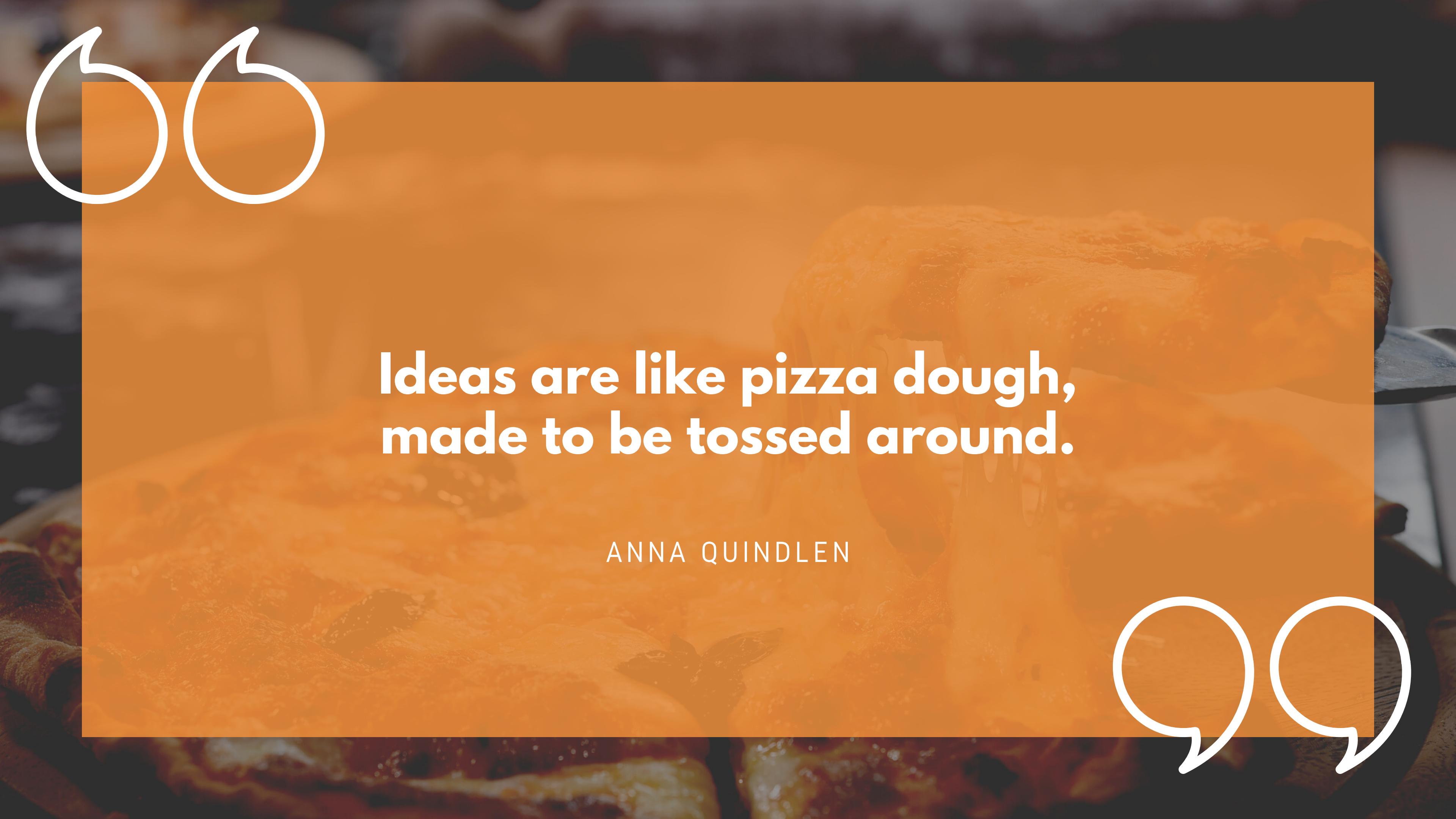




# A PIZZA THE PROFIT

Regression Analysis by Ashley Eakland



**Ideas are like pizza dough,  
made to be tossed around.**

ANNA QUINDLEN



# TOPICS AND HIGHLIGHTS

- Business Understanding
- Data Highlights
- Modeling Process
- Final Results
- Recommendations
- Future Analysis

# BUSINESS UNDERSTANDING

Thank you for trusting me to review and analyze your sales data!

Together, we can find the optimal way forward to increase net sales and customer base.





# DATA HIGHLIGHTS

## REAL WORLD DATA FROM LOCAL FRANCHISE

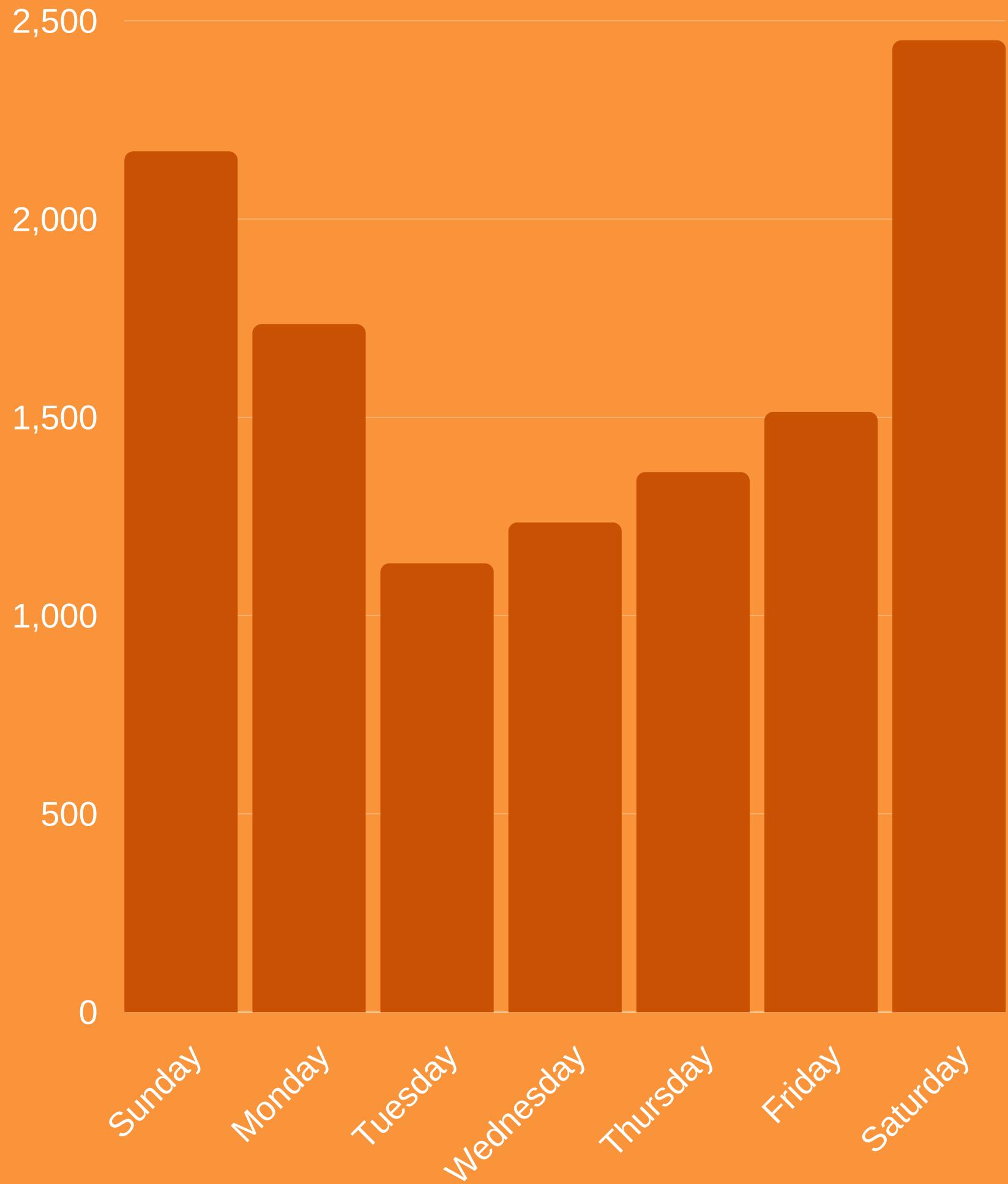
- 9 reports from sales system backend
  - Sales & Order History, Marketing Expenses
  - 5200 unique order numbers & 200+ discount codes
- Sales span August 2020 through Feb 2022
- PII has been scrubbed



# DATA HIGHLIGHTS

## EXPLORATORY DATA ANALYSIS

Extensive data wrangling and feature engineering required.



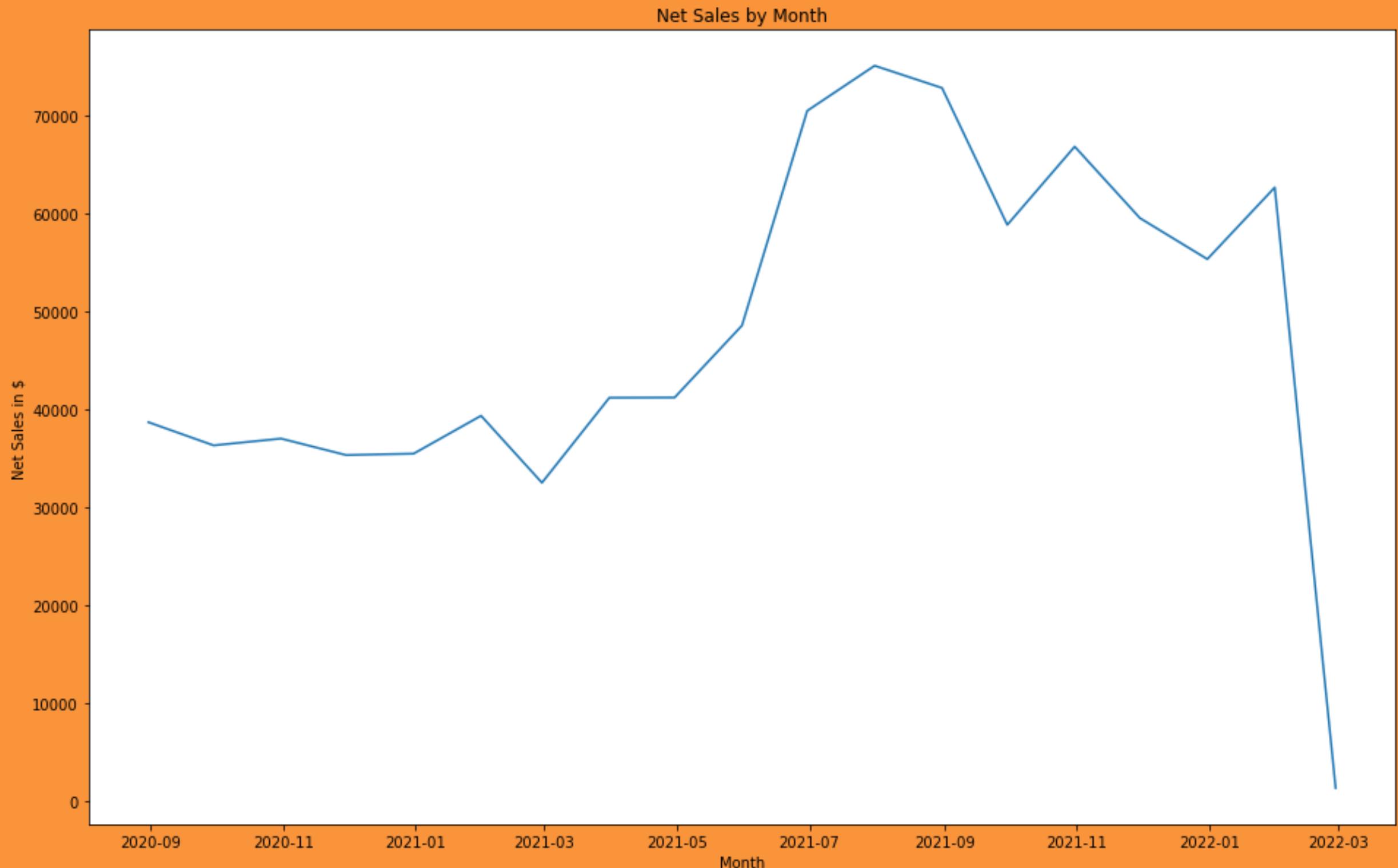
## SALES BY DAY

- Monday beat out Friday for third highest average net sales by day
- Midweek suffers lowest average sales

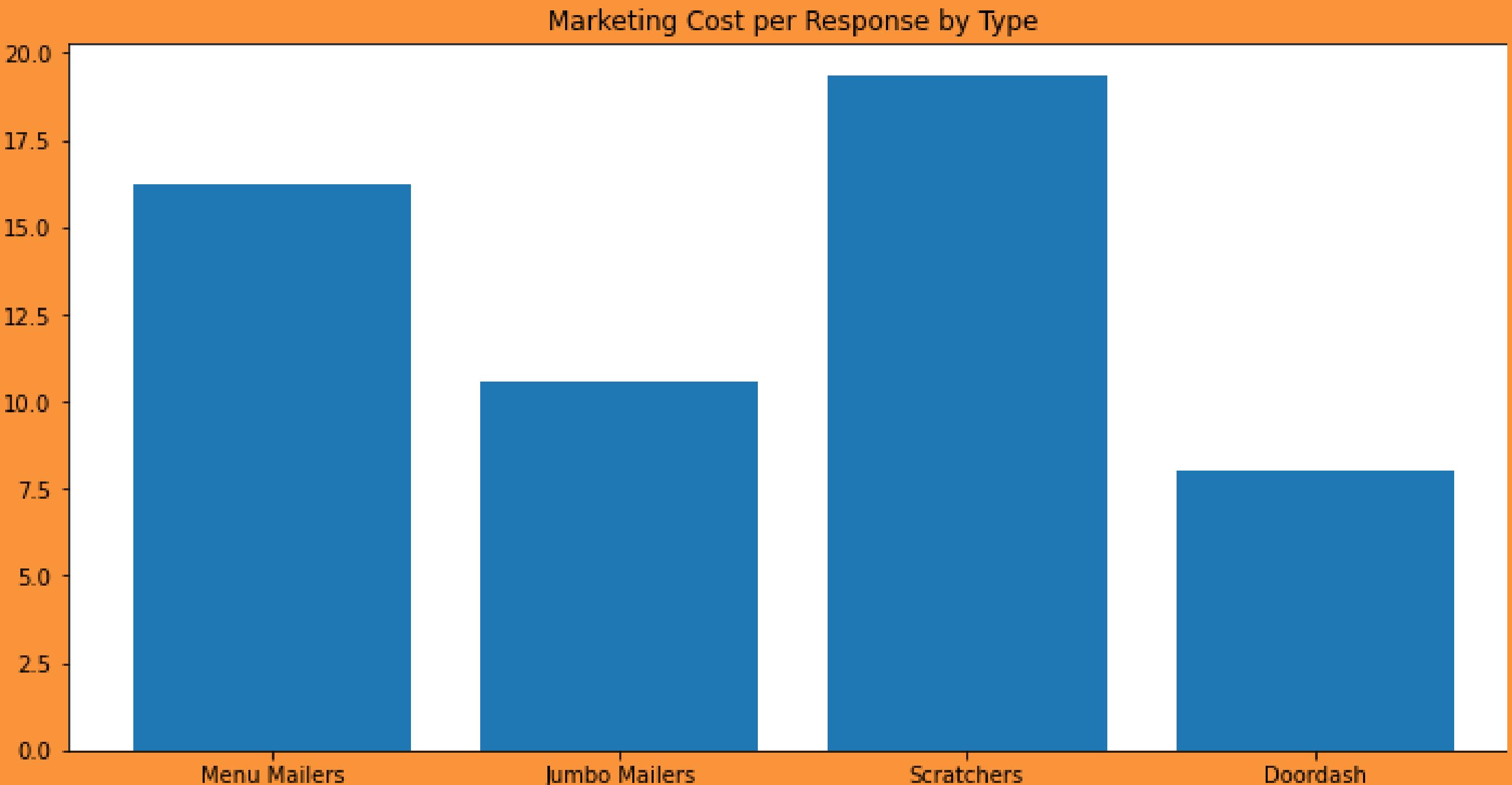
# DATA HIGHLIGHTS

## EDA CONTINUED

- SEASONALITY TREND OBSERVED



# DATA HIGHLIGHTS



EDA CONTINUED

OF MAILER  
MARKETING,  
JUMBO  
POSTCARDS  
LOWEST COST  
PER RESPONSE



\*DOORDASH AND FACEBOOK DIFFICULT TO QUANTIFY RESPONSE RATIO

# MODELING PROCESS



PREP

Clean and prepare data for  
modeling (lots of kneading)



BAKE

Train and tune models  
iteratively



EAT

Select top performer(s) and  
evaluate on holdout

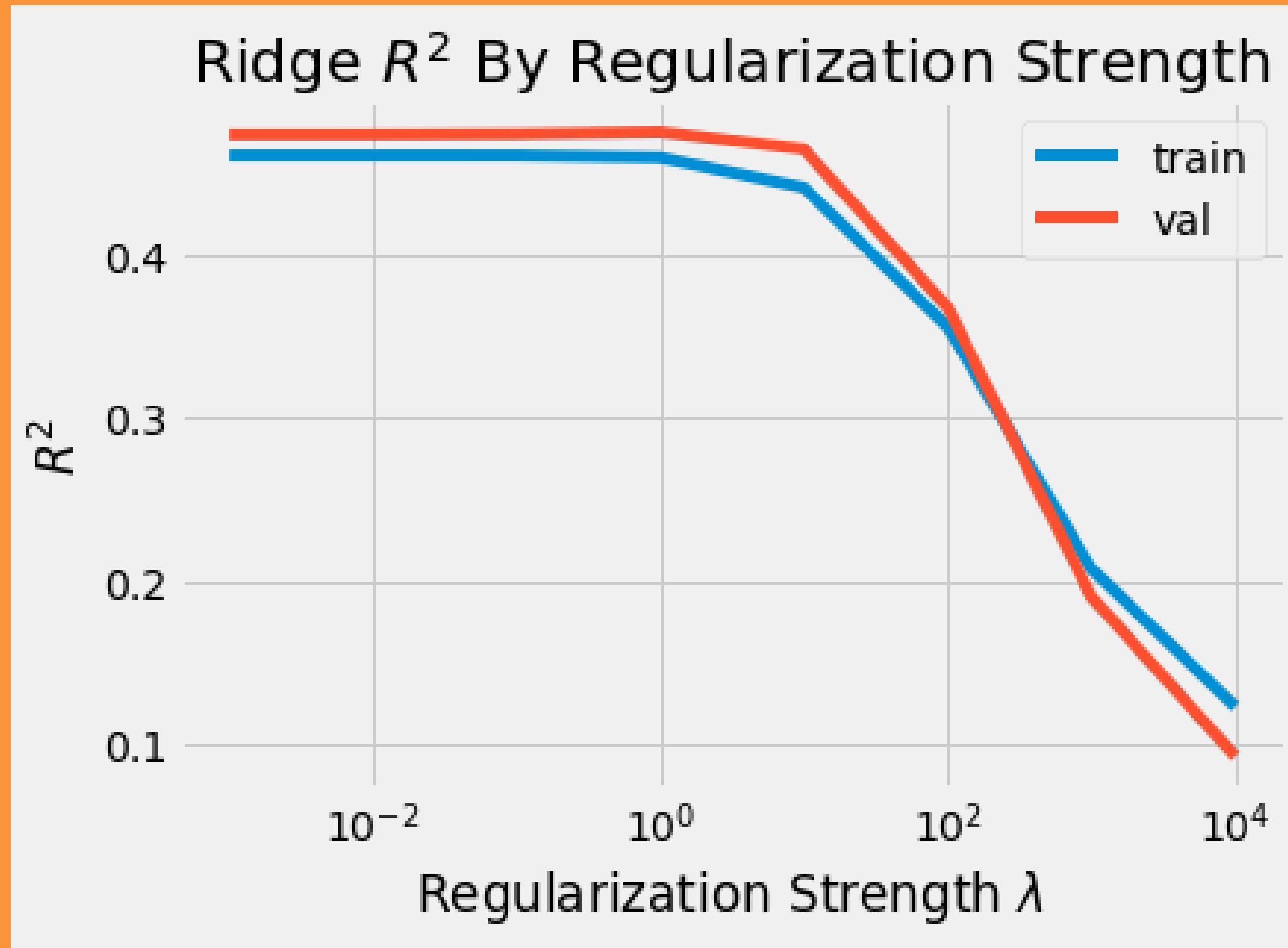
# FINDING OPTIMAL PARAMETERS



## METHODS

- Hyperparameter tuning
- GridSearch
- Visualizations

# MODEL TUNING



MAX R-SQUARED  
SCORE BASED ON  
LAMBDA REG  
STRENGTH



# **FINAL RESULTS**

TOP MODEL

# VOTING REGRESSOR



EXPLAINS 62% OF VARIANCE IN DATA

R-Squared score of .616 overall, 43 features

RMSE OF \$8.53

META ESTIMATOR - AVERAGES PREDICTIONS

Estimators used:

- Tuned RandomForest
- Tuned GradientBoost
- Tuned RidgeRegression

# Identifying Key Features



TOP THREE RIDGE COEFFICIENTS:  
LARGEFAV, SAVE20, GET25MILITARY



TOP ENSEMBLE FEATURES:  
TOTAL DISCOUNT AMOUNT, MAIL25,  
GET25EMPLOYEE



VOTINGREGRESSOR DOESN'T PICK

# LIMITATIONS

- Linear regression assumptions in violation
  - Model could be biased to underestimating
  - Coefficient values affected
- Additional research required



# RECOMMENDATIONS

DUE TO THE UNCERTAINTY SURROUNDING THE RELIABILITY OF THE MODEL PREDICTIONS, I CANNOT, IN GOOD FAITH, MAKE CONCRETE BUSINESS RECOMMENDATIONS AT THIS TIME.



# FUTURE ANALYSIS

## AGGRESSIVE BINNING OF CODES

- By type rather than individual codes

## ADDITIONAL ADJUSTMENTS

- Remove outliers
- Additional non-linear transformations

## IMPLEMENT NLP STRATEGIES

- Customer sentiment analysis



# QUESTIONS?

Please feel free to ask!



**If cauliflower can be pizza, then  
you, my friend, can be anything.**

UNKNOWN



ASHLEY EAKLAND

[ashley@idodata.science](mailto:ashley@idodata.science)

[www.linkedin.com/in/ashley-eakland](https://www.linkedin.com/in/ashley-eakland)

**THANK YOU!**

FOR TECHNICAL INFORMATION AND TO SEE THE  
JUPYTER NOTEBOOKS, PLEASE VISIT

[www.github.com/smashley-eakland/pizza-the-profit](https://www.github.com/smashley-eakland/pizza-the-profit)