

=====

DATA OVERVIEW - COMPLETE DATA

=====

First 5 rows:

	TransactionID	ProductCategory	StoreLocation	UnitPrice	QuantitySold	TotalSale	PaymentMethod	CustomerAge	Returned
	:	:	:	:	:	:	:	:	:
0	T0001	Clothing	Mansoura	1603.67	4	6414.68	Mobile Payment	29	No
1	T0002	Clothing	Mansoura	3698.71	4	14794.8	Credit Card	34	No
2	T0003	Clothing	Alexandria	4189.86	6	25139.2	Credit Card	67	No
3	T0004	Electronics	Mansoura	3088.44	9	27796	Cash	48	No
4	T0005	Groceries	Mansoura	4495.7	10	44957	Mobile Payment	49	No

Data info:  
None

Numerical statistics:

	count	mean	std	min	25%	50%	75%	max
	:	:	:	:	:	:	:	:
UnitPrice	200	2681.79	1399.26	86.78	1725.84	2750.91	3873.68	4987.41
QuantitySold	200	5.88	3.00846	1	3	6	9	10
TotalSale	200	16067.7	12978.9	392.1	4672.22	13203.4	23733.4	49388.8
CustomerAge	200	43.25	15.335	18	30	41.5	56	70

Missing values:

Missing Values	Missing %
:	:

=====

DATA OVERVIEW - DATA WITH MISSING VALUES

=====

First 5 rows:

	TransactionID	ProductCategory	StoreLocation	UnitPrice	QuantitySold	TotalSale	PaymentMethod	CustomerAge	Returned
0	T0001	Clothing	Mansoura	1603.67	4	6414.68	Mobile Payment	nan	No
1	T0002	Clothing	Mansoura	nan	nan	14794.8	Credit Card	34	No
2	T0003	Clothing	Alexandria	4189.86	6	25139.2	Credit Card	67	No
3	T0004	Electronics	nan	nan	9	nan	Cash	nan	No
4	T0005	nan	Mansoura	4495.7	10	44957	Mobile Payment	49	nan

Data info:  
None

Numerical statistics:

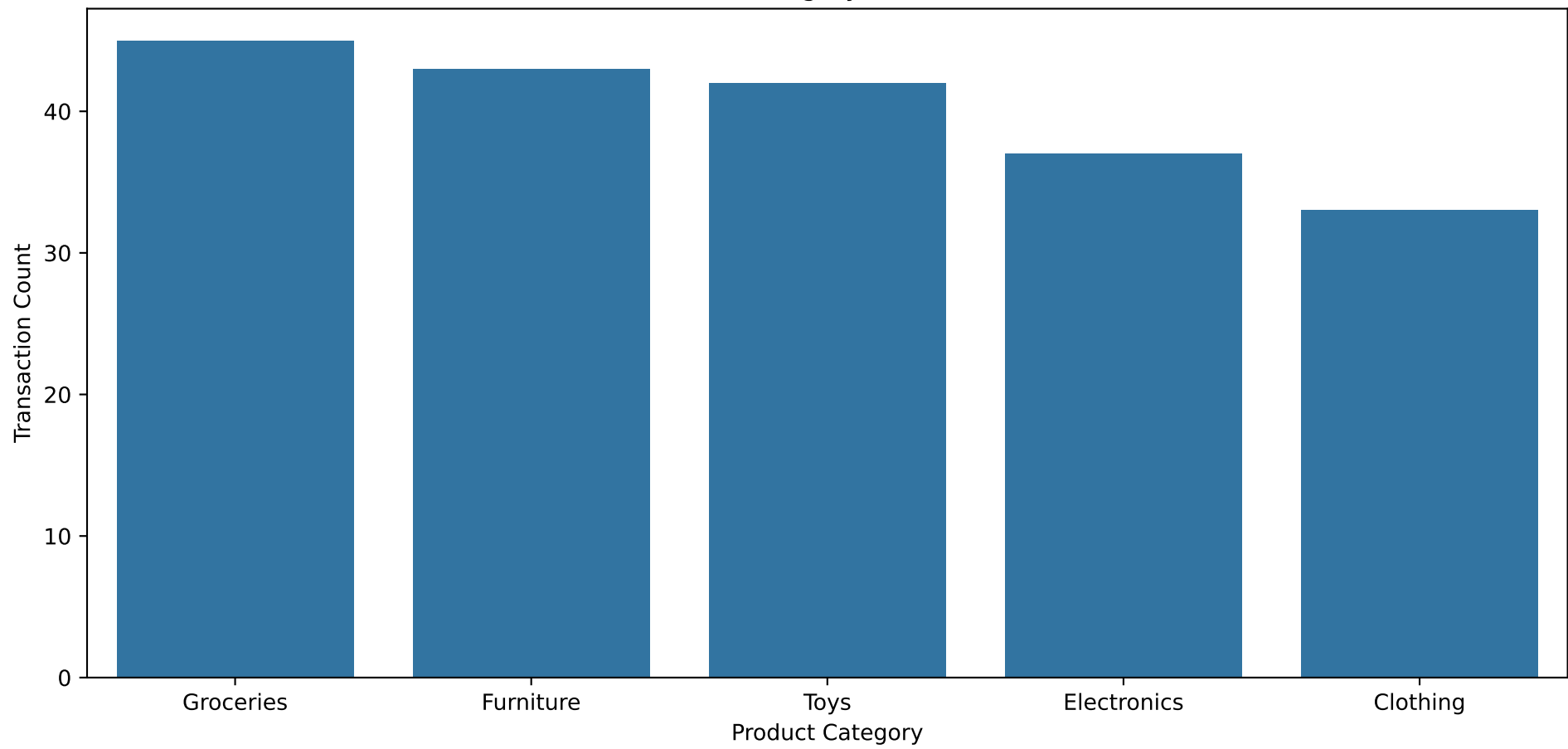
	count	mean	std	min	25%	50%	75%	max
UnitPrice	156	2791.41	1355.5	86.78	1967.55	2777.56	3954.16	4987.41
QuantitySold	165	5.84848	2.95413	1	3	6	8	10
TotalSale	171	16421.8	13023.8	392.1	4863.09	14364.4	23720.7	48541.6
CustomerAge	171	42.848	15.3094	18	30	41	56	70

Missing values:

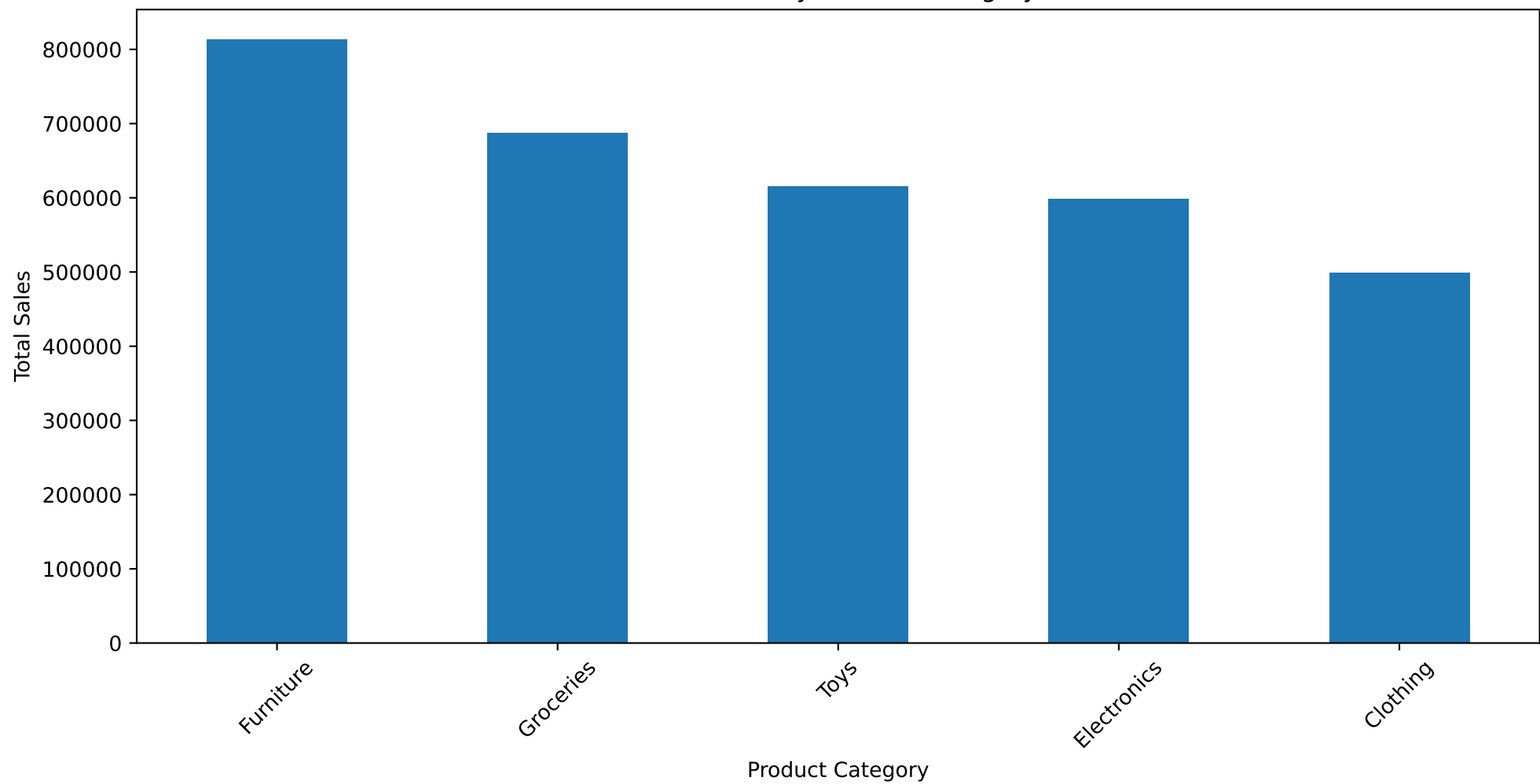
	Missing Values	Missing %
ProductCategory	37	18.5
StoreLocation	38	19
UnitPrice	44	22
QuantitySold	35	17.5
TotalSale	29	14.5
PaymentMethod	31	15.5
CustomerAge	29	14.5
Returned	36	18

# PRODUCT CATEGORY ANALYSIS

Product Category Distribution

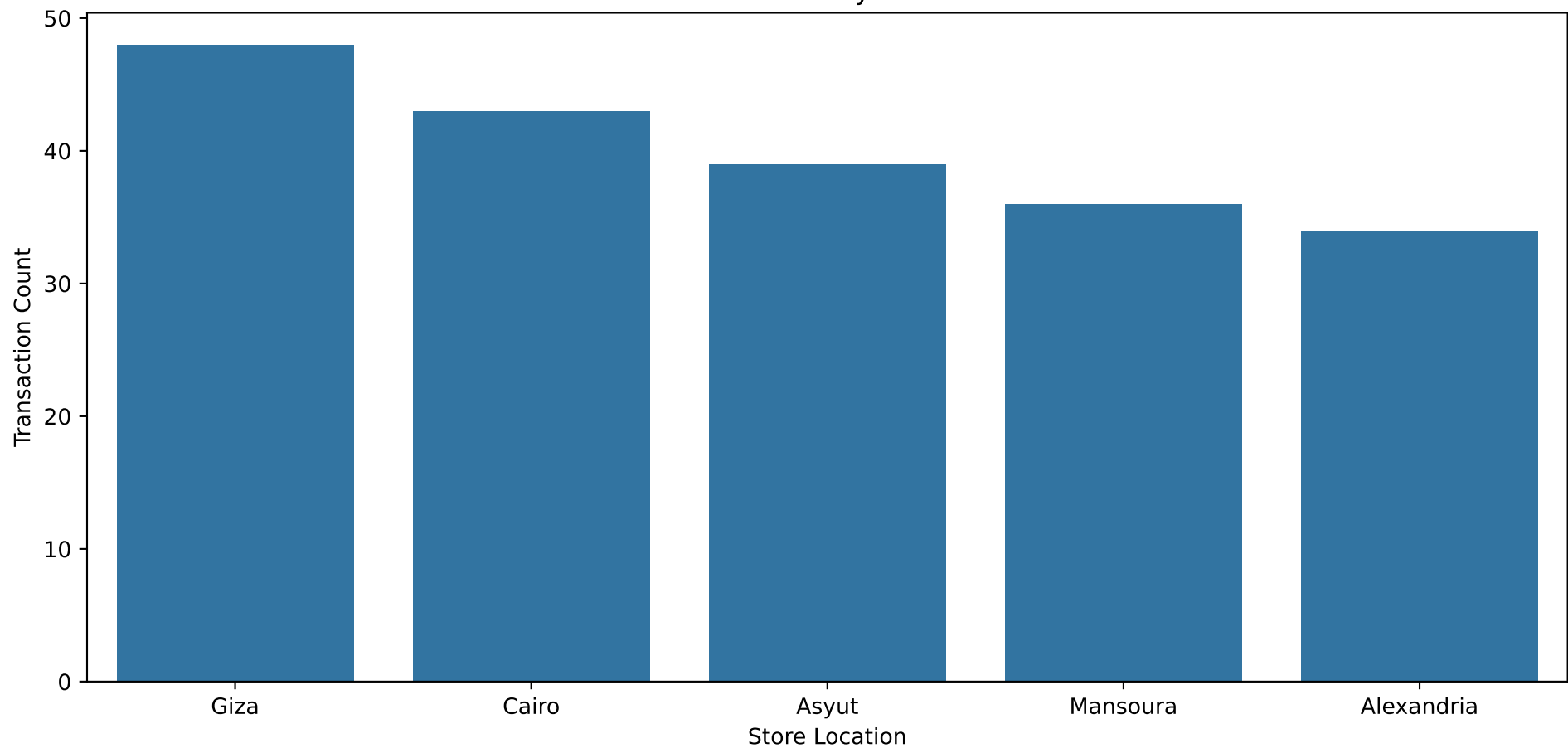


Total Sales by Product Category

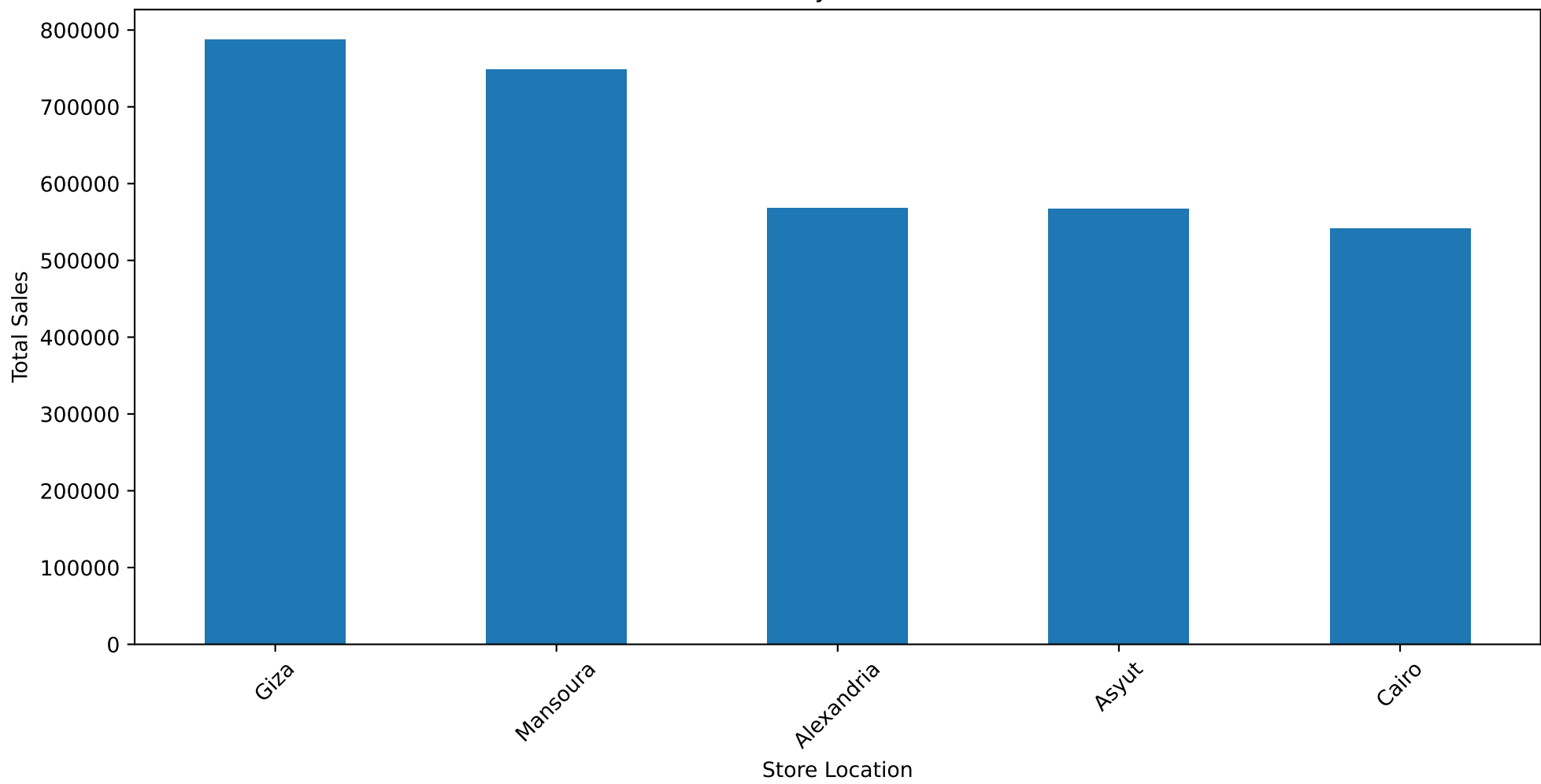


# STORE LOCATION ANALYSIS

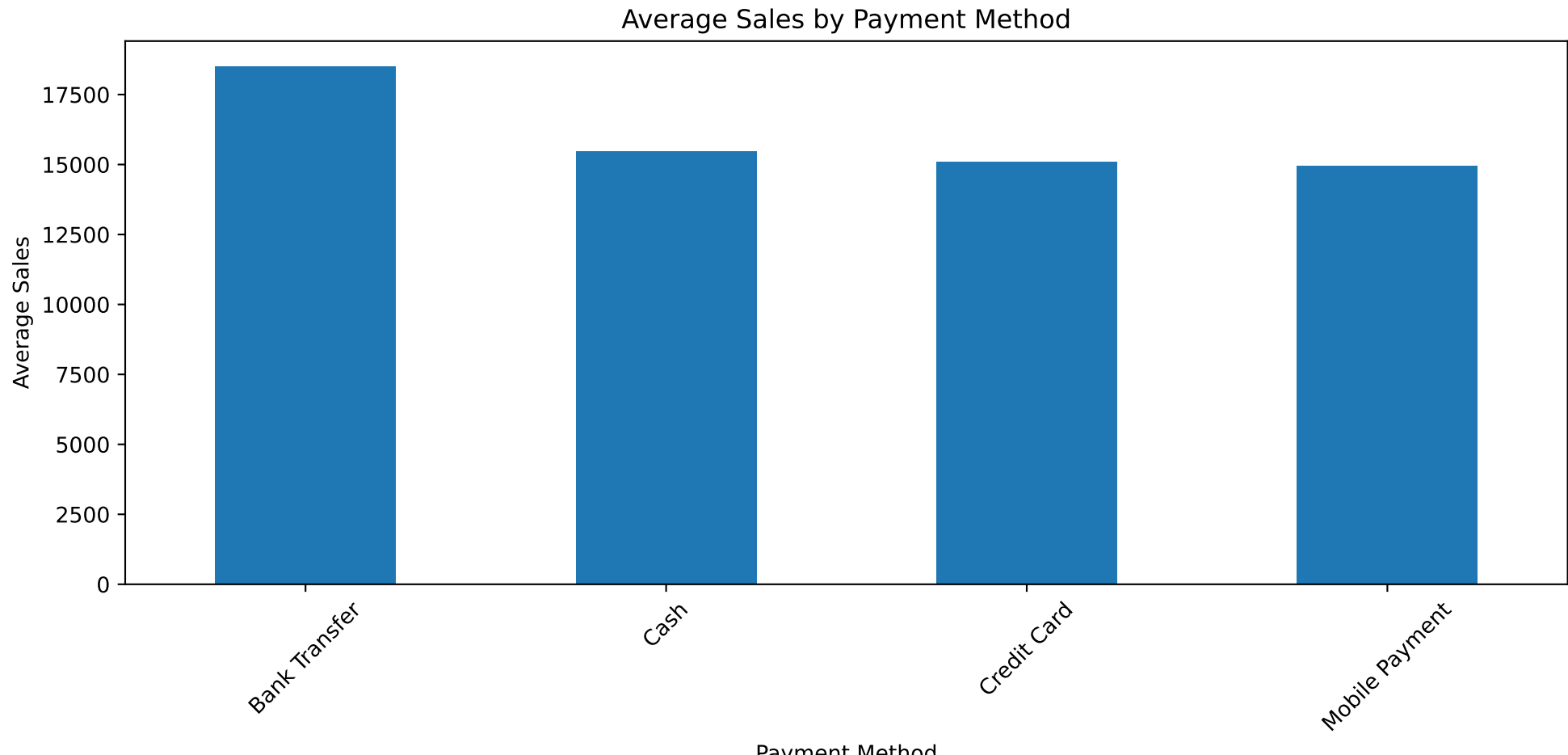
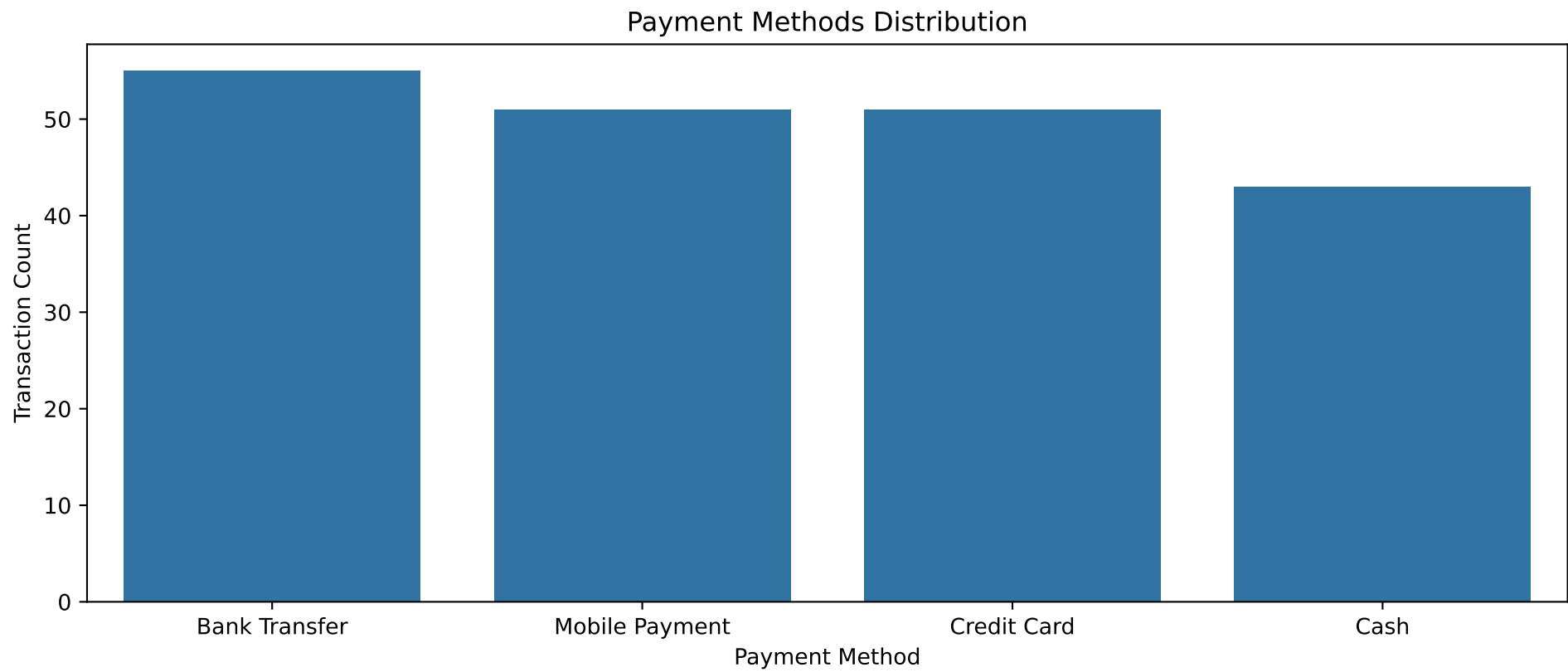
Transactions by Location



Total Sales by Store Location

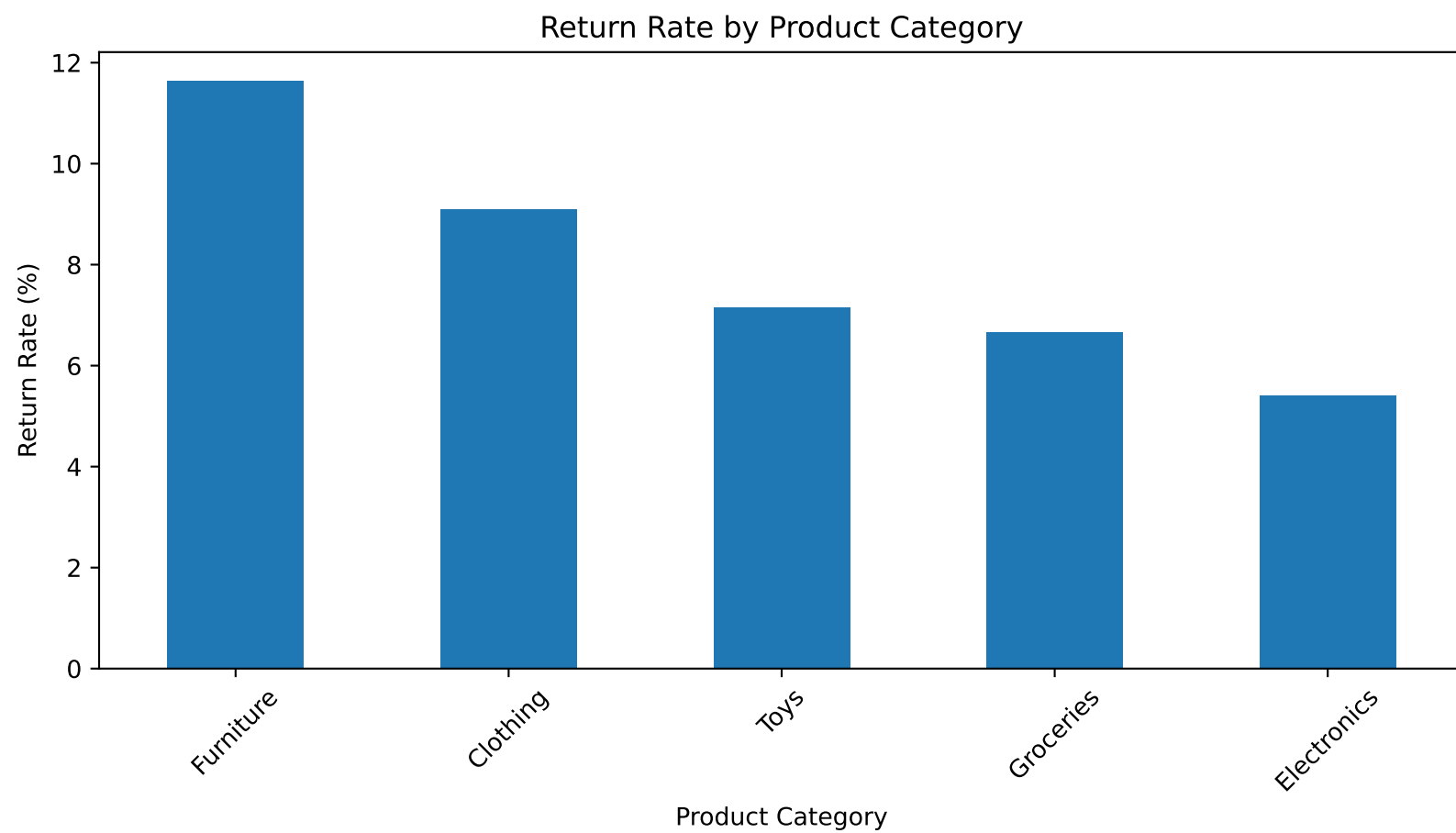
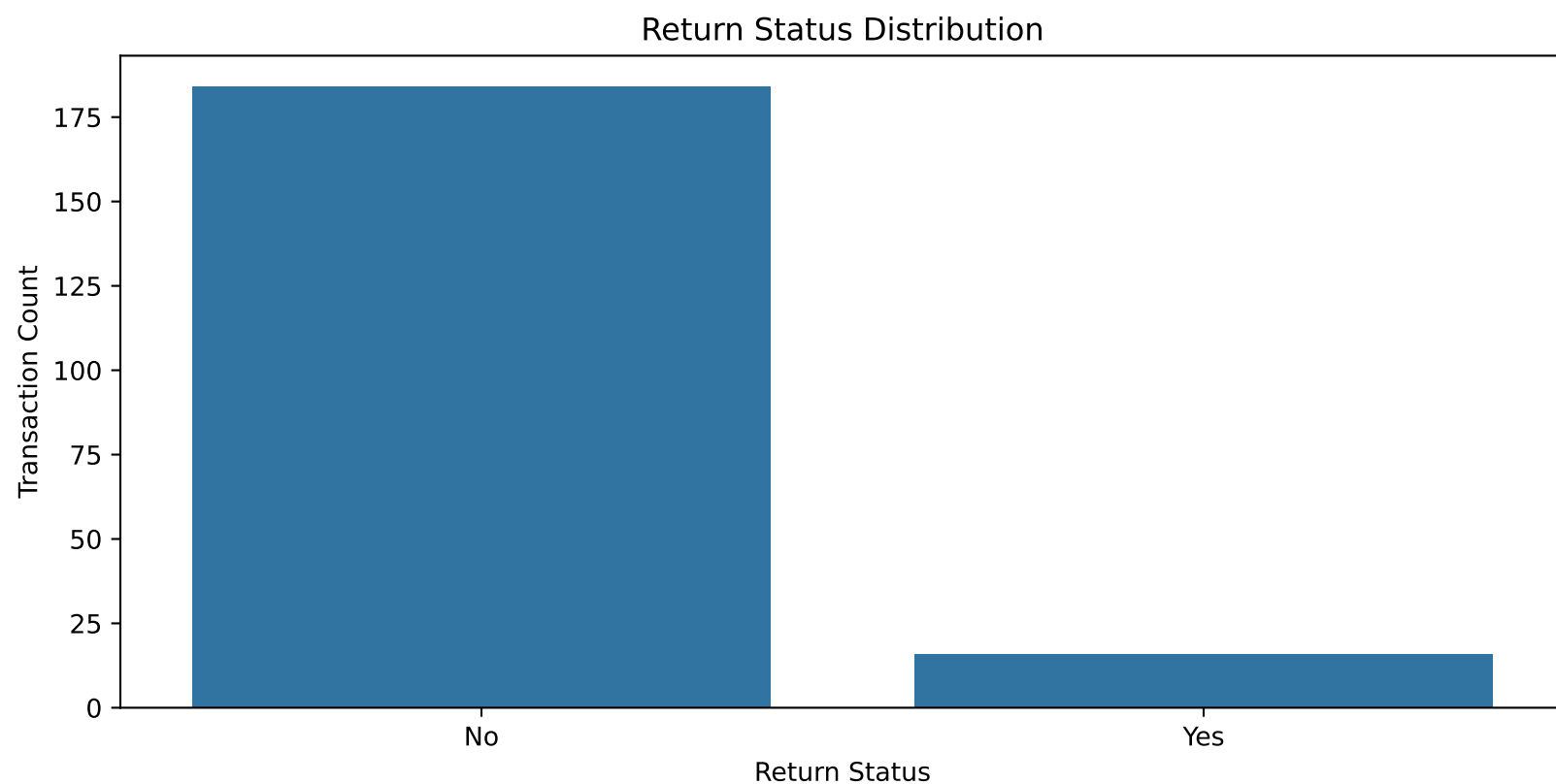


## PAYMENT METHOD ANALYSIS



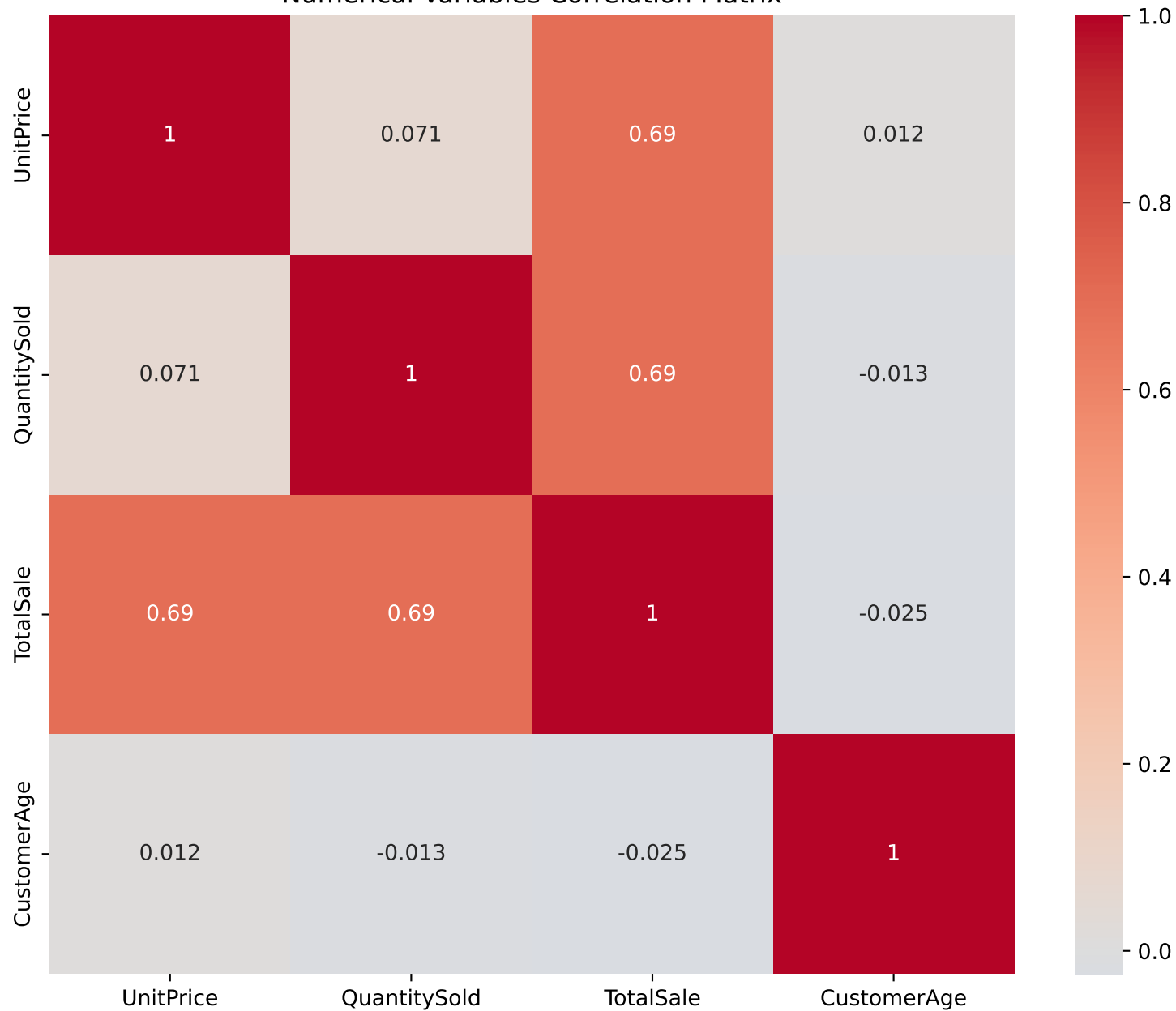


## RETURN RATE ANALYSIS

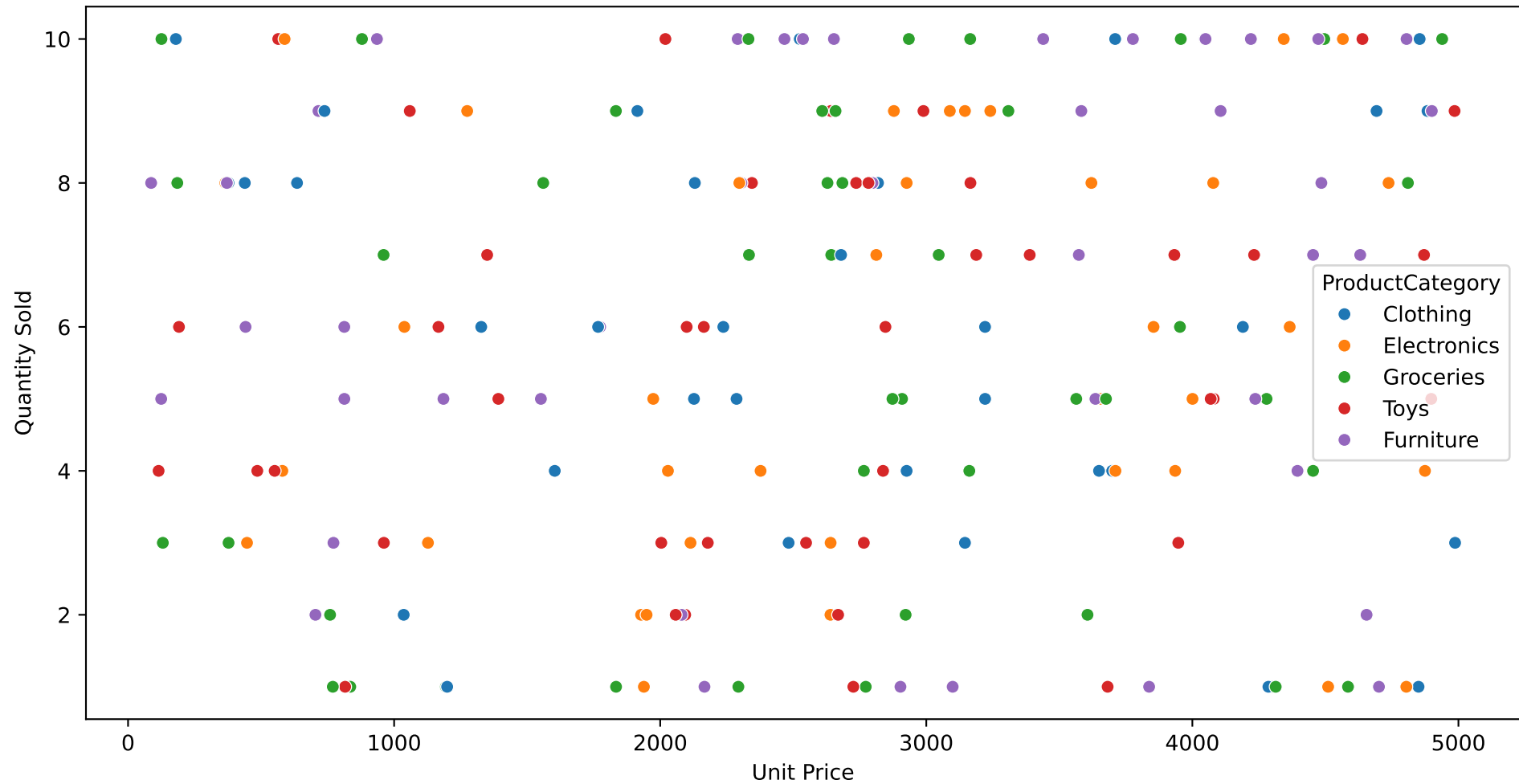


## VARIABLE RELATIONSHIPS ANALYSIS

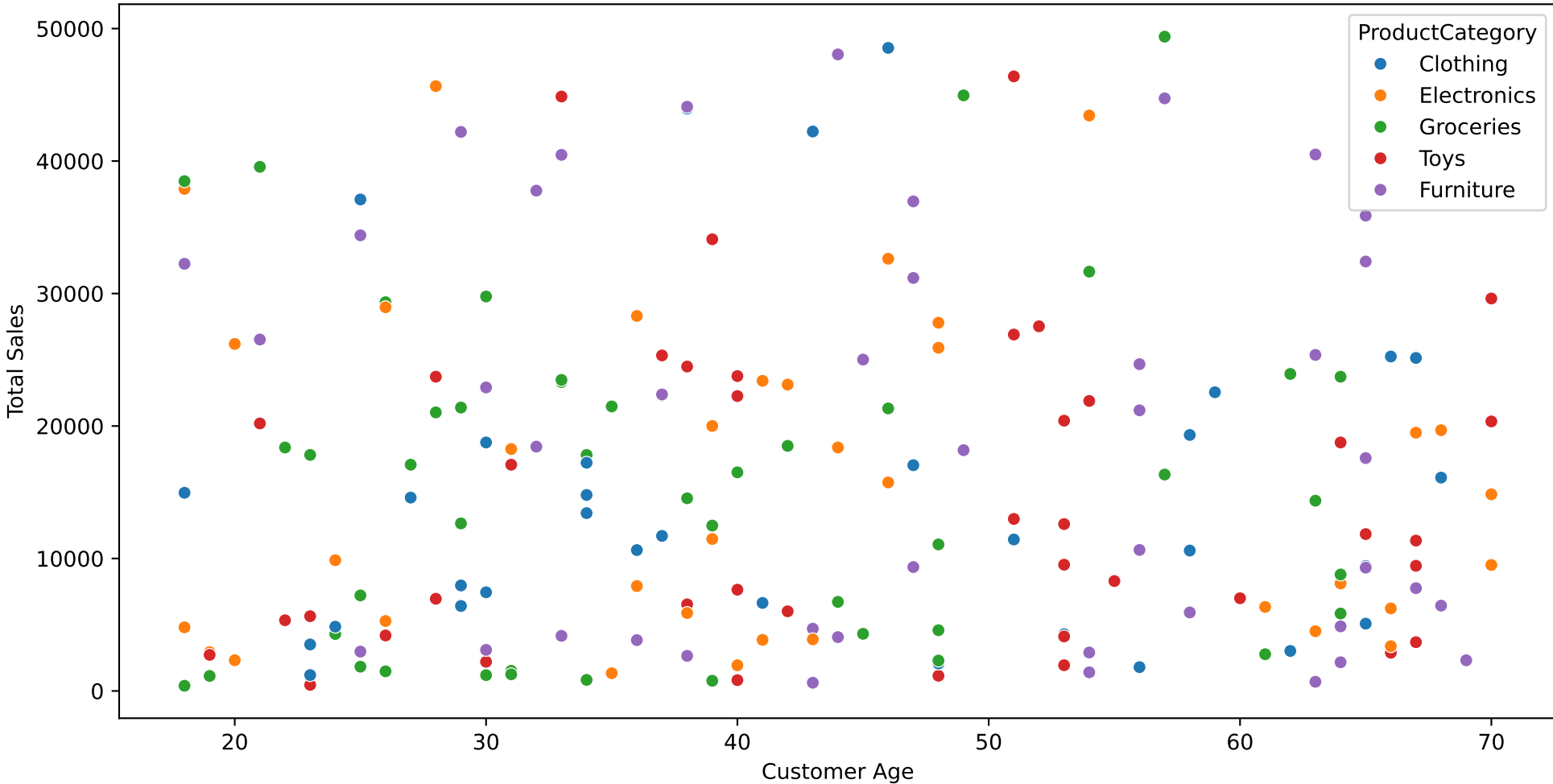
Numerical Variables Correlation Matrix



Unit Price vs Quantity Sold



## Customer Age vs Total Sales



# HANDLING MISSING DATA

# Missing Values After Imputation:

Missing Values	
TransactionID	0
ProductCategory	0
StoreLocation	0
UnitPrice	0
QuantitySold	0
TotalSale	0
PaymentMethod	0
CustomerAge	0
Returned	0



## ORIGINAL VS CLEANED DATA COMPARISON

## Numerical Columns Comparison:

### Column: UnitPrice

	Original Data	Cleaned Data
count	200	200
mean	2681.79	2791.41
std	1399.26	1196.3
min	86.78	86.78
25%	1725.84	2175.72
50%	2750.91	2791.41
75%	3873.68	3650
max	4987.41	4987.41

### Column: QuantitySold

	Original Data	Cleaned Data
count	200	200
mean	5.88	5.84848
std	3.00846	2.68179
min	1	1
25%	3	4
50%	6	5.84848
75%	9	8
max	10	10

### Column: TotalSale

	Original Data	Cleaned Data
count	200	200
mean	16067.7	16421.8
std	12978.9	12037.5
min	392.1	392.1
25%	4672.22	5990.68
50%	13203.4	16421.8
75%	23733.4	22331
max	49388.8	48541.6

### Column: CustomerAge

	Original Data	Cleaned Data
count	200	200
mean	43.25	42.848
std	15.335	14.15
min	18	18
25%	30	31.75
50%	41.5	42.848
75%	56	53.25
max	70	70

## PREPARING DATA FOR KNN ALGORITHM

Data Sample After Encoding:

	TransactionID	UnitPrice	QuantitySold	TotalSale	CustomerAge	ProductCategory_Electronics	ProductCategory_Furniture	ProductCategory_Groceries	ProductCategory_Toys	StoreLocation_Asyut	StoreLocation_Cairo	StoreLocation_Giza	StoreLocation_Mansoura	PaymentMethod_Cash	PaymentMethod_Credit Card	PaymentMethod_Mobile Payment	Returned_Yes
	0	T0001	1603.67	4	6414.68	29	False	False	False	False	True	False	False	True	False		
	1	T0002	3698.71	4	14794.8	34	False	False	False	False	True	False	True	False	False		
	2	T0003	4189.86	6	25139.2	67	False	False	False	False	False	False	True	False	False		
	3	T0004	3088.44	9	27796	48	True	False	False	False	True	True	False	False	False		
	4	T0005	4495.7	10	44957	49	False	False	True	False	False	False	True	False	False		

## PREPARING DATA FOR KNN ALGORITHM

Data Sample After Encoding:

	TransactionID	UnitPrice	QuantitySold	TotalSale	CustomerAge	ProductCategory_Electronics	ProductCategory_Furniture	ProductCategory_Groceries	ProductCategory_Toys	StoreLocation_Asyut	StoreLocation_Cairo	StoreLocation_Giza	StoreLocation_Mansoura	PaymentMethod_Cash	PaymentMethod_Credit Card	PaymentMethod_Mobile Payment	Returned_Yes
	0	T0001	1603.67	4	6414.68	42.848	False	False	False	False	True	False	False	True	False		
	1	T0002	2791.41	5.84848	14794.8	34	False	False	False	False	True	False	True	False	False		
	2	T0003	4189.86	6	25139.2	67	False	False	False	False	False	True	False	False	False		
	3	T0004	2791.41	9	16421.8	42.848	True	False	False	True	False	True	False	False	False		
	4	T0005	4495.7	10	44957	49	False	False	True	False	False	True	False	True	False		