



# The Sales Situation of Liquors in Different Regions of Iowa 2017

# Team Members:

- Mingxuan Zhang
- Lei Teng
- Yuxiang Wang



# Questions sought to answer:

Our project works for analyzing the name, date, kind, price, quantity, and location of sales of individual containers or packages of containers of alcoholic beverages of Iowa 2017.

## 1.Sale situation:

- Which category of alcoholic beverages gains the maximum revenue ?
- Which brand of alcoholic beverages gains the maximum revenue?
- Which category of alcoholic beverages is the most popular?
- As for each vender, which regions gains the maximum revenue?
- How can we predict the category if know the store and vendor?
- Which weekday or month gains the maximum revenue?

## 2.Drink Behavior :

- Which category of alcoholic beverages is the most popular in different regions?
- People in which regions consumed the most alcoholic beverages?

# Data Preparation Work:

## 1. Realize the background knowledge about our data set

## 2. Data Cleaning

- We delete the rows which have missing values , because we need build a accurate Bayesian Classification Model, the missing values have the negative effects about veracity for our model.

## 3. Data Reduction

- The original datasets contains from 2012 to current , its over 12 millions. Our project is working for 2017, so we did the reduction only left datasets about 2017.
- There are 24 attributes in the dataset. Since we don't need all 24 attributes, we have to do Dimensionality reduction. We will remove irrelevant attributes like Invoice/Item Number. We also have to remove redundant attributes, for example, Volume Sold (Liters) and Volume Sold (Gallons) are redundant attributes, we will use Volume Sold(Liters) instead of Volume Sold (Gallons).

## 4. Data Transformation

- Sales numbers contain "\$" which is causing panda to not see them as numbers. So we deleted the "\$" and transfer from type string to type float.



4	S29050301	11/16/2015	3549	Quigley Mart 100 E WASHINGTON 50022	100 E WASH	50022	Page	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	6	\$162.84	4.50	1.19
5	S28867701	11/04/2015	2513	Hy-Vee Fo 812 S 1ST IOWA CITY 52240	812 S 1ST	52	Lee	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	2	\$19.20	0.30	0.08
6	S29050801	11/17/2015	3942	Twin Town 104 HIGH TOLEDO 52342	104 HIGH	86	Johnson	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39
7	S28869201	11/11/2015	3650	Spirits, Str 118 South HOLSTEIN 51025	118 South	47	Tama	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	2	\$19.20	0.30	0.08
8	S28865701	11/09/2015	2538	Hy-Vee Fo 1422 FLAM WATERLC 50702	1422 FLAM	07	Ida	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	1	\$53.34	1.75	0.46
9	S28869501	11/10/2015	3942	Twin Town 104 HIGH TOLEDO 52342	104 HIGH	86	Black Haw	1701100	DECANTE	962	Forbidden	6	1500	\$11.62	\$17.43	6	\$104.58	9.00	2.38
10	S29339301	11/30/2015	2662	Hy-Vee Wi 522 MULB MUSCATIN 52761	522 MULB	70	Tama	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	2	\$106.68	3.50	0.92
11	S29050901	11/16/2015	4307	Crossroad: 117 IOWA DUNLAP 712-2	117 IOWA	43	Muscataine	1701100	DECANTE	65	Laphroaig	12	750	\$19.58	\$29.37	4	\$117.48	3.00	0.79
12	S29049901	11/17/2015	2661	Hy-Vee Fo 1989 PARI SHELTON 51201	1989 PARI	71	Harrison	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	2	\$19.20	0.30	0.08
13	S28868201	11/05/2015	2561	Hy-Vee Fo 4605 FLEDES MOIN 50321	4605 FLE	77	O'Brien	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79
14	S28869601	11/09/2015	4114	After 5 Soi 704 W 7TH ATLANTIC 50022	704 W 7TH	15	Polk	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	2	\$106.68	3.50	0.92
15	S28866901	11/11/2015	3650	Spirits, Str 118 South HOLSTEIN 51025	118 South	47	Cass	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39
16	S29050101	11/19/2015	2806	Osco #8811307 N SE CLINTON 52732	1307 N SE	23	Ida	1701100	DECANTE	962	Forbidden	6	1500	\$11.62	\$17.43	1	\$17.43	1.50	0.40
17	S29049601	11/17/2015	2624	Hy-Vee #2 2395 NW / DUBUQUE 52002	2395 NW	31	Clinton	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79
18	S28868401	11/04/2015	2572	Hy-Vee Fo 6301 UNIV CEDAR FA 50613	6301 UNIV	07	Dubuque	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79
19	S29196301	11/24/2015	2595	Hy-Vee Wi 1620 4TH DENISON 51442	1620 4TH	24	Black Haw	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39
20	S29134301	11/18/2015	3723	J D Spirits 1023 9TH ONAWA 51040	1023 9TH	67	Crawford	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	6	\$162.84	4.50	1.19
21	S28869001	11/10/2015	2665	Hy-Vee / V 1005 E HW WAUKEE 50263	1005 E HW	25	Monona	1081200	CREAM LI	305	Rumchata	1	6000	\$99.00	\$148.50	1	\$148.50	6.00	1.59
22	S29198701	11/24/2015	5093	Cody Mart 1220 N CCLE CLAIRE 52753	1220 N CC	82	Dallas	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	2	\$106.68	3.50	0.92
23	S29282801	11/23/2015	2642	Hy-Vee Wi 512 E OSHEL PELLA 50219	512 E OSH	63	Scott	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	3	\$81.42	2.25	0.59
24	S28868001	11/04/2015	2548	Hy-Vee Fo 100 8TH S ALTOONA 50009	100 8TH S	77	Marion	1701100	DECANTE	962	Forbidden	6	1500	\$11.62	\$17.43	6	\$104.58	9.00	2.38
25	S29199601	11/20/2015	2558	Hy-Vee Fo 1700 E W MOUNT PI 52641	1700 E W	44	Polk	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39
26	S29050501	11/18/2015	3735	C B Liquor 1202 A AVIOSKALOC 52577	1202 A AV	10	Henry	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	18	\$488.52	13.50	3.57
27	S28867001	11/04/2015	3842	Bancroft Li 107 N POB BANCROF 50517	107 N POB	55	Buchanan	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	3	\$28.80	0.45	0.12
28	S28868601	11/09/2015	2650	Hy-Vee Wi 1808 23RD HARLAN 51537	1808 23RD	83	Kossuth	1701100	DECANTE	962	Forbidden	6	1500	\$11.62	\$17.43	3	\$52.29	4.50	1.19
29	S28869101	11/10/2015	2666	Hy-Vee #2 2510 SW / ANKENY 50023	2510 SW	77	Shelby	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39
30	S29197901	11/23/2015	3842	Bancroft Li 107 N POB BANCROF 50517	107 N POB	55	Polk	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39
31	S28865801	11/09/2015	2539	Hy-Vee Fo HIGHWAY IOWA FAL 50126	HIGHWAY	42	Kossuth	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	18	\$488.52	13.50	3.57
32	S28868701	11/10/2015	2651	Hy-Vee / V 1311 4 ST W WEVERLY 50677	1311 4 ST	09	Hardin	1701100	DECANTE	962	Forbidden	6	1500	\$11.62	\$17.43	6	\$104.58	9.00	2.38
33	S29050001	11/17/2015	2666	Hy-Vee #2 2510 SW / ANKENY 50023	2510 SW	77	Bremer	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	2	\$106.68	3.50	0.92
34	S29198101	11/23/2015	4162	Fareway S 4220 16TH CEDAR R/ 52404	4220 16TH	57	Polk	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79
35	S29195801	11/21/2015	2552	Hy-Vee Fo 20 WILSO CEDAR R/ 52404	20 WILSO	57	Linn	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	6	\$162.84	4.50	1.19
36	S29197801	11/23/2015	3650	Spirits, Str 118 South HOLSTEIN 51025	118 South	47	Linn	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	90	\$2442.60	67.50	17.83
37	S29049101	11/16/2015	2538	Hy-Vee Fo 1422 FLAM WATERLC 50702	1422 FLAM	07	Ida	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	18	\$488.52	13.50	3.57
38	S29049501	11/18/2015	2594	Hy-Vee Fo 4500 SER SIOUX CIT 51106	4500 SER	97	Black Haw	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	40	\$384.00	6.00	1.59
39	S29051001	11/19/2015	4988	Happy's W 5925 UNIV CEDAR FA 50613	5925 UNIV	07	Woodbury	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79
40	S28867101	11/09/2015	4604	Pit Stop Li 1324, 1st / NEWTON 50208	1324, 1st	50	Black Haw	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79
41	S29049801	11/16/2015	2643	Hy-Vee Wi 2126 KIME WATERLC 50701	2126 KIME	07	Jasper	1701100	DECANTE	962	Forbidden	6	1500	\$11.62	\$17.43	2	\$34.86	3.00	0.79
42	S29049201	11/17/2015	2544	Hy-Vee Fo 802 SOUT MARSHAL 50158	802 SOUT	64	Black Haw	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79
43	S29196501	11/21/2015	2607	Hy-Vee Wi 520 SO FF SHENAND 51601	520 SO FF	73	Marshall	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79
44	S29191201	11/19/2015	2248	Ingersoll Li 3500 INGEDES MOIN 50312	3500 INGE	77	Page	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	30	\$814.20	22.50	5.94
45	S29050601	11/17/2015	3813	CGI Foods 104 NORT MOUNT A/ 50854	104 NORT	80	Polk	1701100	DECANTE	65	Laphroaig	12	750	\$19.58	\$29.37	36	\$1057.32	27.00	7.13
46	S29199301	11/20/2015	2539	Hy-Vee Fo HIGHWAY IOWA FAL 50126	HIGHWAY	42	Ringgold	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	6	\$57.60	0.90	0.24
47	S29197301	11/21/2015	3443	Super Saw 1141 N BR COUNCIL 51503	1141 N BR	78	Hardin	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	12	\$325.68	9.00	2.38
48	S28868901	11/10/2015	2661	Hy-Vee Fo 1989 PARI SHELTON 51201	1989 PARI	71	Pottawatta	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	6	\$162.84	4.50	1.19
49	S29197501	11/20/2015	3525	Wines and 106 W 2ND WASHING 52353	106 W 2ND	92	O'Brien	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39
50	S29049301	11/19/2015	2567	Hy-Vee Dn 2200 WES DAVENPC 52806	2200 WES	82	Washingto	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	6	\$162.84	4.50	1.19
51	S29196201	11/20/2015	2591	Hy-Vee Wi 1602 E 7TH ATLANTIC 50022	1602 E 7TH	15	Scott	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	6	\$57.60	0.90	0.24
52	S29137201	11/18/2015	2566	Hy-Vee Fo 813 N LINK KNOXVILL 50138	813 N LINK	63	Cass	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	24	\$651.36	18.00	4.76
53	S29198301	11/21/2015	2554	Hy-Vee Fo 802 SOUT MARSHAL 50158	802 SOUT	64	Marion	1701100	DECANTE	962	Forbidden	6	1500	\$11.62	\$17.43	12	\$209.16	18.00	4.76
54	S29198401	11/21/2015	2555	Hy-Vee Fo 802 SOUT MARSHAL 50158	802 SOUT	64	Marion	1701100	DECANTE	962	Forbidden	6	1500	\$11.62	\$17.43	12	\$209.16	18.00	4.76

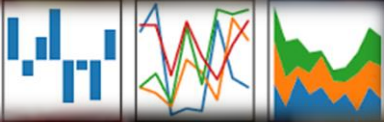


	Date	Number	Number	County	Category	Name	Number	Name	Number	Description	Pack	Volume (ml)	Cost	Retail	Sold
10708015	01/03/2017	4312	78.0	POTTAWATTA	1012200.0	Scotch Whiskies	55.0	SAZERAC NORTH AMERICA	8208	House Of Stuart	6	1750	\$10.52	15.78	
10708016	01/03/2017	4312	78.0	POTTAWATTA	1042100.0	Imported Dry Gins	35.0	BACARDI USA INC	28206	Bombay Dry Gin	12	750	\$10.50	15.75	
10708017	01/03/2017	4312	78.0	POTTAWATTA	1082000.0	Imported Cordials & Liqueurs	259.0	Heaven Hill Brands	65195	Hpnotiq	6	750	\$9.83	14.75	
10708018	01/03/2017	4312	78.0	POTTAWATTA	1081200.0	Cream Liqueurs	260.0	DIAGEO AMERICAS	68037	Bailey's Original Irish Cream	12	1000	\$16.50	24.75	
10708019	01/03/2017	4312	78.0	POTTAWATTA	1012100.0	Canadian Whiskies	65.0	Jim Beam Brands	10627	Canadian Club Whisky	12	1000	\$9.71	14.57	
10708020	01/03/2017	4312	78.0	POTTAWATTA	1062200.0	White Rum	55.0	SAZERAC NORTH AMERICA	44217	Barton Rum Light	12	1000	\$4.00	6.00	
10708021	01/03/2017	4312	78.0	POTTAWATTA	1062400.0	Spiced Rum	260.0	DIAGEO AMERICAS	43338	Captain Morgan Spiced Rum	6	1750	\$18.00	27.00	
10708022	01/03/2017	4312	78.0	POTTAWATTA	1031000.0	American Vodka	55.0	SAZERAC NORTH AMERICA	35318	Barton Vodka	6	1750	\$6.92	10.38	
10708023	01/03/2017	4312	78.0	POTTAWATTA	1032200.0	Imported Flavored Vodka	370.0	PERNOD RICARD USA	34051	Absolut Raspberri	12	1000	\$14.99	22.49	
10708024	01/03/2017	4312	78.0	POTTAWATTA	1032000.0	Imported Vodka	370.0	PERNOD RICARD USA	34007	Absolut Swedish Vodka 80 Prf	12	1000	\$14.99	22.49	
10708025	01/03/2017	4312	78.0	POTTAWATTA	1031000.0	American Vodka	461.0	Skyy Spirits	37987	Skyy Vodka	12	1000	\$12.35	18.53	

# List of tools:

pandas

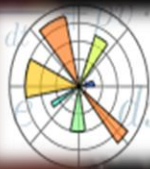
$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



python



NumPy



matplotlib

jupyterhub

- Delimit(datasets)
- Pandas
- Python
- Numpy
- Matplotlib
- Jupyter Notebook

# Classification :

## Bayesian Classification:

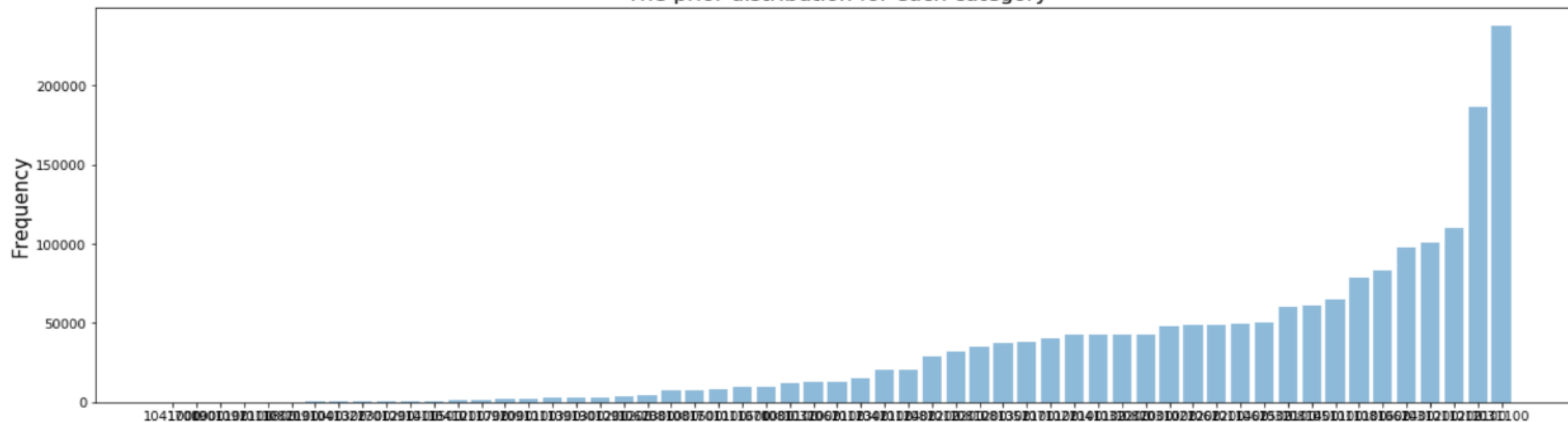
We build a model:

$P(\text{category} | \text{store number, vendor number})$

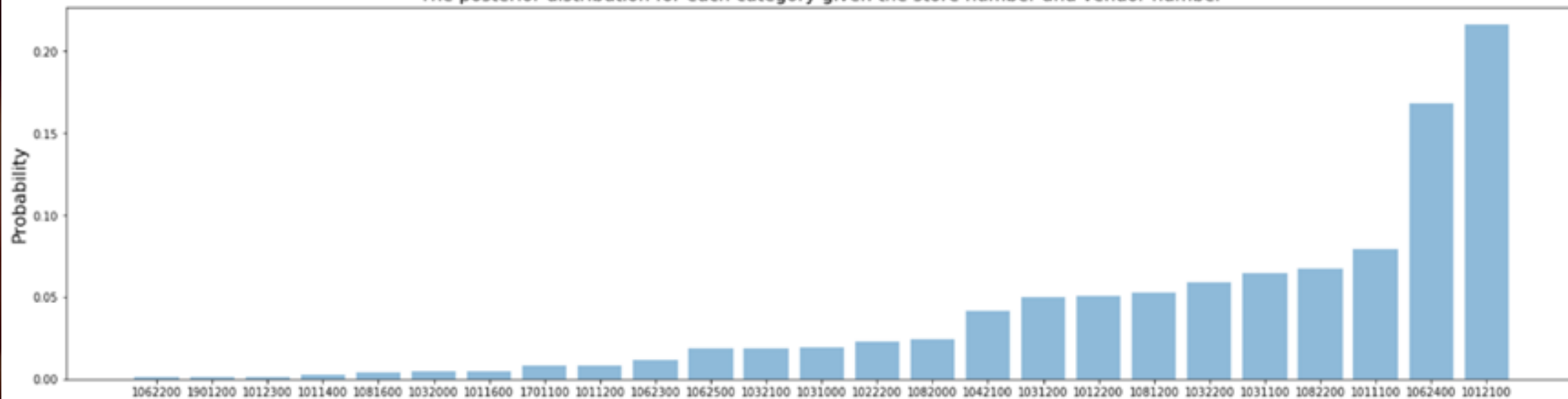
- If we have the store number and vendor number, we can get the posterior probability distribution for the categories. We used this to predict the values for category attribute, we can pick the one with highest probability or we can use this posterior distribution to draw samples for category attribute.
- We can use this to predict the category attribute after 2017 (like 2018), and fill in the missing data.



The prior distribution for each category



The posterior distribution for each category given the store number and vendor number



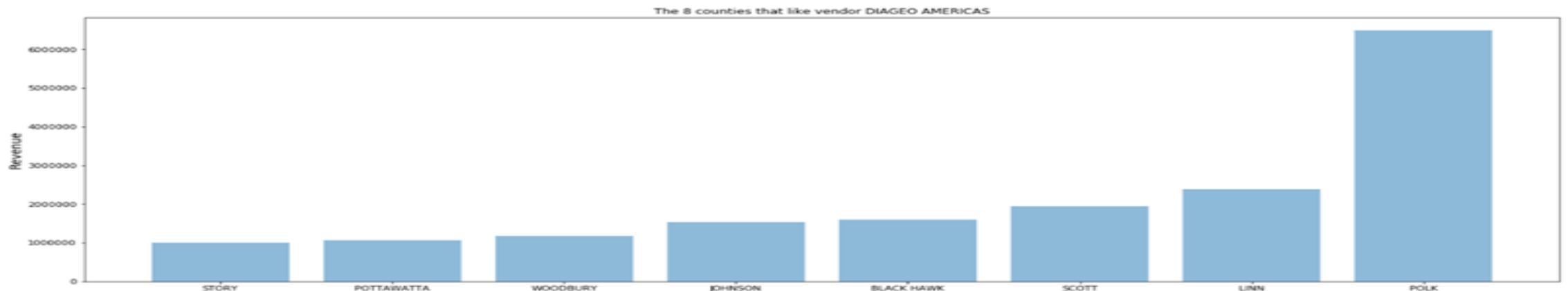
# Classification :

## Build Function to Analyzing:

1. We build a function to analyze the data to get the top 8 counties which brought the best revenue for each vender. Vender number as a input.

```
In [98]: def best8_counties(vendor_number):
sorted_v = sorted(vendor_county[vendor_number].items(), key=operator.itemgetter(1))
x, y = map(list, zip(*sorted_v))
c8 = y[len(y)-8:]
numbers = x[len(x)-8:]
objects = []
for number in numbers:
    objects.append(list(df.loc[df['County Number'] == number, 'County']))
y_pos = np.arange(len(objects))
plt.figure(figsize=(24, 8))
plt.bar(y_pos, c8, align='center', alpha=0.5)
plt.xticks(y_pos, objects)
plt.ylabel('Revenue', fontsize=14)
title = 'The 8 counties that like vendor ' + list(df.loc[df['Vendor Number'] == vendor_number, 'Vendor Name'])[0]
plt.title(title)
plt.show()
```

```
In [99]: best8_counties(260)
```



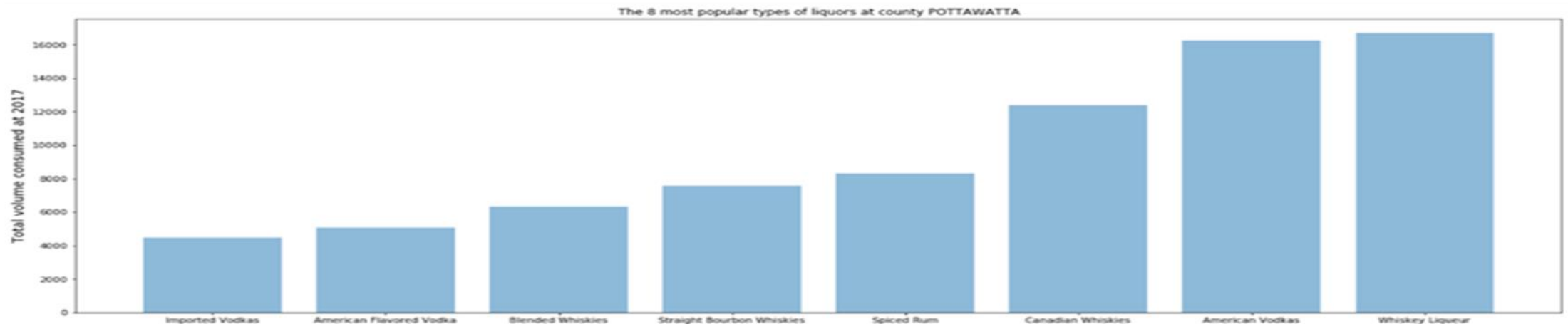


# Classification :

2. We build a function to analyze the data to get the top 8 **categories** consumed in different regions. County number as a input.

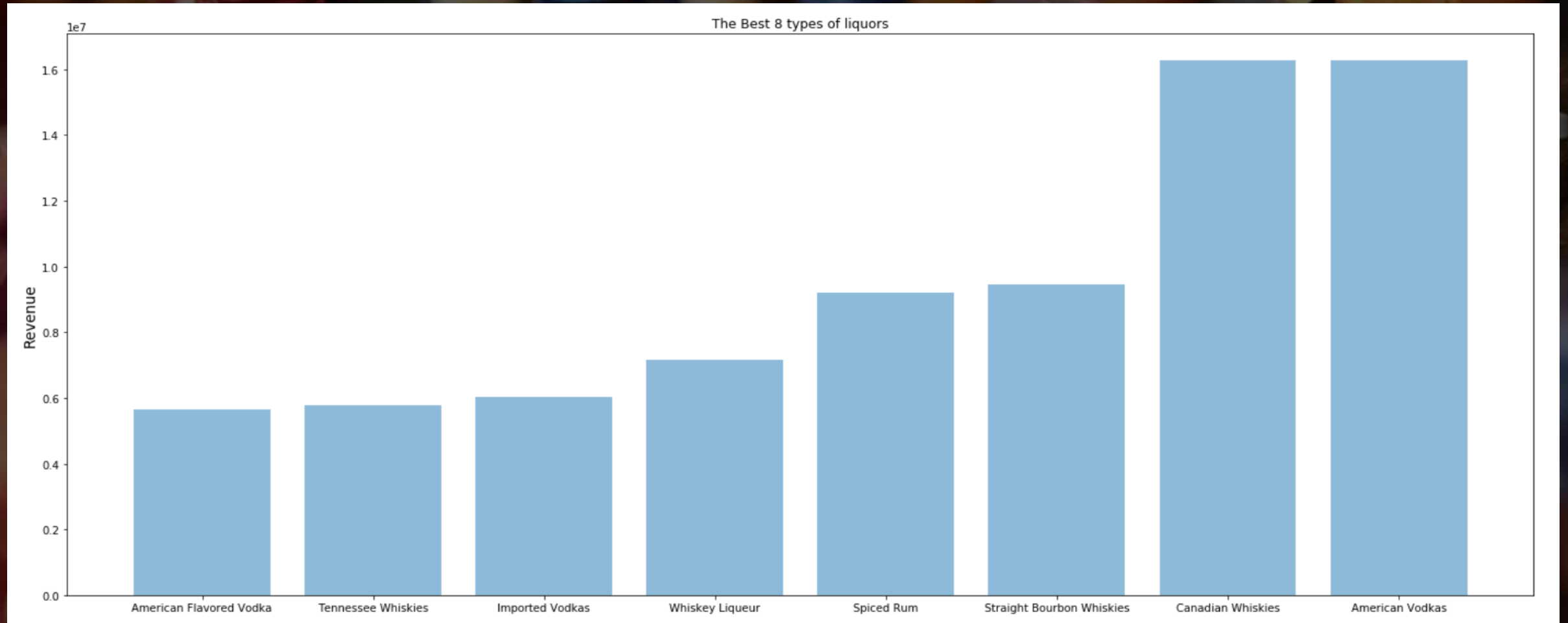
```
In [101]: def best_cate_county(CN):  
    sorted_c = sorted(county_cates[CN].items(), key=operator.itemgetter(1))  
    x, y = map(list, zip(*sorted_c))  
    c6 = y[len(y)-8:]  
    numbers = x[len(x)-8:]  
    objects = []  
    for number in numbers:  
        objects.append(list(df.loc[df['Category'] == number, 'Category Name'])[0])  
    y_pos = np.arange(len(objects))  
    plt.figure(figsize=(24, 8))  
    plt.bar(y_pos, c6, align='center', alpha=0.5)  
    plt.xticks(y_pos, objects)  
    plt.ylabel('Total volume consumed at 2017', fontsize=14)  
    title = 'The 8 most popular types of liquors at county ' + list(df.loc[df['County Number'] == CN, 'County'])[0]  
    plt.title(title)  
    plt.show()
```

```
In [102]: best_cate_county(78)
```



# Classification :

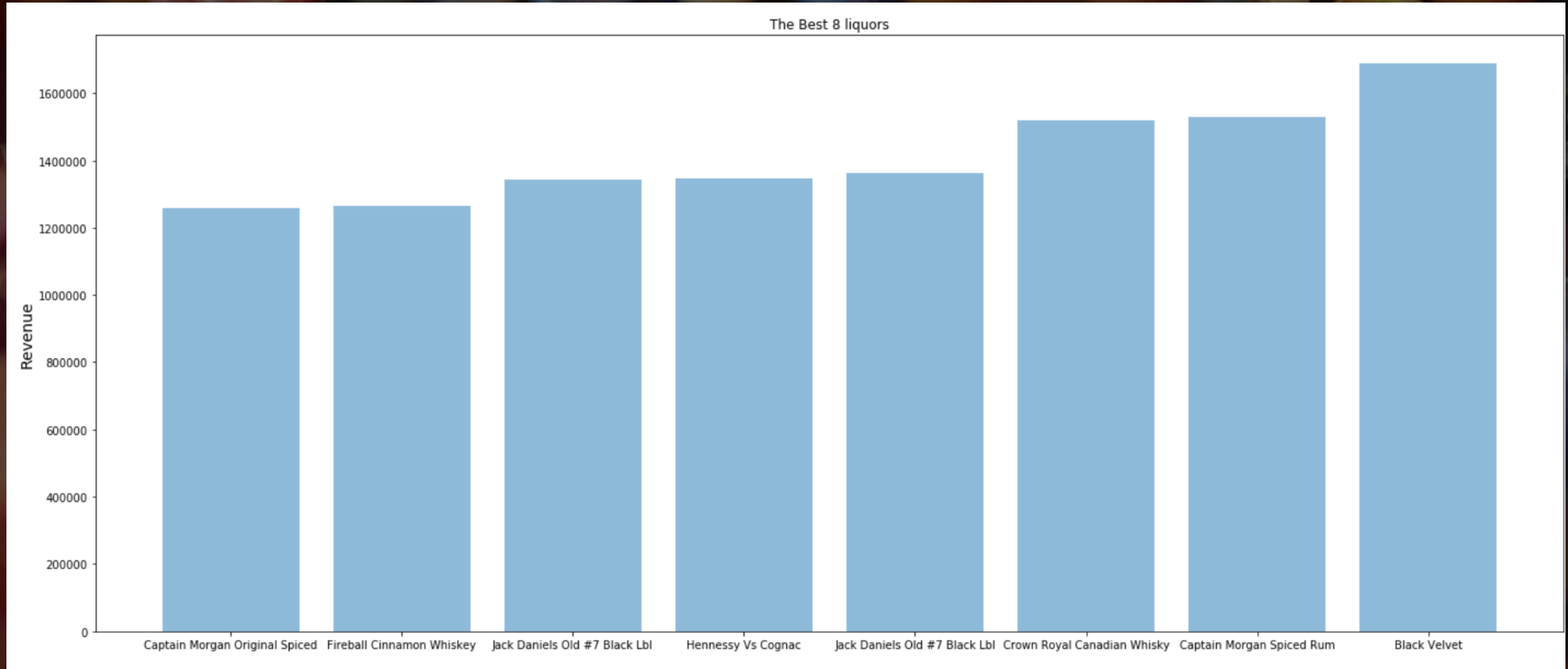
## Revenue and Category analyzing:





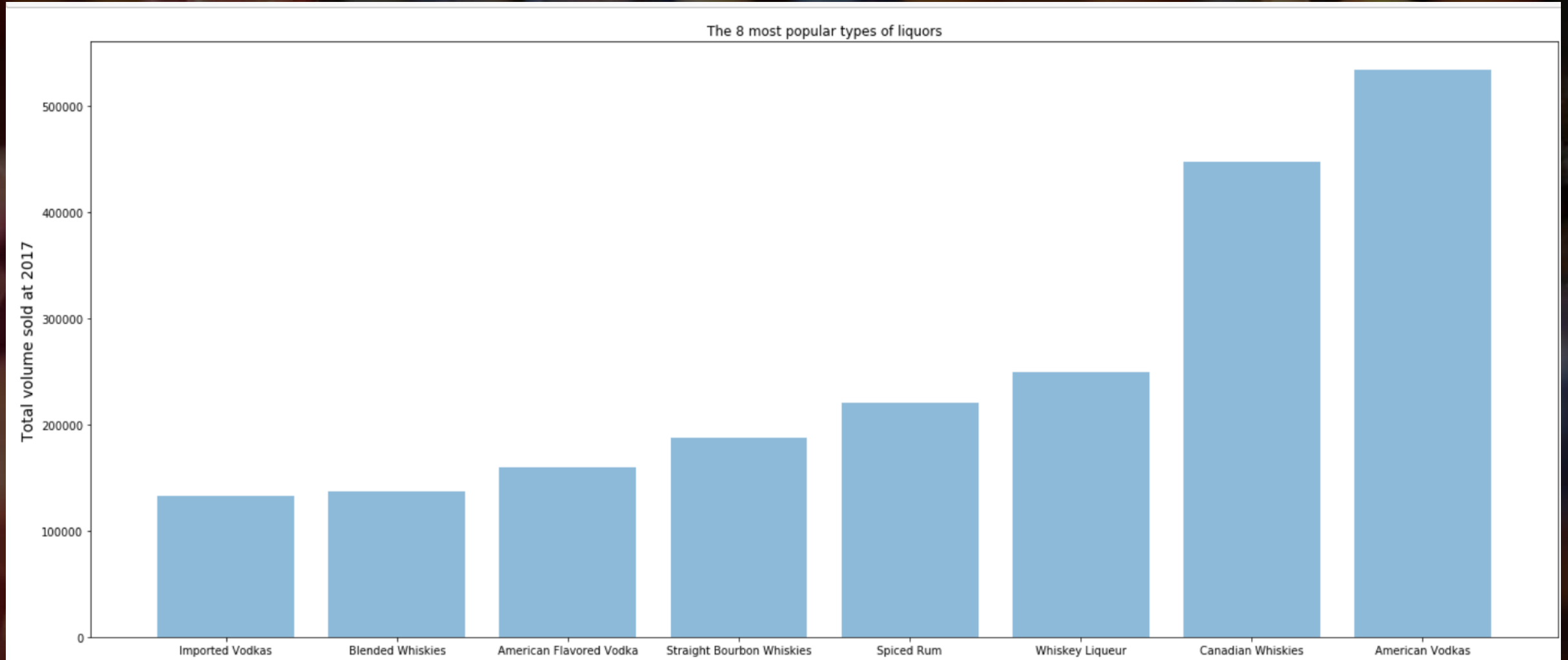
# Classification :

## Revenue and Brand analyzing:



# Classification :

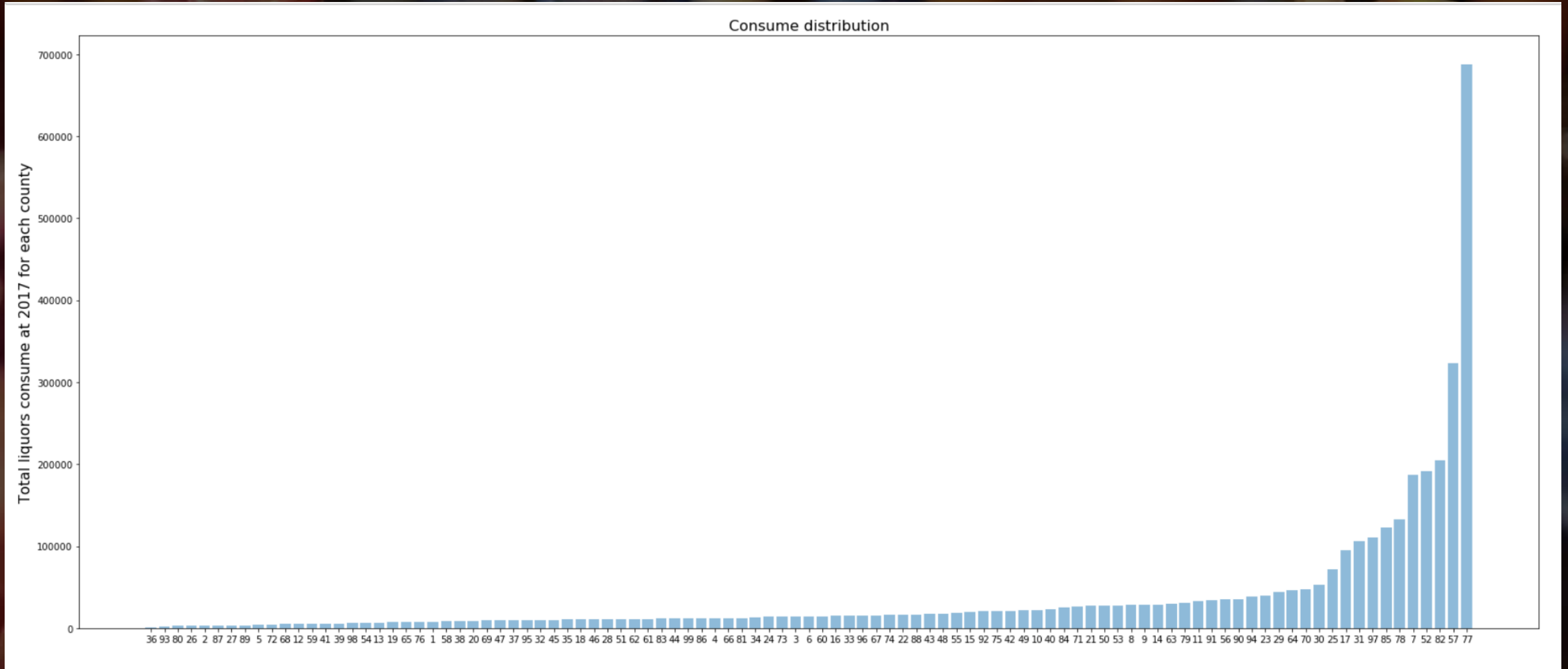
Total volume sold and Category analyzing:





# Classification :

The consume distribution for each county:



# Knowledge gained :

- The American Vodkas is the most popular category of alcoholic beverages in 2017
- because it gains the maximum volume sold and maximum revenue.
- The American Vodkas and Canadian Whiskies have the pretty same volume sold, but they have a wide margin of revenue.
- High Volume sold doesn't mean high revenue.
- The Brand which gains the maximum revenue in 2017 is Black Velvet Which is Canadian Whiskies.
- As for Diageo America Company, the county POLK brought the maximum revenue.
- The Whisker Liqueur is the most popular category of alcoholic beverages in county POTTAWATTA.
- The people in POLK consumed the most alcoholic beverages.



# How that knowledge can be applied:

## Help vender to get more profits:

- Let vender know people in which region consumed more alcoholic beverages then vender can increase the supply for that region.
- Let vender know which region brought less revenue then vender can increase the advertising to attract more users.
- Let vender know which alcoholic beverage can brought more revenue then vender can extend the production.
- Let vender know which alcoholic beverage gains more volume sold then vender can make a good market planning to get more profits.

## Help customers to choose better alcoholic beverage :

- We can let customer know which alcoholic beverage is the most popular to help them to make a right choose.