Feature Selection ML

Rational for feature selection

# Feature Justification

The following table looks at the features in the model and justifies them A significance level of 0.015 wash was chosen as the cutoff point.

|  |  |  |  |
| --- | --- | --- | --- |
| Feature Label | Feature Significance | Kept in Model | Rational |
| media\_type\_labeled | 0.258646 | Yes | These are the types of media campaigns. Each type is labelled as a factor. |
| SRP | 0.069932 | Yes |  |
| net\_weight | 0.068772 | Yes | This is the net weight of items. Smaller units are likely more likely to be sold. This might be because smaller units tend to be featured as on sale’ in media campaigns. |
| units\_per\_case | 0.059543 | Yes | This is the number of units per case may predict the customer acquisition cost to a similar degree to smaller items. |
| sales\_to\_cost\_ratio | 0.055446 | Yes | This is the substitute for the margins. |
| total\_children | 0.042192 | Yes | The number of children may incur additional purchases as one customer (parent) may represent multiple customers. Therefore the number of children is likely influencing how easily a customer is to be acquired as they need additional criteria to account for their children (i.e. a candy bar for them but toys or legos for their children) |
| avg\_cars\_at home(approx) | 0.034914 | Yes | The approximate average number of cars may be a substitute for relative mobility and relative wealth. |
| meat\_to\_total\_ratio | 0.032851 | Yes | This is the ratio of the square feet that the meat section takes up of the total. Meats may be perceived as luxury items that bring in more business and encourage customers to spend more. |
| num\_children\_at\_home | 0.025679 | Yes | The number of children at home may be colinear and similar to the total number of children, where more children mean, torn purchases are attributed to a single customer. |
| unit\_sales(in millions) | 0.024489 | Yes | This is the count of units sold. The more units sold the more. |
| gender\_M | 0.015762 | Yes | Men may be more difficult customers to acquire than women. |
| houseowner\_Y | 0.015488 | Yes | Homeowners are customers who may be paying off mortgages and thus have less disposable income than renters. Their cost of acquisition is higher as a result of requiring more convincing. |
| recyclable\_package | 0.012974 | **No** | The product packaging may impact the purchase based on the type of consumer (i.e. eco-conscious) but is unlikely to directly affect the acquisition cost and instead helps with the public perception. |
| marital\_status\_S | 0.012926 | **No** | Single customers likely have spending habits that are independent of promotions, and media exposure as their disposable cash flows are likely much lower than a married of co-habituating couple. |
| low\_fat | 0.011777 | **No** | Products with low fat are likely not directly involved in media campaigns and therefore do not factor into customer acquisition. |
| store\_state\_OR | 0.011029 | **No** | This geography is too narrow to impact the overall acquisition cost significantly. |
| education\_High School Degree | 0.010666 | **No** | The education of a customer is not likely to affect the customer acquisition cost as education is not likely a predictor of the susceptibility to media or a significant predictor in changing a person’s habits. |
| occupation\_Skilled Manual | 0.010329 | **No** |  |
| occupation\_Manual | 0.010317 | **No** |  |
| education\_Partial High School | 0.009973 | **No** | The education of a customer is not likely to affect the customer acquisition cost as education is not likely a predictor of the susceptibility to media or a significant predictor in changing a person’s habits. |
| avg. yearly\_income\_30-50K | 0.009909 | **No** |  |
| food\_family\_Food | 0.009784 | **No** | The type of food available or purchased is likely widely distributed and highly variable leading to noisy data and a tenuous connection ot the customer acquisition cost. |
| occupation\_Professional | 0.009708 | **No** | A customer’s occupation is not likely to affect the customer acquisition cost as this is not a guarantee of income or disposable income. There is likely a wide distribution of customer’s incomes. |
| avg. yearly\_income\_50-70k | 0.009368 | **No** | Customers who earn below 90K are slightly more consistent as a measure of predicting customer acquisition cost, but likely have their own habits that are independent of media publicity. |
| member\_card\_Normal | 0.008746 | **No** | The membership card is likely independent of the acquisition cost as consumers receive it because of their existing habits and not because of the media. In other words, if a customer has a gold card it is likely that because they are a recurring customer, therefore, their behavior and, thus, acquisition cost will not significantly change. |
| occupation\_Management | 0.008535 | **No** | A customer’s occupation is not likely to affect the customer acquisition cost as this is not a guarantee of income or disposable income. There is likely a wide distribution of customer’s incomes. |
| food\_family\_Non-Consumable | 0.008311 | **No** | The type of product available or purchased is likely widely distributed and highly variable leading to noisy data and a tenuous connection ot the customer acquisition cost. |
| coffee\_bar | 0.008192 | **No** | The type of facilities available is similar to the type of store and thus should not have much of an impact on customer acquisition unless these items are speciflcally featured in media campaigns. |
| member\_card\_Golden | 0.008084 | **No** | The membership card is likely independent of the acquisition cost as consumers receive it because of their existing habits and not because of the media. In other words, if a customer has a gold card it is likely that because they are a recurring customer, therefore, their behavior and, thus, acquisition cost will not significantly change. |
| member\_card\_Silver | 0.008034 | **No** | The membership card is likely independent of the acquisition cost as consumers receive it because of their existing habits and not because of the media. In other words, if a customer has a gold card it is likely that because they are a recurring customer, therefore, their behavior and, thus, acquisition cost will not significantly change. |
| video\_store | 0.007726 | **No** | The type of facilities available is similar to the type of store and thus should not have much of an impact on customer acquisition unless these items are speciflcally featured in media campaigns. |
| avg. yearly\_income\_70-90K | 0.007638 | **No** |  |
| education\_Partial College | 0.007030 | **No** | The education of a customer is not likely to affect the customer acquisition cost as education is not likely a predictor of the susceptibility to media or a significant predictor in changing a person’s habits. |
| sales\_country\_Mexico | 0.007002 | **No** | This geography is too narrow to have a significant impact on the overall acquisition cost. |
| florist | 0.006593 | **No** | The type of product available or purchased is likely widely distributed and highly variable leading to noisy data and a tenuous connection to the customer acquisition cost. |
| store\_state\_Zacatecas | 0.006563 | **No** | This geography is too narrow to have a significant impact on the overall acquisition cost |
| store\_state\_WA | 0.006136 | **No** | This geography is too narrow to have a significant impact on the overall acquisition cost. |
| store\_state\_CA | 0.005533 | **No** | This geography is too narrow to have a significant impact on the overall acquisition cost. |
| prepared\_food | 0.005511 | **No** | The type of food available or purchased is likely widely distributed and highly variable leading to noisy data and a tenuous connection ot the customer acquisition cost. |
| salad\_bar | 0.005409 | **No** | The type of food available or purchased is likely widely distributed and highly variable leading to noisy data and a tenuous connection ot the customer acquisition cost. |
| education\_Graduate Degree | 0.005250 | **No** | The education of a customer is not likely to affect the customer acquisition cost as education is not likely a predictor of the susceptibility to media or a significant predictor in changing a person’s habits. |
| store\_type\_Supermarket | 0.005224 | **No** | The type of store is likely not a differentiator for customers seeking specific products or types of ‘quick service’ products. Customers likely do no understand or prioritize the different in-store types |
| sales\_country\_USA | 0.005213 | **No** | This geography is too wide to have a significant impact on the overall acquisition cost |
| avg. yearly\_income\_130-150K | 0.005082 | **No** | Customers who earn above 90K may be their group with their habits that are independent of customer media acquisition methods. They likely emphasize consuming at non-convenient stores. Their habits may differ for various reasons, such as health, spending habits, or perception. |
| avg. yearly\_income\_110-130K | 0.004942 | **No** | Customers who earn above 90K may be their group with their habits that are independent of customer media acquisition methods. They likely emphasize consuming at non-convenient stores. Their habits may differ for various reasons, such as health, spending habits, or perception. |
| avg. yearly\_income\_90k 110K | 0.004701 | **No** | Customers who earn above 90K may be their group with their habits that are independent of customer media acquisition methods. They likely emphasize consuming at non-convenient stores. Their habits may differ for various reasons, such as health, spending habits, or perception. |
| store\_state\_Yucatan | 0.004219 | **No** | This geography is too narrow to have a significant impact on the overall acquisition cost |
| store\_type\_Gourmet Supermarket | 0.003769 | **No** | The type of store is likely not a differentiator for customers seeking specific products or types of ‘quick service’ products. Customers likely do no understand or prioritize the different in-store types. |
| avg. yearly\_income\_$150K + | 0.003287 | **No** | Customers who earn above 90K may be their group with their habits independent of customer media acquisition methods. They likely emphasize consuming at non-convenient stores. Their habits may differ for various reasons, such as health, spending habits, or perception. |
| store\_type\_Mid-Size Grocery | 0.003101 | **No** | The type of store is likely not a differentiator for customers seeking specific products or ‘quick service’ products. It is likely that customers do no understand or prioritize the different in-store type. |
| store\_state\_DF | 0.002109 | **No** | This geography is too narrow to have a significant impact on the overall acquisition cost. |
| tore\_state\_Guerrero | 0.001901 | **No** | This geography is too narrow to have a significant impact on the overall acquisition cost |
| store\_type\_Small Grocery | 0.001693 | **No** | This geography is too narrow to have a significant impact on the overall acquisition cost |
| store\_state\_Veracruz | 0.001573 | **No** | This geography is too narrow to have a significant impact on the overall acquisition cost |
| store\_state\_Jalisco | 0.000418 | **No** | This geography is too narrow to have a significant impact on the overall acquisition cost |