Statistical_Analytics

Descriptive Statistics
What is the average Sales, Profit spread, and most common Discount?

```
print("Average Sales:", df['Sales'].mean())
print("Profit Std Dev:", df['Profit'].std())
print("Most Common Discount:", df['Discount'].mode()[0])
```

```
Average Sales: 5019.265229999999
Profit Std Dev: 2449.2647104769712
Most Common Discount: 0.22
```

Correlation – Sales vs Profit
Question: Do Sales and Profit move together?

```
print(df[['Sales','Profit']].corr())
```

	Sales_Amount	Profit
Sales_Amount	1.00000	0.97841
Profit	0.97841	1.00000

Is the average Sales significantly higher than 500? (One-sample t-test)
Business case: Management expects ₹500 avg sales. Is it true?

```
from scipy import stats
t_stat, p_val = stats.ttest_1samp(df['Sales'], 500)
print("T-Statistic:", t_stat, "P-value:", p_val)
if p_val < 0.05:
   print(" Sales are significantly different from 500")
else:
   print("No significant difference from 500")
```

```
T-Statistic: 50.201001290357205 P-value: 2.052378436808552e-275
 Sales are significantly different from 500
```

Do customers in East and West have different Discount patterns? (Two-sample t-test)
Business case: Check if discount strategy differs by region.

```
east = df[df['Region']=='East']['Discount']
west = df[df['Region']=='West']['Discount']
t_stat, p_val = stats.ttest_ind(east, west)
print("T-Statistic:", t_stat, "P-value:", p_val)
if p_val < 0.05:
   print("Discounts differ between East & West")
else:
   print("No significant difference in Discounts")
```

```
T-Statistic: 1.857044498644543 P-value: 0.06388683865203344
 No significant difference in Discounts
```

Do Sales differ across sales rep (Bob,Alice,David)? (ANOVA)
Business case: Which sales rep shows more sales?

```
from scipy.stats import f_oneway

Bob = df[df['Sales_Rep']=='Bob']['Sales_Amount']
Alice = df[df['Sales_Rep']=='Alice']['Sales_Amount']
David = df[df['Sales_Rep']=='David']['Sales_Amount']
f_stat, p_val = stats.f_oneway(Bob,Alice,David)
print("F-Statistic:", f_stat, "P-value:", p_val)
if p_val < 0.05:
   print(" Sales differ across Sales_Rep")
else:
   print(" No major difference in Sales across Sales_Rep")
```

```
F-Statistic: 0.17937065862570692 P-value: 0.8358394735784171
 No major difference in Sales across Sales_Rep
```

Is there a relationship between Region and Category? (Chi-Square Test)
Business case: Do certain regions prefer certain product categories?

```
cont_table = pd.crosstab(df['Region'], df['Category'])
chi2, p, dof, expected = chi2_contingency(cont_table)
print("Chi2:", chi2, "P-value:", p)
if p < 0.05:
   print(" Region & Category are related")
else:
   print(" Region & Category are independent")
```

```
Chi2: 7.991425577219436 P-value: 0.5350105452846294
 Region & Category are independent
```

Is **Profit normally distributed? (Shapiro-Wilk Test)**
Business case: Check if profit follows a normal curve (bell-shaped).

```
from scipy.stats import shapiro
stat, p = shapiro(df['Profit'])
print("Statistic:", stat, "P-value:", p)
if p < 0.05:
   print(" Profit is NOT normally distributed")
else:
   print("Profit is normally distributed")
```

```
Statistic: 0.9652355623712482 P-value: 1.0340525235661562e-14
 Profit is NOT normally distributed
```

Is Discount correlated with Sales? (Correlation)
Business case: Do higher discounts increase sales volume?

```
corr = df['Discount'].corr(df['Sales'])
print("Correlation:", corr)
if corr > 0:
   print(" Higher discounts → Higher sales")
else:
   print(" Discounts do not boost sales much")
```

```
Correlation: 0.0231527033411976
 Higher discounts → Higher sales
```

Do Customers in North vs South have different Profits? (Two-sample t-test)
Business case: Compare profitability by region.

```
north = df[df['Region']=='North']['Profit']
south = df[df['Region']=='South']['Profit']
t_stat, p_val = stats.ttest_ind(north, south)
print("T-Statistic:", t_stat, "P-value:", p_val)
if p_val < 0.05:
   print("Profits differ significantly between North & South")
else:
   print(" Profits do not differ much")
```

```
T-Statistic: 0.0186116754327074 P-value: 0.9851584492638695
 Profits do not differ much
```