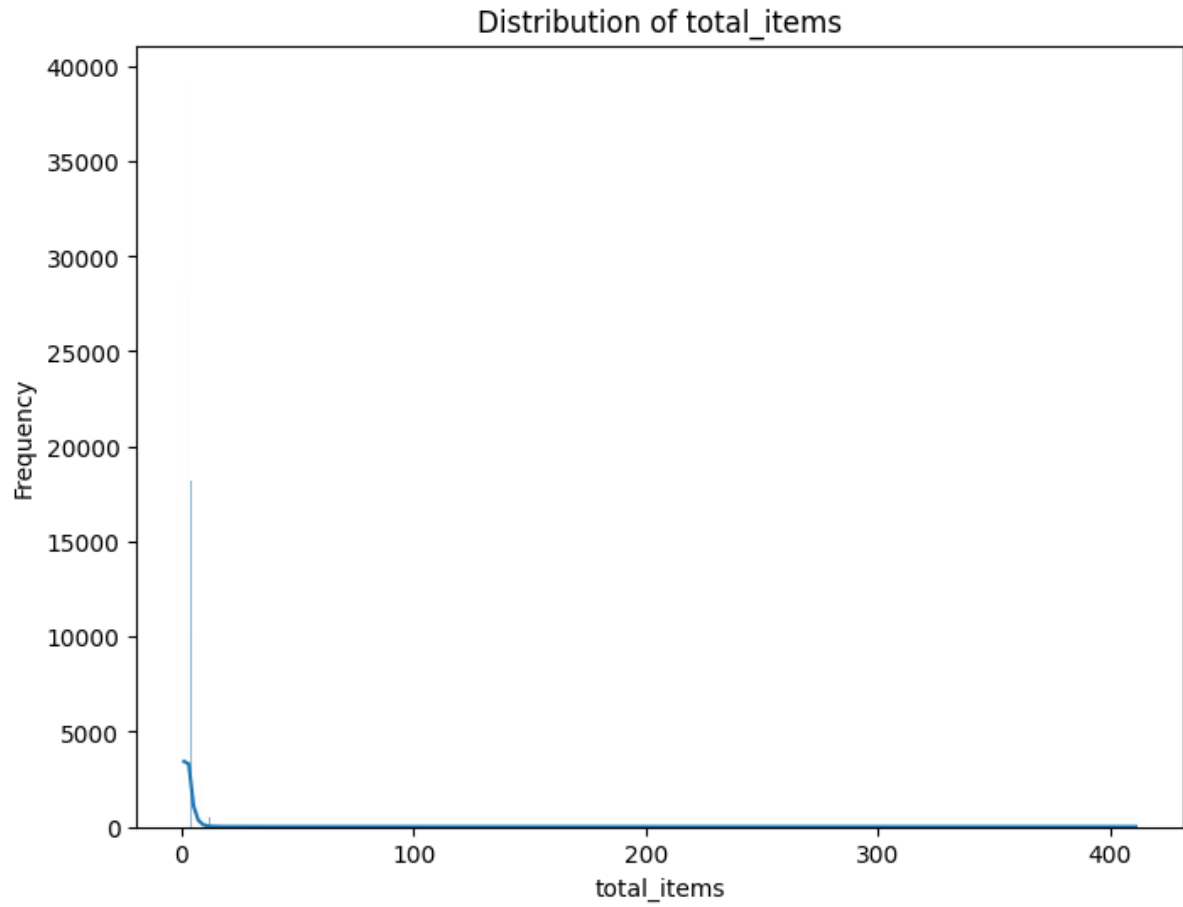


LR Delivery Time Prediction Assignment

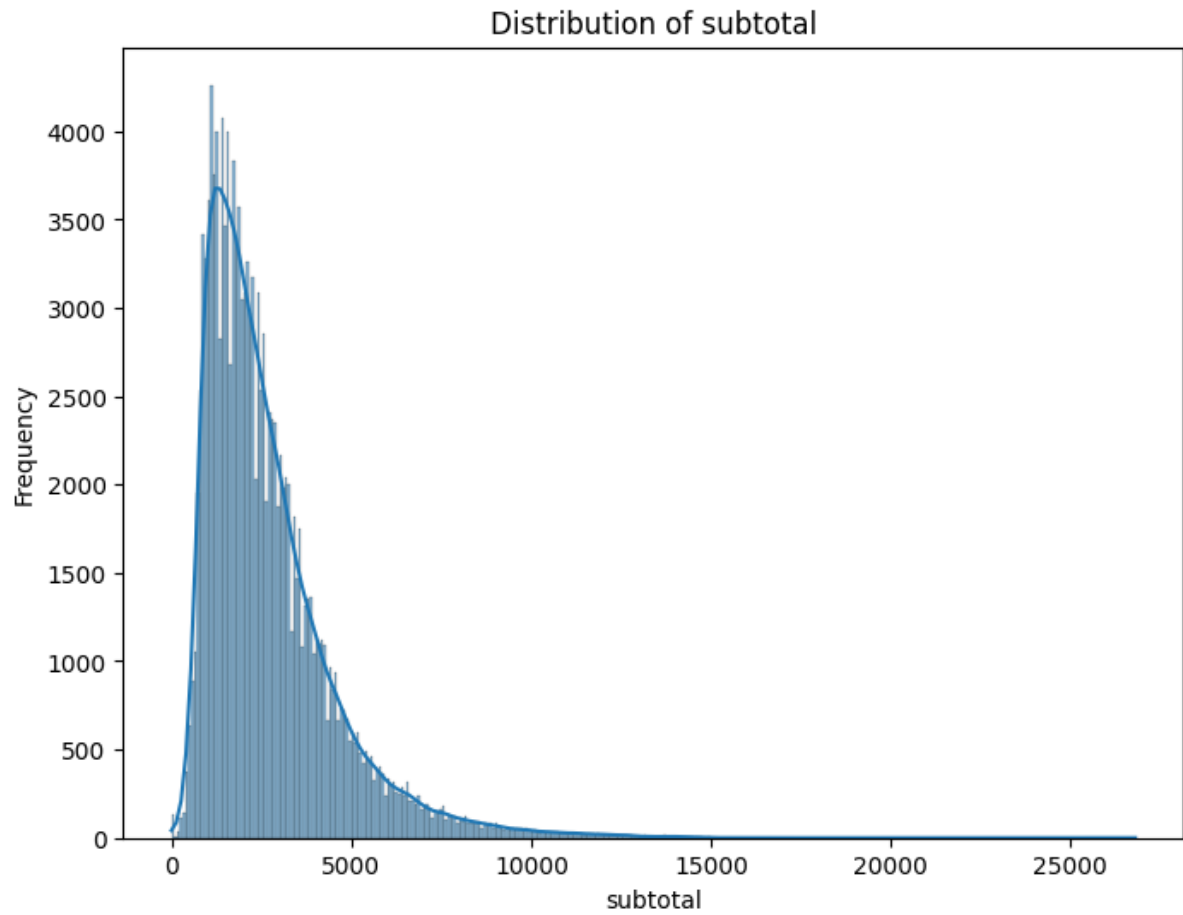
By Rishabh Uniyal

Exploratory Data Analysis on Training Data Graphs



count	140621.000000
mean	3.208582
std	2.727133
min	1.000000
25%	2.000000
50%	3.000000
75%	4.000000
max	411.000000

Name: total_items, dtype: float64



count 140621.000000

mean 2698.872530

std 1830.808391

min 0.000000

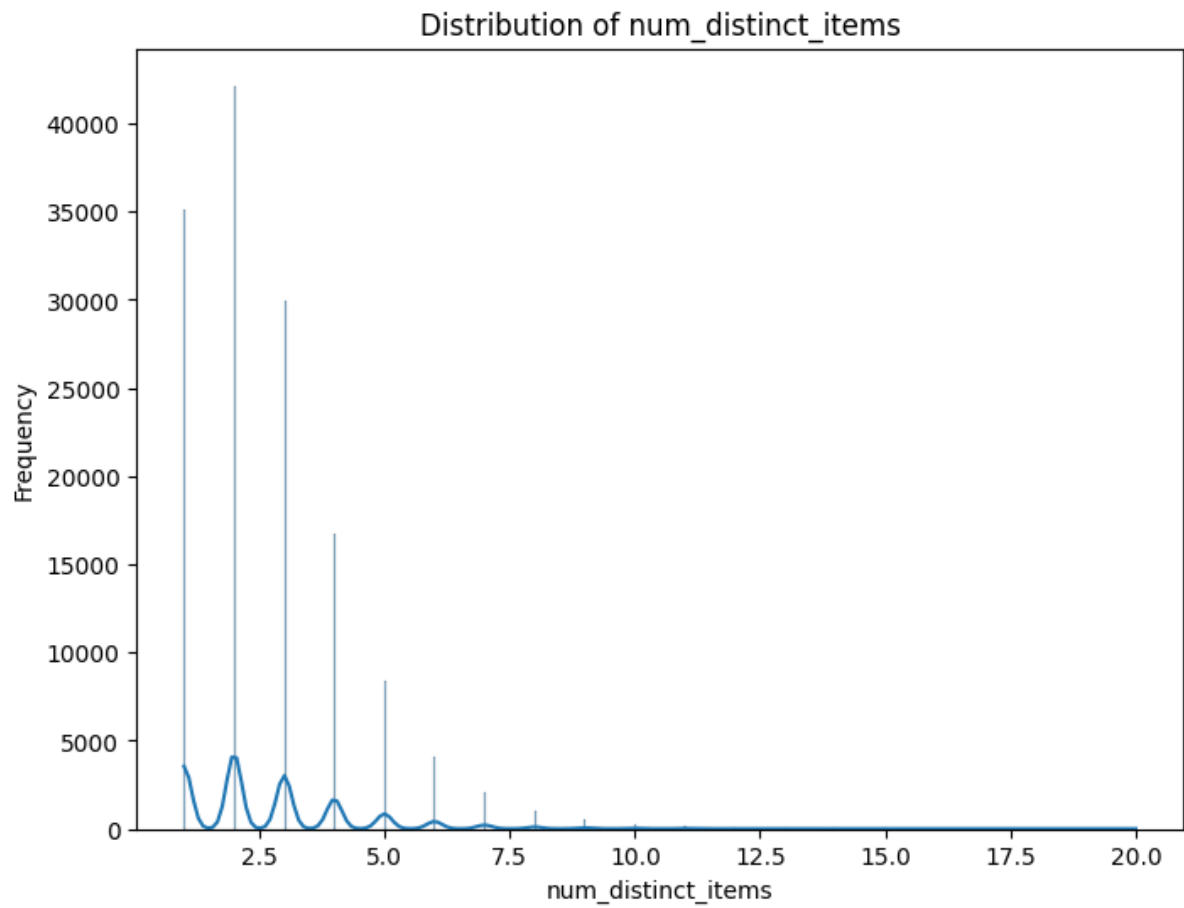
25% 1418.000000

50% 2225.000000

75% 3415.000000

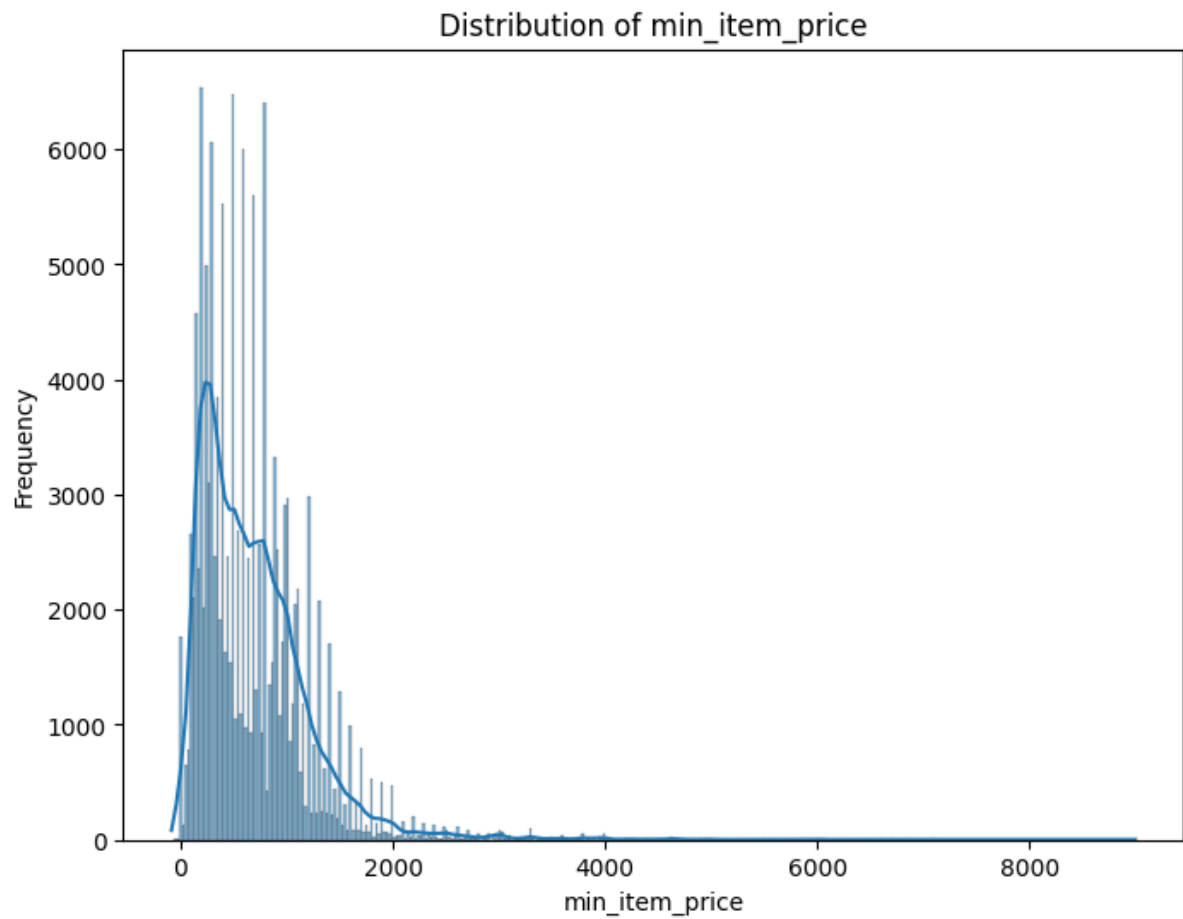
max 26800.000000

Name: subtotal, dtype: float64



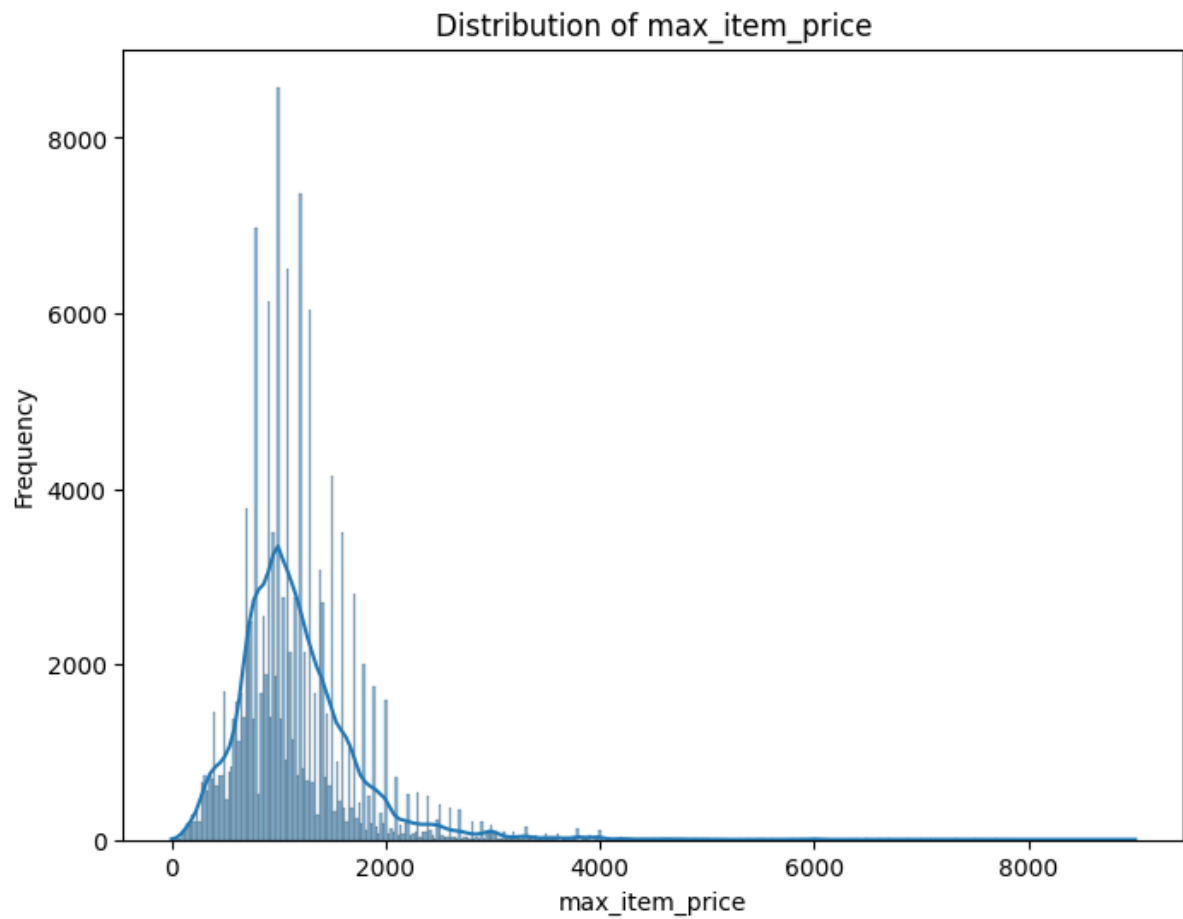
count	140621.000000
mean	2.676464
std	1.626023
min	1.000000
25%	2.000000
50%	2.000000
75%	3.000000
max	20.000000

Name: num_distinct_items, dtype: float64



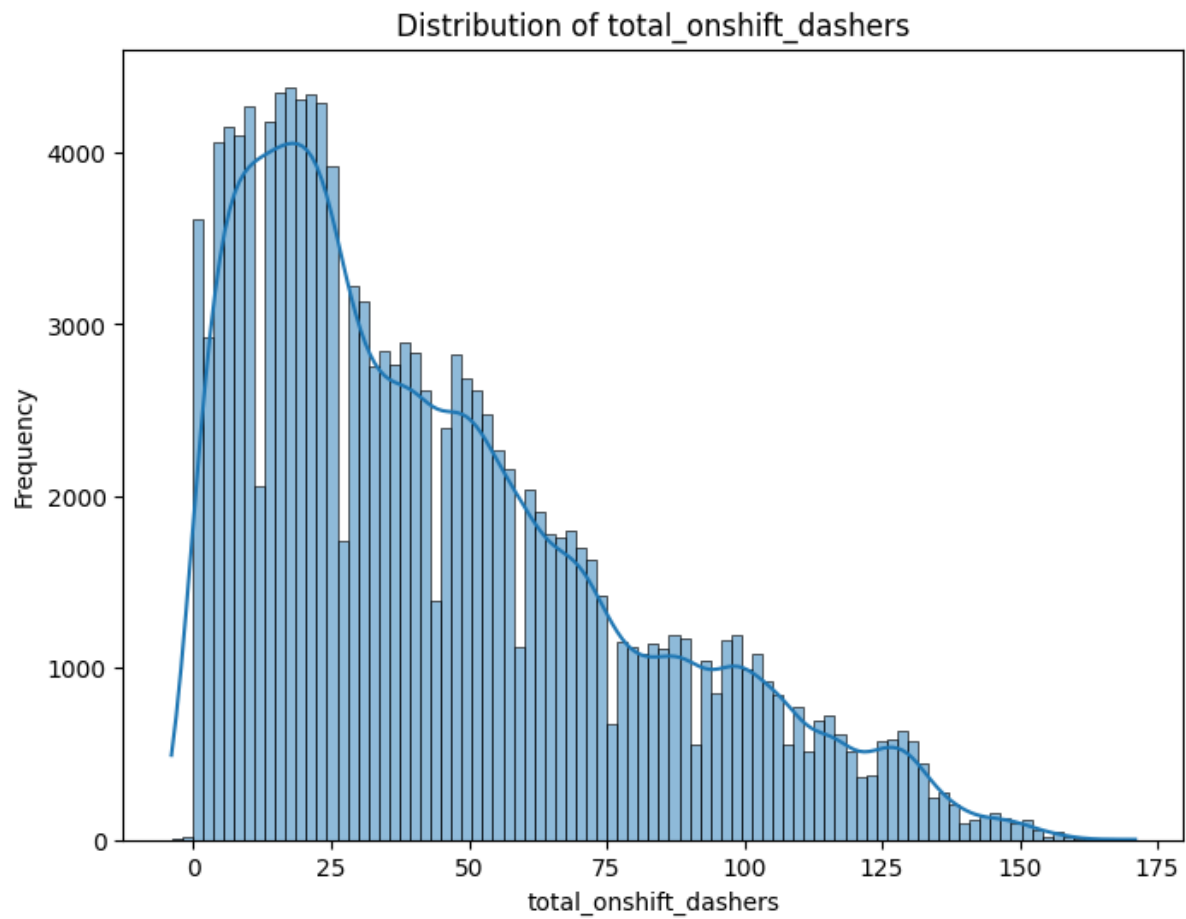
count	140621.000000
mean	684.377589
std	518.633806
min	-86.000000
25%	299.000000
50%	595.000000
75%	940.000000
max	8999.000000

Name: min_item_price, dtype: float64



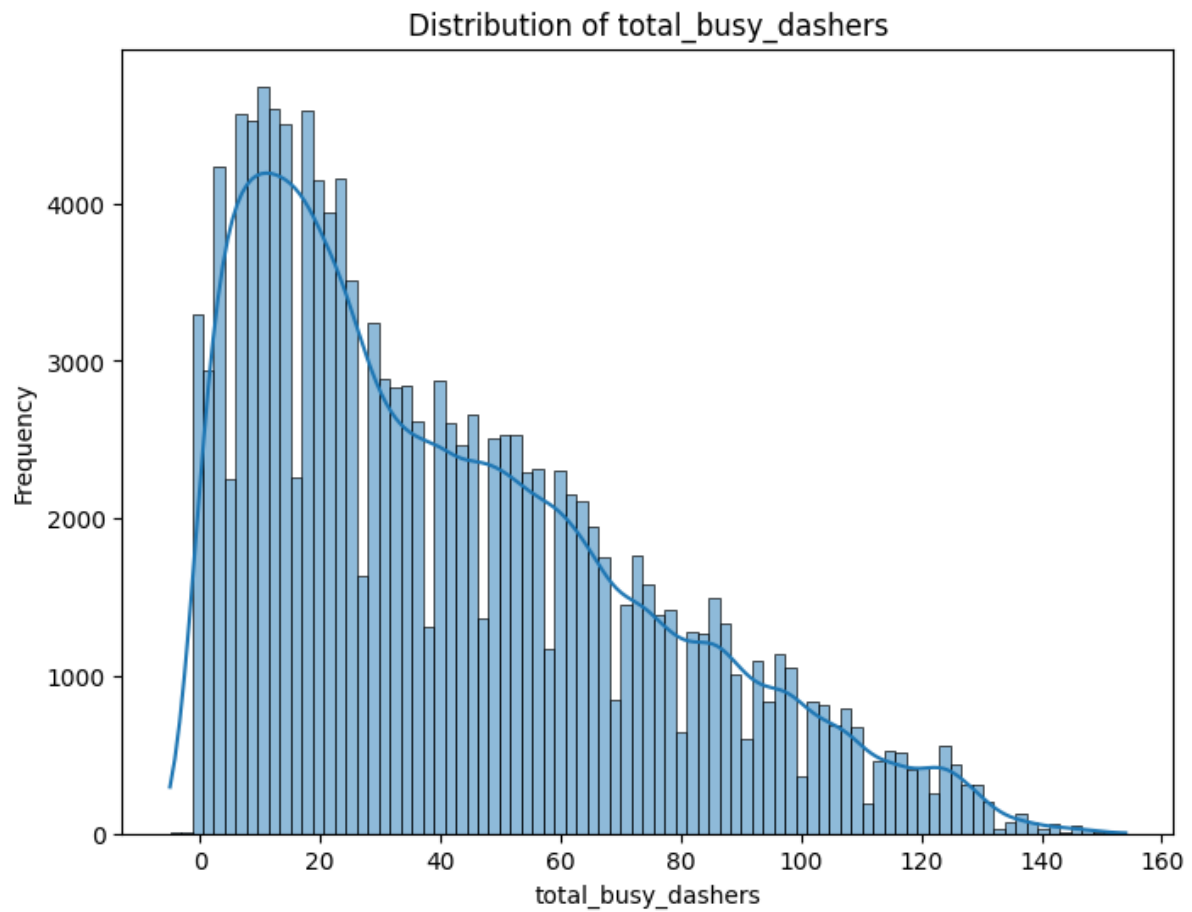
count	140621.000000
mean	1159.948656
std	559.693980
min	0.000000
25%	799.000000
50%	1095.000000
75%	1395.000000
max	8999.000000

Name: max_item_price, dtype: float64



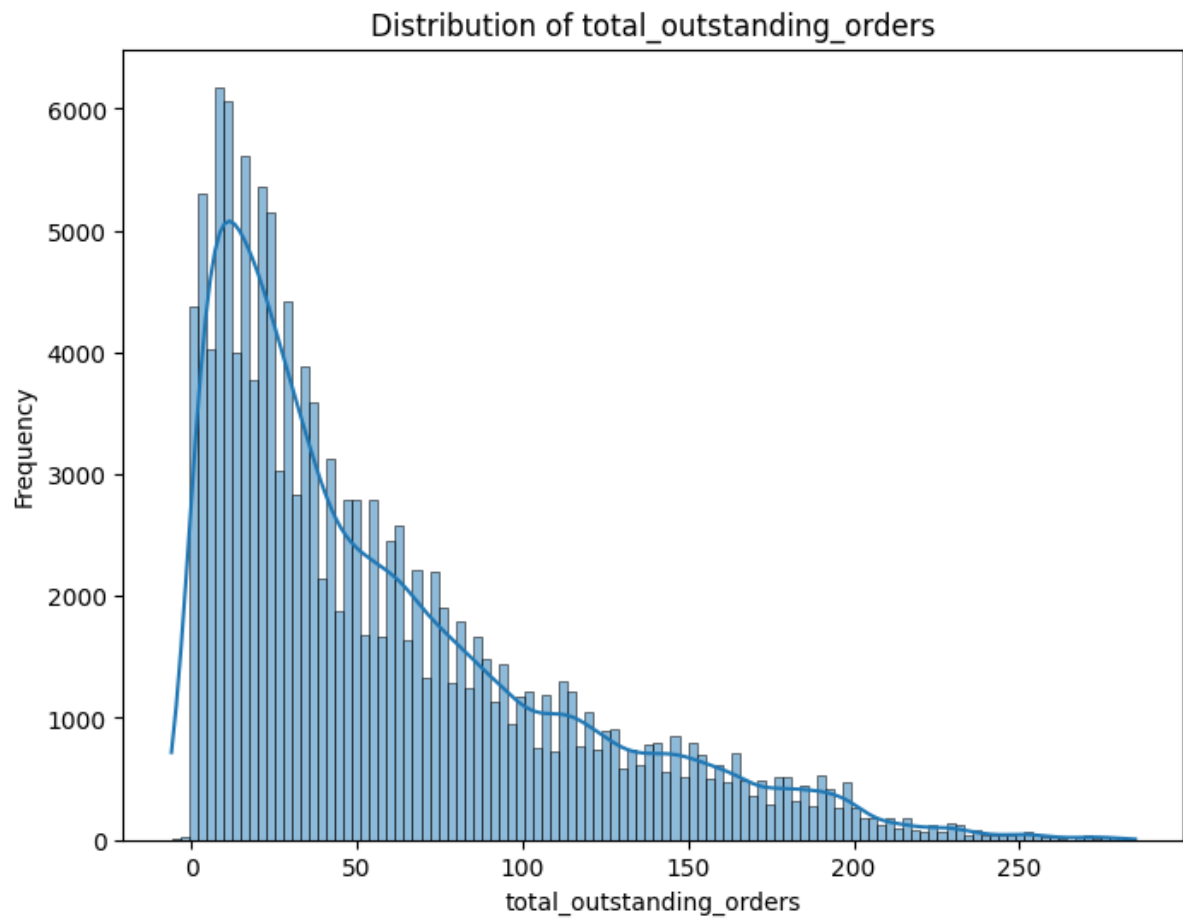
count	140621.000000
mean	44.927948
std	34.572455
min	-4.000000
25%	17.000000
50%	37.000000
75%	66.000000
max	171.000000

Name: total_onshift_dashers, dtype: float64



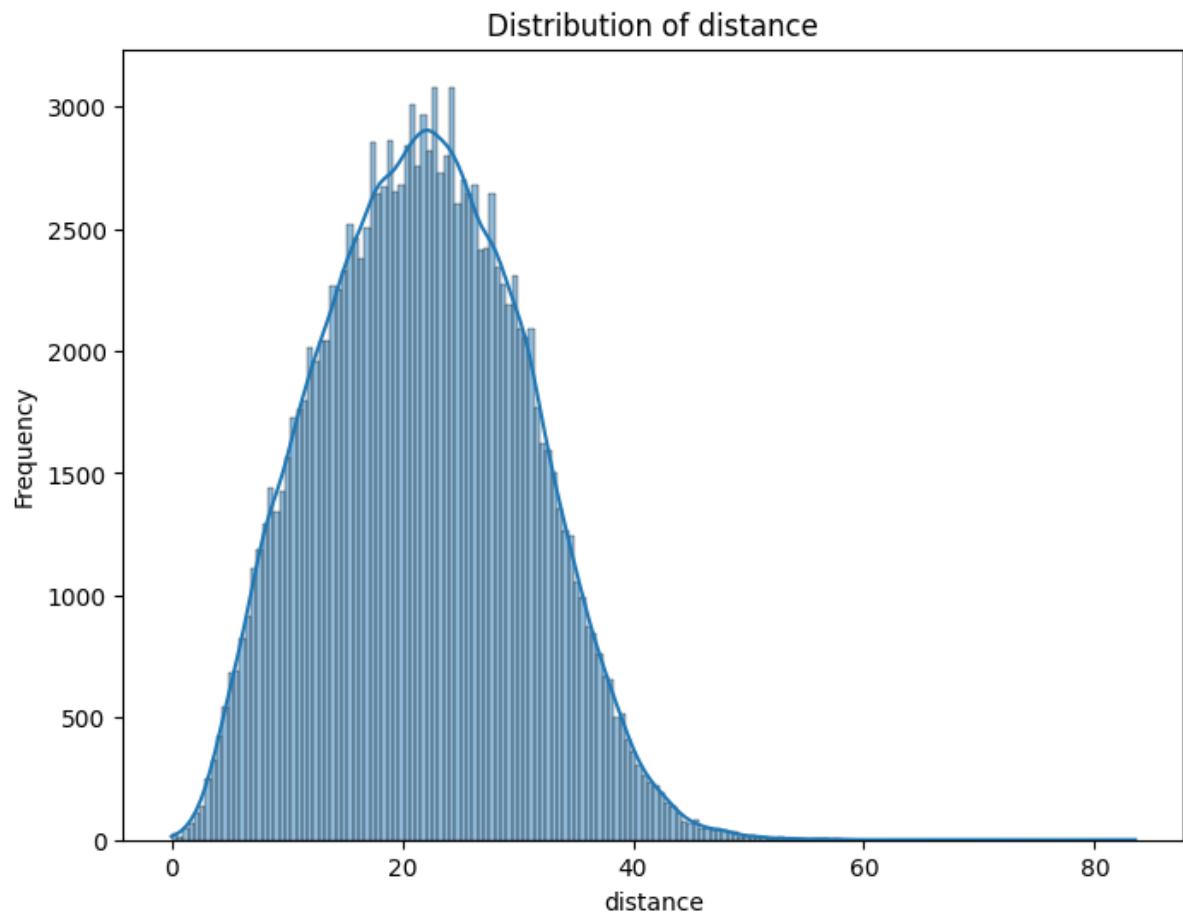
count	140621.000000
mean	41.880025
std	32.197755
min	-5.000000
25%	15.000000
50%	35.000000
75%	63.000000
max	154.000000

Name: total_busy_dashers, dtype: float64



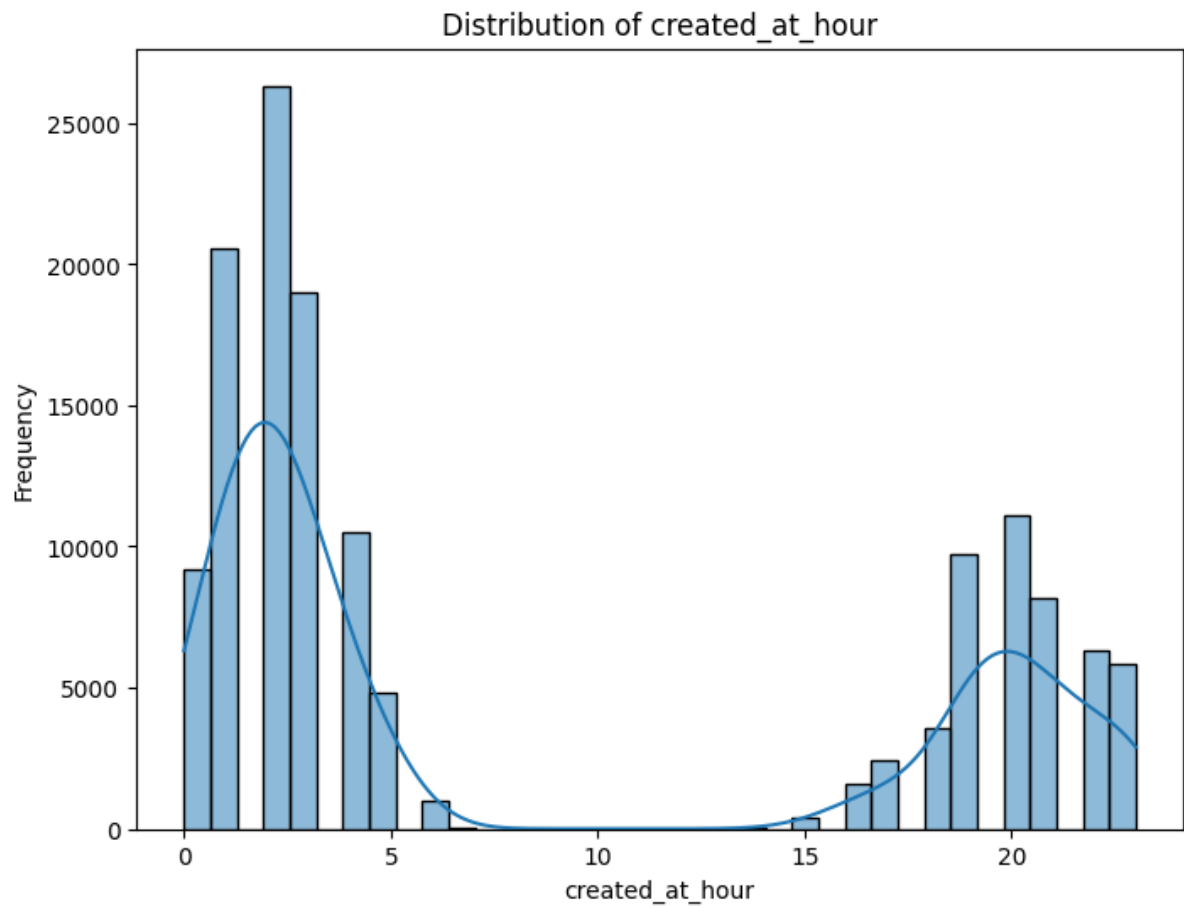
count	140621.000000
mean	58.245148
std	52.773640
min	-6.000000
25%	17.000000
50%	41.000000
75%	85.000000
max	285.000000

Name: total_outstanding_orders, dtype: float64



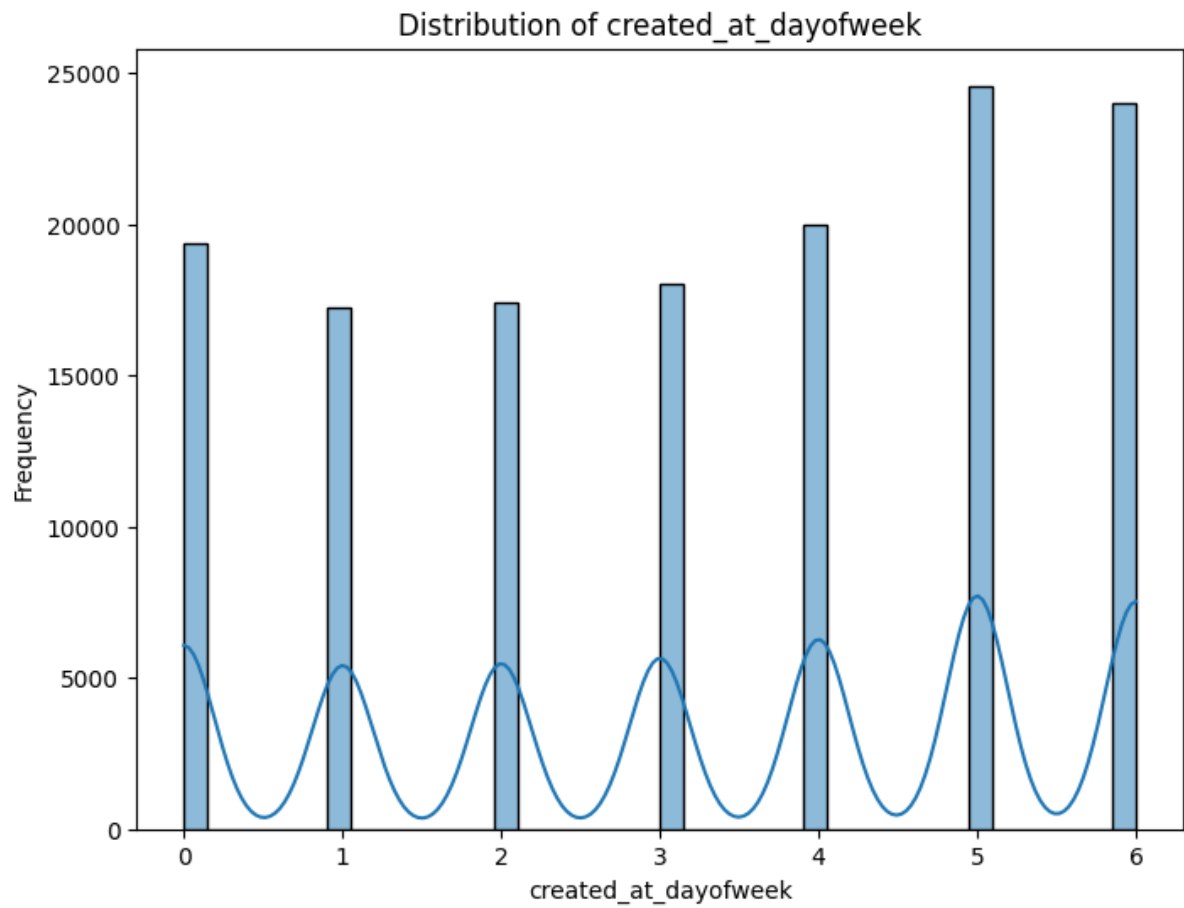
count	140621.000000
mean	21.841054
std	8.751147
min	0.000000
25%	15.320000
50%	21.760000
75%	28.120000
max	83.520000

Name: distance, dtype: float64



count	140621.000000
mean	8.475932
std	8.677293
min	0.000000
25%	2.000000
50%	3.000000
75%	19.000000
max	23.000000

Name: created_at_hour, dtype: float64



```
count    140621.000000
mean         3.221468
std         2.044919
min          0.000000
25%         1.000000
50%         3.000000
75%         5.000000
max          6.000000
```

Name: created_at_dayofweek, dtype: float64



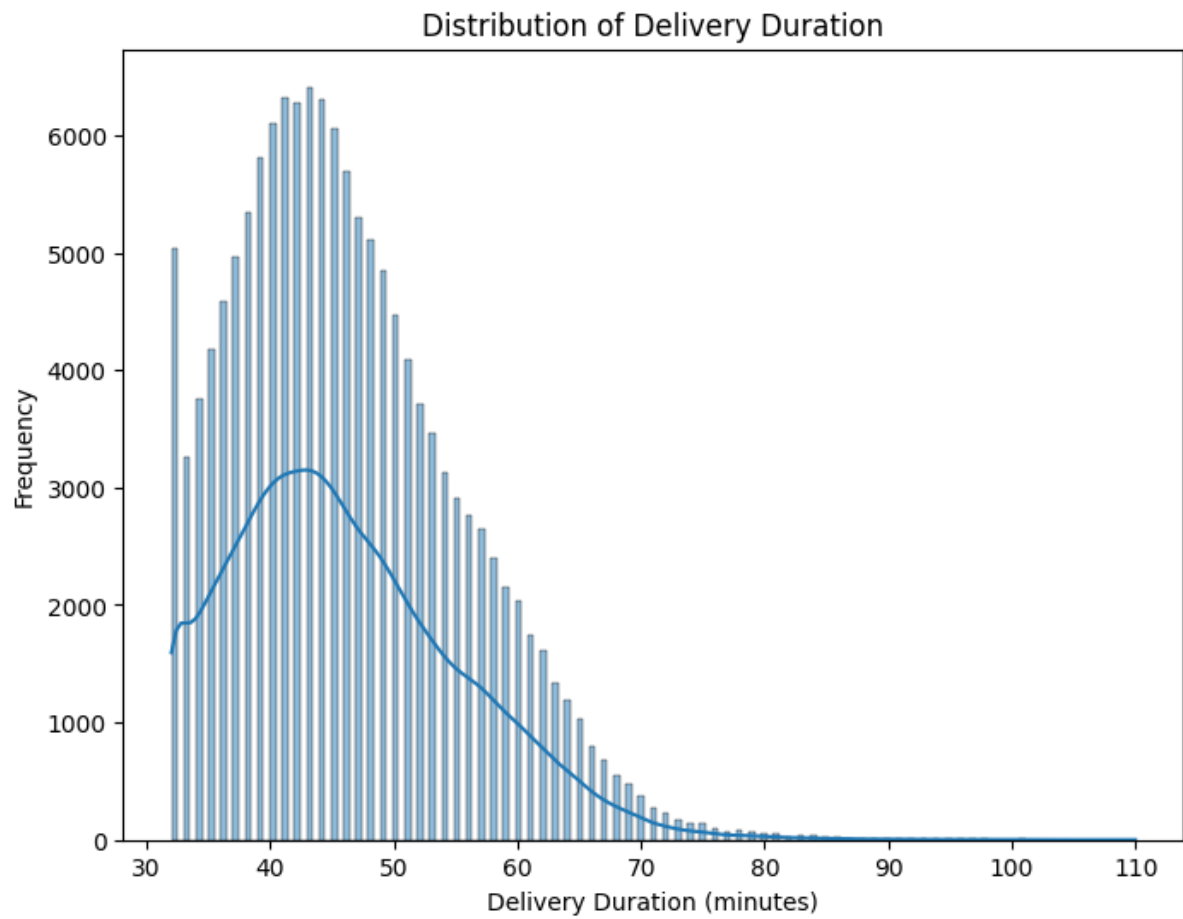
count	140621.000000
mean	8.550224
std	8.383521
min	0.000000
25%	2.000000
50%	4.000000
75%	20.000000
max	23.000000

Name: actual_delivery_time_hour, dtype: float64



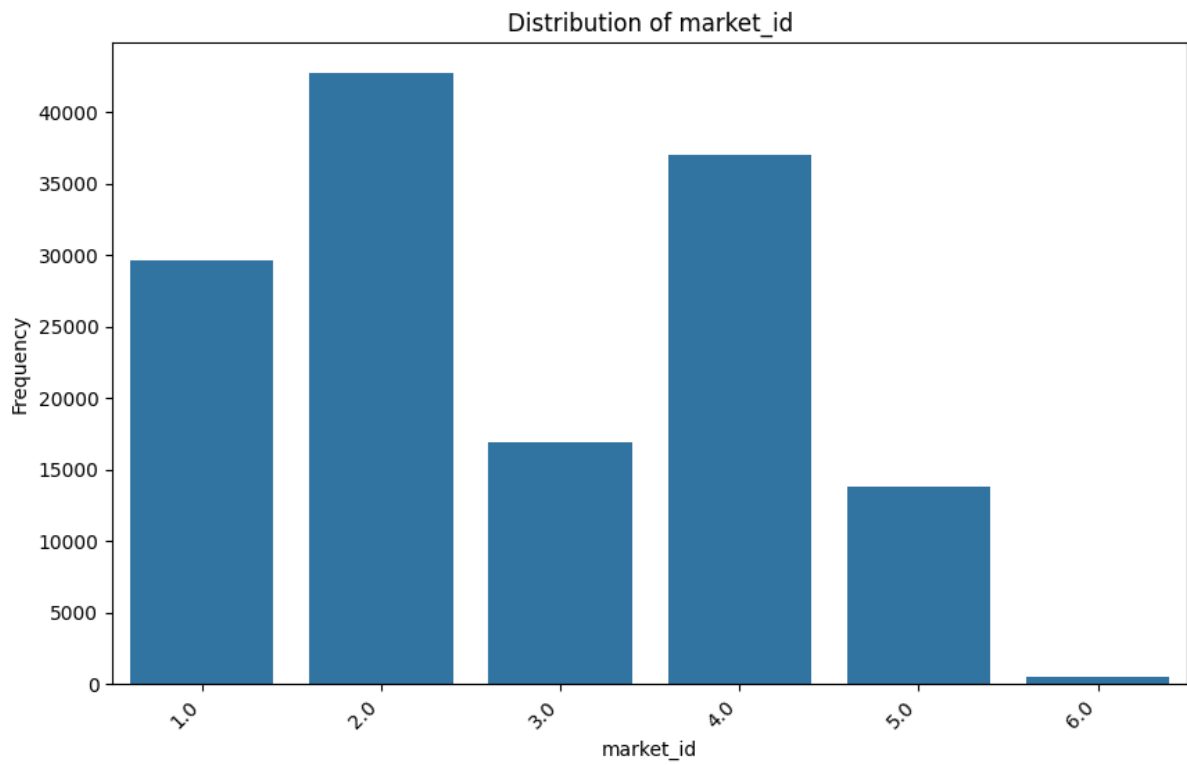
```
count    140621.000000
mean         3.204514
std         2.049327
min          0.000000
25%         1.000000
50%         3.000000
75%         5.000000
max          6.000000
```

Name: actual_delivery_time_dayofweek, dtype: float64



count	140621.000000
mean	46.198071
std	9.320360
min	32.000000
25%	39.000000
50%	45.000000
75%	52.000000
max	110.000000

Name: delivery_duration, dtype: float64



market_id

2.0 42722

4.0 36961

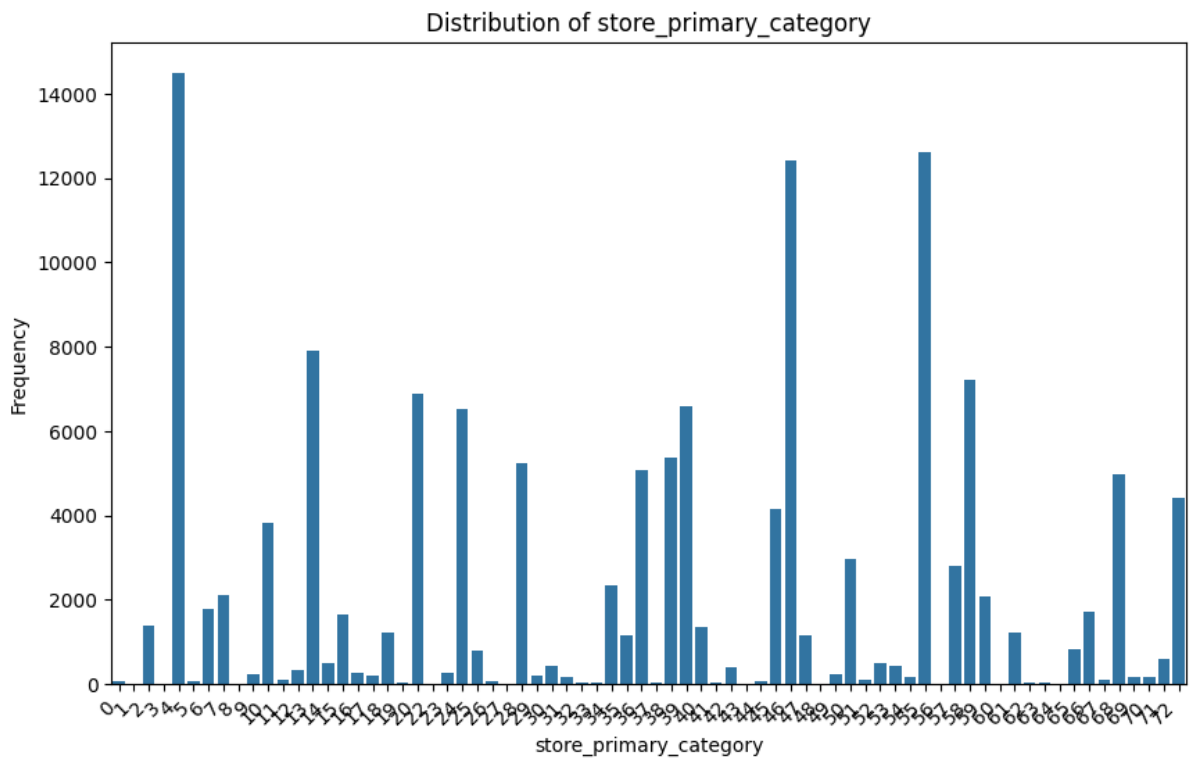
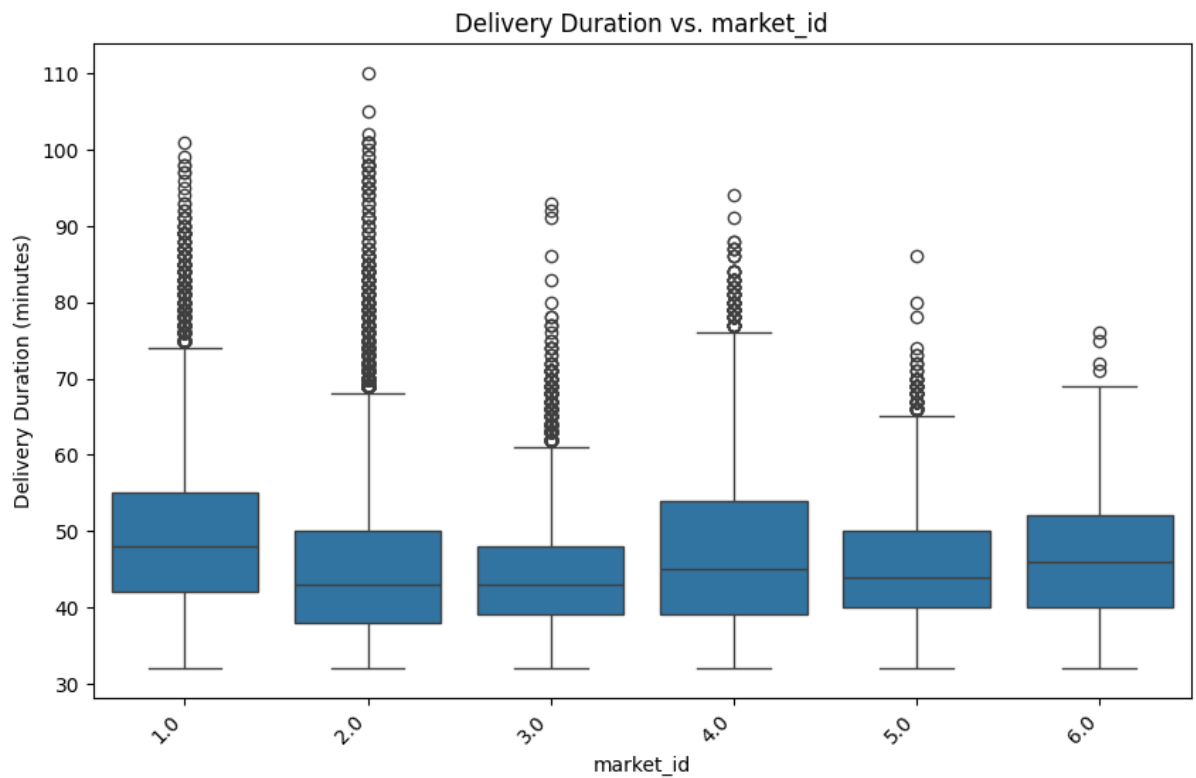
1.0 29662

3.0 16934

5.0 13838

6.0 504

Name: count, dtype: int64



store_primary_category

4 14483

55 12603

46 12427

13 7907

58 7227

...

56 9

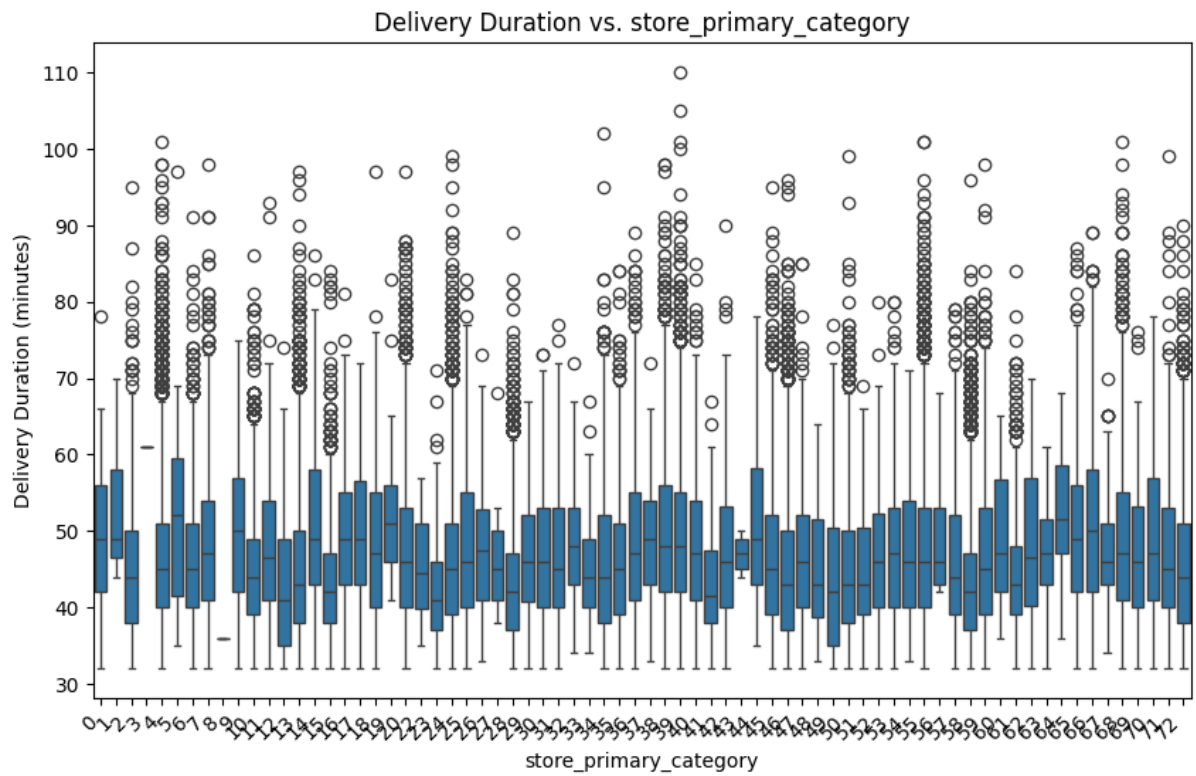
1 7

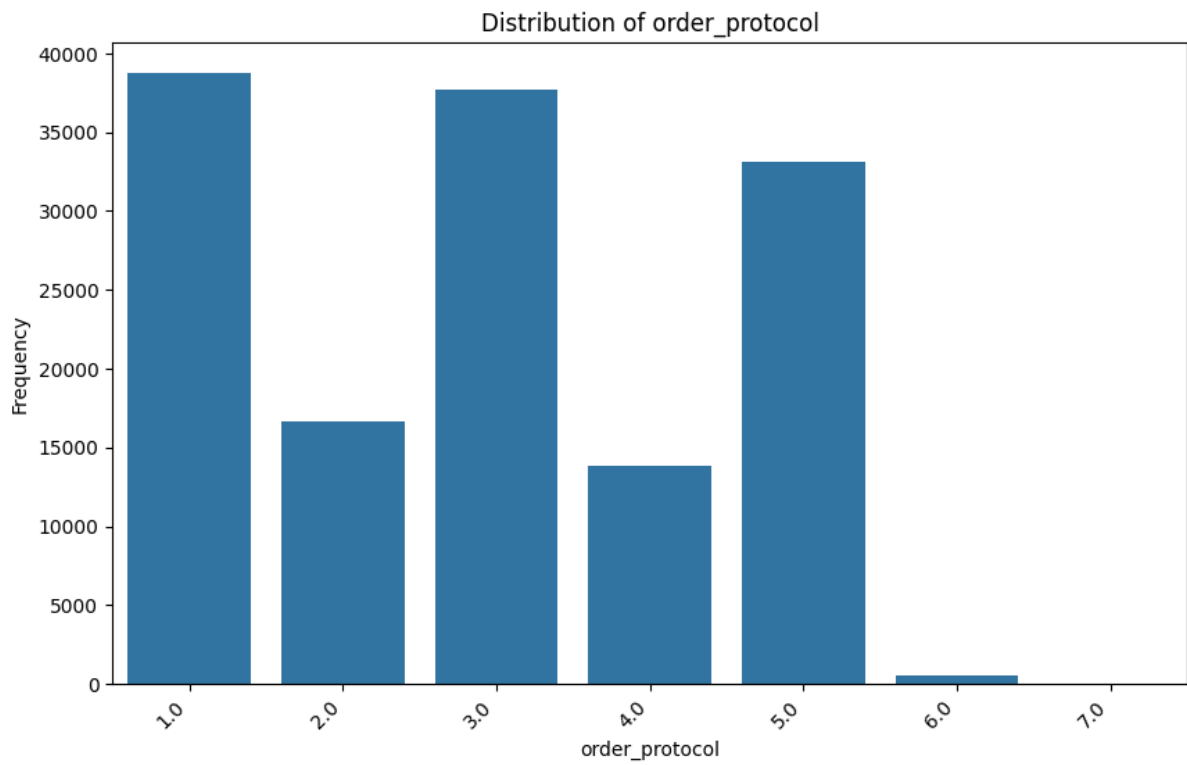
43 6

3 1

8 1

Name: count, Length: 72, dtype: int64





order_protocol

1.0 38746

3.0 37674

5.0 33118

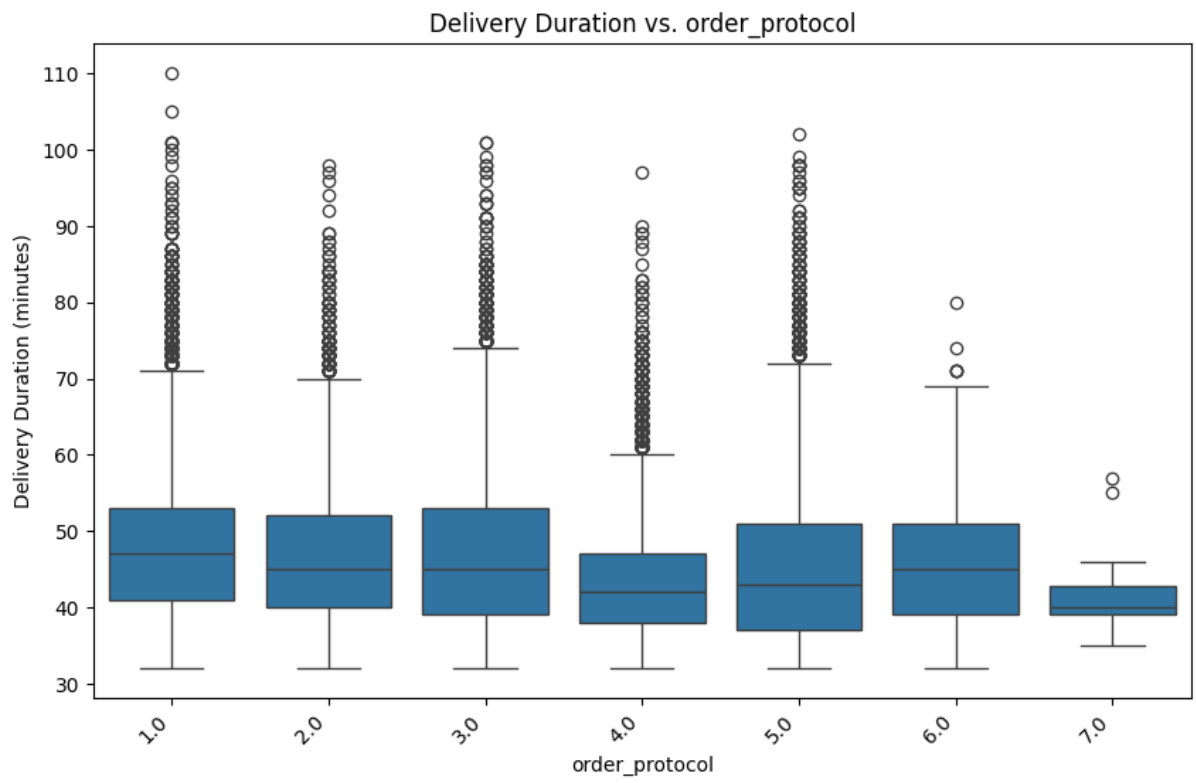
2.0 16689

4.0 13842

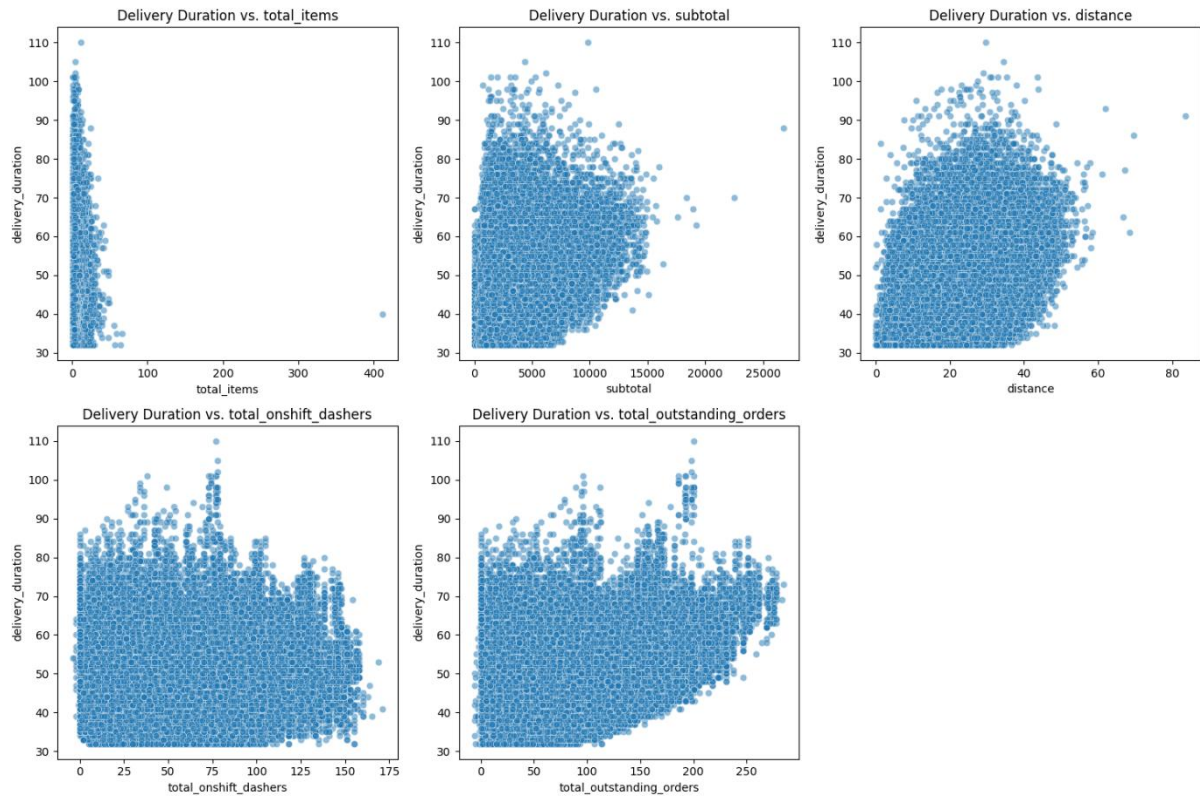
6.0 534

7.0 18

Name: count, dtype: int64



EDL Analysing the data

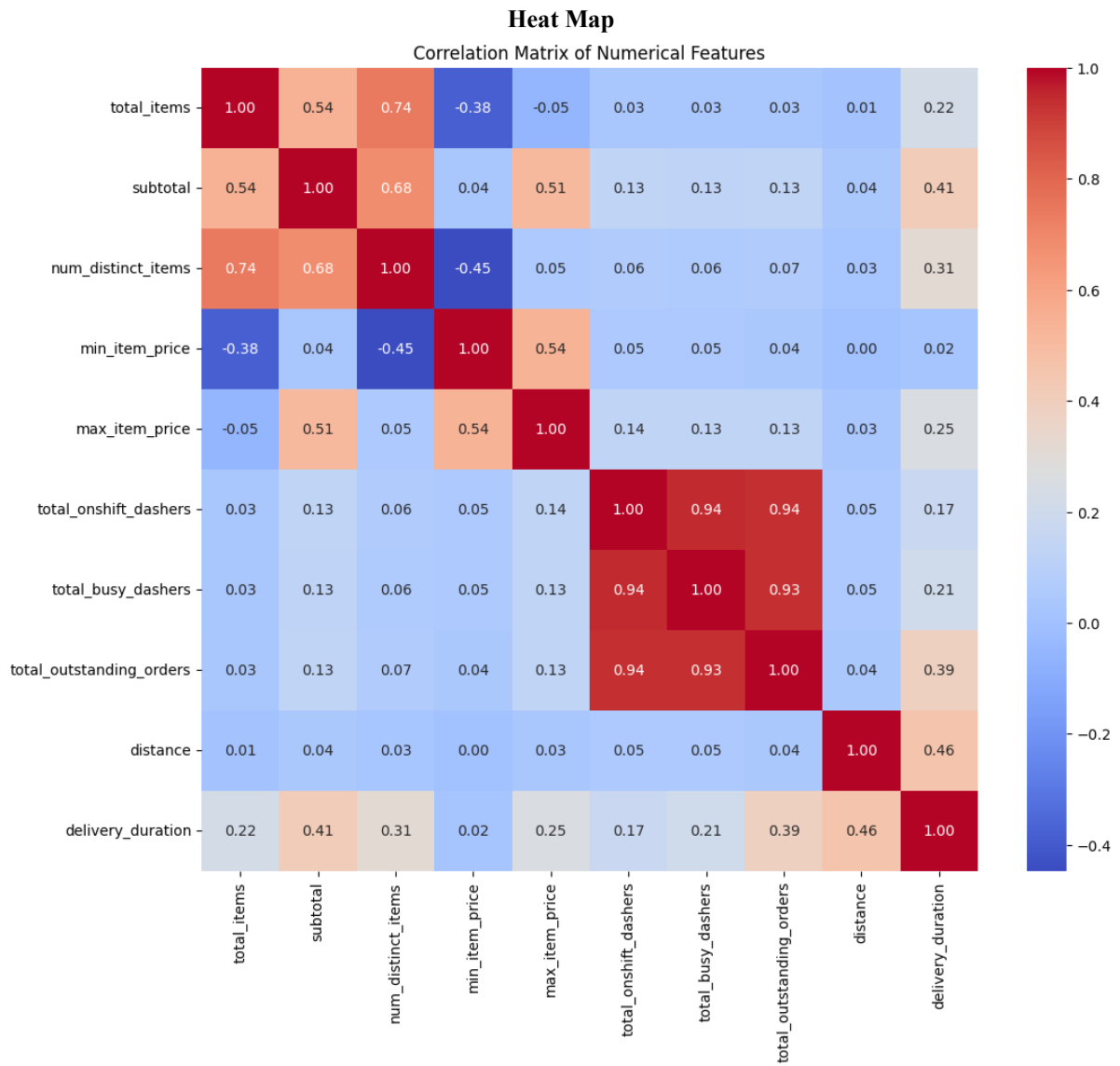


Correlation Coefficients:

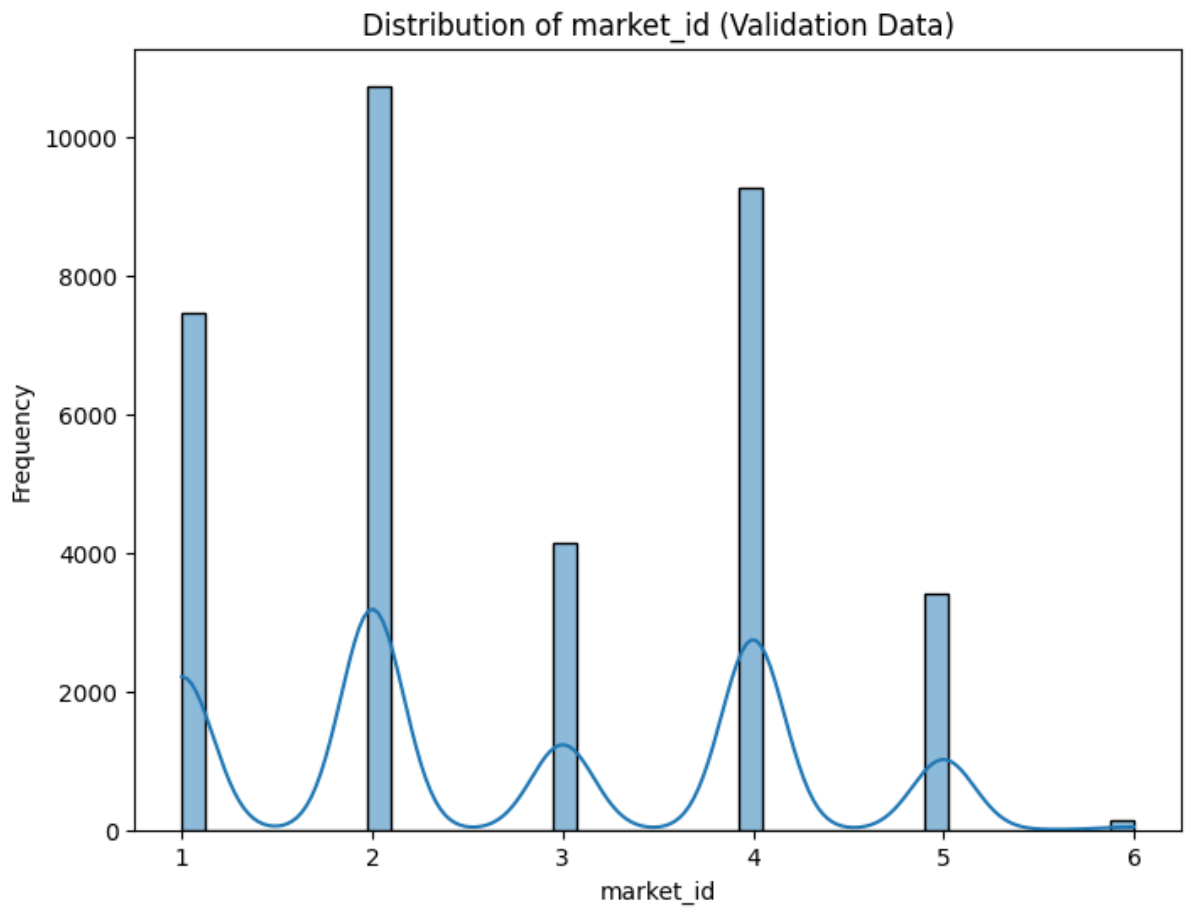
total_items	0.220804
subtotal	0.413392
distance	0.461026
total_onshift_dashers	0.171828
total_outstanding_orders	0.385028
delivery_duration	1.000000

Name: delivery_duration, dtype: float64

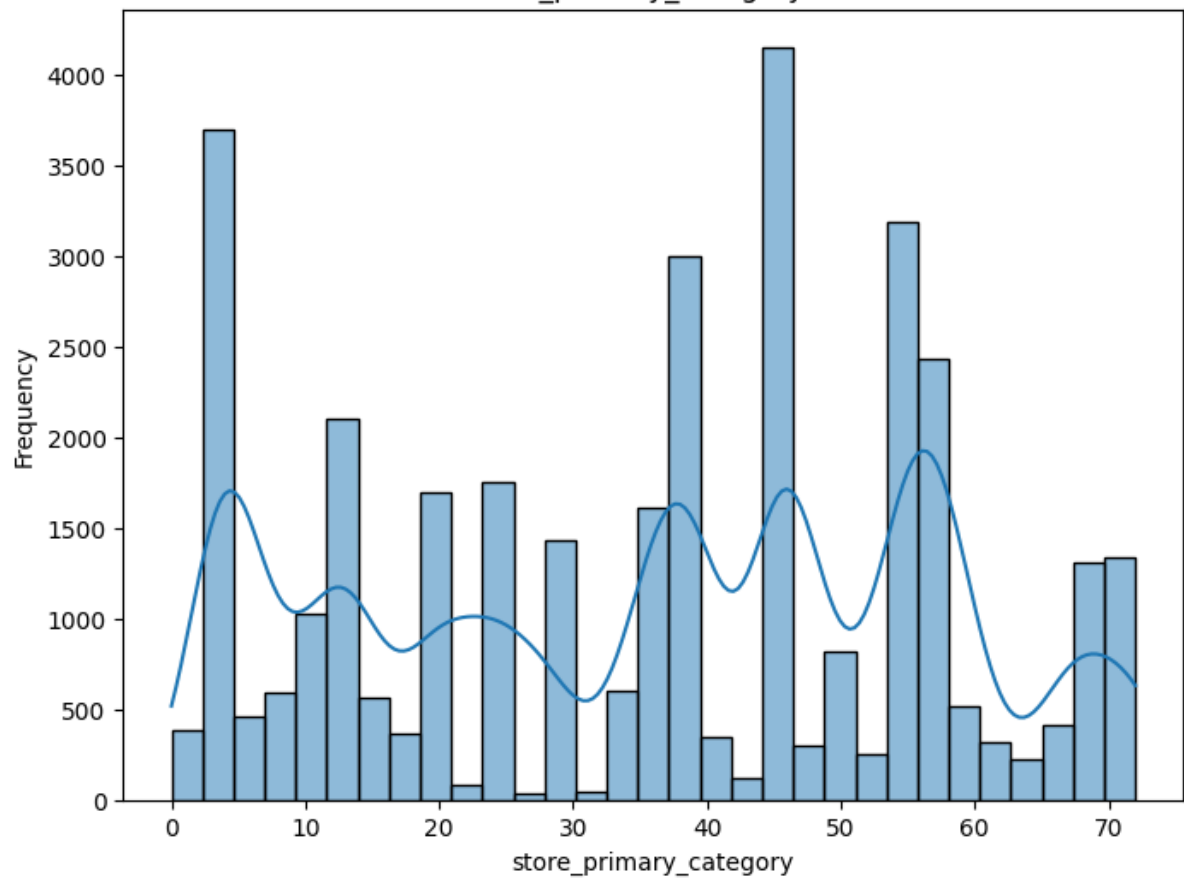


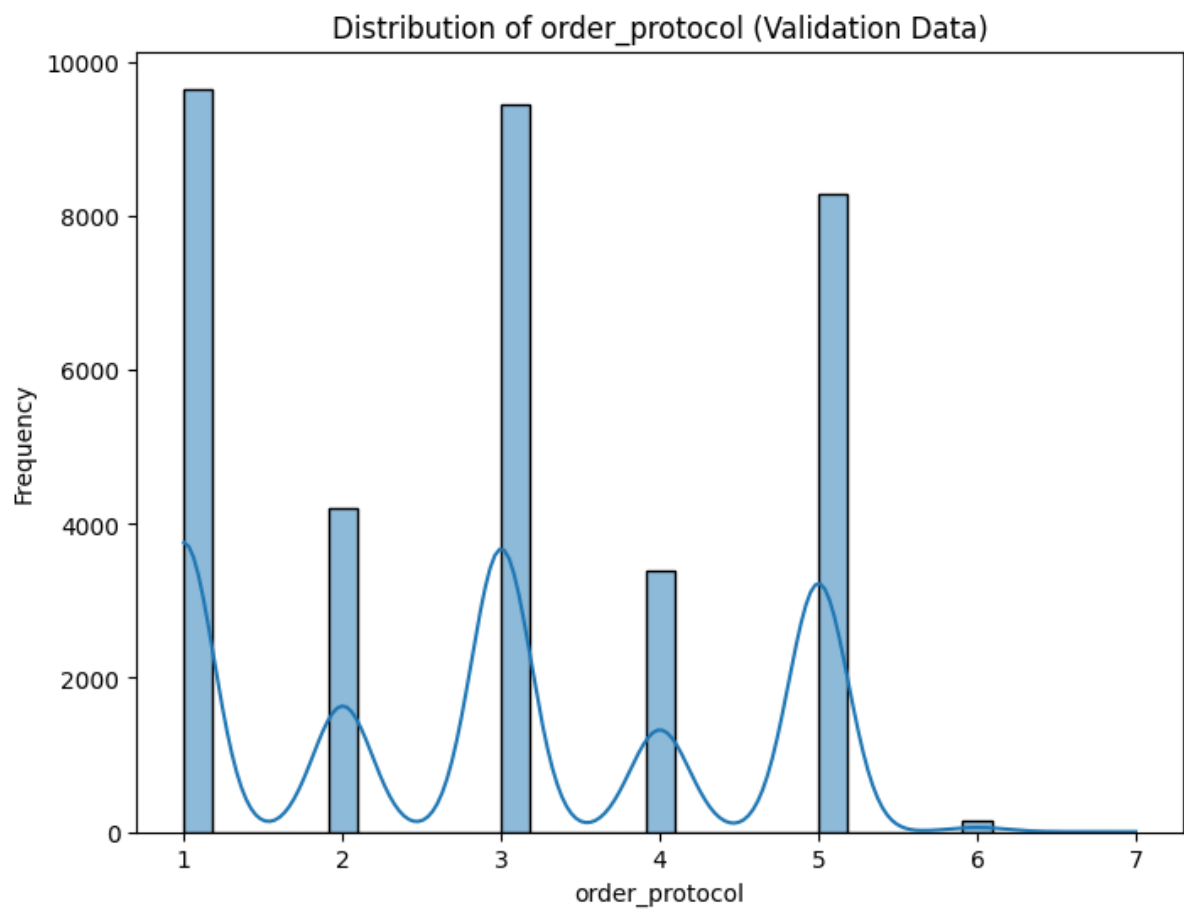


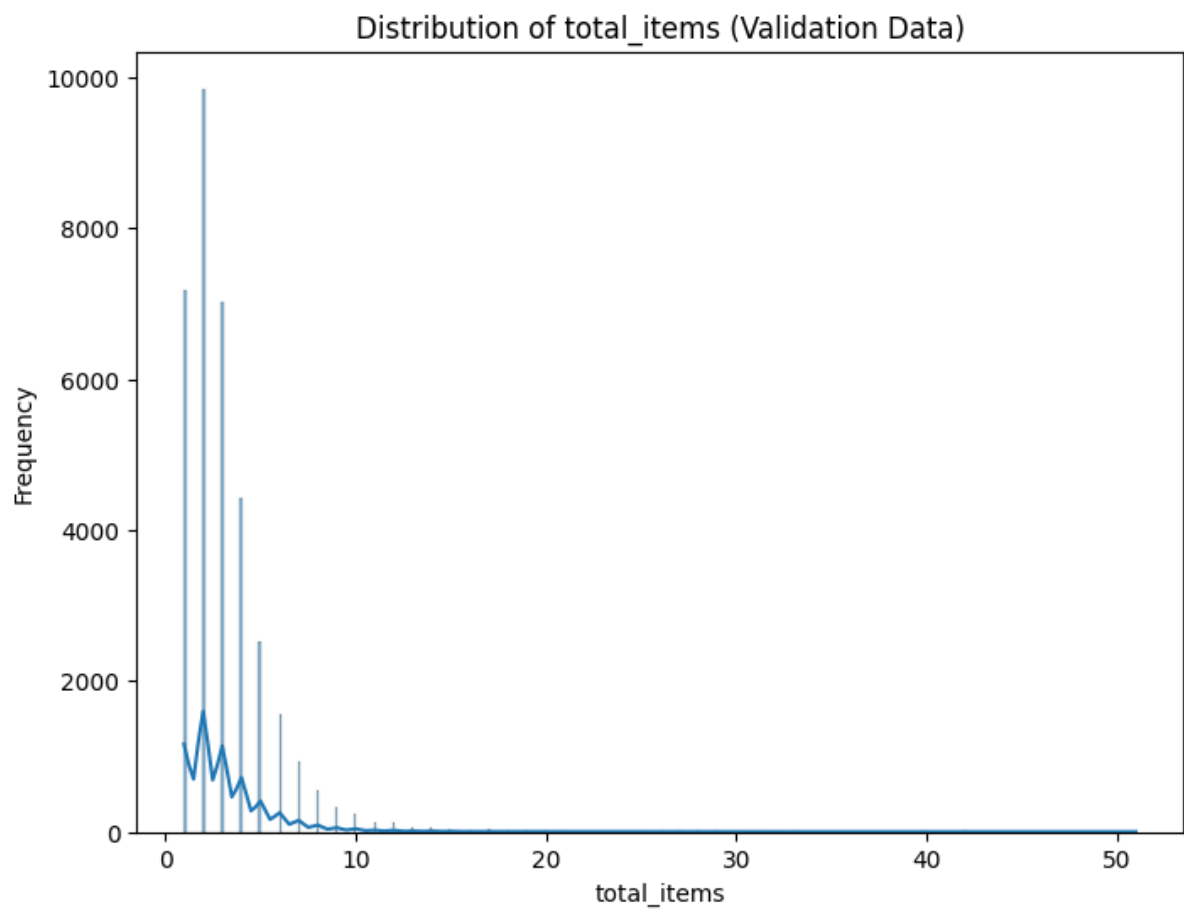
4) Exploratory Data Analysis on Validation Data
4.1 Feature Distribution



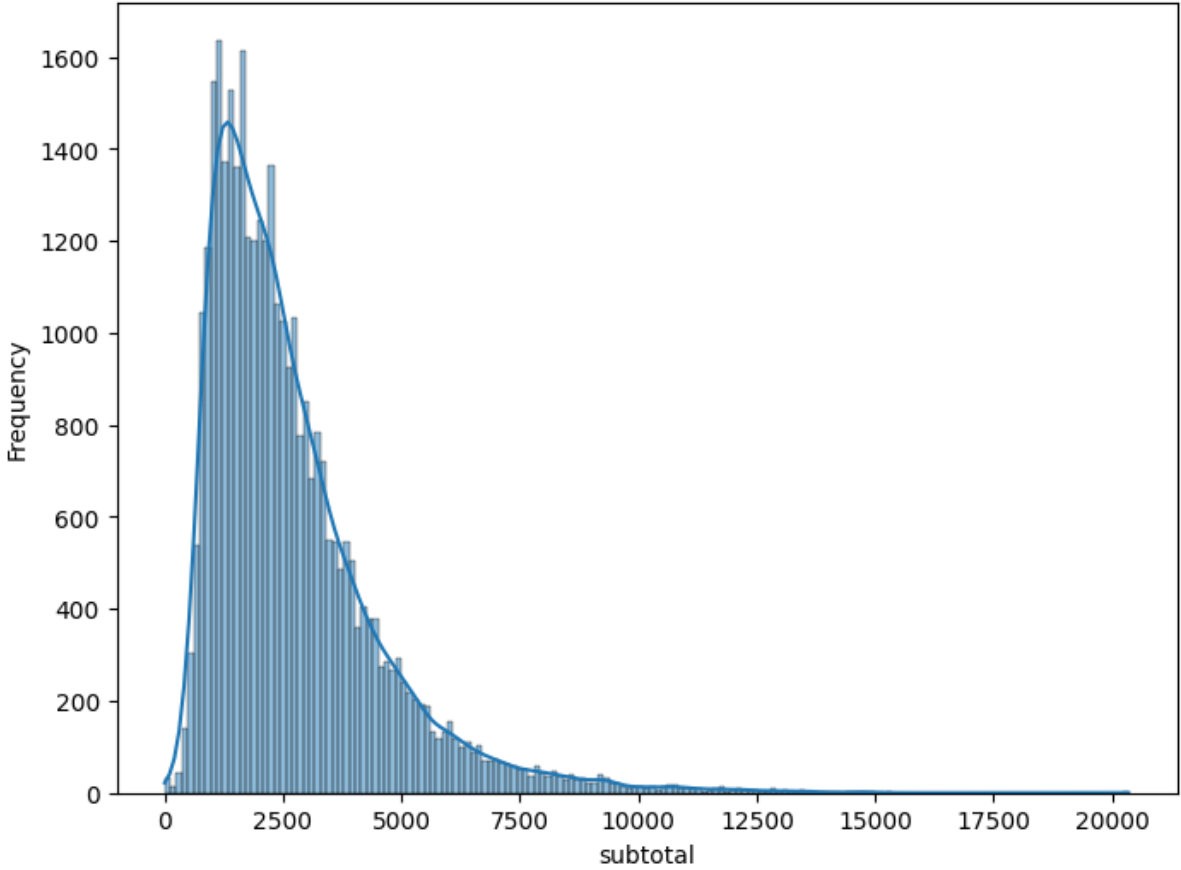
Distribution of store_primary_category (Validation Data)

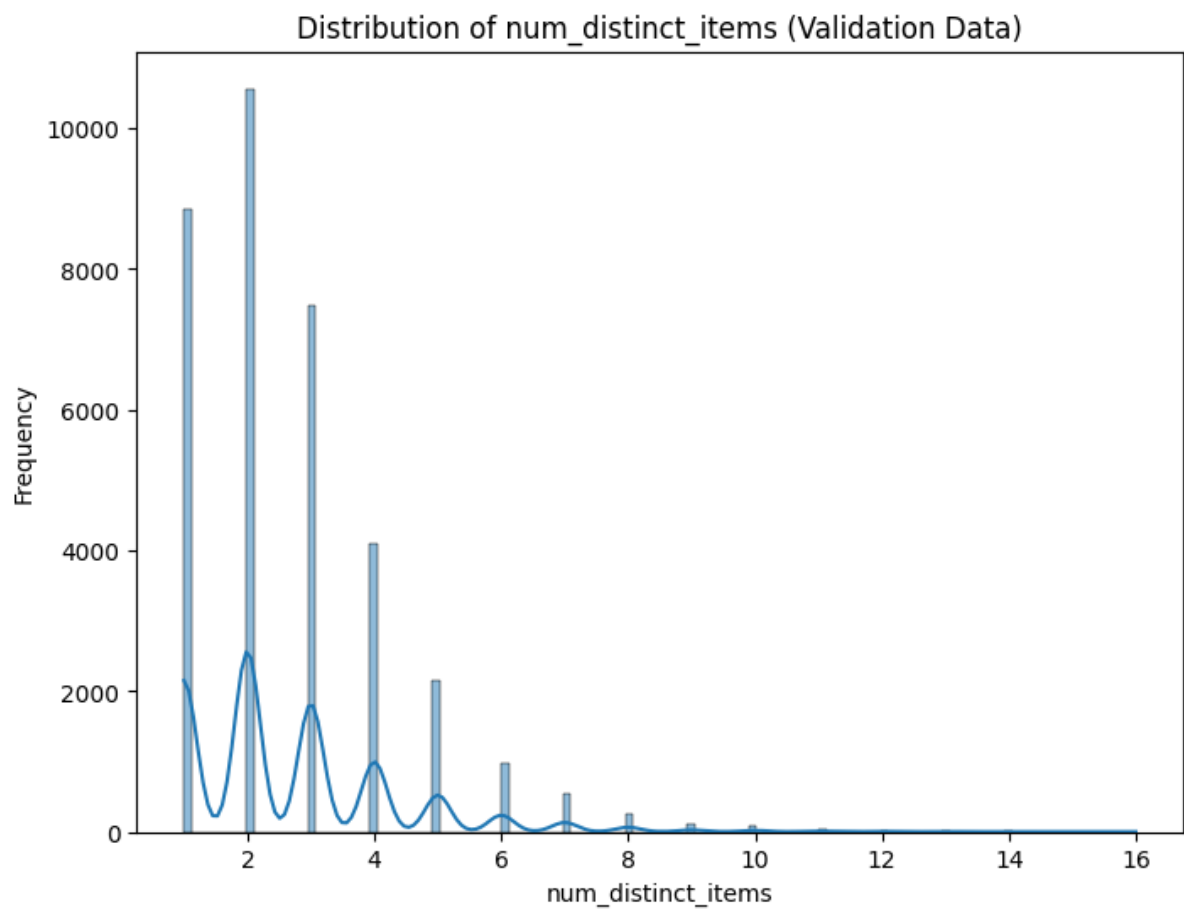




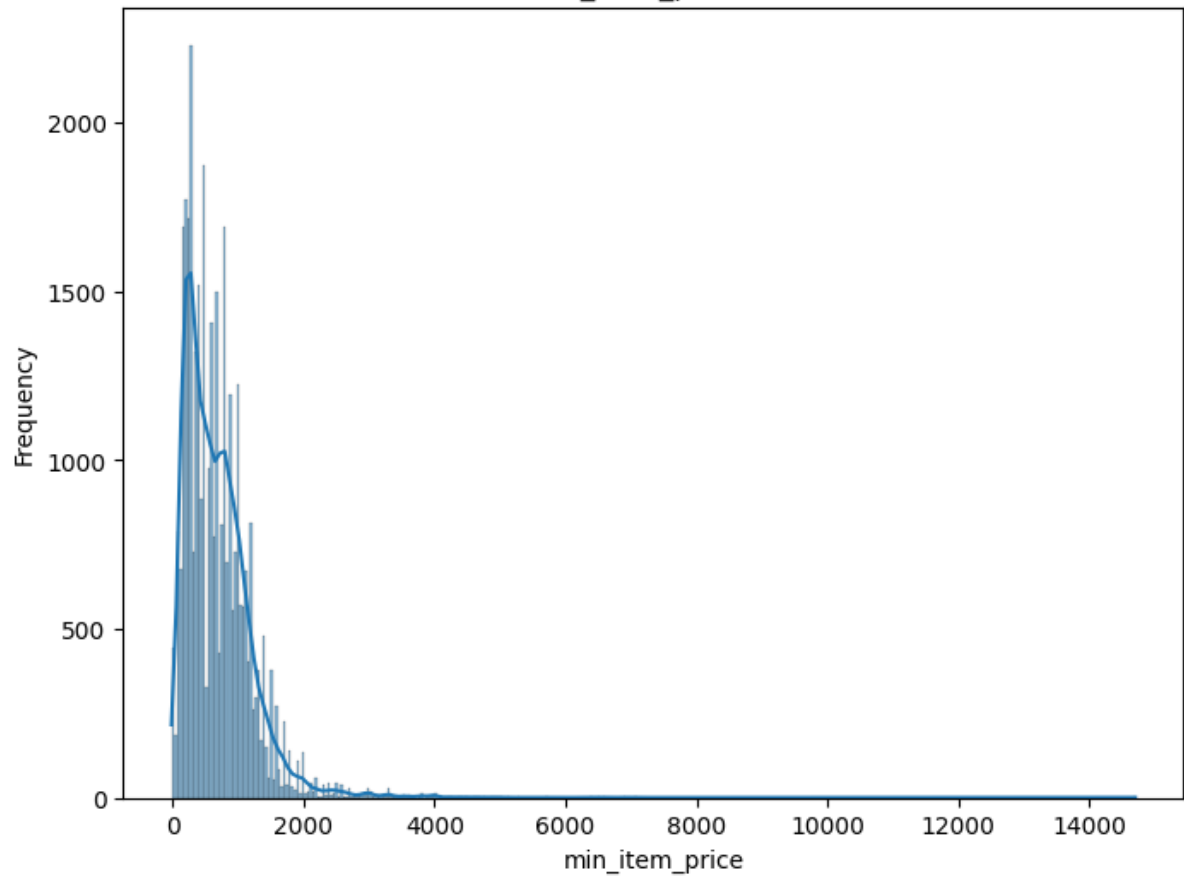


Distribution of subtotal (Validation Data)

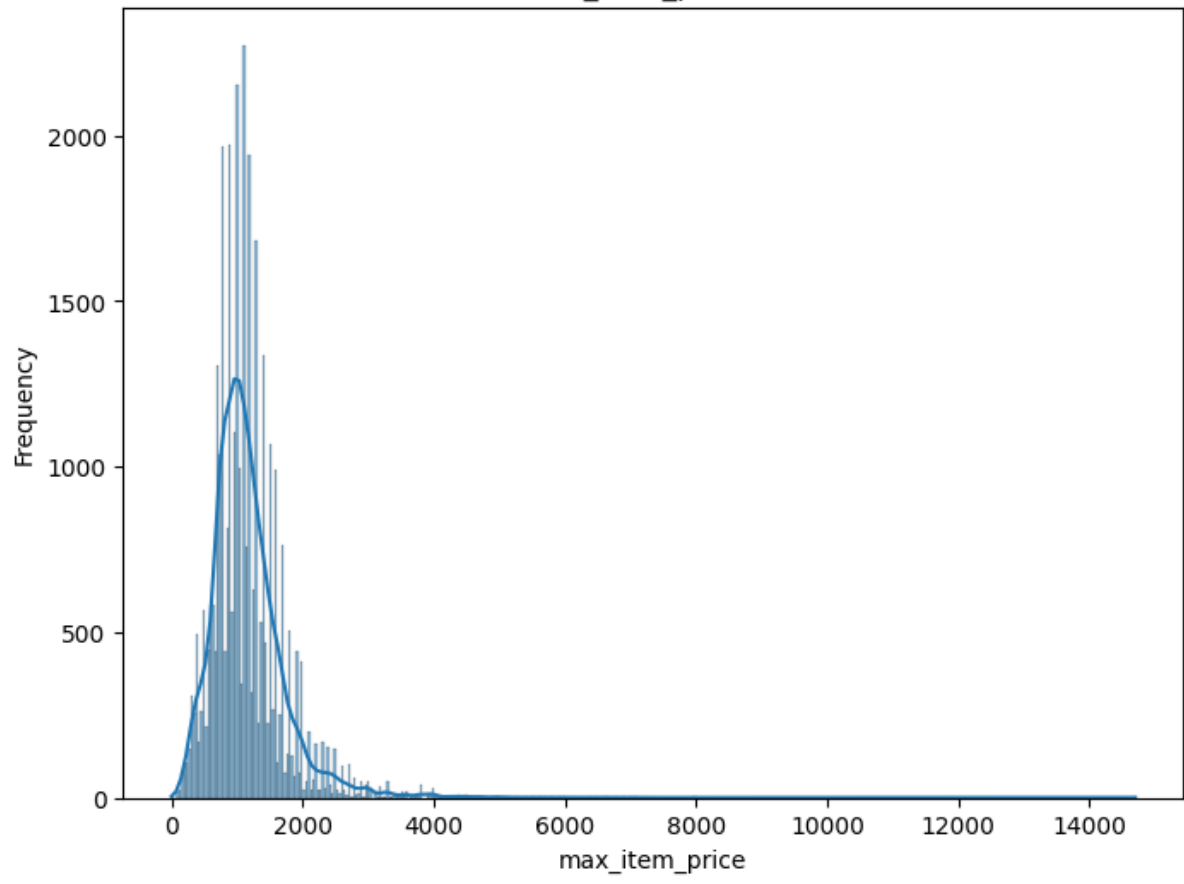




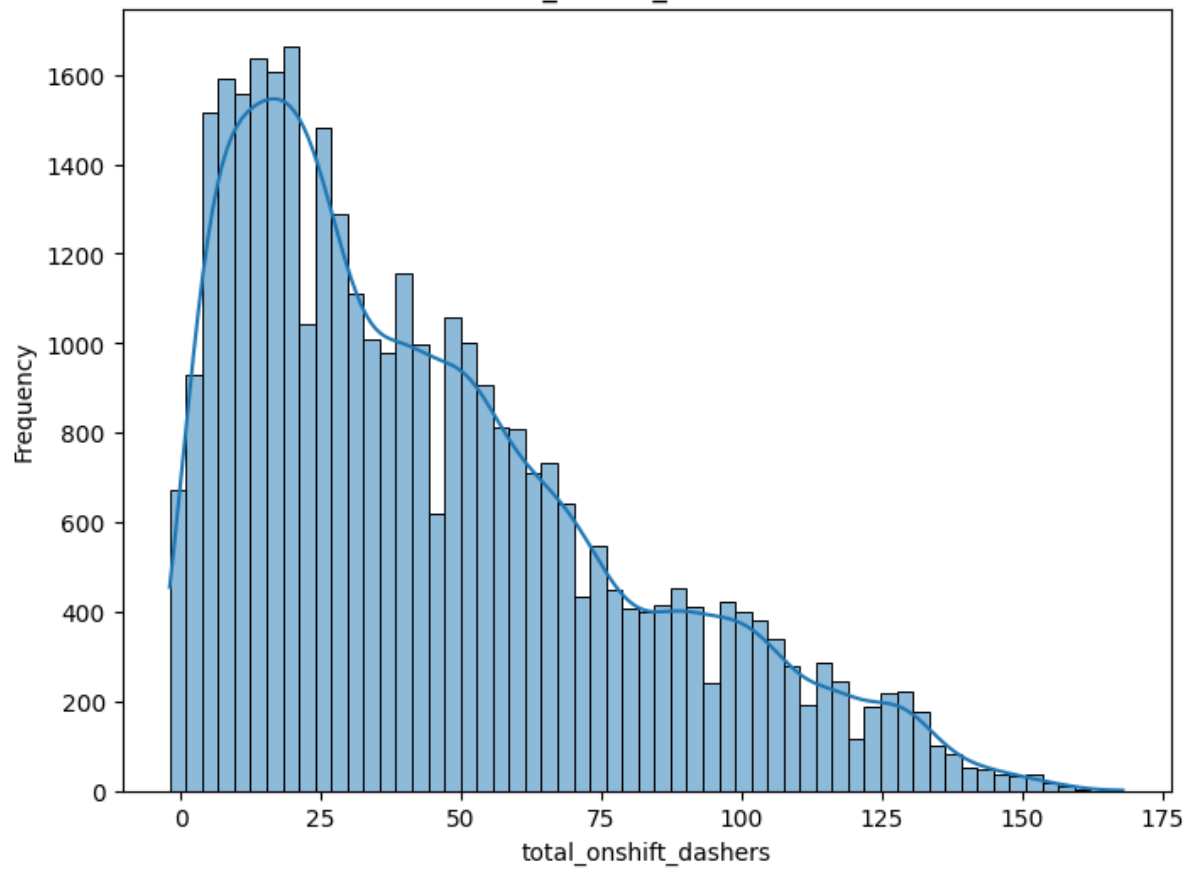
Distribution of min_item_price (Validation Data)

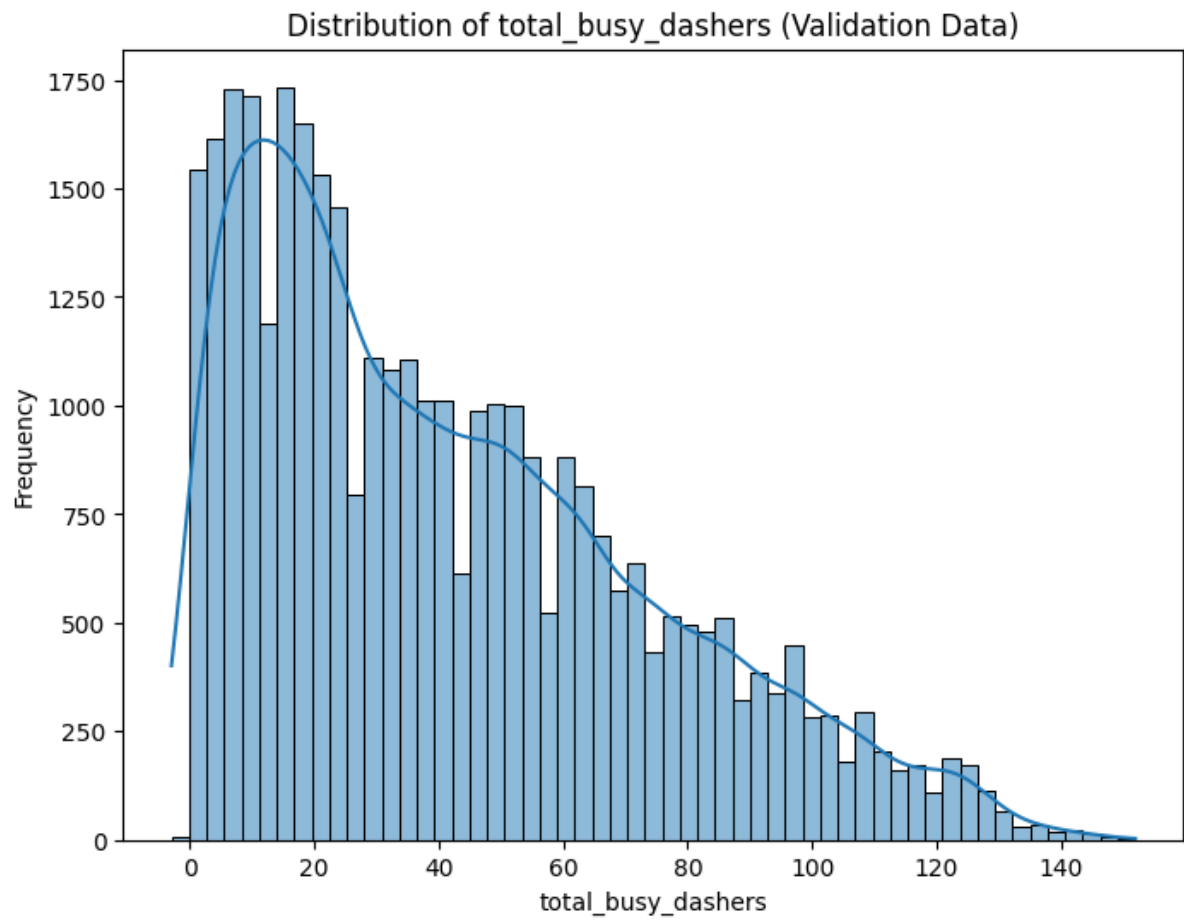


Distribution of max_item_price (Validation Data)

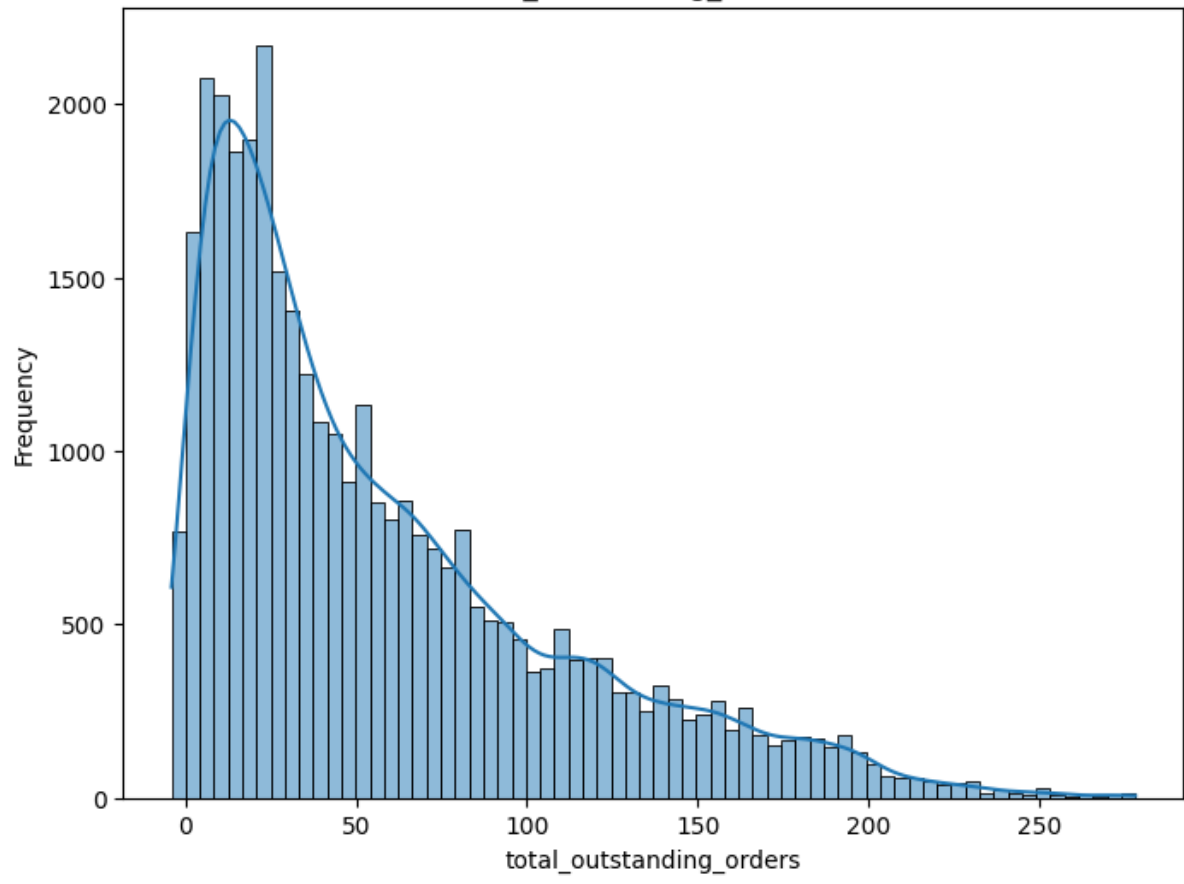


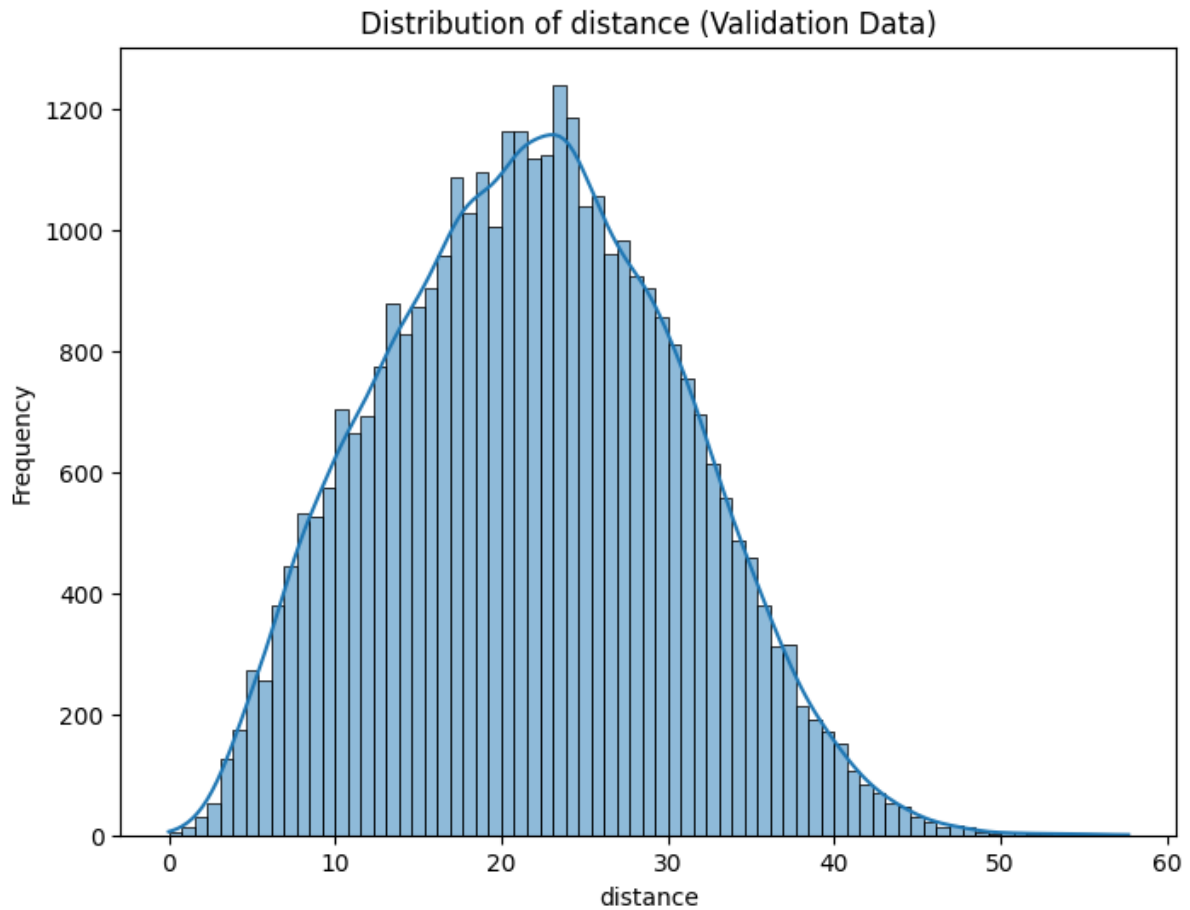
Distribution of total_onshift_dashers (Validation Data)



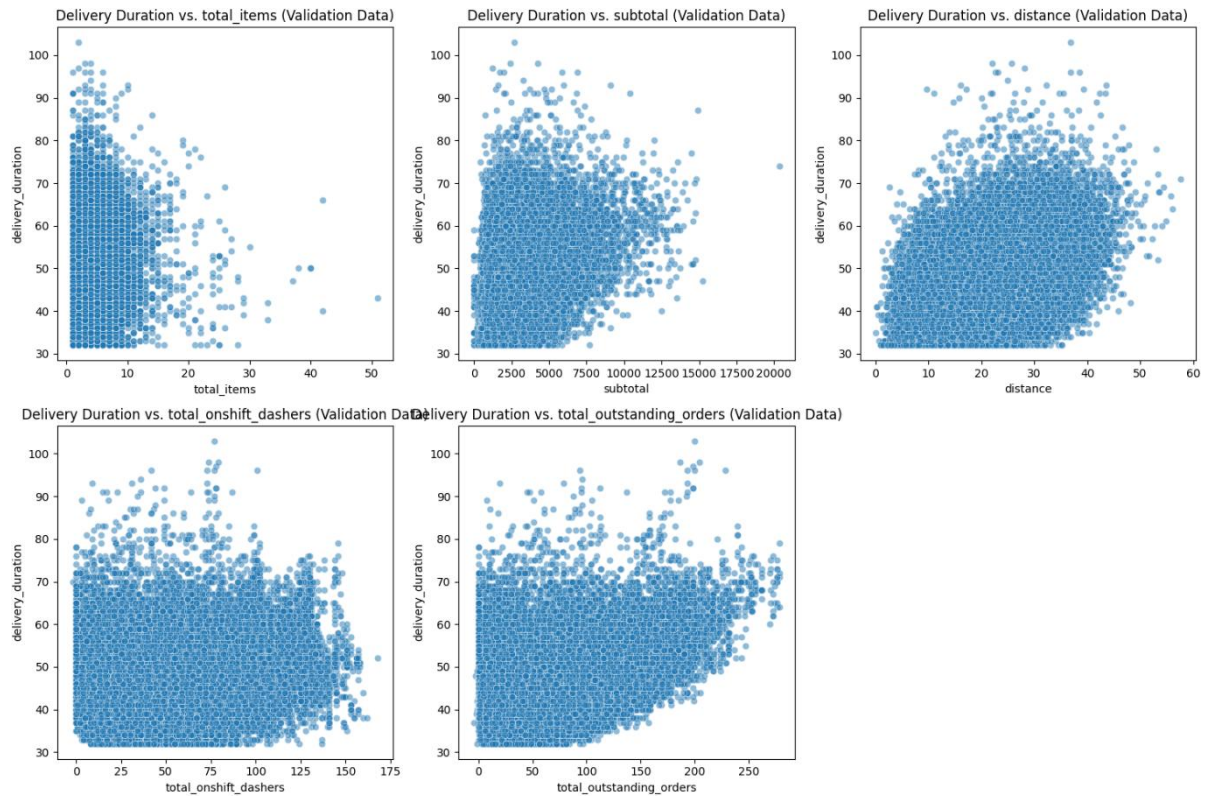


Distribution of total_outstanding_orders (Validation Data)





4.2 Relationships between different features



4.3 Correlation Analysis -Heatmap

