CLUSTERING AND UNDERSTANDING TOY STORES IN LOS ANGELES

IBM Data Capstone Project for Professional Certificate

Abstract

Research, findings, and discussion on using FourSquare Venues API with Python Machine Learning Libraries to Cluster and Analyze Data

By Joshua Pascascio

Joshua.pascascio@gmail.com

TABLE OF CONTENTS

INTRODUCTION 1

BUSINESS PROBLEM 2
AUDIENCE AND USE CASES 2

DATA 3

DATA SOURCES 3
DATA RETRIEVAL 3

METHODOLOGY #

DATA PREPARATION #
CLUSTERING RESULTS AND CLUSTERING STRUCUTRE
CLUSTER RESULTS IN THE TARGET AREA OF LOS ANGELES, CA

RESULTS #

EXPLAINING AND ANALYZING MODEL CLUSTERS
#

DISCUSSION AND RECOMMENDATIONS #

GIVING GREATER UNDERSTANDING TO DIFFERENTIATING TOY STORES #
BUSINESS STRATEGIES AND INSIGHTS FOR CUSTOMERS AND VENDORS
MARKET PENETRATION FOR NEW TOY STORE VENDORS
MARKET LOCATION STRATEGRIES FOR NEW TOY STORE VENDORS #
USING CLUSTER DATA TO IMPROVE TOY SEARCH AND BUYING EXPERIENCE
#
IMPROVEMENTS AND FUTURE RESEARCH RECOMMENDATIONS

CONCLUSION #

INTRODUCTION

This project is designed to provide new categories and ways to group toy stores in Los Angeles as to help provide a more detailed, personal insights for both toy vendors and toy buyers.

BUSINESS PROBLEM

Playing with toys is nearly every person's most cherished memory from his or her childhood.
While some stop playing with toys as they get older and approach adolescence or adulthood. Many more continue to collect and appreciate them as adults. This is because toys extend much further then the cheap action figures or dolls we might remember from our childhood; many can be intricate, handcrafted pieces of art that can be worth a fortune. This makes the vendors, especially those with brick-and-mortar retail shops, vary greatly in their merchandise and what kind of audience and clientele they cater to.

This can make trying to understand toy stores and their customers very difficult because the products and price points can vary so widely. Many toy store customers whether children or parents looking to buy toys for their kids to the seasoned collector with shelves of figures worth more thousands need a better way to separate and categorize all these stores so they can find the store that has the toy for them. Vendors may want to understand how to categorize themselves in relation to other stores so they can find their best customers and optimize their business and location strategies.

AUDIENCE

The audience for this report includes both toy store vendors, meaning companies who have physical stores that sell, sometimes exclusively, some sort of toy products as well as the customers. The customers that are the primary focus are toy collectors especially those that collect specific products like FunKo Pops or anime figurines from a franchise. For vendors, these insights, especially when it comes to what categories and related category of store they belong to, can give guidance on to their marketing and what products they should focus on for promotion. For customer it can provide easier ways to search and navigate stores based on their interest.

DATA

DATA SOURCES

The primary data source will be data retrieved from the FourSquare API on venues and categories. To get a better understanding of the geography and neighborhoods of Los Angeles we will also be using data fetched from the UCLA website. (http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_LA_Neighborhoods_Data#SOCR_Data_-_Los_Angeles_City_Neighborhoods_Data)

FOURSQUARE API

There will be a total of 3 different endpoints that will be queried to fetch the dataset. One will be the *search* or *explore* endpoint to get a list of venues that are toy store related and in Los Angeles. The following input will be to search each endpoint by the VENUE_ID in each result to get more venue-specific data including related categories, store hours, visitors and other data that can be used as features for clustering. The final one would be one to get all the categories and related venue ids that will be used later for clustering.

UCLA DATA

We will need to use either a web crawling or HTML parsing library to fetch the neighborhood and geographical coordinates and use the for cluster analysis an