FORDHAM | Gabelli School of Business

THE JESUIT UNIVERSITY OF NEW YORK

The Alignment of Product Variables to Ecommerce Niches

Data Mining for Business

ISGB/BYGB 7967

Christine Bukovac, Tahsin Rakib Himi, Anh Nguyen, and Olha Sverdel

Date: 12 / 07 / 2020

Fall 2020

Abstract

Problem Statement

- How can small scale online retailers utilize Amazon data of product prices, numeric ratings and written reviews to determine what combination of attributes best aligns with their business model?

Methods

- Data Mining Techniques
  - Cluster Analysis
  - Association Analysis
  - Sentiment Analysis

Data: Amazon

- Data on approximately 9,000 products
- Key Variables include the product category, manufacturer, price, average review rating, and written customer review

Results

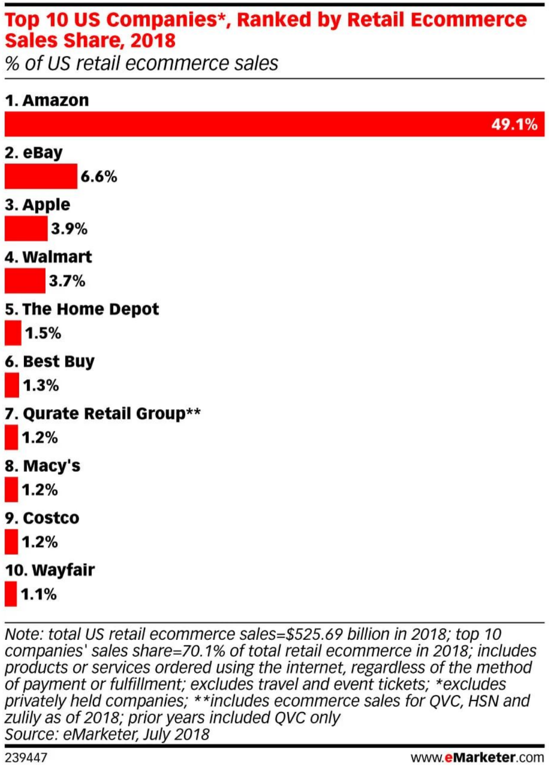
- Retailers should focus on manufacturer relationships, which is linked to customer's satisfaction, rather than product categories
- Retailers can benefit most from selling "Playmobil" products (the highest sentiment rating and clustering group)

Introduction

Online Retail Market

Harder to compete for small-shop online retailers:

- Amazon is ranked #1 for online sales
- Amazon created high barriers of entry to the online retail market for small players
- COVID accelerated shift of Brick-and-Mortar stores online



Introduction

Use of Machine Learning for effective product offering online

Use of Amazon Reviews:

- To determine which type of manufacturers and category of product will bring the most profit
- To which price range of products from each manufacturer to offer in their online store.

Clustering

Sentiment Analysis

Association Rules

Analysing what customers are looking for in each manufacturer can help small online retailers to create an efficient strategy of product selection and pricing.

Data Description

- The dataset used was initially parsed by PromptCloud's in-house web-crawling service and uploaded to data.world
- The dataset originally contained around 10,000 different products
- Key variables include product category, manufacturer (ie Disney, Hasbro, Playmobil), price, average review rating and written review
- We focused our analysis on the products of the top 16 manufacturers (by frequency), which totals to approximately 1,500 products

Dataset sample:

The following represents the first entry of our data set:

ID	Product Name	Manufacturer	Price	Avg. Rating	Categories	Customer Also Bought	Description	Product Info	Product Description	Items Customers Buy After Viewing This Items	Customer Reviews
eac7efa5dbd3d667f26eb3d3ab504464	Hornby 2014 Catalogue	Hornby	3.42	4.9	Hobbies	http://www.amazon.co.uk/Hornby-R8150-Catalogue-2015/dp/B00S9SUUBE | http://www.amazon.co.uk/Hornby-Book-Model-Railways-Edition/dp/1844860957 | http://www.amazon.co.uk/Hornby-Book-Scenic-Railway-Modelling/dp/1844861120 | http://www.amazon.co.uk/Peco-60-Plans-Book/dp/B002QVL16I | http://www.amazon.co.uk/Hornby-Gloucester | http://www.amazon.co.uk/Airfix-5014429781902	Product Description Hornby 2014 Catalogue Box Contains 1 x one catalogue	Technical Details Item Weight640 g Product Dimensions 29.6 x 20.8 x 1 cm Manufacturer recommended age:6 years and up Item model number R8148 (Note: More info under the data excel file)	Product Description Hornby 2014 Catalogue Box Contains 1 x one catalogue	http://www.amazon.co.uk/Hornby-R8150-Catalogue-2015/dp/B00S9SUUBE | http://www.amazon.co.uk/Hornby-Book-Model-Railways-Edition/dp/1844860957 | http://www.amazon.co.uk/Peco-60-Plans-Book/dp/B002QVL16I | http://www.amazon.co.uk/Newcomers-Guide-Model-Railways-Step/dp/1857943295	Worth Buying For The Pictures Alone (As Ever) // 4.0 // 6 April 2014 // By Copnovelist on 6 April 2014 // Part of the magic for me growing up as a boy was to buy (or be given) the new Hornby catalogue every year, even if it included 90% of the same products as the previous year. I've still got my old ones dating back to the 70s and 80s somewhere. These days the catalogue is especially informative in that it tells you the vintage of the rolling stock which is useful if you are dedicating your railway to one particular era and train company. | Amazing detail fabulous photography. // 5.0 // 11 April 2015 // By richard (Note: More reviews under the data excel file)

Problem Statement

The Big Question:

- How can small scale online retailers utilize Amazon data of product prices, numeric ratings and written reviews to determine what combination of attributes best aligns with their business model?

Research Steps:

1. Perform Cluster Analysis, Association Analysis, and Sentiment Analysis on top 16 manufacturers
2. Analyze the correlation of the product attributes (price, review content, review ratings) with categories and manufacturers
3. Inform retailers on which manufacturers to partner or product categories to pursue

Methodology: System Design

Problem Identification	Data Processing	Analysis Model	Compare and Contrast
Identify Business Problem	Clean data	Clustering Model	Association Analysis vs. Clustering Model
Perform Industry Analysis	Manually Label Categories	Association Analysis	Sentiment Analysis vs. Association Analysis
Design Problem Statement	Select Top 16 Most Frequent Manufacturers	Sentiment Analysis	Sentiment Analysis vs. Clustering Model
Select Variables			

python

Results + Conclusion

Methodology: Data Collection

Analysis Model	Variables
Cluster Analysis	Unique ID, Product Name, Manufacturer, Price, Number Available in Stock, Number of Reviews, Average Review Rating, General Categories, Categories
Association Analysis	Unique ID, Product Name, Manufacturer, Price, Number Available in Stock, Number of Reviews, Number of Answered Questions, Average Review Rating, General Categories, Categories, 16 Manufacturers breaking down in binary data.
Sentiment Analysis	Reviews, Rating, Manufactures, General Categories

Methodology: Data Preprocessing

Data Preprocessing

- We performed initial cleaning in Excel, including extracting a substring of the written reviews, removing redundant characters from HTML links and creating a broader General Categories for our products to be grouped into
- We utilized Python & Pandas to narrow our data down to the top 16 manufacturers

	Manufacturer's Name
1	'Amscan',
2	'Corgi',
3	'Disney',
4	'Every-occasion-party-supplies',
5	'Hasbro',
6	'Hornby',
7	'LEGO',
8	'Mattel',
9	'MyTinyWorld',
10	'Oxford Diecast',
11	'Playmobil',
12	'PokÃ©mon',
13	'Scalextric',
14	'Schleich',
15	'Star Wars',
16	'The Puppet Company'

Row Labels

Equipment
- Bedding & Linens
- Camping & Hiking
- Car Parts
- Characters & Brands
- Cooking & Dining
- Gardening
- Home Accessories
- Indoor Lighting
- Lab & Scientific Products
- Laundry, Storage & Organisation
- Medical Supplies & Equipment
- Medication & Remedies
- Novelty & Special Use
- Office Supplies
- panelKey"":""CloudDrivePanel""}
- Party Supplies
- Pens, Pencils & Writing Supplies
- Storage, Cleaning & Ring Sizers

Fashion
- Bags
- Dogs
- Fancy Dress
- Handbags & Shoulder Bags
- Men
- Novelty Jewellery
- Sex & Sensuality
- Supporters' Gear
- Women

Food
- Jams, Honey & Spreads
- Sweets, Chocolate & Gum

Toy
- Arts & Crafts
- Baby & Toddler Toys
- Die-Cast & Toy Vehicles
- Dolls & Accessories
- Educational Toys
- Electronic Toys
- Figures & Playsets
- Games
- Hobbies
- Jigsaws & Puzzles
- Musical Toy Instruments
- Pretend Play
- Puppets & Puppet Theatres
- Sports Toys & Outdoor
- Worlds Apart
- (blank)

Methodology: Variable Selection

Key Variables

- The dataset began with 17 attributes, however we identified 7 to be key in our analysis:
  - manufacturer
  - general category (which we generated)
  - category
  - price
  - average review rating
  - customer reviews
  - customer reviews substring (which we also generated)

product_name	manufacturer	price	number_available_	number_of_reviews	number_of_answered_questions	average_review	general category	categories	subcategories	customer_reviews_substring
Hornby 2014 Cat	Hornby	3.42	5	15	1	4.9	Toy	Hobbies	Model Trains & Railway Sets	Worth Buying For The Pictures Alone (As Ever)
HORNBY Coach F	Hornby	39.99		1	2	5	Toy	Hobbies	Model Trains & Railway Sets	I love it
Hornby 00 Gauge	Hornby	32.19		3	2	4.7	Toy	Hobbies	Model Trains & Railway Sets	Birthday present
Hornby 00 Gauge	Hornby	24.99		2	1	4.5	Toy	Hobbies	Model Trains & Railway Sets	High standard model, well worth the wait. Repl
Hornby Santa's E	Hornby	69.93	3	36	7	4.3	Toy	Hobbies	Model Trains & Railway Sets	Beautiful set
Hornby Gauge W	Hornby	235.58	4	1	1	5	Toy	Hobbies	Model Trains & Railway Sets	Five Stars
Hornby Gauge Ra	Hornby	27.49	6	1	1	5	Toy	Hobbies	Model Trains & Railway Sets	steaming good engine!
Hornby 00 Gauge	Hornby	119.5	2	3	1	5	Toy	Hobbies	Model Trains & Railway Sets	Gods Wonderful Railway.
Hornby R2981 Lc	Hornby		2	4	1	4.3	Toy	Hobbies	Model Trains & Railway Sets	Olympic Gold

Methodology: Model Building

Cluster Analysis

- Group data based on commonalities then analyze each cluster to discover patterns specific to that group
- K-Means Algorithm

Association Rule Mining

- K-Means results were used as a way to preprocess the data (generated clusters) to then perform Association Rule Mining
- Find interesting relationships between attributes to create rules

Sentiment Analysis

- Utilized VADER (Valence Aware Dictionary and sentiment Reasoner) on the text reviews
- Focused on lexicons of sentiment related words, each lexicon being rating either positive, negative neutral or compound

Methodology: Evaluation

Cluster Analysis

- Relied on Silhouette, or the measure of consistency within clusters
- Manually grouped categories data resulted in 0.4 silhouette
- Model based on Amazon Categories generated clusters resulted in 0.2 silhouette

Association Rule Mining

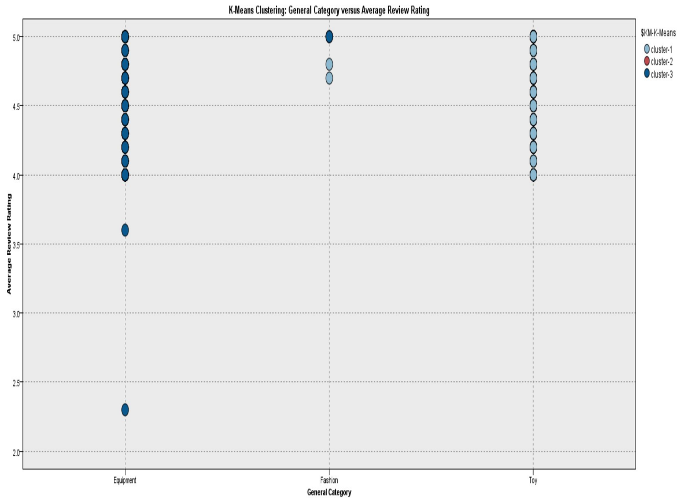
- 17 rules were generated
- Ranges of Key Metrics
  - Support: 1.94% - 66.45%
  - Confidence: 2.38% - 84.30%
  - Lift: 1.01 - 1.39

Sentiment Analysis

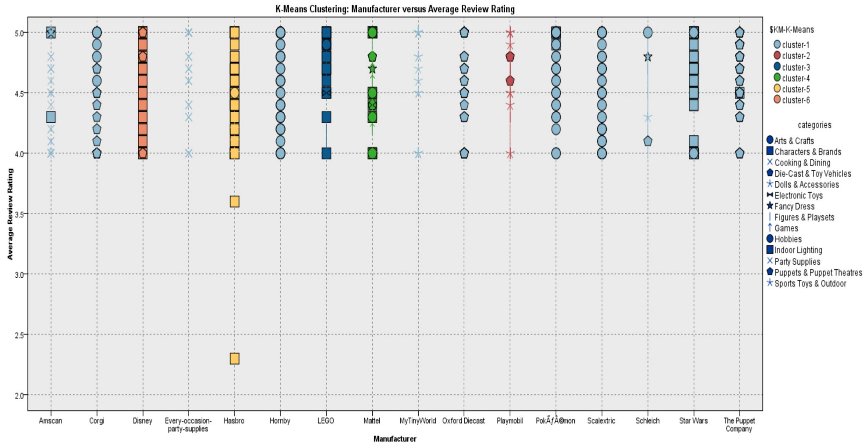
- The compound value is normalized between -1 (most extreme negative) and +1 (most extreme positive) to receive a unidimensional measure sentiment for a review

Clustering Result

- Average Review Rating by General Category is consistent, generally falling between 4.0 to 5.0



- Generally, all Manufacturers earned a review rating between 4.0 and 5.0

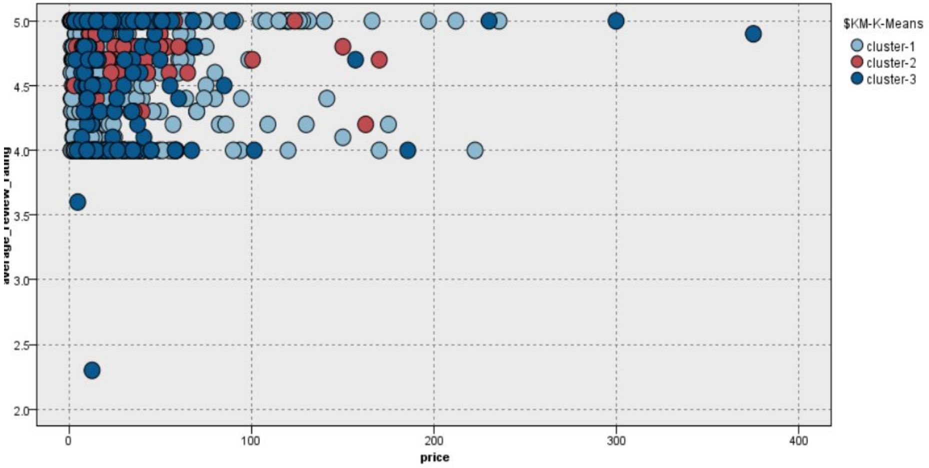


Takeaway: K-means was most useful as a way to preprocess the data, to prepare for Association Rule Mining

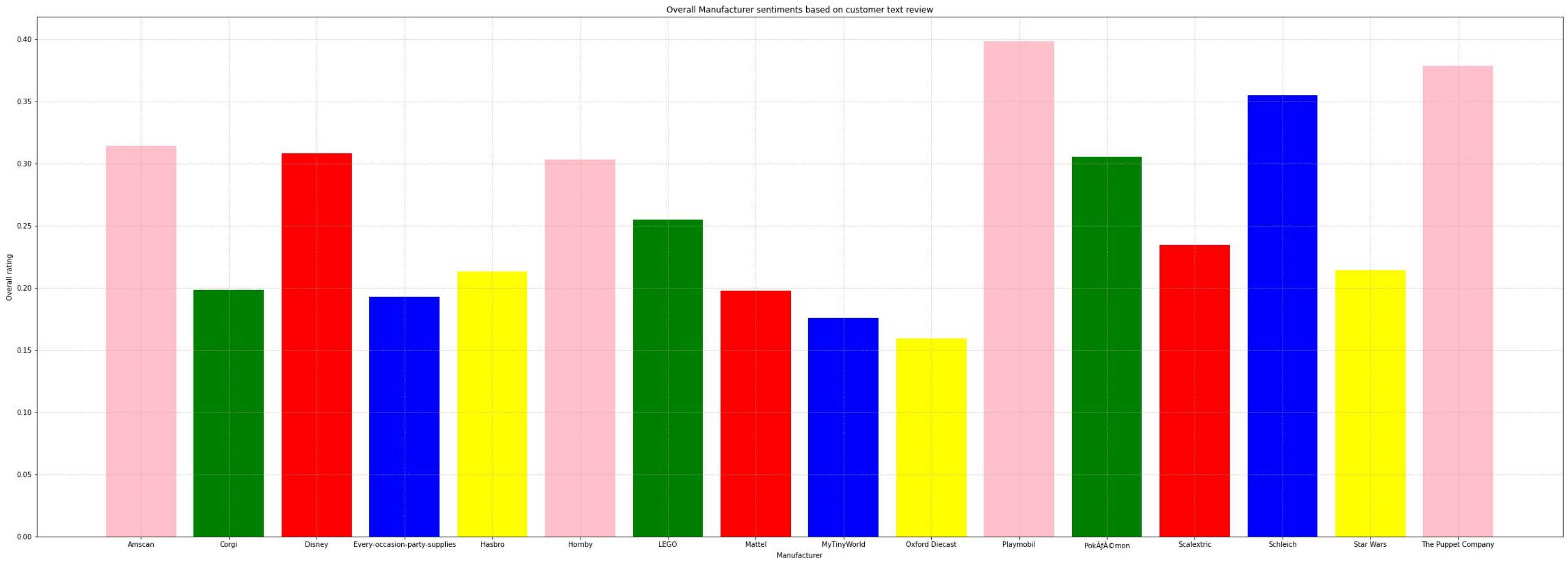
Association Result

Rule ID	Condition	Prediction	Sorted By Confidence(%)	Condition Support (%)	Rule Support (%)	Lift	Deployability (%)
				Other Evaluation Statistics			
1	price ≤ 75.552 average_review_rating > 4.460	$KM-K-Means = cluster-1	84.30	63.80	53.78	1.03	10.02
2	price ≤ 75.552 $KM-K-Means = cluster-2	average_review_rating > 4.460	83.95	5.24	4.40	1.07	0.84
13	price ≤ 75.552 $KM-K-Means = cluster-3	3.920 ≤ average_review_rating < 4.460	29.37	8.14	2.39	1.39	5.75
14	$KM-K-Means = cluster-3	3.920 ≤ average_review_rating < 4.460	28.25	11.44	3.23	1.34	8.21
15	3.920 ≤ average_review_rating < 4.460	$KM-K-Means = cluster-3	15.34	21.07	3.23	1.34	17.84
16	price ≤ 75.552 3.920 ≤ average_review_rating < 4.460	$KM-K-Means = cluster-3	15.16	15.77	2.39	1.33	13.38

- Cluster 1 and 2: The lower the prices the higher the reviews
- Cluster 3: The lower the prices the lower the reviews



Sentiment Result



- The figure indicates that for most manufacturer’s reviews based on lexicon rating were overall positive
- Playmobil has the highest mean compound value based on reviews

Models Comparison

Comparison Metrics		Clustering Analysis	Association Analysis	Sentiment Analysis
Average Review Rating	3 General Categories	Between 4 and 5 stars (with exception of 2 outliers)		Positive Rating of approx. 0.25 (out of 1)
	16 Manufacturers	Between 4 and 5 stars (with exception of 2 outliers)	Rating < 4.46 is considered to be "low"	Varies from 0.15 to 0.4 Highest Rating Manufacturer: PlayMobil Second Highest Rating Manufacturer: The Puppet Company
Manufacturers	Ranking	Most frequent manufacturer: 1. Cluster 1: Disney 2. Cluster 2: Playmobil 3. Cluster 3: Hasbro	Most frequent manufacturer: 1. Cluster 1: Disney 2. Cluster 2: Playmobil 3. Cluster 3: Hasbro	Highest Review Rating Manufacturer: 1. PlayMobil 2. The Puppet Company
Price + Rating Hypothesis	Cluster group and Categories Prediction: • Cluster 1: mostly Toy + some Fashion • Cluster 2: Some Equipment • Cluster 3: Mostly Equipment + some Fashion	• Cluster 1: Price < 174 (except some outliers) • Cluster 3: Price < 100 (except some outliers) • Cluster 1 and 2: 4 < Rating < 5 (except some outliers)	Prices < 75.55 is considered to be "low" • Price < 75 *and* Rating > 4.46 → Cluster 1 or Cluster 2 • 3.92 < Rating < 4.46 *and* Price < 75 → Cluster 3	Rating for all categories > 0.25 (out of 1)

Conclusion

Overall Observation

- Cluster 1(mostly Toy)(mostly Disney): The lower the prices, the higher the reviews for manufacturers
- Cluster 2 (some Equipment)(mostly Hasbro): The lower the prices, the higher the reviews -> in general, this group has higher prices and thus, lower review ratings
- Cluster 3 (mostly Equipment)(mostly Playmobil): Better reviews for most expensive products

Recommendation

- Retailers should focus on manufacturer relationships, which is linked to customer's satisfaction
- Retailers can benefit most from selling only more expensive products from "Playmobil" (the highest sentiment rating and clustering group), and only less expensive products from manufacturers that belong to Cluster 1 and 2 such as Disney and Hasbro

Thank You

Appendix

Model Methodology

Full Description of Source Node

- Data Source: https://data.world/promptcloud/fashion-products-on-amazon-com

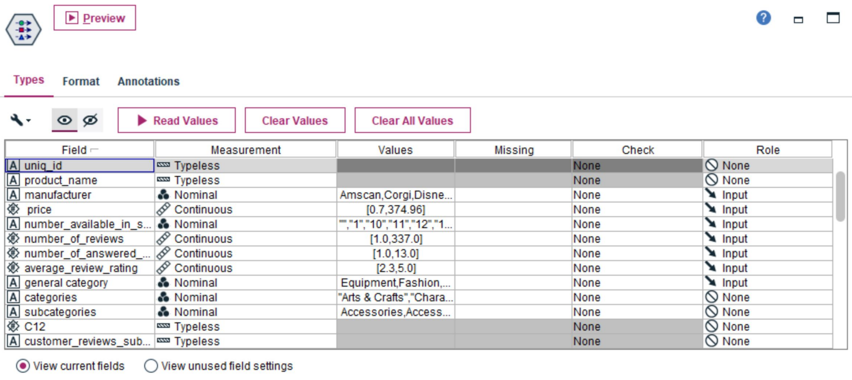
Variables for Analysis	Description
Unique_ID	ID for each product
Product_Name	Name for each product
Manufacturer	Manufacturer who creates the product
Price	Price listed for each product
Number_Available_in_ Stock	Number of Product Available on Shelve
Number_of_Reviews	Number of Reviews scraped for each Products
Number_of_Answered _Questions	The number of Questions Answered for each product
Average_Review_Rating	Average Rating out of 5 stars for each product
General Categories	Manual Labelled General Categories for each product based on Amazon Categories.
Categories	Amazon define the product's categories
customer_reviews	A list of reviews for each product in each cell

Variables EXCLUDED from Analysis	Description
subcategory	Amazon defined subcategories after main categories for each product We decided not use this variable because we have general categories.
customers_who_bought_this_item_also_bought description	List of links to the of products that the same customer purchased after buying the product listed We decided not to use this variable because this variable does not support our analysis
product_information	The information of the product (weights, length, Materials, etc.) We decided not using hi
product_description	The description of the products based on the
items_customers_buy_after_viewing_this_item	Links to the items that were bought after viewing given product
customer_questions_and_answers	List of questions that customers had before buying the product
sellers	The list of sellers for the product

Full Description of Clustering Model Nodes

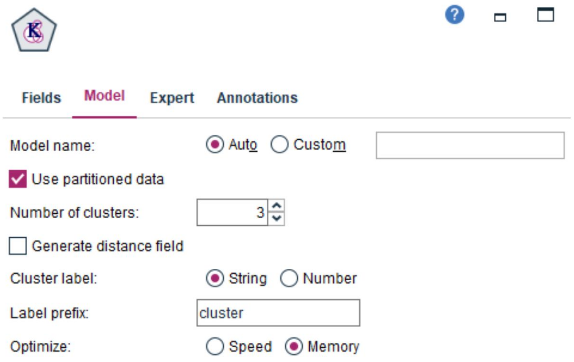
K-Means Cluster by 3 General Categories

Type Node - used to define each variable's role. This model was derived from the "general category" therefore the other category fields are marked none.



Field	Measurement	Values	Missing	Check	Role
uniq_id	Typeless			None	None
product_name	Typeless			None	None
manufacturer	Nominal	Amscan,Corgi,Disne...		None	Input
price	Continuous	[0.7,374.96]		None	Input
number_available_in_s...	Nominal	"","1","10","11","12","1...		None	Input
number_of_reviews	Continuous	[1.0,337.0]		None	Input
number_of_answered_...	Continuous	[1.0,13.0]		None	Input
average_review_rating	Continuous	[2.3,5.0]		None	Input
general category	Nominal	Equipment,Fashion,...		None	Input
categories	Nominal	"Arts & Crafts","Chara...		None	None
subcategories	Nominal	Accessories,Access...		None	None
C12	Typeless			None	None
customer_reviews_sub...	Typeless			None	None

K-Means Node - used to select the desired cluster size. We chose 3 since there are 3 general categories



Full Description of Clustering Model Nodes

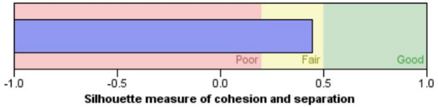
K-Means Cluster by 3 General Categories (continued)

K-Means Nugget - shows the results of our model, including cluster sizes, silhouette and predictor importance

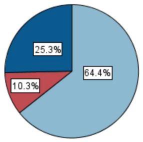
Model Summary

Algorithm	K-Means
Inputs	6
Clusters	3

Cluster Quality

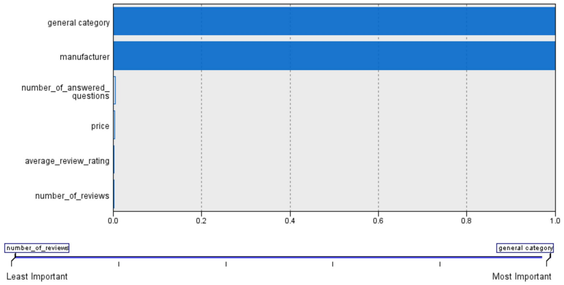


Cluster Sizes

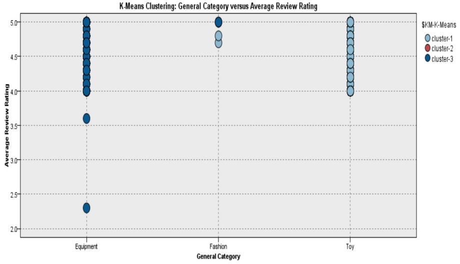


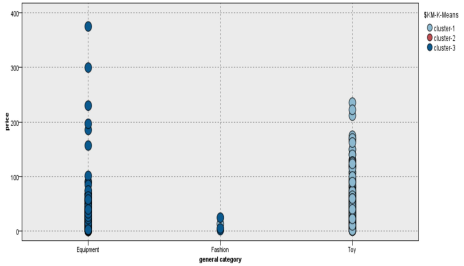
Size of Smallest Cluster	159 (10.3%)
Size of Largest Cluster	996 (64.4%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	6.26

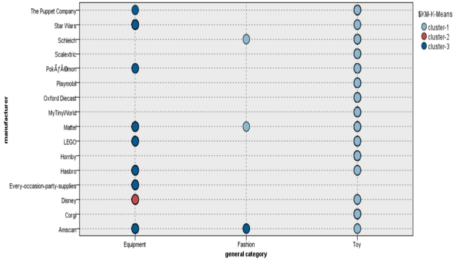
Predictor Importance



Plot Nodes - used to graph our results



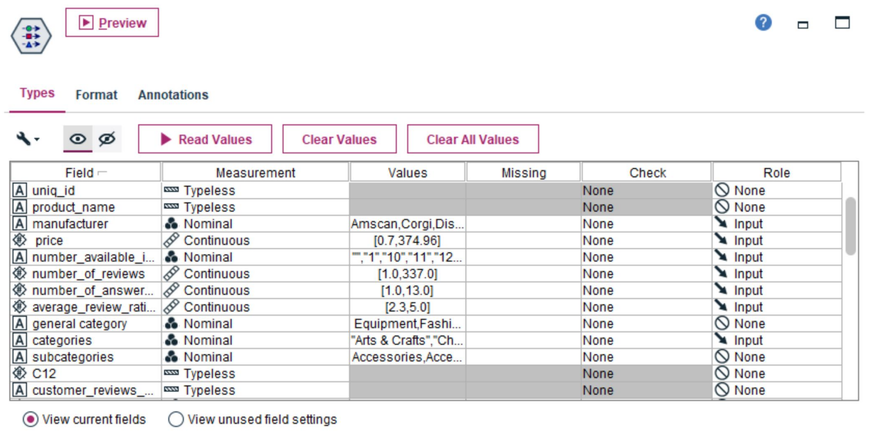




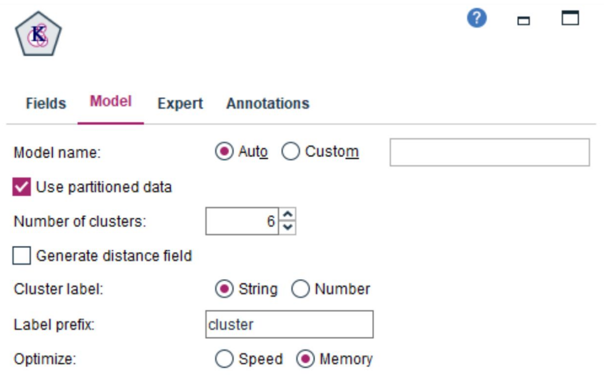
Full Description of Clustering Model Nodes

K-Means Cluster by Amazon Categories

Type Node - used to define each variable’s role. This model was derived from the categories of the original dataset, therefore “general category” which was our manually grouping was given the role none



K-Means Node - used to select the desired cluster size. We chose 6 since 6 clusters gave the highest silhouette



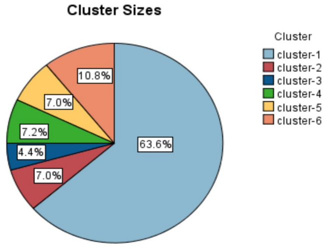
Full Description of Clustering Model Nodes

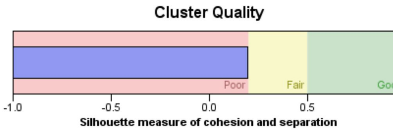
K-Means Cluster by Amazon Categories (continued)

K-Means Nugget - shows the results of our model, including cluster sizes, silhouette and predictor importance

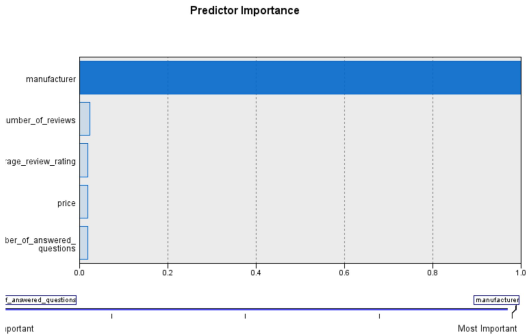
Model Summary

Algorithm	K-Means
Inputs	6
Clusters	6

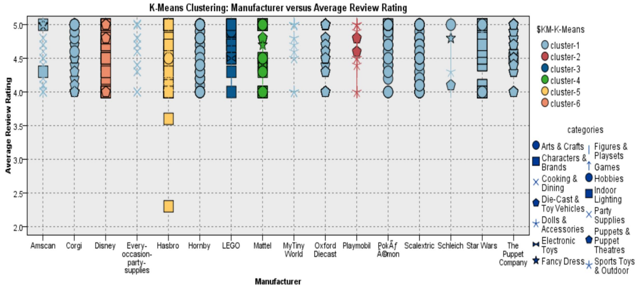


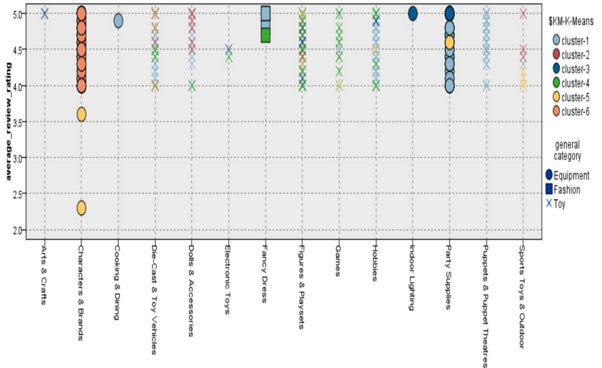


Size of Smallest Cluster	68 (4.4%)
Size of Largest Cluster	984 (63.6%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	14.47



Plot Nodes - used to graph our results

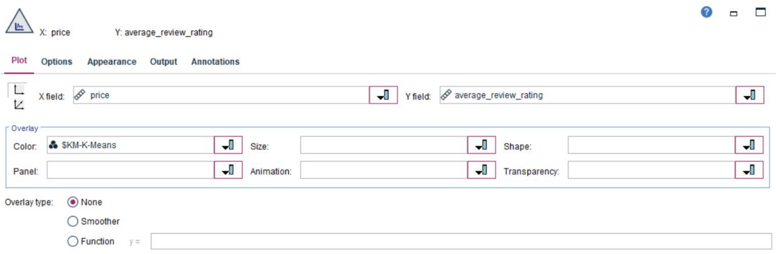




Full Description of the Association Rules Node

Clustering manufacturers into three groups

Used 1% as minimum Support and Confidence to see all the possible results



Price	**Continuous**	**Target and Input**
Average_Review_Rating	**Continuous**	**Target and Input**
$KN-K-Means (Cluster)	**Nominal**	**Target and Input**

Rule Statistics[a,b]

Measurements	Minimum	Maximum	Mean	Standard Deviation
Condition Support (%)	2.33	81.58	48.62	34.08
Confidence (%)	2.38	84.30	63.23	30.56
Rule Support (%)	1.94	66.45	33.89	30.31
Lift	1.01	1.39	1.10	0.14
Deployability (%)	0.39	79.64	14.73	17.64

a. Number of Rules is 17

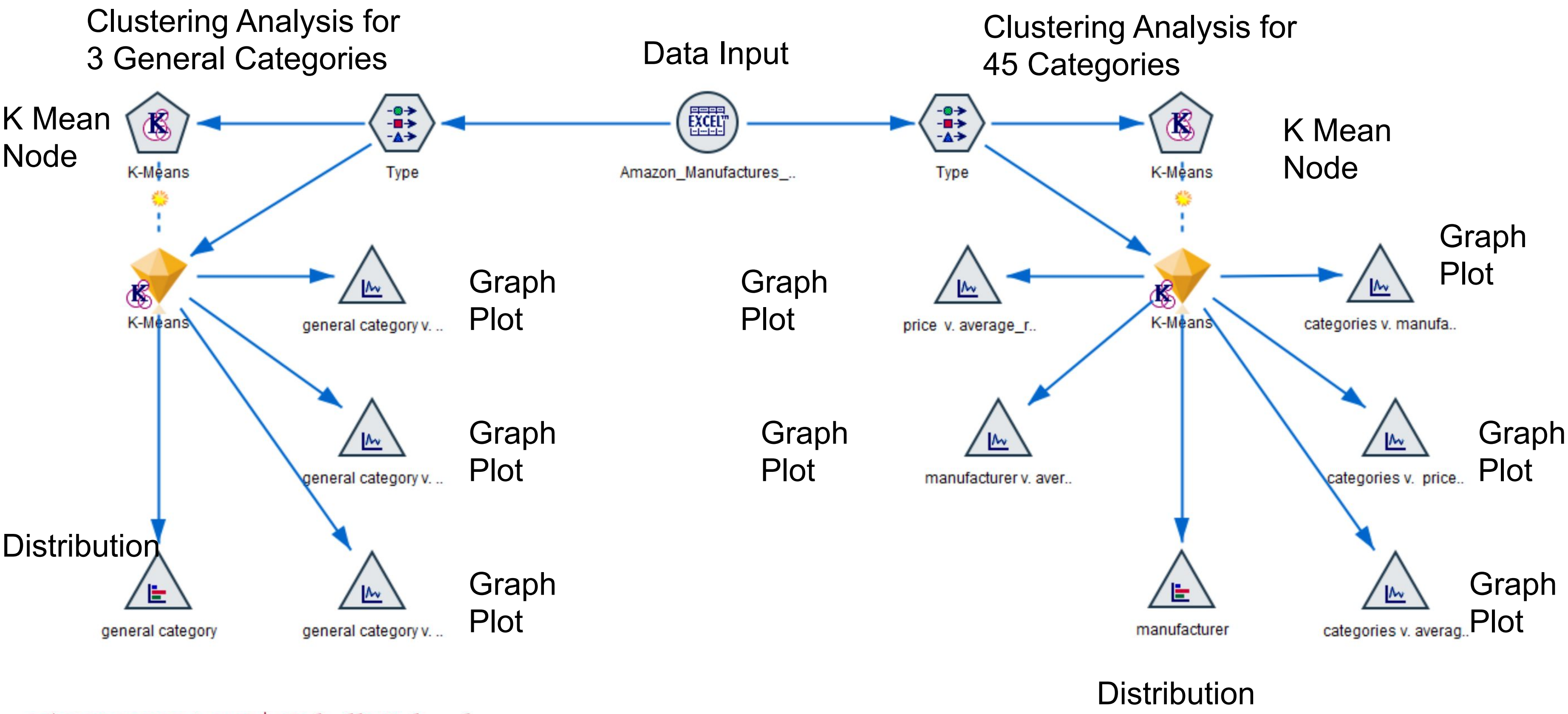
b. Number of Valid Events Data Source Records is 1,547

Build Settings[a]

Maximum Number of Rules	1,000
Minimum Condition Support	0.01
Minimum Confidence	0.01
Minimum Rule Support	0.01
Minimum Lift	1.00
Maximum Number of Items in a Rule	10
Maximum Number of Items in a Condition	5
Maximum number of Items in a Prediction	1
Use only True Value for Flag Fields	True
Allow Rules without Conditions	False
Evaluation Measure Sorting the Rules	Confidence

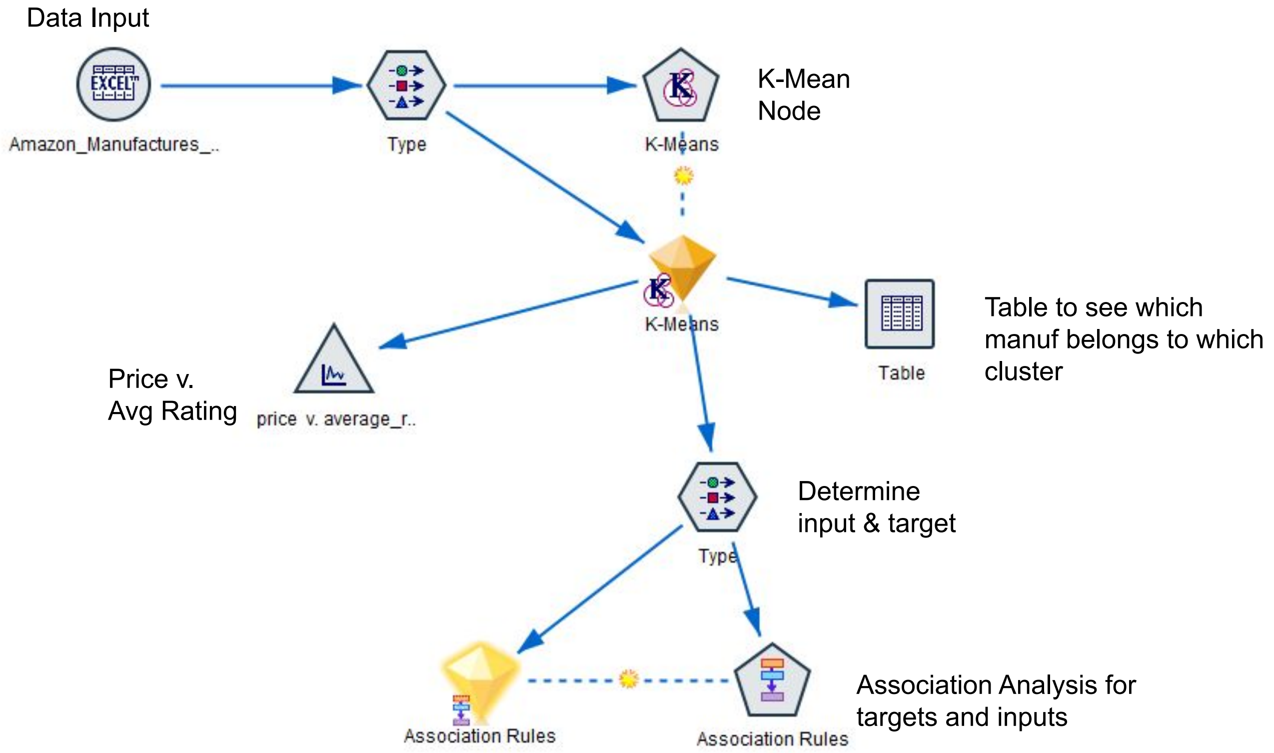
a. The specified maximum number of items in a rule was not reached due to insufficient number of frequent itemsets at previous levels.

K-Mean Model Analysis:



FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK
Gabelli School
of Business

Association Analysis



Sentiment Analysis

Code for Manufacturer

```
# -*- coding: utf-8 -*-
"""
Created on Sun Nov 29 19:49:43 2020

@author: himit
"""

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from datetime import datetime
import time
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from collections import defaultdict
import seaborn as sns
from nltk.corpus import stopwords
nltk.download('vader_lexicon')
nltk.download('punkt')

sentences = pd.read_excel(r"C:\Users\himit\Desktop\Fall 2020\Data Mining\Project\Amazon_Manufactures_GenCategories.xlsx", index_col= 0)
sentences.head()


sid = SentimentIntensityAnalyzer()

# Array to hold sentiment
sentiments = []

# Declare variables for sentiments
compound_list = []
positive_list = []
negative_list = []
neutral_list = []



from nltk import sentiment
from nltk import word_tokenize


sentences.isnull().sum()

sentences_up=sentences.dropna(subset=['customer_reviews_substring'])

bodies = sentences_up['customer_reviews_substring'].to_list()
names = sentences_up['manufacturer'].to_list()
categories = sentences_up['general category'].to_list()

sentences_up.isnull().sum()

for index,body in enumerate(bodies):
    #body = bodies ['body']
    compound = sid.polarity_scores(body)['compound']
    pos = sid.polarity_scores(body)['pos']
    neg = sid.polarity_scores(body)['neg']
    neu = sid.polarity_scores(body)['neu']

    x=({"Name": names[index], "Body":body,"Compound": compound,
                       "Positive": pos,
                       "Negative": neu,
                       "Neutral": neg})
    sentiments.append(x.copy())

print(sentiments)
sentiments_pd = pd.DataFrame.from_dict(sentiments)
sentiments_pd.to_csv("Results_manufactures.csv")
sentiments_pd.head()

# ploting the results

sent = sentiments_pd.pivot_table(index = 'Name', values ='Compound', aggfunc = np.mean)
sent

# Bar Graph

colors = ["pink","green","red","blue","yellow"]
x_axis = np.arange(len(sent.index.values))
tick_locations = [value+0.4 for value in x_axis]
fig,ax=plt.subplots(figsize=(40, 14))
plt.xticks(tick_locations, sent.index.values, rotation="horizontal")

plot=plt.bar(sent.index.values, sent["Compound"], color=colors, alpha=1, align="edge")
plt.grid()


plt.title("Overall Manufacturer sentiments based on customer text review" )
plt.xlabel("Manufacturer")
plt.ylabel("Overall rating")
ax.grid(linestyle="dotted")

plt.show()
```

Code for categories

```
# -*- coding: utf-8 -*-
"""
Created on Fri Dec  4 19:12:20 2020

@author: himit
"""

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from datetime import datetime
import time
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from collections import defaultdict
import seaborn as sns
from nltk.corpus import stopwords
nltk.download('vader_lexicon')
nltk.download('punkt')

sentences = pd.read_excel(r"C:\Users\himit\Desktop\Fall 2020\Data Mining\Project\Amazon_Manufactures_GenCategories.xlsx", index_col= 0)
sentences.head()


sid = SentimentIntensityAnalyzer()

# Array to hold sentiment
sentiments_cat = []

# Declare variables for sentiments
compound_list = []
positive_list = []
negative_list = []
neutral_list = []



from nltk import sentiment
from nltk import word_tokenize


sentences.isnull().sum()

sentences_up=sentences.dropna(subset=['customer_reviews_substring'])

bodies = sentences_up['customer_reviews_substring'].to_list()
categories = sentences_up['general category'].to_list()

sentences_up.isnull().sum()

for index,body in enumerate(bodies):
    #body = bodies ['body']
    compound = sid.polarity_scores(body)['compound']
    pos = sid.polarity_scores(body)['pos']
    neg = sid.polarity_scores(body)['neg']
    neu = sid.polarity_scores(body)['neu']

    x=({"Category": categories[index], "Body":body,"Compound": compound,
                       "Positive": pos,
                       "Negative": neu,
                       "Neutral": neg})
    sentiments_cat.append(x.copy())

print(sentiments_cat)
sentiments_cat = pd.DataFrame.from_dict(sentiments_cat)
sentiments_cat.to_csv("Results_cat.csv")
sentiments_cat.head()


sentiments_cat.Category.unique()

# ploting the results

sent_cat = sentiments_cat.pivot_table(index = 'Category', values ='Compound', aggfunc = np.mean)
sent_cat

# Bar Graph

colors = ["pink","green","red"]
x_axis = np.arange(len(sent_cat.index.values))
tick_locations = [value+0.4 for value in x_axis]
fig,ax=plt.subplots(figsize=(10,7))
plt.xticks(tick_locations, sent_cat.index.values, rotation="horizontal")

plot=plt.bar(sent_cat.index.values, sent_cat["Compound"], color=colors, alpha=1, align="edge")
plt.grid()


plt.title("Overall category sentiments based on customer text review" )
plt.xlabel("Category")
plt.ylabel("Overall rating")
ax.grid(linestyle="dotted")

plt.show()
```