

# NỘI DUNG CHÍNH

**1. GIỚI  
THIỆU VỀ  
IOWA**

**2.  
DATASET**

**3. TIỀN XỬ  
LÝ DỮ LIỆU**

**4. PHÂN  
TÍCH**

**5. KẾT  
LUẬN & ĐỀ  
XUẤT**





# 1. GIỚI THIỆU

**Iowa (phát âm là Ai-ô-va) là một tiểu bang thuộc miền Trung Tây Hoa Kỳ.**

## Iowa giáp với Minnesota về phía bắc

## Nebraska và Nam Dakota về phía tây

## Missouri về phía nam

## Wisconsin và Illinois về phía đông





## 2. DATASET

	invoice_and_item_number	date	store_number	store_name	address	city	zip_code	store_location	county_number	county	...	item_number	item_description	pack	bottle_volume_ml	state_bottle_cost	state_bottle_retail	bottles_sold	sale_dollars	volume_sold_liters	volume_sold_gallons
0	INV-37504900003	2021-06-15	5645	Phillips 66 / Grinnell	1031 West St	Grinnell	50112.0	POINT (-92.727379 41.745903)	79.0	POWESHIEK	...	38178	Titos Handmade Vodka	6	1750	19.00	28.50	12	342.00	21.00	5.54
1	INV-35039100017	2021-03-16	3488	Bender's Foods	400 S Second	Bellevue	52031.0	POINT (-90.42292 42.255449)	49.0	JACKSON	...	43337	Captain Morgan Original Spiced	12	1000	11.75	17.63	36	634.68	36.00	9.51
2	INV-37617900100	2021-06-18	2649	Hy-Vee #3 / Dubuque	400 Locust St	Dubuque	52001.0	POINT (-90.666497 42.497219000000001)	31.0	DUBUQUE	...	64529	Bumbu Rum	6	750	18.50	27.75	1	27.75	0.75	0.19
3	INV-36572700003	2021-05-12	3682	Jumbo's	110 E Main St	Wesley	50483.0	POINT (-93.202775000000002 43.286886)	55.0	KOSSUTH	...	25608	Seagrams 7 Crown	6	1750	12.96	19.44	6	116.64	10.50	2.77
4	INV-37602100172	2021-06-17	2621	Hy-Vee Food Store #3 / Sioux City	3301 Gordon Dr	Sioux City	51105.0	POINT (-96.364108 42.487752)	97.0	WOODBURY	...	86873	Southern Comfort 80prf	12	750	11.50	17.25	6	103.50	4.50	1.18
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
99995	INV-38877000020	2021-08-03	5909	EZ Stop 1 / Dubuque	1101 Rhomberg Ave	Dubuque	52001.0	POINT (-90.660269 42.519502)	31.0	DUBUQUE	...	65256	Jagermeister Liqueur	12	750	13.00	19.50	5	97.50	3.75	0.99
99996	INV-34702500010	2021-03-02	5863	K-Zar Inc - Waterloo	1761 Independence Avenue	Waterloo	50707.0	POINT (-92.302514 42.498404)	7.0	Black Hawk	...	48105	Hennessy VS	12	375	11.49	17.24	12	206.88	4.50	1.18
99997	INV-41403100015	2021-10-27	2545	Hy-Vee Drugstore / Iowa City	310 N 1st St	Iowa City	52245.0	NaN	52.0	JOHNSON	...	18196	Ezra Brooks Sour Mash Bourbon	12	750	8.00	12.00	12	144.00	9.00	2.37
99998	INV-41391900038	2021-10-27	6154	Off Campus Beer and Spirits / Ames	4518 Mortensen Rd STE 109	Ames	50014.0	POINT (-93.679385 42.012137)	85.0	STORY	...	39300	Revelton Honey Vodka	6	750	13.50	20.25	2	40.50	1.50	0.39
99999	INV-34285100011	2021-02-15	4594	Quik Trip #554 / SW 63rd DM	901 SW 63rd St	Des Moines	50312.0	POINT (-93.703165 41.569715)	77.0	POLK	...	26824	Jack Daniels Old #7 Black Label Square	24	375	7.50	11.25	6	67.50	2.25	0.59
100000 rows x 24 columns																					

### 2.1 TỔNG QUAN VỀ DATASET

Bộ dữ liệu bán hàng rượu của Bang Iowa là tập hợp dữ liệu bao gồm 24 cột thể hiện tình hình đặt hàng của các cửa hàng tại khu vực trong thời gian từ tháng 01/2021 đến tháng 01/2022.



## 2. DATASET

### 2.2. Các cột data chính xây dựng mô hình RFM

1 - date: Ngày đặt hàng

2 - store\_number: Mã số cửa hàng đặt hàng

3 - store\_name: Tên cửa hàng đặt đơn

4 - invoice\_and\_item\_number: Mã hóa đơn

5 - bottles\_sold: Số chai rượu cửa hàng đặt

6 - sale\_dollars: Tổng chi phí đặt hàng rượu (số chai nhân với số chai bán lẻ của nhà nước)

#	Column	Non-Null Count	Dtype
0	invoice_and_item_number	100000 non-null	object
1	date	100000 non-null	object
2	store_number	100000 non-null	int64
3	store_name	100000 non-null	object
4	address	99995 non-null	object
5	city	99995 non-null	object
6	zip_code	99995 non-null	float64
7	store_location	88226 non-null	object
8	county_number	99995 non-null	float64
9	county	99995 non-null	object
10	category	100000 non-null	int64
11	category_name	100000 non-null	object
12	vendor_number	100000 non-null	int64
13	vendor_name	100000 non-null	object
14	item_number	100000 non-null	int64
15	item_description	100000 non-null	object
16	pack	100000 non-null	int64
17	bottle_volume_ml	100000 non-null	int64
18	state_bottle_cost	100000 non-null	float64
19	state_bottle_retail	100000 non-null	float64
20	bottles_sold	100000 non-null	int64
21	sale_dollars	100000 non-null	float64
22	volume_sold_liters	100000 non-null	float64
23	volume_sold_gallons	100000 non-null	float64

dtypes: float64(7), int64(7), object(10)



## 2.3 Business Understanding

### Mục tiêu làm rõ các câu hỏi sau

- 1, Xu hướng số lượng chai rượu được bán ra từ tháng 1 năm 2021 đến tháng 1 năm 2022 như thế nào?
- 2, Iowa kiếm được bao nhiêu doanh thu từ hoạt động kinh doanh mỗi tháng? Loại rượu nào thúc đẩy doanh số bán hàng nhiều nhất?
- 3, Có bao nhiêu cửa hàng đã đặt hàng rượu? Cửa hàng nào có đơn đặt hàng nhiều nhất?
- 4, Những cửa hàng nào có khả năng rời bỏ ?
- 5, Các loại rượu mới nên bán cho tệp khách hàng nào trước ?
- 6, Có sự khác biệt nào giữa các khu vực về loại rượu phổ biến nhất không?



# 3. TIỀN XỬ LÝ DỮ LIỆU

## CÔNG CỤ SỬ DỤNG:

3.1 - Xây dựng mô hình RFM bằng SQL

3.2 - Sử dụng Power BI tạo các biểu đồ so sánh và phân tích





# 3.1 Xây dựng mô hình RFM bằng SQL

```
79
80 -- Xây dựng mô hình RFM phân khúc khách hàng
81 with A as(
82     select store_number,
83            invoice_and_item_number,
84
85            [date],
86            sale_dollars
87 from [Iowa Liquor])
88 --select* from A
89 , RFM_Base as(
90     select store_number,
91            datediff(day,Max([date]), GETDATE() ) as Recency_value,
92            count(distinct [date] ) as Frequency_value,
93            round(sum(sale_dollars),2) as Monetary_value
94     from A
95     group by store_number
96 )
97 --select * from RFM_Base
98 ,RFM_Score
99 AS
100 (
101     SELECT *,
102            NTILE(5) OVER (ORDER BY Recency_Value DESC) as R_Score,
103            NTILE(5) OVER (ORDER BY Frequency_Value ASC) as F_Score,
104            NTILE(5) OVER (ORDER BY Monetary_Value ASC) as M_Score
105     FROM RFM_Base
106 )
107 --SELECT * FROM RFM_Score
108 , RFM_Final
109 AS
110 (
111     SELECT *,
112            Cast(CONCAT(R_Score, F_Score, M_Score) as int) as RFM_Overall
113     from RFM_Score)
114 SELECT a.* , b.segment
115 FROM RFM_Final a join segment_scores b
116 on a.RFM_Overall=b.scores
117
```

## 1. Tính toán các chỉ số R, F, M của khách hàng

- R: Tính số ngày kể từ ngày cuối cùng khách hàng mua hàng đến hiện tại
- F: Số lần các cửa hàng đã đặt hàng
- M: Tổng số tiền khách hàng đã bỏ ra để mua hàng

## 2. Tính điểm và chia rank các biến

## 3. Phân loại khách hàng theo bảng Segment Score (bảng bao gồm 11 phân khúc khách hàng)



# Bảng Segment Score

Segment	Activity	Scores
Champion	Đã mua gần đây, đặt hàng thường xuyên và chi tiêu nhiều nhất.	555, 554, 544, 545, 454, 455, 445
Loyal	Đặt hàng thường xuyên. Có tương tác các chương trình khuyến mãi.	543, 444, 435, 355, 354, 345, 344, 335
Potential Loyalist	Đã mua gần đây, chi tiêu tốt.	553, 551, 552, 541, 542, 533, 532, 531, 452, 451, 442, 441, 431, 453, 433, 432, 423, 353, 352, 351, 342, 341, 333, 323
Promising	Từng là khách trung thành. Chi tiêu thường xuyên và khá nhiều tiền. Nhưng lần mua cuối cùng đã cách đây vài tuần.	525, 524, 523, 522, 521, 515, 514, 513, 425,424, 413,414,415, 315, 314, 313
New Customers	Khách hàng mới	512, 511, 422, 421, 412, 411, 311
Needs attention	Khách hàng cốt lõi có lần mua gần đây nhất đã xảy ra hơn một tháng trước.	535, 534, 443, 434, 343, 334, 325, 324
About to sleep	Khách phổ biến, mua hàng cách đây không lâu.	331, 321, 312, 221, 213, 231, 241, 251
At Risk	Giống như “Can’t lose them but losing” nhưng lượng tiền và tần suất thấp hơn	255, 254, 245, 244, 253, 252, 243, 242, 235, 234, 225, 224, 153, 152, 145, 143, 142, 135, 134, 133, 125, 124
Can’t lose them	Thực hiện các đơn đặt hàng lớn nhất và thường xuyên. Nhưng đã lâu không trở lại.	155, 154, 144, 214,215,115, 114, 113
Hibernating customers	Đã thực hiện lần mua hàng cuối cùng của họ cách đây khá lâu nhưng trong 4 tuần qua, họ đã truy cập trang web hoặc mở email.	332, 322, 233, 232, 223, 222, 132, 123, 122, 212, 211
Lost customers	Đã thực hiện lần mua hàng gần đây nhất cách đây khá lâu và hoàn toàn không tương tác trong 4 tuần qua.	111, 112, 121, 131,141,151



## 3.2 Data Prepare with Power BI

### TẠO BẢNG CALENDER

- Sử dụng Dax trong PBI tạo bảng Calendar:

**Calendar=**  
**CALENDAR(min('Iowa Liquor'[date]),max('Iowa Liquor'[date]))**

- Tạo thêm các cột: Month, Quarter, Year

### TẠO BẢNG RFM

- Import data từ SQL Server vào Power BI.
- Lấy data bao gồm các cột chứa giá trị R, F, M và phân chia khách hàng theo 11 phân khúc với số điểm chia rank tương ứng của từng khách hàng

### TÍNH TOÁN THÊM GIÁ TRỊ PROFIT

- Tính thêm giá trị Profit  
**Profit = 'Iowa Liquor'[sale\_dollars] - 'Iowa Liquor'[state\_bottle\_cost] \* 'Iowa Liquor'[bottles\_sold]**
- Tính biên lợi nhuận gộp trên doanh thu (Gross Margin)

### TÍNH TOÁN GIÁ TRỊ TRUNG BÌNH R, F, M

Tính thêm giá trị trung bình của các chỉ số R, F, M theo từng phân khúc khách hàng với mục tiêu so sánh đánh giá





## 4. PHÂN TÍCH

---

### 4.1 TỔNG QUAN



Từ tháng 01/2021 đến tháng 01/2022

- Tổng doanh thu khoảng \$ 168.84M
- Tổng lợi nhuận khoảng \$ 157.02M
- Tổng số lượng chai bán ra thị trường khoảng 12M chai.  
Trong đó có 1927 cửa hàng đặt mua với khoảng hơn 1M hoá đơn



Num of Store

1.927K



Num of sold bottles

12M



Total Sales

168.84M



Total Profit

157.02M

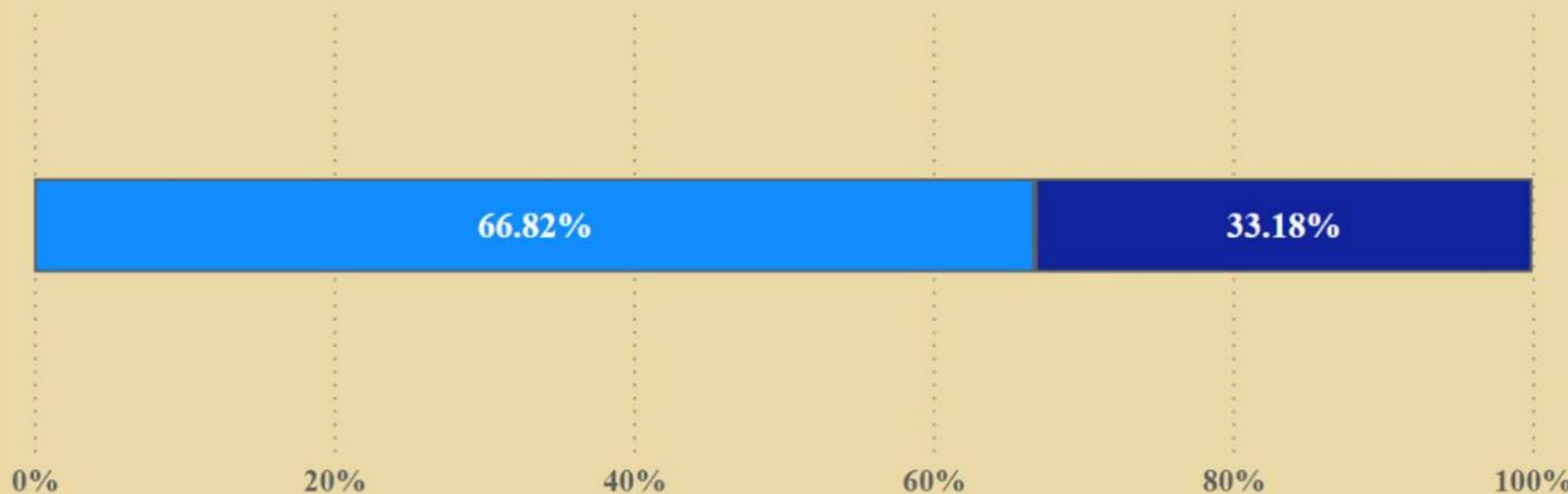


Num of Invoice

1.048575M

### % Tổng doanh thu và Biên lợi nhuận gộp

● Tổng doanh thu ● Biên lợi nhuận gộp



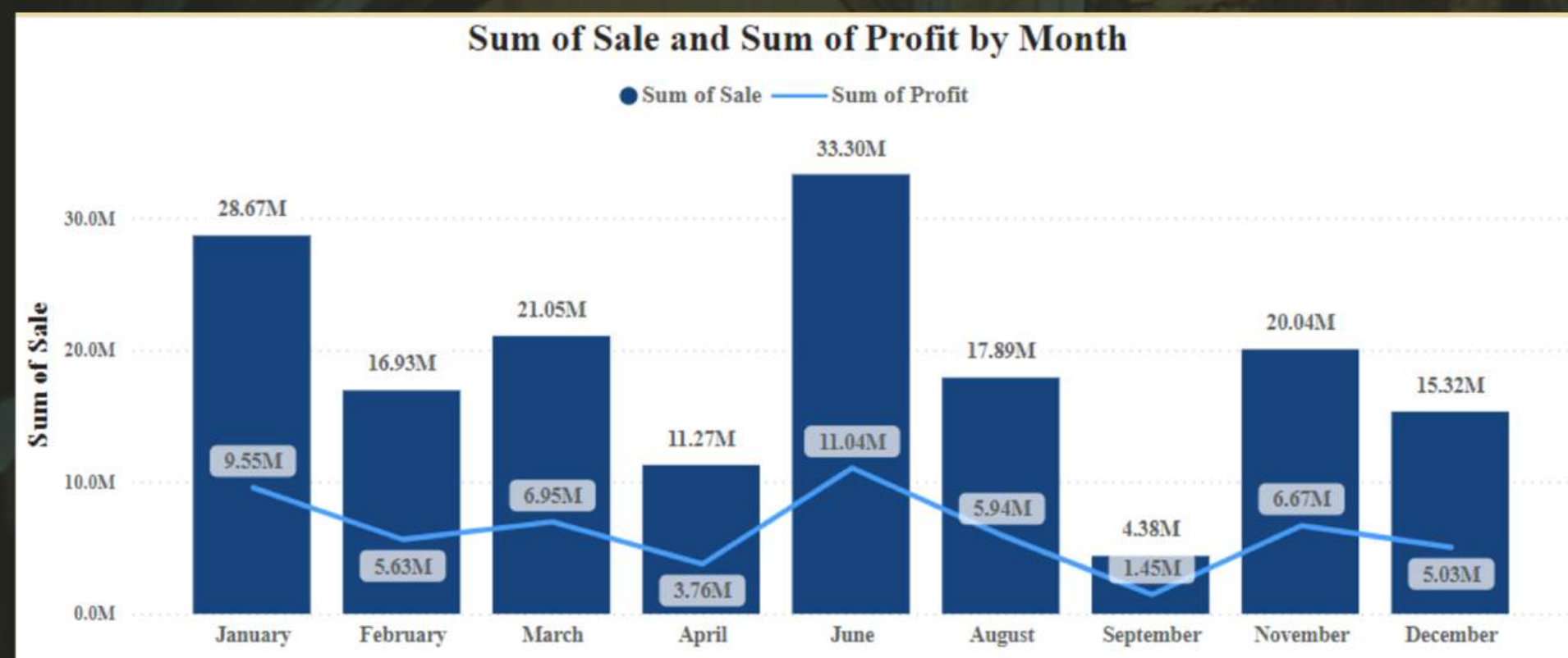
- Tỷ suất Biên lợi nhuận gộp trên doanh thu khoảng 33.18%. Chỉ số này có ý nghĩa tích cực khi so với trung bình ngành hàng tại khu vực (khoảng 25%)



## 4.2 Xu hướng bán hàng theo thời gian



Tình hình tiêu thụ rượu tại khu vực Iowa thấp nhất vào tháng 9 chỉ với số lượng bán ra khoảng hơn 330 ngàn chai. Xu hướng số lượng rượu bán ra trong thời gian này không ổn định. Nguyên nhân xu hướng tăng giảm liên tục có thể do năm 2022 tình hình dịch Covid bùng nổ tuy lượng tiêu thụ tăng nhưng nguồn cung không đáp ứng đủ do các chính sách kiểm soát dịch của chính phủ gây ra.

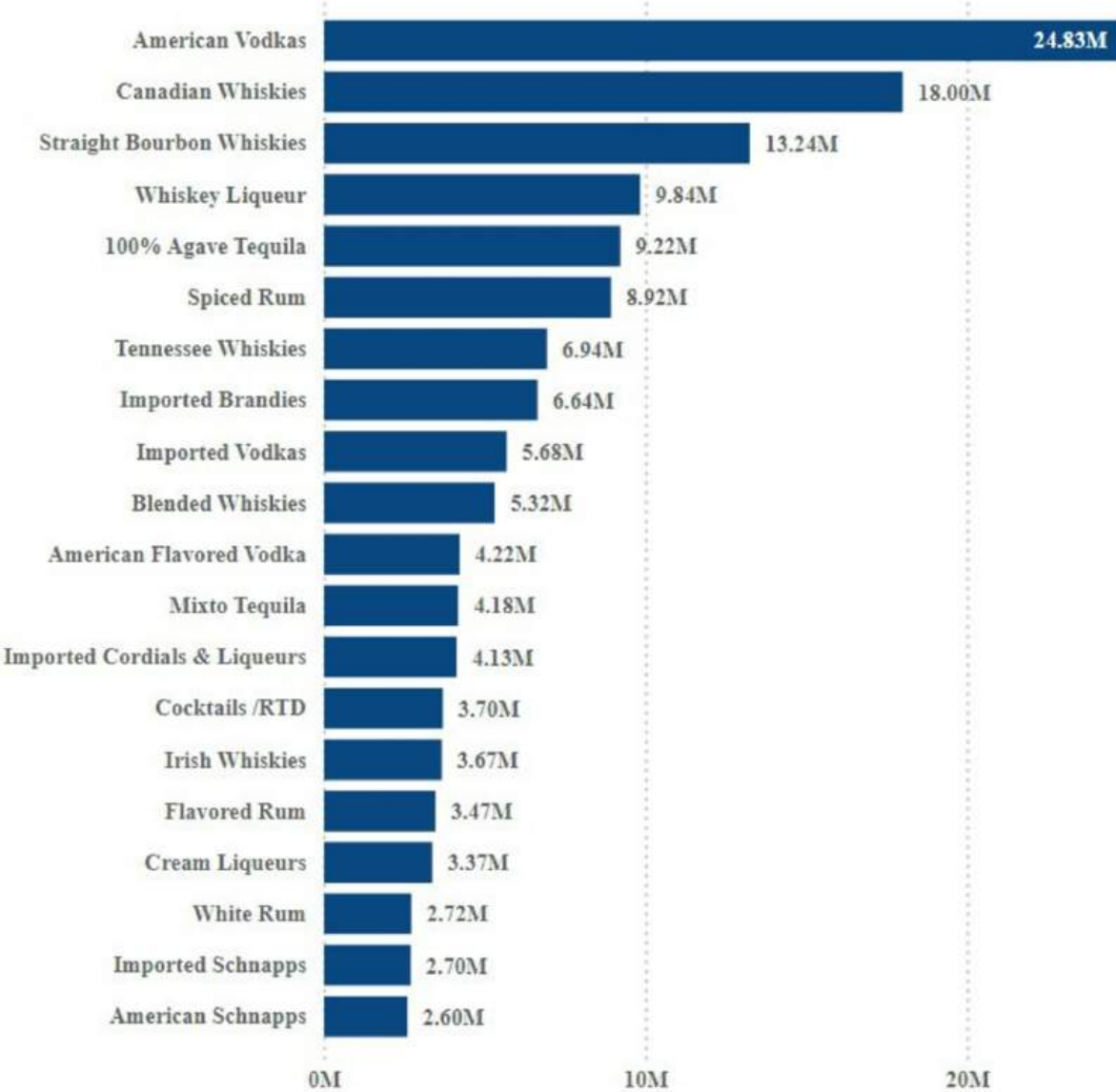


Số lượng rượu bán ra đã ảnh hưởng trực tiếp đến doanh thu và lợi nhuận của Iowa. Vì vậy có thể thấy doanh thu các tháng có xu hướng không đồng đều, tăng giảm liên tục.



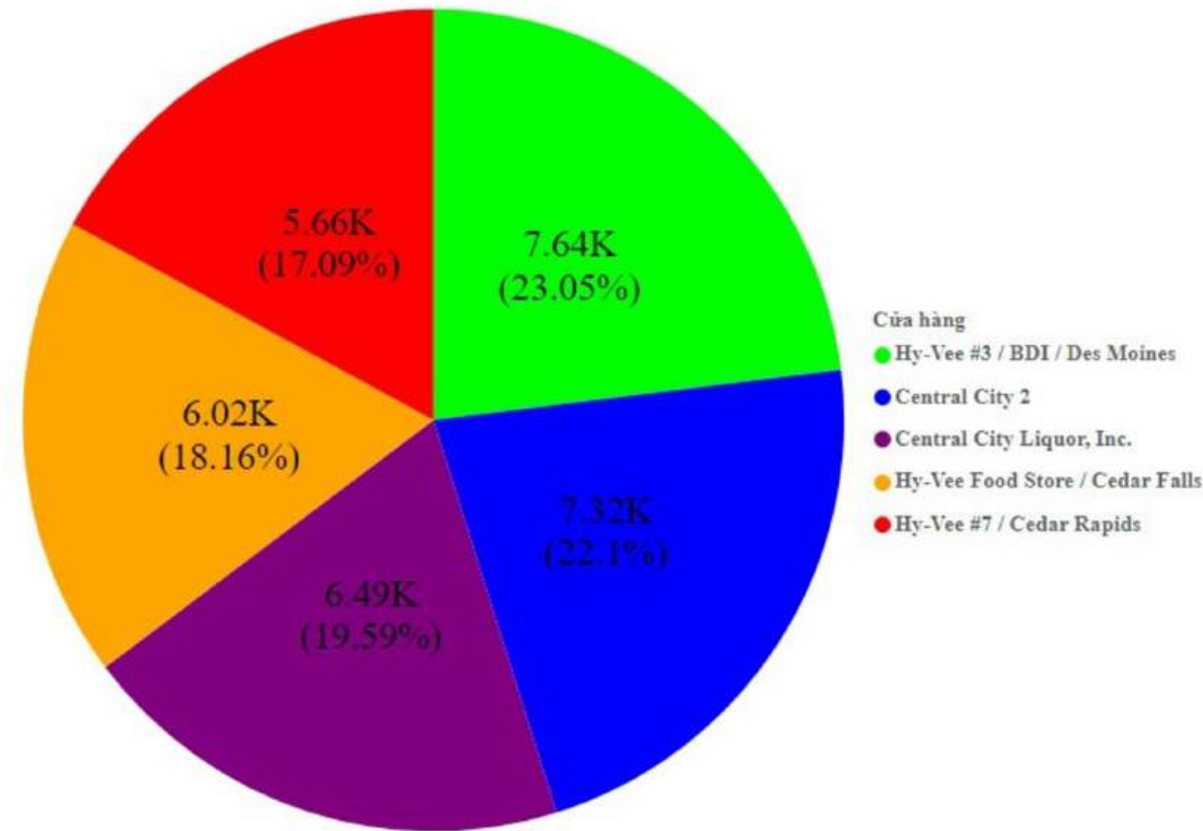
## 4.3 Tình hình doanh thu và đơn đặt hàng

Top 20 loại rượu có doanh thu cao nhất



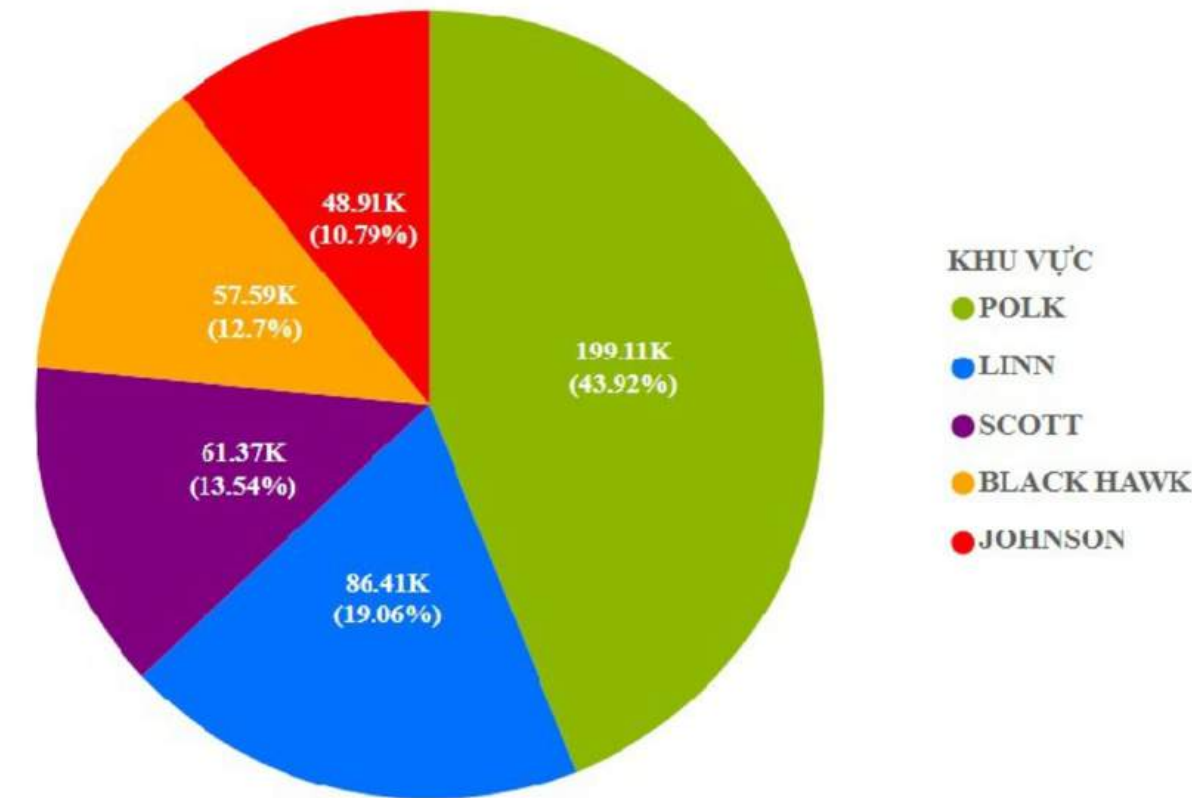
Loại rượu được ưa chuộng nhất ở Iowa là American Vodkas với tổng doanh thu khoảng \$ 24.83M

Top 5 cửa hàng đặt nhiều đơn hàng nhất



Cửa hàng có đơn đặt hàng nhiều nhất là Hy-Vee #3 với tổng số lượng đơn đặt hàng trong năm 2021 khoảng 7.64K đơn hàng

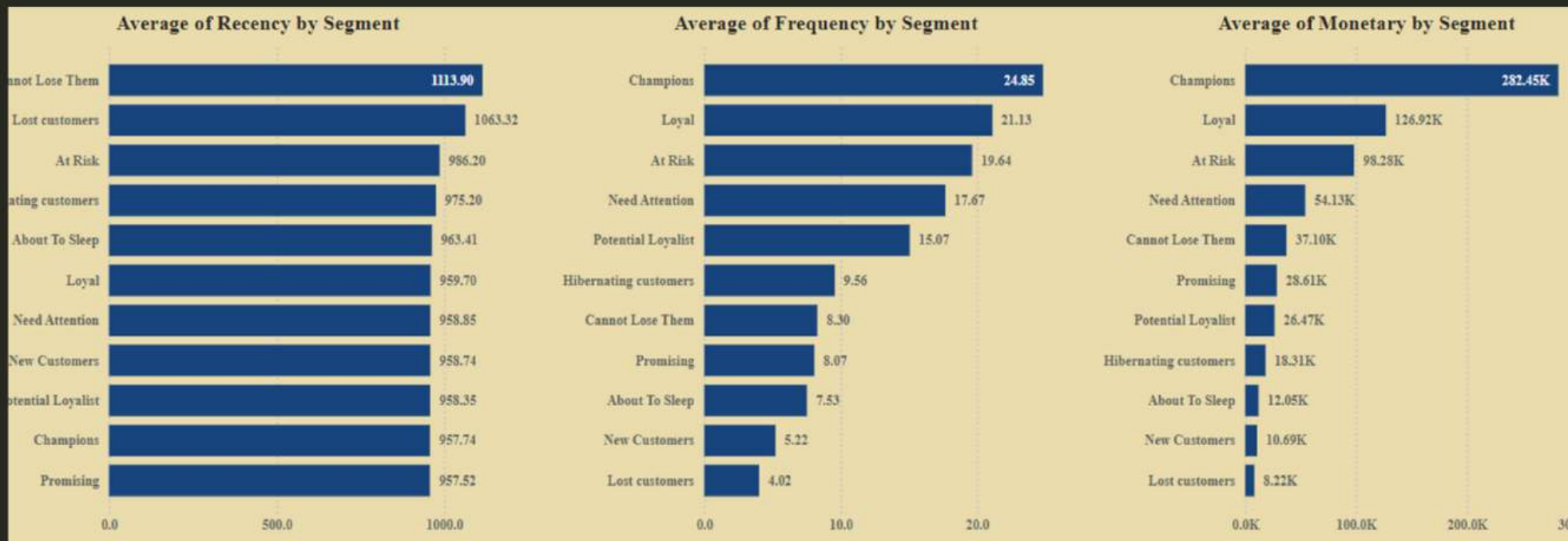
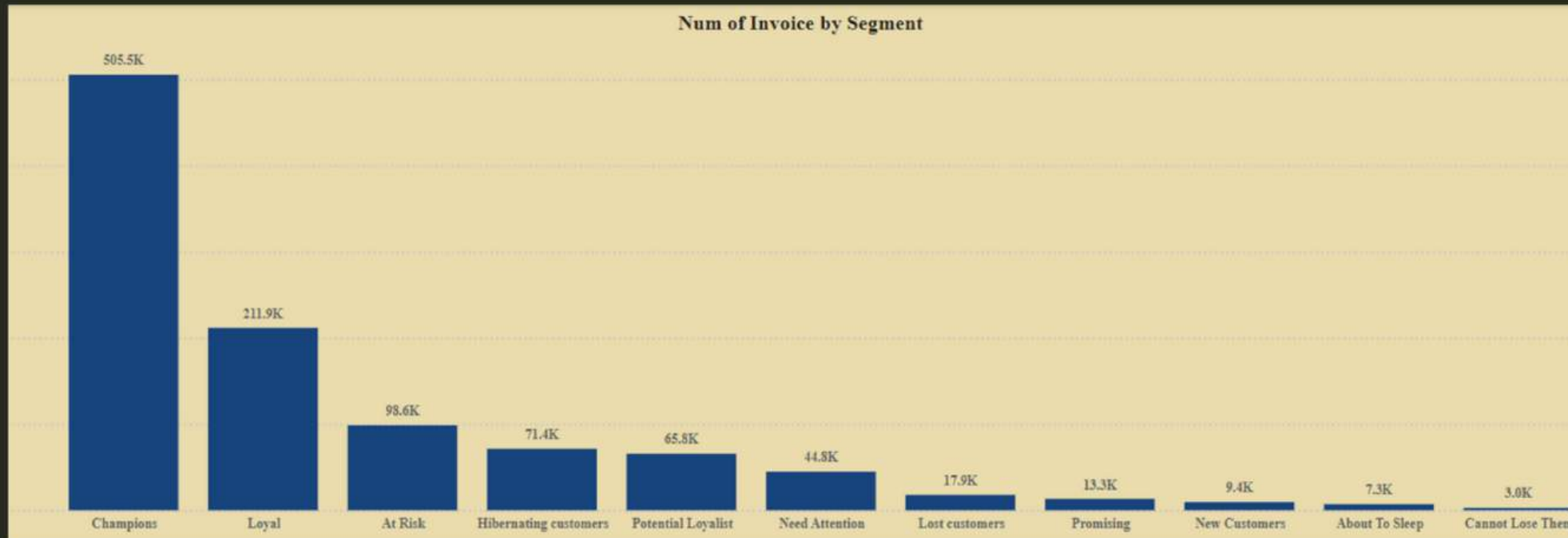
Top 5 khu vực có đơn đặt hàng nhiều nhất



Khu vực Polk là khu vực đặt hàng lớn nhất với số lượng đơn đặt hàng là 199.11K đơn hàng



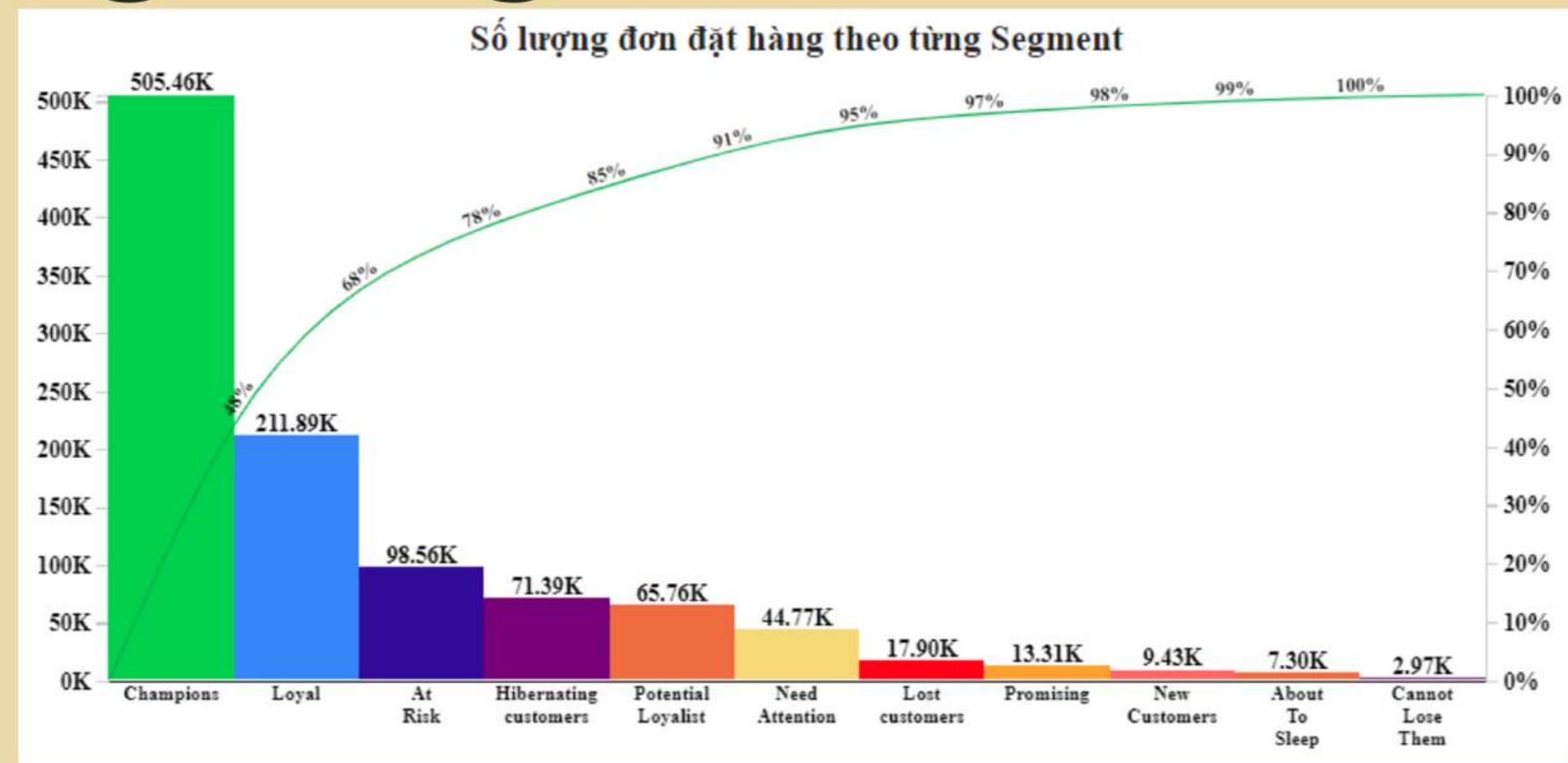
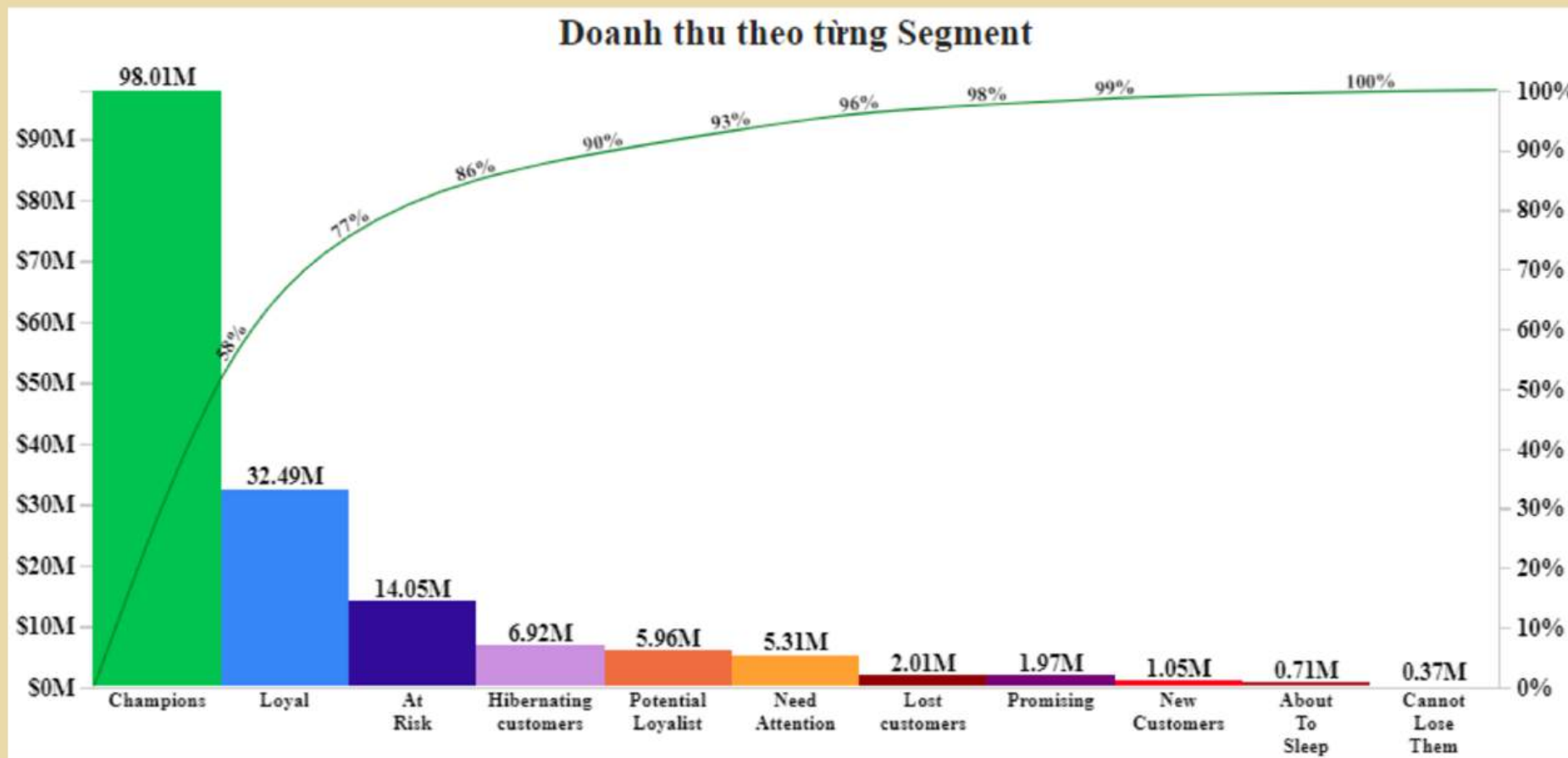
# 4.4 NHỮNG CỬA HÀNG CÓ KHẢ NĂNG RỜI BỎ



- Nhóm Cửa hàng Lost Customers, Promising, New Customers, About to sleep và Cannot Lose Them là nhóm cửa hàng có ít đơn đặt hàng cùng với doanh thu thấp.
  - Đặc biệt nhóm cửa hàng Lost Customers có thời gian mua hàng cuối cùng cách hiện tại khoảng hơn 1000 ngày, số lần mua hàng rất ít khoảng dưới 4 lần và doanh số thấp nhất so với các nhóm cửa hàng còn lại.
- > Từ đó có thể thấy nhóm cửa hàng Lost Customers sẽ có khả năng rời bỏ dịch vụ.



## 4.5 Nhóm cửa hàng trung thành



- Nhóm cửa hàng đem lại doanh thu nhiều nhất là Champions và Loyal chiếm hơn 77% tổng doanh thu trong Iowa
- Ngoài ra, nhóm cửa hàng Champions và Loyal có số lượng đơn đặt hàng chiếm khoảng hơn 68% trong tổng số đơn đặt hàng tại khu vực

Nhìn chung, tuy chỉ chiếm gần 20% số lượng khách hàng nhưng đem lại gần 80% giá trị doanh thu.

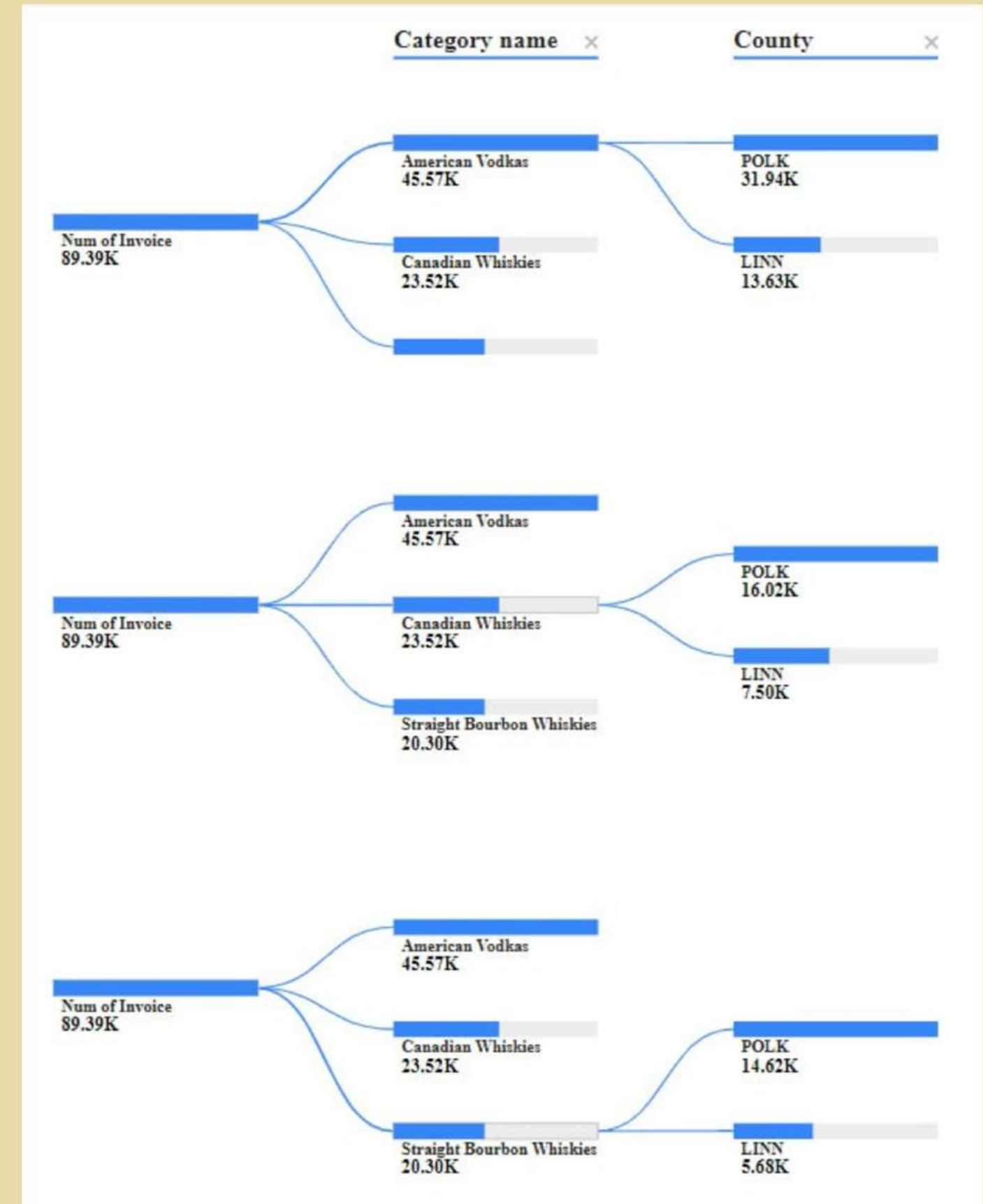
Vì vậy, đây là nhóm cửa hàng cần được tập trung khai thác.

Nếu có sản phẩm mới, 2 nhóm cửa hàng này cần được ưu tiên giới thiệu trước.



## 4.6 Sự khác nhau giữa các khu vực về sản phẩm phổ biến

- + Đây là 3 loại rượu có số lượng đơn đặt hàng nhiều nhất trong năm 2021. Cả 3 loại rượu này có số lượng đặt hàng nhiều nhất là ở khu vực Polk thứ hai là khu vực Linn
- + Ngoài khu vực Polk có lượng đặt hàng lớn và doanh thu cao nhất. Các quận còn lại không có nhiều sự chênh lệch về số lượng đặt hàng cũng như doanh thu





# 5. KẾT LUẬN VÀ ĐỀ XUẤT

## 5.1 KẾT LUẬN

1. Nhóm cửa hàng đem lại doanh thu cao nhất cho Iowa là nhóm cửa hàng Champions và Loyal, đây cũng là nhóm cửa hàng trung thành và tiềm năng nhất.
2. Nhóm cửa hàng có khả năng rời bỏ dịch vụ là nhóm Lost Customers, About to Sleep và New Customers. Đây là các nhóm cửa hàng mua hàng số lượng ít, thời gian mua hàng cách đây rất lâu và số tiền bỏ ra mua hàng không nhiều.
3. Đặc biệt đối với nhóm New Customers là nhóm cửa hàng mới, họ đang trong quá trình tìm kiếm sản phẩm phù hợp với nhu cầu tại khu vực của họ, nên việc có quay lại và trở thành khách hàng tiềm năng hay không còn phụ thuộc nhiều vào các yếu tố về chất lượng, giá thành của sản phẩm cũng như các chương trình khuyến mãi, hay thậm chí là cả hậu mãi.





## 5.2 Đề xuất



1. Khi có sản phẩm mới có thể ưu tiên chào bán cho nhóm cửa hàng Champions và Loyal trước vì đây là nhóm khách hàng trung thành và đem về doanh thu cao nhất cho Iowa. Ngoài ra nhóm cửa hàng này cũng không cần phải bỏ ra nhiều chi phí để tiếp cận, quảng bá khi có sản phẩm mới.
2. Đối với các nhóm cửa hàng Promising, About to sleep, Cannot Lost Them, và đặc biệt là nhóm cửa hàng Lost Customer có khả năng rời bỏ thì không nên tốn quá nhiều chi phí để thực hiện các chiến dịch quảng cáo, vì nhóm cửa hàng này không đem lại nhiều doanh thu. Tuy nhiên đối với các nhóm Promising, Cannot Lost Them, nên duy trì 1 khoản chi phí quảng cáo tối thiểu; mặc dù các nhóm này hiện tại chưa mang lại nhiều doanh thu, nhưng họ hoàn toàn có thể trở thành khách hàng tiềm năng trong tương lai.
3. Để thu hút nhóm cửa hàng New Customers, nên có các chương trình khuyến mãi dành riêng cho nhóm này, tăng cường giới thiệu các sản phẩm mới để nhóm cửa hàng này tiếp cận được với nhiều loại rượu hơn.





Thank's for  
watching