

The Sales Situation of Liquors in Different Regions in Iowa 2017

CSCI 4502 Final Project Presentation



Team Members

Lei Teng

Mingxuan Zhang

Yuxiang Wang

Questions sought to answer



1.Sale situation:

- **How can we classify the category attribute if know the store number and vendor number?**
- As for each vender, which regions gains the maximum revenue?
- Which category of alcoholic beverages is the most popular in different regions?
- Which category of alcoholic beverages gains the maximum revenue ?
- Which brand of alcoholic beverages gains the maximum revenue?

2.Drinking Behavior:

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

Data Preparation Work



Data Cleaning

We delete the rows which have missing values, in order to build an accurate Bayesian Classification Model, the missing values may make some negative effects to the veracity for our model.

Data Reduction

- The original datasets contains from 2012 to current, its over 12 millions while our project is working for 2017. Thus, we did the reduction only left datasets about 2017.
- We did the dimensionality reduction. We remove irrelevant attributes like 'Invoice/Item Number'. We also remove redundant attributes. For example, Volume Sold (Liters) and Volume Sold (Gallons). We can use Volume Sold(Liters) instead of Volume Sold (Gallons).

Data Transformation

- "Sales numbers" contains "\$" which causes panda to fail to identify them as numbers. So we deleted the "\$" and transform it from type string to type float.

4	S29050301	11/16/2015	3549	Quicker Li 1414 48TH FORT MAJ 52627	1414 48TH	56	Lee	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	2	\$19.20	0.30	0.08		
5	S28867701	11/04/2015	2513	Hy-Vee Fo 812 S 15TH IOWA CITY 52240	812 S 15TH	52	Johnson	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39		
6	S29050801	11/17/2015	3942	Twin Town 104 HIGHV TOLEDO 52342	104 HIGHV	86	Tama	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	2	\$19.20	0.30	0.08		
7	S28869201	11/11/2015	3650	Spirits, Str 118 South HOLSTEIN 51025	118 South	47	Ida	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	1	\$53.34	1.75	0.46		
8	S28865701	11/09/2015	2538	Hy-Vee Fo 1422 FLAM WATERLC 50702	1422 FLAM	07	Black Haw	1701100	DECANTE	962	Duggan's	1238	Forbidden	6	1500	\$11.62	\$17.43	6	\$104.58	9.00	2.38
9	S28869501	11/10/2015	3942	Twin Town 104 HIGHV TOLEDO 52342	104 HIGHV	86	Tama	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	2	\$106.68	3.50	0.92		
10	S29339301	11/30/2015	2662	Hy-Vee Wi 5222 MULB MUSCATIN 52761	5222 MULB	70	Muscataine	1701100	DECANTE	65	Jim Beam	173	Laphroaig	12	750	\$19.58	\$29.37	4	\$117.48	3.00	0.79
11	S29050901	11/16/2015	4307	Crossroad: 117 IOWA DUNLAP 712-2	117 IOWA	43	Harrison	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	2	\$19.20	0.30	0.08		
12	S29049901	11/17/2015	2661	Hy-Vee Fo 1989 PARI SHELTON 51201	1989 PARI	71	O'Brien	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79		
13	S28868201	11/05/2015	2561	Hy-Vee Fo 4605 FLEDES MOIN 50321	4605 FLEJ	77	Polk	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	2	\$106.68	3.50	0.92		
14	S28869601	11/09/2015	4114	After 5 Sor 704 W 7TH ATLANTIC 50022	704 W 7TH	15	Cass	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39		
15	S28866901	11/11/2015	3650	Spirits, Str 118 South HOLSTEIN 51025	118 South	47	Ida	1701100	DECANTE	962	Duggan's	1238	Forbidden	6	1500	\$11.62	\$17.43	1	\$17.43	1.50	0.40
16	S29050101	11/19/2015	2806	Oscos #881307 N SE CLINTON 52732	1200 N SE	23	Clinton	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79		
17	S29049601	11/17/2015	2624	Hy-Vee #2 2395 NW DUBUQUE 52002	2395 NW	31	Dubuque	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	Black Haw	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39		
19	S29196301	11/24/2015	2595	Hy-Vee Wi 1620 4TH DENISON 51442	1620 4TH	24	Crawford	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	6	\$162.84	4.50	1.19		
20	S29134301	11/18/2015	3723	J D Spirits 1023 9TH ONAWA 51040	1023 9TH	67	Monona	1081200	CREAM LI	305	MHW Ltd	258	Rumchata	1	6000	\$99.00	\$148.50	1	\$148.50	6.00	1.59
21	S28869001	11/10/2015	2665	Hy-Vee / V 1005 E HK WAUKEE 50263	1005 E HK	25	Dallas	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	2	\$106.68	3.50	0.92		
22	S29198701	11/24/2015	5093	Cody Mart 1220 N CCLE CLAIR 52753	1220 N CC	82	Scott	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	3	\$81.42	2.25	0.59		
23	S29282801	11/23/2015	2642	Hy-Vee Wi 512 E OSI PELLA 50219	512 E OSI	63	Marion	1701100	DECANTE	962	Duggan's	1238	Forbidden	6	1500	\$11.62	\$17.43	6	\$104.58	9.00	2.38
24	S28868001	11/04/2015	2548	Hy-Vee Fo 100 8TH S ALTOONA 50009	100 8TH S	77	Polk	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39		
25	S29199601	11/20/2015	2558	Hy-Vee Fo 1700 E W/MOUNT PI 52641	1700 E W	44	Henry	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	18	\$488.52	13.50	3.57		
26	S29050501	11/18/2015	3735	C B Liquor 1202 A AVIOSKALOC 52577	1202 A AV	10	Buchanan	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	3	\$28.80	0.45	0.12		
27	S28867001	11/04/2015	3842	Bancroft Li 107 N POF BANCROF 50517	107 N POF	55	Kossuth	1701100	DECANTE	962	Duggan's	1238	Forbidden	6	1500	\$11.62	\$17.43	3	\$52.29	4.50	1.19
28	S28868601	11/09/2015	2650	Hy-Vee Wi 1808 23RD HARLAN 51537	1808 23RD	83	Shelby	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39		
29	S28869101	11/10/2015	2666	Hy-Vee #2 2510 SW ANKENY 50023	2510 SW	77	Polk	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 POF BANCROF 50517	107 N POF	55	Kossuth	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	18	\$488.52	13.50	3.57		
31	S28865801	11/09/2015	2539	Hy-Vee Fo HIGHWAY IOWA FAL 50126	HIGHWAY	42	Hardin	1701100	DECANTE	962	Duggan's	1238	Forbidden	6	1500	\$11.62	\$17.43	6	\$104.58	9.00	2.38
32	S28868701	11/10/2015	2651	Hy-Vee / V 1311 4 ST WAYERLY 50677	1311 4 ST	09	Bremer	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	2	\$106.68	3.50	0.92		
33	S29050001	11/17/2015	2666	Hy-Vee #2 2510 SW ANKENY 50023	2510 SW	77	Polk	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79		
34	S29198101	11/23/2015	4162	Fareway S4220 16TH CEDAR R 52404	4220 16TH	57	Linn	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	6	\$162.84	4.50	1.19		
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	90	\$2442.60	67.50	17.83		
36	S29197801	11/23/2015	3650	Spirits, Str 118 South HOLSTEIN 51025	118 South	47	Ida	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	18	\$488.52	13.50	3.57		
37	S29049101	11/16/2015	2538	Hy-Vee Fo 1422 FLAM WATERLC 50702	1422 FLAM	07	Black Haw	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	40	\$384.00	6.00	1.59		
38	S29049501	11/18/2015	2594	Hy-Vee Fo 4500 SER SIOUX CIT 51106	4500 SER	97	Woodbury	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79		
39	S29051001	11/19/2015	4988	Happy's W 5925 UNIV CEDAR FA 50613	5925 UNIV	07	Black Haw	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	Jasper	1701100	DECANTE	962	Duggan's	1238	Forbidden	6	1500	\$11.62	\$17.43	2	\$34.86	3.00	0.79
41	S29049801	11/16/2015	2643	Hy-Vee Wi 2126 KIME WATERLC 50701	2126 KIME	07	Black Haw	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	20	\$192.00	3.00	0.79		
42	S29049201	11/17/2015	2544	Hy-Vee Fo 802 SOUT MARSHAL 50158	802 SOUT	64	Marshall	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 5250 SO FF SHENANC 51601	5250 SO FF	73	Page	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	30	\$814.20	22.50	5.94		
44	S29191201	11/19/2015	2248	Ingersoll Li 3500 INGEDES MOIN 50312	3500 INGE	77	Polk	1701100	DECANTE	65	Jim Beam	173	Laphroaig	12	750	\$19.58	\$29.37	36	\$1057.32	27.00	7.13
45	S29050601	11/17/2015	3813	CGI Foods 104 NORT MOUNT A 50854	104 NORT	80	Ringgold	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	6	\$57.60	0.90	0.24		
46	S29199301	11/20/2015	2539	Hy-Vee Fo HIGHWAY IOWA FAL 50126	HIGHWAY	42	Hardin	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	12	\$325.68	9.00	2.38		
47	S29197301	11/21/2015	3443	Super Sav 1141 N BR COUNCIL 51503	1141 N BR	78	Pottawatt	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	6	\$162.84	4.50	1.19		
48	S28868901	11/10/2015	2661	Hy-Vee Fo 1989 PARI SHELTON 51201	1989 PARI	71	O'Brien	65	Jim Beam	237	Knob Cree	3	1750	\$35.55	\$53.34	3	\$160.02	5.25	1.39		
49	S29197501	11/20/2015	3525	Wines and 106 W 2ND WASHINGTON 52353	106 W 2ND	92	Washington	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	6	\$162.84	4.50	1.19		
50	S29049301	11/19/2015	2567	Hy-Vee Dr 2200 WESDAVENPC 52806	2200 WES	82	Scott	130	Disaronno	249	Disaronno	20	150	\$6.40	\$9.60	6	\$57.60	0.90	0.24		
51	S29196201	11/20/2015	3549	Hy-Vee Wi 1602 E 7TH ATLANTIC 50022	1602 E	71	Cass	255	Wilson Da	297	Templeton	6	750	\$18.09	\$27.14	24	\$651.36	18.00	4.76		
52	S29137201	11/18/2015	2566	Hy-Vee Fo 813 N LINK KNOXVILL 50138	813 N LINK	63	Marion	1701100	DECANTE	962	Duggan's	1238	Forbidden	6	1500	\$11.62	\$17.43	12	\$209.16	18.00	4.76

ID	Date		Number	Number	Quantity	Category	Name	Number	Name	Number	Description	Units	Volume (ml)	Cost	Retail	Sold
	Start	End														
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	Skyv Spirits	37987	Skvv Vodka	12	1000	\$12.35	18.53	

List of Tools

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





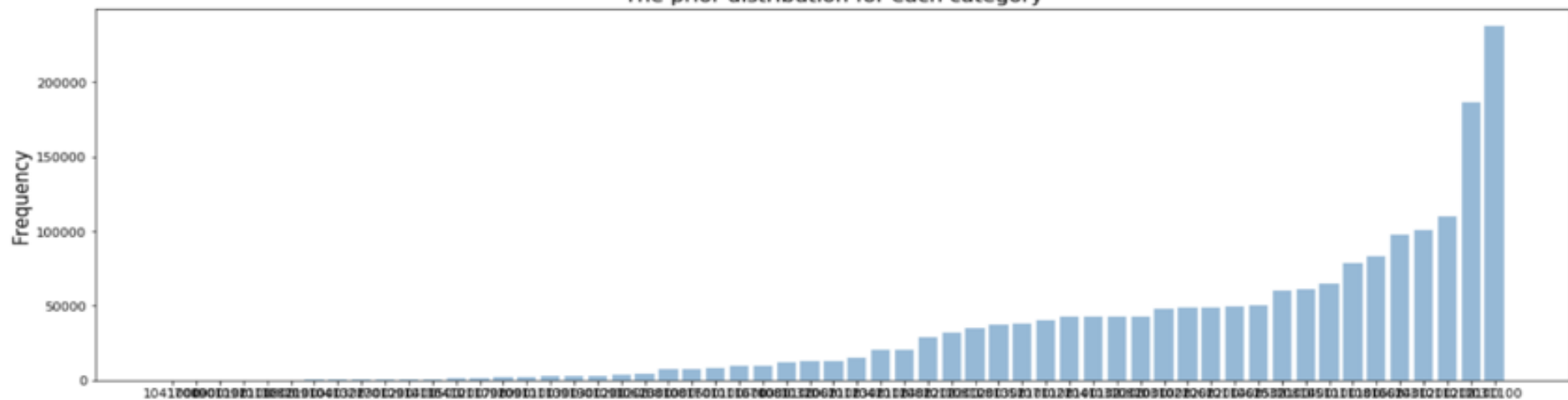
Classification: Bayesian Classification

We built a model:

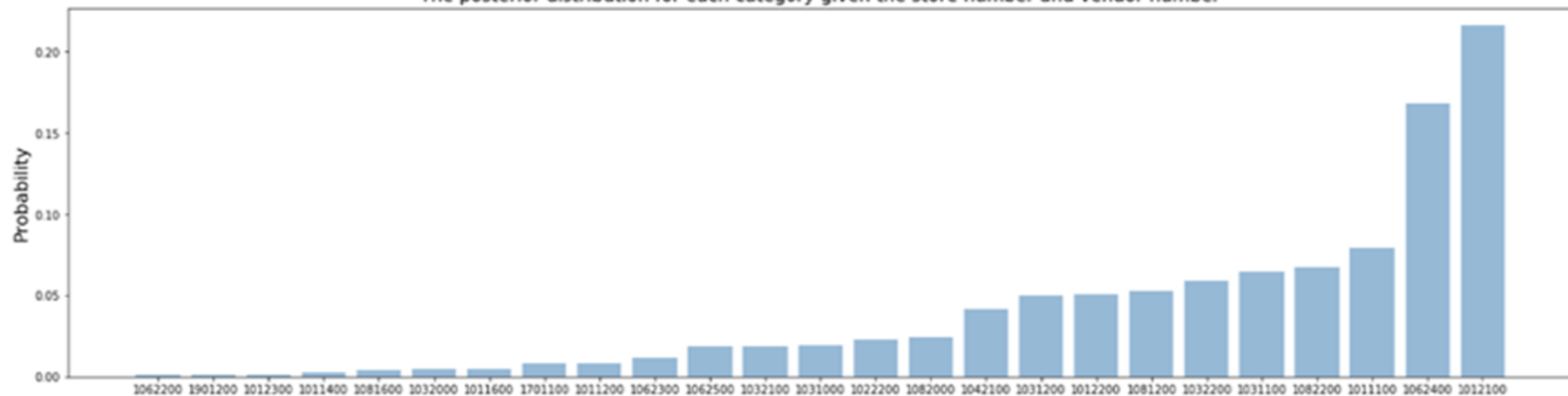
$$P(\text{category} \mid \text{store number, vendor number})$$

If we have the store number and vendor number, we can get the posterior probability distribution for the categories. We used this to classify the category attribute, we can pick the one with highest probability or we can use this posterior distribution to draw samples for category attribute.

The prior distribution for each category



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

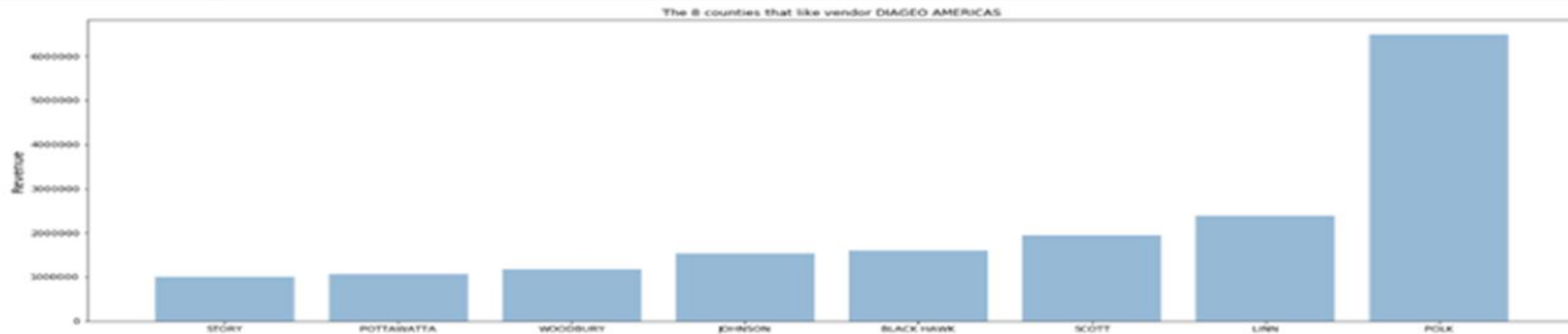


Data Analysis

- 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'])[0])
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)
```



Data Analysis

- 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)
```

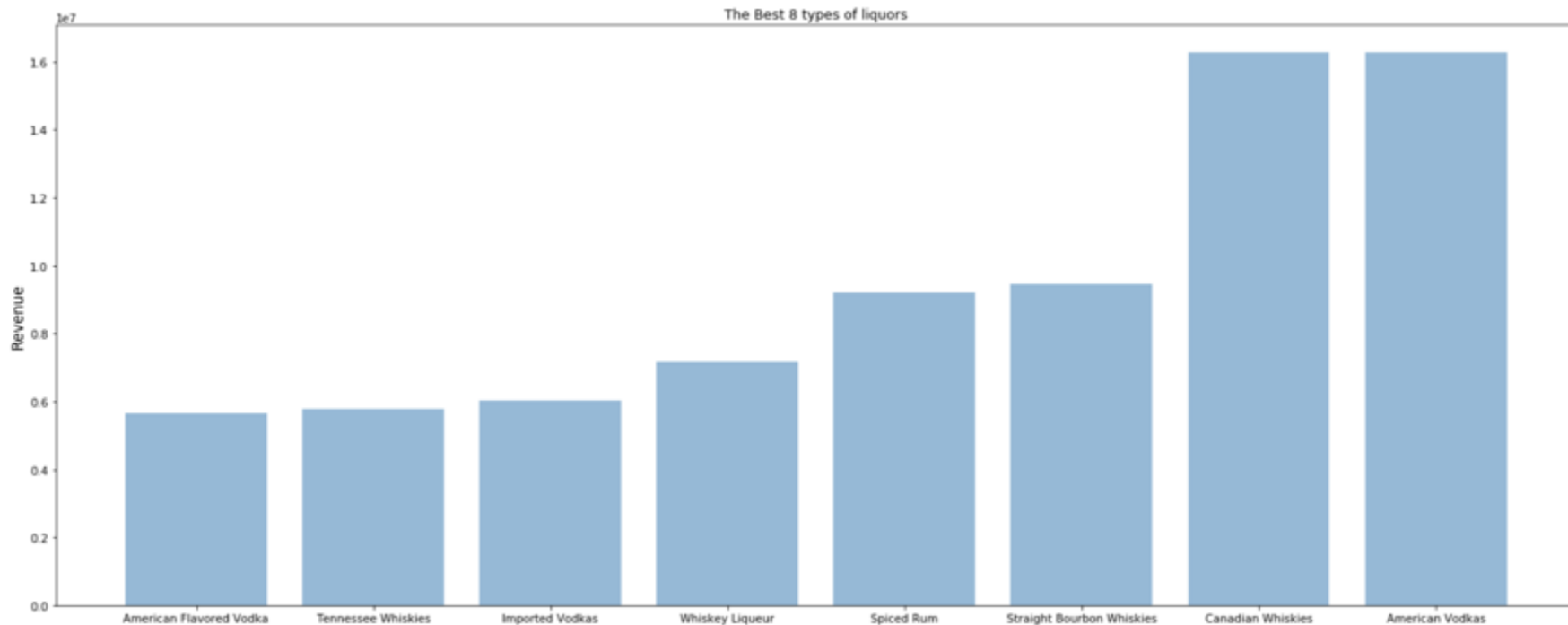




Some others Analysis

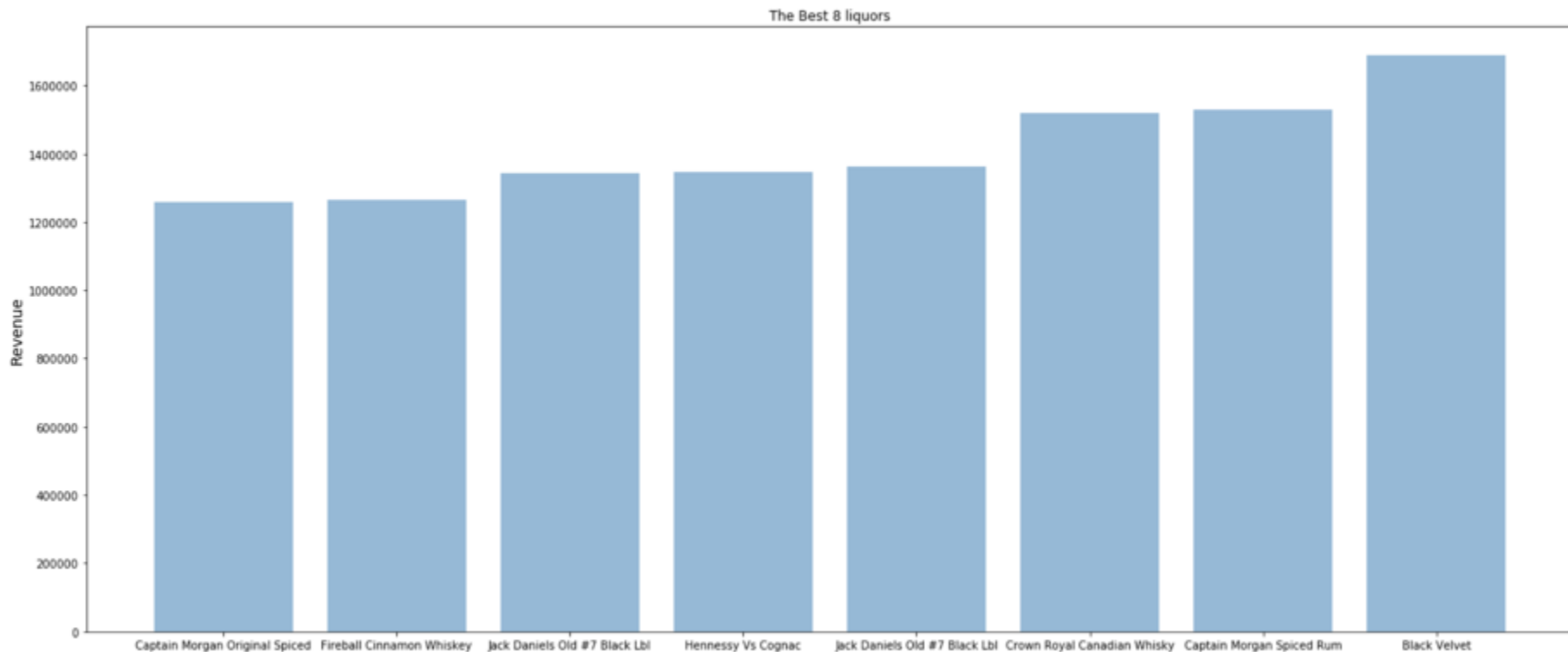


Revenue and Category analyzing



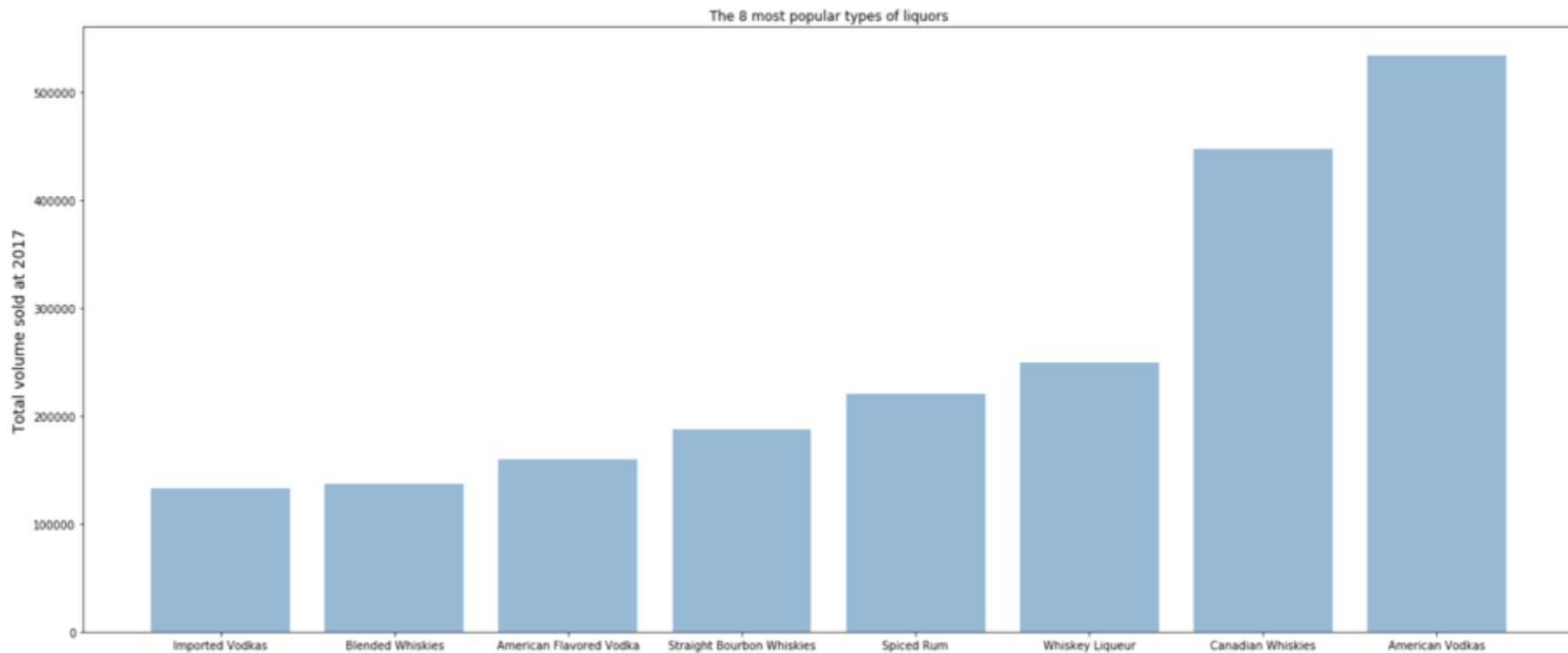


Revenue and Brand analyzing



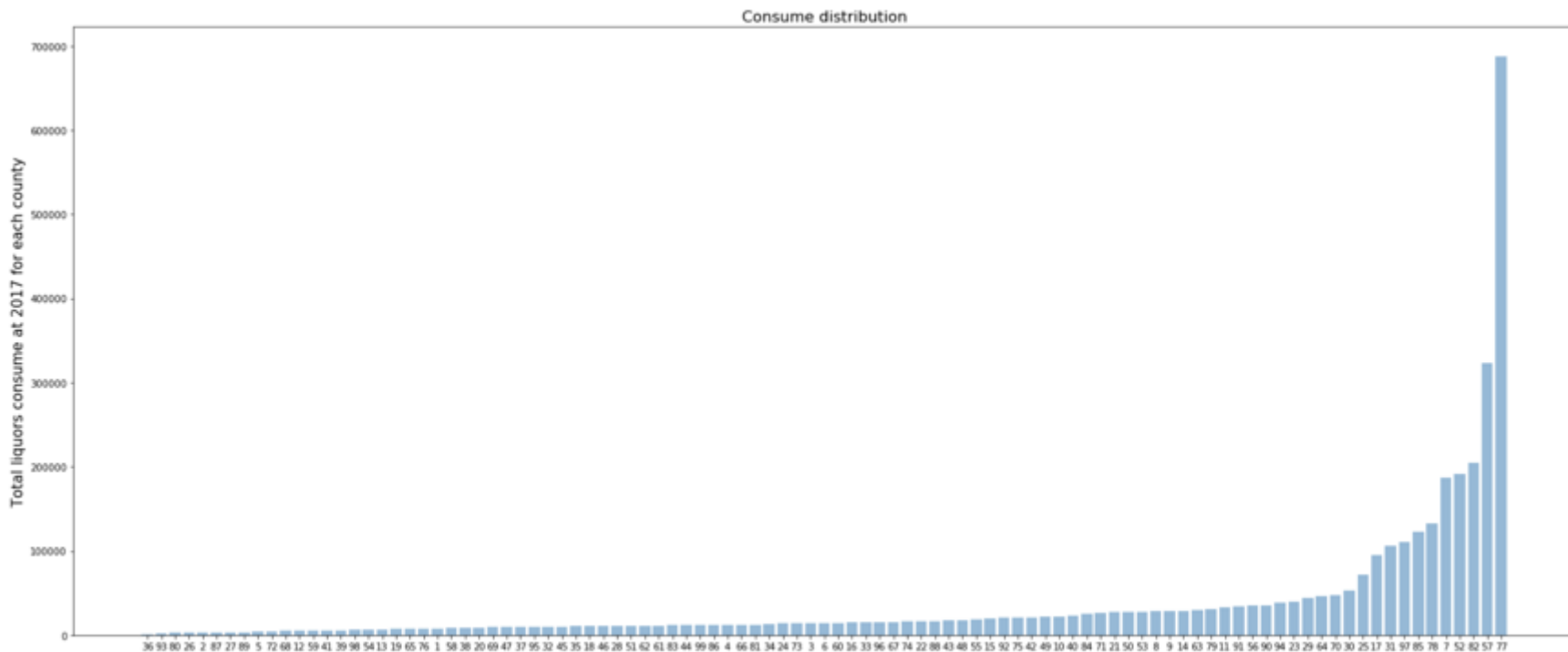


Total volume sold and Category analyzing






The consume distribution for each county



Knowledge Gained

- 
- The American Vodkas is the most popular category of alcoholic beverages in 2017, 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 the Black Velvet. Which is a 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.
 - People in POLK consumed the most alcoholic beverages.



How that knowledge can be applied

- **Help vender to get more profits**
- **Help customers to choose better alcoholic beverage**



Some other Applications

- Use the Bayesian classification to predict the missing value (such as those alcoholic beverage which has no category label)
- Use the Bayesian classification to classify the category attribute after 2017(like 2018)