Investigating the Feasibility of Opening Video Game Stores in St. Louis, MO, USA

Applied Data Science Capstone Week 4

Taher Hajilounezhad

November 2019

Introduction

St. Louisis an independent city (city not in a county) in the U.S. state of Missouri. It is the second largest city in the state of Missouri behind Kansas City. It is situated along the western bank of the Mississippi River, which forms the state line between Illinois and Missouri. The estimated 2018 population of the city proper was 302,838 and the bi-state metropolitan area was 2,804,724. Greater St. Louis is the largest metropolitan area in Missouri, second-largest in Illinois, 7th-largest in the Great Lakes Megalopolis, and the 20th-largest in the United

States.The architecture of St. Louis exhibits a variety of commercial, residential, and monumental architecture. St. Louis is known for the Gateway Arch, the tallest monument constructed in the United States at 630 feet (190 m). The Arch pays homage to Thomas Jefferson and St. Louis's position as the gateway to the West.

The computer and video-game industry has grown from focused markets to mainstream. They took in about 9.5 billion USD in the US in 2007, 11.7 billion in 2008, and 25.1 billion in 2010 (ESA annual report) and is estimated to be around 140 billion USD in 2018. Although, The games industry's shift from brick and mortar retail to digital downloads led to a severe sales decline at video game retailers, but retail game stores diversified its services by purchasing chains that repair wireless devices and expanding its trade-in program through which customers trade used games for credit towards new games. Other chain stores revamped its stores so customers would spend time playing games there. . It built a gaming arena for events and tournaments.

Also, here is some figures about how much children play video games:

- 66% of tweens aged 8 to 12 play video games for an average of 2 hours per day

- 56% of teens ages 13 to 17 play video games for an average of 2.5 hours per day

- Over 80% of both tweens and teens have a gaming console

- Young children aged 2 to 4 play for 21 minutes per day,

- Young children aged 5 to 8 play for 42 minutes per day

Although there are a few video game stores, but the increased interest in video gaming specially among teenagers or even younger childer opens a huge door to start new video game stores. A video game retail chain stores is interested to investigate possible locations for new video game stores.

Global revenues of the video game industry from 1971 to 2018, not adjusted for inflation.[40]
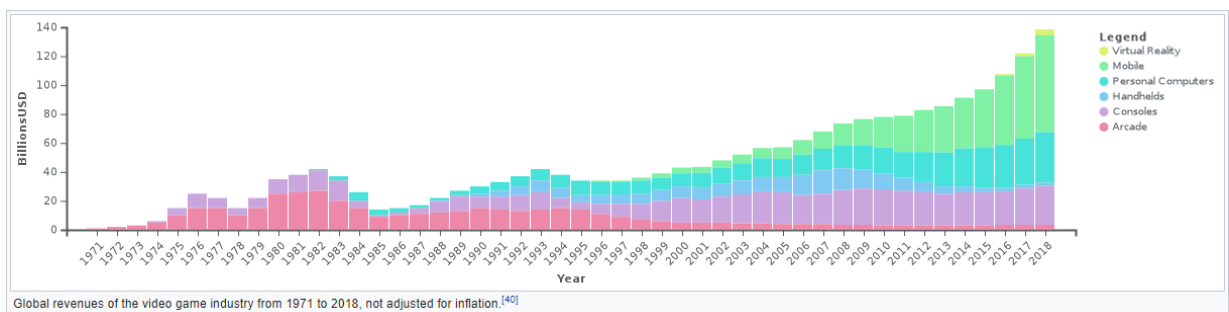
Business Problem

This capstone project aims to study the feasible locations in St. Louis, MO to open new video game stores. Various data science techniques like clustering by machine learning is employed to provide a comprehensive study to suggest to the city hall what locations are best candidates for the video game stores.

The Target Audience

This project is primarily defined to look for mutiple alternatives the city of St. Louis has to open new department stores. This study can also be used by other chain retail video gaing companies to evaluate the investment opportunities.

Data

Following data are used to answer business questions:

- List of neighborhouds in St. Louis, MO. This determins the scope of project in the heart of Missouri

- Latitude and longitude of the neighborhouds to plot the map and to get the venue data

- Venue data specially data related to other local businesses. It will be uised for clustering the neigborhoods.

Source of Data

This page on wikipedia "https://en.wikipedia.org/wiki/List_of_neighborhoods_of_St._Louis" (https://en.wikipedia.org/wiki/List_of_neighborhoods_of_St._Louis") provides the list of neighborhoods in St. Louis, MO. Web scraping techniques will be employed to extract data from this page, using Python requests and Beautifulsoup package. Later, Python Geocoder will be used to obtain the geographical coordinates of the neighborhoods. Then, the Foursquare API is used to get the venue data. It will provide various categories of venue data including department stores. This will help us tackle our business problem. After that, data preparation and wrangling will be applied to finally buid a clustering machine learning model (K-means for this proeject). Finally, Folium will be used for map visualization.

Methodology

At first, the list of neighborhoods in St. Luois, MO, USA is obtained. I have used (https://en.wikipedia.org/wiki/List_of_neighborhoods_of_St._Louis#List_of_neighborhoods

(https://en.wikipedia.org/wiki/List_of_neighborhoods_of_St._Louis#List_of_neighborhoods))
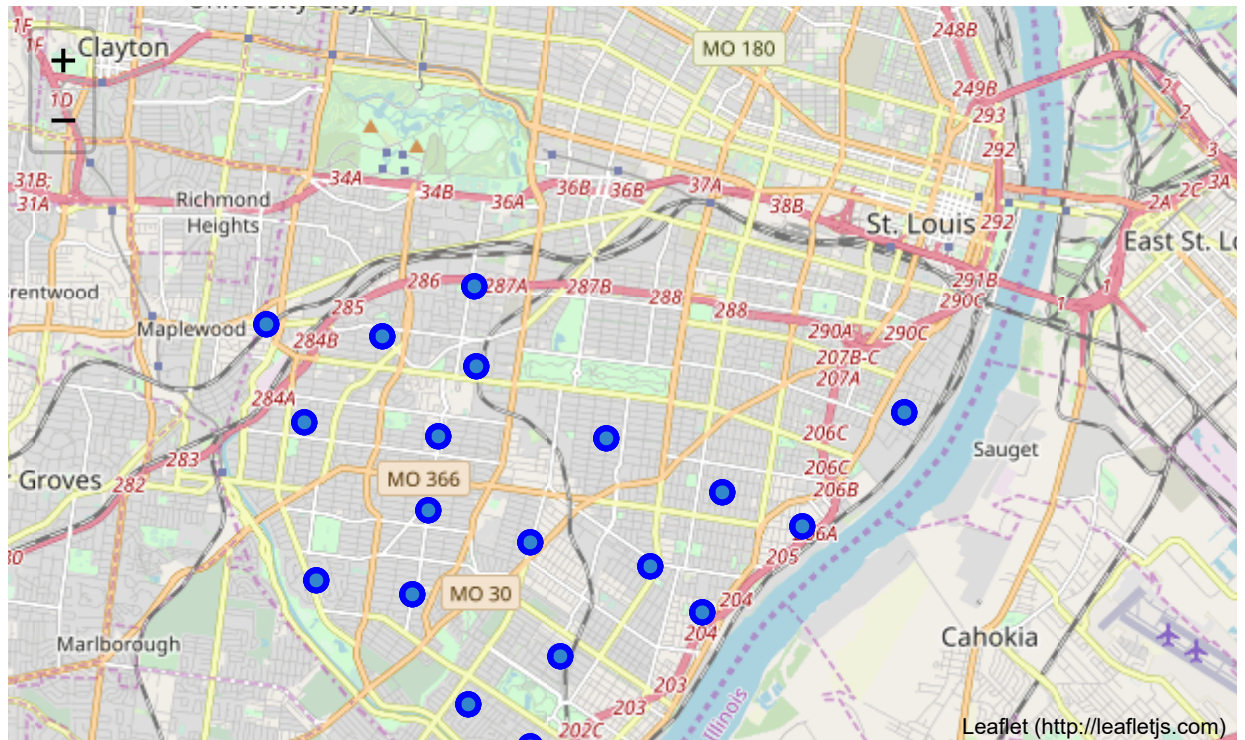to access data. Later, web scrapping is performed by Pytho nrequests and beautiful soup
packages to extract the list of neighborhood data. In order to be able to use Foursquare
API, we need to get the geographical coordinates as latitude and longitude. Accordingly,
Geocoder package is employed that enables converting addresses into coordinates. Then,
data is populated into a dataframe. Folium package is used to visualize the neighborhoods
in map. This is a double check of the accuracy of geographical data extracted by
Geocoder.

Then, Foursquare API is used again to get the top 100 venues within the radius of 2000
meters. API calls are utilized to to Foursquare passing in the geographical coordinates of
the neighborhoods within a loop. Foursquare will return the venue data in JSON format
from which venue name, category, latitude and longitude is extracted. The the number of
returned venues for each neighborhood is checked and the. The number of unique
categories is also determined. Neighborhoods are grouped and the mean frequency of
occurance of each unique venue category is calculated.

K-means clustering algorithm is applied on data with three number of clusters. It is a
common unsupervised clustering technique and is a good fit to our type of project.
NEighborhoods are clustered into 3 clusters based on their frequency of occurance for
"Video Game Store". IT is clear from the results which neighborhoods have higher
concentration of video game stores. These results are used to answer the buisness
questions.

In [266]:

Out[266]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: