CSE 519 - Final Report - Retail Sales Data Analysis

Project Overview

The aim of every retail business is to attract new customers, retain existing customers and sell more to each customer. Through this project, we are trying to provide a local chain, Costello Ace with insights that will help them with their business.The insights will help them to make customer-centric, strategic and operational decisions.

Objective

In order to help Costello drive more sales, we have focussed our efforts on:

1. Loyalty Membership Analysis - Can more ACE membership help Costello with increased customer loyalty and revenue?
2. Time Series Modelling of Sales - Model to closely monitor customer buying patterns and forecast inventory levels for a store.
3. Product Placement Strategy - What items should be placed together in the stores. How does product placement strategy increase sales?
4. Membership and Product Promotion - Can we promote ACE membership subscription through product promotion based on Basket Analysis?

Dataset

The dataset provided by Costello‘s Ace spans over 4 years, 2015-2018. The following metadata follows after cleaning the data.

	2015-2016	2017-2018	Total
Product Transactions	14,725,900	15,545,819	30,036,805
Non Merchandize Transactions	901,571	1,782,201	2,918,686
Overall Sales Transactions'	15,627,471	17,328,020	32,955,491
Features	39	39	39
Number of Orders	6,554,802	7,069,644	13,378,901
Number of Stores	29	31	32
Categories of Products	45	49	57

Data Preprocessing

- Converting columns with type: **object** to **numeric.** For instance these columns : "Net Sales","Gross Margin", "$ Off Retail", "Line #" were treated with this method.
- Removed special characters from columns. E.g. % sign from Gross Margin column, $ or () sign from "Actual-Retail" column. We dealt with the brackets that indicated negative value in 'Actual-Retail' column separately.
- Removed rows which had all columns values as NaN or column names as their values. 24 rows were removed using this criteria.
- *Date* - Checked if the date column adhered to MM/DD/YYYY format.
- *Transaction Time* - Ensured that all the transaction time adhered to %H:%M:%S.
- Customer Number - Imputed the NaN values with *5 value which are customers without a registered account with the retailer.
- *Scanned UPC* - Used Item Number to impute the NaN values.
- *Department Code* - Used Department Name to impute the NaN values.
- For the preliminary analysis, we did not consider departments that were not related to merchandize. For instance - 'LIPA INSTANT REBATE', 'GIFT CARD'. We identified 32 such departments.
- All other columns didn't have a meaningful mapping for imputation, hence we used string NaN as a placeholder: Promo/Discount, Class Name, Class Code, Fineline Code, Loyalty ID.

Implementation and Insights

1. Loyalty Membership Analysis

For this objective we segregated the data into two categories: ACE members and NON_ACE members. Records with Loyalty ID were considered to be ACE members, NON_ACE otherwise. We initially checked the membership distribution at the company level **(Fig. 1)** as well as store level**(Fig. 2)**.

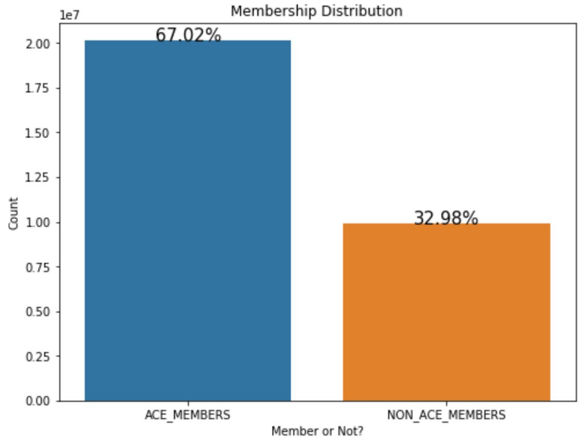


Fig (1) **Fig (2)**

In the four year period, we could see that ~33% of the transactions were done by non-members. From **(Fig. 2)** we see that most of the stores have higher ACE members than non-members with few exceptions like store **'D','T', 'a', 'z'**. Store 'a' has more non-members than members as it is a new store and like any other store, we can expect some delay in gaining potential members. Store 'z' 0 ACE_members in the data. This could possibly be because of very less data at our disposal for 'z' i.e. 27 records. ~33% was a potential to utilize and we wanted to base our recommendation to convert non-members to members based on further analysis.

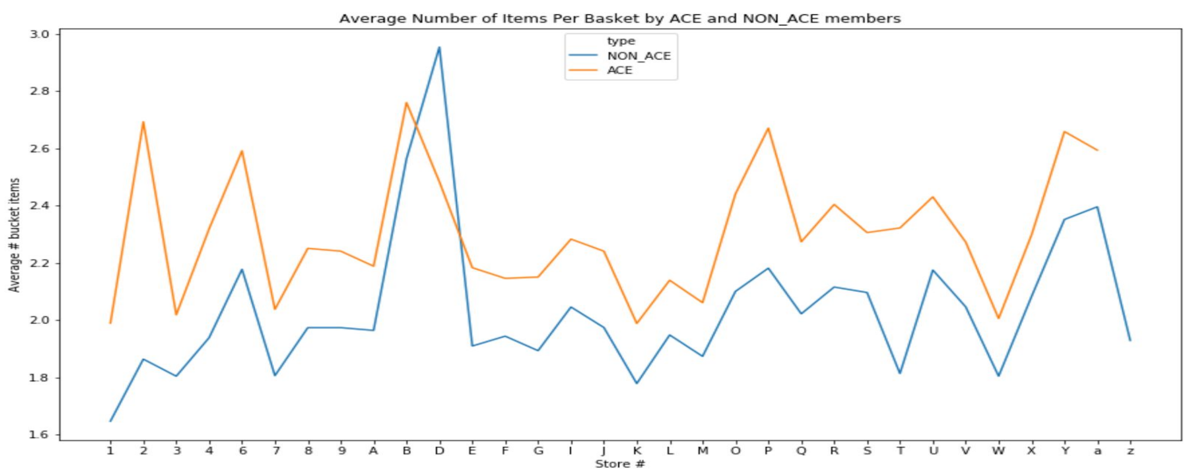
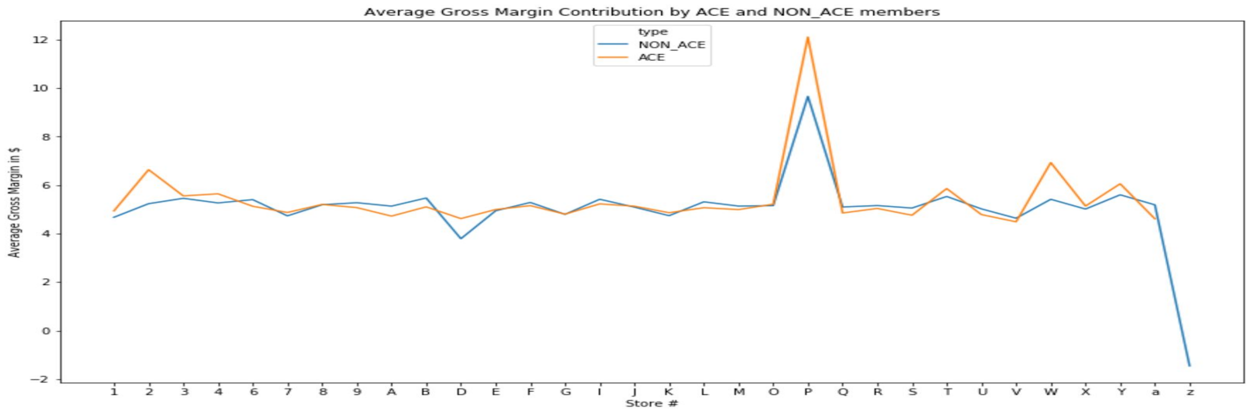


Fig (3)

As the primary motive behind a loyalty program is to turn occasional buyers into loyal customers, we wanted to see if the ACE_members purchase more than NON_ACE members. For this we did basket level analysis to see what was the average number of items bought by members and non-members. From **(Fig. 3)**, It was interesting to observe that ACE members buy more than non-members on an average at every store **except 'D' and 'z'**.

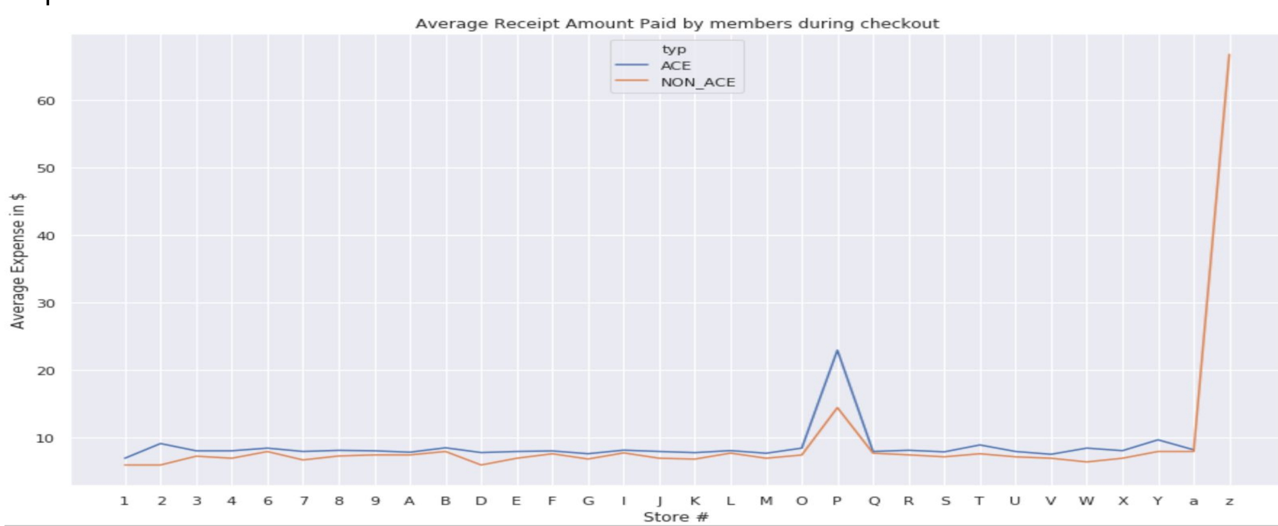
<u>Sniff Test</u> - 'z' is an exception as the data is very less and there are no ACE members yet as the store is relatively new. Hence, if we convince non-members to take the membership, we can expect more purchases as the membership provides attractive offers if people buy more. Since store 'D' is closed before 2017, there are not enough records to compare margin of D with other stores in an unbiased way. Hence it as an exception.



Fig(4)

Next, we were interested to see the Gross Margin contribution by members and non-members. For each category, we got the average gross margin per Receipt **(Fig. 4)**. Later, we grouped the records based on store, to get an insight on the average gross margin contribution per purchase in a store. It was interesting to see that stores **'2', '4', 'P', 'W' and 'Y'** seem to have notable margin contribution by members over non-members. These stores could be the immediate focus by the retailer to start the promotion of ACE membership.

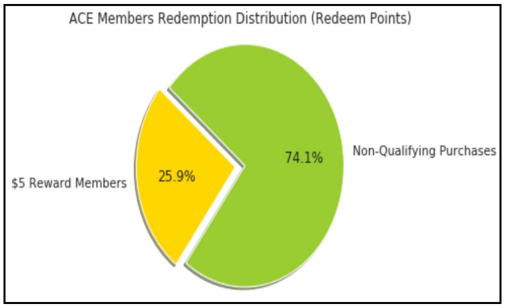
Also, to ensure the ACE membership contribution to the bucket items or margin are in fact significant and not by chance, we considered the members based on **'Repeating Customers**'. We wanted to see how many times an ACE member visit in a four year period and what is the ACE member contribution to the margin of the company. We found it interesting to see that members visit the store **10 times** on average, which was more than what we guessed it would be as it is a hardware store and **66.47%** of the total margin is driven by the members of the loyalty club. Hence we see a huge potential to increase margin, by convincing non-members to become members by providing exciting offers on their initial purchases.

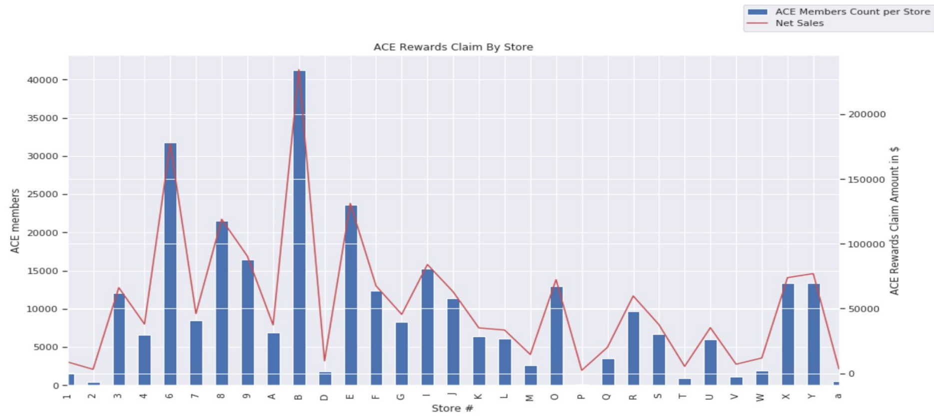


Fig(5)

<u>Increase Reward points based on Average checkout amount</u>

In order to boost the sales of the stores, we wanted to check the average checkout in each store. From **(Fig. 5)** we observe that on average, across all stores, people pay anywhere between \$5 to \$10 with the exception of store 'P' and store 'z'. From this plot, Costello's ACE can add reward points or provide instant savings coupons to customers who buy more than a threshold checkout amount. Say, customer A buys more than \$12 in Store '1', then reward the user with 10 points. This will drive more sales.





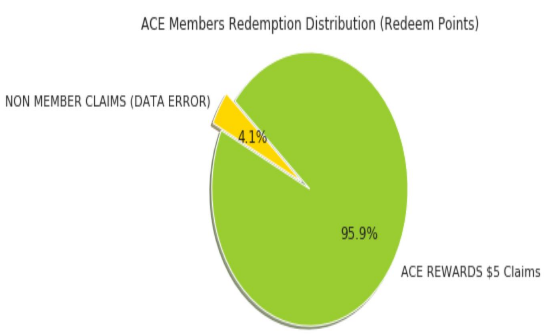
Fig(6)

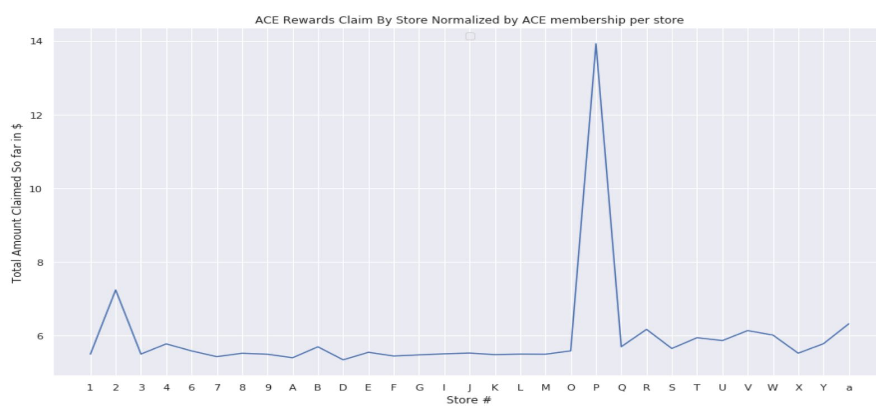
Fig(7)

<u>ACE Rewards Program</u>

ACE Rewards Program offers 10 points for every $1 purchase and gives a $5 reward when a customer accrues 2500 points. From the given data **(Fig. 6)**, we see that only 25.9% purchases over the four year period qualify for a $5 reward i.e. cumulative purchases made for at least $250 by an ACE member. 74.1% of the remaining ACE members have non qualifying purchases.
(Fig.7) shows the correlation of ACE rewards claim made in each store with respect to count of number of ACE members in the corresponding store. We can see that it is well related and for few stores the distance between the line and the corresponding top of bar graph is wide indicating that, even though there are less members, the claims made are more i.e. Store 2 and P.

<u>Sniff Test</u> - This behavior is confirmed with **(Fig.9)** ACE Rewards by Claim (Normalized Plot) where stores 2 and P have higher claim amount.





Fig(8)

Fig(9)

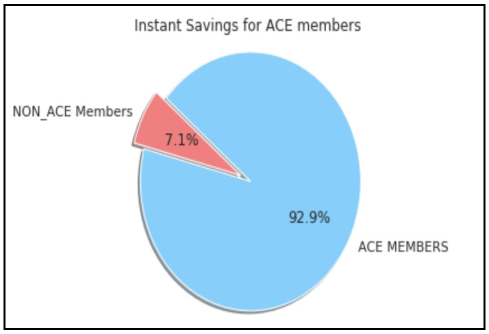
<u>Anomaly</u> - (Fig. 8) From the total transactions that involved a $5 reward claim, 4.1% were claimed by users not having a ‘Loyalty ID’ i.e. a NON_ACE member. We were not exactly sure about the reason for this and wanted to flag this. Maybe the clerk incorrectly added rewards or it is data error.

<u>Average amount claimed through ACE Rewards Program</u>

From **(Fig. 9)**, we see that on an average, all the ACE members of a store claim once i.e. $5 except store ‘2’ and ‘P’ that show higher claim.

Instant Savings

Instant Savings [program](#) are for ACE rewards member only. When we analyzed the data, we could see that even non members have claimed instant savings **(Fig. 11)**. This is another **anomaly** we found.



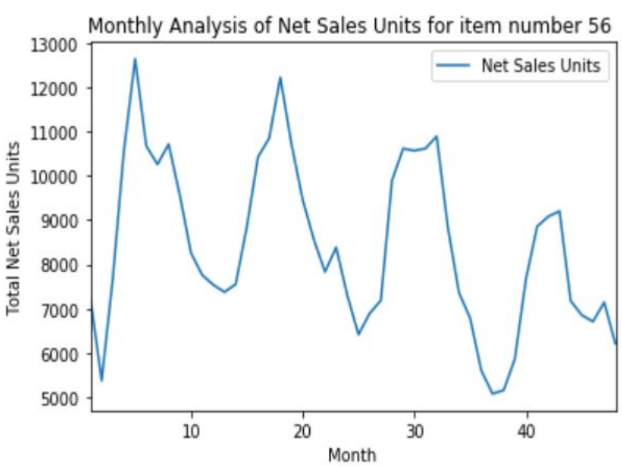
Fig(11)

Fig(12)

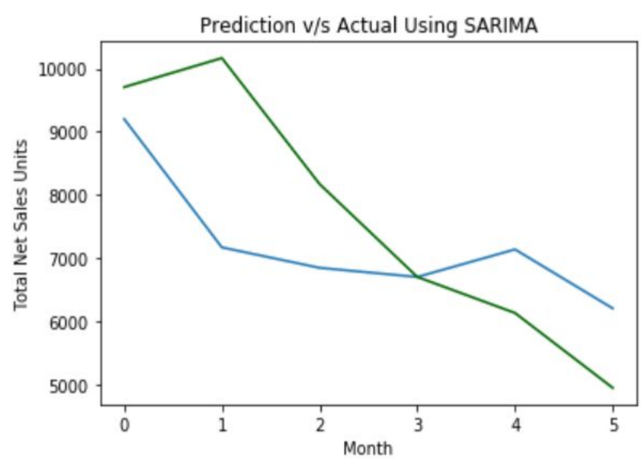
(Fig.12) shows the average savings done by ACE members on their purchases through Instant Savings along with the average expenditure during checkout. As expected, the checkout amount is more than the savings that could be applied with the exception of Store P and T, which suggests that ACE members are active on savings plan and buy more items with instant savings.

2. Time Series Modelling of Sales:

For this objective, finding out what products should a store stock and how much of it should be stocked, we employed ARIMA(autoregressive integrated moving average) model to our utility. ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. A variant of ARIMA called SARIMA was chosen owing to the seasonal nature of the data.Data being used has the following pattern:



Fig(13)

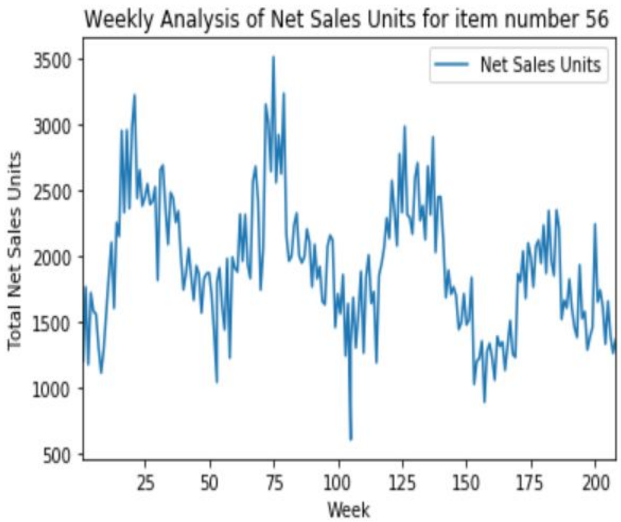


Fig(14)

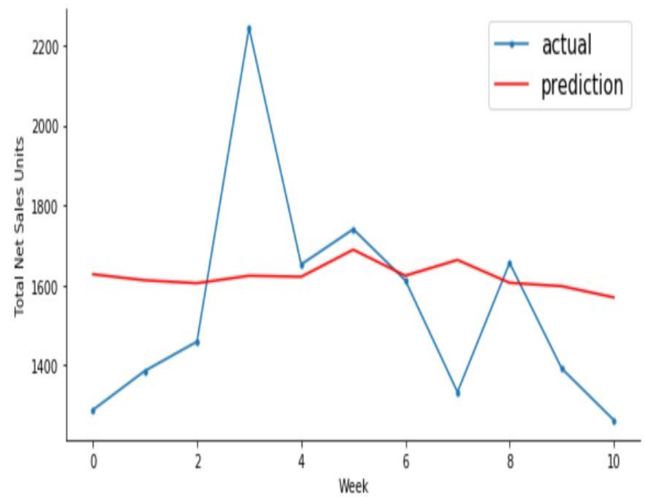
From the **Fig(13)** we observed that the data had a yearly seasonality and declining sales trend. The total sales count grouped by the month. Hence we had 48 total instances from the data. Of them we used 42 for training and remaining 6 for test. Below is the comparison of Actual vs Predicted data:

From the **Fig(14)** we could see that model didn't perform well, because of less data trained to the model. Although we have 4 years of data the model still couldn't get the pattern thereby not performing to the expectations.

After exploring classical statistical tool SARIMA we were exploring if we can get good results from a machine learning model. *LSTM* seemed to be a good choice since it can remember patterns spanning a long period of time. The sales data was aggregated by week. Since we had 4 years of data, 208 data points were there. Of them 200 were used for training and for remaining 8 we tried to predict.



Fig(15)



Fig(16)

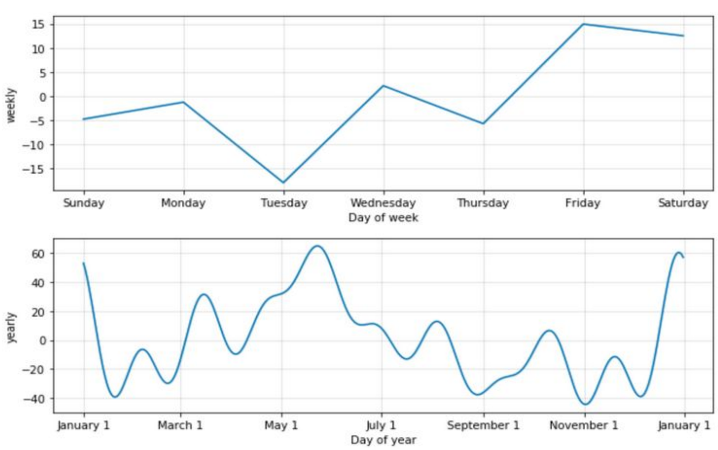
From the **Fig(16)**, we could see that LSTM model didn't perform that well even after passing enough number of training records to the model, this can be due to model architecture and it's hyper parameters.

We continued our investigation around store#9 and product: fasteners item code 56 for the whole dataset. The sales count for this product were grouped by the day now.
Now we consider **PROPHET by Facebook.** A brief of how prophet works?[As mentioned here]

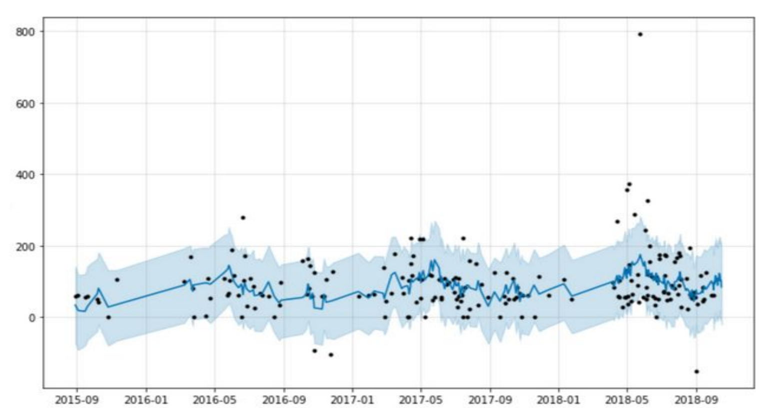
At its core, the Prophet procedure is an additive regression model with four main components:

- A piecewise linear or logistic growth curve trend. Prophet automatically detects changes in trends by selecting changepoints from the data.
- A yearly seasonal component modeled using Fourier series.
- A weekly seasonal component using dummy variables.

Also along with predicted values it provides us with a range of possible values, the upper and the lower bound, here we tried to address this for the product named 'Screen' and store 'T'.



Fig(17)

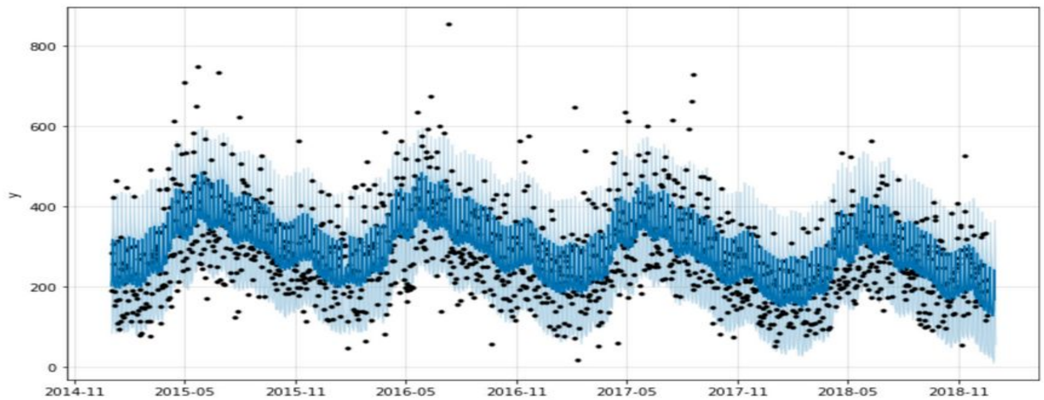


Fig(18)

The above **Fig(17)** shows the weekly trend of sales and the second one shows the yearly trend of sales. We can observe that during weekends there's usually a surge of sales. The product is also popular during the summer months[April to July] as well as in January. In the Fig(18) *black dots* represent the actual sales recorded.The *dark blue* line represents the value predicted by the model.The *light blue* line region represent the upper and the lower bound predicted by the model.Observe that a great majority of black dots are within the light blue region generated by the model. So till the point store has the stock within this region they would mostly never witness the stock shortage problem.

The **Fig(18)** shows the performance of prophet model over 215 data points. 200 were used for training while we predicted the values for the remaining 15 days. About 93% percent of the points are within the lower bound and upper bound value predicted by Prophet.

A similar analysis was performed for Store: 9 Item Number: 56(Fasteners). For it we had 1452 data points, of which 1440 data points were used for training, 12 used for testing. Observe from the following **Fig(19)** that a great majority of actual data points represented by black dots fall within the blue region.



Fig(19)

Summary of time series analysis

Of all the models we explored, Prophet seemed to be performing the best in terms of accuracy, performance, ease of use and intuitiveness. Apart from these Prophet's USP is the upper and lower bound predicted by it along with the value forecasted.

3. Product Placement Strategy

For this objective we implemented the **APRIORI and FP** algorithm. In this section, we will discuss the baseline algorithm we used and also discuss the shortcomings of each and why we considered other approaches.

APRIORI -

Apriori is an algorithm for frequent item set mining and association rule learning over relational databases.To understand association rules, this blog has a very good example.

The key concept in the APRIORI algorithm is that it assumes all subsets of a frequent itemset to be frequent. Similarly, for any infrequent itemset, all its supersets must also be infrequent.

An association rule has two parts: an **antecedent** (if) and a **consequent** (then). An antecedent is an item found within the data. A consequent is an item found in combination with the antecedent.

Association rules are created by searching data for frequent if-then patterns and using the criteria/metric ***support***, ***confidence, lift, conviction and leverage*** to identify the most important relationships.

Implementation Steps

Since the product placement analysis is based on the purchasing behavior of the customers in a particular store, it makes more sense to analyze store wise. The implementation steps discussed below will apply for each store.

1. Filter out all the transactions that do not correspond to any merchandize/product being sold. This step is very important to avoid associations that include gift cards or instant savings that are not products per se.
2. For the store under analysis, get all the orders (Item numbers) with receipt information i.e. per basket.
3. Calculate the individual item frequency and remove items that fall below a specified threshold. This will help eliminate all individual items that are below a support threshold, meaning, products that are not purchased often can be ignored.
4. Filter a basket with less than 2 items. This will eliminate all baskets that cannot have an association rule established. This step is to avoid unnecessary processing.
5. For each basket, get all pairs of possible items.
6. Next, count the number of times each item pair appears in the orders.
7. Calculate the Association Rule Mining metrics to derive insights for product placement.

Shortcomings of the APRIORI-pairwise approach -

- The above algorithm is the implementation of APRIORI for itemset of size exactly 2 i.e. (A->B). When we implemented for itemsets of size 3 or above (say A,B -> C) the processing took a lot of time as the matrix generated for processing grows extremely large when we increase more items. Hence we explored other existing implementations to handle large scale transactional data. We decided to proceed with the implementation in this research paper as it is more relevant and has proven results for large database with sales transactions.

Shortcomings of APRIORI-k approach

- Using Apriori needs a generation of candidate itemsets. These itemsets may be large in number if the itemset in the database is huge.
- Apriori needs multiple scans of the data to check the support of each itemset generated and this leads to high costs.

Frequent Pattern Growth Algorithm (FP)

This algorithm is an improvement to the Apriori method. A frequent pattern is generated without the need for candidate generation. It builds its own tree called the FP tree on the transactions.
The purpose of the FP tree is to mine the most frequent pattern. Each node of the FP tree represents an item of the itemset. The root node represents null while the lower nodes represent the itemsets. The association of the nodes with the lower nodes that is the itemsets with the other itemsets are maintained while forming the tree.
More information on FP can be found here and algorithm can be found here.

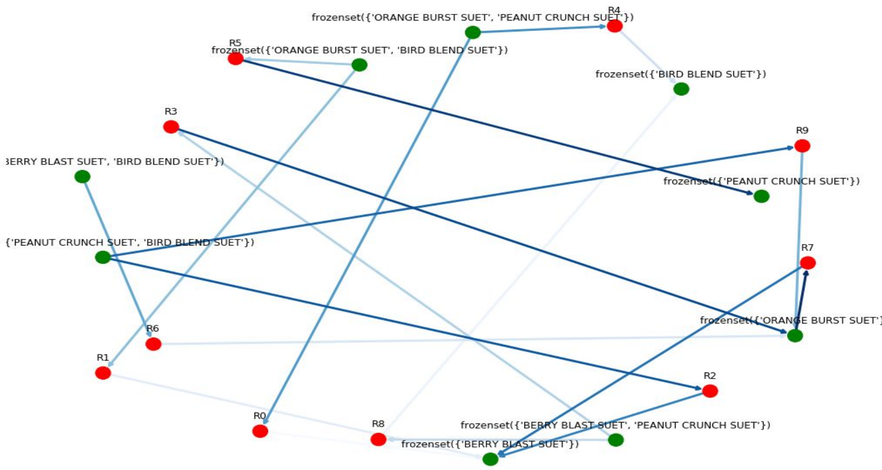
Evaluation Metrics (Support - 0.001, Confidence - 0.1, Lift - >1, Leverage - closer to 0, Conviction - >1)

For instance, In STORE # Y

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
32	(PEANUT CRUNCH SUET, ORANGE BURST SUET)	(BERRY BLAST SUET)	0.001613	0.003580	0.001258	0.779768	217.807098	0.001252	4.524414
0	(BIRD BLEND SUET, ORANGE BURST SUET)	(BERRY BLAST SUET)	0.001640	0.003580	0.001278	0.779275	217.669234	0.001272	4.514297
20	(PEANUT CRUNCH SUET, BIRD BLEND SUET)	(BERRY BLAST SUET)	0.001673	0.003580	0.001181	0.706301	197.285982	0.001175	3.392655
33	(PEANUT CRUNCH SUET, BERRY BLAST SUET)	(ORANGE BURST SUET)	0.001812	0.003070	0.001258	0.694184	226.110983	0.001252	3.259900
82	(PEANUT CRUNCH SUET, ORANGE BURST SUET)	(BIRD BLEND SUET)	0.001613	0.003709	0.001090	0.675448	182.096651	0.001084	3.069740
81	(BIRD BLEND SUET, ORANGE BURST SUET)	(PEANUT CRUNCH SUET)	0.001640	0.003311	0.001090	0.664249	200.588787	0.001084	2.968532
1	(BIRD BLEND SUET, BERRY BLAST SUET)	(ORANGE BURST SUET)	0.001935	0.003070	0.001278	0.660808	215.239870	0.001272	2.939135
64	(ORANGE BURST SUET)	(BERRY BLAST SUET)	0.003070	0.003580	0.002028	0.660576	184.513956	0.002017	2.935619

In the **Fig(20)**, red dot represents rules and green dots represents itemsets. We can observe the relation between 2 or 3 items. For example, Rule R5 represents the relation between itemset (ORANGE BURST SUET, BIRD BLEND SUET) and

PEANUT CRUNCH SUET. Arrow coming into a rule represents an antecedent and an arrow coming out of a rule (red dot) represents a consequent. This means people buying ORANGE BURST SUET and BIRD BLEND SUET will most probably buy PEANUT CRUNCH SUET. For Store Y (from the above table), we found itemsets with combinations of more than 2 items with a lift score >1.



Fig(20)

SEASONAL BUYING TRENDS BASED ON FREQUENT ITEM SETS FOR EACH SEASON

We assumed that buying seasons can be segregated into four different seasons based on four seasons of America divided into different months.

We ran FP algo on each store and found the seasonal trends in the behaviour of the customers.

For Instance, In STORE # 2

Spring

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
2	(WEED & FEED 5M)	(LAWN FOOD 5M)	0.002818	0.003623	0.002379	0.844156	232.978486	0.002369	6.393417
3	(LAWN FOOD 5M)	(WEED & FEED 5M)	0.003623	0.002818	0.002379	0.656566	232.978486	0.002369	2.903559
1	(SUFFOLK COUNTY BAG CHARGE)	(FASTENERS)	0.006478	0.086996	0.001610	0.248588	2.857450	0.001047	1.215050
0	(FASTENERS)	(SUFFOLK COUNTY BAG CHARGE)	0.086996	0.006478	0.001610	0.018511	2.857450	0.001047	1.012260

Summer

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(KEY KWIKSET KW1-ACE250PK)	(KEY SCHLAGE SC1-ACE250PK)	0.005526	0.011899	0.001031	0.186667	15.687513	0.000966	1.214878
1	(KEY SCHLAGE SC1-ACE250PK)	(KEY KWIKSET KW1-ACE250PK)	0.011899	0.005526	0.001031	0.086687	15.687513	0.000966	1.088865

Fall

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
4	(REGISTER 4WAY 8X8 WHT, DRYER VENT DUCT4"X8FT AL)	(REGISTER 4WAY 6X6 WHT)	0.001133	0.001133	0.001133	1.000000	882.642857	0.001132	inf
6	(DRYER VENT DUCT4"X8FT AL, REGISTER 4WAY 6X6 WHT)	(REGISTER 4WAY 8X8 WHT)	0.001133	0.001133	0.001133	1.000000	882.642857	0.001132	inf
7	(REGISTER 4WAY 8X8 WHT)	(DRYER VENT DUCT4"X8FT AL, REGISTER 4WAY 6X6 WHT)	0.001133	0.001133	0.001133	1.000000	882.642857	0.001132	inf
9	(REGISTER 4WAY 6X6 WHT)	(REGISTER 4WAY 8X8 WHT, DRYER VENT DUCT4"X8FT AL)	0.001133	0.001133	0.001133	1.000000	882.642857	0.001132	inf

Winter

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
2	(ICE MELT 40# PAIL ACE)	(ICE MELT 40# BAG ACE)	0.001186	0.004998	0.001186	1.000000	200.067797	0.001180	inf
5	(PLASTIC PL LID F/3.5&5G)	(PLSTC BUCKET 5G WHT ACE)	0.001059	0.002245	0.001017	0.960000	427.616604	0.001014	24.943875
4	(PLSTC BUCKET 5G WHT ACE)	(PLASTIC PL LID F/3.5&5G)	0.002245	0.001059	0.001017	0.452830	427.616604	0.001014	1.825651
3	(ICE MELT 40# BAG ACE)	(ICE MELT 40# PAIL ACE)	0.004998	0.001186	0.001186	0.237288	200.067797	0.001180	1.309556

Follows the seasonal buying pattern. WEED & FEED 5M & LAWN FOOD 5M are mostly bought in spring season. This is intuitively correct as during spring people maintain gardens and parks.
Also, during the summer and fall , items related to construction or repairs works are mostly bought together like KEY KWIKSET KW1-ACE250PK) and KEY SCHLAGE SC1-ACE250PK, DRYER VENT DUCT4"X8FT AL, REGISTER 4WAY 6X6 WHT and REGISTER 4WAY 8X8 WHT. In winter, ICE MELT 40# PAIL ACE and ICE MELT 40# BAG ACE or PLASTIC PL LID F/3.5&5G and PLASTIC BUCKET 5G WHT ACE, are most frequently bought.

INSIGHTS FROM FP-ALGORITHM:

We applied FP-Algorithm on all the stores, we found that most of the stores placement strategy is fine as we got itemsets same classes and fine-lines, and below are some interesting insights from particular stores:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
3	(COLOR SAMPLE BM BASE2 PT)	(COLOR SAMPLE BM BASE1 PT)	0.002390	0.006932	0.001115	0.466281	67.267353	0.001098	1.860659
6	(KEY SCHLAGE SC1-ACE250PK)	(KEY KWIKSET KW1-ACE250PK)	0.011632	0.015727	0.001928	0.165710	10.536923	0.001745	1.179773
2	(COLOR SAMPLE BM BASE1 PT)	(COLOR SAMPLE BM BASE2 PT)	0.006932	0.002390	0.001115	0.160797	67.267353	0.001098	1.188759
7	(KEY KWIKSET KW1-ACE250PK)	(KEY SCHLAGE SC1-ACE250PK)	0.015727	0.011632	0.001928	0.122566	10.536923	0.001745	1.126429
4	(KEY KWIKSET KW1-ACE250PK)	(FASTENERS/SCREWS/NAILS/BOLTS ETC)	0.015727	0.057257	0.001331	0.084639	1.478223	0.000431	1.029914

For **store 4** “FASTENERS/SCREWS/NAILS/BOLTS ETC“ belongs to the different department(FASTENERS) and class(HILLMAN GROUP, INC) when compared to the antecedent item “KEY KWIKSET KW1-ACE250PK”, which belongs to HARDWARE department and class of keys, from the above table we can see that they have a lift score >1, So these items could be placed together which could increase sales for the Store.

For **store T** APPLIANCES & HOME ENTERTAINMENT and FASTENERS/SCREWS/NAILS/BOLTS ETC belong to different departments and classes. APPLIANCES & HOME ENTERTAINMENT belongs to department APPLIANCES & HOME ENTERTAINMENT and FASTENERS/SCREWS/NAILS/BOLTS ETC belongs to department FASTENERS. Despite being from different departments they have good lift score greater than 1, which indicates that whenever someone does buy APPLIANCES & HOME ENTERTAINMENT, he is very likely to FASTENERS/SCREWS/NAILS/BOLTS ETC as well, as inferred from a high lift value **> 1.**

Similarly, for **store 6 ,** items RECEPTACLE 4-WIRE50A250V and ELBOW PULL EMT 90D 3/4" belongs to different classes Receptacles/Outlets and electrical fittings respectively but department is same(Electrical supplies).

SNIFF TEST FOR FP-ALGORITHM:

We found that most of the stores follow similar seasonal buying trends. Stores L, K, J, I, G, F, E, 9, 8, 7,2,1 has similar trends of items that customers buy in different seasons. In Spring, most frequently bought items were related to lawns and Gardens like WEED & FEED 5M and LAWN FOOD 5M. During summer and fall, the frequent items bought by customers were mostly related to hardware like CLAMP 11/16 TO 1-1/2"SS, COUPLE INSERT POLY 1" or KEY KWIKSET and KEY SCHLAGE.

While in winter, items related to melting or removing snow or related to feeding birds were bought like ICE MELT 40# PAIL ACE and ICE MELT 40# BAG ACE or PLASTIC PL LID F/3.5&5G and PLASTIC BUCKET 5G WHT ACE, PEANUT CRUNCH SUET . Feeding birds is intuitively correct as birds avoid migration in winter and rescue centers for birds buy food for birds for their survival.

Some stores have similar buying patterns throughout the year across different seasons. For example, Stores Y(Edgewater MD), W(Glen Burnie, Maryland), X (PASADENA MD) are all located in **Anne Arundel County in Maryland** and most of the products bought are products for feeding birds such as different types of suet like BERRY BLAST SUET, ORANGE BLAST SUET etc. This is intuitively correct as Maryland has over 400 species of birds and various bird sanctuaries are present in **Anne Arundel County.**https://www.kaggle.com/datatheque/association-rules-mining-market-basket-analysis

4. Membership Promotion

We did non-seasonal and seasonal analysis of items frequently bought by non ace customers only. We believe if the conviction of these items across different seasons is high then we can give rewards on the antecedent items for the ace members. This will compel the customers to join membership program of ace.

SEASONAL REWARDS FOR ENCOURAGING NON ACE CUSTOMERS TO TAKE ACE MEMBERSHIP

For Instance, In STORE # 1

Spring

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
2	(KEY KWIKSET KW1-ACE250PK)	(KEY SCHLAGE SC1-ACE250PK)	0.024198	0.018107	0.003567	0.147392	8.140074	0.003128	1.151635
3	(KEY SCHLAGE SC1-ACE250PK)	(KEY KWIKSET KW1-ACE250PK)	0.018107	0.024198	0.003567	0.196970	8.140074	0.003128	1.215150
0	(WEED & FEED 5M)	(LAWN FOOD 5M)	0.003292	0.003018	0.002524	0.766667	254.045455	0.002514	4.272781
1	(LAWN FOOD 5M)	(WEED & FEED 5M)	0.003018	0.003292	0.002524	0.836364	254.045455	0.002514	6.090992

Summer

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
2	(KEY KWIKSET KW1-ACE250PK)	(KEY SCHLAGE SC1-ACE250PK)	0.022889	0.016895	0.002428	0.106061	6.277793	0.002041	1.099745
3	(KEY SCHLAGE SC1-ACE250PK)	(KEY KWIKSET KW1-ACE250PK)	0.016895	0.022889	0.002428	0.143695	6.277793	0.002041	1.141078
0	(COUPLE INSERT POLY 1")	(CLAMP 11/16 TO 1-1/2"SS)	0.004558	0.004310	0.001784	0.391304	90.782609	0.001764	1.635776
1	(CLAMP 11/16 TO 1-1/2"SS)	(COUPLE INSERT POLY 1")	0.004310	0.004558	0.001784	0.413793	90.782609	0.001764	1.698107

Fall

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(KEY KWIKSET KW1-ACE250PK)	(KEY SCHLAGE SC1-ACE250PK)	0.024410	0.021058	0.003059	0.125301	5.950407	0.002545	1.119177
1	(KEY SCHLAGE SC1-ACE250PK)	(KEY KWIKSET KW1-ACE250PK)	0.021058	0.024410	0.003059	0.145251	5.950407	0.002545	1.141376
2	(CAP KEY ASST 275/PKG)	(KEY KWIKSET KW1-ACE250PK)	0.003294	0.024410	0.001059	0.321429	13.167728	0.000978	1.437711
3	(KEY KWIKSET KW1-ACE250PK)	(CAP KEY ASST 275/PKG)	0.024410	0.003294	0.001059	0.043373	13.167728	0.000978	1.041897

Winter

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(KEY KWIKSET KW1-ACE250PK)	(KEY SCHLAGE SC1-ACE250PK)	0.028132	0.024992	0.004711	0.167442	6.699866	0.004007	1.171099
1	(KEY SCHLAGE SC1-ACE250PK)	(KEY KWIKSET KW1-ACE250PK)	0.024992	0.028132	0.004711	0.188482	6.699866	0.004007	1.197592

For example, for Store # 1, in spring season, we can give reward points on antecedent WEED & FEED 5M for ace members as they are more frequently bought by customers than any other products and its consequent LAWN FOOD 5M is highly dependent on its consequent as its conviction value (2.651297) is high. This will compel the non ace customers to join the membership program as these items are most frequently bought even by non ace customers. We can also apply rewards on LAWN FOOD 5M as well , as they are also frequently bought by customers and have good conviction score w.r.t WEED & FEED 5M.

For other seasons, we can similarly give reward points to items like CLAMP 11/16 TO 1-1/2"SS, COUPLE INSERT POLY 1" or KEY KWIKSET and KEY SCHLAGE which are most frequently bought during summer.

Conclusion:

In this report, we tried to suggest some strong reasons for Non Ace members to be converted to Ace members thereby increasing revenue to Costello, also we worked on Market Basket Analysis where we used Apriori and FP Algorithms to get rules which is used to get meaningful product placement strategies for a particular store. Apart from that we worked on time series analysis using different models like SARIMA, LSTM, PROPHET to forecast net sales of a product for particular store which will help with inventory management.

References:

1. https://www.farmersalmanac.com/the-seasons
2. https://dnr.maryland.gov/wildlife/Pages/plants_wildlife/mdwllists.aspx
3. https://www.alkaline-ml.com/pmdarima/tips_and_tricks.html
4. https://people.duke.edu/~rnau/411arim.htm
5. https://www.kaggle.com/datatheque/association-rules-mining-market-basket-analysis
6. https://www.yelp.com/biz/costellos-ace-hardware-bethpage-bethpage-2
7. https://hub.packtpub.com/time-series-modeling-what-is-it-why-it-matters-how-its-used/
8. https://www.softwaretestinghelp.com/fp-growth-algorithm-data-mining/