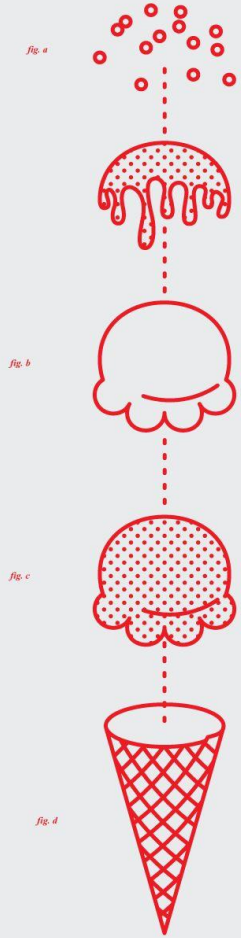
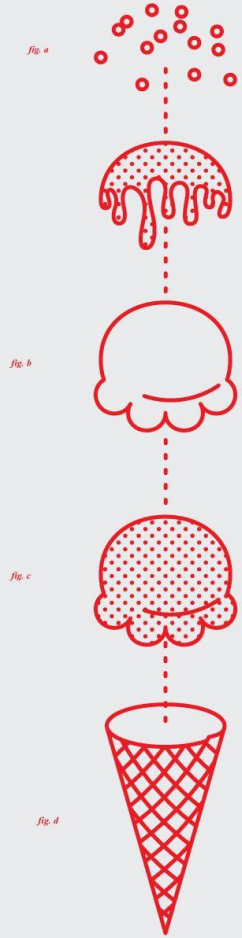


Cross-Sell for HVC



Agenda

- Business Objective
- Recommendations
- Data Dictionary
- ERD
- Codeflow chart
- Function - Next Purchase
- Function - Overall Purchases

The Objective

The mission is to provide two recommendation tools with the highest achievable accuracy for HVC Ice Cream. These two tools will aid in predicting

(1) which product will a customer buy next

(2) all the products a customer will buy in the dependent period

Recommendations: Model 1

Product Predictions:

CustomerID	Predictions
70849559	23678
20785084667436	23678
19455464224844	23685
22681060166636	23678
1404616	23685
1402180	23678
21983352286840	23678
1399764	23716
1248668	23678
726691	23678

Top 10 Products:

ProductID	Frequency
23678	11459
23716	8269
23685	7231
23734	4159
23714	3021
8609011	2962
20296288686840	2731
21792988536842	2592
23721	2545
23686	2021

Recommendations: Model 2

CustomerID	Products
1251478	23704 23690 23681 23687 23769 23678 8609011 59894425 49267982 71927445
62220789	21341176376844 46260776 18326144633838 23734 23678 23679 8609011
1459122	21341176376844 46260776 18326144633838 23734 23678 23679 8609011
1250285	21341176376844 46260776 18326144633838 23734 23678 23679 8609011
2.25167E+13	21341176376844 46260776 18326144633838 23734 23678 23679 8609011
1246731	21341176376844 46260776 18326144633838 23734 23678 23679 8609011
876247	21341176376844 46260776 18326144633838 23734 23678 23679 8609011
2.19329E+13	21341176376844 46260776 18326144633838 23734 23678 23679 8609011
62211041	21341176376844 46260776 18326144633838 23734 23678 23679 8609011

Next Product

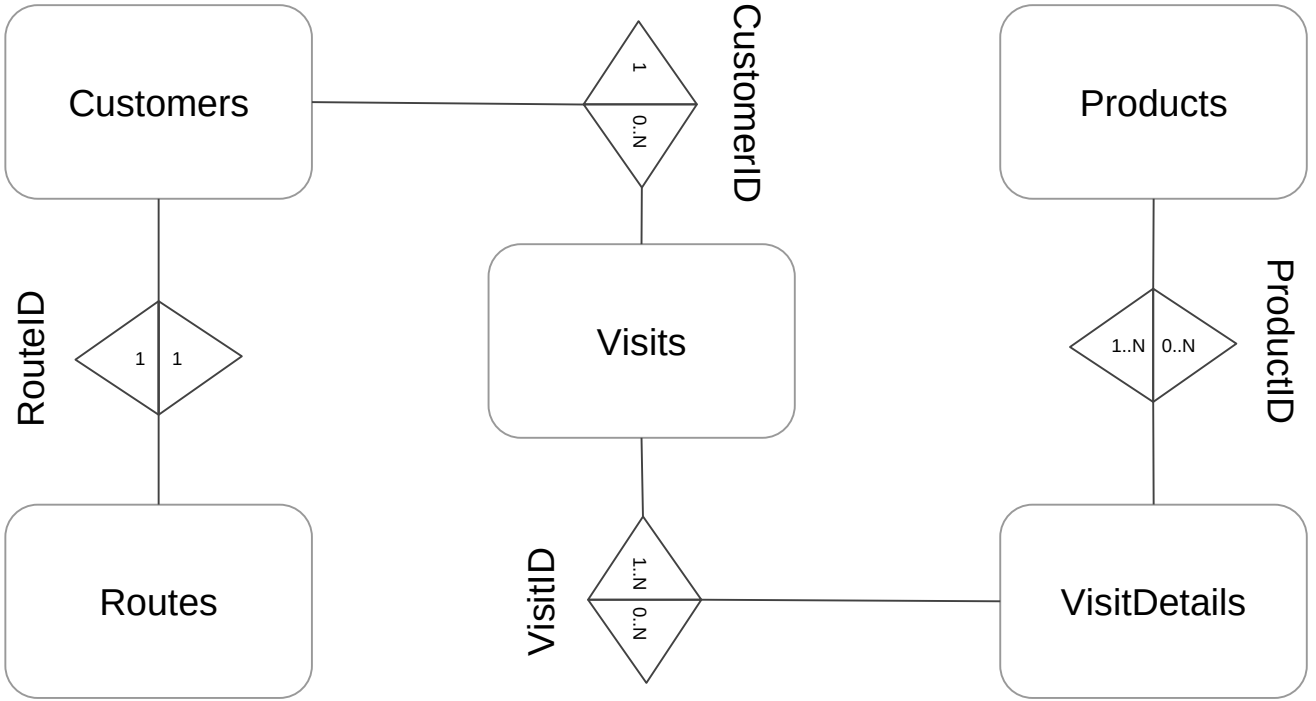
Variable	Description
numVisits	Total number of visits,count visitID
TotalAmount	Total amount spent
TotalItems	Total number of items purchased
AverageSpent	Average spent per item by visit ID
RecentNumproducts	Total number of unique products of most recent visit
[ProductID]	If each product was purchased in independent period = 1, else =0
FirstProdDep	First product customer purchased in dependent period

All Products

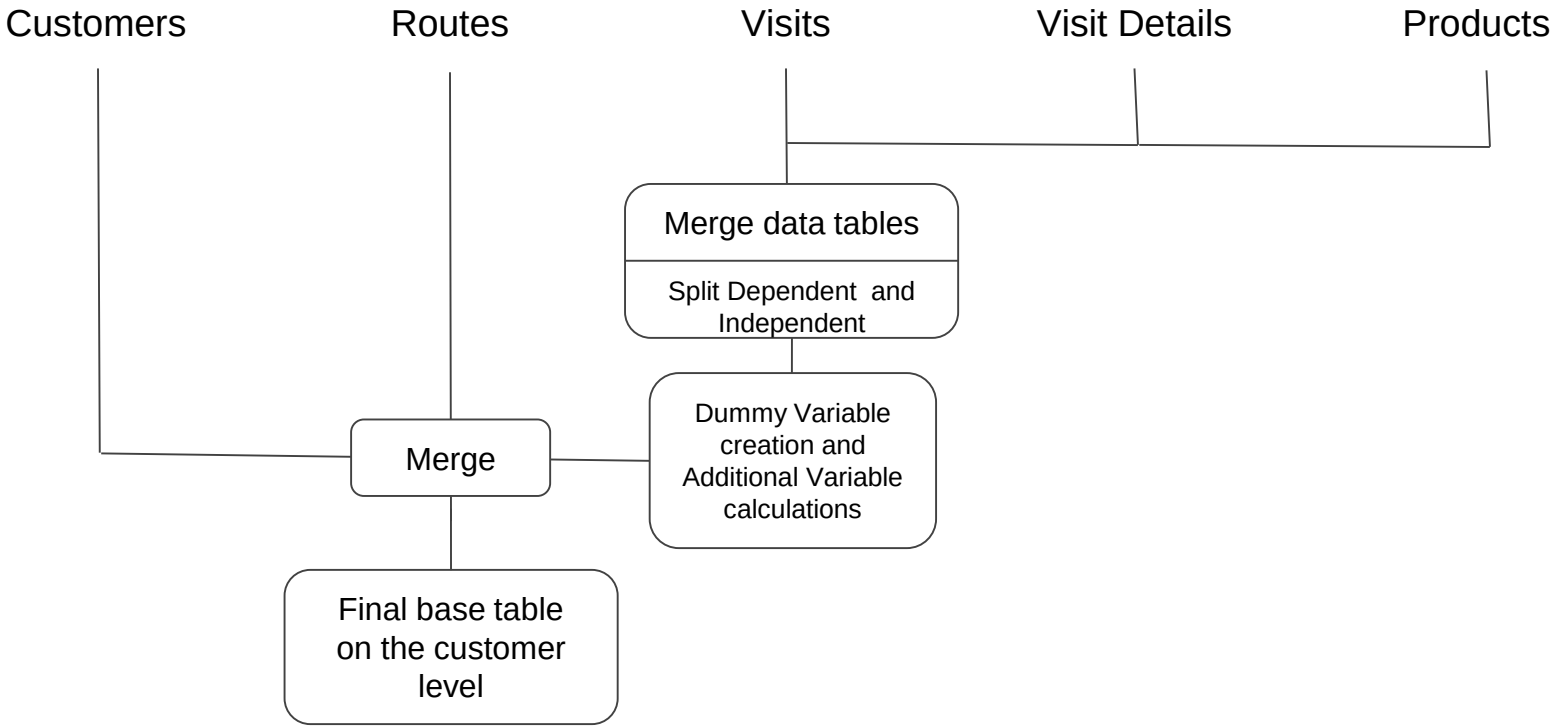
Variable	Description
numVisits	Total number of visits,count visitID
TotalAmount	Total amount spent
TotalItems	Total number of items purchased
AverageSpent	Average spent per item by visit ID
RecentNumproducts	Total number of unique products of most recent visit
[ProductIDIND]	If each product was purchased in independent period = 1, else =0
[ProductIDDep]	If customer purchased product in dependent period = 1, else= 0
Region	Customer Region
WeekOrder	Customer Week order
Day	Day of Visit

Variable	Description
MostRecentProdInd	Total number of visits,count visitID
SecMostRecentProdInd	Total amount spent
FirstProdBought	Total number of items purchased
MostRecentProdFam	Average spent per item by visit ID
SecRecentProdFam	Total number of unique products of most recent visit
FirstProdFam	If each product was purchased in independent period = 1, else =0
MostRecentRep	If customer purchased product in dependent period = 1, else= 0
MostFreqProduct	Most frequent product purchased in independent period
MostFreqSalesRep	Most frequent sales rep in independent period
MostFreqPayType	Most frequent pay type in independent period
CustomerType	Customer type

ERD


Customers
1
0..N
CustomerID
Products
RouteID
1
1
Visits
1..N
0..N
ProductID
Routes
VisitID
1..N
0..N
VisitDetails

Code Flowchart



Next Product Function

Runtime: 8 minutes

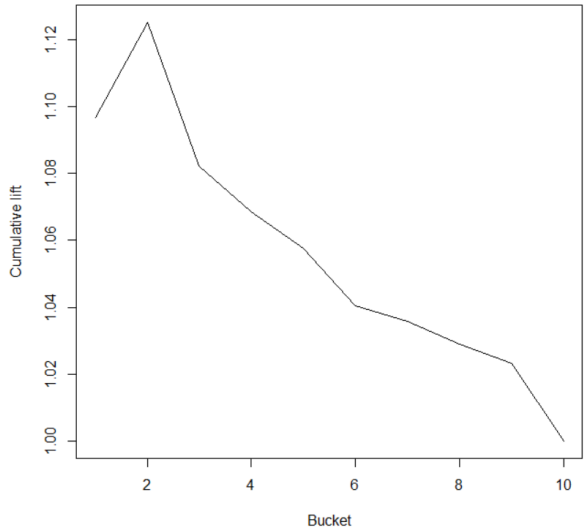
50% of that time is dedicated to building the model

Number of Lines: 250

Performance Metrics:

Accuracy: 23%

Top Decile Lift: 1.13



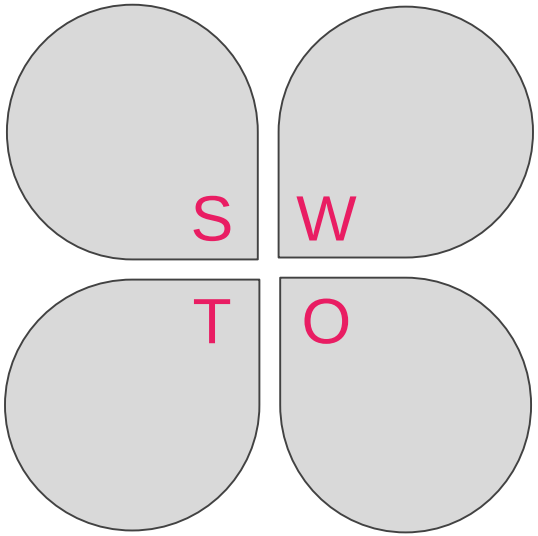
SWOT Analysis: Decision Tree

Strengths:

- Less variables: simpler model
- Accuracy increased to 23%

Weaknesses:

- Numeric variables so maybe missing an important predictor



Threats:

- Low Accuracy
- Only 1.13 times better than random when selecting 20% of our records

Opportunities:

- With adjusted predictors and base table format, could yield good results

All Products Function

Runtime: 40 minutes

Most of time is dedicated to find the optimum K value.

Number of Lines: ~400

Cutoff for probability: 0.1

Performance Metrics: 42% of times correctly predicting 1.

92% of times correctly predicting 0.

Top Decile Lift:1.1

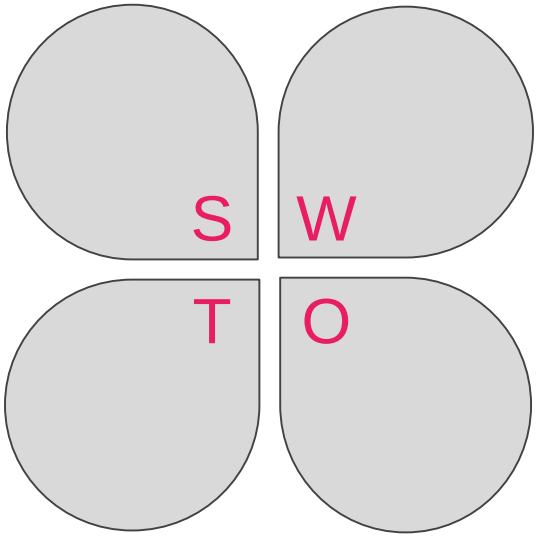
SWOT Analysis: K-Nearest

Strengths:

- TDL above 1
- 92% accurate on predicting no purchase

Weaknesses:

- 40 minute runtime



Threats:

- 42% accurate predicting correct product purchase

Opportunities:

- More flexibility with data type

Additional Analysis

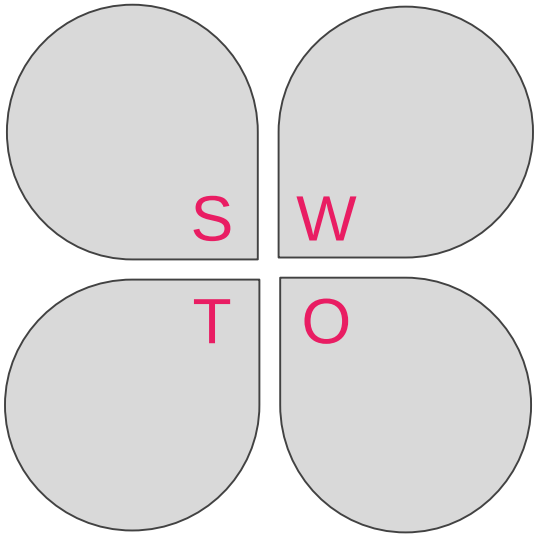
SWOT Analysis: Random Forest

Strengths:

- Runtime: ~20 min
- Tested on building just 10 trees

Weaknesses:

- Computational complexity



Threats:

- Cost of time for fitting is pretty significant when comparing to other models

Opportunities:

- Potential to increase performance metrics

Model 1: Random Forest Recommendations

CustomerID	Product
18864364293638	23766
724219	20004276638042 23740 49685718
1396505	23678 23783 23727
1248904	23788 23699 23714
1251622	17604188674640
871583	23695 58880561 23716
21622964227638	23684 21792988536842 23728
720608	23750 23734 23777

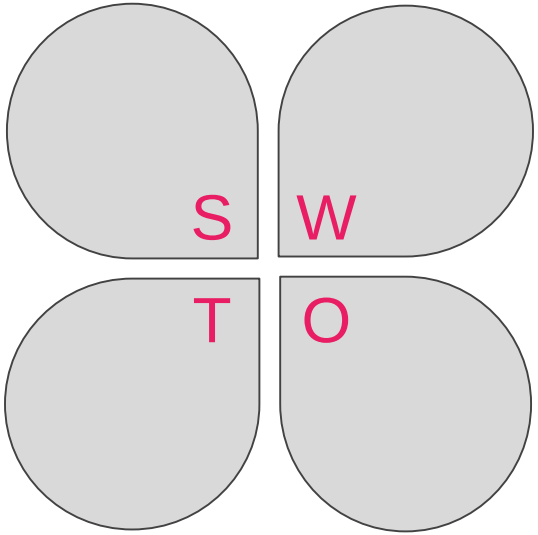
SWOT Analysis: Multivariate Regression

Strengths:

- Runtime: < 2min

Weaknesses:

- Wouldn't generate predictions across all observations
- Adjusted R-squared: 0.3544



Threats:

- Not as reliable as other models
- Amount of overfitting
- Some irrelevant metrics

Opportunities:

- Could use this to find variable importance and adjust model
- First Product ever purchased had high p-value