

Sales Data Analysis

elizabethh.com

ELIZABETH HO

Data Exploration and Feature Engineering

Sales Transactions Data

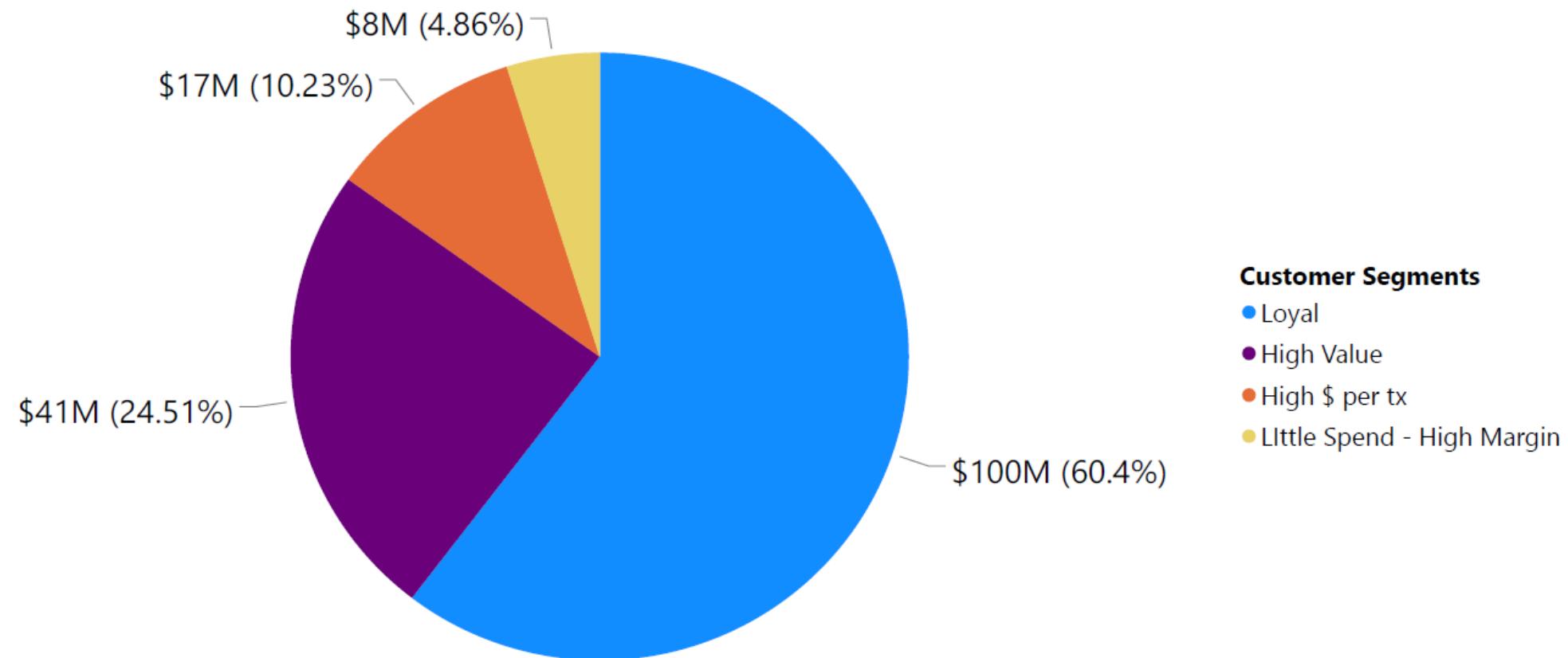
- 2,201,620 Line items
- Customer – 1,300
- Category - 14
- Subcategory - 293
- Part – 40,885
- Sales Price
- Sales Quantity
- Cost of Goods Sold
- Customer Number
- 329,554 Sales Orders
- Jan 1, 2015 – Oct 17, 2017

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

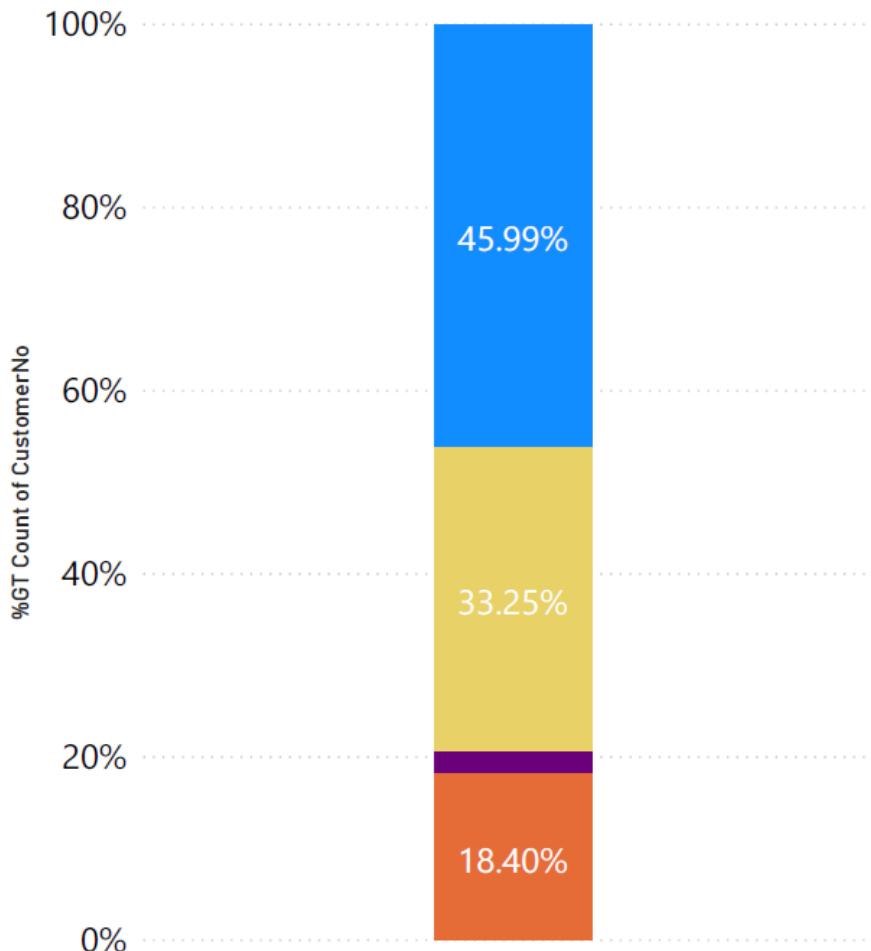


- Customer Segments**
- Loyal
 - High Value
 - High \$ per tx
 - Little Spend - High Margin

Customer Segmentation

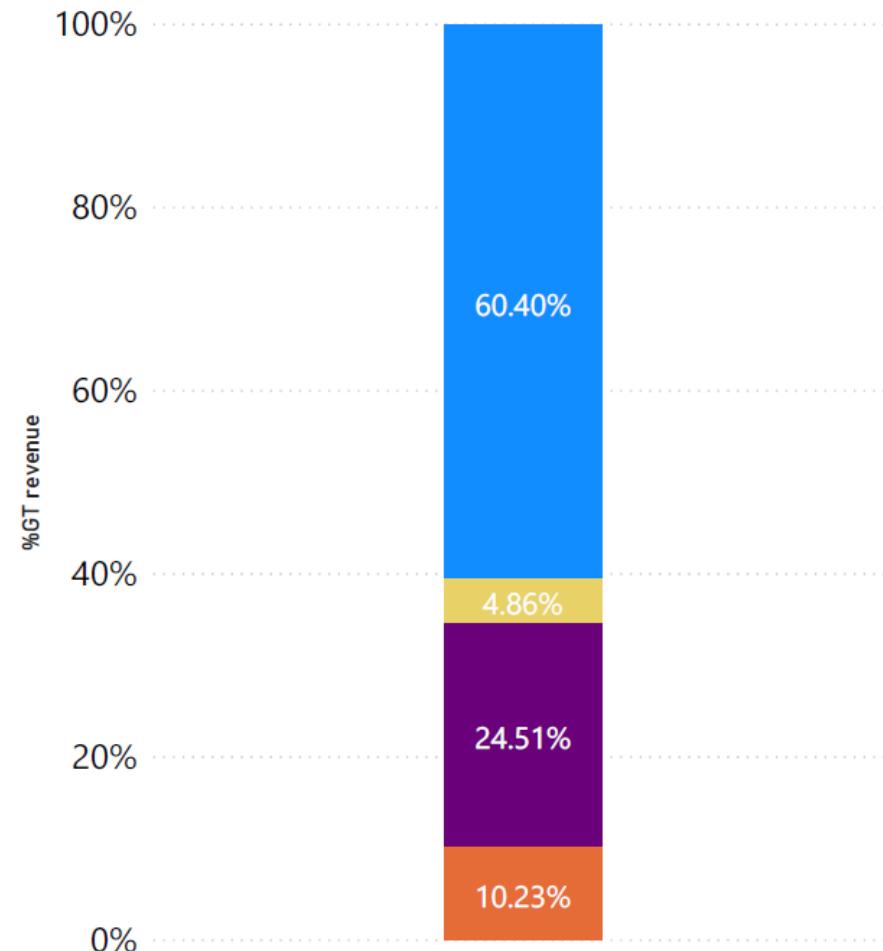
Segment Count

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



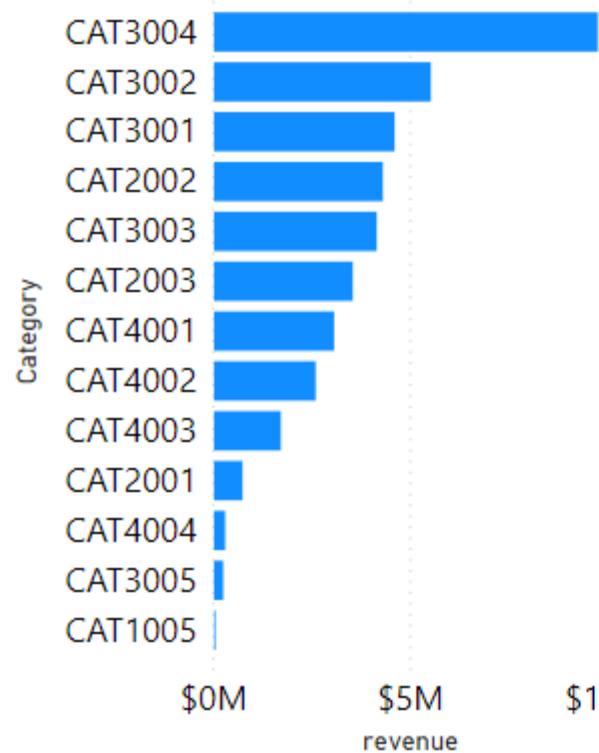
LTV by Description

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

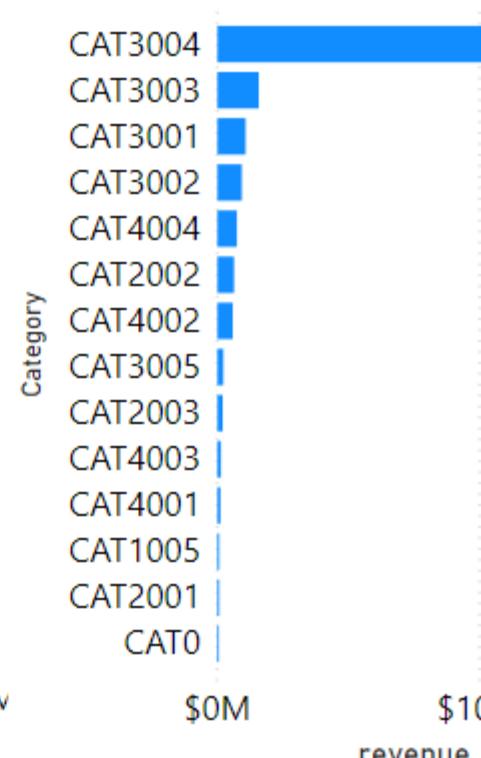


Customer Segmentation

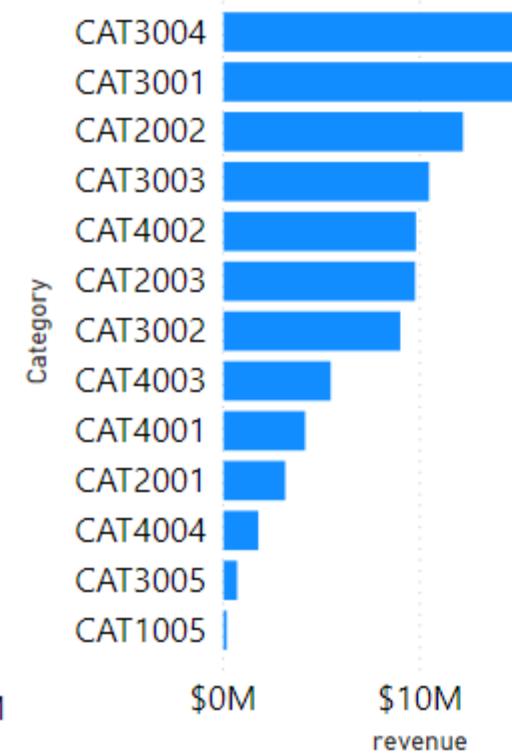
High Value



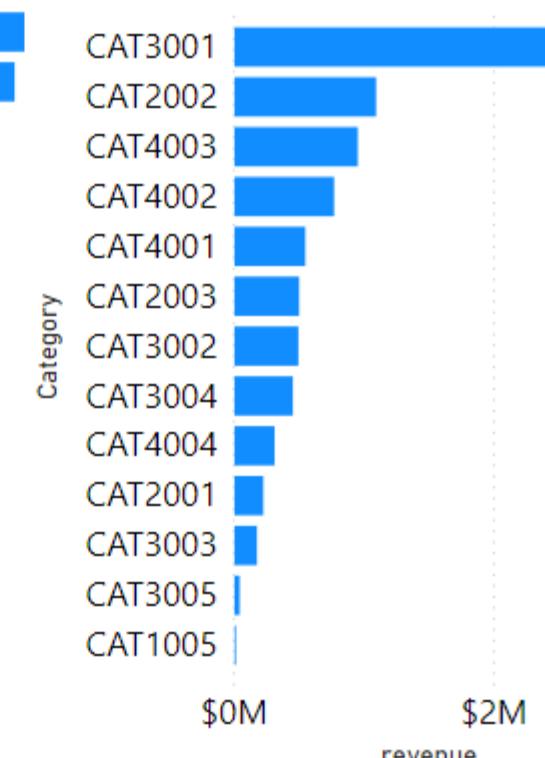
High \$ / TXN



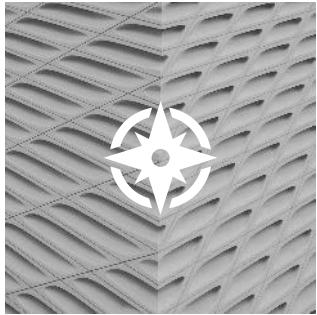
Loyal



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

ELIZABETH HO

Sales Data Analysis

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 = $P(\text{Milk} \& \text{Butter}) = 6/100 = 0.06$
- confidence = support/ $P(\text{Butter}) = 0.06/0.08 = 0.75$
- lift = confidence/ $P(\text{Milk}) = 0.75/0.10 = 7.5$

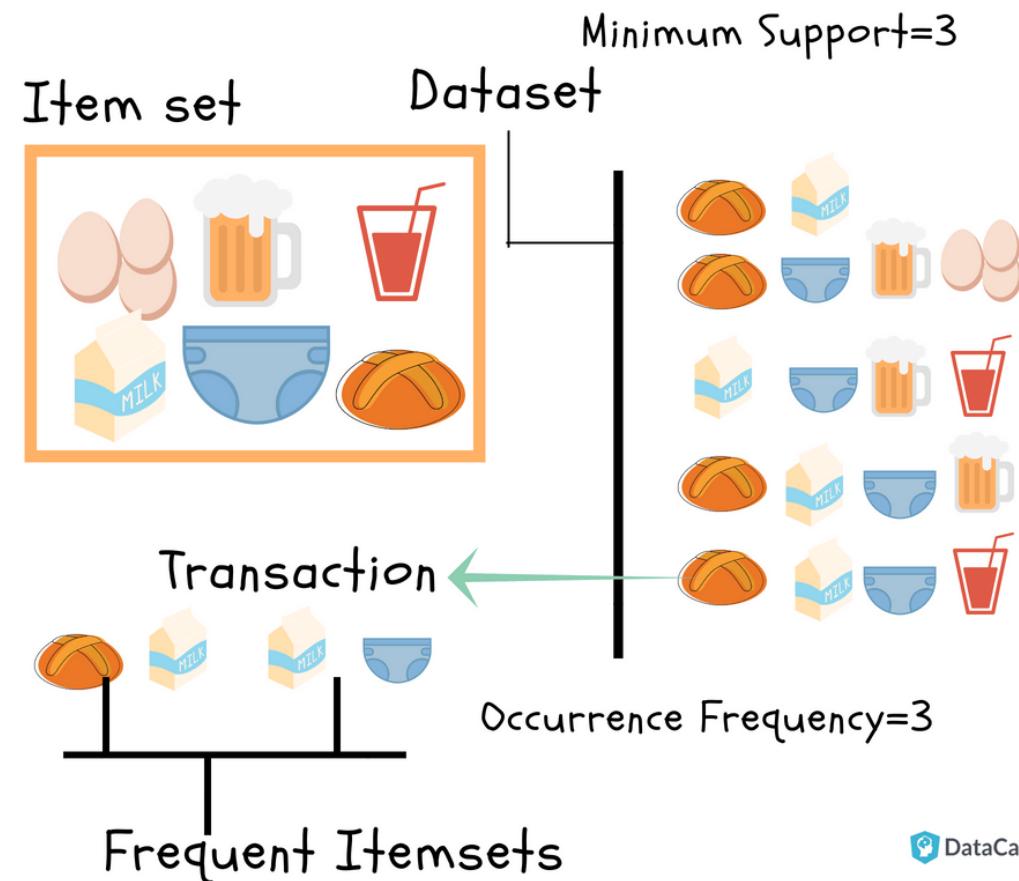


Most Popular Baskets

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

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

Rule: $X \Rightarrow Y$

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

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



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

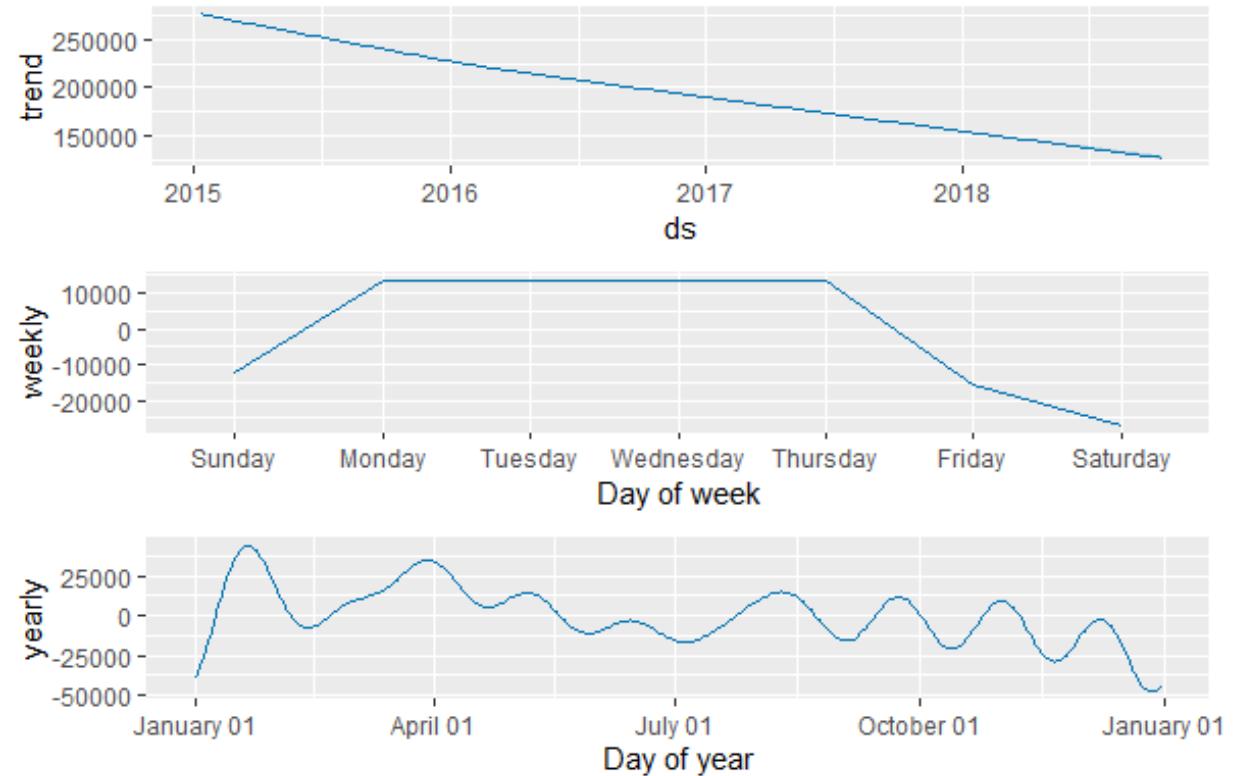
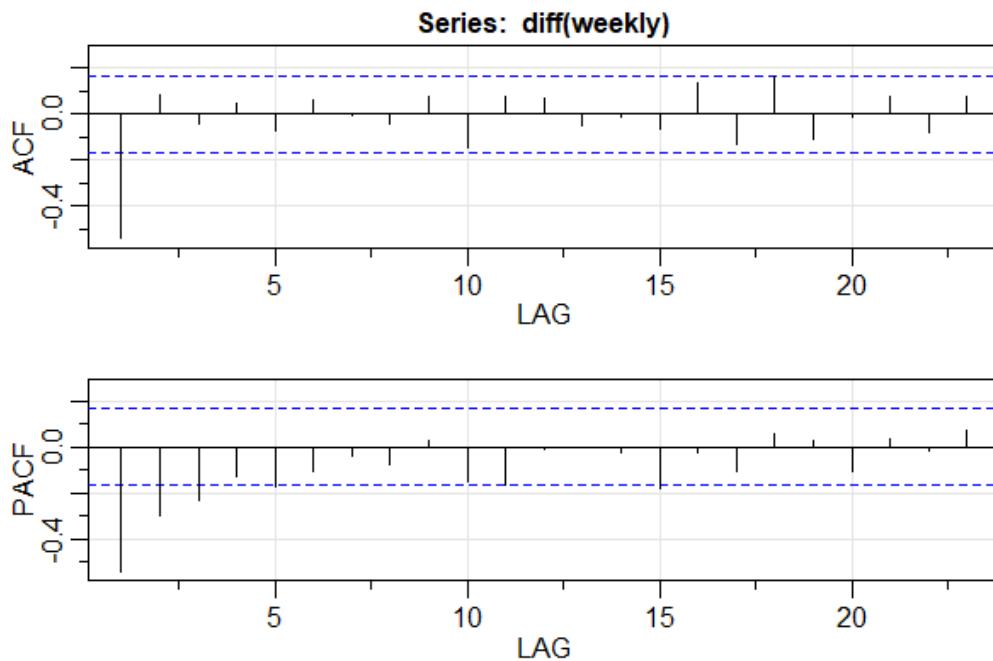
Profit

Sales By Month



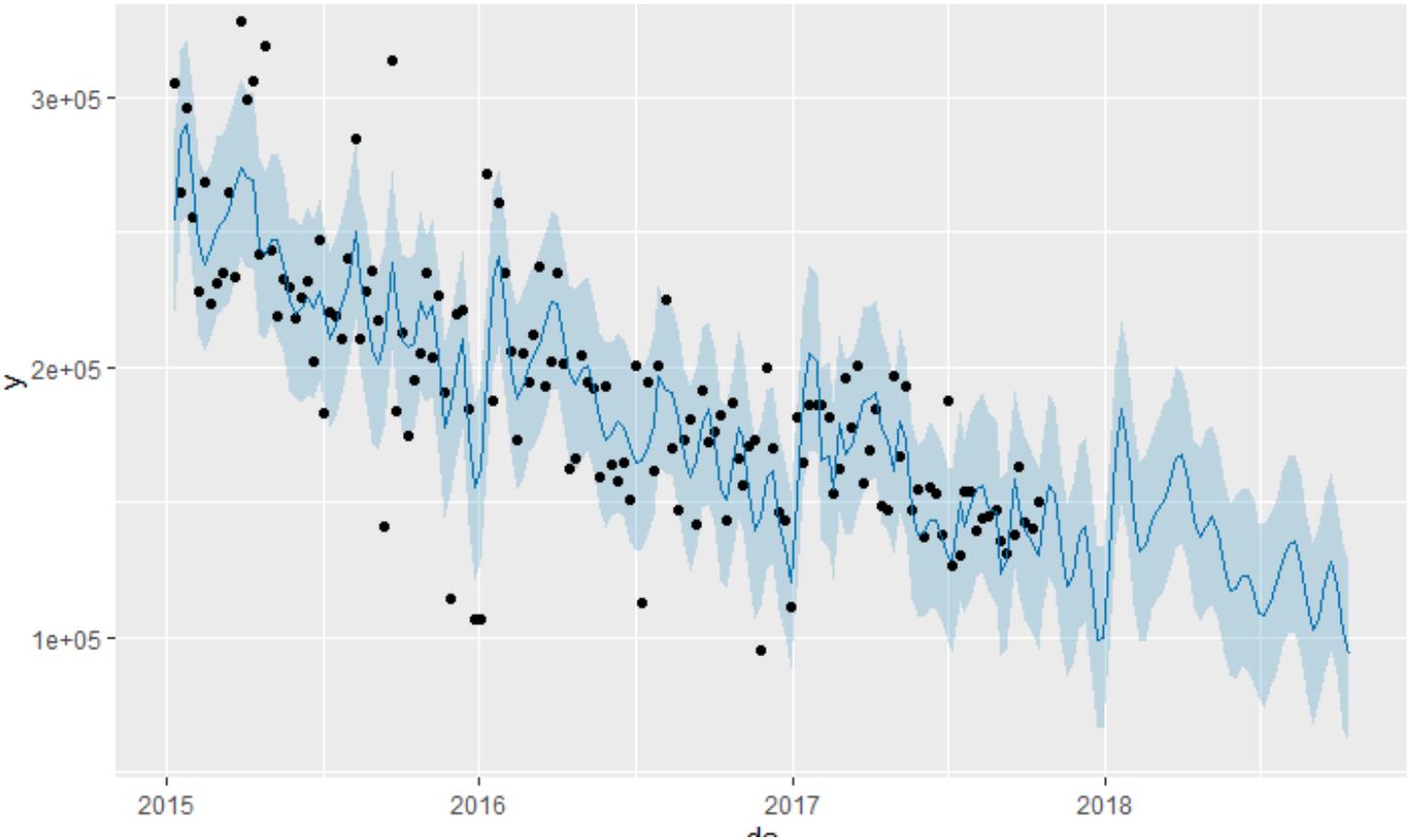
Sales Forecast

- 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



ELIZABETH HO

Sales Data Analysis

Thank You

• Elizabeth Ho



Appendix

Customer Segmentation