TOPIC
Big Data Analytics & Machine Learning 101

DATE
April 26, 2017
Agenda

› 2017 Big Data trends

› What is Big Data? Why now?

› What is Machine Learning? Why now?

› Statistical Analysis - 101

› Case Studies
  ○ A Machine Learning Algorithm to Recommend Personalized Offers for a Leading Retailer
  ○ How Machine Learning Can Help Optimize Frame Assortment for an Optical Retailer?
  ○ Healthcare Analytics for Maryland Department of Human Resources
BIG DATA TRENDS & FRAMEWORK
Oil produces more than diesel, butane, and kerosene.

E.g. golf balls

Data produces more than monthly BI reports.

E.g. analytics embedded apps for increased revenue & profit
Cloud Enabled FAST Analytics
Insights & Outcomes vs Tools
Visual Data Discovery tools want to go Enterprise
Edge Analytics (Embedded BI)
Data Storytelling

Enterprise gets Comfortable with Open Source
Predictive Analytics goes mainstream, impacting multiple business functions
Discovery Analytics vs Operational BI
Real-time, Machine Learning & Artificial Intelligence
Self Service Data Prep and Analytics

10 DATA ANALYTICS TRENDS FOR 2017
ANALYTICAL MATURITY ROADMAP

SENSE & RESPOND

PREDICT & ACT

RAW DATA
CLEANED DATA
STANDARD REPORTS
AD HOC REPORTS & OLAP
GENERIC PREDICTIVE ANALYTICS
PREDICTIVE MODELING
OPTIMIZATION
OPERATIONALIZED & PRESCRIPTIVE ANALYTICS

PROFITABILITY & COMPETITIVE ADVANTAGE

QUALITY & COMPLETENESS OF INSTRUMENTATION, DATA PLATFORMS, ANALYTICS, & DATA SCIENCE

What Happened?

Why Did It Happen?

What Will Happen?

What’s the Best Possible Outcome?

Can we Automate the Action to Take?

Make avail to the most people.

Embed within Mobile Apps, for real-time decisioning support in context
Reference Architecture - Data Flow and Access

- **Information Portal**
  - Business Operations
  - Information Analyst

- **Analytics Workbench**
  - Decision Makers
  - Data Discovery
  - Data Mining
  - Advanced Analytics

- **Data Science Laboratory**
  - Business Operations
  - Data Scientist

- **IT**
  - Report Writer
  - Enterprise Data warehouse
  - Local Data Marts
  - Ad Hoc Files
  - Social
  - Open Data
  - External Data Providers
  - Text

- **Information**
  - Information Producers
  - Information Consumers

- SOURCE: GARTNER
BIG DATA ARCHITECTURE TO INTERACT ACROSS EVERY DIGITAL TOUCHPOINT

- ALERTS & APPLICATIONS
- OPERATIONAL REPORTING
- BUSINESS INSIGHTS
- CUSTOMER INSIGHTS
- WEB & MOBILE REPORTS & AD-HOC ANALYSIS
- HIGH PERFORMANCE DB (IMDB)

- REAL TIME OFFER ENGINE
- LOYALTY OFFER ANALYSIS
- PREDICTIVE DEMAND MODELS
- PROFILING
- SEGMENTING
- REPORTING
- TRENDING
- MINING
- MODELING
- USAGE ANALYTIC MODELS

- ENTERPRISE DATA WAREHOUSE
- BIG DATA STORAGE

CLIENT INTERNAL DATA SOURCES (STRUCTURED)
- TRANSACTIONAL DATA (POS)
- CUSTOMER PROFILE
- CUSTOMER PREFERENCES
- LOYALTY PROGRAMS DATA
- BILLING & AUTHORIZATION
- ASSETS USAGE DATA

CUSTOMER INFO VIA EXTERNAL SOURCES (UNSTRUCTURED)
- SOCIAL LISTENING
- EMAIL, MOBILE
- WEB CLICKS
- CALL CENTER / CRM SURVEYS
Operationalize Data Science

Solution Framework For Automated/Real-time Processing

Big Data Platform

Aggregated Reference Data

Cached Data for real-time

Presentation Layer & Applications

Visualizations

Mobile

Events

Real-time Decision Engine

Event Stream Processing

Predictive Model updates

Business Analysts
Configure Business Rules

Data Scientist
Predictive Analytics/R

BI Developers
KPI Visualizations & Reports

Alerts

Cloud, Mobile

Applications

Sensor_Log

Cooling_Control_TREE

Cooling_Control

Usage & Service Level Analytics
WHAT IS BIG DATA?
WHY NOW?
The 3 V’s of Big Data

Data Volume
- Transactions
- Terabytes
- Petabytes
- Records
- Files

Data Velocity
- Real-Time
- Near Real-Time
- Batches
- Streams

Data Variety
- Structured
- Unstructured
- Semi-Structured
- Social
- Mobile

BIG DATA
The Emergence of New Data Sources in Recent Times Has Made it Imperative for Organizations to Adopt Big Data Analytics

Why Now?

Clickstream Data
(Online Behavior)

Emergence of New Data Sources

Sensor Data
(GPS, M2M)

Social Media
(FB, Twitter, LinkedIn, etc.)

HYPERCONNECTIVITY AND MOBILITY ARE CHANGING EVERYTHING PROVIDING NEW OPPORTUNITIES
WHAT IS MACHINE LEARNING?
What is Machine Learning?

"Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed."

- Arthur Samuel
(Pioneer in Computer Science)

- **Supervised**
  - Algorithm is "trained" on a pre-defined set of data, to facilitate predictive modeling on new data. Example are:
    - Linear Regression
    - Logistic Regression
    - Classification
    - Neural Network

- **Unsupervised**
  - Algorithm finds inherent patterns and relationships in the data:
    - Clustering
    - Association Rules
Why Now?

1. **MATURED FIELD** - The field of Machine Learning has matured a lot in the last decade relying heavily on the field of statistics and is moving towards “Statistical Machine Learning”. The tools that implement these algorithms or methods have also been maturing, such as R and Python.

2. **ABUNDANT DATA** - There is an abundance of data right now, from email, social networking, blogs, etc. Machine Learning provides the methods to convert data collected from disparate sources into meaningful information to drive decision making.

3. **ABUNDANT COMPUTATION** - Abundant and cheap computation has driven the abundance of data we are collecting and the increase in capability of machine learning methods.
STATISTICAL ANALYSIS 101
Supervised Vs Unsupervised Algorithms

In supervised learning, the output datasets are provided which are used to train the machine and get the desired outputs whereas in unsupervised learning no datasets are provided, instead the data is clustered into different classes.

- **Supervised learning**: discover patterns in the data that relate data attributes with a target (class) attribute.
  - These patterns are then utilized to predict the values of the target attribute in future data instances.

- **Unsupervised learning**: The data have no target attribute.
  - We want to explore the data to find some intrinsic structures in them.

- Regression, Decision tree
- K-Means, PCA
What it answers?

1. Is this A or B?
   Classification

2. Is this weird?
   Anomaly Detection

3. How much – or – How many?
   Regression

4. How is this organized?
   Clustering

5. What should I do next?
   Reinforcement Learning
Is this A or B?

**Classification:** Identify what category new information belongs in

**Predict Between Two Categories**

Two-Class Classification

- **Answers simple**
  - two-choice questions, like yes-or-no, true-or-false.
  - Is this tweet positive?
  - Will this customer renew their service?
  - Which of two coupons draws more customers? A/B

**Predict Between Several Categories**

Multi-Class Classification

- **Answers complex**
  - questions with multiple possible answers.
  - What is the mood of this tweet?
  - Which service will this customer choose?
  - Which of several promotions draws more customers? A/B/C

**Algorithms:** Two-class logistic regression, Two-class boosted decision tree, Multiclass neural network, Multiclass random forest etc.
**Find unusual occurrences**

**Anomaly Detection:** Identify and predict rare or unusual data points

**Algorithms:** One Class SVM, PCA-Based Anomaly Detection etc.

*Is this weird?*
How much – or – How many?

**Regression**: Forecast the future by estimating the relationship between variables.

**Algorithms**: Linear Regression, Bayesian Linear Regression etc.
How is this organized?

**Clustering:** Separate similar data points into intuitive groups.

**Algorithms:** K-Means (Centroid based), Expectation-Maximization (Distribution based) etc.
What should I do next?

Reinforcement Learning: These algorithms learn from outcomes and decide on the next action.

If I’m a temperature control system for a house: Adjust the temperature or leave it where it is?

Predict next decision for a self-driving car: At a yellow light, brake or accelerate?

For a robot vacuum: Keep vacuuming, or go back to the charging station?

Algorithms: Monte-Carlo method, Genetic algorithms, Neural network etc.
A FEW EXAMPLES OF OUR WORK
Our Work Samples

**Data Driven Personalization** - A Machine Learning Algorithm to Recommend Personalized Offers

**Frame Assortment Optimization** - How Machine Learning Can Help Optimize Frame Assortment

**A Data Driven Approach to Enable “Independence from Support”** - Healthcare Analytics for Maryland Department of Human Resources
Data Driven Personalization
Data Driven Personalization

SITUATION
Victoria's Secret was interested in leveraging Clickstream (Coremetrics) & Customer Demographics data to analyze Path to Purchase and uncover insights & opportunities with specific focus on optimizing shipping costs, as customer research indicated that shipping charges were the primary reason for not purchasing.

SOLUTION
DMI developed a Machine Learning algorithm to create customer segments based on key dimensions like customer loyalty, price sensitivity and product value groups. A "shipping sensitivity score" was established for each customer based on past purchase behavior. For each customer segment, DMI recommended the optimal shipping promotion or discount against constraints such as minimum order size, price sensitivity & shipping margin.

RESULTS
1. With the proposed shipping promotion to all price sensitive customers, the estimated increase in net sales is 13%
2. Savings of up to 2% is expected by not providing shipping offers to non-price sensitive customers
3. This resulted in an overall estimated increase of 15% net sales, while protecting the shipping revenue.
We started by getting a better understanding of VSD’s customers

How loyal they are?
Integrated view of their shopping behavior across omnichannel platforms (in-store, online, mobile)

How Price Sensitive they are?
Understanding their willingness to pay by looking at their shopping discounts, products purchased, offers redeemed

Customer Centric Shipping Strategy
Developing customer centric shipping propositions to optimize shipping revenue and maximize satisfaction.
Designed a **customer centric** approach to optimize shipping promotions

- Segment based on key dimensions: sales volume and frequency of purchase over the period of analysis
- Establish normative benchmarks for each segment
- Analyze key measures across segments

- Establish “shipping sensitivity score” for each customer based on past shipping promotions, discounts availed on-line & in-store and their purchase behavior across product groups

- For each customer segment, identify optimal shipping promotion threshold against constraints such as minimum order size & shipping margins
- Build sensitivity model to impute potential lift in shipping & sales volume across customer segments?

- Produce actionable shipping recommendations & promotions for granular & finite customer segments based on:
  - What are the optimal threshold(s) for basket order shipping promotions?
  - What are shipping promotion levels eg.. 50% off shipping for orders above $50?
Current Solution Architecture & Future consideration for Operationalization

DATA SOURCES
- MAINFRAMES
- COREMETRICS
- CDM

MANTA STORAGE
Data Processing & Transformations using CLI (Unix) & R SDKs

32 Bit Files with Summarized/Consolidated data

64 Bit
Customer Segmentation & Shipping Sensitivity analysis

Data File with Segmentation & Analytical Flags
Deliverable with analysis results
Presentation with methodology, results & recommendations

FUTURE CONSIDERATION
Display Customer centric offers on Victoria’s Secret website & App/ Send offers via e-mail
Customer Insights Portal & KPIs

Data table with Segmentation & Analytical Flags (periodically refreshed)
Analyzed High Volume, High Velocity data from three source systems (~10bn rows)

**Mainframes**
- Order header table (~32M rows)
- Order detail table (~75M rows)
- Allocation table (~88M rows)
- Customer Demographics (~124M rows)
- Offers Table (~20M rows)

**Coremetrics**
- Page View table (~6B rows)
- Order table (~18M rows)
- Cart Item Addition (~483M rows)
- Conversion Event (~3B rows)
- Registration (~452M rows)
- Geography (~888M rows)

**CDM**
- Customer Monthly In Store Sales Summary data table (~90M rows)

*We looked into 12 months of data from all the sources (July 2014 – June 2015)

**Key Takeaways**

1. Looking at the breadth and depth of data, VSD is well positioned to leverage BIG DATA to drive better decisions

2. We see significant opportunity to not only build upon “Shipping Promotions” but leverage such insights to influence other sections of business (i.e. product promotions, product assortment, labor optimization, etc.)
Data suggested **7 Distinct** Customer Loyalty segments

- **# of Orders - Online**
  - Customer Value Score - 1: Shoppers buying low value products; but Online Avg Spend is greater than In Store Avg Spend
- **$Spend - Online**
  - Customer Value Score - 2: Shoppers buying moderate value products; but Online Avg Spend is greater than In Store Avg Spend
- **# of Orders – In Store**
  - Customer Value Score - 4: Valuable Customers - High In Store shoppers with lesser Online shopping; And Shop multiple departments
- **$Spend – In Store**
  - Customer Value Score - 5: Very Loyal Customers - High In Store shoppers with lesser Online shopping; And Shop multiple departments
- **# of Departments shopped**
  - Very Loyal Customers - High Shopping Online than In Store; And Shop multiple departments
- **Average Price per Item (before discounts)**
  - Premium Shoppers - High Shopping both Online & In Store; And Shop multiple departments

**Customer Value Score Scale:** 1 - Low → 7 - Very High
Similarly we created 7 different price sensitivity segments

1. **Price Sensitivity Score - 1**: Very Low Price Sensitive (both Product & Shipping)
2. **Price Sensitivity Score - 2**: Low Product price sensitive; but Very high Shipping price sensitive;
3. **Price Sensitivity Score - 3**: Medium Product Price Sensitive; and Low Shipping Price Sensitive
4. **Price Sensitivity Score - 4**: Medium Price Sensitive towards both Product & Shipping: More discounts In Store
5. **Price Sensitivity Score - 5**: High product price sensitive; And Very High Shipping price sensitive
6. **Price Sensitivity Score - 6**: Very high price sensitive towards Products & High price sensitive towards Shipping; Redeems lot of Offers
7. **Price Sensitivity Score - 7**: Very High Product price sensitive: but Low Shipping Price Sensitive

**Price Sensitivity Score Scale**: 1 - Low → 7 - Very High
Customer Value & Price Sensitivity score combination is an effective way to identify targets for shipping promotions.

**Target 1:** High Customer Value & High Price Sensitive

**Target 2:** Medium Customer Value or High Price Sensitive

**Target 3:** Medium Customer Value & Medium Price Sensitive

**Target 4:** Low Customer Value

**Target 5:** Low Price Sensitive

%Customers by Customer Value Score vs. Price Sensitivity Score:

<table>
<thead>
<tr>
<th>Customer Value Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.17%</td>
<td>9.37%</td>
<td>7.02%</td>
<td>6.79%</td>
<td>7.43%</td>
<td>0.00%</td>
<td>6.66%</td>
</tr>
<tr>
<td>2</td>
<td>10.53%</td>
<td>6.74%</td>
<td>2.99%</td>
<td>3.74%</td>
<td>2.96%</td>
<td>0.00%</td>
<td>0.84%</td>
</tr>
<tr>
<td>3</td>
<td>0.75%</td>
<td>0.21%</td>
<td>0.03%</td>
<td>0.17%</td>
<td>0.05%</td>
<td>0.01%</td>
<td>0.01%</td>
</tr>
<tr>
<td>4</td>
<td>1.39%</td>
<td>1.27%</td>
<td>0.80%</td>
<td>4.13%</td>
<td>1.77%</td>
<td>0.63%</td>
<td>0.59%</td>
</tr>
<tr>
<td>5</td>
<td>0.05%</td>
<td>0.04%</td>
<td>0.04%</td>
<td>0.13%</td>
<td>0.03%</td>
<td>0.99%</td>
<td>0.01%</td>
</tr>
<tr>
<td>6</td>
<td>0.70%</td>
<td>1.56%</td>
<td>1.75%</td>
<td>3.17%</td>
<td>2.41%</td>
<td>1.53%</td>
<td>0.26%</td>
</tr>
<tr>
<td>7</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.01%</td>
<td>0.01%</td>
<td>0.00%</td>
<td>0.25%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

%Customers: 24% 19% 13% 18% 15% 3% 8% 100%

*Customer Value Score Scale: 1 - Low → 7 - Very High

*Price Sensitivity Score Scale: 1 - Low → 7 - Very High*
We developed an optimization model that is designed to **drive engagement** and **protect** shipping revenue.

Customers in each Target group are classified based on their Avg. Basket size band:

- **Target 1:** High Customer Value & High Price Sensitive
- **Target 2:** Medium Customer Value or High Price Sensitive
- **Target 3:** Medium Customer Value & Medium Price Sensitive
- **Target 4:** Low Customer Value
- **Target 5:** Low Price Sensitive

$0- $50  \quad >$50- $100

>$100- $150  \quad >$150- $200

>$200- $250  \quad >$250- $300

>$300

- Stretch the basket size of each customer to upper bound by providing shipping discounts for a min $ purchase (i.e. upper bound of the basket size band).
- Provide higher shipping discounts for high value customers or high price sensitive customers and lower discounts for low price sensitive or low value customers.
The model allocates **relevant** shipping discount for each customer group maximizing the revenue

**Objective Function:**
Maximize change in total revenue (i.e. incremental revenue from stretching basket size + change in shipping revenue)

**Constraints:**

1. Min/Max discounts for each customer group:

<table>
<thead>
<tr>
<th>Avg Basket Size (S$Spent per Order) band</th>
<th>Target1</th>
<th>Target2</th>
<th>Target3</th>
<th>Target4</th>
<th>Target5</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 &gt; 0-50</td>
<td>Min 100%</td>
<td>Max 100%</td>
<td>Min 100%</td>
<td>Max 100%</td>
<td>Max 100%</td>
</tr>
<tr>
<td>02 &gt; 50-100</td>
<td>Min 75%</td>
<td>Max 100%</td>
<td>Min 50%</td>
<td>Max 100%</td>
<td>Max 100%</td>
</tr>
<tr>
<td>03 &gt; 100-150</td>
<td>Min 100%</td>
<td>Max 100%</td>
<td>Min 100%</td>
<td>Max 75%</td>
<td>Max 100%</td>
</tr>
<tr>
<td>04 &gt; 150-200</td>
<td>Min 100%</td>
<td>Max 100%</td>
<td>Min 100%</td>
<td>Max 75%</td>
<td>Max 100%</td>
</tr>
<tr>
<td>05 &gt; 200-250</td>
<td>Min 100%</td>
<td>Max 100%</td>
<td>Min 100%</td>
<td>Max 100%</td>
<td>Max 100%</td>
</tr>
<tr>
<td>06 &gt; 250-300</td>
<td>Min 100%</td>
<td>Max 100%</td>
<td>Min 100%</td>
<td>Max 75%</td>
<td>Max 100%</td>
</tr>
<tr>
<td>07 &gt; 300</td>
<td>Min 100%</td>
<td>Max 100%</td>
<td>Min 100%</td>
<td>Max 100%</td>
<td>Max 100%</td>
</tr>
</tbody>
</table>

2. Change in Shipping charges paid <= 10%

**Recommendation:**

The model allocates relevant shipping discount for each customer group maximizing the revenue.
An uplift of \(~15\%\) in net revenue (order amount+shipping rev) is estimated with the recommended shipping promotions.
Frame Assortment Optimization
OPTIMIZING PRODUCT CATEGORIES TO MAXIMIZE GROSS MARGIN

A retail client asked us to determine the optimal mix of product categories in a specific store brand to maximize gross margins.

SOLUTION
DMI built nonlinear regression models to estimate gross margin as a function of number facings displayed by product category. We also developed an optimization model to find the optimal mix of categories constrained by the facings in each category. Constraints varied by product category. Finally, we created a tool to execute the model at a store level.

RESULTS
The gross margin is expected to increase by 14% across targeted stores with the recommended optimal mix of frame categories for display.
Frame Category Optimization - Background and Objective

**Background:** For Pearle Vision stores, the ratio of frames sold and faced indicates that there is an opportunity to optimize the %facings by frame category. Profit per facing data also shows that there is a potential to increase the facings for a few categories with higher profit compared to others.

**Objective:** Determine optimal mix of frame categories to maximize $Gross Margin

<table>
<thead>
<tr>
<th>Store Segment</th>
<th>Store Count</th>
<th>% Luxury</th>
<th>% Premium</th>
<th>% Lifestyle</th>
<th>% Sport</th>
<th>% Fashion</th>
<th>% Private Label</th>
<th>Profit Per Facing</th>
<th>Luxury $GM/Facing</th>
<th>Premium $GM/Facing</th>
<th>Lifestyle $GM/Facing</th>
<th>Sport $GM/Facing</th>
<th>Fashion $GM/Facing</th>
<th>Private Label $GM/Facing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada High Affluence</td>
<td>22</td>
<td>0.95</td>
<td>1.08</td>
<td>1.88</td>
<td>1.23</td>
<td>0.93</td>
<td>0.87</td>
<td>$173</td>
<td>$272</td>
<td>$371</td>
<td>$179</td>
<td>$181</td>
<td>$143</td>
<td></td>
</tr>
<tr>
<td>Canada Low Affluence</td>
<td>34</td>
<td>0.45</td>
<td>0.94</td>
<td>1.64</td>
<td>1.16</td>
<td>0.98</td>
<td>1.07</td>
<td>$119</td>
<td>$218</td>
<td>$295</td>
<td>$224</td>
<td>$171</td>
<td>$156</td>
<td></td>
</tr>
<tr>
<td>US High Affluence</td>
<td>18</td>
<td>0.74</td>
<td>1.04</td>
<td>1.59</td>
<td>1.35</td>
<td>0.77</td>
<td>0.94</td>
<td>$199</td>
<td>$227</td>
<td>$283</td>
<td>$243</td>
<td>$130</td>
<td>$127</td>
<td></td>
</tr>
<tr>
<td>US Mid-High Affluence</td>
<td>50</td>
<td>0.67</td>
<td>0.98</td>
<td>1.56</td>
<td>1.19</td>
<td>0.83</td>
<td>1.07</td>
<td>$154</td>
<td>$190</td>
<td>$245</td>
<td>$184</td>
<td>$127</td>
<td>$126</td>
<td></td>
</tr>
<tr>
<td>US Mid-Low Affluence</td>
<td>43</td>
<td>0.66</td>
<td>0.95</td>
<td>1.38</td>
<td>1.25</td>
<td>0.85</td>
<td>1.09</td>
<td>$141</td>
<td>$174</td>
<td>$204</td>
<td>$178</td>
<td>$120</td>
<td>$122</td>
<td></td>
</tr>
<tr>
<td>US Low Affluence</td>
<td>12</td>
<td>0.58</td>
<td>0.85</td>
<td>1.2</td>
<td>1.2</td>
<td>0.81</td>
<td>1.39</td>
<td>$106</td>
<td>$135</td>
<td>$153</td>
<td>$143</td>
<td>$95</td>
<td>$135</td>
<td></td>
</tr>
</tbody>
</table>
Frame Category Optimization - Approach

Conducted analysis for US Mid High Affluence Segment (50 Stores)

1. Developed Regression Models to estimate Gross Margin$
   - Built non-linear regression models to estimate $Gross Margin dependent on #facings for each frame category
   - A diminishing return of Gross Margin is noticed beyond certain #facings
Frame Category Optimization - Approach

2. Built Optimization Model to find Optimal %Facings

- Grouped stores based on #facings (high, medium, low) and applied constraints by frame category based on management discretion
- Developed optimization models at a store level to find optimal %facings by maximizing $Gross Margin subject to constraints. Created excel automation to run the models at store level

Maximize:
Total Gross Margin = f_1(L*T) + f_2(P*T) + f_3(F*T) + f_4(LS*T) + f_5(S*T) + f_6(PL*T)

Subject to Constraints:
1. L + P + F + LS + S + PL = 100%
2. Min_L <= L*T <= Max_L
3. Min_P <= P*T <= Max_P
4. Min_F <= F*T <= Max_F
5. Min_LS <= LS*T <= Max_LS
6. Min_S <= S*T <= Max_S
7. Min_PL <= PL*T <= Max_PL
8. 0% <= L, P, F, LS, S, PL <=100%

T - #Total Facings available;
L - %Luxury; P - %Premium; F - %Fashion; LS - %LifeStyle;
S - %Sport; PL - %Private Label facings
Applying the model against the average medium facing store produces the following projected results.
Non Linear Curves & Lower/Upper Bounds

**Luxury**

**Equation:**
\[ 27556.17 - 635.71 \times \text{Nbr\_Facings} + 6.74 \times \text{Nbr\_Facings}^2 - 0.02 \times \text{Nbr\_Facings}^3 \]

**R-Square:** 0.459

- **Minimums**
  - Low: 47
  - Medium: 75
  - High: 80

- **Maximums**
  - Low: 160
  - Medium: 160
  - High: 220


**Premium**

**Equation:**
\[ 189926.8 - 2435.72 \times \text{Nbr\_Facings} + 11.59 \times \text{Nbr\_Facings}^2 - 0.016 \times \text{Nbr\_Facings}^3 \]

**R-Square:** 0.353

- **Minimums**
  - Low: 160
  - Medium: 160
  - High: 160

- **Maximums**
  - Low: 330
  - Medium: 250 or 27%
  - High: 330 or 28%


**Fashion**

**Equation:**
\[ -534930.28 + 13262.71 \times \text{Nbr\_Facings} - 118.95 \times \text{Nbr\_Facings}^2 + 0.47 \times \text{Nbr\_Facings}^3 - 0.0007 \times \text{Nbr\_Facings}^4 \]

**R-Square:** 0.330

- **Minimums**
  - Low: 100
  - Medium: 100
  - High: 100

- **Maximums**
  - Low: 133
  - Medium: 133
  - High: 133
Non Linear Curves & Lower/Upper Bounds

**Lifestyle**

Equation:
\[ 130108.59 \cdot 3468.72 + 32.44 \cdot \text{Nbr Facings}^2 + 0.0898 \cdot \text{Nbr Facings}^3 \]

R-Square: 0.220

Minimums
- Low: 90
- Medium: 90
- High: 90

Maximums
- Low: 160
- Medium: 160
- High: 160

**Sport**

Equation:
\[ 6600.395 - 135.73 \cdot \text{Nbr Facings} + 5.01 \cdot \text{Nbr Facings}^2 - 0.026 \cdot \text{Nbr Facings}^3 \]

R-Square: 0.176

Minimums
- Low: 50 or 9%
- Medium: 75
- High: 30

Maximums
- Low: 111
- Medium: 111
- High: 111

**Private Label**

Equation:
\[ -138839.42 + 1444.61 \cdot \text{Nbr Facings} - 3.21 \cdot \text{Nbr Facings}^2 \]

R-Square: 0.126

Minimums
- Low: 140
- Medium: 140
- High: 140

Maximums
- Low: 190
- Medium: 190 or 21%
- High: 190
An increase of ~14% in Gross Margin was estimated across all stores with the recommended optimal %facings.

The avg %facings recommended for High, Medium & Low Facings stores are displayed below:

- **High Facing Stores**: Increase Premium; Decrease Fashion
- **Medium Facing Stores**: Decrease Luxury and Fashion; Increase Premium, Lifestyle and Private Label
- **Low Facing Stores**: Increase Lifestyle; Decrease Luxury, Fashion, Sport and Private Label
A Data Driven Approach to Enable “Independence from Support”
Key Business Questions that needs to be addressed through Analytics

1. Who are we? (What is the demographics and profile of our constituent population)

2. How much are we spending by each profile segment? – looking at claims and administrative costs and comparing it with the benchmarks/goals?

3. Where are we spending? – focusing on the utilization of services (including where the services are provided)

4. What can we change? – focusing on targeted actions that drives the most impact: training, placement and other programs to enable independence from support
A DATA DRIVEN APPROACH
To Analyze Existing Constituent Profiles & Proactively Accelerate “Independence from support”

1. Constituent Segmentation
   - Create constituent segments based on service (adoption, child support, etc.) and financial assistance type (emergency, medical assistance, etc):
     a. New Constituents, Temporary Assistance
     b. Economically Disadvantaged
     c. Repeat Constituents
     d. Permanent Beneficiaries
     e. Fraud & Program Abuse

2. Profiling & Analysis
   - Profile segments by demographics (Age, Income, Gender, Occupation, Location, etc.)
   - For example, 25% of constituents are located in Howard County that contribute to 40% of $spend, having a median age of 55
   - Identify steps or measures that drive “self-sustenance”

3. Outcomes & Potential Actions
   - Proactively manage and prioritize spend stemming out of segmentation and profiling insights
   - Recommend most frequent path to “Independence from support” in each segment
   - Tweak enrollment and eligibility rules based on usage insights
Segmentation with Disjoint Clustering – Example (using K-Means) Algorithm

- Repeat Constituents
- Permanent Beneficiaries
- Fraud & Program Abuse
- New Constituents, Temporary Assistance
- Economically Disadvantaged

Time to “Independence from Support” →

$\rightarrow$ Total Cost of Support

Analyze Historical Data
Systematic Analytical Process

Turning Data Into Action

- Analyze Historical Data
- Assess Current Constituents
- Conduct Profiling (Demographics & Digital Behavior)
- Identify Measures (Self-Sustenance)
- Recommend “Path to Independence From Support”

What happened?
Who are my constituents? (Identify Segments)
How do I proactively drive self-sustenance? (Analyze & Recommend)
DATA DRIVEN OUTCOMES
Accelerate “Independence from Support” - Based on Historical Trends (Illustrative)

Segment: New Constituents, Temporary Assistance
Micro-Segment: Income <$20K, Age 45-55, Montgomery County Residents

New Constituents (Temporary Assistance)  
- 20% Food Supplement Program
- 40% Soft Skills Training
- 55% Independence from support

Temporary Cash Assistance  
- 25% Energy Assistance
- 65% Supplemental Energy Program
- 75% Job Search & Placement

Energy Assistance
- 55% Supplemental Energy Program
- 65% Job Search & Placement
- 75% Independence from support

Food Supplement Program
- 40% Soft Skills Training
- 55% Independence from support

Independence from support
- 60% Supplemental Energy Program
- 75% Job Search & Placement

53
Proactively Drive Independence From Support Using Data Analytics

Proactively Manage and Support Constituents Towards “Self-Sustenance”
Visualization & Insights

**Geographical Distribution by County ($ Spend)**

- [Image of a map showing distribution by county]

**Constituent Segment Insights**

- Howard County

<table>
<thead>
<tr>
<th>Constituent Segment</th>
<th>Constituents</th>
<th>Total Spend</th>
<th>Spend %</th>
<th>Average Spend</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Constituents: Temporary Assistance</td>
<td>4,400</td>
<td>$900,000</td>
<td>30%</td>
<td>$210</td>
</tr>
<tr>
<td>Economically Disadvantaged</td>
<td>2,000</td>
<td>$800,000</td>
<td>40%</td>
<td>$400</td>
</tr>
<tr>
<td>Repeat Constituents</td>
<td>2,300</td>
<td>$800,000</td>
<td>23%</td>
<td>$346</td>
</tr>
<tr>
<td>Permanent Beneficiaries</td>
<td>1,000</td>
<td>$100,000</td>
<td>5%</td>
<td>$100</td>
</tr>
<tr>
<td>Fraud &amp; Program Abuse</td>
<td>300</td>
<td>$40,000</td>
<td>2%</td>
<td>$1,333</td>
</tr>
</tbody>
</table>

**Segment Distribution by Segment (# and $)**

- [Image of a bar chart showing distribution by segment]

**Demographics of Constituents**

- [Image of a demographic pie chart]

  - Age
  - Gender
  - Income
THANK YOU

THIAG LOGANATHAN
President, DMI Big Data Insights
tloganathan@dminc.com
Ph: 513.315.0260

© 2015 DMI CONFIDENTIAL & PROPRIETARY