

Sales Data Analysis

elizabethh.com

ELIZABETH HO

Data Exploration and Feature Engineering

Sales Transactions Data

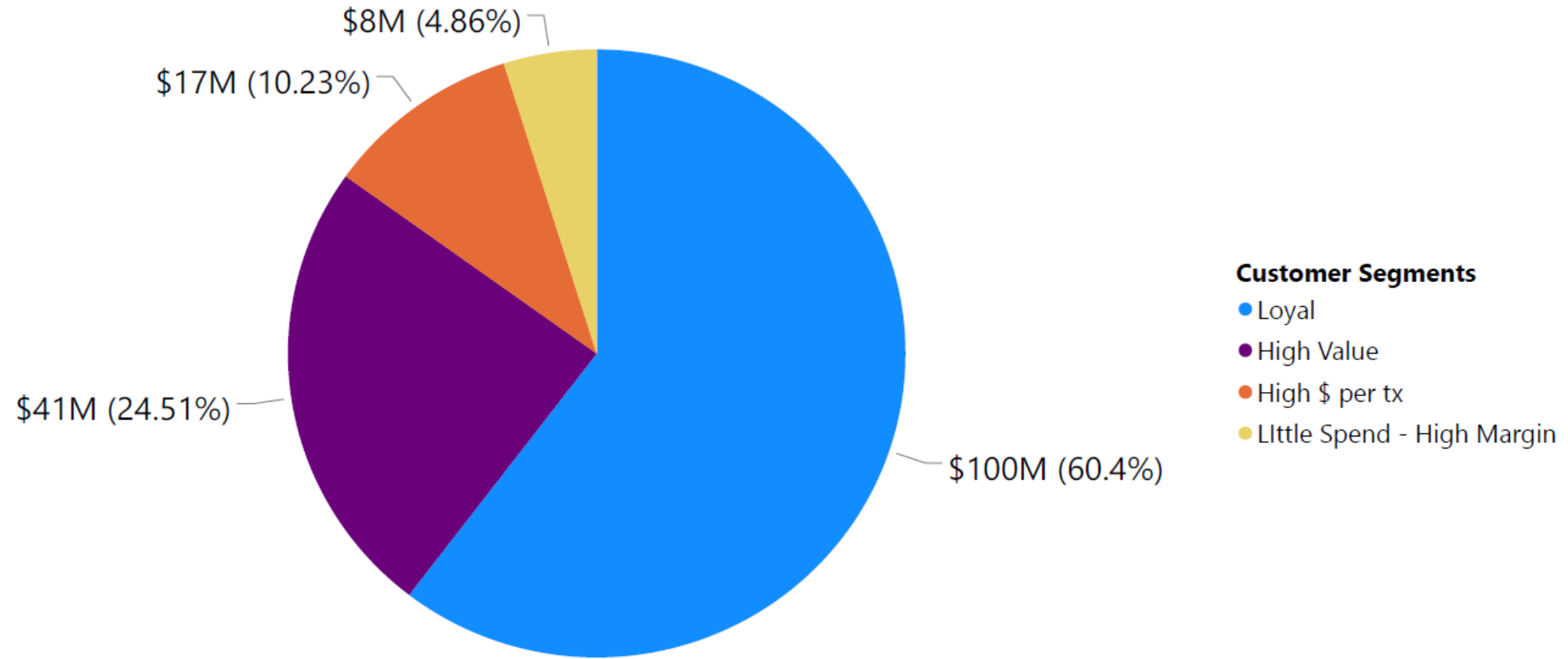
- 2,201,620 Line items
- Customer – 1,300
- Category - 14
- Subcategory - 293
- Part – 40,885
- Sales Price
- Sales Quantity
- Cost of Goods Sold
- Customer Number
- 329,554 Sales Orders
- Jan 1, 2015 – Oct 17, 2017

Feature Engineered New Columns

- Profit
- Revenue
- Customer Attributes
 - Total Spend
 - Profit Margin
 - Total Transactions
 - Tenure
 - Recency
 - Average Spend / Transaction
 - % Spent in Category

Customer Segmentation – Lifetime Value

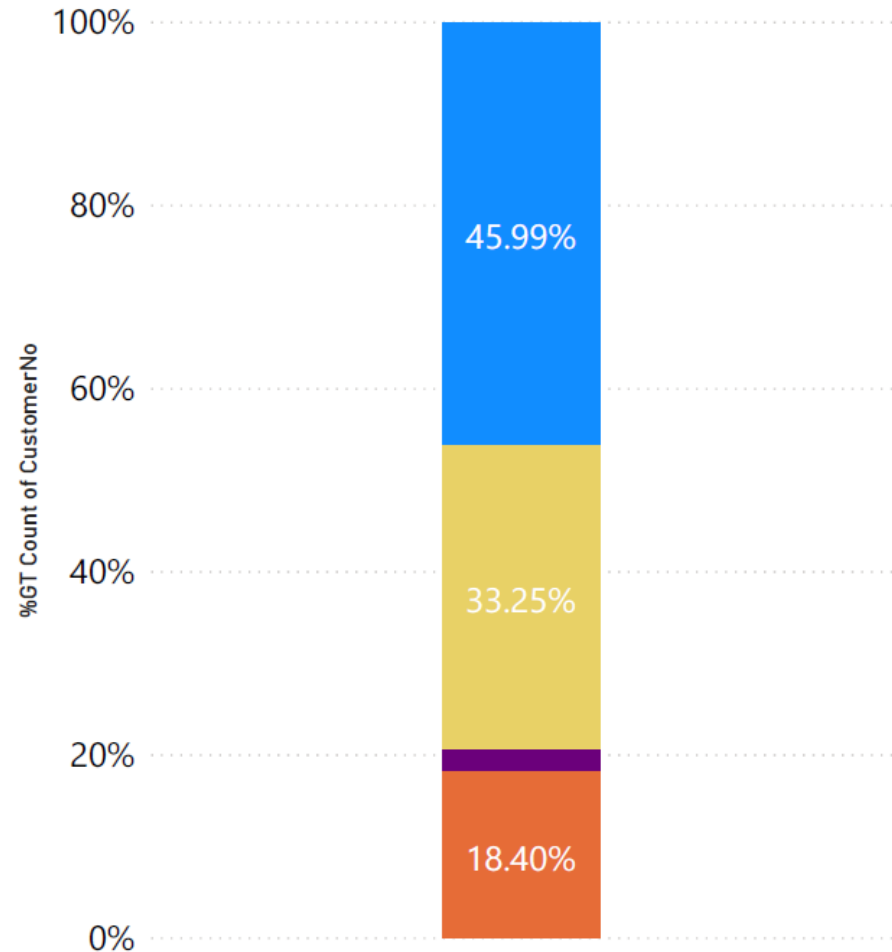
LTV BY Segment



Customer Segmentation

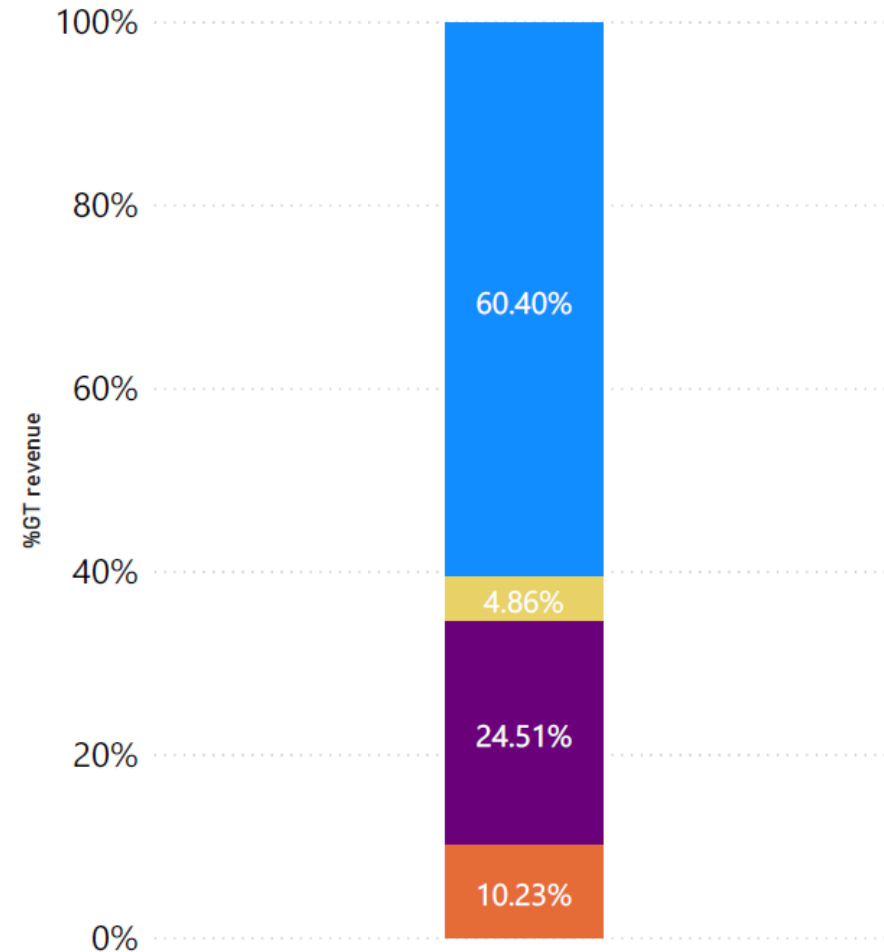
Segment Count

Description ● High \$ per tx ● High Value ● Little Spend ... ● Loyal



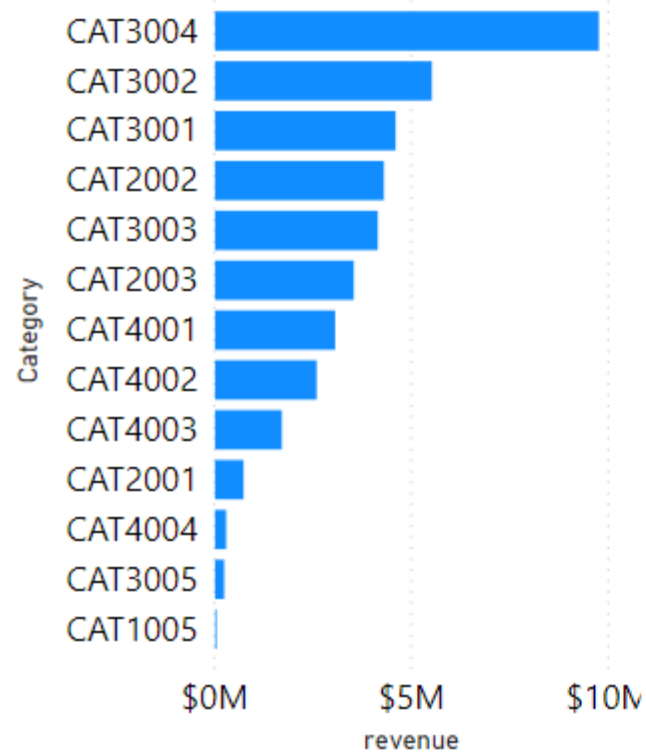
LTV by Description

Description ● High \$ per tx ● High Value ● Little Spend ... ● Loyal

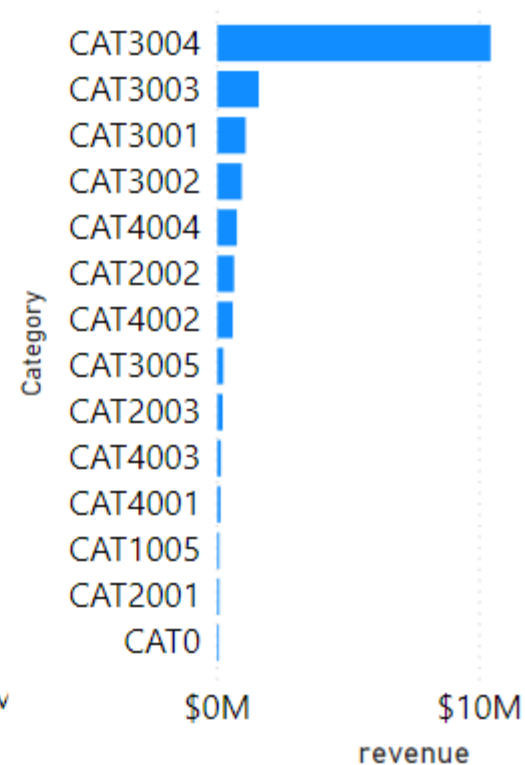


Customer Segmentation

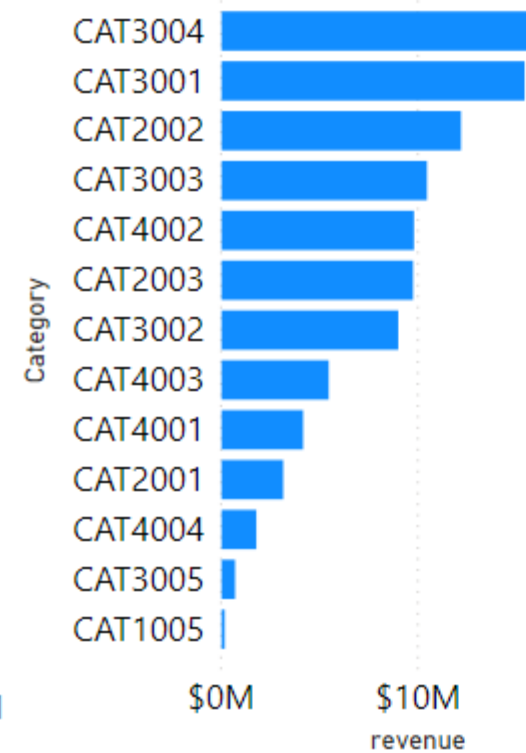
High Value



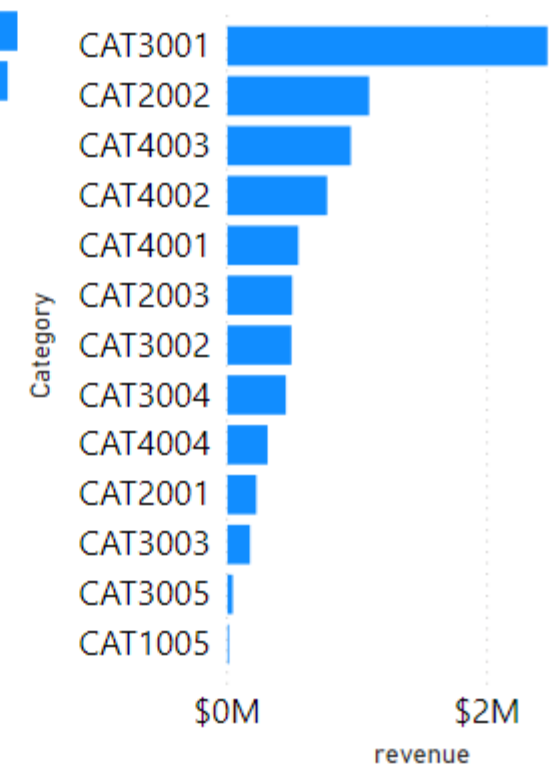
High \$ / TXN



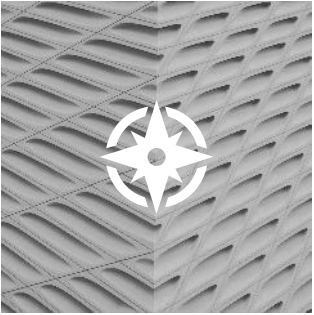
Loyal



Little Spend - High M...



Customer Segmentation



Highest Value

- highest spend
- **\$2M**
- highest total transactions
- **7,170**
- highest \$ / txn
- **\$9,200**
- average tenure



Loyal

- Longest tenure
- **1003 Days**
- 2nd highest total spend
- **\$250k**
- most recent purchasers



CAT3004 Buyers

- **47%** of spending on CAT3004 Products
- 2nd Highest \$ / txn
- **\$2500**



High Profit Margin

- **37%** Profit Margin
- **42%** of Total Spend on CAT3001



Small Spenders

- Total Spend under **\$30k**
- Total Transactions
- **~50**
- Represents **5%** of Customers

Price Elasticity

How sensitive are my customers to price?

$$\text{Price Elasticity} = \frac{\% \text{Change in Demand}}{\% \text{Change in Price}}$$



Customer
Segments



By Product
Category



By Top Selling
Products

Price Elasticity

High Value Customers



Top Products Purchased

- P62PA30868
 - 2%
- P64PA16383
 - - 0.3%



By Category

○ CategoryCAT2001	-1.2	○ CategoryCAT4001	-0.7
○ CategoryCAT2002	-1.4	○ CategoryCAT4002	-0.3
○ CategoryCAT2003	-0.7	○ CategoryCAT4003	-0.3
○ CategoryCAT3001	-0.7	○ CategoryCAT4004	-1.0
○ CategoryCAT3002	-0.7	○ CategoryCAT1005	-0.4
○ CategoryCAT3003	0.0		
○ CategoryCAT3004	-0.5		
○ CategoryCAT3005	-0.9		



Market Basket Analysis

Example of Association Rules:

- Assume there are 100 customers
- 10 of them bought milk, 8 bought butter and 6 bought both of them.
- bought milk => bought butter
- support = $P(\text{Milk \& Butter}) = 6/100 = 0.06$
- confidence = $\text{support} / P(\text{Butter}) = 0.06 / 0.08 = 0.75$
- lift = $\text{confidence} / P(\text{Milk}) = 0.75 / 0.10 = 7.5$

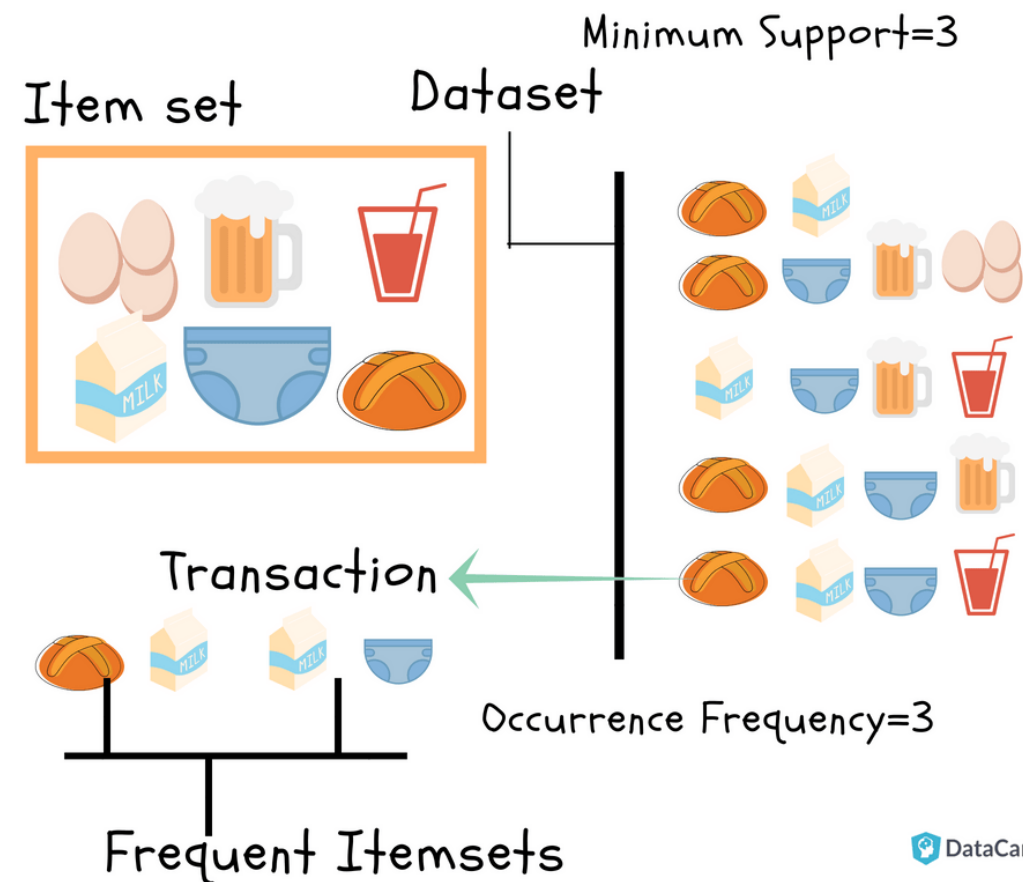
Most Popular Baskets

Products Sold Together

["P380PA11986","P2330PA9482","P562PA35589"]
 ["P380PA11977","P2330PA9482","P562PA35589"]
 ["P9PA32318","P2330PA9482","P562PA35589"]
 ["P380PA11977","P380PA11986","P2330PA9482"]
 ["P380PA11977","P380PA11986","P562PA35589"]
 ["P115PA11463","P2330PA9482","P562PA35589"]
 ["P380PA11986","P9PA32318","P2330PA9482"]
 ["P380PA11986","P9PA32318","P562PA35589"]
 ["P380PA11977","P9PA32318","P2330PA9482"]
 ["P380PA11977","P9PA32318","P562PA35589"]
 ["P115PA11463","P380PA11986","P2330PA9482"]
 ["P380PA11977","P380PA11986","P9PA32318"]
 ["P115PA11463","P380PA11977","P2330PA9482"]
 ["P115PA11463","P380PA11986","P562PA35589"]
 ["P15PA11465","P2330PA9482","P562PA35589"]
 ["P115PA11460","P2330PA9482","P562PA35589"]
 ["P115PA11463","P9PA32318","P562PA35589"]
 ["P115PA11462","P2330PA9482","P562PA35589"]
 ["P115PA11463","P9PA32318","P2330PA9482"]
 ["P115PA11463","P380PA11977","P562PA35589"]

Frequency

2426
 2418
 2385
 2309
 2210
 2035
 2031
 2025
 1984
 1969
 1877
 1832
 1808
 1765
 1756
 1727
 1700
 1696
 1695
 1677



Rules for Market Basket

if these Products in Basket	then Recommend	confidence	lift
["P660PA1519","P852PA1815"]	["P852PA1824"]	1.00	428.78
["P660PA34758","P852PA1824"]	["P660PA1519"]	1.00	466.79
["P852PA1824","P6226PA32320","P499PA3205"]	["P852PA1815"]	1.00	429.37
["P660PA1519","P660PA34758","P852PA1824"]	["P852PA1815"]	1.00	429.37
["P660PA1519","P660PA34758","P852PA1815"]	["P852PA1824"]	1.00	428.78
["P852PA1824","P499PA3205"]	["P852PA1815"]	1.00	429.37
["P660PA34758","P852PA1815","P852PA1824"]	["P660PA1519"]	1.00	466.79
["P660PA1519","P852PA1824"]	["P852PA1815"]	1.00	429.37
["P852PA1824","P5PA2538"]	["P852PA1815"]	1.00	429.37
["P852PA1815","P499PA3205"]	["P852PA1824"]	1.00	428.78
["P852PA1815","P6226PA32320","P499PA3205"]	["P852PA1824"]	1.00	428.78
["P852PA1824","P6226PA32320"]	["P852PA1815"]	1.00	429.37
["P660PA34758","P852PA1815"]	["P852PA1824"]	1.00	428.78
["P660PA34758","P852PA1815"]	["P660PA1519"]	1.00	466.79
["P660PA34758","P852PA1824"]	["P852PA1815"]	1.00	429.37
["P852PA1815","P6226PA32320"]	["P852PA1824"]	1.00	428.16
["P852PA1815","P5PA2538"]	["P852PA1824"]	1.00	428.12
["P852PA1815"]	["P852PA1824"]	1.00	427.04
["P852PA1824"]	["P852PA1815"]	0.99	427.04
["P660PA1519","P852PA1824"]	["P660PA34758"]	0.99	450.38

Rule: $X \Rightarrow Y$

$$\text{Support} = \frac{\text{freq}(X, Y)}{N}$$

$$\text{Confidence} = \frac{\text{freq}(X, Y)}{\text{freq}(X)}$$

$$\text{Lift} = \frac{\text{Support}}{\text{Supp}(X) \times \text{Supp}(Y)}$$



Rule	Support	Confidence	Lift
$A \Rightarrow D$	2/5	2/3	10/9
$C \Rightarrow A$	2/5	2/4	5/6
$A \Rightarrow C$	2/5	2/3	5/6
$B \& C \Rightarrow D$	1/5	1/3	5/9

Sales Forecast

28M

SellQty

\$127M

total_cogs

\$50M

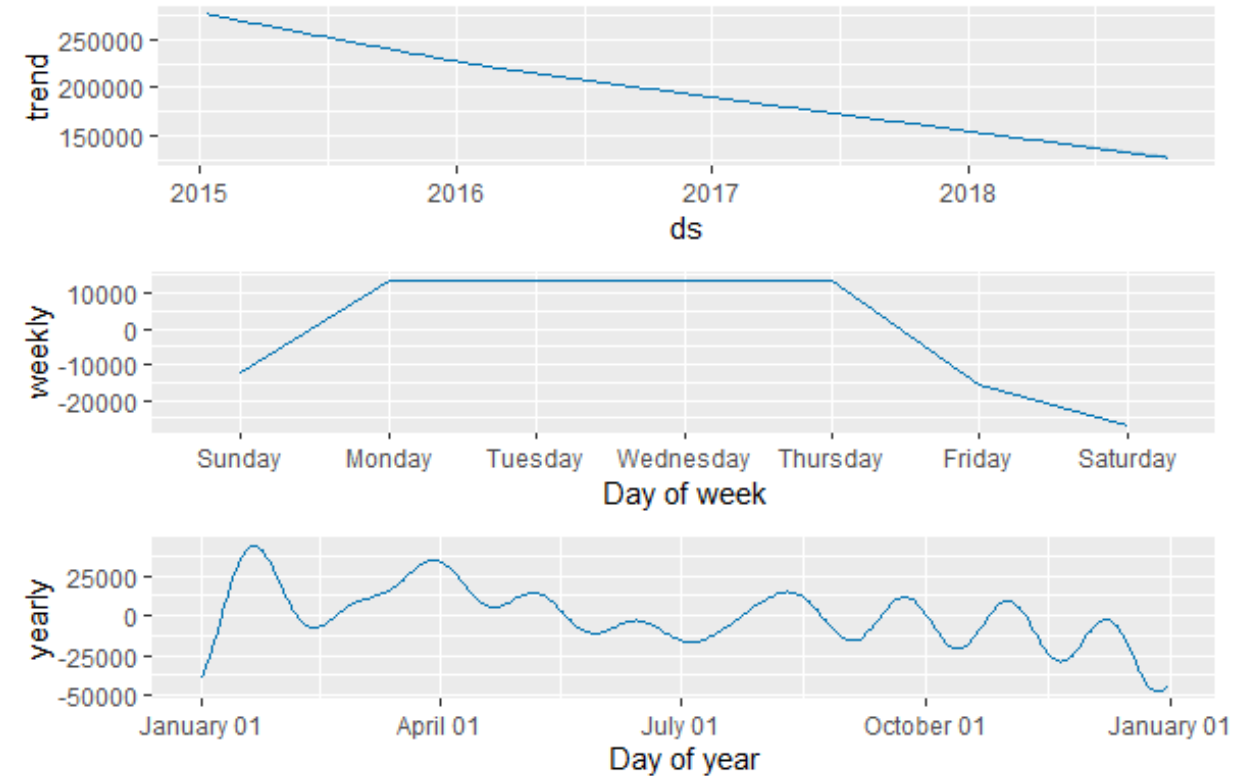
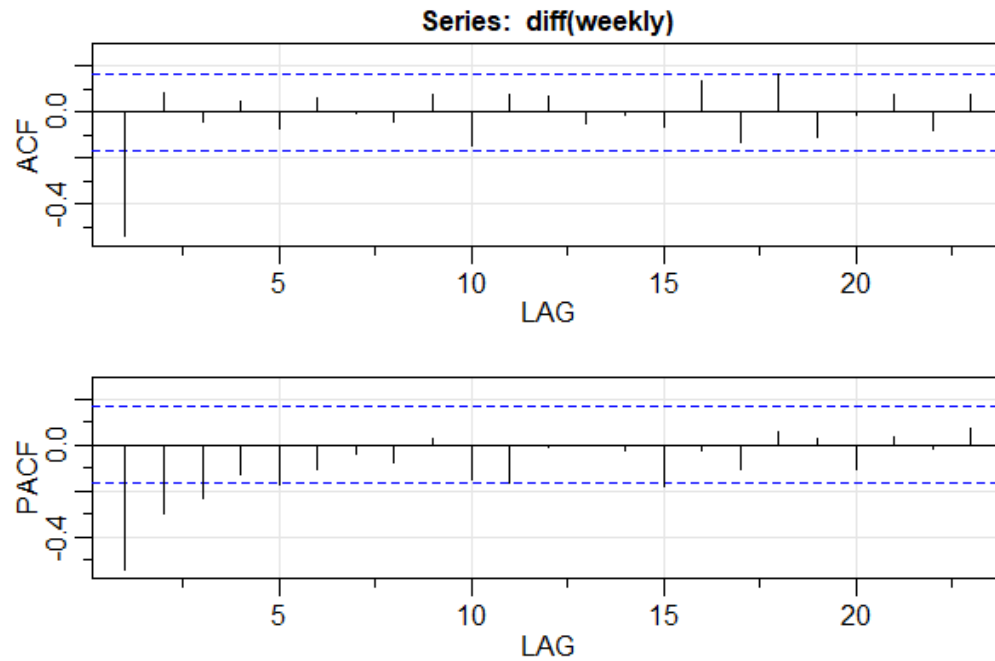
Profit

Sales By Month



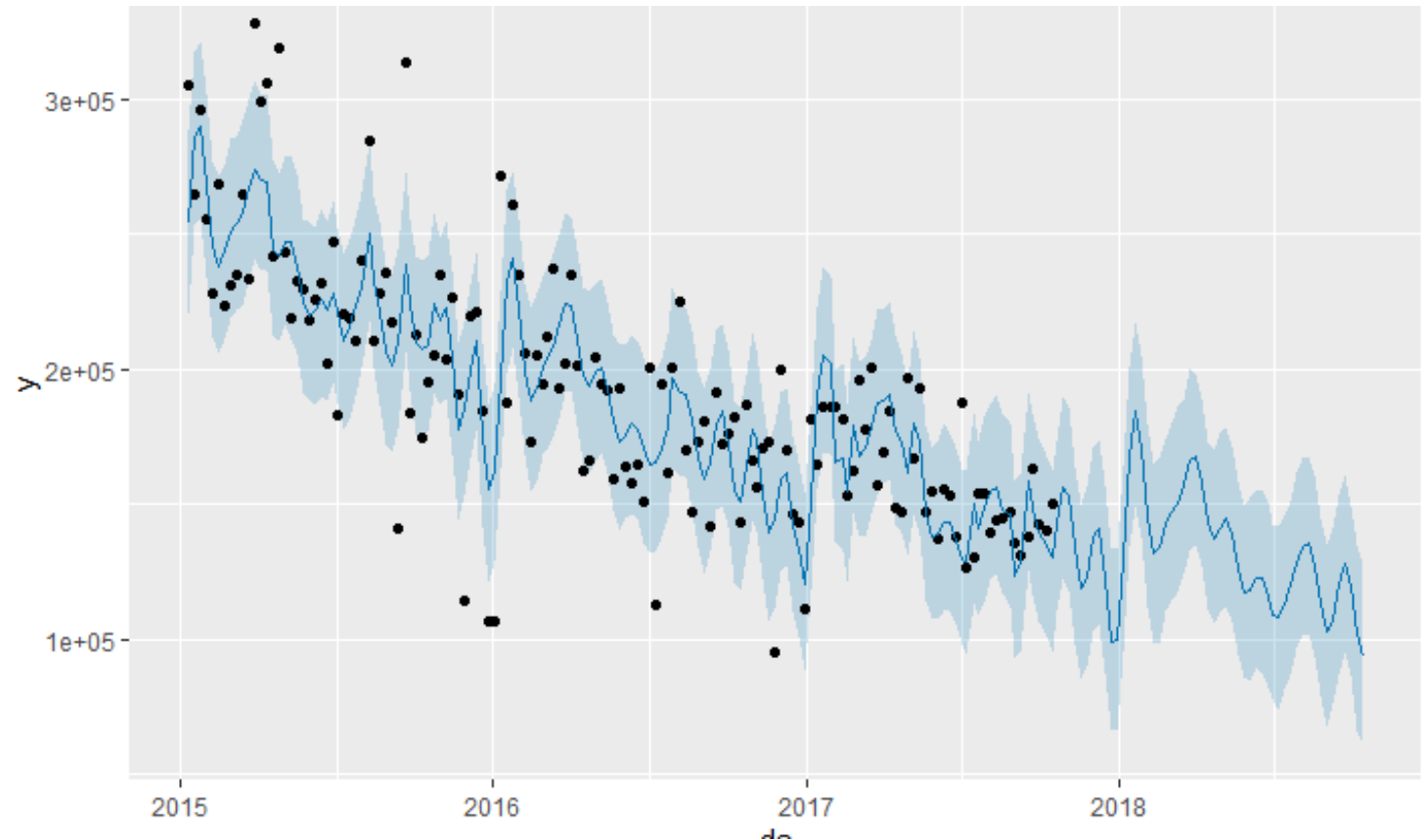
Sales Forecast

- No Sales on Sunday
- Aggregate by Week



Sales Forecast

- Prophet library open source
- Factor Yearly, Weekly & Daily Seasonality & Holiday Effects
- Better results than arima or arima.x



Key Insights & Recommendations



High Value Segment
Promote Product
P62PA30868 and
increase price

Average Sales Price of
P62PA30868 **\$830**

1% ↑ \$ P62PA30868
results in 2% ↑ in
quantity sold

This segment spends
19% of their
spending in this
category



Implement new
Market Basket Rules
for Product
Recommendations

Top 20 Market Basket
Rules provide lift
over 400.

Top 2802 rules, avg
confidence of 0.5,
average lift of 25.8



Adjust labor and
inventory to reflect
lowest sales cycle in
December.

Sales are lowest at
the end of the year
then spike in January



Thank You

 Elizabeth Ho

Appendix

Customer Segmentation

			segment3	total_spend	profit_ma rgin	total_txns	tenure	recency	spend_p er_txn	CAT1005	CAT2001	CAT2002	CAT2003	CAT3001	CAT3002	CAT3003	CAT3004	CAT3005	CAT4001	CAT4002	CAT4003	CAT4004
55%	1,091,241	Loyal Customer - Longest tenure, 2nd highest total spend, most recent purchasers	2	\$256,220	33%	365	1003	11	\$768	0%	3%	14%	14%	18%	9%	8%	13%	0%	4%	11%	5%	1%
4%	84,228	Small Spender -High Profit Margin CAT3001 Buyers	5	\$28,539	37%	52	623	233	\$569	0%	1%	14%	5%	42%	4%	0%	2%	0%	5%	9%	15%	2%
2%	33,450	CAT3004 Buyers - 2nd Highest \$/txn	4	\$108,515	26%	51	528	157	\$2,505	1%	1%	4%	2%	6%	7%	14%	47%	5%	1%	5%	1%	9%
2%	36,891	Small Spender - CAT2002 and CAT4002 Buyers	6	\$29,237	36%	44	577	203	\$680	6%	0%	25%	6%	8%	2%	1%	8%	0%	2%	36%	5%	0%
3%	63,402	Smallest Total Spend - Most spend in CAT3002 and CAT2003	1	\$23,002	34%	50	505	237	\$589	0%	9%	8%	20%	12%	19%	4%	6%	1%	10%	7%	3%	0%
34%	680,156	High Value (highest spend, highest total transactions, highest spend per transaction, average tenure	3	\$2,027,412	22%	7170	598	155	\$9,209	8%	5%	7%	5%	6%	7%	7%	19%	0%	5%	5%	3%	2%
	1,989,368																					