Final Report

Will They Reorder This Season?: Customer-SKU Reorder Prediction in Wholesale

MGSC 401
Statistical Foundations of Data Analytics

Shanez Aithadi
Ermina Ashraf
Daksh Chawla
Zaid Zubiedi

We initially explored a synthetic, one-year dataset with K-Means clustering, but we abandoned it because it lacked the richness needed to capture real reorder and seasonal patterns (Details in Exhibit 0A). This pushed us to refine our problem: we shifted to a real wholesale setting where we predict whether a customer will reorder a specific SKU. The new dataset has retail transactions from a UK-based giftware company that covers data from December 2009 to December 2011, where most of the company's clients are wholesalers.

Our main challenge was reframing the project. We ultimately decided to build three models: two supervised learning models using logistic regression, and one unsupervised learning model using K-Means for customer segmentation to refine our initial supervised approach.

**Preprocessing & feature engineering in Exhibit 0B.

Refined hypothesis: We expect the probability that a customer reorders a given SKU within **42 days** (Exhibit 0C) is not constant throughout the year. We expect this 42-day reorder probability to be higher in peak seasons like October/November as wholesalers begin stocking up for the holiday season. On the other hand, we expect reorders of specific SKUs to be lower in other months, as wholesalers will probably only replenish inventory throughout the year for more general gifting needs such as birthdays and smaller seasonal events.

By understanding how reorder probabilities change across seasons, the distributor can buy smarter, stock the right products at the right time, reduce stockouts during peak months, cut overstock in slow months, and time their marketing and operations around when customers are actually most likely to reorder.

Our Models

The first model was a logistic regression designed to analyze the reorder probability based on holidays, country, price and quantity (Exhibit 1A). The model produced a total of 6 meaningful predictors: price, all country indicators and secondary-holiday.

Price had the most meaningful impact, with a unit increase in price increasing the probability of reorder by 48%. When it comes to countries, a customer from Germany is 20% more likely to reorder than a French customer, and an Irish customer is 170% more likely. Countries in the 'other' category are 21% more likely to reorder than France-based customers, and those in the UK are 86% more likely than those in France.

The final significant variable is secondary holidays, which make wholesalers 40% less likely to reorder than when there is no holiday (Exhibit 1B).
The results are intuitive: wholesalers typically stock up before the holiday periods, reducing reorder likelihood immediately afterwards.

Despite meaningful coefficients, overall model performance was weak. The confusion matrix showed an almost even distribution of false positives and true positives as well as false negatives and true negatives (Exhibit 1c). The weakness is further reflected by the AUC score of 0.531 (Exhibit 1C).

To improve performance, we introduced new variables. The first was 'has_previous_purchase', a binary variable identifying whether a customer has ordered a specific SKU in the last 42 days. Since the main customers are wholesalers, we assume perhaps regular order cycles, which means if a customer has ordered in the past, they may be more likely to reorder in the same coming interval. The second addition was 'months', based on the assumption that wholesalers stock up heavily in October and November in preparation for holiday demand in December.

The second model identified a total of 18 meaningful predictors. The secondary holiday variable was no longer significant; however, the other variables from model 1 remained significant with similar odds ratios (Exhibit 2A). As expected, the presence of prior purchases increased the likelihood of a reorder by 72%. As for the months, relative to January, orders placed in February and March were less likely to be reordered. On the other hand, October and November purchases increase reorder likelihood by over 200% relative to January. December purchases also increased reorder likelihood by 145% relative to January.

The accuracy scores also increased, with sensitivity going from 0.520 in the previous model to 0.640 in the second one. Most importantly, the AUC score rose to 0.658, which is more consistent with scores of comparable research using a similar classification structure and AUC-based evaluation (Liu et al., 2016).

The third and final model is a logistic regression with customer segmentation.

In a wholesale setting, the assumptions taken up to this point are unrealistic: different buyers, including small boutique stores, medium retailers, and large wholesalers, exhibit drastically different reorder behaviours. While our Model 2 improved performance by incorporating recency and seasonal patterns, it still treated all customers as if they behaved the same. To capture meaningful differences in customer purchasing behaviour, our Model 3 incorporated customer segments generated through K-Means clustering on customer-level purchasing features.

We first aggregated the dataset at the customer level, computing total quantity purchased, total spend, number of invoices, and number of unique SKUs purchased. These four features summarize core purchasing behaviour (volume, value, order frequency, and product breadth). We standardized these features and applied K-Means with K=4, producing four distinct customer segments. These clusters were then merged back into the transaction-level dataset so that each SKU purchase was associated with its customer's segment. We performed K-Means clustering on customer-level features and merged the resulting segment labels back into the transaction-level dataset (Exhibit 3A).

The 4 clusters we ended up with revealed meaningful behavioural differences. Cluster 0 represented small/occasional buyers with low spend and low invoice frequency. Cluster 1 consisted of high-spend specialty retailers that purchased expensive SKUs with a high variety of SKUs. Cluster 2 captured moderate, steady buyers with consistent but mid-range purchasing behaviour, while Cluster 3 represented large wholesalers with extremely high spend and high order quantities. The scatter plot of spend vs quantity (Exhibit 3B) supported these interpretations, where clusters were clearly differentiated in scale and purchasing intensity.

We then converted the cluster label into dummy variables (cluster_1, cluster_2, and cluster_3), with Cluster 0 serving as the baseline. Model 3 includes all predictors from Model 2: price, previous purchase, holidays, country, and monthly dummy variables, as well as the new cluster indicators. The logistic regression coefficients (Exhibit 3C) showed two strong customer-segment effects: Cluster 1 and Cluster 3 had large, statistically significant positive coefficients (+1.12 and +1.05, respectively), indicating that these groups reorder far more frequently than the baseline. Cluster 2 showed a mild but positive effect (+0.16), consistent with our initial interpretation that they are moderate regular buyers.

Impact of Customer Segmentation: Model 2 vs. Model 3

Comparing Model 3 to Model 2, while the overall accuracy remained similar (0.630 in Model 3 vs. 0.615 in Model 2), the balance between sensitivity and specificity shifted significantly. Model 3 became more conservative in predicting reorders, with a decrease in sensitivity from 0.640 to 0.598, indicating that the model produced fewer false reorder predictions. At the same time, specificity increased to 0.664 (Exhibit 3E), indicating that Model 3 became more effective at identifying customers who would not reorder within the 42-day window. This is helpful for distributors, as the reduction in false positives enables them to avoid unnecessary stock allocation to customers who are unlikely to reorder.

Moving on to the ROC curve comparison (Exhibit 3D). Model 3's ROC curve lies above that of Model 2, across almost the entire threshold range, thus having a higher AUC score (0.677 vs. 0.655). This suggests that incorporating customer-segment information enables the model to more effectively distinguish between likely and unlikely reorders. The performance comparison chart (Exhibit 3E) summarizes these changes across all key statistical metrics: accuracy, precision, recall, F1 score, and AUC. This further highlights the incremental but consistent improvements attributed to segmentation. Overall, Model 3 produces more realistic reorder probabilities, thereby reducing the overconfident predictions observed in Model 2.

By incorporating behavioural segmentation on top of seasonality and recency, our final Model offers a more operationally useful tool, improving reliability in inventory planning and replenishment decisions. However, it is apparent that the improvement is modest rather than dramatic because short-term purchase and reorder behaviour contains a high degree of inherent randomness. This point is highlighted in works such as Fader and Hardie (2009), which show that even well-specified customer models face substantial unpredictability in non-contractual purchasing behaviour.

Final Results and Interpretations: **Model Performance**

Both our models showed a predictive power that is moderate but meaningfully better than random guessing. This proves that, despite wholesale reorder behaviour containing significant firm-specific noise, the model variables still capture important systematic signals (Exhibit 4A for full performance metrics). Overall, both models show that the reorder patterns are not random and that there are existing predictable patterns in customers, like price, seasonal cycles and country, which guide their purchasing behaviour (Exhibit 4B for main drivers of reorder probability).

Recommendations

1) **Prioritize repeat-purchase, high-value SKUs**

- Use the price and purchase history of customers to find the SKUs that are important and have a high probability of being reordered quickly.
- Maintain a shorter reorder cycle and higher safety stock for these SKUs.

2) **Integrate model results into operational tools**

- Find the important SKUs from the predicted probability given by the model and integrate these into a system to make important daily operational decisions

3) **Embed seasonality into reorder policy**

- Increase baseline stock and decrease reorder limits for the months with the highest reorder probability within 45 days, which in our case are October and November.
- Additionally, we should also relax the limits on low probability months like February and March.

4) **Use customer clusters for customized service levels**

- For clusters with large positive coefficients, wholesalers should maintain a more proactive approach towards reordering and keeping inventory ready during peak seasons.

5) **Enhance data to improve model accuracy**

- In the future, data should include more operational variables like promotion, lead times, and existing stock to improve the R^2(explanatory power) and improve the predictive power of the model.

Appendix

Exhibit 0A:

Our initial PoC, "How Customers Shop in Seasons: Exploring Age, Gender, & Product Preferences in Sales," aimed to use K-Means clustering on a synthetic, one-year dataset to see how customers' habits changed across seasons and to identify groups like "holiday shoppers" or "loyal customers". However, upon further reflection, we didn't want to use synthetic data as it may lack the full complexity of real-world data. Moreover, the data provided was only for a single year, with only 1000 iterations and unique customer IDs, making it harder to track true reorders or seasonal patterns. We also lacked key information, like country and demographics, which limits cross-referencing with reality.

Exhibit 0B:
Preprocessing

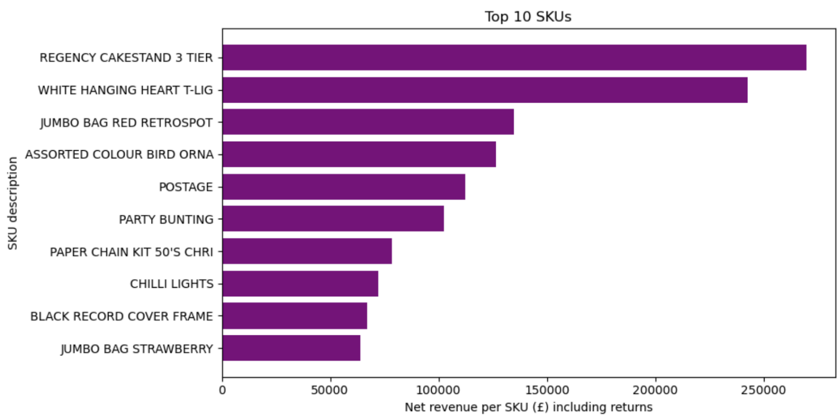
→ Started with 1M rows, and deleted rows with null values, missing customer IDs (~23% of data), as well as rows concerning return transactions of a SKU, to avoid problems while doing the project, leaving us with ~800K rows.
→ Converted the datetime column to just its date component to help with feature engineering.
→ Grouped by customer ID and sorted by SKU and invoice date to ensure chronological order.
→ Computed next purchase date and days until reorder to find average interval between orders of the same SKU & customer ID.
→ Used the median to find the y variable
Since the mean is pulled by outliers, use the median for the benchmark (brought up to be a more logical interval).
→ Computed previous purchase date (for the other model)
→ Computed days since last purchase.

Feature engineering:

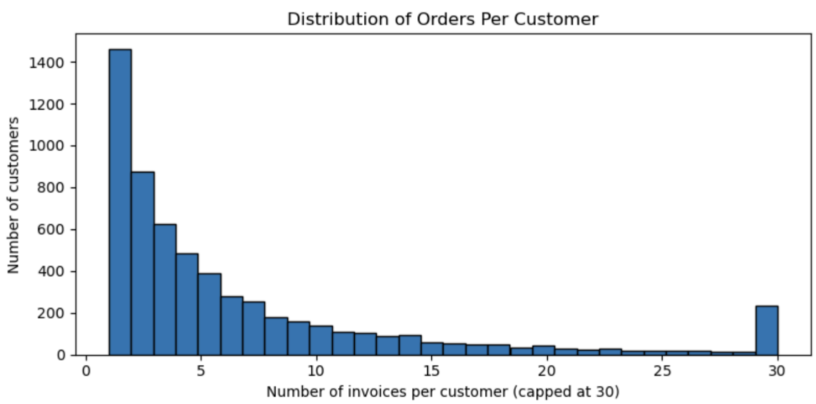
→ Started with 41 countries total, so we decided to create a new column called "country_simple", which contains the top four countries + all the other 37 combined together.
→ Created a new column for every month to facilitate the detection of seasonal patterns.

Exhibit 0C: We set 42 days as our baseline reorder window because it is the median time between successive purchases of the same SKU by the same customer in our data.

Exhibit 0D:
General overview of data.



Top 10 Most Popular Gift SKUs



Orders Per Wholesaler (customer)

Seasonality Patterns

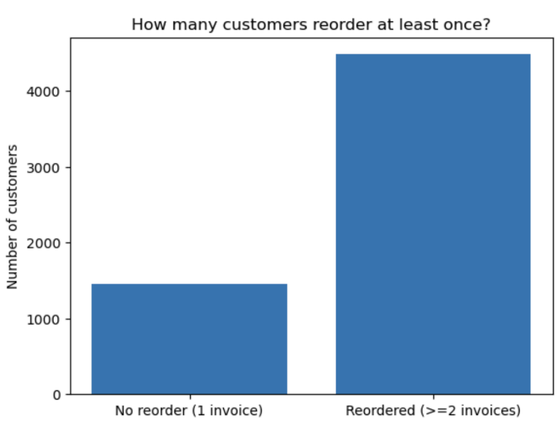


Exhibit 1A: Model 1 significant outputs and odds ratio

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]	odds_ratio
const	-0.608079	0.027015	-22.509327	3.363389e-112	-0.661027	-0.555132	0.544395
Price	0.393276	0.034267	11.476690	1.727690e-30	0.326114	0.460439	1.481828
Country_simple_Germany	0.198753	0.036380	5.463291	4.673865e-08	0.127450	0.270056	1.219880
Country_simple_Ireland	0.996641	0.033049	30.156716	8.758033e-200	0.931867	1.061415	2.709166
Country_simple_Other	0.196587	0.033123	5.935141	2.935938e-09	0.131668	0.261507	1.217242
Country_simple_UK	0.624258	0.027273	22.889221	5.949443e-116	0.570804	0.677712	1.866860
holiday_type_secondary_holiday	-0.431752	0.043843	-9.847740	7.010254e-23	-0.517682	-0.345822	0.649371

Exhibit 1B: Model 1 performance measures

Truth	**0**	**1**
Predicted		
0	85536	76892
1	81937	88665

```
Accuracy:    0.523
Sensitivity: 0.520
Specificity: 0.527
```

AUC: 0.531

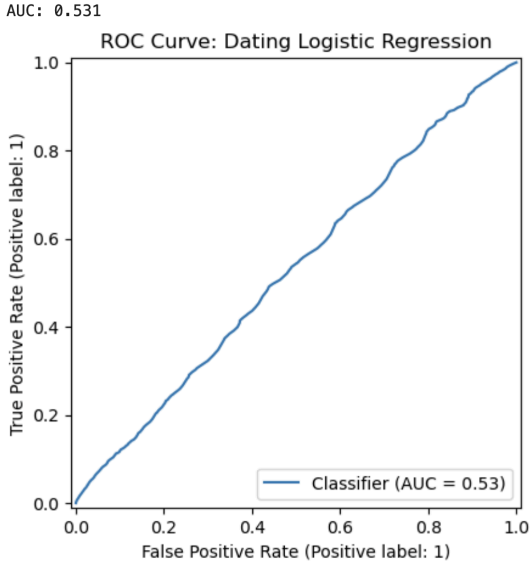


Exhibit 2A: Model 2 significant outputs and odds ratio

	coef	std err	z	P>|z|	odds_ratio
const	-1.4221	0.032	-44.706	0.000	0.241207
Price	0.3837	0.035	11.053	0.000	1.467705
has_prev_purchase	0.5448	0.008	65.083	0.000	1.724263
Country_simple_Germany	0.2026	0.038	5.379	0.000	1.224583
Country_simple_Ireland	0.9809	0.034	28.665	0.000	2.666855
Country_simple_Other	0.1764	0.034	5.145	0.000	1.192915
Country_simple_UK	0.5697	0.028	20.171	0.000	1.767737
month_2	-0.0488	0.021	-2.357	0.018	0.952372
month_3	-0.1114	0.019	-5.752	0.000	0.894581
month_4	0.0464	0.020	2.314	0.021	1.047493
month_5	0.0842	0.019	4.362	0.000	1.087846
month_6	0.1118	0.019	5.797	0.000	1.118289
month_7	0.2590	0.020	13.258	0.000	1.295634
month_8	0.2922	0.020	14.587	0.000	1.339371
month_9	0.5053	0.019	26.879	0.000	1.657483
month_10	1.1331	0.019	60.917	0.000	3.105268
month_11	1.2383	0.018	67.837	0.000	3.449744
month_12	0.8969	0.018	50.634	0.000	2.451990

Exhibit 2B: Model 2 performance measures

Truth	**0**	**1**
Predicted		
0	117862	77497
1	49611	88060

```
Accuracy:    0.618
Sensitivity: 0.640
Specificity: 0.603
```

```
AUC: 0.658
```

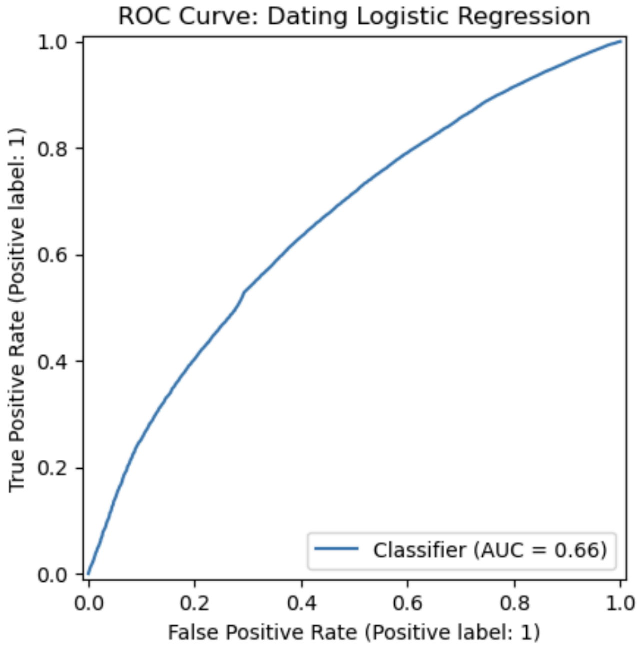


Exhibit 3A:

It is important to note that the K-Means clustering was performed in a separate notebook, and the resulting customer–cluster mapping was imported via CSV (under the name "customer_clusters.csv") into the main notebook.

```python
# Build customer-level features
df_pos = df[df["Quantity"] > 0].copy()

cust_features = df_pos.groupby("Customer ID", as_index=False).agg(
    total_quantity=("Quantity", "sum"),
    total_spend=("Price", "sum"),
    unique_skus=("StockCode", "nunique"),
    num_invoices=("Invoice", "nunique")
)
```

```python
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans

# Scale numeric features
scaler = StandardScaler()
cust_scaled = scaler.fit_transform(
    cust_features[["total_quantity","total_spend","unique_skus","num_invoices"]]
)

# Run K-Means (choose number of clusters, e.g., 4)
kmeans = KMeans(n_clusters=4, random_state=0, n_init=10)
cust_features["cluster"] = kmeans.fit_predict(cust_scaled)
```

```python
cluster_map = cust_features[["Customer ID", "cluster"]]
cluster_map.to_csv("customer_clusters.csv", index=False)
```

```python
cluster_map.to_csv("C:/Users/zaida/Documents/customer_clusters.csv", index=False)
```

Exhibit 3B:

Customer Segments Visualized (Log-Scaled Spend vs Quantity)

Exhibit 3C: Logistic Regression Coefficients (Model 3)

	coef	std err	z	P>|z|
const	-1.4750	0.032	-45.414	0.000
Quantity	-0.0074	0.005	-1.558	0.119
Price	0.5131	0.039	13.274	0.000
has_prev_purchase	0.5289	0.008	62.641	0.000
Country_simple_Germany	0.2218	0.037	5.944	0.000
Country_simple_Ireland	-0.0098	0.037	-0.264	0.792
Country_simple_Other	-0.0442	0.035	-1.281	0.200
Country_simple_UK	0.4408	0.028	15.743	0.000
month_2	-0.0609	0.021	-2.919	0.004
month_3	-0.0753	0.019	-3.868	0.000
month_4	0.0851	0.020	4.222	0.000
month_5	0.1246	0.019	6.414	0.000
month_6	0.1392	0.019	7.176	0.000
month_7	0.2794	0.020	14.182	0.000
month_8	0.3228	0.020	15.994	0.000
month_9	0.5201	0.019	27.366	0.000
month_10	1.2186	0.019	64.318	0.000
month_11	1.2105	0.018	65.532	0.000
month_12	0.9070	0.018	50.690	0.000
holiday_type_secondary_holiday	0.0807	0.045	1.796	0.072
cluster_1	1.1212	0.016	68.017	0.000
cluster_2	0.1604	0.009	18.414	0.000
cluster_3	1.0517	0.017	60.623	0.000

Exhibit 3D: ROC Curve Comparison (Model 2 vs Model 3)

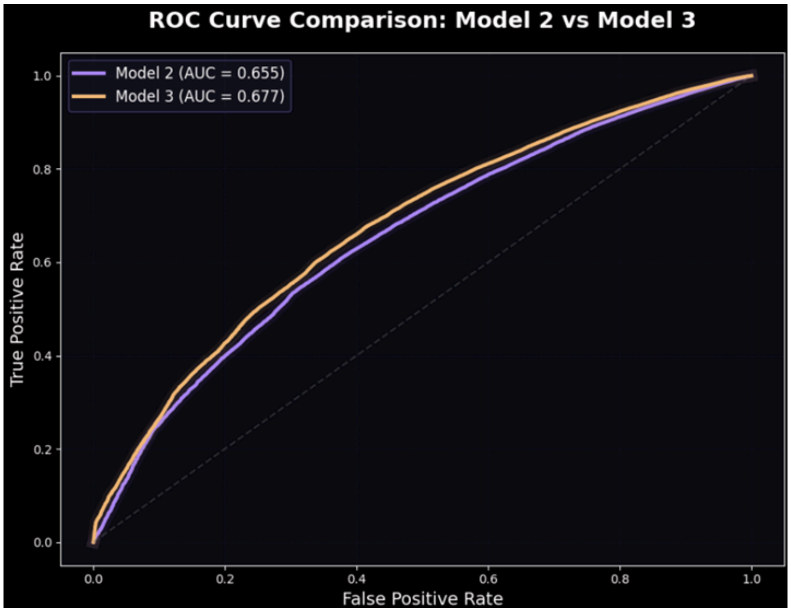


Exhibit 3E: Performance Comparison (Model 2 vs Model 3)

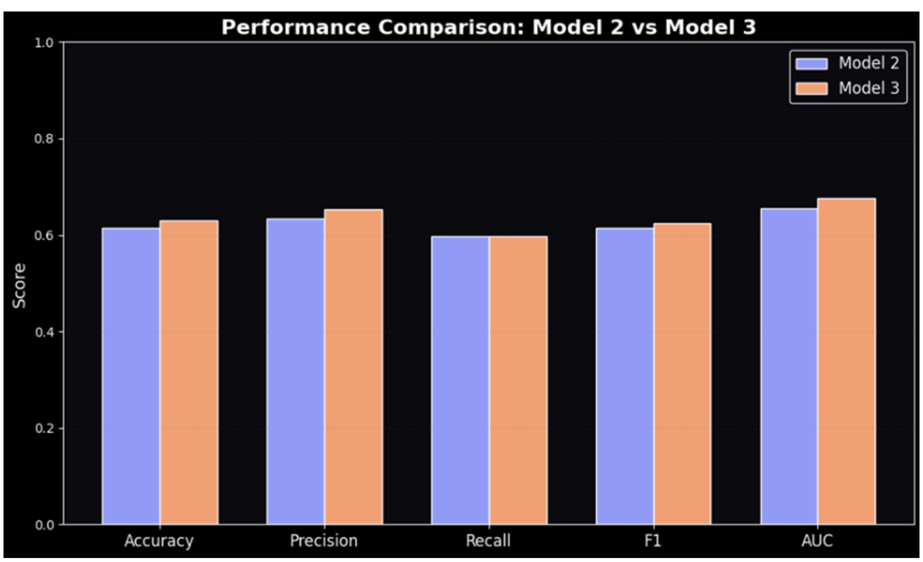


Exhibit 3F: Confusion Matrix & Metrics (Model 3)

```
=== Model 3 Metrics (with Clusters) ===
Accuracy:  0.630
Precision: 0.653
Recall:    0.598
F1 Score:  0.624
ROC-AUC:   0.677

Confusion Matrix:
[[107514  54423]
 [ 68765 102328]]

ISLP Confusion Table:
Truth           0       1
Predicted
0          107514   68765
1           54423  102328
```

```
Model 3 Accuracy:    0.630
Model 3 Sensitivity: 0.598
Model 3 Specificity: 0.664
```

Exhibit 4A:

Model 2 (Logistic Regression):

- **Accuracy → 0.615**
- **Recall/Sensitivity →0.633**
- **Specificity → 0.598**
- **ROC-AUC → 0.655**

Model 3 (Logistic Regression+ Clustering)

We can see improvement in Model 3 AUC and specificity, which shows that adding the customer-based clustering marginally enhanced the explanatory power of our model and helped us understand the behavioural patterns and operational drivers of customers. This suggests that the primary driver of reorder likelihood is product and temporal characteristics, with the customer specificity playing a secondary role.

Exhibit 4B:

Main Drivers of Reorder Probability

1) **Prior Purchase (has_prev_purchase)**
    - Throughout both models, the prior-purchase variable is one of the most important positive predictors of reorder within the given 42 days. This tells us that the customers who have recently purchased a SKU are more likely to repurchase it over the short term.
    - This confirms why these SKUs are the most valuable to retailers and are thus reordered by wholesalers more frequently.
2) **Price**
    - Price is another variable that is one of the most important predictors of reorder. The model assigns a sizable and positive coefficient, which shows that higher-priced SKUs are the ones most probable to be reordered quickly.
    - This coefficient given by the model might be confusing, as price is intuitively taken to be negatively correlated with purchasing demand. However, in this case, higher-value SKUs refer to the bundled higher-margin items that wholesalers consider strategically important. These SKUs are restocked and ordered more frequently by the wholesalers.
3) **Seasonality**
    - Earlier months (February and March) are found to have a lower reorder probability.
    - The reorder probability increases steadily from late spring to summer.
    - The model shows that September and November have the largest positive effect, and October and November have the largest coefficient.

These findings follow the common intuition that the retailers/wholesalers stock up before mega events like Black Friday and Christmas, leading to shorter reorder gaps and more frequent orders.

4) **Customer Clusters**

- In Model 3, some clusters have higher reorder probabilities than the baseline cluster.
- Some customer segments behave as high-intensity buyers, reordering the same SKUs frequently over the short term
- Whereas other segments act more randomly and have slower reorder patterns.

This segmentation explains why Model 3 performs better than Model 2 as we include the behavioural aspects into the model as well.

References

Liu, G., Nguyen, T., Zhao, G., Zha, W., Yang, J., Cao, J., Wu, M., Zhao, P., & Chen, W. (2016). *Repeat buyer prediction for E-Commerce*. https://www.semanticscholar.org/paper/Repeat-Buyer-Prediction-for-E-Commerce-Liu-Nguyen/e89d0012e0d000c55f5ae926fb28bf670ce79dfe

Fader, P.S., & Hardie, B.G. (2009). Probability Models for Customer-Base Analysis. *Journal of Interactive Marketing, 23*, 61-69.https://www.semanticscholar.org/paper/Probability-Models-for-Customer-Base-Analysis-Fader-Hardie/d43972cf774649a3cb5320dc4c52323ef964b206

Dataset: http://archive.ics.uci.edu/dataset/502/online+retail+ii