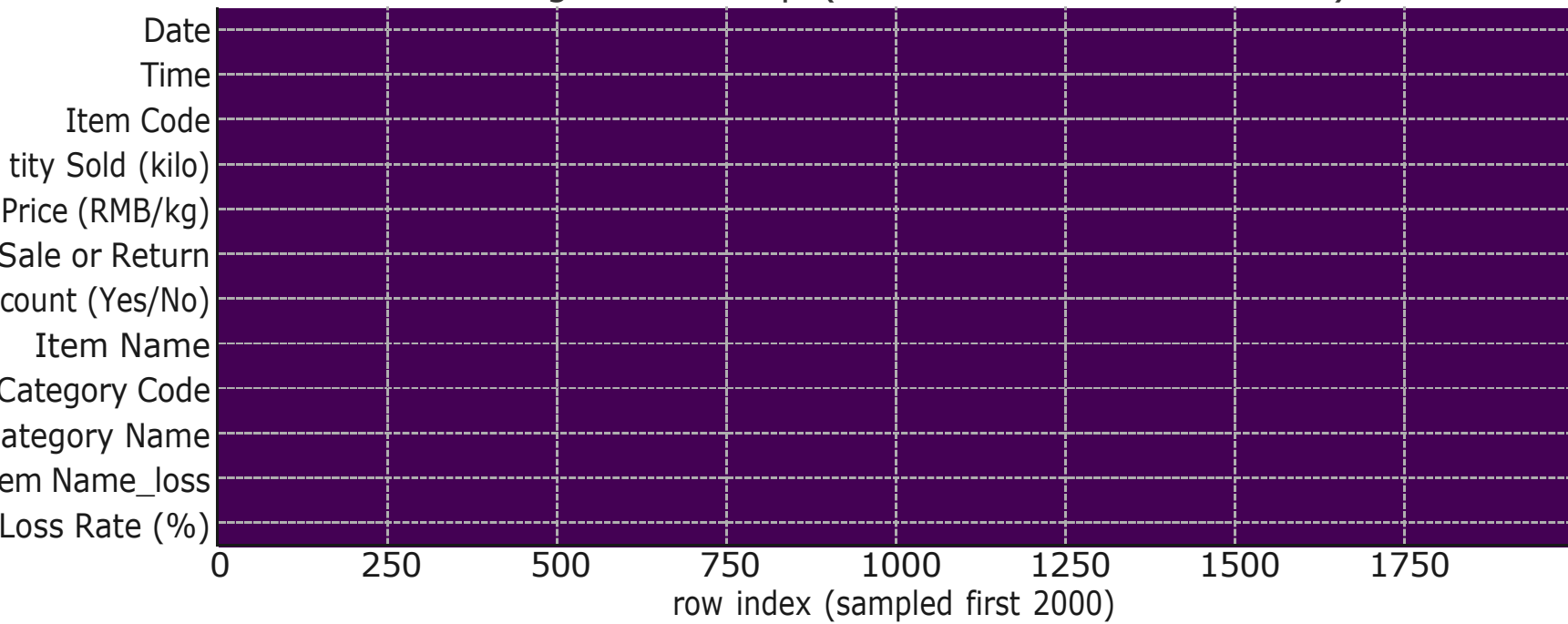
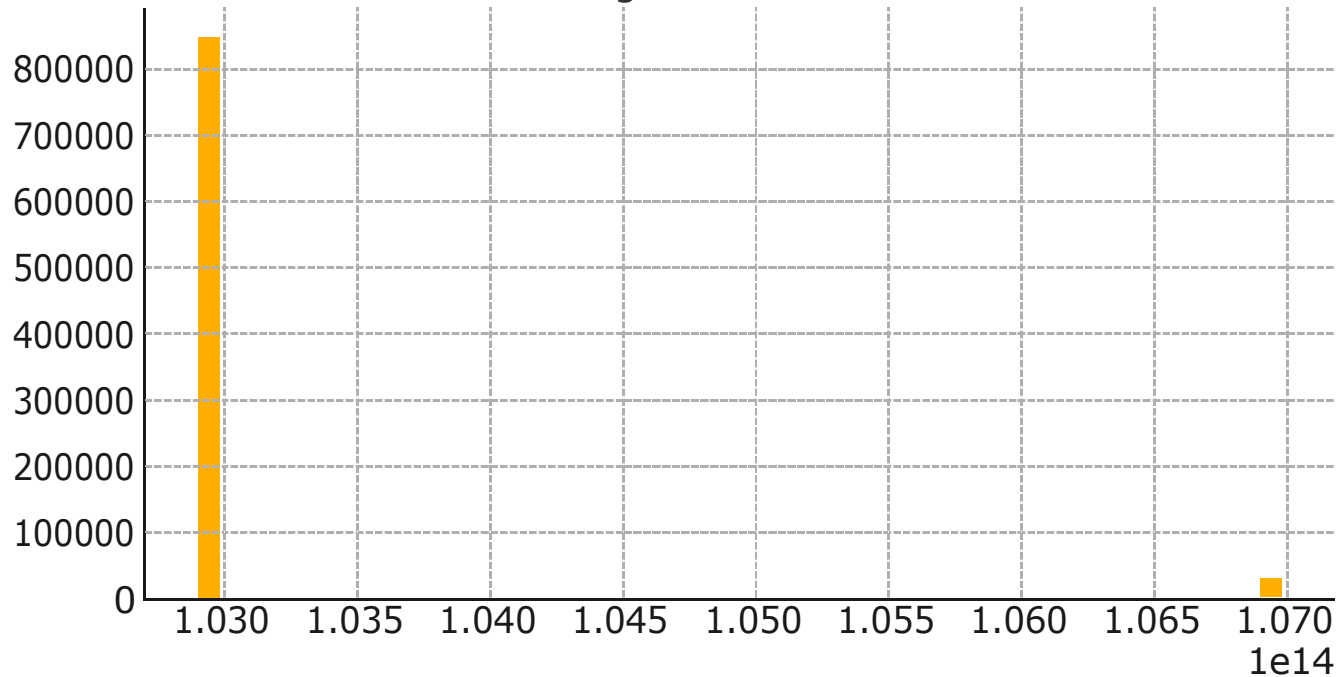


## **EDA Report\_Merged Dataset**

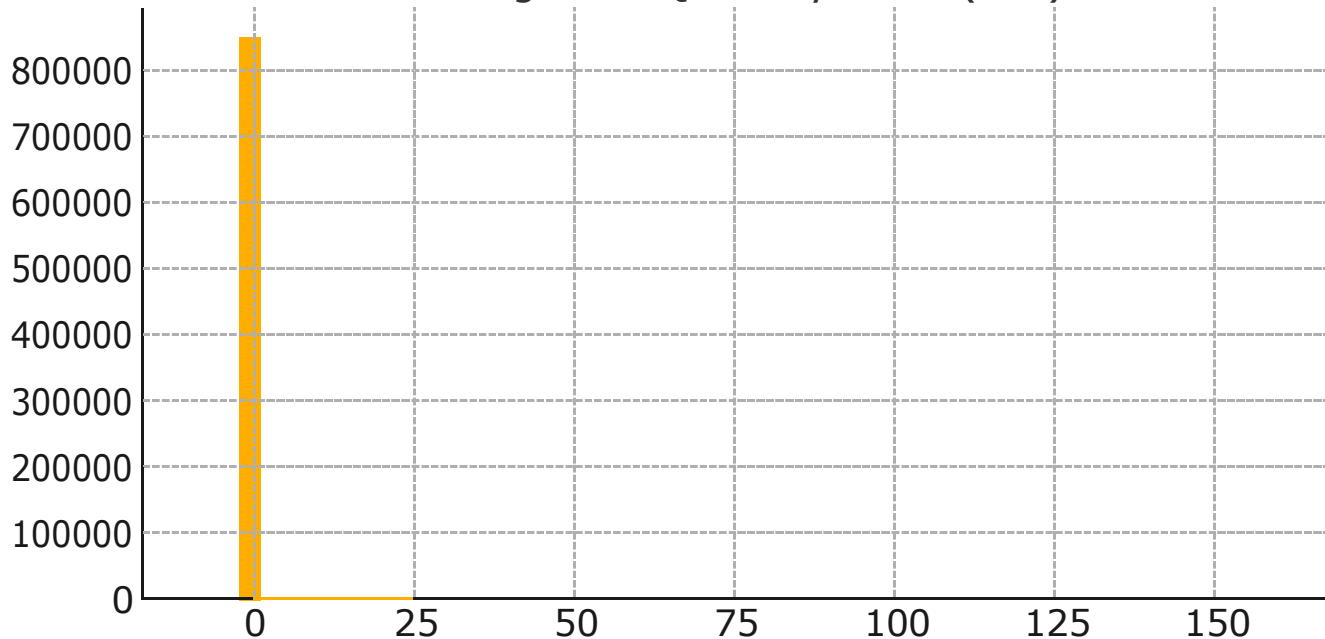
Missing values map (columns x first 2000 rows)



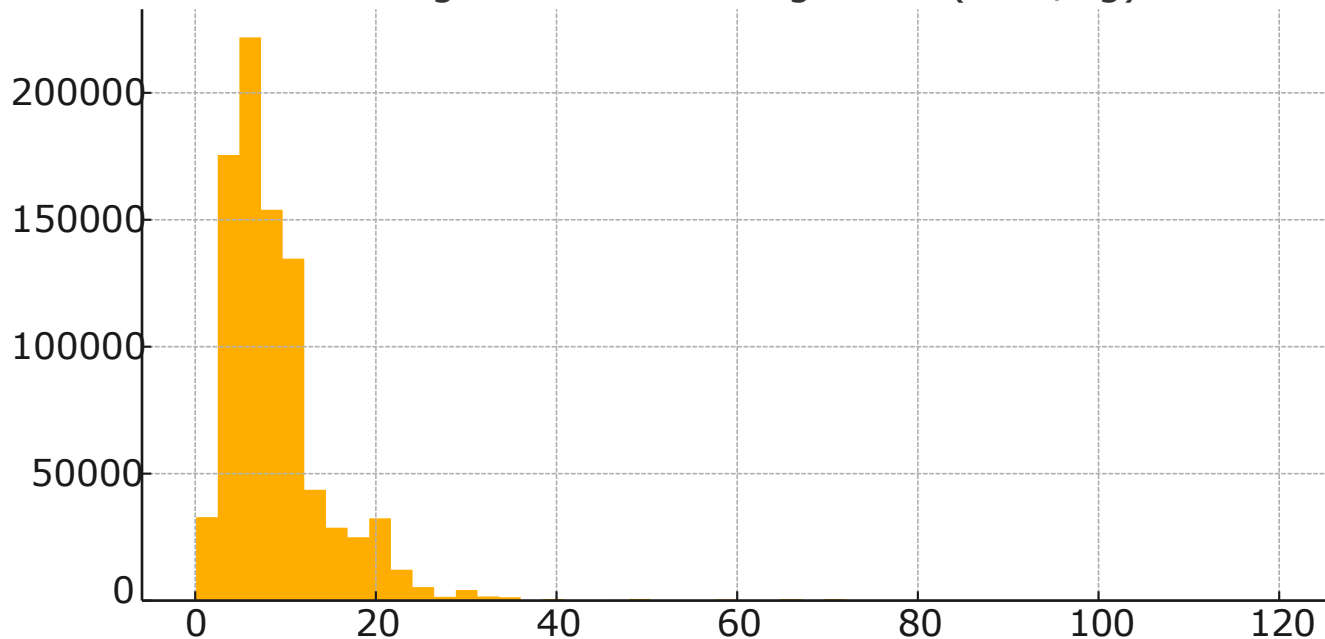
# Histogram: Item Code



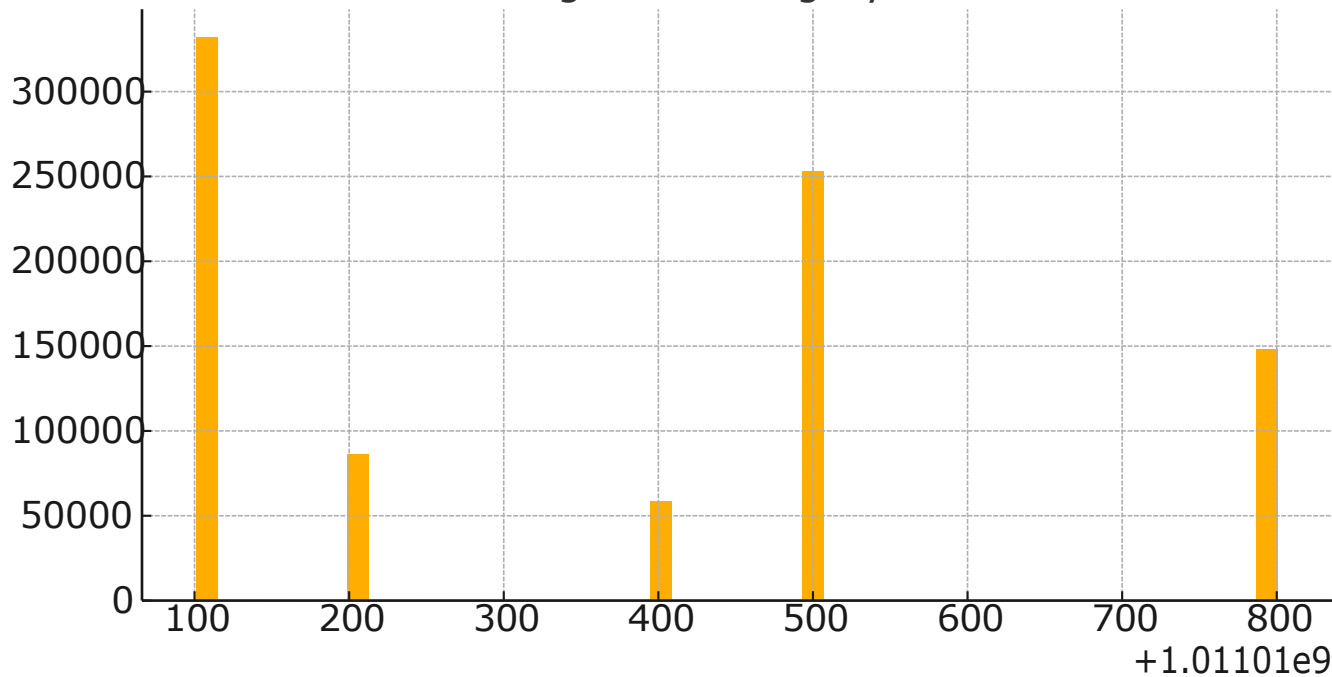
Histogram: Quantity Sold (kilo)



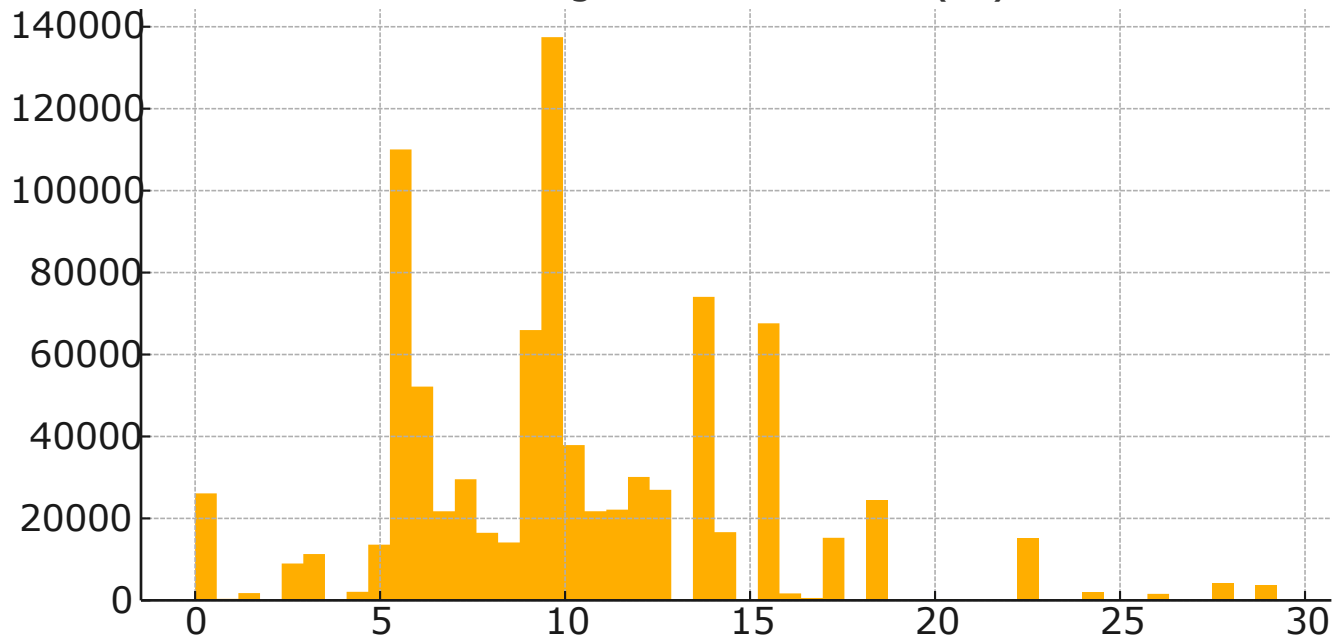
Histogram: Unit Selling Price (RMB/kg)



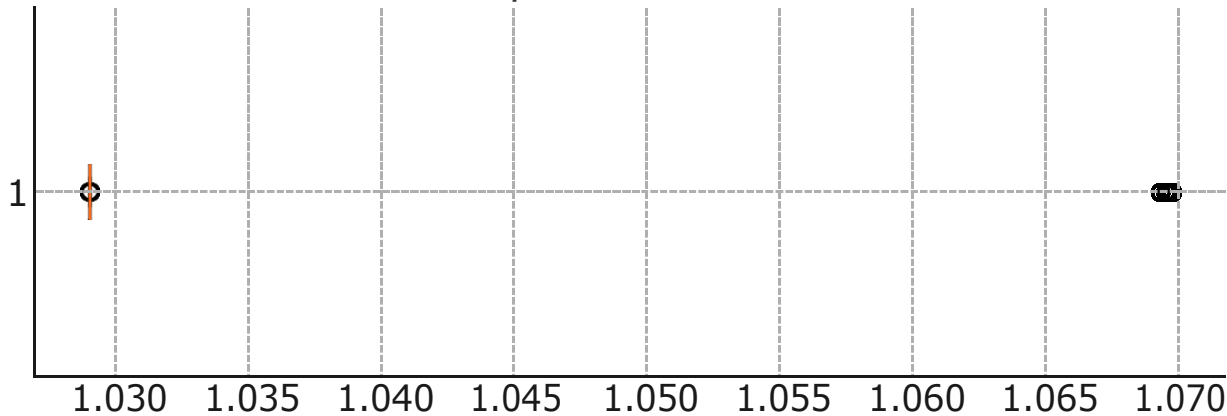
# Histogram: Category Code



# Histogram: Loss Rate (%)

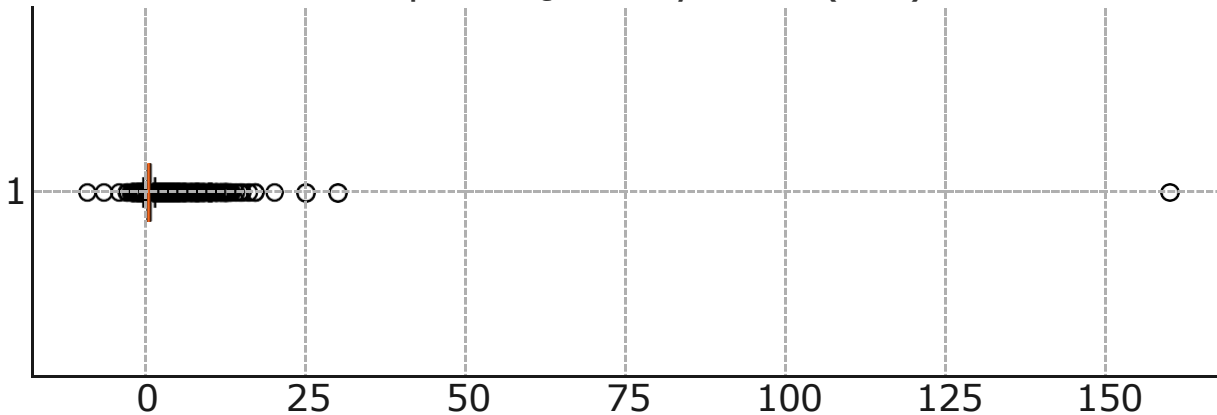


Boxplot: Item Code

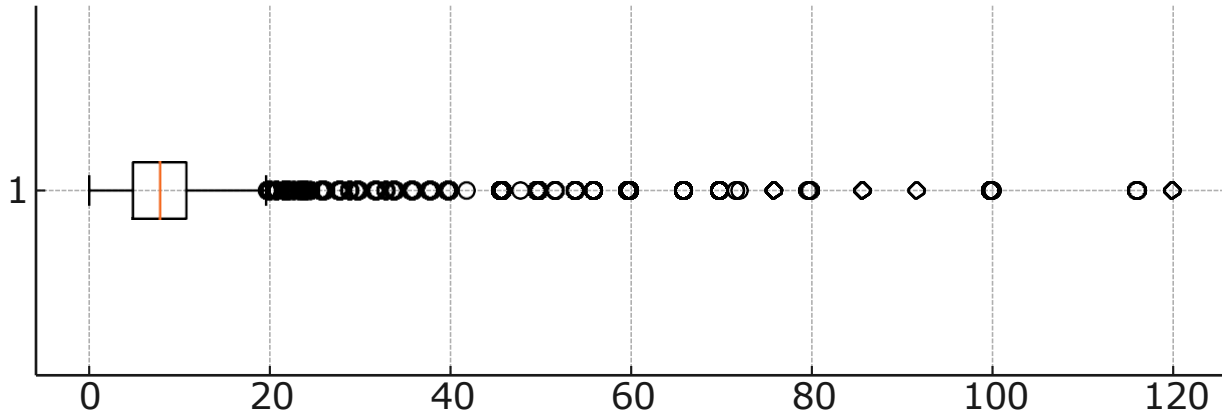




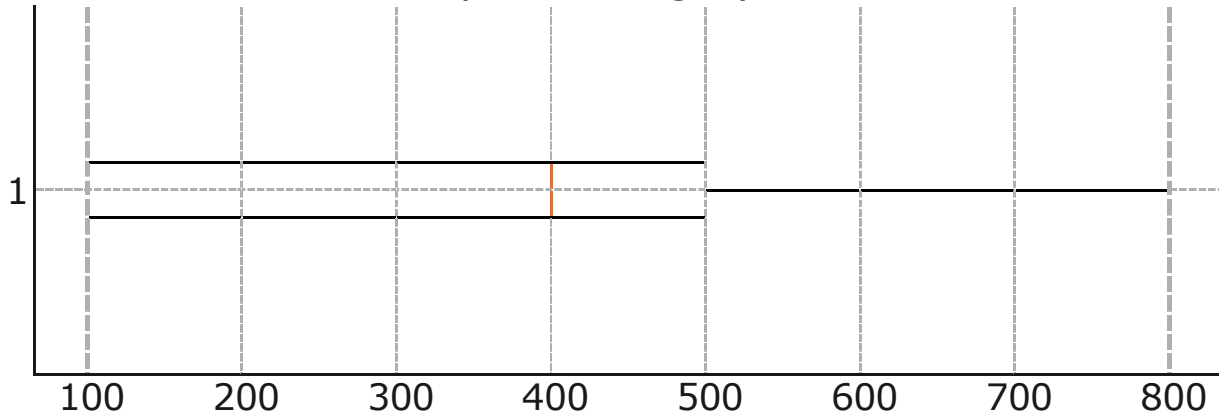
Boxplot: Quantity Sold (kilo)



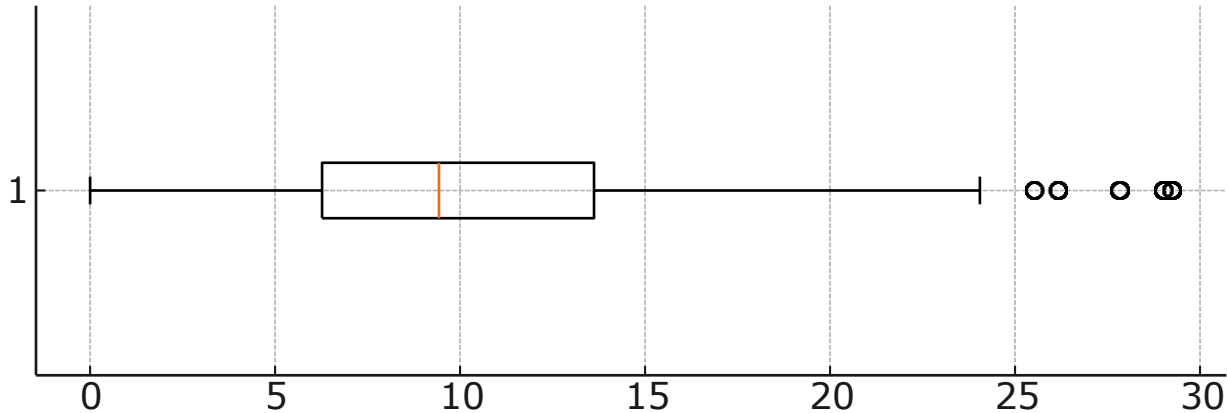
Boxplot: Unit Selling Price (RMB/kg)



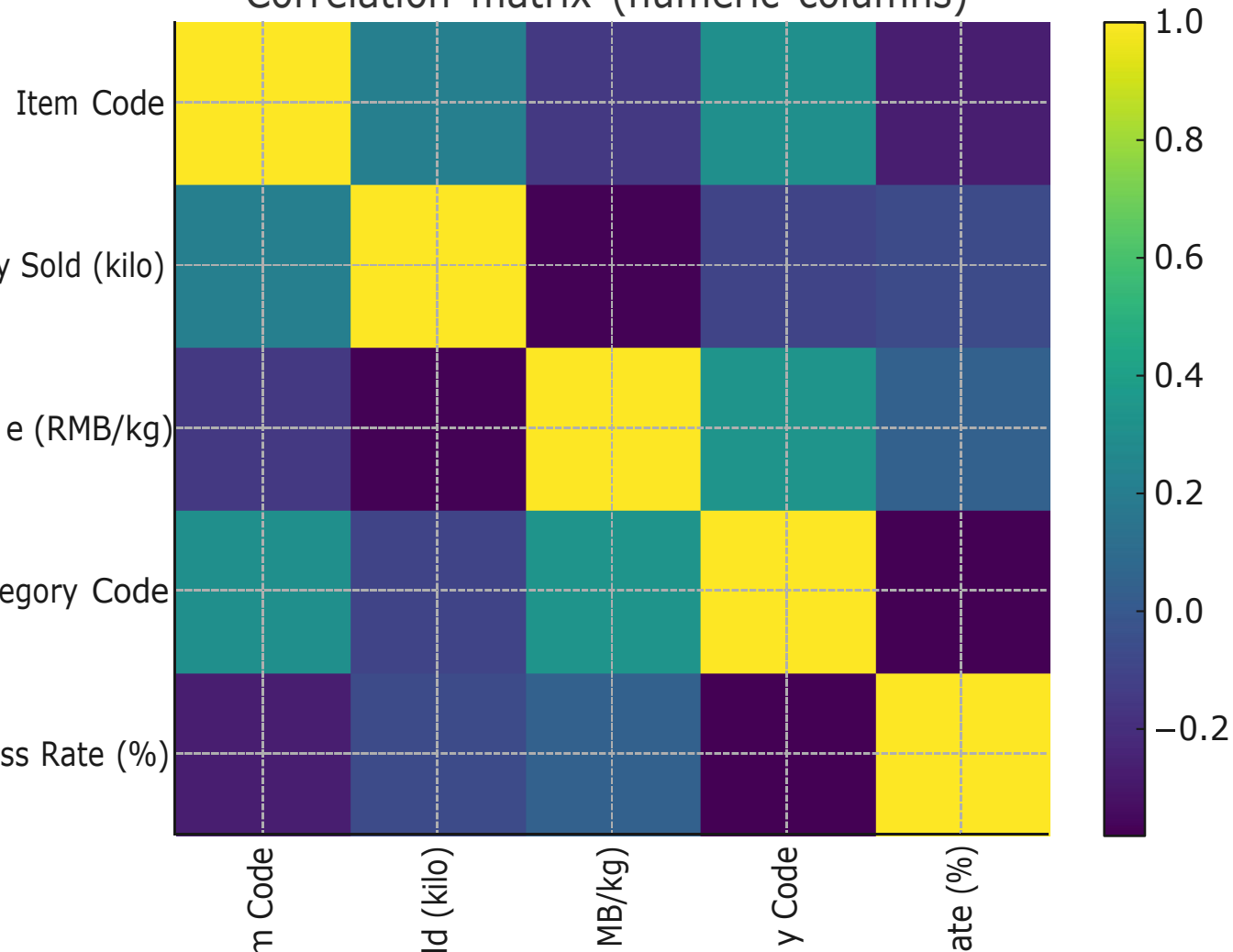
Boxplot: Category Code

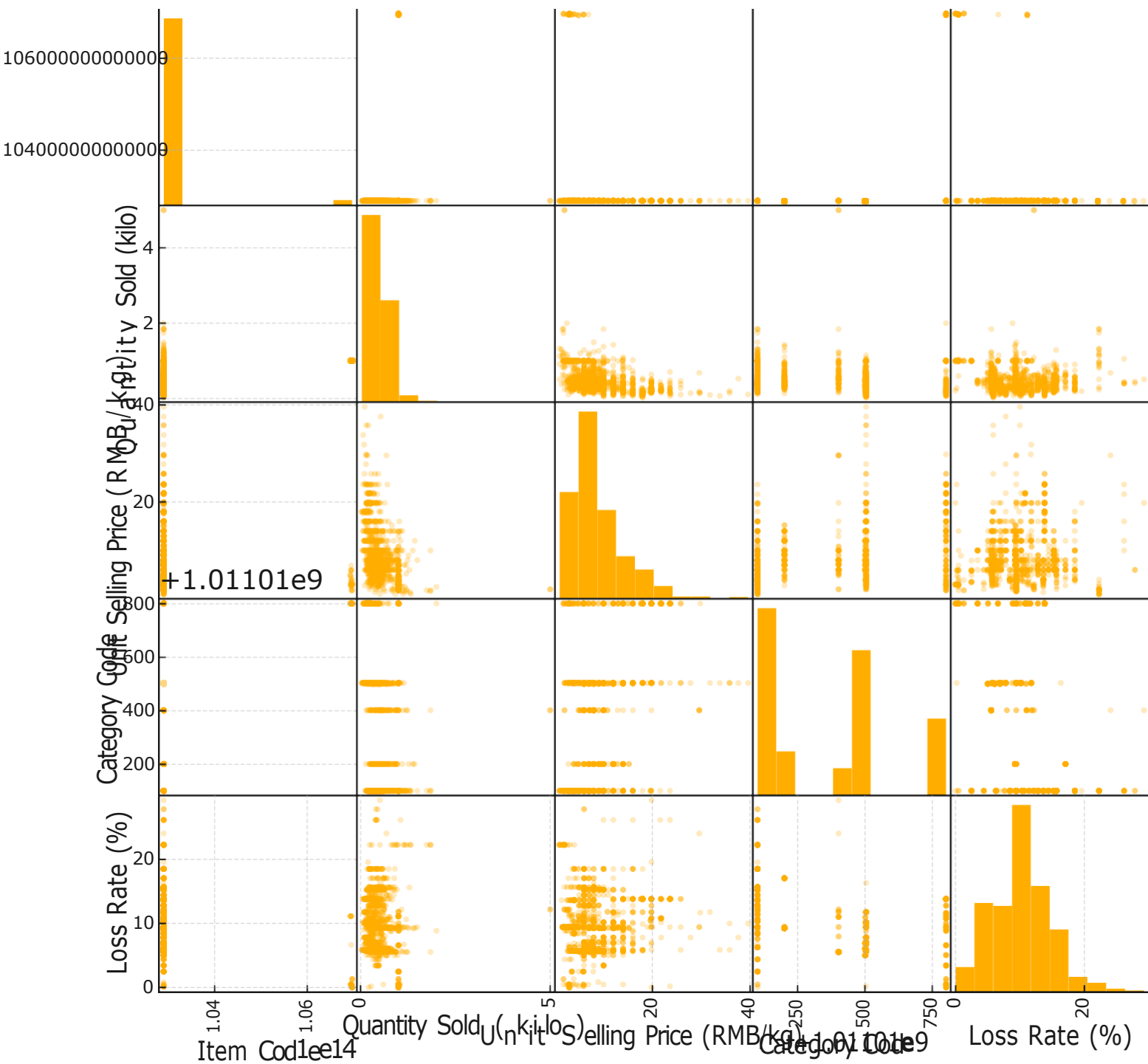


Boxplot: Loss Rate (%)

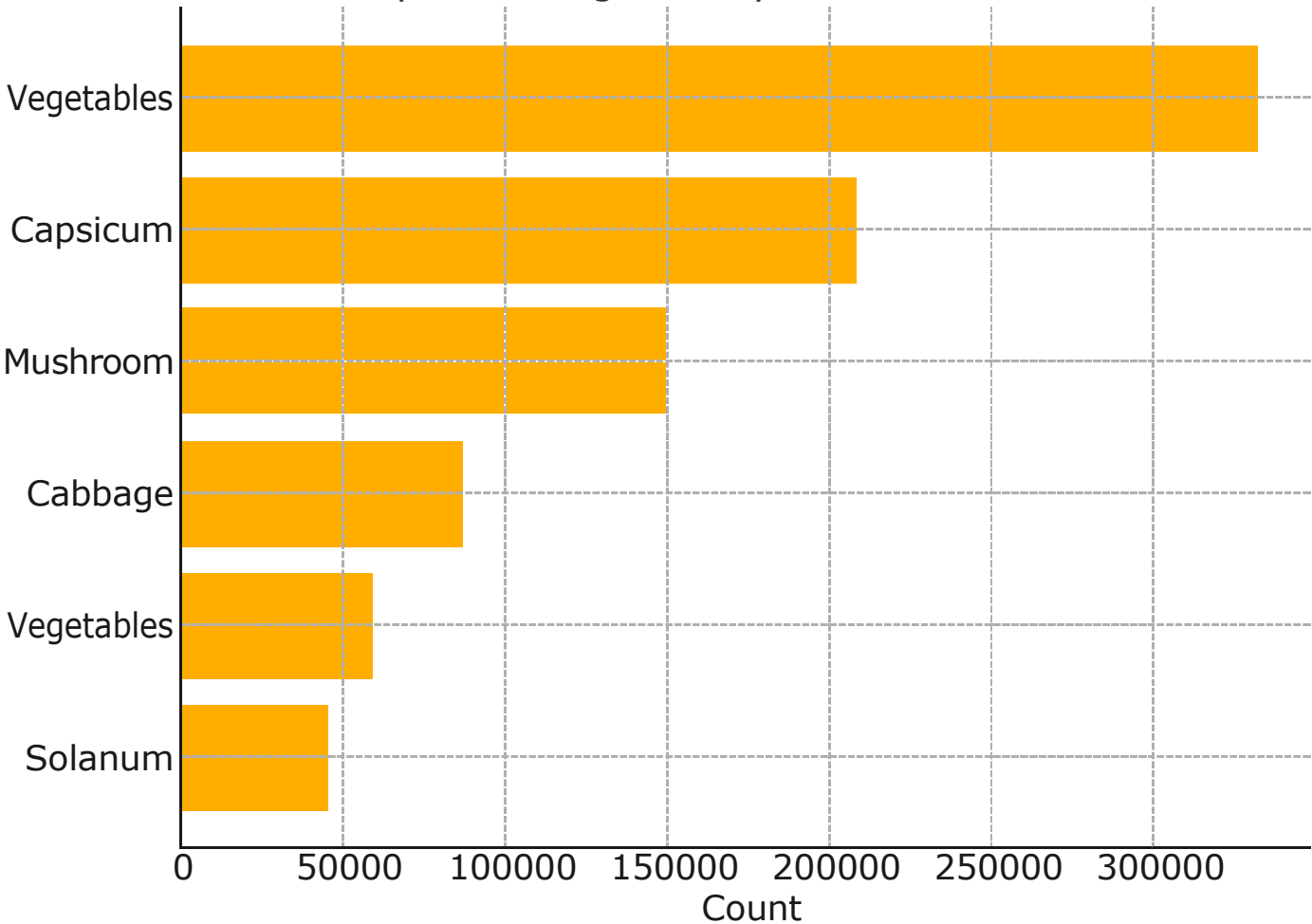


Correlation matrix (numeric columns)

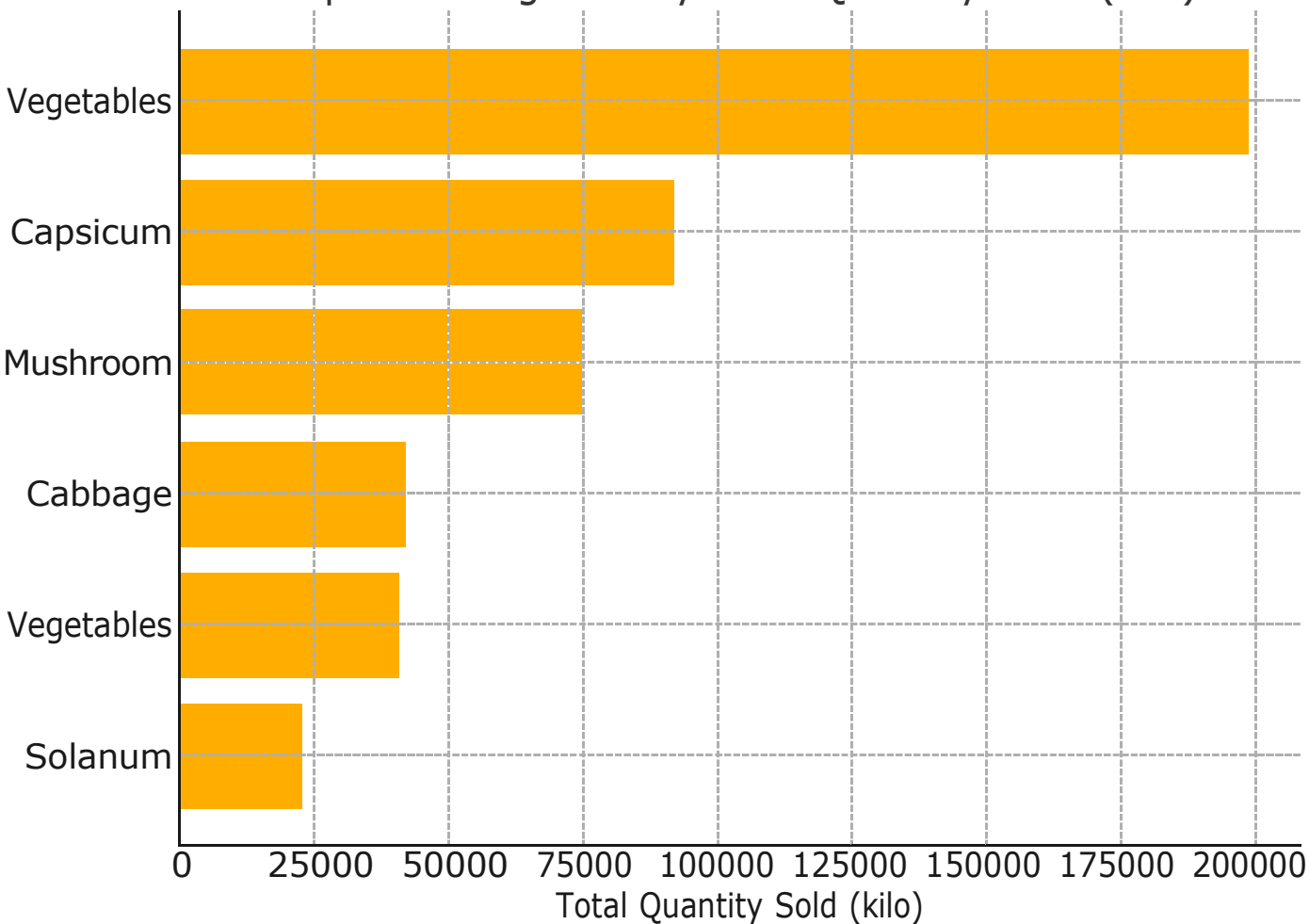




Top 20 Categories by transaction count



Top 20 Categories by total Quantity Sold (kilo)





# Outliers detected (z>|3|) in column: Item Code

	Date	Time	Item Code	Quantity Sold (kilo)	Unit Selling Price (RMB/kg)	Sale or Return	Discount (Yes/No)	
9178	2020-07-10	09:50:55.311	106956146480203	1.0	3.5	sale	No	The Crab
9662	2020-07-10	17:09:39.740	106956146480203	1.0	3.5	sale	No	The Crab
9768	2020-07-10	18:08:47.436	106956146480203	1.0	3.5	sale	No	The Crab
9858	2020-07-10	18:43:36.757	106956146480203	1.0	3.5	sale	No	The Crab
12449	2020-07-12	15:59:37.086	106956146480197	1.0	3.5	sale	No	The
12569	2020-07-12	17:02:24.967	106956146480197	1.0	3.5	sale	No	The
12571	2020-07-12	17:02:34.498	106956146480197	1.0	3.5	sale	No	The
12574	2020-07-12	17:02:42.030	106956146480197	1.0	3.5	sale	No	The
14260	2020-07-14	15:00:40.451	106956146480197	1.0	3.5	sale	No	The
14261	2020-07-14	15:00:59.389	106956146480197	1.0	3.5	sale	No	The
14672	2020-07-14	19:04:27.545	106956146480197	1.0	3.5	sale	No	The
14825	2020-07-14	21:38:42.610	106956146480197	1.0	3.5	sale	No	The
15237	2020-07-15	14:52:59.282	106956146480197	1.0	3.5	sale	No	The
15323	2020-07-15	15:54:17.389	106956146480197	1.0	3.5	sale	No	The
15402	2020-07-15	16:51:52.758	106956146480197	1.0	3.5	sale	No	The
16395	2020-07-16	15:48:28.452	106956146480197	1.0	3.5	sale	No	The
16431	2020-07-16	16:23:48.655	106956146480197	1.0	3.5	sale	No	The
16666	2020-07-16	18:38:45.748	106956146480197	1.0	3.5	sale	No	The
16695	2020-07-16	18:49:59.260	106956146480197	1.0	3.5	sale	No	The
16703	2020-07-16	18:55:54.063	106956146480197	1.0	3.5	sale	No	The

# Skewness of numeric columns (high->low)

Quantity Sold (kilo)	75.842577
Item Code	5.282436
Unit Selling Price (RMB/kg)	3.060216
Loss Rate (%)	0.780182
Category Code	0.427622