# Parameters driving consumer demand in Brazil

**Authors:** Krishna Rajendran
**Advisor:** Dr. Matthias Winkenbach

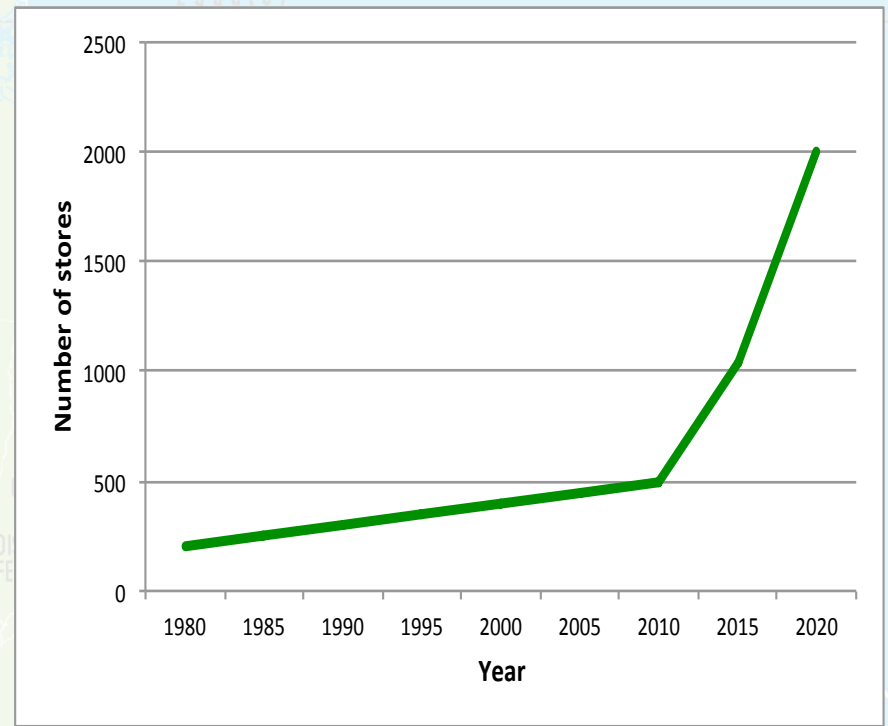**MIT SCM Research FEST**

May 19, 2016

**MIT** Supply Chain
M A N A G E M E N T

# Introduction



**Current distribution of Lojas Americanas stores in Brazil**



**Total number of stores year over year**

Source: Google Maps

# Introduction



Source: http://www.vagaemprego.com.br/2014/12/27/lojas-americanas-abre-inscricoes-para-programa-de-estagio-2015-em-brasilia/



Source: https://commons.wikimedia.org/wiki/File:Lojas_Americanas_-_Lapa.jpg

**Large stores with sales area close to 1500-2000 square meters**

**Smaller stores with sales area under 300 square meters**

**Research focus : Determine the ideal product assortment for stores with smaller shelf space.**

**MIT** Supply Chain
M A N A G E M E N T

# High-level approach
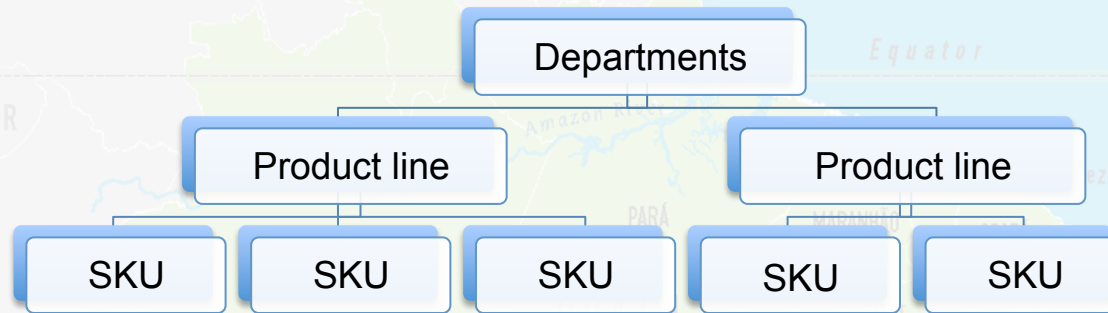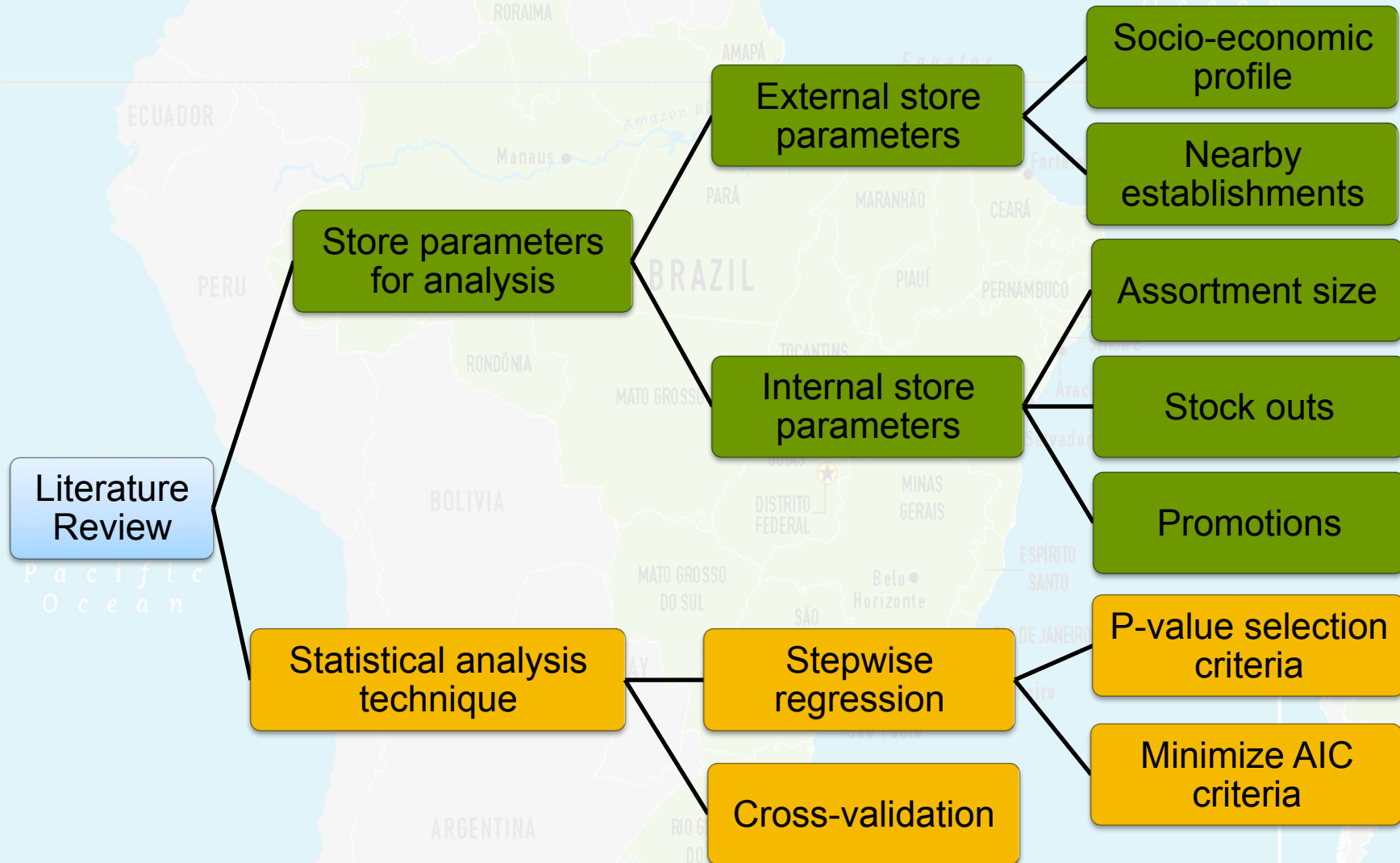


- Company has 36 departments, 1000 product lines, and 22,000+ SKUs.

- Departments have many product lines and each product line has multiple SKUs.

- Company policy – all departments and product lines must be in all stores. Entire departments and product lines cannot be removed.

**Focus of this thesis:**

**For each department find the store parameters that most influence sales.**

# Literature Review

```
Literature
Review
├── Store parameters for analysis
│   ├── External store parameters
│   │   ├── Socio-economic profile
│   │   └── Nearby establishments
│   └── Internal store parameters
│       ├── Assortment size
│       ├── Stock outs
│       └── Promotions
└── Statistical analysis technique
    ├── Stepwise regression
    │   ├── P-value selection criteria
    │   └── Minimize AIC criteria
    └── Cross-validation
```

MIT Supply Chain
MANAGEMENT
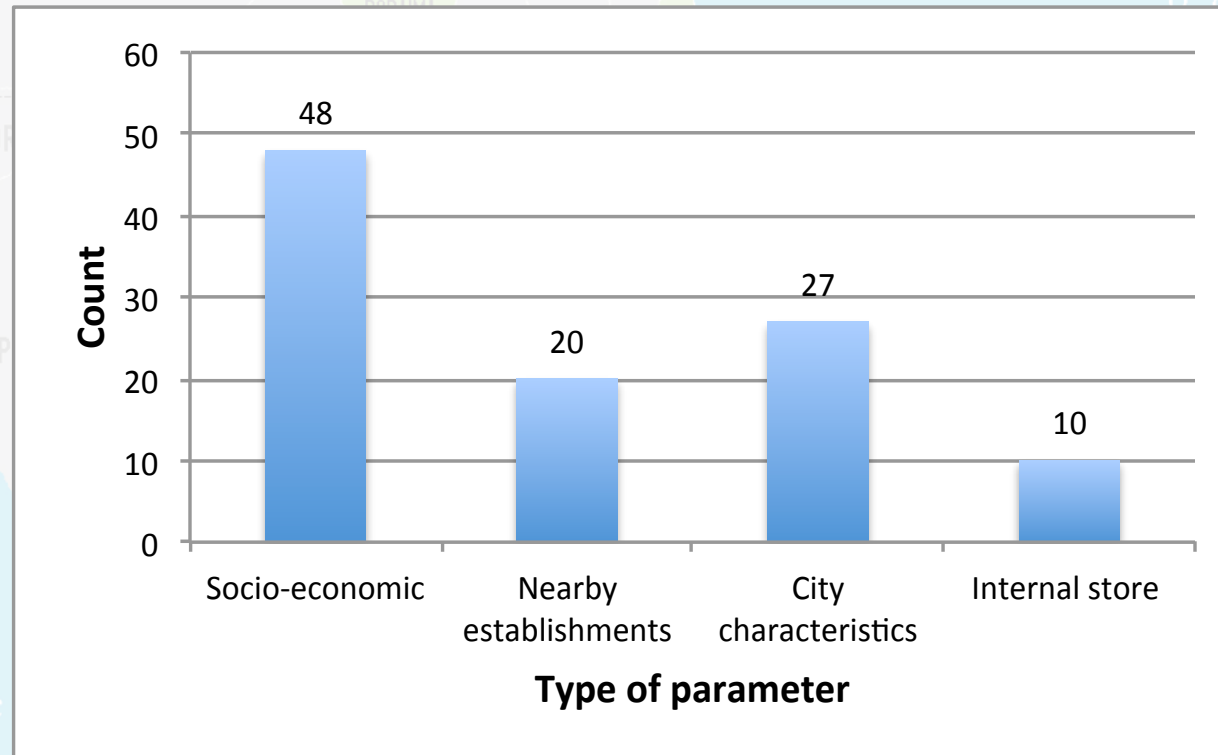
# Data Analyzed



- For each of the 36 departments, 105 parameters across 1,041 stores were analyzed. In total there were about 3.9 million data points.

- For each department, the focus was to identify the key parameters that had a statistically significant impact on that department's sales.

MIT Supply Chain
MANAGEMENT

# Store parameters

**Socio-economic**

- Workers (salaried people)
- Inactive residents (retired)
- Economically active people
- Permanent households
- Number of people in different age groups

**Establishments nearby**

- Schools
- Supermarkets
- Shopping malls
- Hotels
- Cosmetic/ Perfume shops

**Store Parameters**

**City parameters**

- Apartments/ Homes
- Population density
- Tourists
- University Segment

**Internal store parameters**

- Assortment size
- Sales area
- Inventory storage area
- Temperature

**MIT** Supply Chain
M A N A G E M E N T

# Model features

**Stepwise backward elimination regression**

- Analyze influence of parameters on department sales.
- All parameters are initially taken into the model.
- Parameters are eliminated in a stepwise sequence.
- Process is run based on selection criteria – minimum p-value threshold and minimize AIC.

**Minimize multicollinearity**

- Correlation matrices – variables with correlation coefficients magnitude greater than 0.7 eliminated.
- Variance Inflation Factor (VIF) – variables with VIF > 10 are eliminated.

**Cross-validation**

- Holdout cross validation method is used.
- Four different training to validation ratios used to find optimal output. Optimal output is model with highest R-squared adjusted for validation.

**MIT** Supply Chain
MANAGEMENT

# Methodology

```
┌─────────────────────────────────────────────────┐
│ Remove highly correlated predictor variables    │
│ and divide data into training and validation set │
└─────────────────────────────────────────────────┘
                        ↓
┌─────────────────────────────────────────────────┐
│           Start Stepwise Regression             │
└─────────────────────────────────────────────────┘
          ↓                           ↓
┌──────────────────────┐      ┌──────────────┐
│ P-value 0.05,0.1,0.15│      │     AIC      │
└──────────────────────┘      └──────────────┘
          ↓                           ↓
┌─────────────────────────────────────────────────┐
│       Select model with highest R-squared       │
│              validation set value                │
└─────────────────────────────────────────────────┘
                        ↓
┌─────────────────────────────────────────────────┐
│               Rerun the model                    │
└─────────────────────────────────────────────────┘
                        ↓
┌─────────────────────────────────────────────────┐
│ Remove predictor with highest VIF and rerun      │
│ model each time until highest VIF < 10           │
└─────────────────────────────────────────────────┘
                        ↓
┌─────────────────────────────┐    ┌──────────────────┐
│ Note R-square validation set│ →  │ Alter training to │
│ value for final model       │    │ validation ratio  │
└─────────────────────────────┘    └──────────────────┘
```

8

MIT Supply Chain
MANAGEMENT

# Regression Output Tables

| Summary of Fit | |
|---|---|
| R-squared | 0.7524 |
| R-squared adjusted | 0.7385 |
| Root Mean Square Error | 439688.2 |
| Number of observations | 473 |

| Cross validation | |
|---|---|
| Source | R-squared adjusted |
| Training Set | 0.7385 |
| Validation Set | 0.7224 |

- R-squared : Indicates explanatory power of the regression model

- R-squared adjusted: Similar to R-squared but imposes a penalty when new variables are added that do not improve explanatory power

- **R-squared adjusted validation value: This is the key performance metric. Explains how well the model works for new data.**

MIT Supply Chain
MANAGEMENT
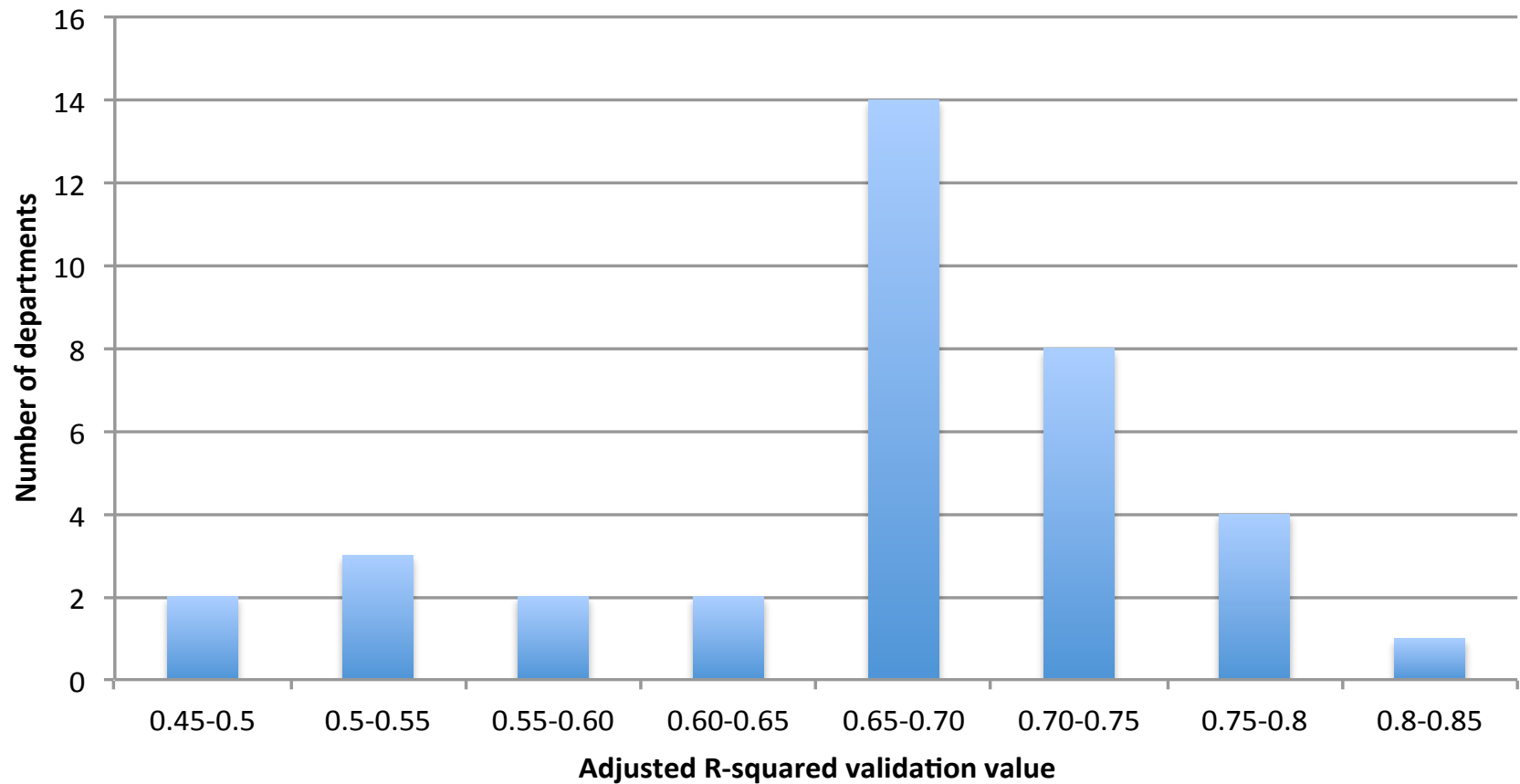
# Results



**Majority of R-squared values are above 0.65, indicating good results**

# Regression Output Tables

| Parameter | Parameter estimate | P-value |
|---|---|---|
| University Segment: Level A | 136,946.16 | <. 0001 |
| Hospitals (1km) | -23,494.65 | <. 0001 |
| Public schools (1km) | 23,143.16 | <. 0001 |
| Sales Area (in meter square) | 879.9767 | <. 0001 |
| Days since Inauguration of store | 43.7731 | <. 0001 |
| Store Type: Conventional | -289,896.95 | <. 0001 |
| Geographic Region - NOT South East | -293,664.49 | <. 0001 |
| PIB per Capita Total (2012) | -3.9253 | 0.0153 |
| Store Status: NOT Premium | -63,477.44 | 0.0115 |
| Geographic Region - NOT Central | -129,384.90 | 0.0121 |
| Shopping Center (1km) | 53,401.39 | 0.0129 |
| Stock area (in meter square) | 208.7668 | 0.0215 |
| Store Location - Street | -61,159.85 | 0.0324 |
| Private schools (1km) | 4,145.23 | 0.0532 |
| Tourist Segment: Level C | -77,980.03 | 0.0541 |
| Days since last refurbishment | -13.8445 | 0.0689 |
| PEA Density (2007) | 18.9676 | 0.0754 |
| Perfume and Cosmetic shops (1km) | -17,589.19 | 0.0767 |
| Intercept | 573,766.37 | 0.0812 |
| Average summer temperature (C°) | -11,664.53 | 0.0842 |
| Geographic Region - NOT North East | -54,772.23 | 0.0963 |
| University Segment: Level C | -74,425.14 | 0.0985 |

- Table shows the set of statistically significant parameters from the initial list of 105.

- Cells highlighted in green indicate parameters identified as noteworthy based on site visit and discussions with company's executives.

- For each department, a separate regression output table was obtained.

- Parameter estimates indicate the amount sales will increase for a unit increase in that parameter.

MIT Supply Chain
MANAGEMENT

# Results

| Department | Most influential and noteworthy parameters |
|---|---|
| Audio & Video Equipment | Education level, supermarkets, retired individuals |
| Baby products | Age group 25-34, level of education, store status (premium/non-premium) |
| Bed and bath | Temperature, pharmacies |
| Beverages | Hotels, tourists, temperature |
| Biscuits and Cookies | Supermarkets, income level |
| Books | Schools, income level |
| Briefs and underwear | Sales area, inventory storage area, income level |
| Candies | Hotels, population density, residential complexes |
| Children's clothing | Age group 0-9, income level, age of store, schools |
| Chips and Cereal | Gyms, residential complexes |
| Chocolates | Schools, education level, shopping centers |
| Christmas | Age group 50-59, income level |
| Cleaning products | Super markets, temperature, income level |
| Cosmetics | Perfume shops, temperature, inventory storage area |
| Diapers | Inventory storage area, perfume shops, age of store |
| Easter | Tourists, shopping centers, income level, temperature |
| Electronics/IT Equipment | Shopping centers, inventory storage area, income level |
| Footwear | Temperature, income level, supermarkets, age group 15-19 |
| Hair coloring products | Inventory storage area, income level, perfume/cosmetic shops |
| Home appliances | Age group 35-49, Age group 50-59, shopping centers, income level |
| Imported Toys | Age group 10-14, Age group 50-59, income level |
| Kitchen utensils | Age group 25-34, supermarkets |
| Lingerie | Perfume/cosmetic shops, income, super and hyper markets |
| Men's clothing | Sales area, shopping centers, age group 50-59 |
| Mobile phones | Store status (premium/non-premium), income level |
| Movies | Age group 50-59, store status (premium/non-premium), income level |
| Music | Assortment size, education level, income level |
| Perfumes and Hygiene | Supermarkets, inventory storage area |
| Plastic utilities | Supermarkets, income level, temperature |
| Socks and Scarves | Sales Area, assortment size, temperature, income level, age of store |
| Sports & Beach Equipment | Gyms, schools, income level, temperature |
| Stationary items | Age group 15-19, income level |
| Table and bar | Income level, shopping centers |
| Toys | Sales area, assortment size, income level |
| Videogames | Income level, age group 25-34 |
| Women's clothing | Sales area, age of store, income level, shopping centers |

MIT Supply Chain
MANAGEMENT

# Results

| Department | High income | Low income | Super market | Sales Area | Shopping centers | Temp. | Inventory Area | Age of store | Education level | School | 50-59 age group |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Chocolate | | | ✓ | | ✓ | ✓(red) | | ✓ | ✓ | ✓ | |
| Children clothes | ✓ | | ✓(red) | | | | | ✓(red) | | ✓ | |
| Lingerie | ✓ | | | | ✓(red) | | | | | | |
| Socks & Scarves | | ✓(red) | ✓ | ✓ | | ✓(red) | | ✓ | | | |
| Footwear | ✓ | | ✓(red) | | | ✓ | | | | | |
| Innerwear | ✓(red) | ✓ | | ✓ | | | ✓ | | | | |
| Baby Products | | | | | | | | | ✓ | | |
| Women's clothing | | ✓ | | ✓ | ✓(red) | | | ✓ | | | |
| Bed and Bath | | | | | | ✓ | | | | | |
| Men's clothes | ✓ | | | ✓ | ✓ | | | | | | ✓ |
| Stationary | ✓ | | | | | | | | | | |
| Books | ✓ | ✓ | | ✓ | | | | | | ✓ | |
| Cleaning | | ✓ | ✓ | | | ✓ | | | | | |
| Kitchen utensils | | | ✓ | | | | | | | | |
| Table and Bar | ✓ | | | | ✓ | | | | | | |
| Toys | ✓ | | | ✓ | | | | | | | |
| Music | ✓ | | | | | | | | ✓ | ✓ | |
| Audio & Video | ✓ | | ✓ | | | | | | ✓ | | |
| Home appliance | ✓ | ✓ | ✓(red) | | ✓ | | ✓ | | | | ✓ |
| Perfumes | ✓ | ✓ | ✓(red) | ✓ | ✓ | | ✓ | ✓(red) | | | |
| Electronics& IT | | ✓ | | ✓ | ✓ | | ✓ | | | | |
| Hair Coloring | | ✓ | ✓(red) | ✓ | | | ✓ | | | | |
| Cosmetics | | | ✓(red) | ✓ | | | ✓ | ✓ | | | |
| Sports & Beach | ✓ | | | | | ✓ | | | | ✓ | |
| Imported Toys | ✓ | | | | | | | | | | ✓ |
| Candies | ✓ | | | | | | | ✓(red) | | | |
| Easter | | ✓ | | | ✓ | ✓(red) | | | | | |
| Plastic utilities | | ✓ | ✓(red) | | | | | | | | |
| Video games | | ✓(red) | | ✓ | | | | | | | |
| Chips& Cereal | ✓ | ✓ | | | | | | | | | |
| Beverages | | | | | | ✓(red) | | | | | |
| Biscuits & Cookies | ✓ | ✓(red) | ✓(red) | | | | | | | | |
| Diapers | | | | | ✓(red) | | ✓ | ✓(red) | ✓ | | |
| Christmas | ✓ | | | | | | | | | | ✓ |
| Telephones | | ✓ | | | | | | | | | |
| Movies | ✓ | | | | | | | | | | ✓ |

MIT Supply Chain MANAGEMENT

# Conclusion

**Model Extension**
- The methodology developed can be used to analyze data at the product line and SKU level to devise optimal product assortment policy.

**Key findings**
- The insights developed here could be used by retail companies (in general) who could benefit by focusing their research on analyzing the eleven key factors listed in the previous slide.

**Future Scope**
- The model could be improved by factoring in promotions.
- Some findings did not have a practical explanation. Consumer surveys and additional statistical analysis tools may help better understand the reasons for these findings.

**MIT** Supply Chain
M A N A G E M E N T

# Questions?

**MIT** Supply Chain
M A N A G E M E N T