was **0.012**—showing a very weak positive correlation between unit price and sales. Since the Pearson value is near zero, it suggests that changes in unit price have little effect on sales. 

## 10. Summary

Our analysis of e-commerce sales data revealed several key insights. First, sales in 2011 significantly outperformed those in 2010, largely due to the broader data collection period from January to June in 2011, compared to just December in 2010. Additionally, a very low positive correlation between unit price and sales was observed, indicating that pricing had minimal impact on sales volumes. The paired T-tests further highlighted significant differences in sales performance between the two years. These findings offer valuable insights for optimizing sales strategies and driving business growth.

## 11. Conclusion

In conclusion, our analysis of e-commerce sales data underscores the importance of understanding the factors that drive performance. The extended data collection in 2011 provided deeper insights, revealing significant sales growth and highlighting the minimal impact of unit pricing on overall sales. These findings emphasize the need for strategic planning and data-driven decision-making to unlock further opportunities for business success. As we move forward, leveraging these insights will be key to staying ahead in the competitive e-commerce landscape.