Data Science capstone project

Project name: Project_TopUS_Revenue
This is a capstone project for IBM Data Science Professional Certificate.

The idea is to find top affluential neighborhoods in the US and compare those neighborhoods and determine how similar or dissimilar they are based on their revenue generating venues/businesses. This idea is interesting to business investors to decide on the next location/business to invest.

Introduction/Business Problem:

Target audience: business investors who are interested in investing in the US local businesses

Every day, people interest and activities change based on their personal preference and the neighborhood they choose to reside. Business investors need to know what are the main revenue streams in the US and how those are geographically segregated across different neighborhoods in different cities. The knowledge of what main hubs are in each neighborhood and what main business streams are in each neighborhood and in the US in the big picture are important as that guides business investors on how to invest at different locations/cities. Information on what type of business are booming and what are the main social hubs at different locations can be very useful prior to the investment. Hence, the question is what are the main revenue generating local businesses within the US? And how those top revenue generating businesses geographically change within the US for most affluential cities?

Data Requirement:

First, I need to find main income generators in the US. This requires the knowledge of most affluent cities/regions and the most popular businesses in a specific city. I am going to scrape Wikipedia for this step and government websites for different states in the US. I am going to rank the top most affluential states based on their income. Second, I need to find the latitude and longitude of the most affluential neighborhoods across the US using geolocator. Third, I need to get information/data about the venues and businesses for these neighborhoods based on their specific locations. I am going to use Foursquare API to extract data about venues/businesses. Fourth, I am going to compare the top revenue generating revenues for the neighborhoods.

- List of affluential neighborhoods in the US (Wikipedia and State websites)
- The latitude and longitude of those neighborhoods (Geolocator)
- The average income for those neighborhoods (Wikipedia and State websites)
- Venue data for different neighborhoods (Foursquare API)