Store Location Recommendation: Iowa Study

DC-DSI 4
March 2017

Joshua Sparks

Roadmap

- Background
- Introductory exploration and rationale for data analysis
- Assumptions
- Report on “successful” stores
- Model building methodology
- Takeaways

Source: https://github.com/ga-students/DC-DSI4/blob/master/projects/01-project/data/sat_scores.csv

Background

- Objective: Given Iowa data, recommend possible locations
- Data collection: Provided, sourced from link below
- Data scope: 2015 transactions for analysis

Source: https://data.iowa.gov/Economy/Iowa-Liquor-Sales-by-Year-and-County/ahiv-u4uz

Introductory Exploration of Data

Introductory Analysis

- Location target for analysis: Zip code
- Drawbacks:
  - Somewhat scattershot
  - Physical areas covered may vary

Introductory Analysis

- After exploring the data:
  - Highest individual location classifier: Zip code
  - Least number of missing values: Zip code

Assumptions

Key Assumptions

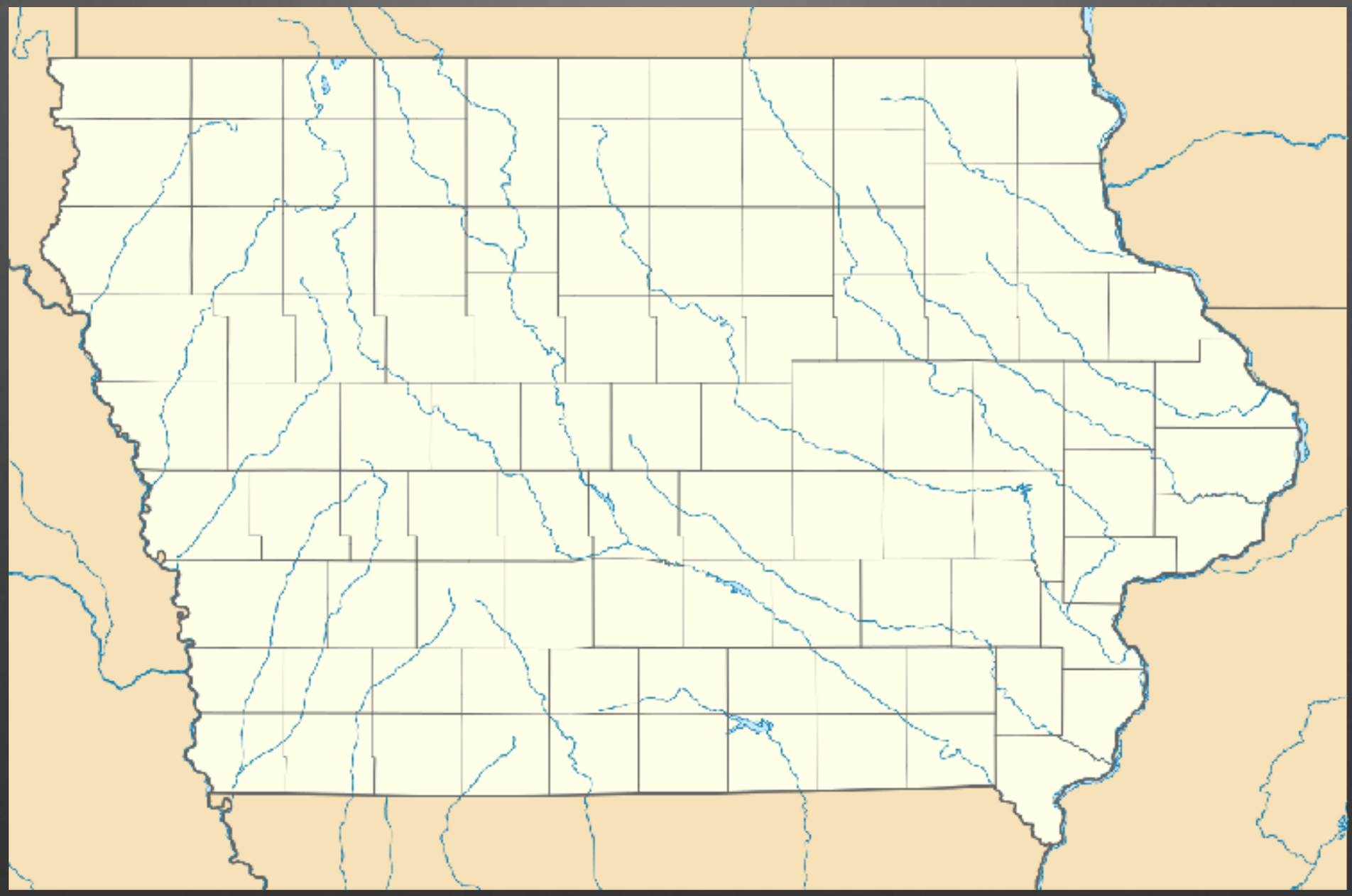
- Data are as accurate as Iowa provides
- Meaningful inferences based on State Sales

(These are data from state's point-of-view)

- Bottles purchased can achieve good sell-through

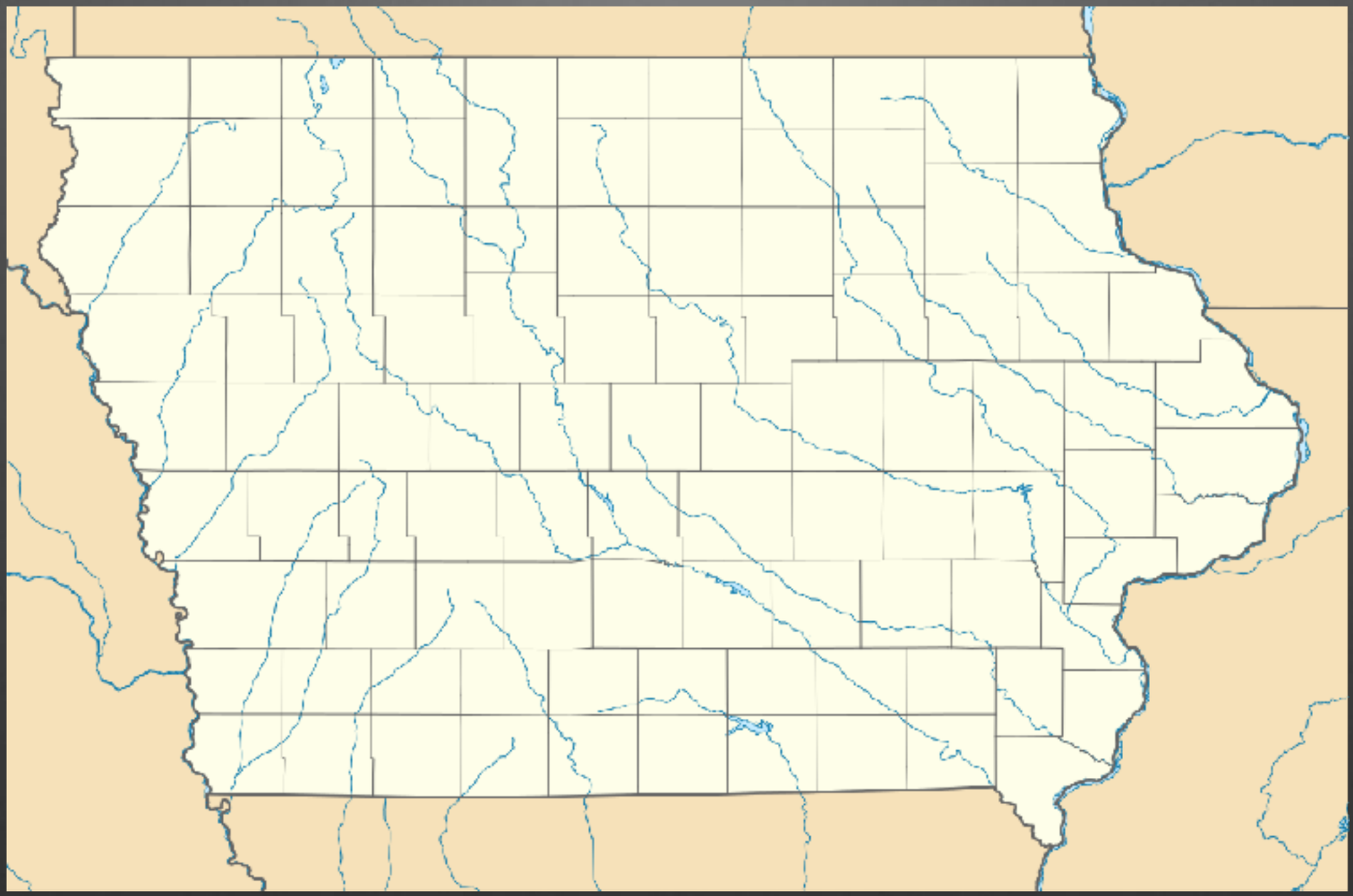
State and Store Reports

State Annual Sales (Bottles Sold to Licensees)



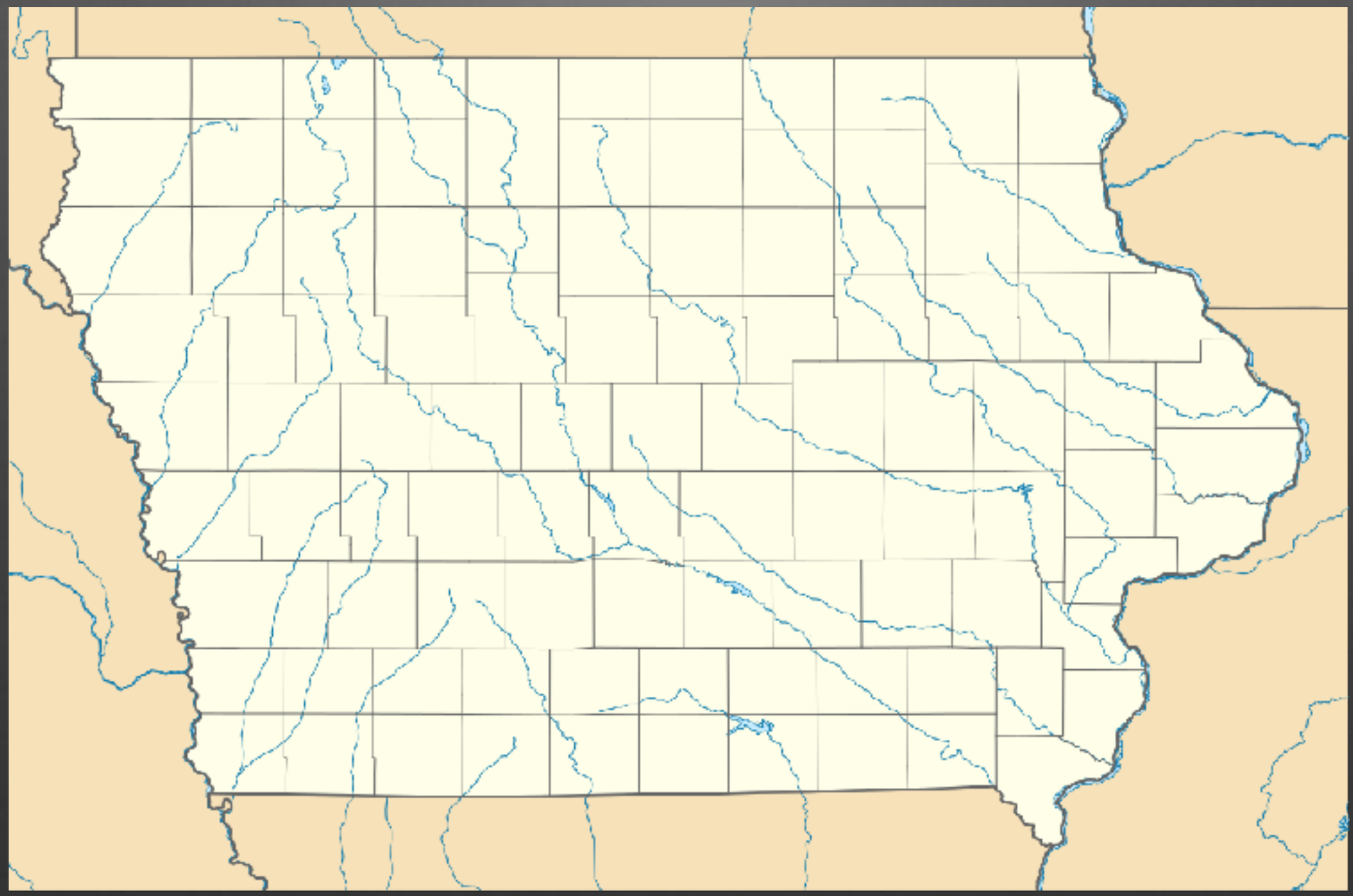
21,544,320

State Annual Sales (Gallons Sold to Licensees)



5,175,835

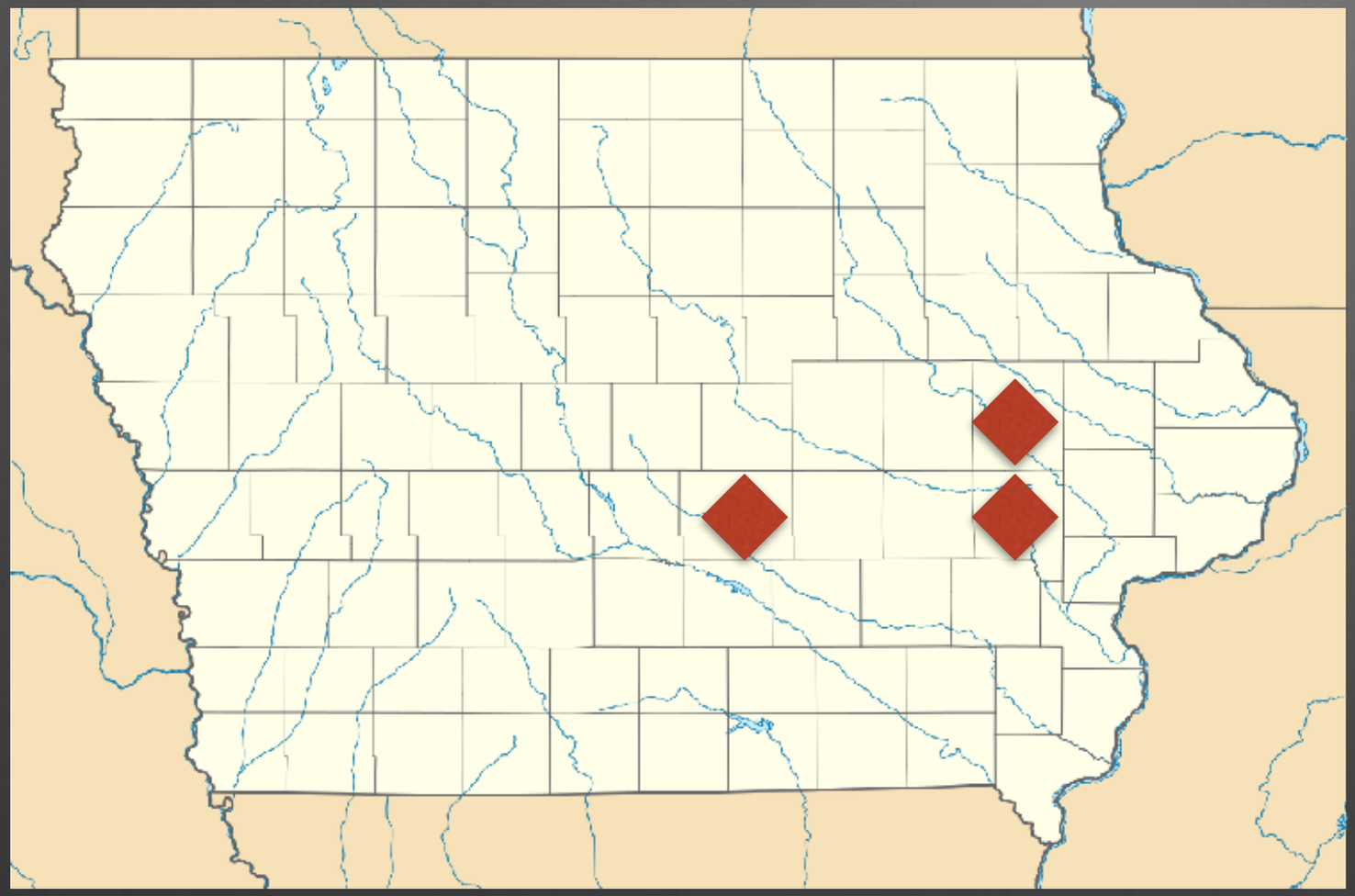
State Annual Profit (Sold - Cost)



$ 94,936,918

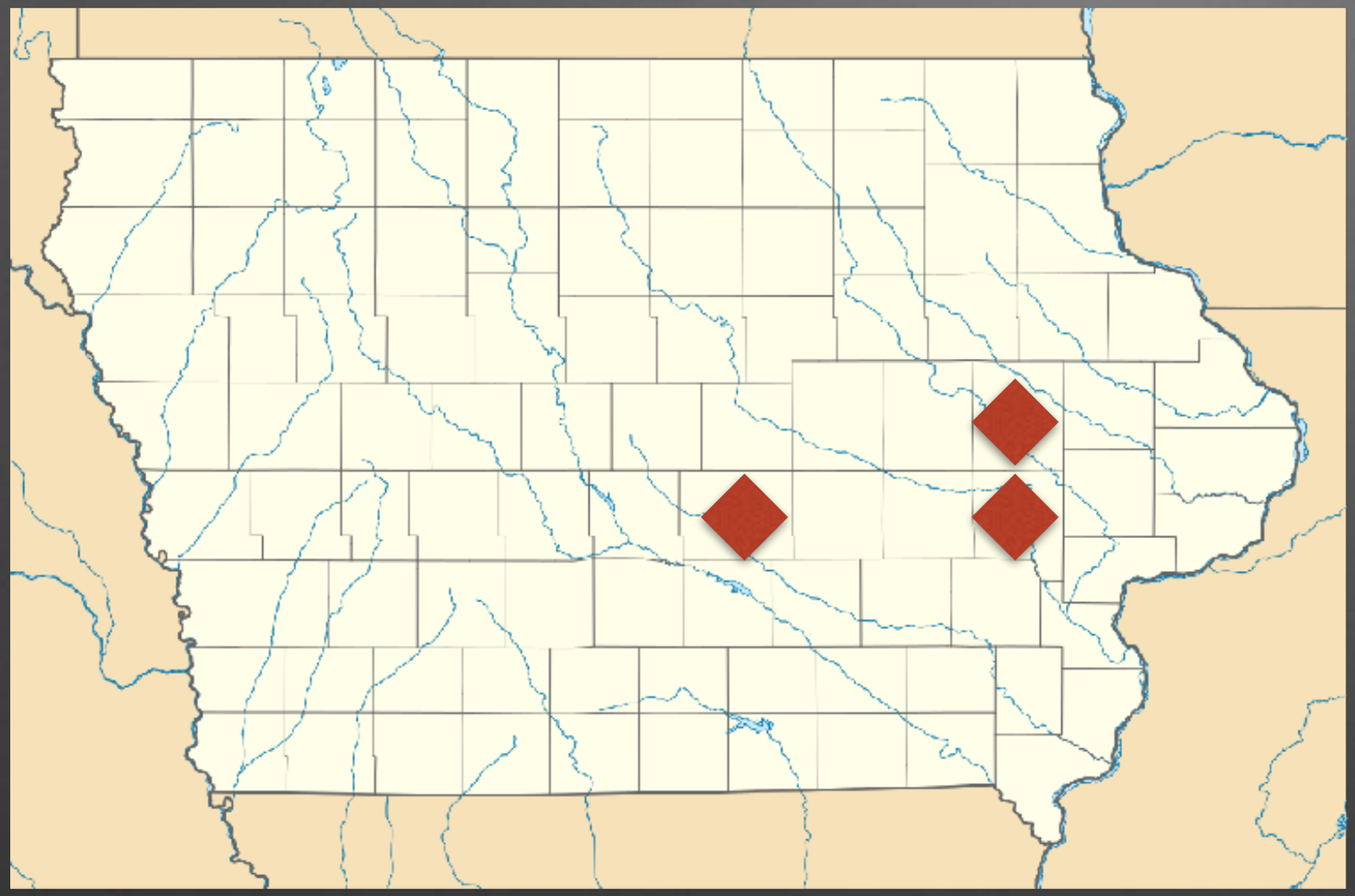
STORE NUMBER	LOCATION	Name
2633	DES MOINES	Hy-Vee #3
4829	DES MOINES	Central City 2
2512	IOWA CITY	Hy-Vee Wine and Spirits
3385	CEDAR RAPIDS	Sam’s Club 8162
3420	WINDSOR HEIGHTS (DES MOINES)	Sam’s Club 6344

Top Performers: Bottles Purchased



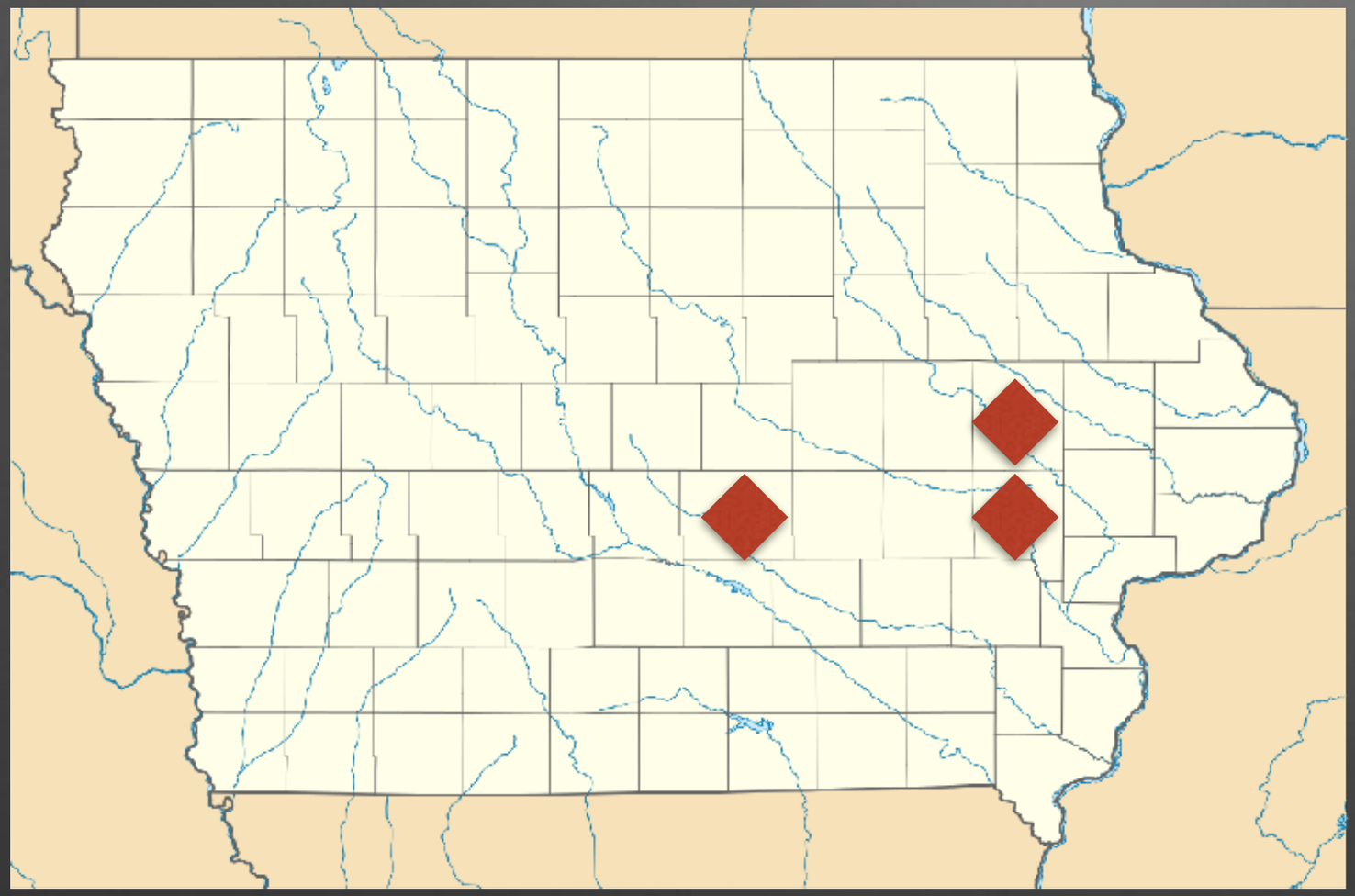
STORE #	LOCATION	Name	Bottle Count
2633	DES MOINES	Hy-Vee #3	595083
4829	DES MOINES	Central City 2	517487
2512	IOWA CITY	Hy-Vee Wine and Spirits	282804
3385	CEDAR RAPIDS	Sam's Club 8162	244868
3420	WINDSOR HEIGHTS (DES MOINES)	Sam's Club 6344	192460

Top Performers: Gallons Purchased



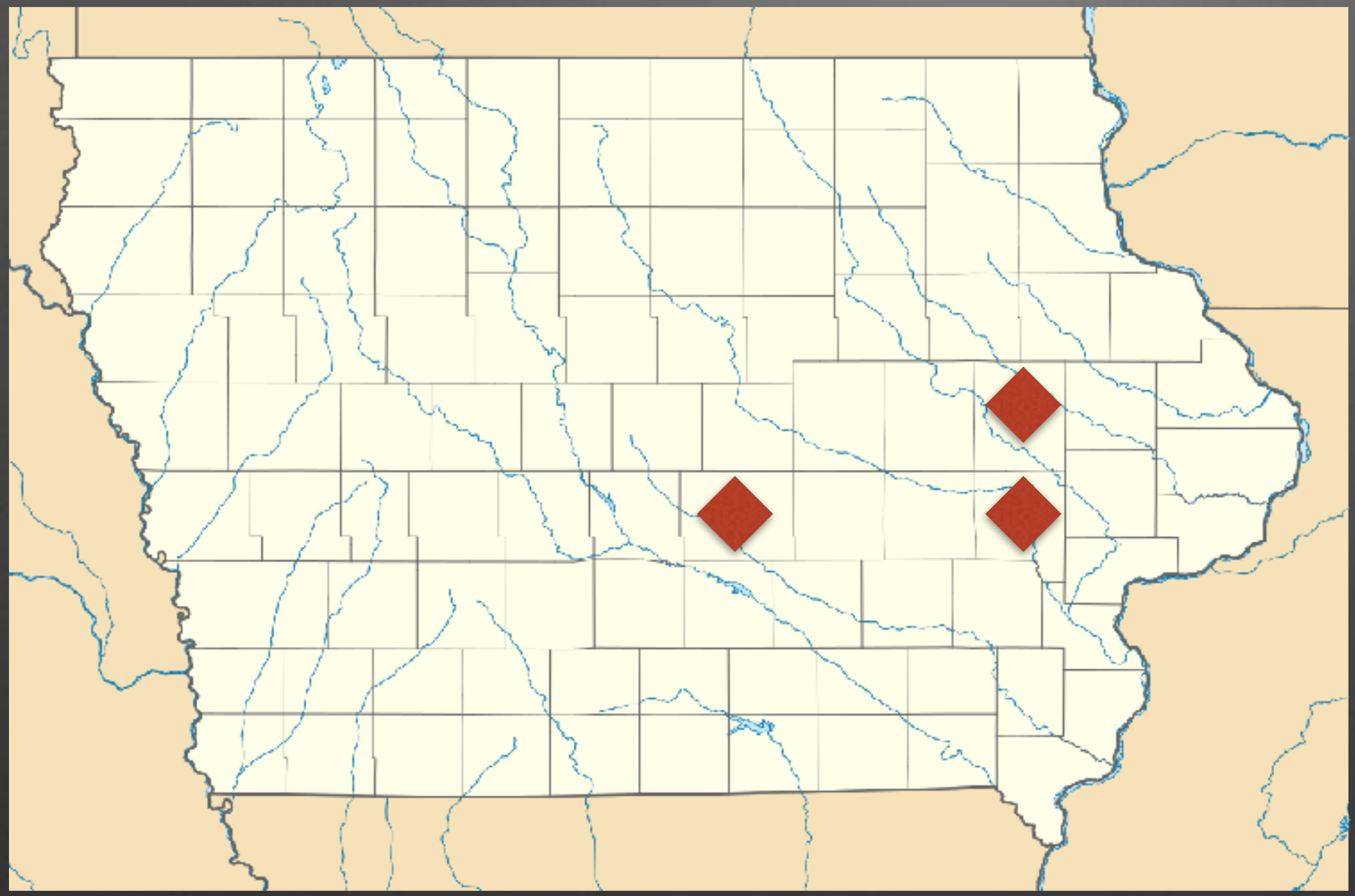
STORE #	LOCATION	Name	Gallons
2633	DES MOINES	Hy-Vee #3	153957
4829	DES MOINES	Central City 2	132091
2512	IOWA CITY	Hy-Vee Wine and Spirits	72084
3385	CEDAR RAPIDS	Sam's Club 8162	70668
3420	WINDSOR HEIGHTS (DES MOINES)	Sam's Club 6344	62649

Top Performers:
$ Spent



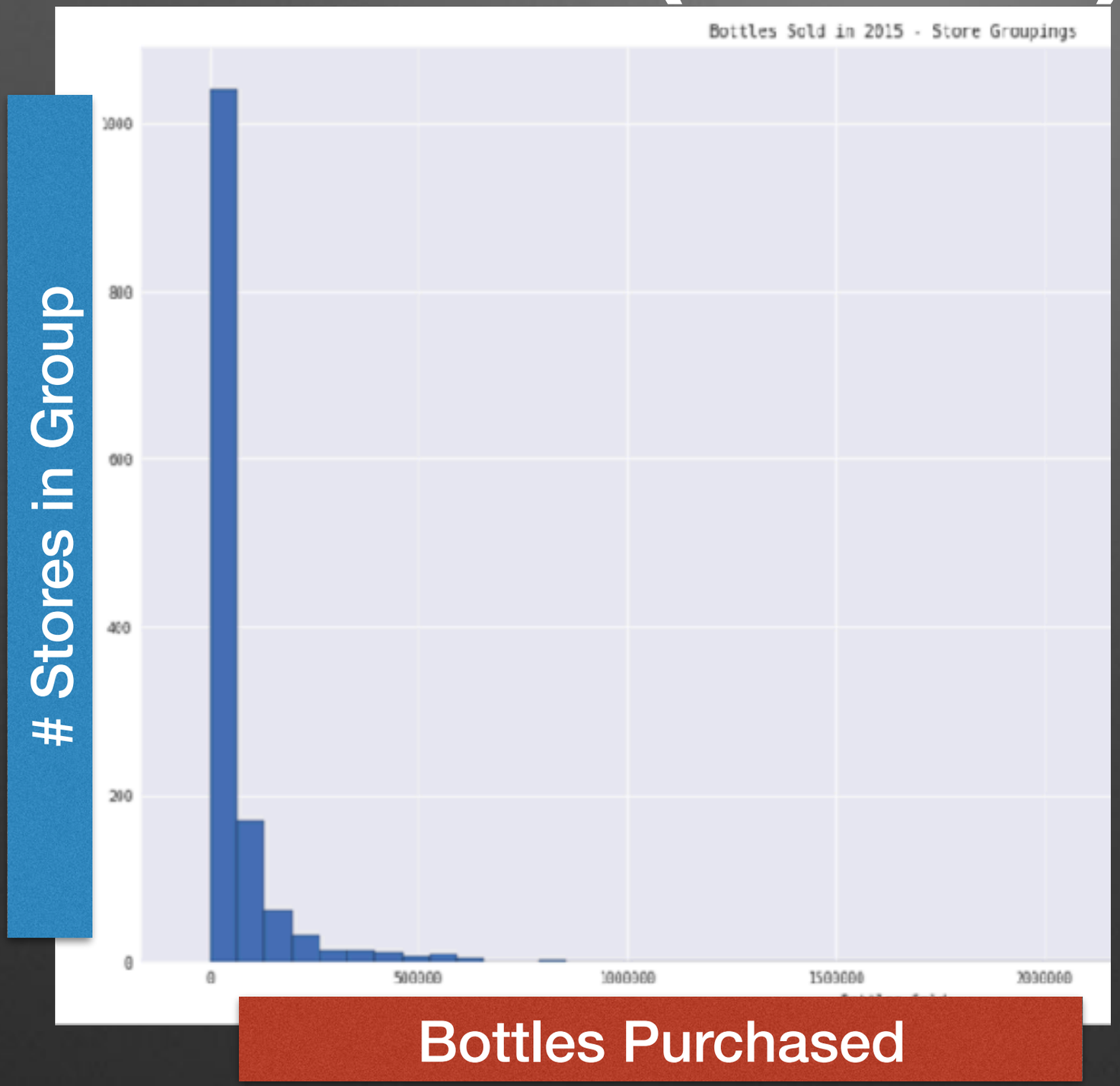
STORE #	LOCATION	Name	$ Spent
2633	DES MOINES	Hy-Vee #3	$9,837,698
4829	DES MOINES	Central City 2	$8,696,730
2512	IOWA CITY	Hy-Vee Wine and Spirits	$4,155,665
3385	CEDAR RAPIDS	Sam's Club 8162	$3,941,460
3420	WINDSOR HEIGHTS (DES MOINES)	Sam's Club 6344	$3,418,898

Top Performers:
$ Purchased



2633	9837698.17
4829	8696730.63
2512	4155665.47
3385	3941460.17
3420	3418898.19

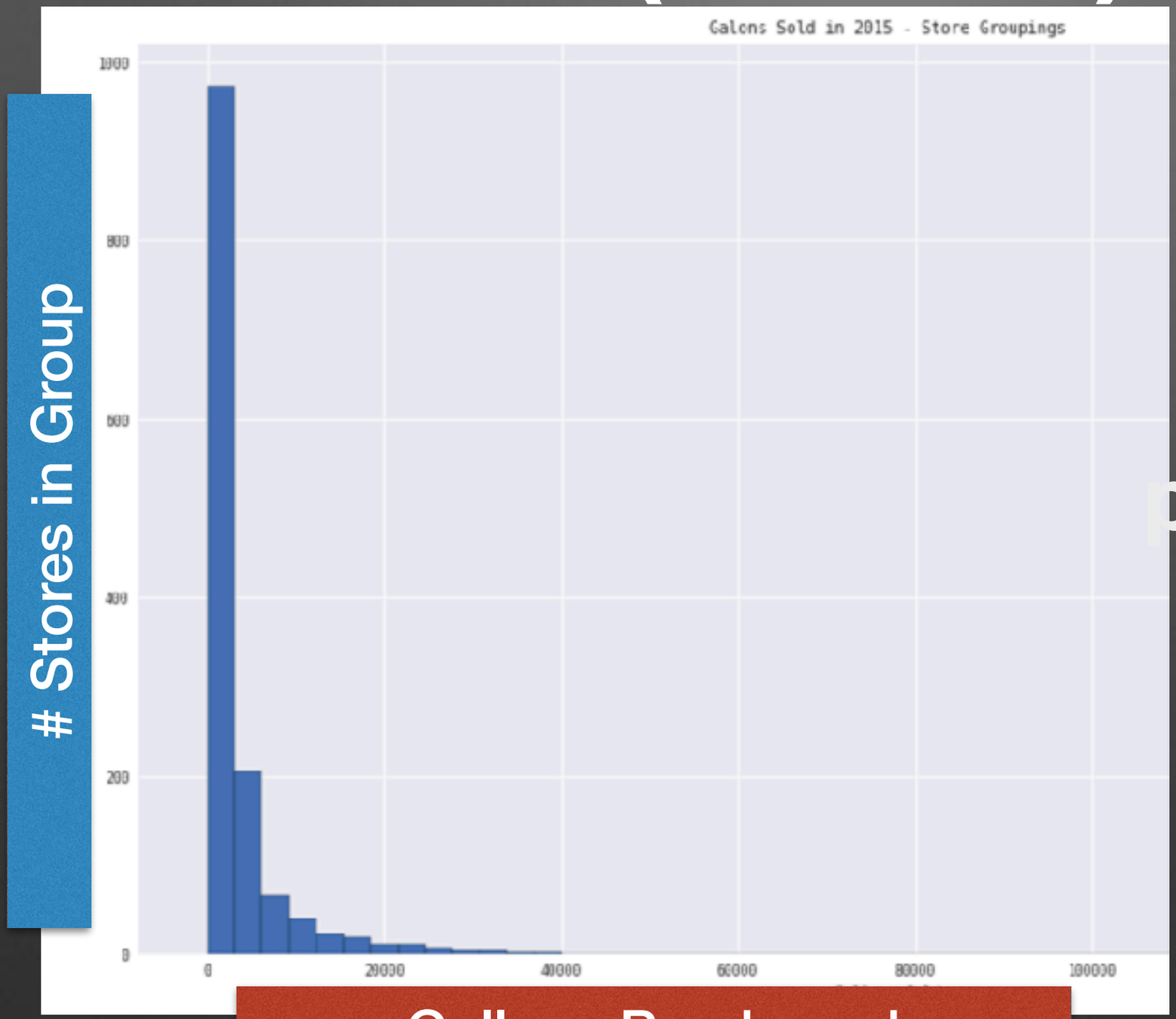
Store Annual Purchases (Bottles)



Most stores purchase fewer bottles

Source: http://www.edweek.org/ew/articles/2001/06/13/40actbox.h20.html

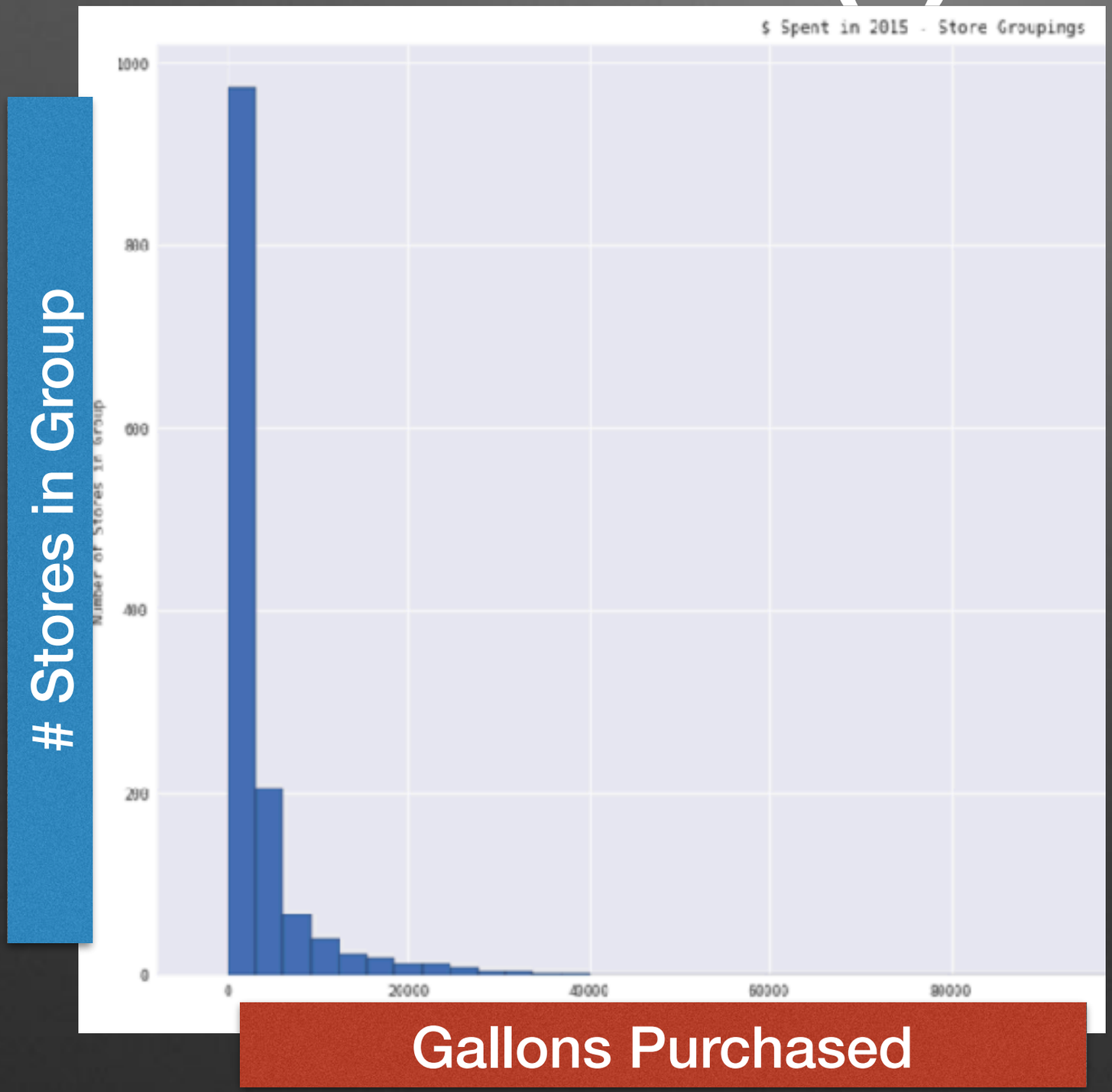
Store Annual Purchases (Gallons)



Most stores purchase lower volumes

Source: http://www.edweek.org/ew/articles/2001/06/13/40actbox.h20.html

Store Annual Purchases ($)



Most stores purchases in lower $ range

Source: http://www.edweek.org/ew/articles/2001/06/13/40actbox.h20.html

Predictive Model

Methodology

- Goal: Suggest store locations
- Point of a model: Prediction
- Idea: Using data, can I provide model-based predictions of locations based on your store's sales data?

Methodology

- Inverse approach:
- May be able to take an area and predict sales, but has no basis on your actuals
- As owner, you have expertise on historical sales and annual targets

Methodology

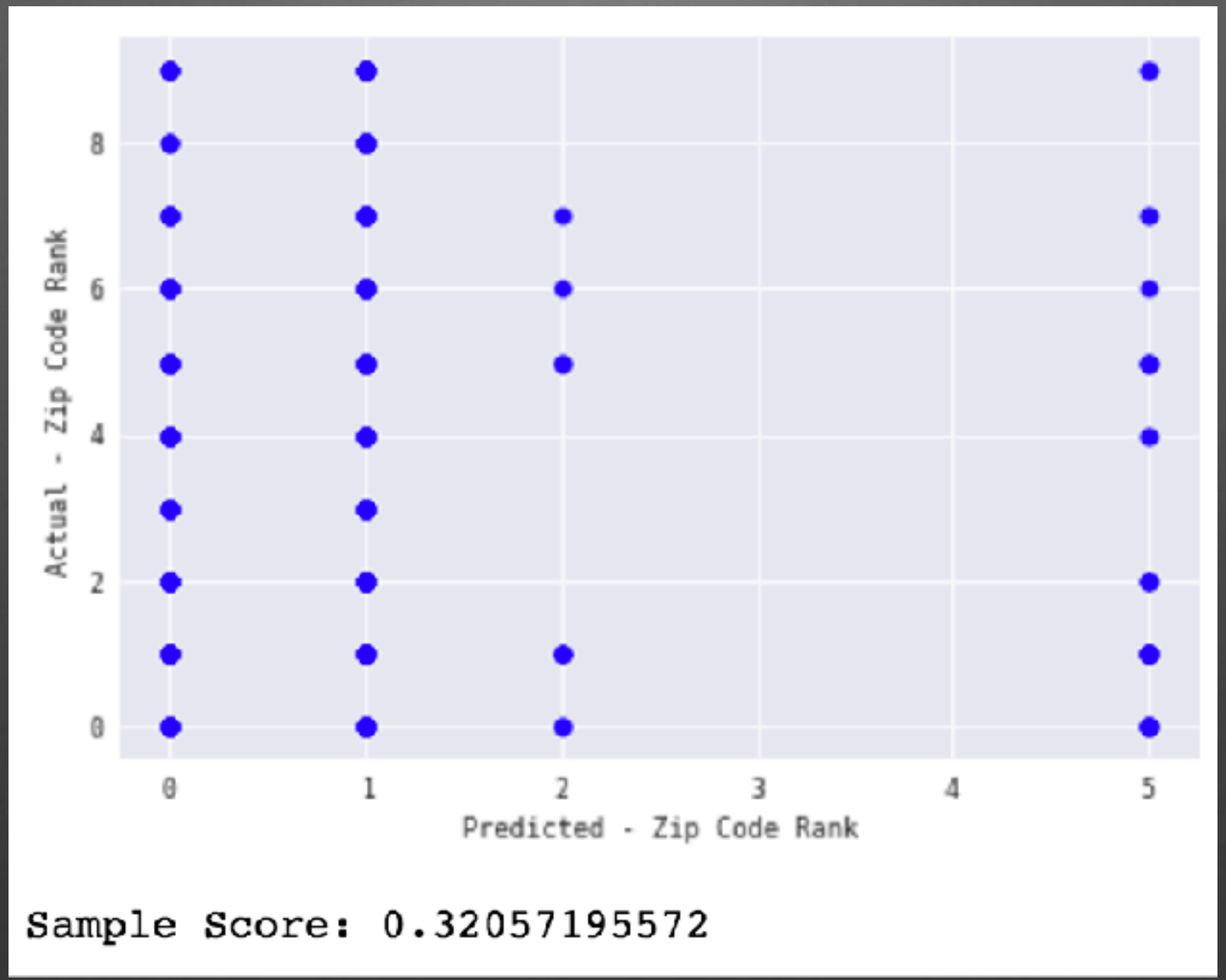
<u>Therefore:</u>

- Likely recommendation just from "eyeballing": Look at areas of high population
- Using a model, which can account for the dependences between input, find a more objective recommendation?

Methodology

- Group zip codes based on how many stores order - higher ranked zip codes imply higher market potential
- Categorical “buckets” of zip codes based on representation
- 10 bucket target: This already suggests 40 zip codes
- Then feed in sales targets, predict which zip code group it matches

Model Results



Sample Score: 0.32057195572

Takeaways

Takeaways

Data based inference

- Target the high population centers
- Expectations: Perhaps obvious, but unlikely to become very high-volume purchaser

Takeaways

<u>Model-based prediction:</u>

- TBD based on your targets
- TBD based on model refinement

Executive Summary

- In general, stores do not purchase in high quantities, or spend too much on alcohol. Targeting high-population centers with a small to mid-size approach with regard to spending should provide reasonable business stability and revenue. Further investigation for inverted predictions, of revenue per zip code, will be delivered in follow-up dependent on your chain's areas of interest.

Future Directions

- Re-run analysis with prediction of store sales based on locations of interest, follow-up
- Continue to refine model, adjust setup