PRESENTATION AGENDA

- About Us
- History
- The Need for R
- Dimension Reduction Techniques
- How MCA Helps
- R and MicroStrategy REST APIs
- MicroStrategy Demo
- Start to Finish / Big Picture
- Questions
MEET OUR TEAM

TIFFANY DELANO
Business Analyst II

JON GREENE
Business Analyst II

YASH SHAH
Data Scientist

ALISHA THOMPSON
Director BI, Analytics

DANIEL EGBE
Data Warehousing Programmer

SADANAND B R
Data Scientist Intern

Domtar
PRODUCTS PEOPLE RELY ON EVERY DAY

We design, manufacture, market, and distribute a wide variety of pulp, paper and personal care products from copy paper to baby diapers.

2016 Sales
US $5.1 Billion

18%
PERSONAL CARE

82%
PULP & PAPER

CUSTOMERS IN
50+ COUNTRIES

APPROXIMATELY
10,000 EMPLOYEES

STOCK SYMBOL: UFS
DOMTAR PERSONAL CARE TIMELINE

A series of diverse acquisitions serving customers in healthcare, retail, and directly to their homes.


- Attends Healthcare Products (U.S.)
- EAM (U.S.)
- AHP (U.S.)
- INDAS (Europe)
- HDIS (U.S.)
HISTORY
### THE RAW DATA

![Excel Spreadsheet]

The image shows a Microsoft Excel spreadsheet titled "Au_Factor_DiagData.xlsx". The spreadsheet contains a dataset with columns labeled A to S, and rows numbered from 148 to 187. The data appears to include various demographic and financial information. The spreadsheet is in a section titled "THE RAW DATA" in the document, suggesting it is part of a presentation or report at MICROSTRATEGY WORLD 2018.
### THE RAW DATA

**DATA SAV FILE**

<table>
<thead>
<tr>
<th>A - Z</th>
<th>..................</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA - AZ</td>
<td>..................</td>
</tr>
<tr>
<td></td>
<td>..................</td>
</tr>
<tr>
<td>ZZ - ZZ</td>
<td>AAZ ----- AZA</td>
</tr>
<tr>
<td></td>
<td>..................</td>
</tr>
<tr>
<td>AAA -- AAZ</td>
<td>HSG</td>
</tr>
</tbody>
</table>

8,000 COLUMNS

1460 ROWS

**FIND INSIGHTS FROM SURVEY RESULTS**

**DEVELOP BRAND STRATEGY BASED ON INSIGHTS FROM DATA**

MICROSTRATEGY WORLD 2018
THE NEED FOR R

Fat Data structure

Use algorithms to find insights from data quick – rather than trial and error

Column Names like D3_1_STANDARD

Meaningful Insights ASAP
DIMENSION REDUCTION TECHNIQUES
DIMENSION REDUCTION TECHNIQUES

- MULTIPLE CORRESPONDENCE ANALYSIS

- Appropriate for Survey Data: Categorical Data

- MCA: Works on a structure $X_{ik} = (Y_{ik} / P_k - 1)$
  - Tabular Contingency Tables – Finds Patterns – correlates similar to Chi-Sq methodology
  - MCA can be thought of as a technique to identify the dimensions that will maximize the inertia associated with a category
  - MCA groups individuals that are similar to each other and separates ones that aren’t – especially ones that don’t match on rare categories.
HOW MCA HELPS

Meaningful Information from three lifestyle questions:

<table>
<thead>
<tr>
<th>3 QUESTIONS</th>
<th># VARIABLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lifestyle Information</td>
<td>98 Variables (14 x 7 options)</td>
</tr>
<tr>
<td>Career Information</td>
<td>112 Variables (16 x 7 options)</td>
</tr>
<tr>
<td>Buy Insights Information</td>
<td>98 Variables (14 x 7 options)</td>
</tr>
</tbody>
</table>

- How important is career?
- How much time on NET?
- How tech savvy are you?
- What your friends think about you?
- Best for the baby?

DEAL SEEKERS
BUSY PARENTS – CONVENIENCE SEEKERS
ECO FRIENDLY – BEST FOR THE ENVIRONMENT
MAINSTREAM – WHAT THEY SAW ON TV
HOW MCA HELPS

Variance by Dimensions (Lifestyle)

How many dimensions can I reduce while preserving most information?

Answers to what questions set people apart from each other?

Having identified the questions that were most important, can I then start clustering users in logical business segments?
PRESERVING VARIANCE WHILE REDUCING VARIABLES

Contribution by Lifestyle Questions

% Contribution DMI

FIRST OF MY FRIENDS TO BUY THE LATEST PRODUCT
I AM MORE DIGITALLY CONNECTED THAN ANY OF MY FRIENDS
THE BRAND I BUY SAY A LOT ABOUT ME
OTHER QUESTIONS

Contribution by Lifestyle Questions

% Contribution DMI

Tried and Tested Rather Than Latest Trends
I Don’t Worry About Having the Latest Technology
There Is Risk in Buying Something I Am Not Familiar With
OTHER QUESTIONS
MCA OUTPUT – GROUPS FORMING

- Opposite Quadrants – Far away
- Same Quadrant – Similar Profiles
- Too close to center: Equally probably (low chi-square)
- We can take this MCA data and find the Euclidean Distance and cluster them
R - MICROSTRATEGY REST APIs
R - MICROSTRATEGY REST APIs

OVERVIEW
- PLUMBER
- AGILE FOR CHANGE AND DEPLOYMENT
- QUICKER IMPLEMENTATION

LIMITATIONS
- VALIDATION
- CUSTOM CODE
- MONITORING
- SECURITY
- COOKIES, TOKEN
- SCALABILITY - DOCKER
DEMO
MICROSTRATEGY DEMO
MICROSTRATEGY DEMO
MICROSTRATEGY AND R INTEGRATION

What does the Store Brand baby diaper market look like?

Which customers should I target?

How should I reach them?
START TO FINISH - BIG PICTURE

1460x8000 Data
Multiple Correspondence Analysis
350 Contrasting Important Variables
Clustering 1 Meaningful - Customer Segmentation
Rest API - Send Data
MicroStrategy Dashboard

R helped us do the math, REST APIs send the information, & MicroStrategy helped us present it
QUESTIONS?
We are looking for a seasoned MicroStrategy Architect to join our team!

Come see us after the presentation.
Vote for

Singing Data Ninjas

@ Datathon!
THANK YOU!

facebook.com/DomtarEveryday/

@DomtarEveryday

www.domtarpersonalcare.com