Data Description

Target Audience :

To solve this problem, as a data scientist my objective is to locate the best place with higher sales revenue, higher population and lowest rent value in order to get profitable business with lowest investment value.

1. Sales Data

The State of Iowa provides different datasets and visualizations about different categories such Communities, Commerce, Health... in the following website : https://data.iowa.gov/ .

In the Sales and Distribution category, specifically in the following website https://data.iowa.gov/Sales-Distribution/2019-Iowa-Liquor-Sales/38x4-vs5h , contains our first dataset. This filtered view contains the spirits purchase information of Iowa Class “E” liquor licensees by product and date of purchase for calendar year 2019. The dataset can be used to analyze total spirits sales in Iowa of individual products at the store level to get an idea on alcohol consumption and Sales. (.csv)

Column Name	Description
Invoice/Item Number	Concatenated invoice and line number associated with the ...
Date	Date of order
Store Number	Unique number assigned to the store who ordered the liqu...
Store Name	Name of store who ordered the liquor.
Address	Address of store who ordered the liquor.
City	City where the store who ordered the liquor is located
Zip Code	Zip code where the store who ordered the liquor is located
Store Location	Location of store who ordered the liquor. The Address, City...
County Number	Iowa county number for the county where store who order...
County	County where the store who ordered the liquor is located
Category	Category code associated with the liquor ordered
Category Name	Category of the liquor ordered.

Category Name	Category of the liquor ordered.
Vendor Number	The vendor number of the company for the brand of liquor...
Vendor Name	The vendor name of the company for the brand of liquor o...
Item Number	Item number for the individual liquor product ordered.
Item Description	Description of the individual liquor product ordered.
Pack	The number of bottles in a case for the liquor ordered
Bottle Volume (ml)	Volume of each liquor bottle ordered in milliliters.
State Bottle Cost	The amount that Alcoholic Beverages Division paid for eac...
State Bottle Retail	The amount the store paid for each bottle of liquor ordered
Bottles Sold	The number of bottles of liquor ordered by the store
Sale (Dollars)	Total cost of liquor order (number of bottles multiplied by t...
Volume Sold (Liters)	Total volume of liquor ordered in liters. (i.e. (Bottle Volume...
Volume Sold (Gallons)	Total volume of liquor ordered in gallons. (i.e. (Bottle Volu...

For more info, you can check the link to the datasets. This dataset contains a lot of information. I'll focus my study on Sales, Volume Sold for each store, city and county.

2. Population

The population dataset was taken also from [data.iowa.gov](). This dataset contains city population in Iowa from 2010 to 2018.

Column Name	Description	Type
FIPS	County FIP – A five-digit code for counties based on Federal...	Number #
County		Plain Text T
City		Plain Text T
Year	For population estimates the date of reference is always Jul...	Date & Time
Estimate		Number #
Primary Point	Primary latitude and longitude in decimal degrees for the c...	Point

FIPS	County	City	Year	Estimate	Primary Point
1966720	Keokuk	Richland	July 01, 2018	567	POINT (-91.9960251 41.1...
1924600	Scott	Eldridge	July 01, 2018	6,813	POINT (-90.5805427 41.6...
1906805	Benton	Blairstown	July 01, 2018	662	POINT (-92.0823291 41.9...
1912630	Dubuque	Centralia	July 01, 2018	136	POINT (-90.8365993 42.4...
1902350	Woodbury	Anthon	July 01, 2018	560	POINT (-95.8656328 42.3...
19031	Cedar	Balance of Cedar County	July 01, 2018	7,406	POINT (-91.1324125 41.7...
1965550	Fayette	Randalia	July 01, 2018	55	POINT (-91.8864465 42.8...
1980580	Cerro Gordo	Ventura	July 01, 2018	716	POINT (-93.4621042 43.1...
1962040	Mahaska	Pella (pt.)	July 01, 2018	2	POINT (-92.9177547 41.4...
1956055	Benton	Newhall	July 01, 2018	845	POINT (-91.9673411 41.9...
1944985	Jefferson	Libertyville	July 01, 2018	343	POINT (-92.0501957 40.9...
1963840	Pocahontas	Plover	July 01, 2018	70	POINT (-94.6222266 42.8...
1928020	Floyd	Floyd	July 01, 2018	318	POINT (-92.7402544 43.1...
1972345	Story	Sheldahl (pt.)	July 01, 2018	153	POINT (-93.6967164 41.8...

File source (.csv) is available in :
https://data.iowa.gov/Community-Demographics/City-Population-in-Iowa-by-County-and-Year/y8va-rhk9

3. Average Rent value

The average rent value is presented by the following Website :
https://www.rentdata.org/states/iowa/2019

This Website gives us the average rent and the population for every county in the state of Iowa in 2019. Using BeautifulSoup in python we will extract the data from the Website

County	0 BR	1 BR	2 BR	3 BR	4 BR	Est. Population
Adair County	$481	$502	$664	$877	$947	7,682
Adams County	$481	$517	$664	$960	$1,163	4,029
Allamakee County	$481	$506	$664	$865	$977	14,330
Appanoose County	$481	$502	$664	$840	$911	12,887
Audubon County	$481	$502	$664	$863	$898	6,119
Benton County Metro	$522	$526	$683	$871	$966	26,076
Black Hawk County Metro	$554	$662	$836	$1,087	$1,355	131,090
Boone County	$520	$603	$718	$980	$983	26,306
Bremer County Metro	$496	$564	$740	$927	$1,198	24,276
Buchanan County	$501	$516	$682	$918	$922	20,958
Buena Vista County	$444	$509	$664	$914	$1,020	20,260
Butler County	$481	$502	$664	$848	$898	14,867

Using the three features we can cluster our data and conclude using Foursquare API the best places to fit for our new pub.

4. Data Preparation

After cleaning and assigning data types on each variable of the datasets, From the first dataset I started by adding a column that calculates the revenue of each operation made by every liquor store.

The revenue variable is equal to the Sales made minus the vendor price. The vendor price is the multiplication of the bottle sold and State Bottle Cost :

```
liq['Revenu']=liq['Sale (Dollars)']-liq['Bottles Sold']*liq['State Bottle Cost']
liq.head()
```

tore ame	Address	City	Zip Code	Store Location	County Number	County	...	Item Description	Pack	Bottle Volume (ml)	State Bottle Cost	State Bottle Retail	Bottles Sold	Sale (Dollars)	Volume Sold (Liters)	Volume Sold (Gallons)	Revenu
tfish rlie's	1630 East 16th St	Dubuque	52001.0	NaN	31.0	DUBUQUE	...	Crown Royal	12.0	1000.0	18.89	28.34	1.0	28.34	1.00	0.26	9.45
ntral City quor, Inc.	1460 2ND AVE	Des Moines	50314.0	POINT (-93.619787 41.60566)	77.0	POLK	...	Rumchata	12.0	375.0	7.00	10.50	2.0	21.00	0.75	0.19	7.00
ım & f573 / SE DM	5830 SE 14th St	Des Moines	50315.0	NaN	77.0	POLK	...	Titos Handmade Vodka	12.0	750.0	9.64	14.46	6.0	86.76	4.50	1.18	28.92

The Revenue features represent our target variable that will help us take the decision to choose which city to open our Pub.

After having the revenu value on each operation. We should add the Population variable.

This dataset is csv file that contains each city population in the state of Iowa from 2010 to 2018. We cleaned our data to get the population on each city in the state of Iowa :

```
pop=pop[['City','Population']]
npop.head()
```

	City	Population
0	Orleans	589.0
1	Garwin	497.0
2	Mechanicsville	1133.0
3	Johnston	22040.0
4	Walnut	774.0

After filtering our data to get the average population in every city, let's add it to our first dataset. This column will be considered in the city analysis.