

**Project:**

**Amazon**

**Movies & TV Products Analysis**

Abstract

Amazon.com is an electronic commerce company where consumers can pick and choose from millions of product available on the website. Majority of the products are from third-party sellers/vendors.

For both vendors and consumer are keen for best sellers/popular products. The Amazon sales rank (ASR) provides an indication of the popularity of a product sold on any Amazon locale. Effectively, it is a "best sellers list" for the millions of products stocked by Amazon.

Products that appear in the best seller list enjoy additional exposure on the Amazon website and this may lead to an increase in sales.

Project Introduction

Amazon derives many of its sales from third-party sellers who sell products on Amazon (around 40% in 2008). Third-party Sellers receive commission by Amazon if the product results in a sale. Third-party Sellers can access the Amazon catalog directly through AWS. Products brought together are based on customers' shopping history.

Amazon sales rank

The Amazon sales rank (ASR) provides an indication of the popularity of a product sold on any Amazon locale. Effectively, it is a "best sellers list" for the millions of products stocked by Amazon. Products that appear in these lists enjoy additional exposure on the Amazon website and this may lead to an increase in sales. In particular, products that experience large jumps (up or down) in their sales ranks may be included within Amazon's lists of "movers and shakers"; such a listing provides additional exposure that might lead to an increase in sales.

Through this project we are going to explore.

1. Best Seller: Top 10 of product in the category based on their Sales Rank
2. Co-purchasing : Top 3 popular products that are bought together with best seller
3. Price positioning: Explore price range of the best seller compared to the price range of category.

About the dataset

The dataset consist of a JSON file.

Downlaoded from : http://snap.stanford.edu/data/amazon/productGraph/categoryFiles/meta\_Movies\_and\_TV.json.gz

For this project we are using Metadata that includes descriptions, price, sales-rank, brand info, and co-purchasing links:

The size of the database is 264.39 mb and contain information about 200,000 products

A snapshot of one piece of product meta-data information:

{ "asin": "0000031852",

"title": "Girls Ballet Tutu Zebra Hot Pink",

"price": 3.17,

"imUrl": "http://ecx.images-amazon.com/images/I/51fAmVkTbyL.\_SY300\_.jpg", "related":

{ "also\_bought": ["B00JHONN1S", "B002BZX8Z6", "B00D2K1M3O", "0000031909", "B00613WDTQ", "B00D0WDS9A", "B00D0GCI8S", "0000031895", "B003AVKOP2", "B003AVEU6G", "B003IEDM9Q", "B002R0FA24",],

"also\_viewed": ["B002BZX8Z6", "B00JHONN1S", "B008F0SU0Y", "B00D23MC6W", "B00AFDOPDA", "B00E1YRI4C", "B002GZGI4E", "B003AVKOP2", "B00D9C1WBM", "B00CEV8366", "B00CEUX0D8", "B0079ME3KU",],

"bought\_together": ["B002BZX8Z6"] },

"salesRank": {"Toys & Games": 211836},

"brand": "Coxlures",

"categories": [["Sports & Outdoors", "Other Sports", "Dance"]] }

* asin – product Id
* title – product Name
* price – product price in us dollar
* imUrl - product image’s url
* related - also bought, also viewed, bought together, buy after viewing
* salesRank - sales rank information
* brand – brand name
* categories - categories the product belongs to

Best Sellers:

The Best Sellers are filtered based on Sales Rank. The business wants to know the top ten products that are popular among the customers. The Sales tank indicator is used for this purpose.

Approach

The following steps were followed to answer the business question:

* Loaded the dataset (meta\_Movies\_and\_TV. json) file into HDFS.
* Created a spark sql program and have declared four variables-­‐ asin, description, price realted, title and ID.
* reading JSON file to create rdd to get salesrank and associated category: aDataRdd referencing complete path of actual data file (meta\_Movies\_and\_TV. json)
* Reviewing the first 100 rows.
* amoviescat variable holds the 'Indicator Name' Column from meta\_Movies\_and\_TV. json file
* amoviessalesRank variable holds 'Sales Rank' column from meta\_Movies\_and\_TV. json. .
* Filtered the dataset with Asin and Sales Rank
* Limited the sorted result set to Top 10 rows.

Findings

Here are the Top ten Seller based on Sales Rank

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Asin : 0005041104  Sales Rank : 865267Camp Harmony [VHS] | Asin : 0005059836  Sales Rank : 570349  Product Details | Asin : 000510372X  Sales Rank : 701863Product Details | Asin : 0005165687  Sales Rank : 936793  Product Details | Asin : 0005200288  Sales Rank : 183133  Product Details |
| Asin : 0005253446  Sales Rank : 673158  Product Details | Asin : 0005257891  Sales Rank : 738164  Product Details | Asin : 0005419263  Sales Rank : 390213Product Details | Asin : 0005445825  Sales Rank : 544159  Product Details | Asin : 0005452287  Sales Rank : 475950  Product Details |

From the list of best seller on Amazon, none them are popular movies or tv show that we think of.

Co-Purchasing:

The co-purchasing is about add-on purchase in a single transaction. This is where the “Frequently Bought Together” section comes in. By indicating to customers what other item has been frequently purchased along with the one they are buying, business increases the chances of the order to grow in size, resulting in increased financial returns. The business wants to know the top three products that are frequently brought together by customers. The “bought\_together” column is used for this purpose.

Approach

The following steps were followed to answer the business question:

* Registering the product dataframe as temporary table.
* Since Bought together field entries are not available for every top selling product we are selecting one which has entries.
* Fetching “Related” field for the record selected.
* “also\_bought” field is coming as a row of multiple entries. So we split these and created a dataframe of products which were bought together.
* We want to find the 3 popular products which were bought together. We are taking sales rank as the metric of popularity.
* We are getting sales rank of these products by making a join wth aproduct table.
* By making the join we get the 3 products with sales rank in ascending order.

Findings:

|  |  |  |
| --- | --- | --- |
| Bought Frequently:1 | Bought Frequently : 2 | Bought Frequently: 3 |
| 1606830090Product Details | 1599754169  Product Details | 031068546X  Product Details |

The most popular items are from same category. A business can look into cross selling similar type to items.

Price positioning:

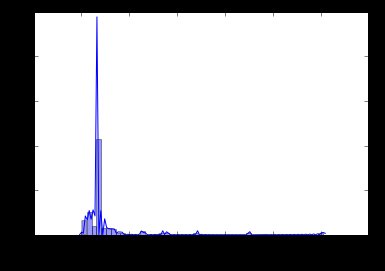
Pricing influences how the market perceives the products offering. In a Price-Value Matrix many factors influence price of product, mostly compared with competitors’ rates and products. The business wants to know the best sellers pricing to ascertain its relative position in the competition. The business then needs to apply different pricing strategies like Skim, Match, Surround or Undercut to influence its products positioning.

Approach

The following steps were followed to answer the business question:

* We took product ids and sales rank from aProduct table, arranged in ascending order using sales rank , and created a sql dataframe.
* We converted this dataframe into a pandas dataframe so that we can perform some graphical analysis over it.
* Most of the products have missing values for price. We need to handle these as we want to see the price distribution of products sold according to the sales rank.
* We filled NAN using the median of the price column ( Median seemed a good choice considering the dataset is skewed) .
* Using seaborn’s distplot we plotted the price distribution .

Findings:

From all the prices in the category range the following plot is drawn.

the above graph shows a right skewed distribution , which suggests that low sales ranked(Frequently bought) items have low prices having the 75% products below price 16.08 $. Total 50% products fall between 13.46 and 15.95. Above inferences are based on the taken sample only.

May be for category Movies & TV people tend to buy less expansive options.(one time watch may be) :)

Conclusion:

The availability of E-commerce has turned people into savvy online shoppers, and the power has shifted from business to consumer. Often, shoppers bypass the idea of brand loyalty if the product is available at a lower price somewhere else. Determining a unique selling proposition is a foundational step in any good brand strategy. The goal is to differentiate your product from your competitors, and give your customers a reason to purchase. Citation

<https://en.wikipedia.org/wiki/Amazon_(company)>

Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering

R. He, J. McAuley

WWW, 2016

pdf

Image-based recommendations on styles and substitutes

J. McAuley, C. Targett, J. Shi, A. van den Hengel

SIGIR, 2015

pdf