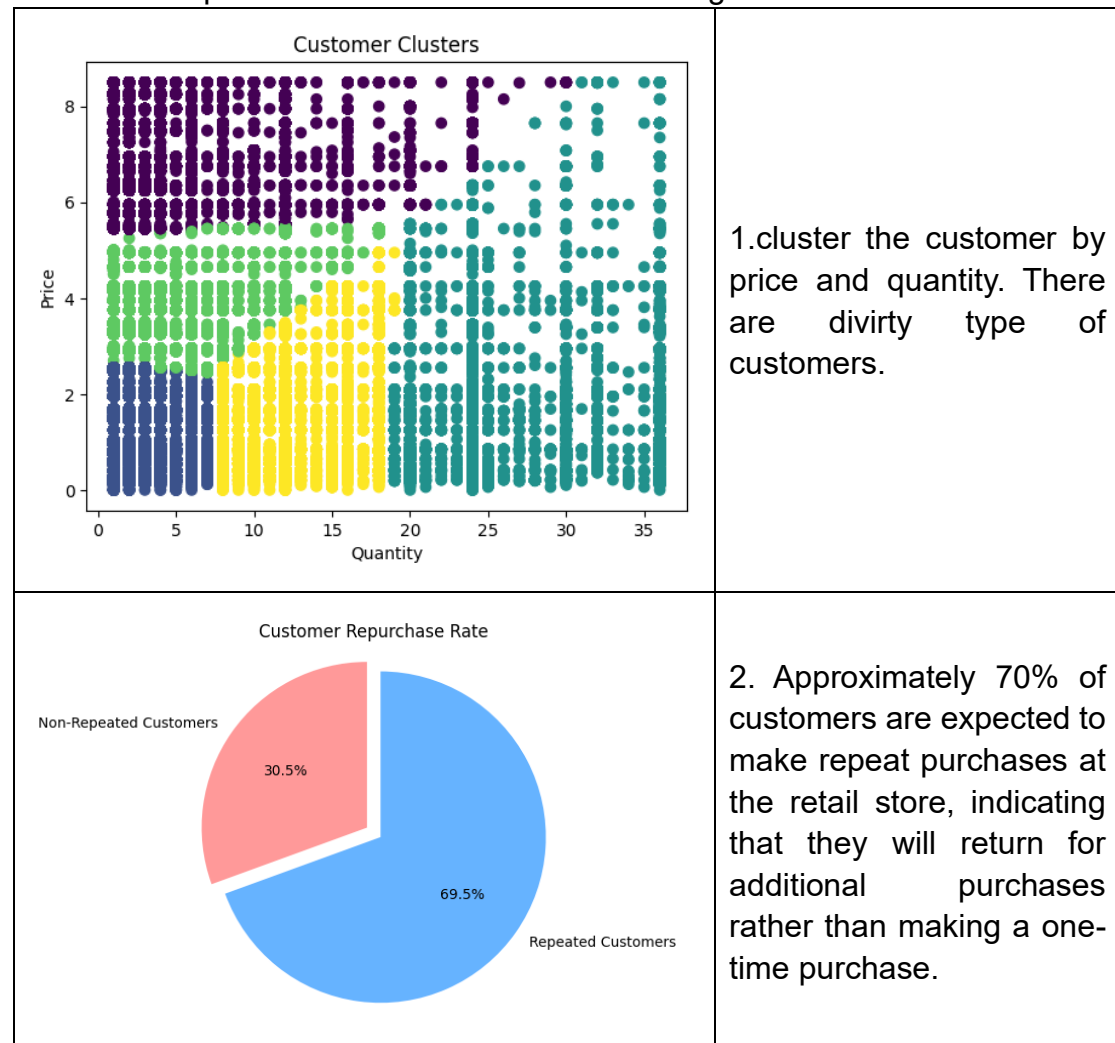


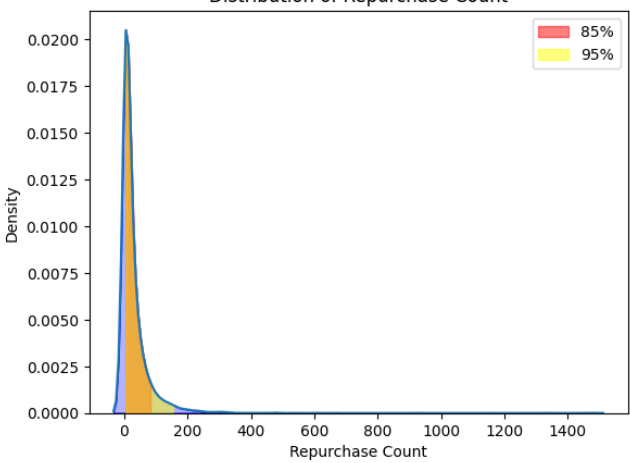
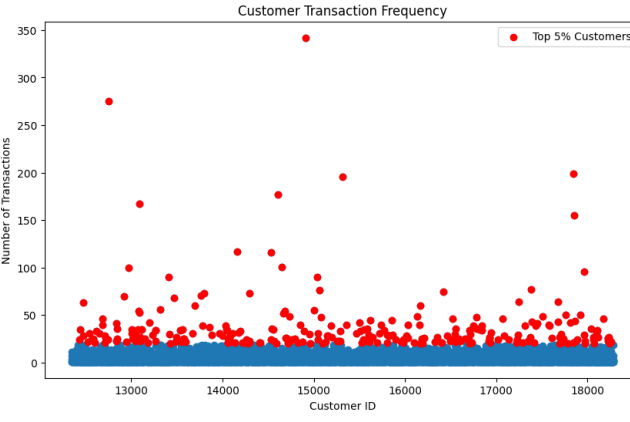
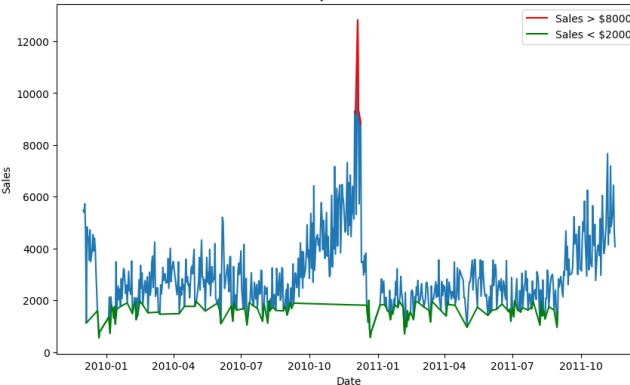
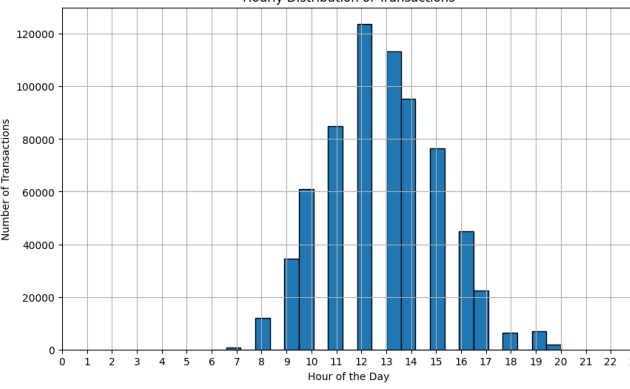
Q4. Recommendation and Business Analysis

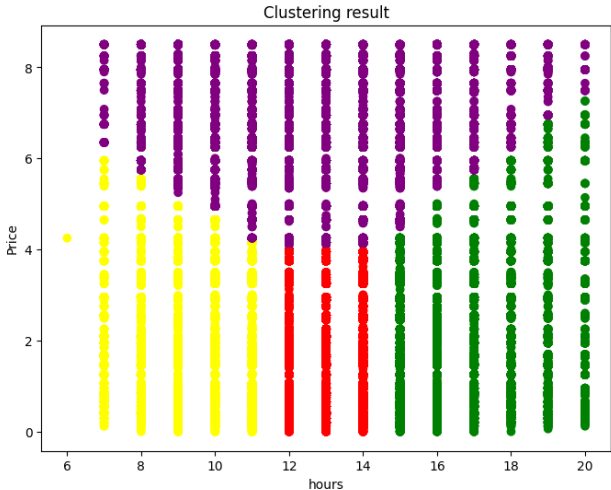
Preprocessing

- Remove NAN
- delete negative values
- keep values below 95% and remove values above 95%

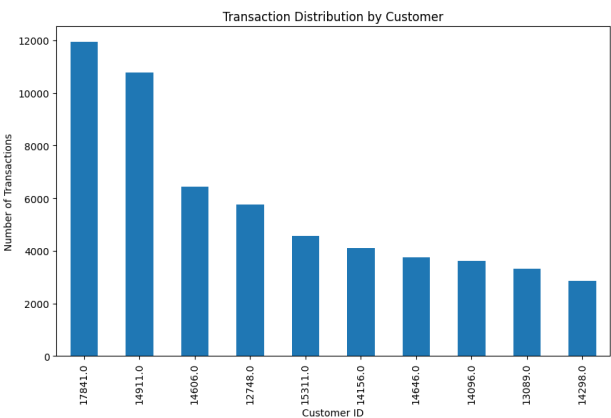
a. At least 8 pictures and at least 8 business insights



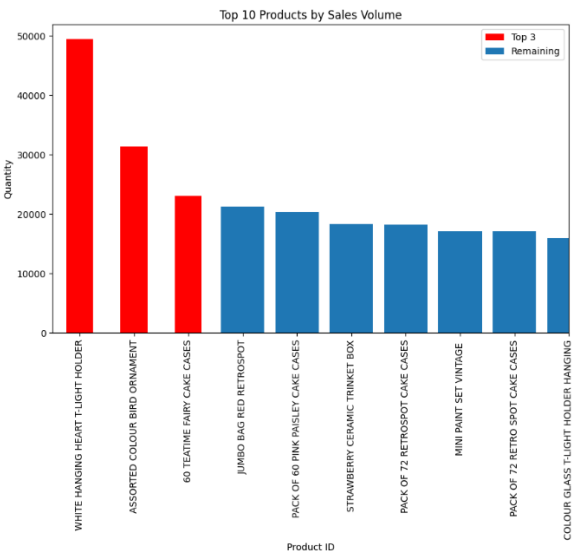
 <p>The chart, titled 'Distribution of Repurchase Count', shows the density of repurchase counts on the y-axis (ranging from 0.0000 to 0.0200) against the repurchase count on the x-axis (ranging from 0 to 1400). Two curves are plotted: an orange curve for the 85% group and a yellow curve for the 95% group. Both curves are highly right-skewed, peaking near zero repurchases and tapering off as the count increases. The 95% curve is slightly broader than the 85% curve.</p>	<p>3. Within the group of customers who make repeat purchases, approximately 95% of them are expected to make more than 150 purchases.</p>
 <p>The chart, titled 'Customer Transaction Frequency', displays the number of transactions on the y-axis (0 to 350) against customer ID on the x-axis (13000 to 18000). Red dots represent the 'Top 5% Customers'. Most dots are clustered below 50 transactions, with a few outliers reaching up to 350. A blue shaded area at the bottom represents the rest of the customer base.</p>	<p>4. The top 5 most loyal customers are primarily engaged in small quantity transactions.</p>
 <p>The chart, titled 'Daily Sales Trend', shows sales volume on the y-axis (0 to 12000) over time on the x-axis (from 2010-01 to 2011-10). A blue line represents daily sales, showing significant fluctuations and a major peak in early 2011. A green line represents sales below \$2000, which remains relatively flat and low throughout the period. A red line segment at the peak of the blue line indicates sales above \$8000.</p>	<p>5. There was a peak in transactions at the end of 2010, followed by continued fluctuations in transaction activity.</p>
 <p>The chart, titled 'Hourly Distribution of Transactions', shows the number of transactions on the y-axis (0 to 120000) against the hour of the day on the x-axis (0 to 23). The distribution is bell-shaped, starting near zero at midnight, rising to a peak of over 120,000 transactions around noon (hour 12), and then gradually declining towards the end of the day.</p>	<p>6. People tend to concentrate their purchases during the midday hours, and the distribution of these purchase times follows a normal distribution pattern.</p>



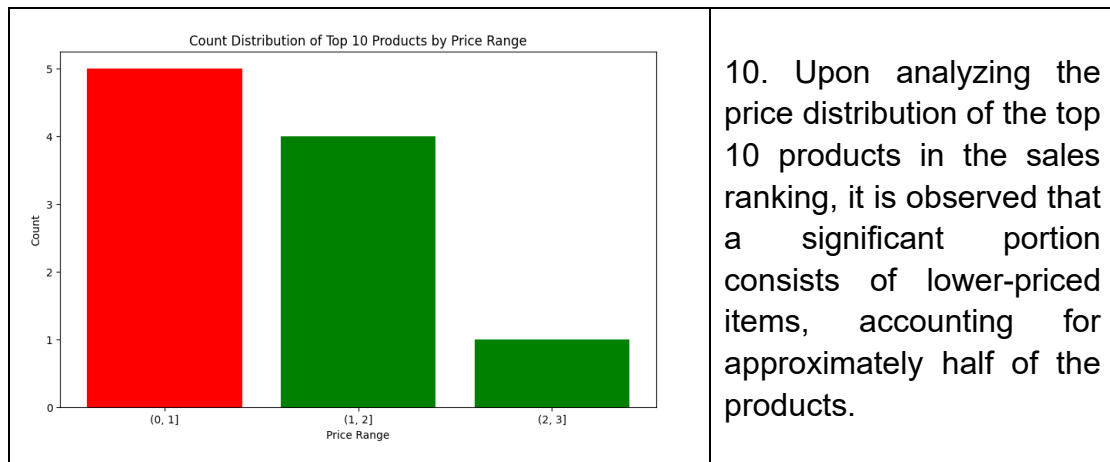
7. To cluster users based on different time periods and offer corresponding pricing for products during those time periods, such as offering significant discounts on fruits in the evening



8. The top 10 customers with the highest purchase volume are considered loyal customers.



9. Increasing the inventory quantity of the best-selling products can help improve sales performance.



b. The algorithm details, process, and results of association rule analysis, along with providing no fewer than 5 sales/recommendation suggestions.

Find The transaction count is 32,098, the number of product categories is 4,366, and the number of customers is 5,611.

Steps:

I converts product combinations into **one-hot encoded format** and uses the **Apriori** algorithm to find **frequent itemsets**. Then, based on the frequent itemsets, it calculates the **confidence, support, and lift of association rules**. Finally, the rules are sorted based on confidence, lift, and support to identify high-quality association rules.

Suggestions:

Most confident

- 1.item (22697) and (22699) should put together
- 2.item (21231) and (21232) should put together
3. item (21699) and (21697) should put together
- 4.item (21733) and (85123A) should put together
- 5.item (85099F) and (85099B) should put together

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhangs_metric
62	(22697)	(22699)	0.027136	0.030376	0.021466	0.791045	26.042005	0.020641	4.640345	0.988422
1	(21231)	(21232)	0.032463	0.052215	0.023896	0.736084	14.097159	0.022201	3.591243	0.960236
63	(22699)	(22697)	0.030376	0.027136	0.021466	0.706667	26.042005	0.020641	3.316583	0.991725
5	(21733)	(85123A)	0.046888	0.138856	0.032370	0.690365	4.971808	0.025859	2.781162	0.838166
13	(85099F)	(85099B)	0.040439	0.082310	0.024394	0.603236	7.328789	0.021066	2.312934	0.899944

sorted_rules_by_lift

1. some one buy item (22699) could have more chance to buy (22697)
2. some one buy item (22697) could have more chance to buy(22699)
3. some one buy item (21231) could have more chance to buy (21232)
4. some one buy item (82482) could have more chance to buy (82494L)
5. some one buy item(82494L) could have more chance to buy (82482)

63	(22699)	(22697)	0.030376	0.027136	0.021466	0.706667	26.042005	0.020641	3.316583	0.991725
62	(22697)	(22699)	0.027136	0.030376	0.021466	0.791045	26.042005	0.020641	4.640345	0.988422
1	(21231)	(21232)	0.032463	0.052215	0.023896	0.736084	14.097159	0.022201	3.591243	0.960236
47	(82482)	(82494L)	0.048632	0.051312	0.028787	0.591928	11.535952	0.026291	2.324808	0.960002
46	(82494L)	(82482)	0.051312	0.048632	0.028787	0.561020	11.535952	0.026291	2.167223	0.962713

sorted_rules_by_support:

1. The items (85123A) and (21733) are frequently purchased together.
2. The items (21733) and (85123A) are frequently purchased together.
3. The items (22384) and (20725) are frequently purchased together.
4. The items (20725) and (22384) are frequently purchased together.
5. The items (82482) and (82494L) are frequently purchased together.

4	(85123A)	(21733)	0.138856	0.046888	0.032370	0.233116	4.971808	0.025859	1.242838	0.927680
5	(21733)	(85123A)	0.046888	0.138856	0.032370	0.690365	4.971808	0.025859	2.781162	0.838166
29	(22384)	(20725)	0.053212	0.071936	0.028818	0.541569	7.528490	0.024990	2.024436	0.915909
28	(20725)	(22384)	0.071936	0.053212	0.028818	0.400606	7.528490	0.024990	1.579576	0.934387
47	(82482)	(82494L)	0.048632	0.051312	0.028787	0.591928	11.535952	0.026291	2.324808	0.960002

Reference

[1]<https://zhuanlan.zhihu.com/p/440147093> 关联规则