BETTER IN BULK?

WHAT CONSUMER HABITS OF CAP 5771 REVEAL   COMPARED ALONGSIDE THE AVERAGE CUSTOMER

Hayden Kirkeide and Alan Van Etten

INTRODUCTION

INTRODUCTION

- Costco Wholesale: membership-based wholesale retail chain founded in 1983
  - 1976: The first official store opens in San Diego, California, operating under a different company that merged with the Costco corporation in 1993
  - Today: locations in 14 different countries across 4 continents

- Mission statement: to provide products ranging from produce, furniture, and even insurance and tires of the "best value" it can possibly provide

- 76.2 million Costco membership holders - 2024 fiscal year documents
  - 7.3% YOY increase compared to 2023
  - Also taking into account a $5.00 membership price increase in September 2024

- Demographics tend to skew towards middle-aged individuals with families, but recent demographic inquiries have shown consistent growth in new memberships belonging to millennials (30 - 35 y/o) and Generation Z (20 - 25)

RESEARCH

RESEARCH QUESTIONS

Our guiding research questions:

1. What do the motivations and benefits of shopping at Costco versus other stores tell us about our demographic as students that are young adults?

2. Do our class's Costco and general shopping habits provide insight into the global trends of Costco consumers when compared against them?

3. What can the limited study of our consumer habits – while of varied socioeconomic circumstances – as students in this time of economic uncertainty reveal about consumer habits in the modern day?

NOVELTY

- Narrowed analytical focus to one specific demographic group

- Emphasized Costco in particular rather than analysis over a group of stores

- Change in trends over a short period time in a market facing a negative outlook

- Goal of comparing against global trends rather than determining global trends

EXPERIMENTS ALGORITHMS &

ALGORITHMIC PROCESSES

- Using Python via Google Colab

- Storing the survey data in Google Sheets, importing as a dataframe

- Exploratory analysis provided by Google Forms

- Statistical analysis: Chi-square tests, ANOVA tests

- Modelling: linear regression, logistic regression, k-means clustering, multi-layer perceptron

EXPERIMENTS AND ANALYSIS

- Chi-square tests for independence between age and membership status

- ANOVA tests to search for correlations between membership status and the shopping habits
  - Relationship between demographics and membership status established in chi-square test

- Linear regression to visualize the relationship between the separate shopping attributes

- K-means clustering to find common groupings of demographic attributes within a range of shopping attribute values

- Logistic regression and a multi-layer perceptron for predictive analysis, cross-validated and using previous tests as a baseline for which attributes to include and their starting weights in order to predict membership/non-membership based on demographic characteristics

DATASETS

EARNINGS DOCUMENTS

Include growth in membership and revenue

SYNTHETIC DATASETS

Artificially-generated datasets for scalable impacts of our analysis

YOU!

Survey responses of demographic and transaction data

PRELIMINARY

Demographics, membership status, and privacy acknowledgements

TRIP-BY-TRIP

Submitting general transaction data and grocery shopping figures

SURVEY DESCRIPTIONS

CHI-SQUARE TEST OF INDEPENDENCE

- H_0: no relationship between age and membership status

- H_1: observed relationship between age and membership status

KEEP NULL HYPOTHESIS

CAP 5771

Chi-Square Test	Results
Pearson Chi-Square	1.98385
p-value	0.3794
Cramer's V	0.3721

SYNTHETIC

Chi-Square Test	Results
Pearson Chi-Square	5.8679
p-value	0.1182
Cramer's V	0.1532

ANOVA

- H_0: no relationship between Costco membership status and money spent

- H_1: statistically significant relationship between Costco membership status and money spent

KEEP NULL HYPOTHESIS
CAP 5771
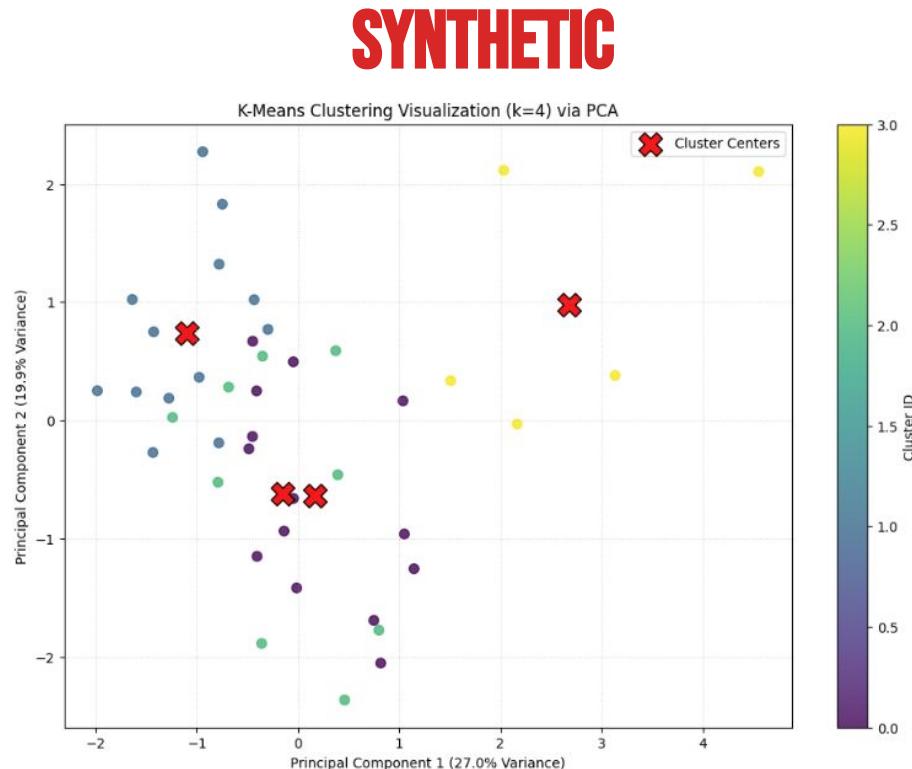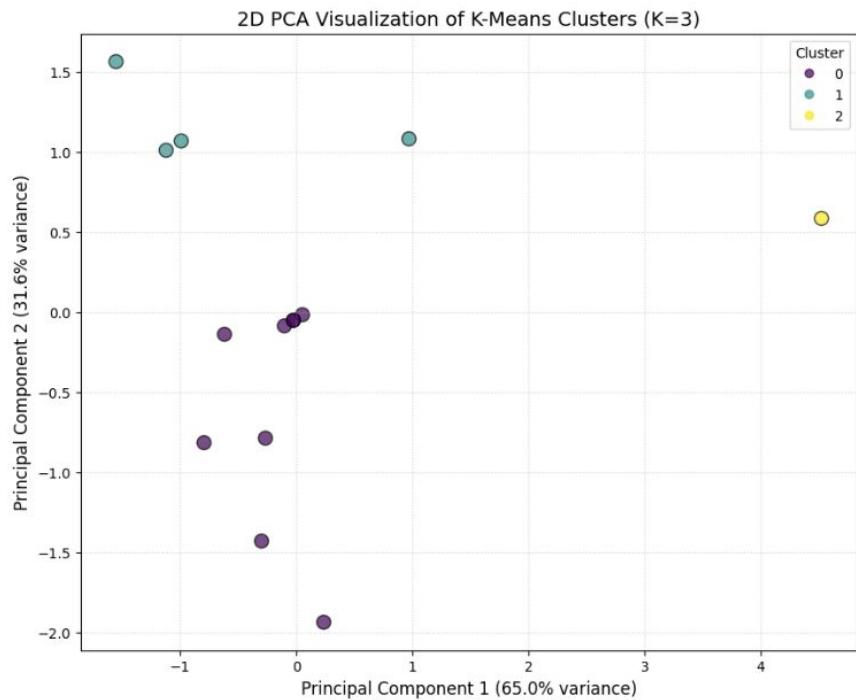
ANOVA Statistic	Results
F-statistic	0.3652
p-value	0.5567
Effect size (η^2)	0.0295

REJECT NULL HYPOTHESIS
SYNTHETIC
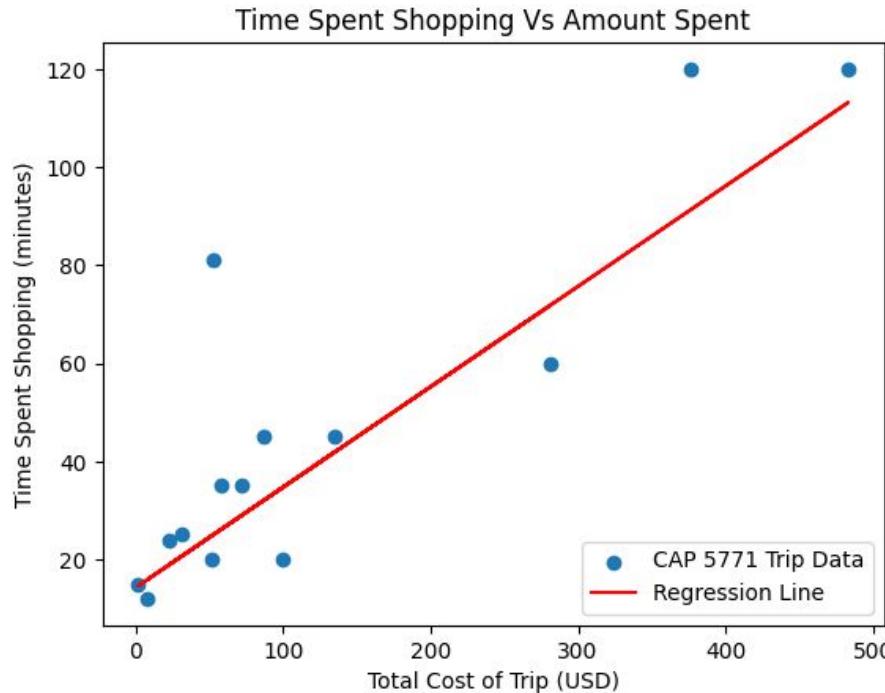
ANOVA Statistic	Results
F-statistic	702.4277
p-value	0.0000001
Effect size (η^2)	0.7391

K-MEANS CLUSTERING

CAP 5771

SYNTHETIC
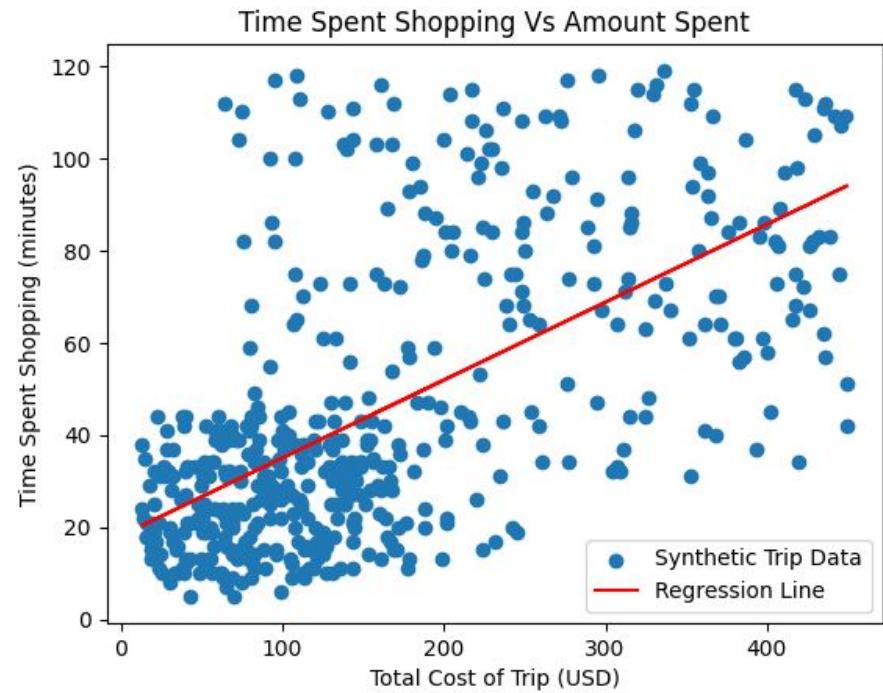
LINEAR REGRESSION - TIME SPENT VS COST

Time Spent Shopping Vs Amount Spent (left)
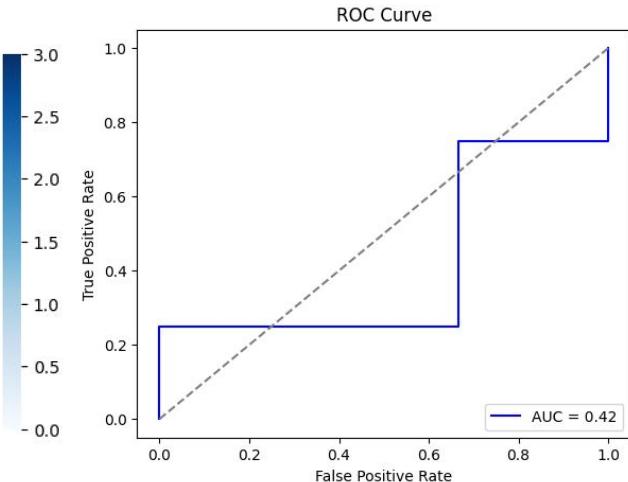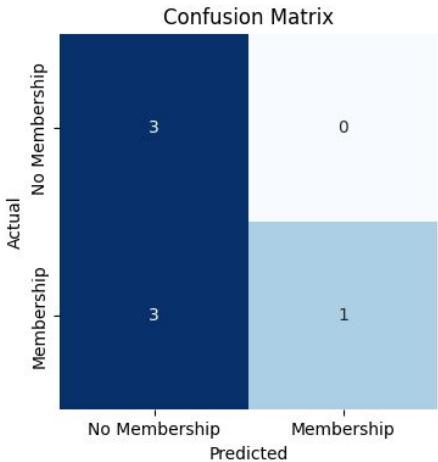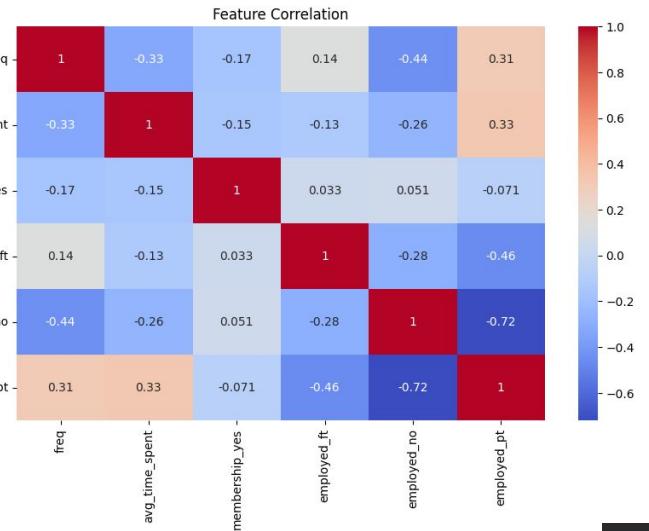
Y = 0.21341018 X + 20.09818627
R = 0.87291251 P-value: 4.60042326e-05

Time Spent Shopping Vs Amount Spent (right)

Y = 0.16509734 X + 19.07143373
R = 0.61405076 P-value: 3.8095749e-53

LOGISTIC REGRESSION - CLASS DATA

Predicting Membership Status Using Employment, Visit Frequency, and Time Spent Shopping

Feature Correlation

Confusion Matrix

ROC Curve

AUC = 0.42
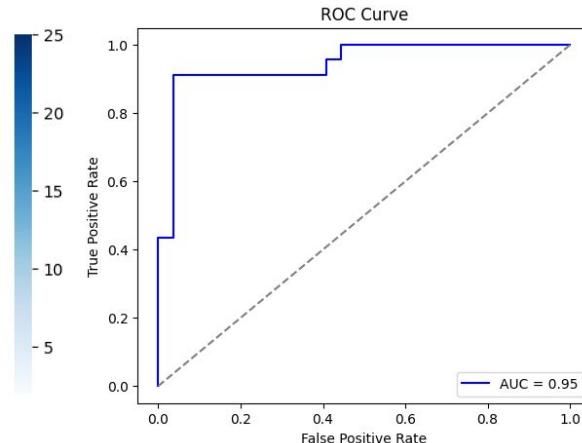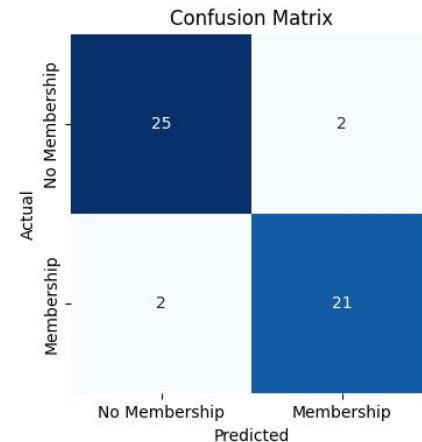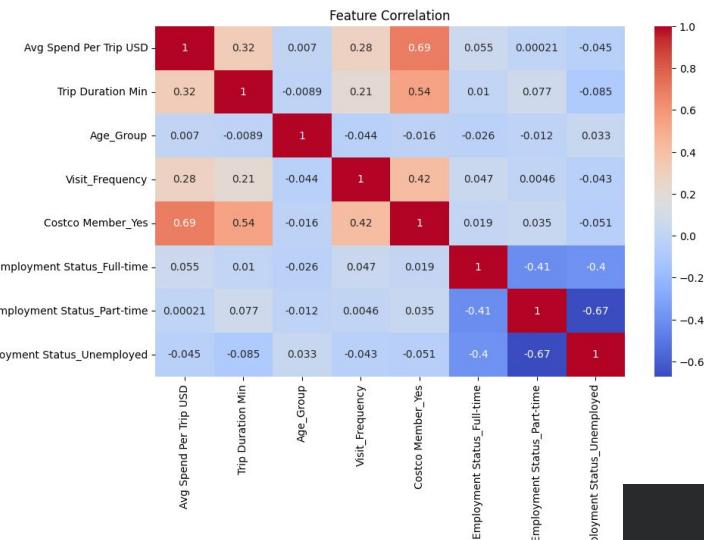
	Feature	Coefficient
3	employed_no	0.167106
4	employed_pt	0.072077
2	employed_ft	-0.239403
0	freq	-0.477752
1	avg_time_spent	-0.619890

	precision	recall	f1-score	support
False	0.50	1.00	0.67	3
True	1.00	0.25	0.40	4
accuracy			0.57	7
macro avg	0.75	0.62	0.53	7
weighted avg	0.79	0.57	0.51	7

LOGISTIC REGRESSION - SYNTHETIC DATA

Predicting Membership Status Using Employment, Visit Frequency, Age Group, and Time Spent Shopping


Feature Correlation

Confusion Matrix

ROC Curve

Feature	Coefficient
Avg Spend Per Trip USD	2.229029
Trip Duration Min	1.425968
Visit_Frequency	1.006230
Employment Status_Unemployed	0.030392
Employment Status_Full-time	-0.003151
Employment Status_Part-time	-0.027002
Age_Group	-0.101394

	precision	recall	f1-score	support
False	0.93	0.93	0.93	27
True	0.91	0.91	0.91	23
accuracy			0.92	50
macro avg	0.92	0.92	0.92	50
weighted avg	0.92	0.92	0.92	50

MULTI-LAYER PERCEPTRON - CLASS DATA

Predicting Membership Status Using Employment, Visit Frequency, and Time Spent Shopping

```python
clf = MLPClassifier(hidden_layer_sizes=(4, 2),max_iter=1000)

clf.fit(X_trainarray, Y_train.ravel())

Y_pred = clf.predict(X_testarray)
print(Y_pred)
print(Y_test.T)


accuracy = accuracy_score(Y_test, Y_pred)
print(f"Accuracy: {accuracy*100:.2f}%")

class_report = classification_report(Y_test, Y_pred)
print("Classification Report:\n", class_report)
```

```
['no' 'no' 'no' 'no' 'yes' 'yes']
[['yes' 'no' 'yes' 'no' 'no' 'yes']]
Accuracy: 50.00%
Classification Report:
              precision    recall  f1-score   support

          no       0.50      0.67      0.57         3
         yes       0.50      0.33      0.40         3

    accuracy                           0.50         6
   macro avg       0.50      0.50      0.49         6
weighted avg       0.50      0.50      0.49         6
```

Accuracy: 50.00
Low quality due to low amounts of data to train on

MULTI-LAYER PERCEPTRON - SYNTHETIC DATA

Predicting Membership Status Using Employment, Visit Frequency, Age Group, and Time Spent Shopping

```python
clf = MLPClassifier(solver='lbfgs', alpha=1e-5,
                    hidden_layer_sizes=(40, 20), max_iter=1000)

clf.fit(X_trainarray, Y_train.ravel())

Y_pred = clf.predict(X_testarray)
print(Y_pred)
print(Y_test.T)


accuracy = accuracy_score(Y_test, Y_pred)
print(f"Accuracy: {accuracy*100:.2f}%")

class_report = classification_report(Y_test, Y_pred)
print("Classification Report:\n", class_report)
```

```
['No' 'Yes' 'No' 'Yes' 'No' 'No' 'No' 'Yes' 'Yes' 'Yes' 'No' 'Yes' 'No'
 'No' 'Yes' 'No' 'Yes' 'Yes' 'No' 'Yes' 'Yes' 'No' 'No' 'No' 'No' 'Yes'
 'Yes' 'Yes' 'Yes' 'No' 'No' 'Yes' 'Yes' 'No' 'Yes' 'No' 'No' 'Yes' 'No'
 'Yes' 'No' 'No' 'Yes' 'Yes' 'No' 'Yes' 'No' 'No' 'No' 'No']
[['No' 'Yes' 'No' 'Yes' 'No' 'No' 'Yes' 'Yes' 'Yes' 'Yes' 'No' 'Yes' 'No'
  'No' 'Yes' 'No' 'Yes' 'Yes' 'No' 'Yes' 'Yes' 'No' 'No' 'No' 'No' 'Yes'
  'Yes' 'No' 'Yes' 'No' 'No' 'Yes' 'Yes' 'No' 'Yes' 'No' 'No' 'Yes' 'No'
  'Yes' 'Yes' 'No' 'Yes' 'Yes' 'No' 'Yes' 'No' 'No' 'Yes' 'No']]
Accuracy: 92.00%
Classification Report:
              precision    recall  f1-score   support

          No       0.89      0.96      0.92        25
         Yes       0.96      0.88      0.92        25

    accuracy                           0.92        50
   macro avg       0.92      0.92      0.92        50
weighted avg       0.92      0.92      0.92        50
```

Accuracy: 92.00
Larger amount of training data allows for more robust models

CONCLUSIO

CONCLUSION

- No statistically significant relationship between demographic and membership status

- Statistically significant relationship between membership status and money spent per transaction
    - Further research to be done on value for goods, market basket, etc.

- The more time people spent in stores, the more they spent

- K-means clustering revealed high frequency shoppers spend more money
    - Includes money spent on Costco membership as well as grocery transactions

- Logistic regression and a multi-layer perceptron showed weak results for the survey data due to low number of responses. In the synthetic data, they showed high accuracy, similar accuracies between models

LIMITATIONS & FUTURE RESEARCH

LIMITATIONS

- Narrowed data collection by focusing on our class specifically
  - Larger, more robust student types and demographics (i.e. freshmen, first generation students, etc.) provide larger picture

- Short time frame reduced amount of data collection

- Only students present submitted forms

FUTURE RESEARCH

- By paying for the membership, validity of Costco shoppers' purchase value for goods, market basket, etc.

- Quality of high frequency shopper basket
  - Possible connections to overconsumption

- Possible connection between increasing rates of inflation leading to more young adults enrolling in a Costco membership

- Even more research into shopping habits of students and the effect of student employment on consumer habits
  - Expansion into fields beyond groceries like hobbies, dining out, etc.

- Worldwide student trends of consumerism +

REFERENCES

- Bai, Junfei, et al. "Consumer Choice of Retail Food Store Formats in Qingdao, China." Journal of International Food & Agribusiness Marketing, vol. 20, no. 2, 2008, pp. 89–109, https://doi.org/10.1080/08974430802186217.

- Bobbitt, Zach. "K-Means Clustering in Python: Step-by-Step Example." Scientific Blog. Statology, https://www.statology.org/k-means-clustering-in-python/. Accessed 19 Nov. 2025.

- Chen, Jiangpei. Marketing Strategy Management of Costco : Analysis and Comparison to S-Group. 2021, Theseus, https://www.theseus.fi/bitstream/handle/10024/504238/Chen_Jiangpei.pdf?sequence=2&isAllowed=y. Centria University of Applied Sciences.

- "Chi-Square Test of Independence." Informational Blog. Python for Data Science, https://www.pythonfordatascience.org/chi-square-test-of-independence-python/. Accessed 17 Nov. 2025.

- Crissone, Chloe Danica. Costco: A Strategic Analysis. 2015, https://louis.uah.edu/cgi/viewcontent.cgi?article=1274&context=honors-capstones. University of Alabama, Honor's.

- Foust, Lyden. "Costco Brand Teardown: Seizing A Cultural Moment." Blog. Spatial.Ai, 21 Aug. 2025, https://www.spatial.ai/post/costco-brand-teardown-seizing-cultural-moment.

- Fox, Edward J., et al. "Consumer Shopping and Spending across Retail Formats." The Journal of Business, vol. 77, no. S2, 2004, pp. S25-60, https://doi.org/10.1086/381518.

REFERENCES

- Nilsson, Elin, Tommy Gärling, et al. "Who Shops Groceries Where and How? – The Relationship between Choice of Store Format and Type of Grocery Shopping." The International Review of Retail, Distribution and Consumer Research, vol. 25, no. 1, 2014, pp. 1–19, https://doi.org/10.1080/09593969.2014.940996.

- "One-Way ANOVA." Scientific blog. Python for Data Science, https://www.pythonfordatascience.org/anova-python/. Accessed 17 Nov. 2025.

- Prasad, Ch. Jayasankara. "Effect of Consumer Demographic Attributes on Store Choice Behaviour in Food and Grocery Retailing - An Empirical Analysis." Management and Labour Studies, vol. 35, no. 1, 2010, pp. 35–58, https://doi.org/10.1177/0258042X1003500103.

- Wholesale Corporation, Costco. "About Us." Retail. Costco.Com, https://www.costco.com/about.html.

- ——. "ANNUAL REPORT ON FORM 10-K FOR THE FISCAL YEAR ENDED SEPTEMBER 1, 2024." Costco Wholesale Corporation, 8 Oct. 2024.

? QUESTION