Customer Acquisition Cost Prediction

→ How smart media planning helps growth

INSY 695 GROUP PROJECT

Group 4

4 Ps

4 Ps

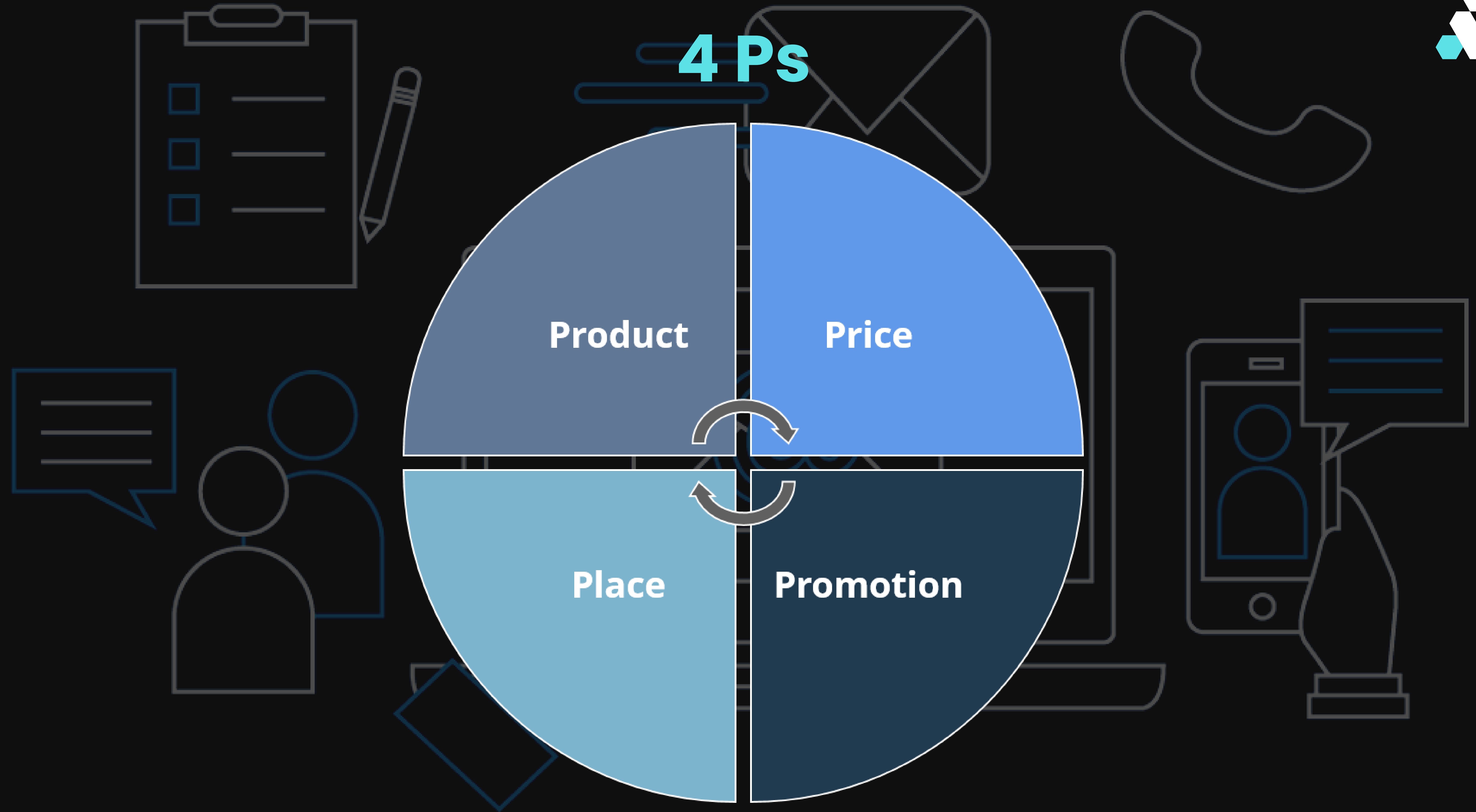
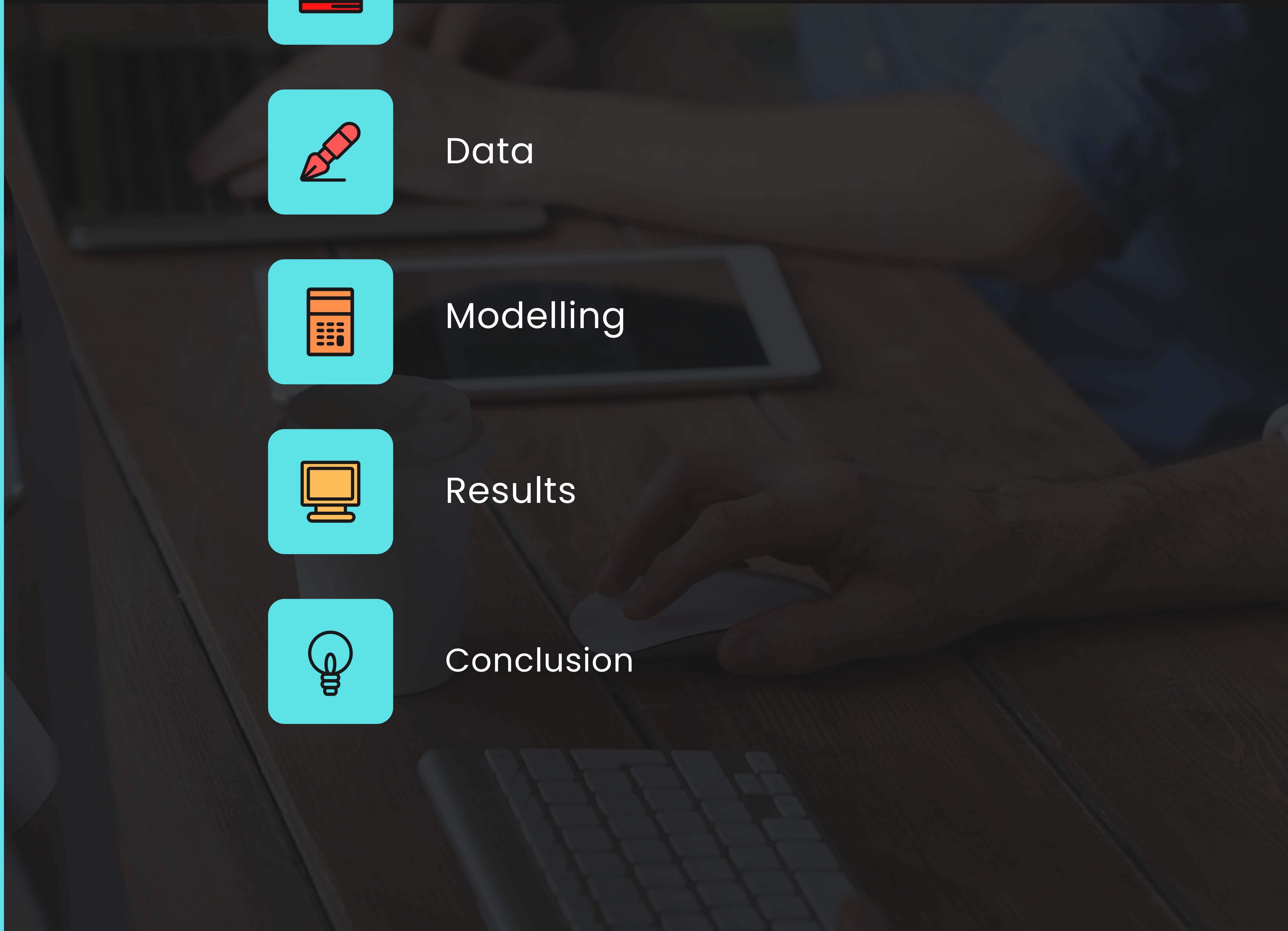


Table of Content



INSY 695 GROUP PROJECT

Who Are We?

Emery

Lead-Project Manager

Lucie

Lead-Business Analyst

Chongho

Data Analyst

Bennett

Data Scientist

Julie

Lead-Product Manager

Priyanka

Lead-Data Analyst

Raman

Lead-Data Scientist

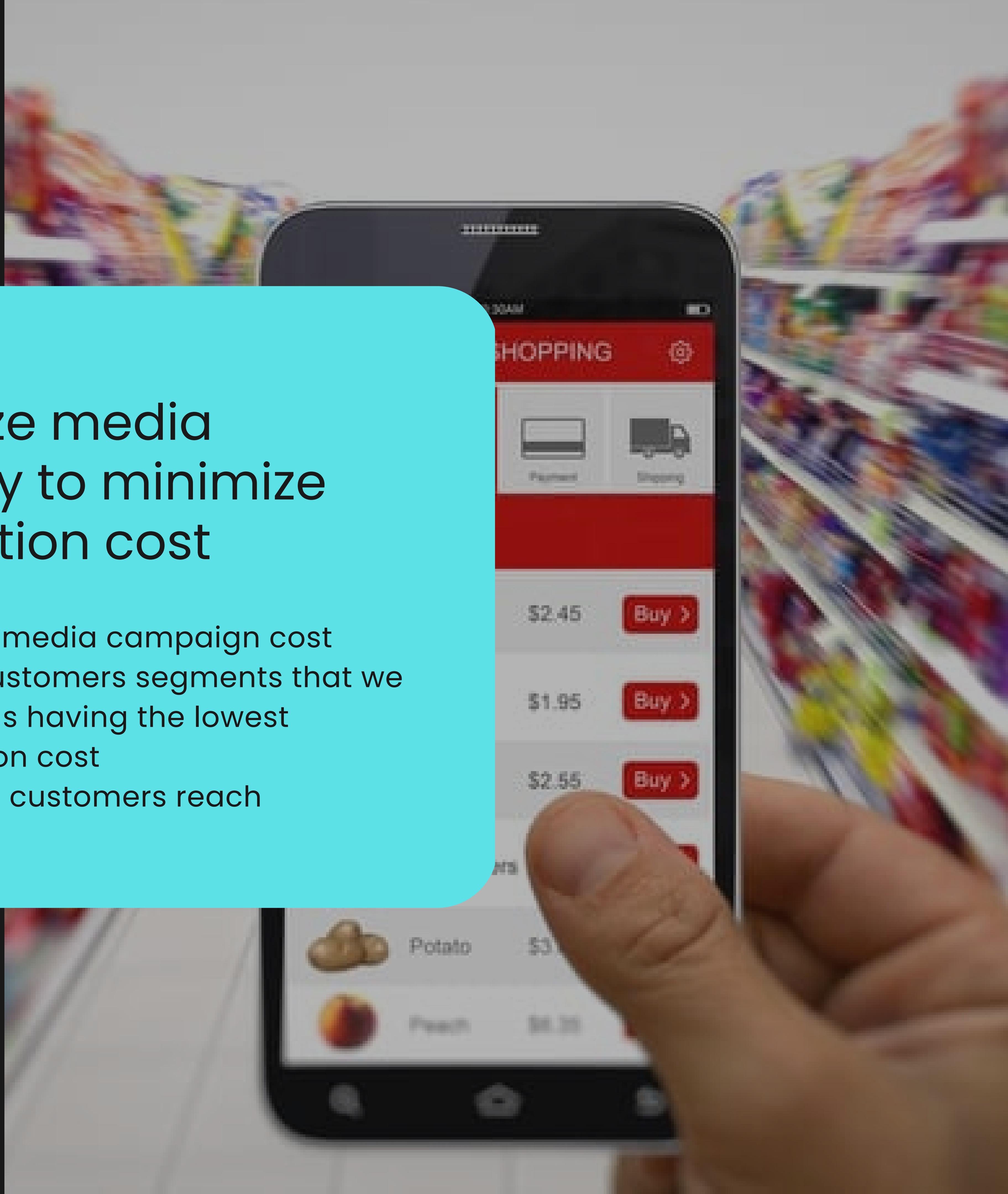
Use Case: Framing the Problem

Context - Problem Statement - Strategy

Context

- Post- COVID
- Race for attention
- Horizontal growth
  - International extension
  - +300 stores in Canada
- Vertical investments
  - Customer segmentation
  - Loyalty Programs

Problem Statement



Optimize media strategy to minimize acquisition cost

- Minimize media campaign cost
- Target customers segments that we identify as having the lowest acquisition cost
- Maximize customers reach

USE CASE

Strategy

Customer segmentation

- Identify customer features
- Investigate our customer segmentation

Predict customer acquisition cost

- Defined the relevant factors in our predictions
- Identify cost of acquisition per segment

Lower cost of acquisition

- Create a strategy around an optimal customer cost of acquisition and segment to grow our buisness

Data

Acquisition – Preparation – Exploration

Data Aquisition

- Customer's acquisition Cost

- Based on more than 60k observations
  - Customers' demographic data (gender, marital status...)
  - Stores' data (sales, store's type)
  - Media campaigns' data (type, acquisition cost)

Data Preparation Steps

Some Key Steps

Key Assumptions:
Cost - The Customer Acquisition Cost
Media Type- Type of media used for advertising. encoded by label encoding

NAs

There were no detectable missing or null values in the datasets.

Removed rows: 0

Outlier Detection

Using an isolation forest we identified and removed outliers from the datasets

Removed rows: 3,005

Feature Scaling

Multi-collinearity distorts model performance. By scaling features we can reduce this effect.

Synthesized Attributes

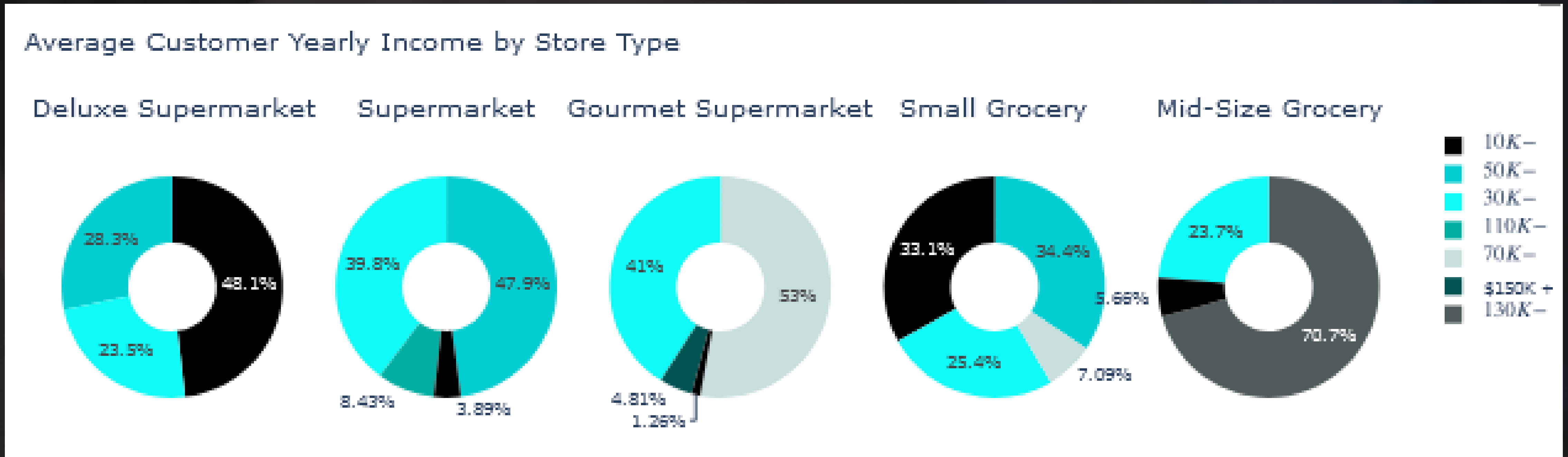
Attribute whose parse tree node value is determined by the attribute value at child nodes.

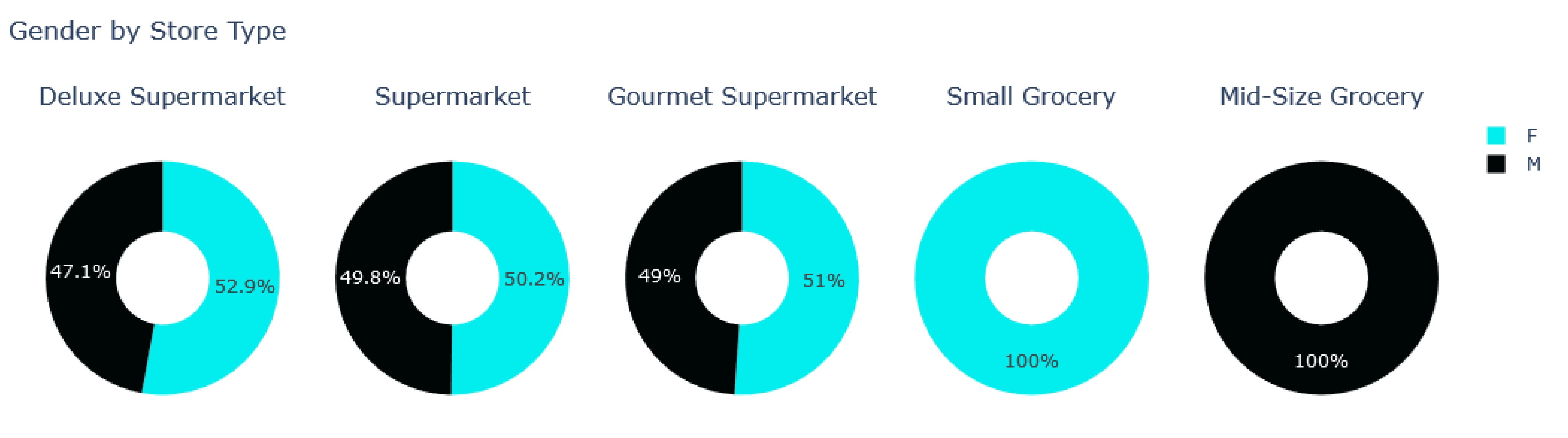
Dummification

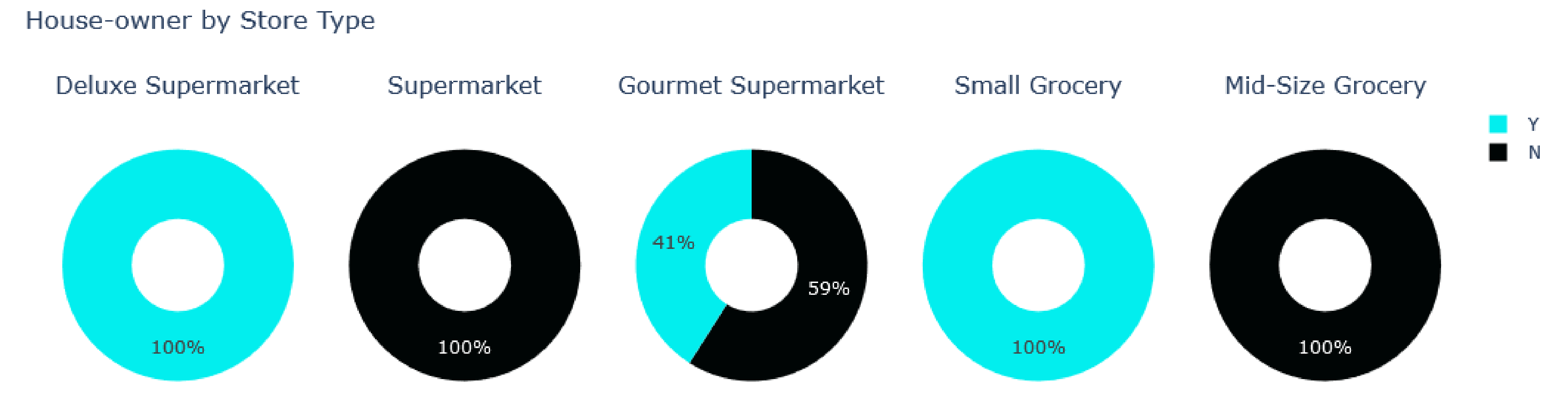
Any Categorical Variables were dummified based on one hot encoding

Data Exploration

Customer Segments

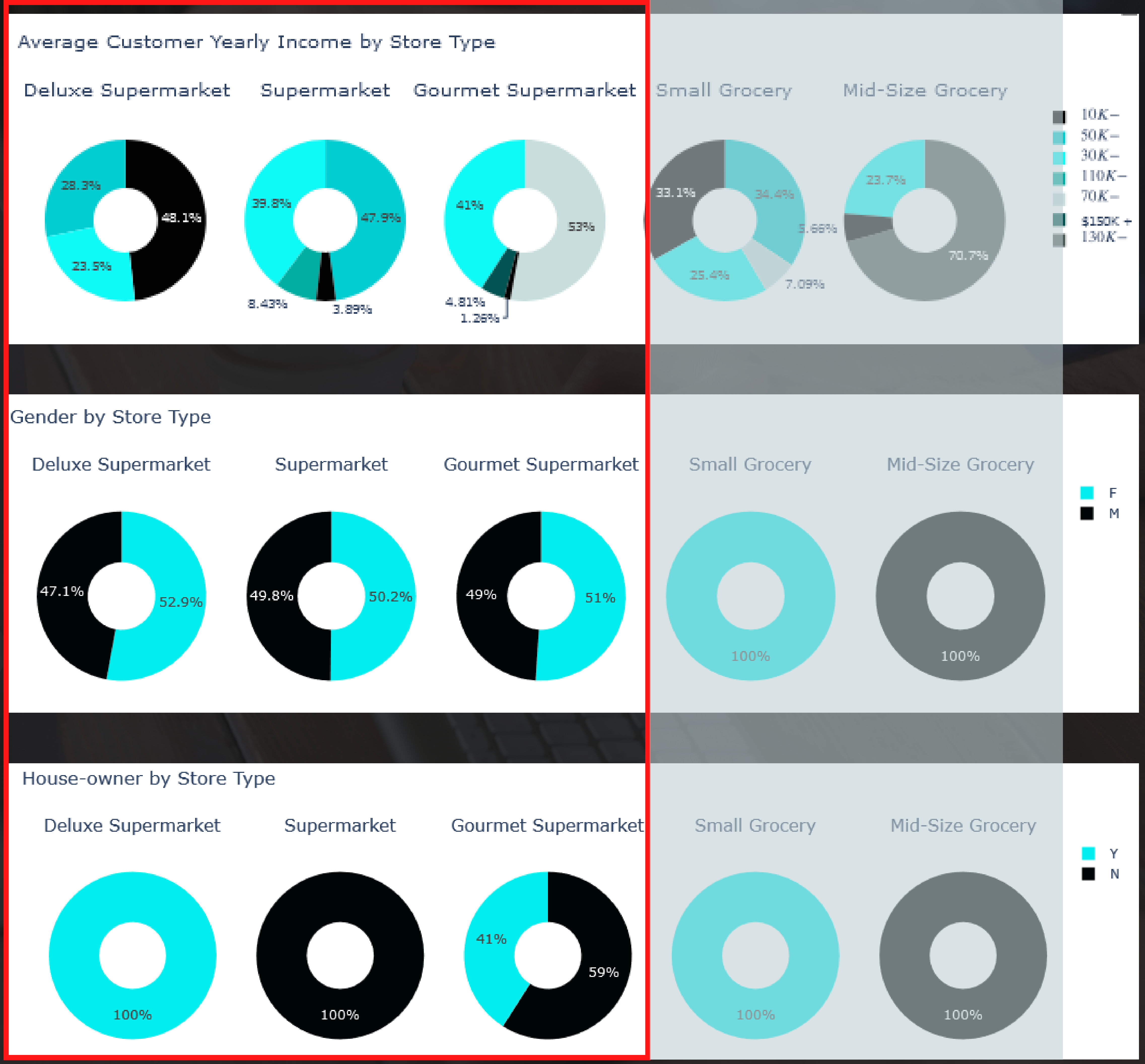






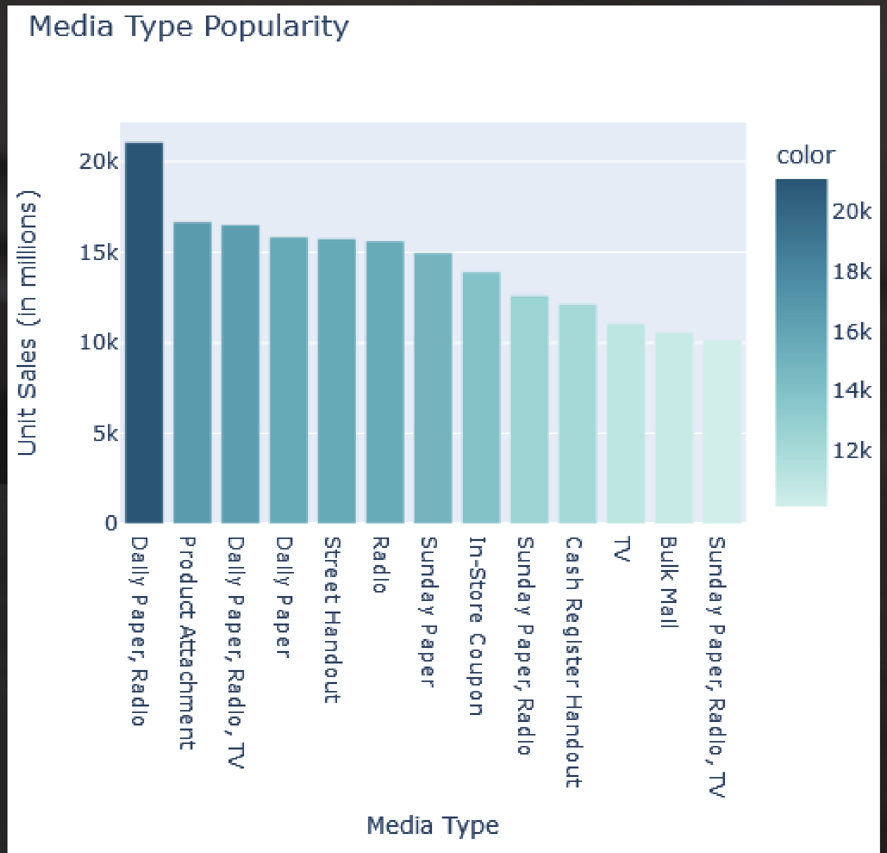
Data Exploration

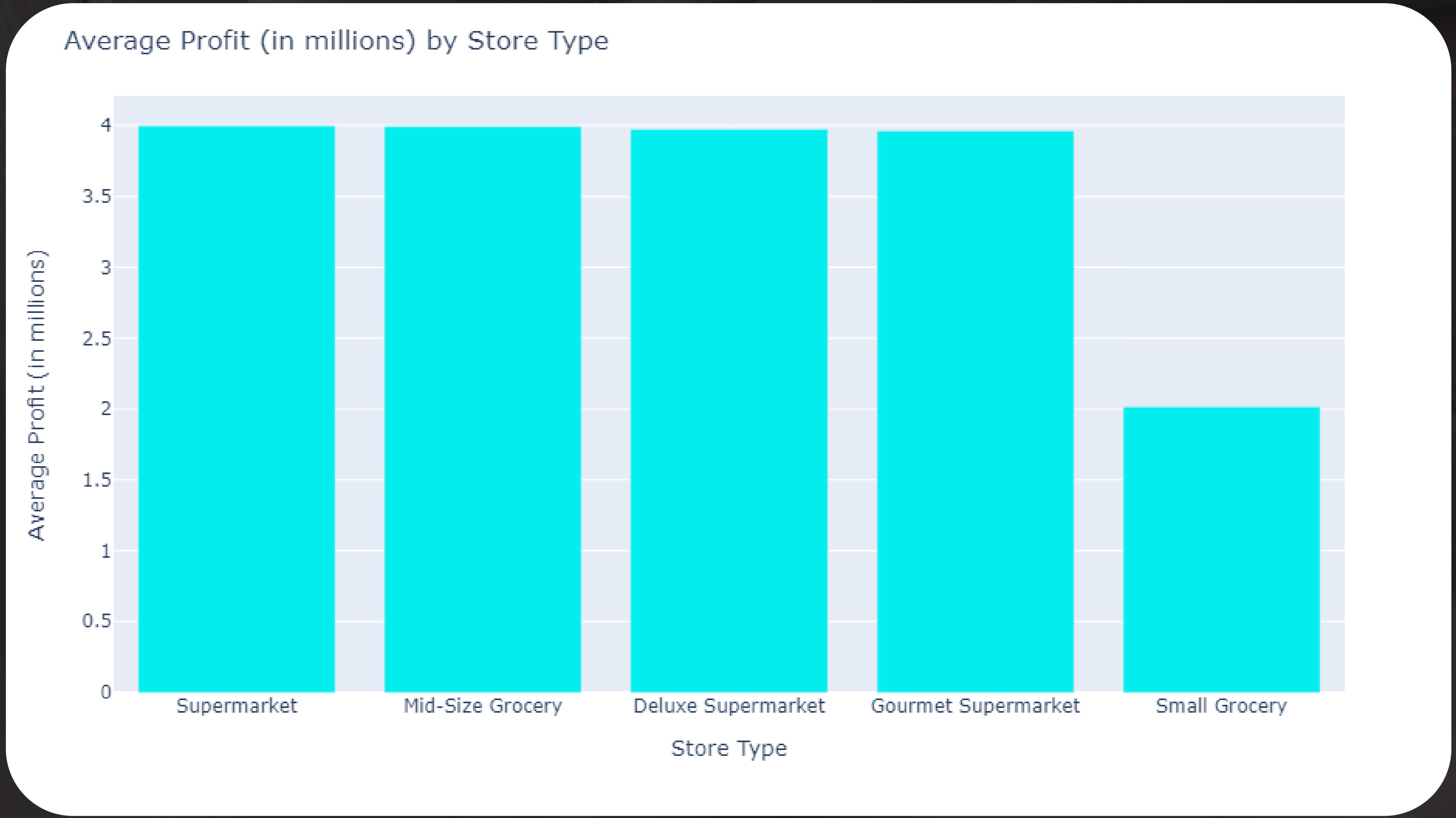
Customer Segments



Data Exploration

Stores and Media segment

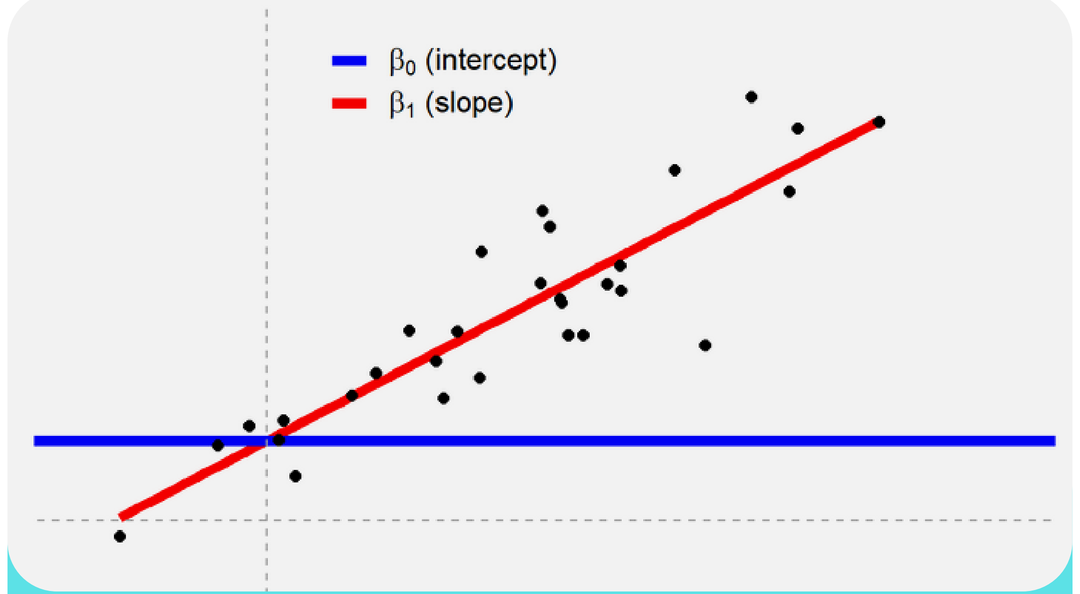




Modeling Strategy

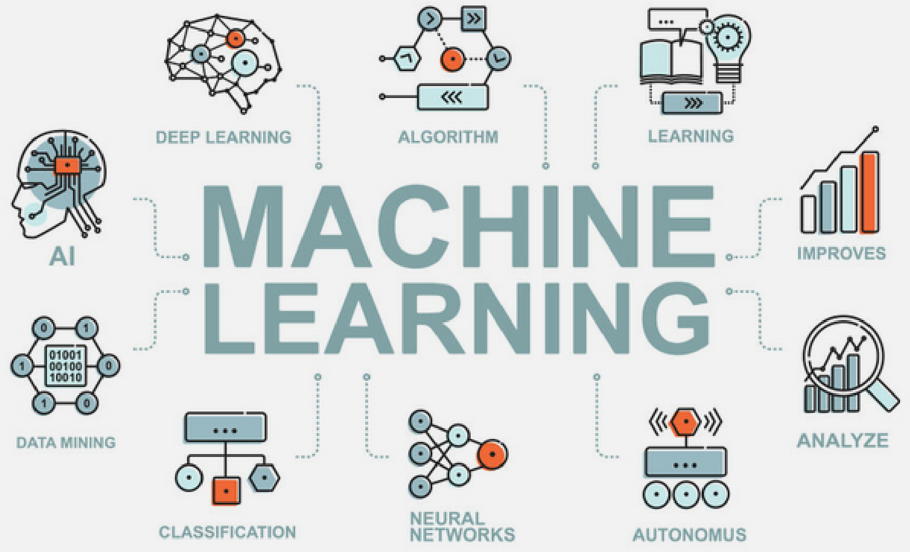
Approach - Evaluation - Selection

Modeling Strategies



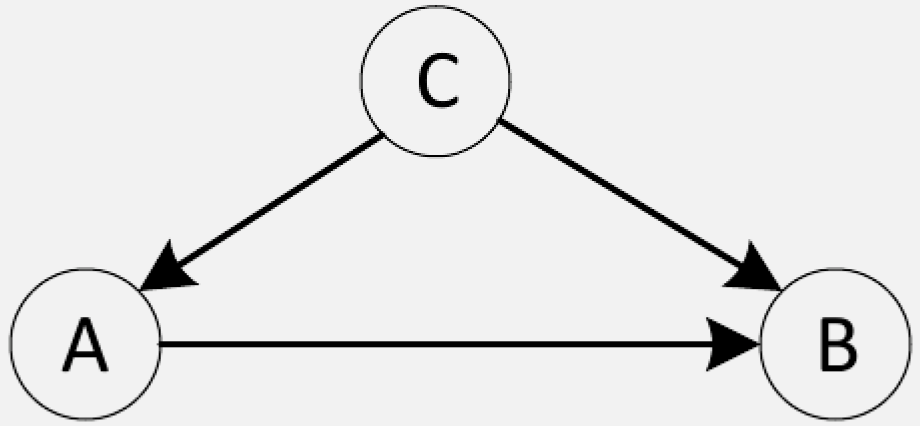
Linear Models

We used a dummy regressor to establish a baseline and understand some factors that predict cost the error.



Machine Learning

-Tree-Based Methods
-XGBoost
-Neural Networks
- & More



Causal Inference

We want to examine the effect of human metrics on costs.

Objective: Reduce the cost of acquisition of our customers

Causal Inference

	Effect of Gender	Effect of Loyalty	Effect of Media Type
Outcome Variable	Cost	Cost	Cost
Treatment	Gender	Member Card	Media Type
Control Variables	All Other Variables	All Other Variables	All Other Variables

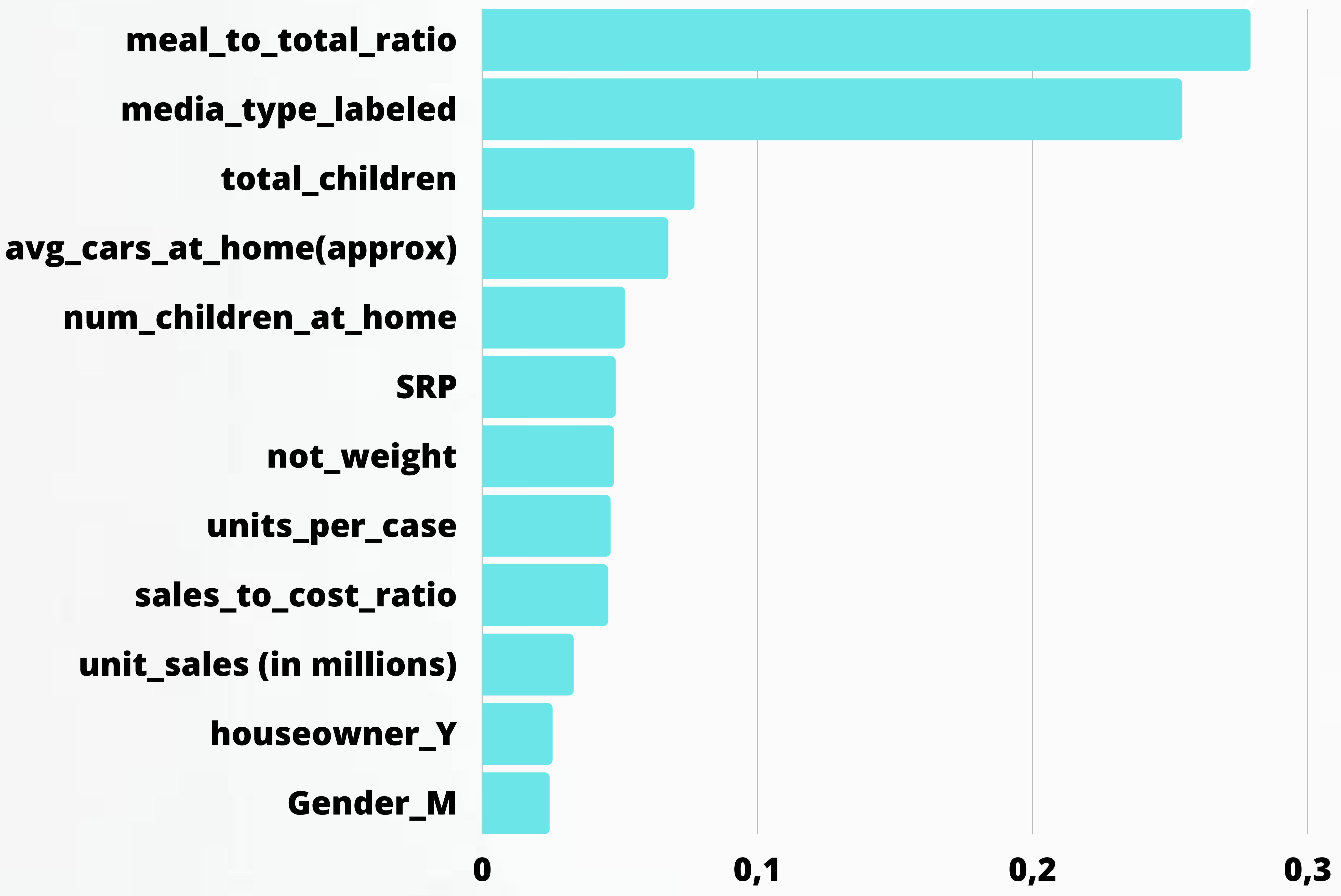
Evaluation & Selection

Model Accuracy (MSE)

Results

Interpretation - Insights

Feature significance



Key Insights

→ **Less meat**

→ **Media cost**

Conclusion

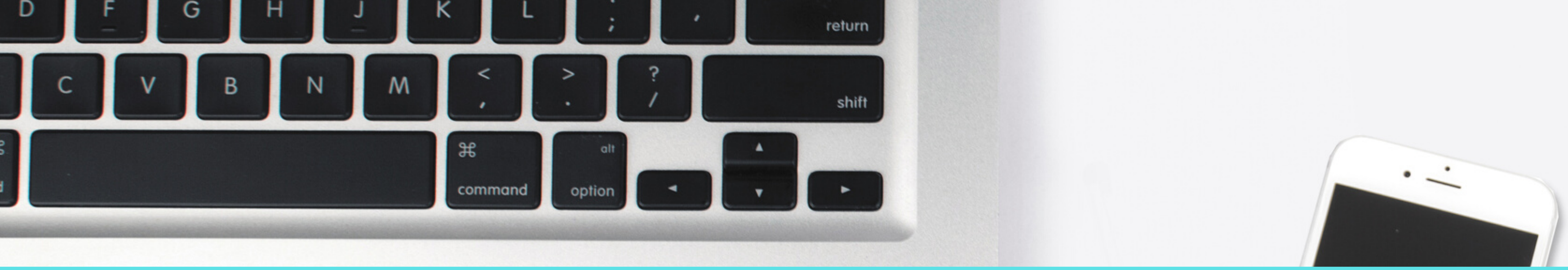
Lessons Learned - Next Step

Appendix

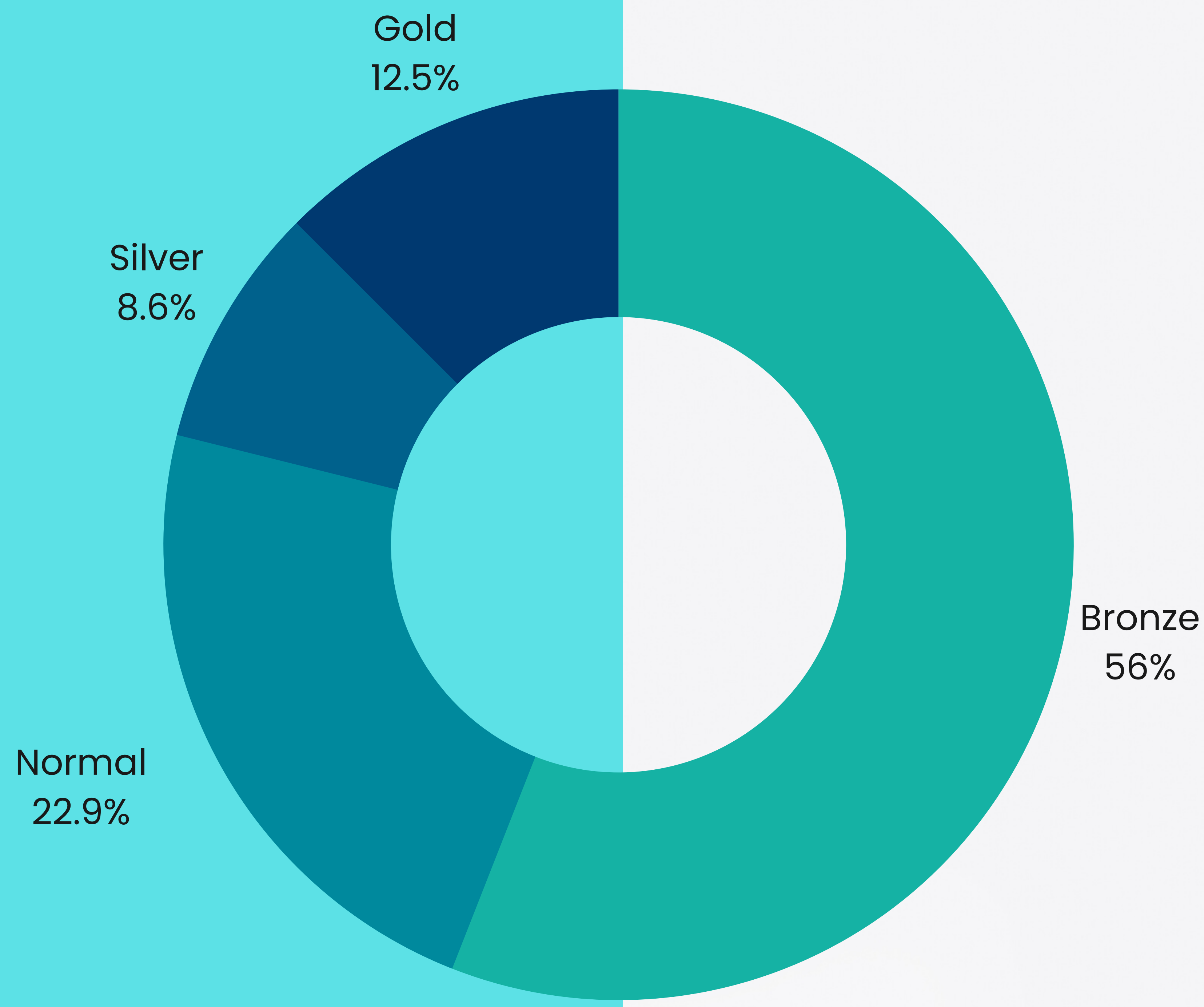
Data Structure

Column	Name	Count		Type
---	------	-----	---------	-----
0	food_category	60428	non-null	object
1	food_department	60428	non-null	object
2	food_family	60428	non-null	object
3	store_sales(in millions)	60428	non-null	float64
4	store_cost(in millions)	60428	non-null	float64
5	unit_sales(in millions)	60428	non-null	float64
6	promotion_name	60428	non-null	object
7	sales_country	60428	non-null	object
8	marital_status	60428	non-null	object
9	gender	60428	non-null	object
10	total_children	60428	non-null	float64
11	education	60428	non-null	object
12	member_card	60428	non-null	object
13	occupation	60428	non-null	object
14	houseowner	60428	non-null	object
15	avg_cars_at home(approx)	60428	non-null	float64
16	avg. yearly_income	60428	non-null	object
17	num_children_at_home	60428	non-null	float64
18	avg_cars_at home(approx).1	60428	non-null	float64
19	brand_name	60428	non-null	object

	Column	Name	Count		Type
	20	SRP	60428	non-null	float64
	21	gross_weight	60428	non-null	float64
	22	net_weight	60428	non-null	float64
	23	recyclable_package	60428	non-null	float64
	24	low_fat	60428	non-null	float64
	25	units_per_case	60428	non-null	float64
	26	store_type	60428	non-null	object
	27	store_city	60428	non-null	object
	28	store_state	60428	non-null	object
	29	store_sqft	60428	non-null	float64
	30	grocery_sqft	60428	non-null	float64
	31	frozen_sqft	60428	non-null	float64
	32	meat_sqft	60428	non-null	float64
	33	coffee_bar	60428	non-null	float64
	34	video_store	60428	non-null	float64
	35	salad_bar	60428	non-null	float64
	36	prepared_food	60428	non-null	float64
	37	florist	60428	non-null	float64
	38	media_type	60428	non-null	object
	39	cost	60428	non-null	float64



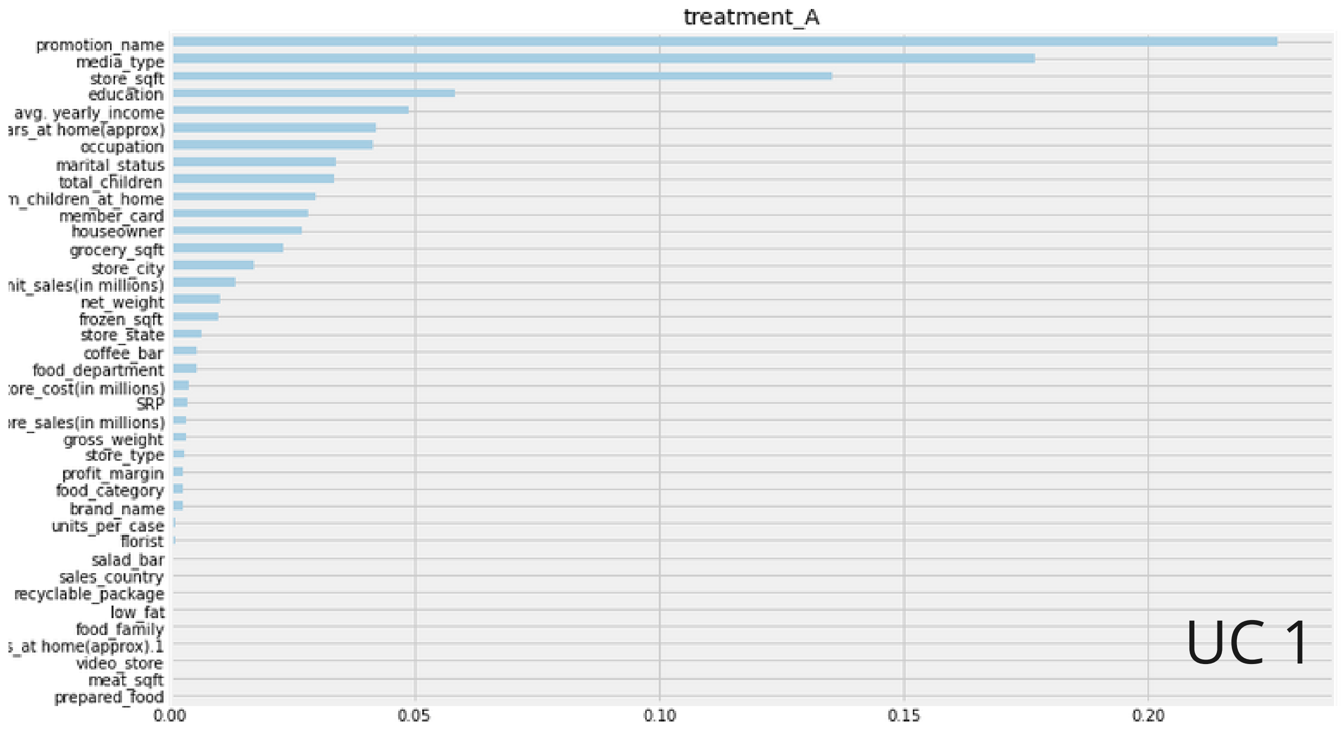
Loyalty Program

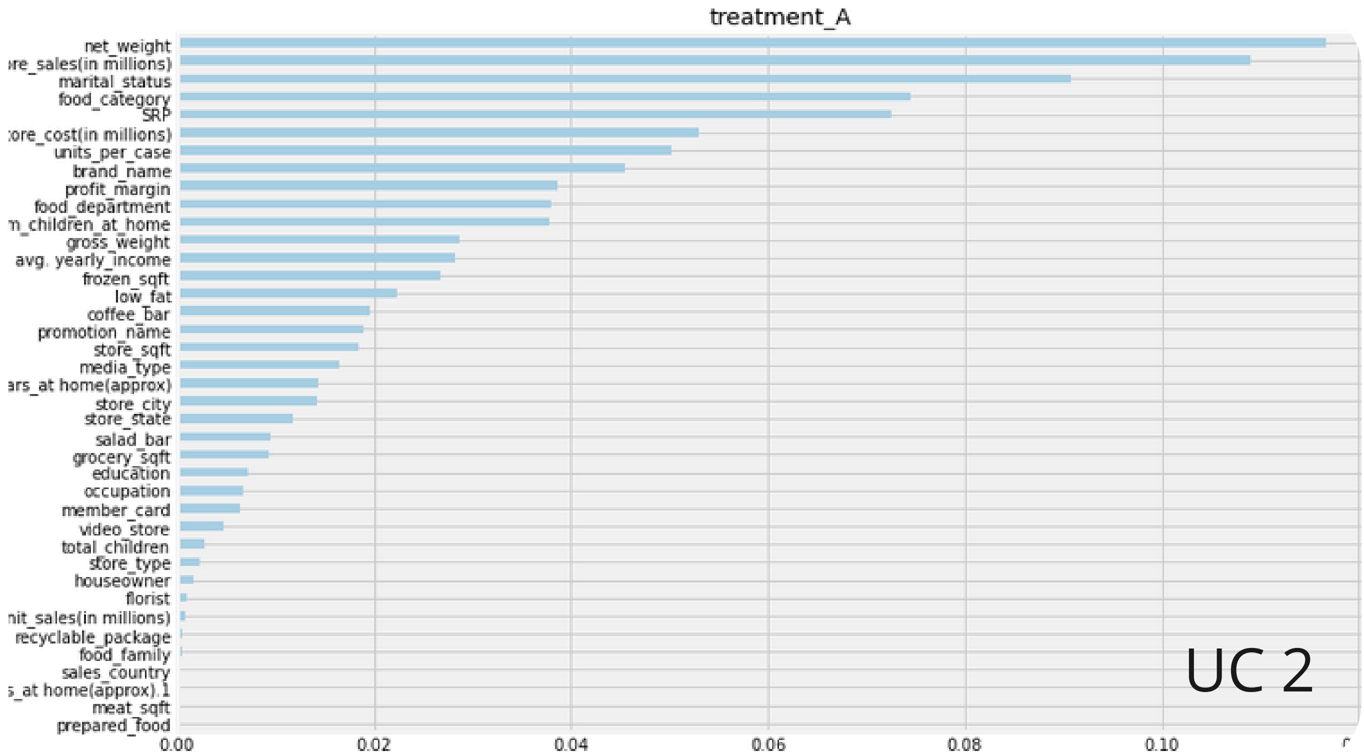


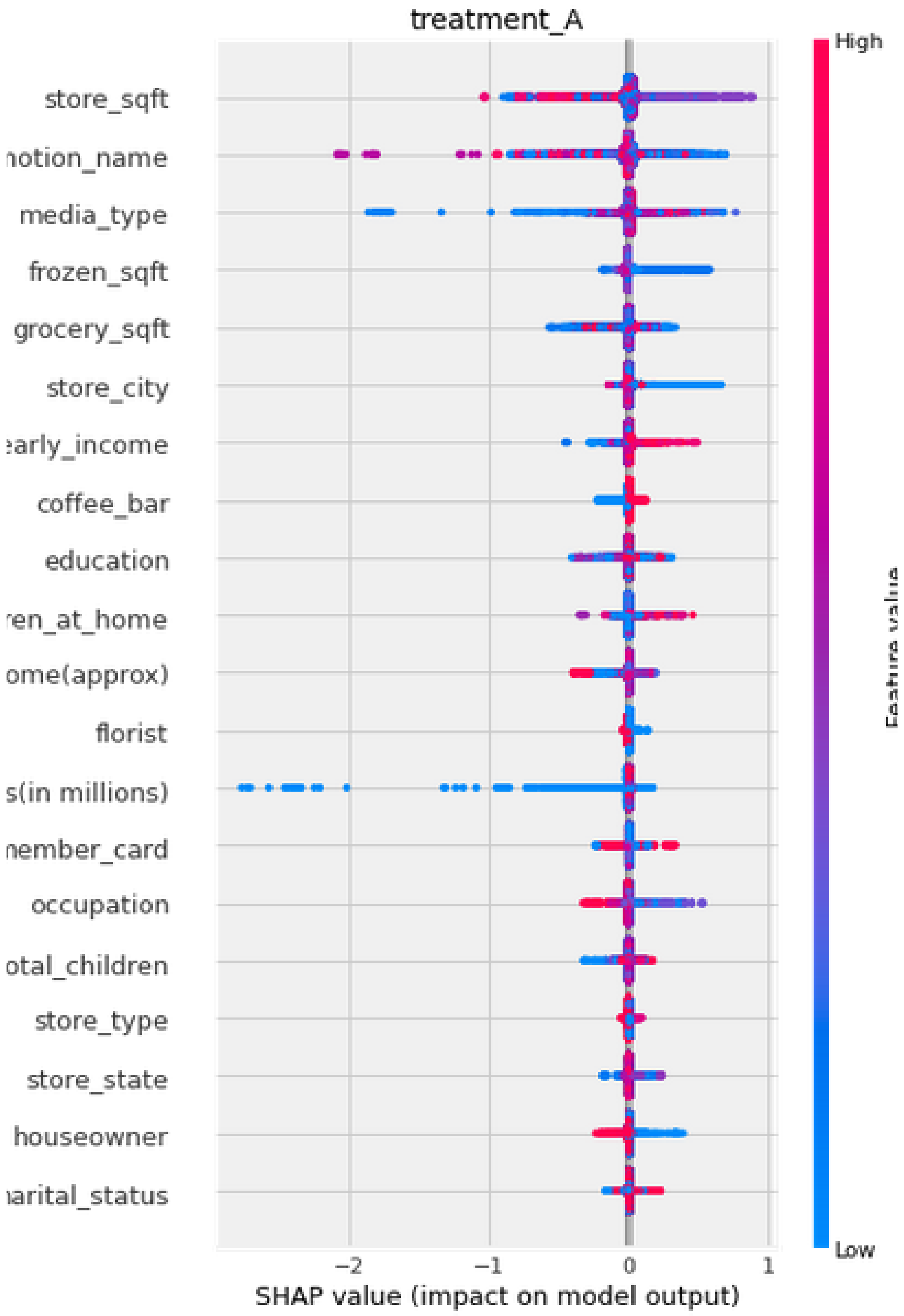
Distribution of cutomers' membership status

77% of customers have enrolled in the loyalty program

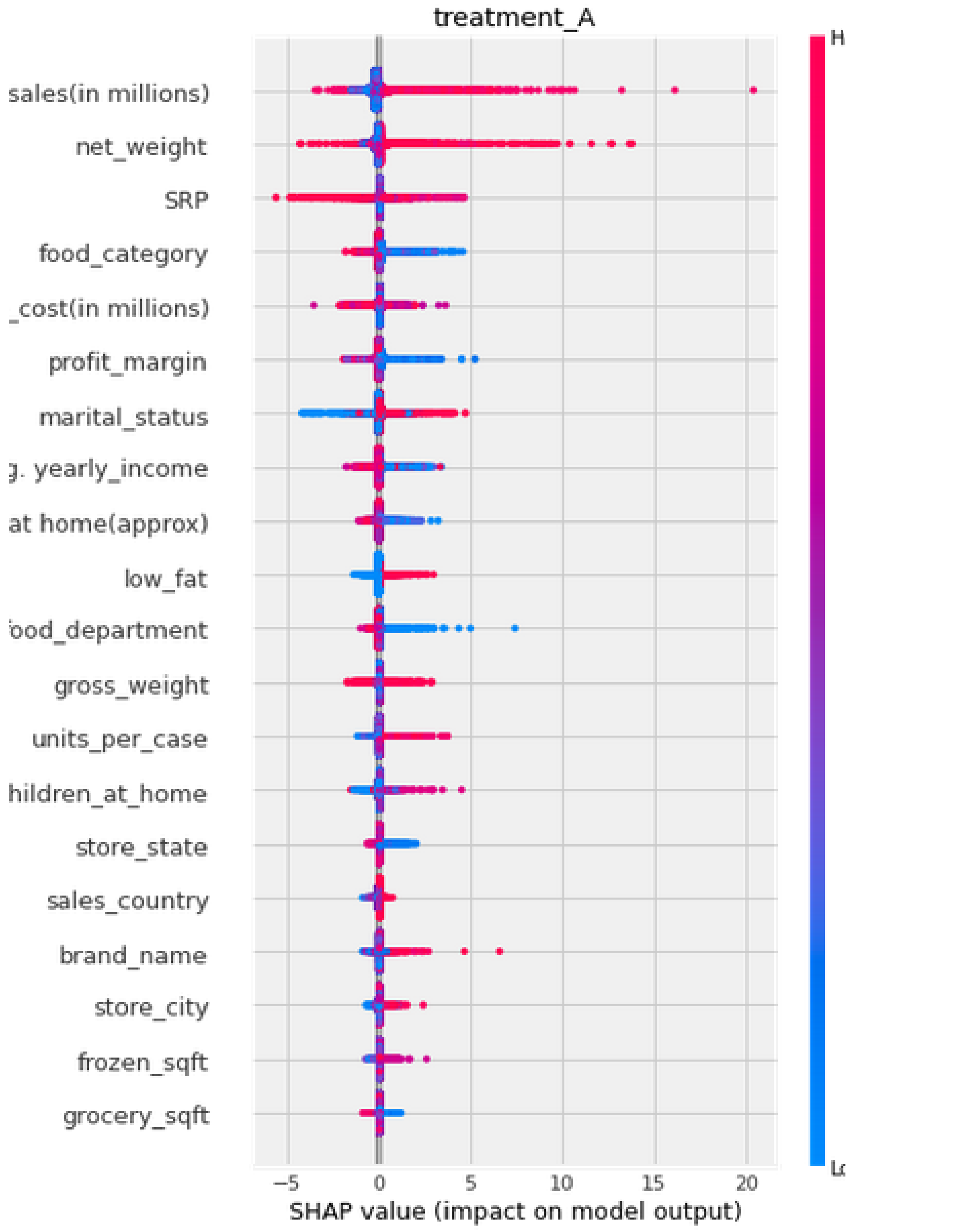
Causal Inference





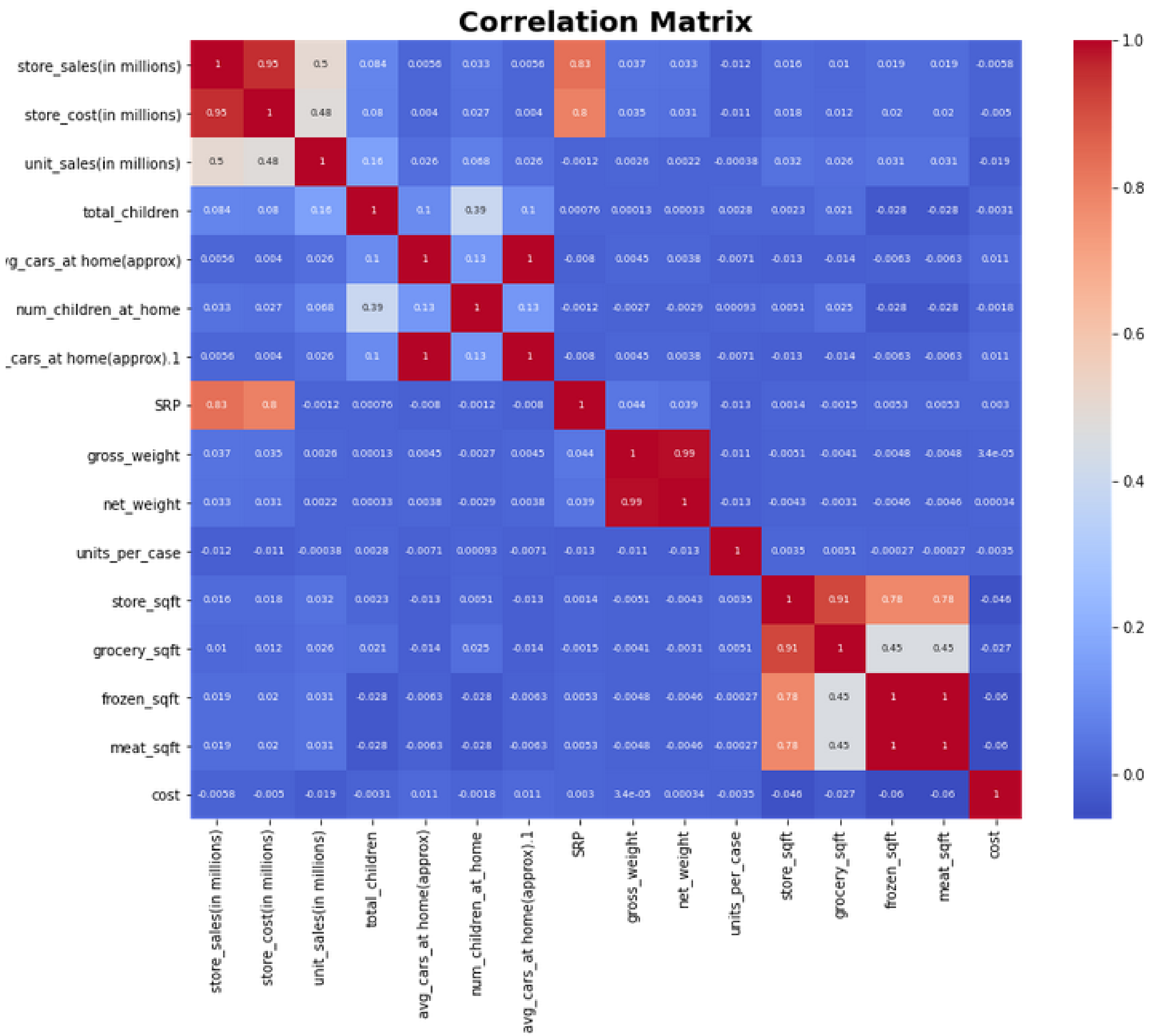


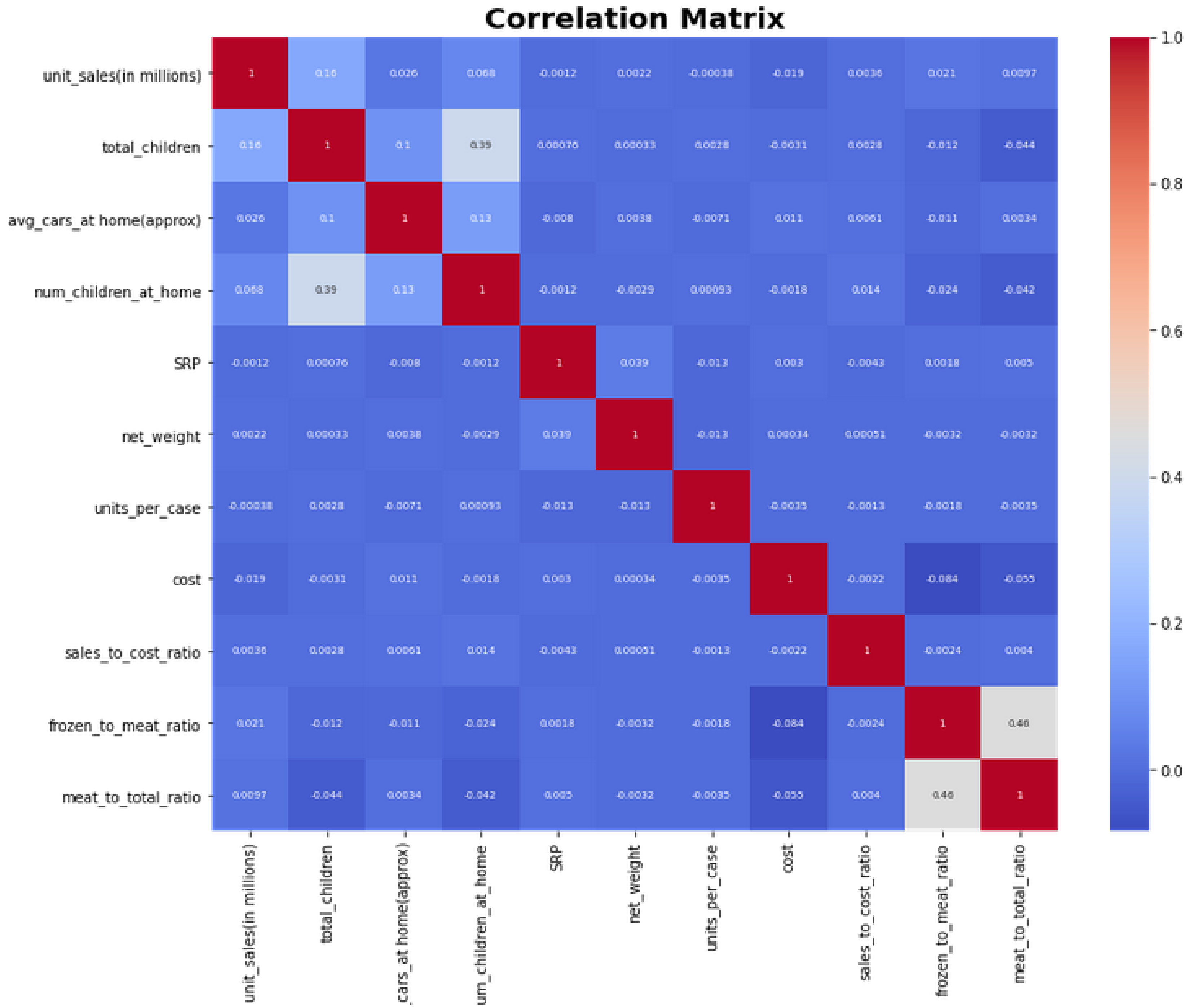
UC 1: Gender



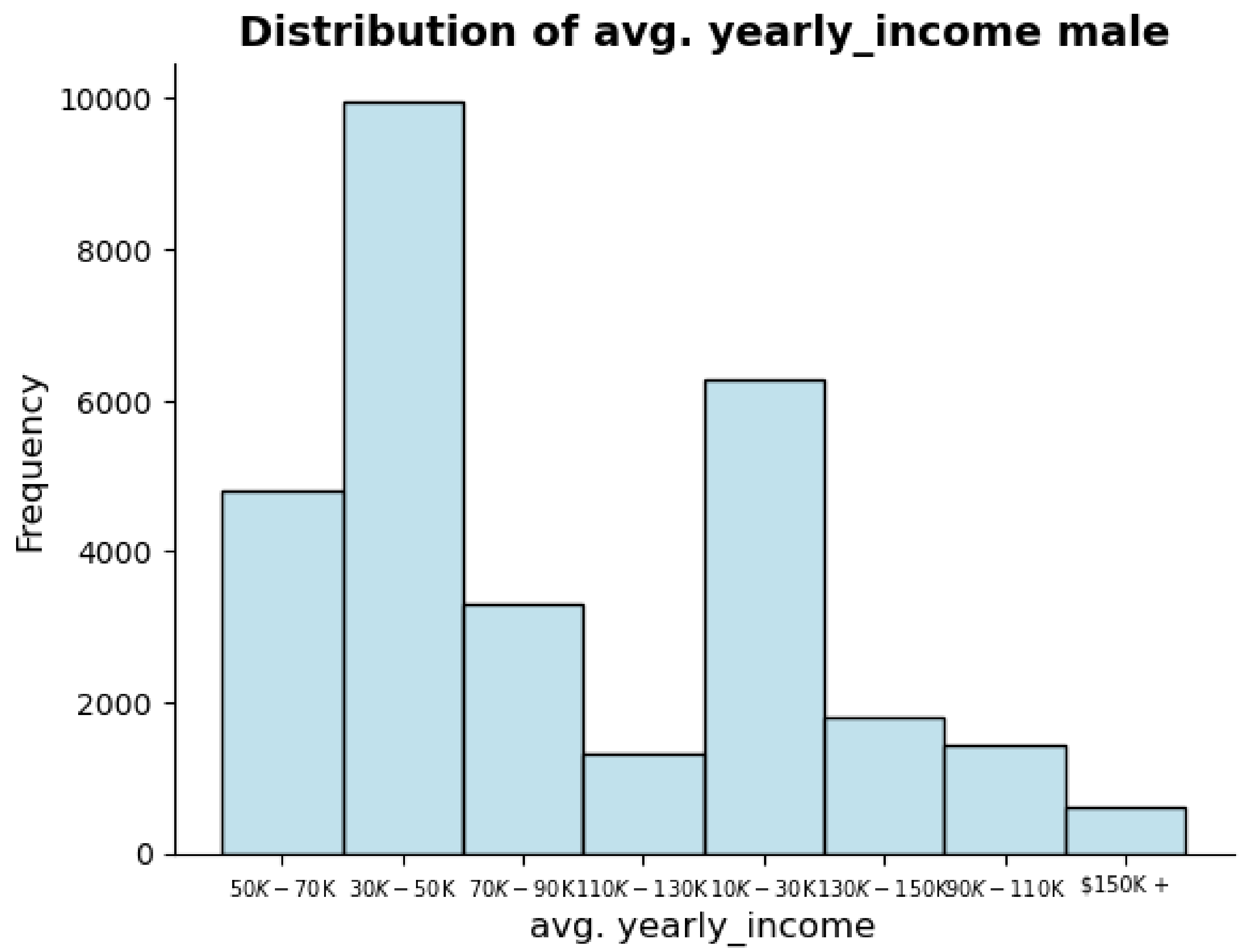
UC 2: Loyalty

Data Correlation

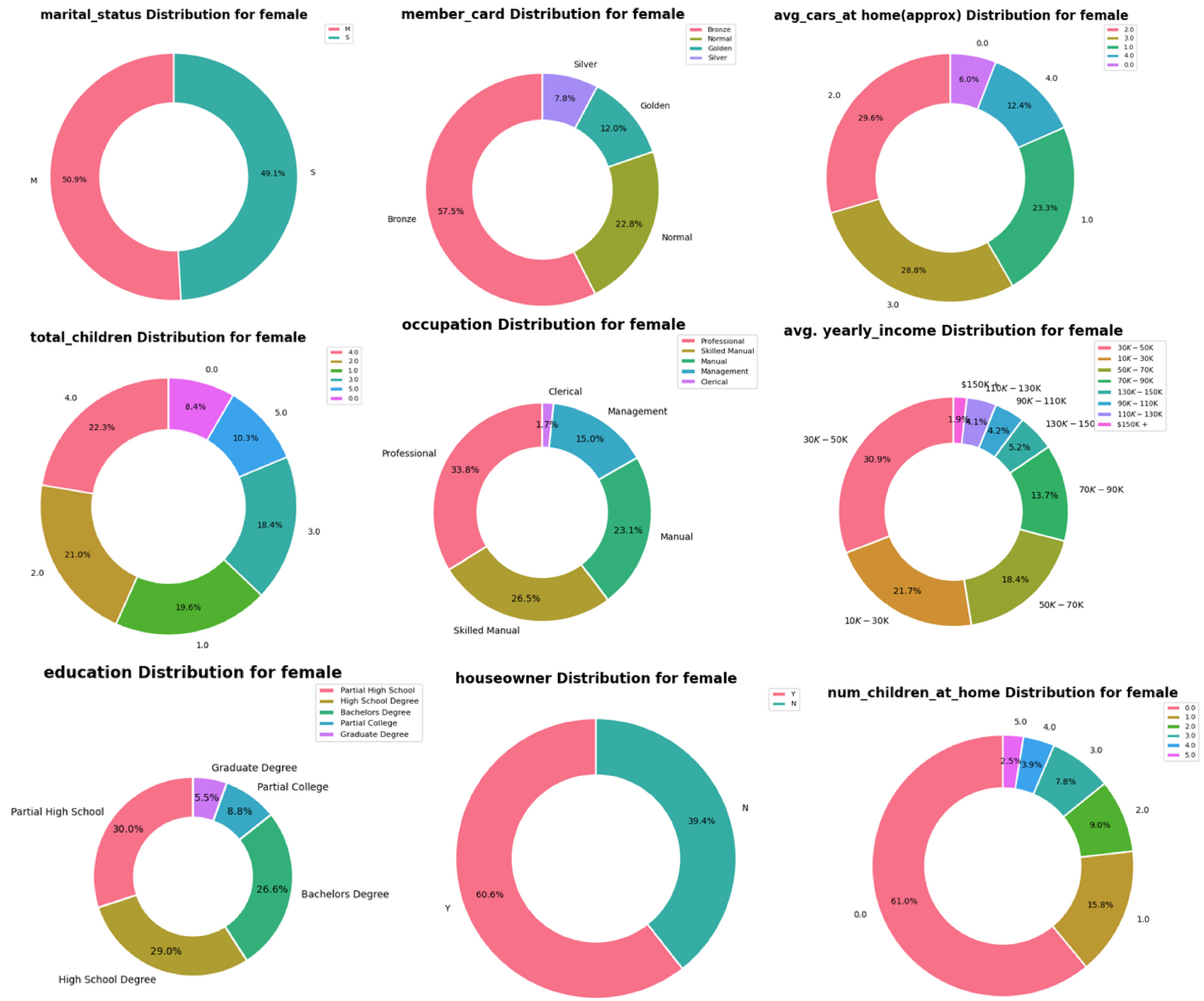


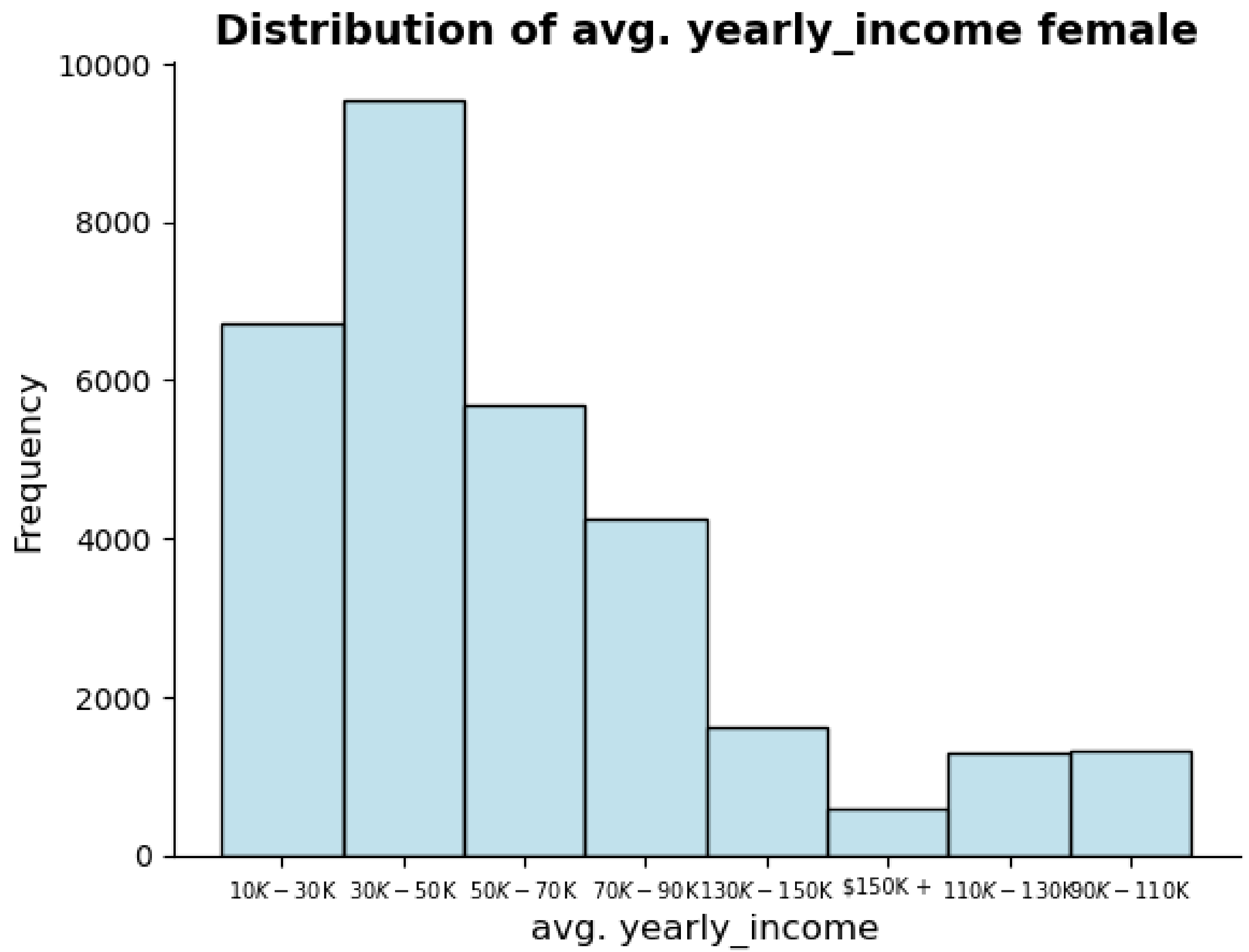


Data Evaluation: Male

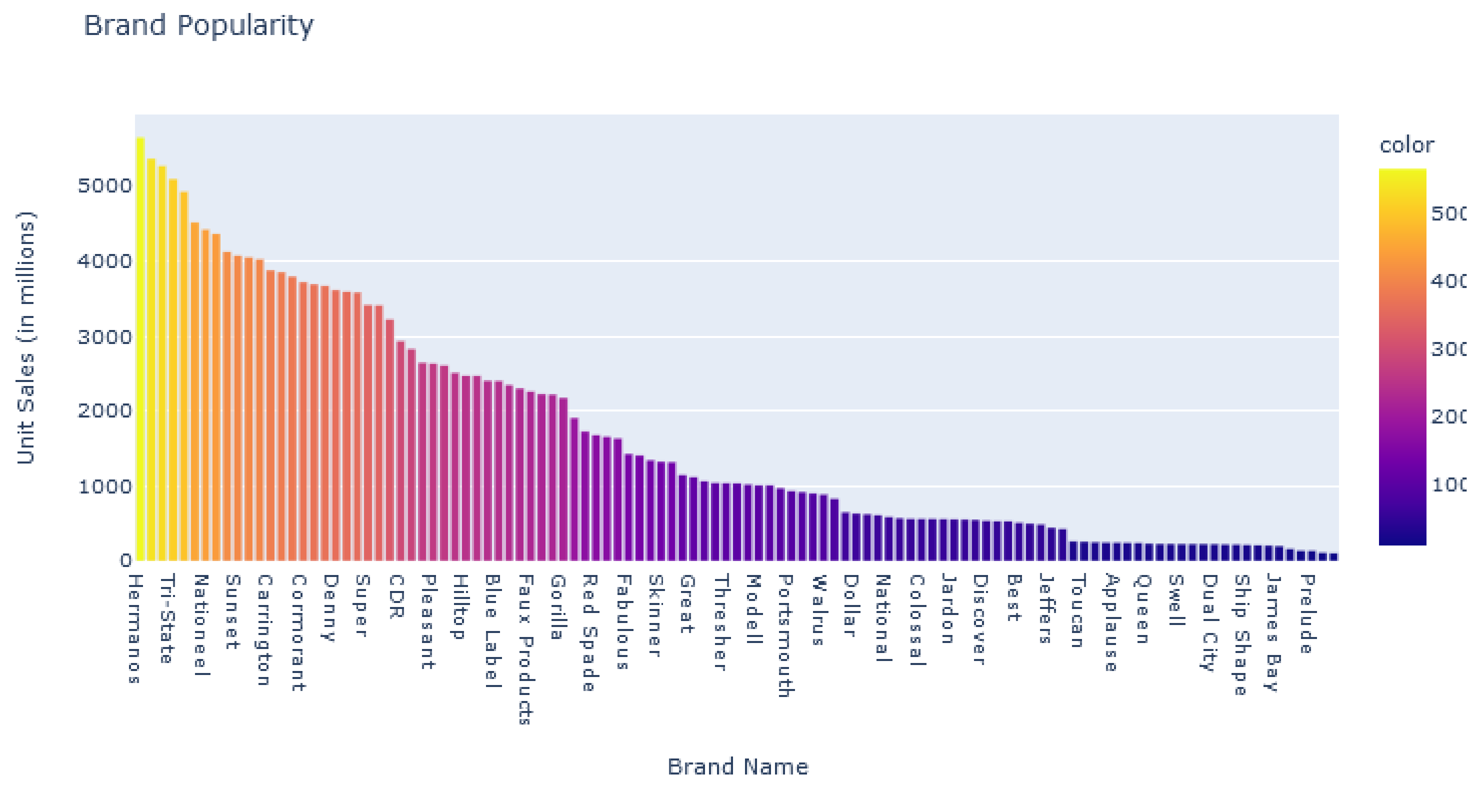


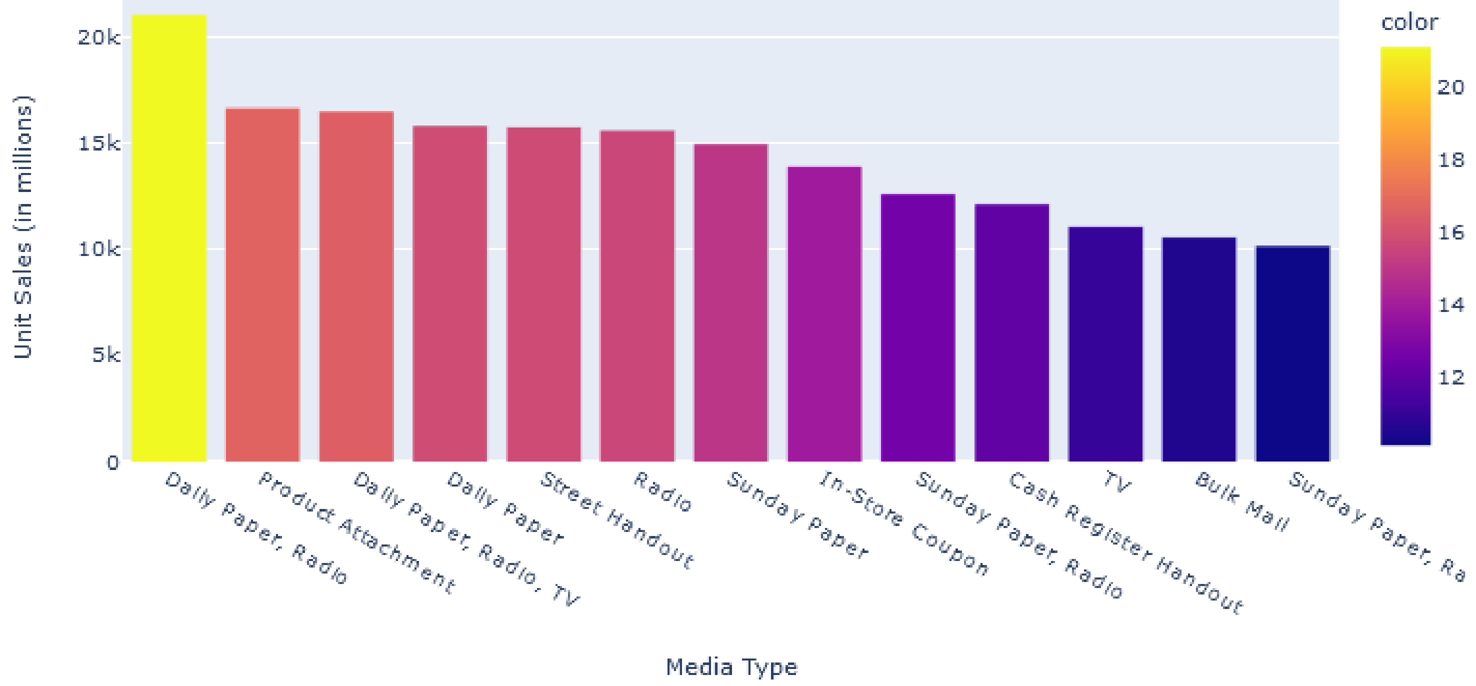
Data Evaluation: Female



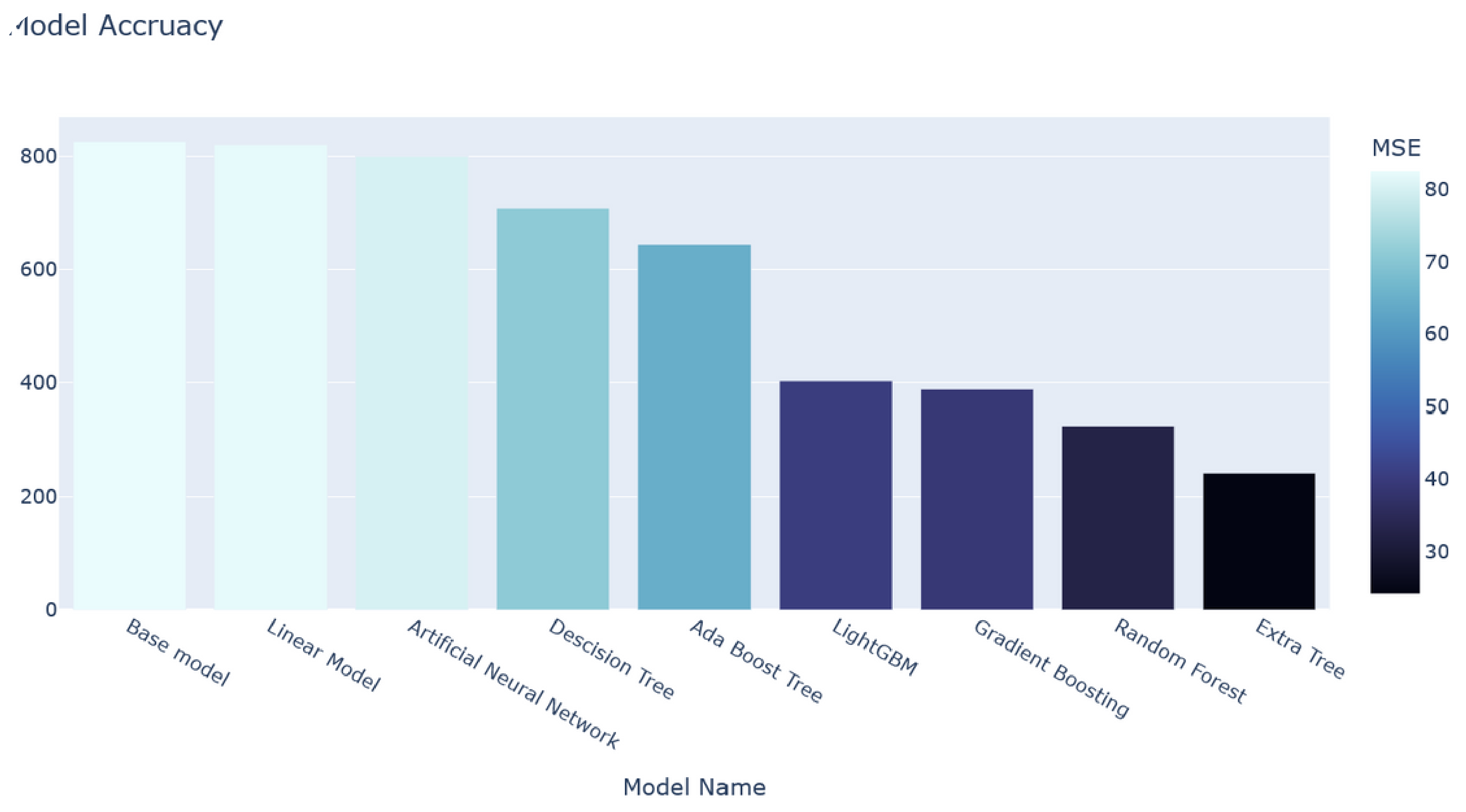


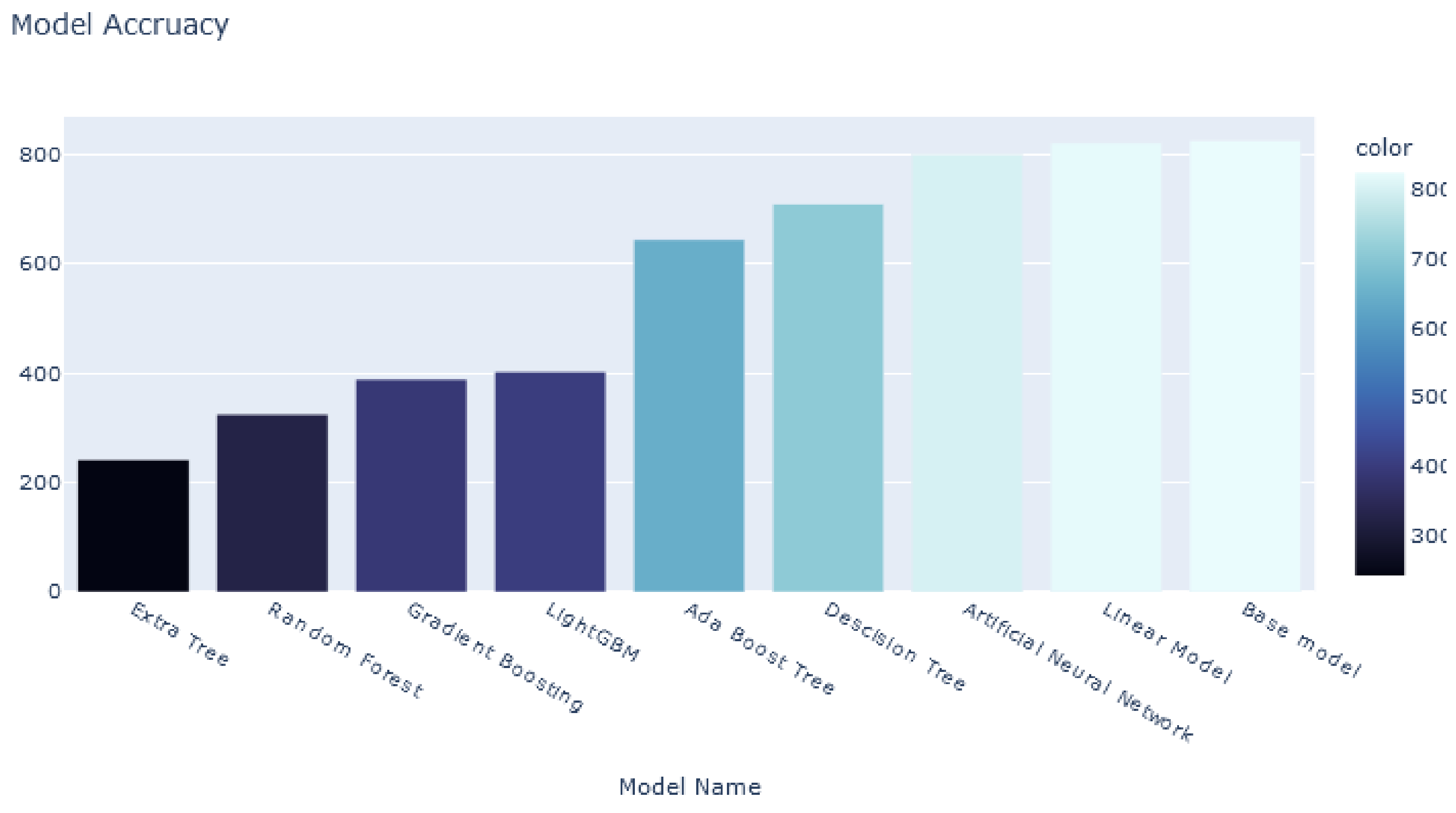
Data Evaluation



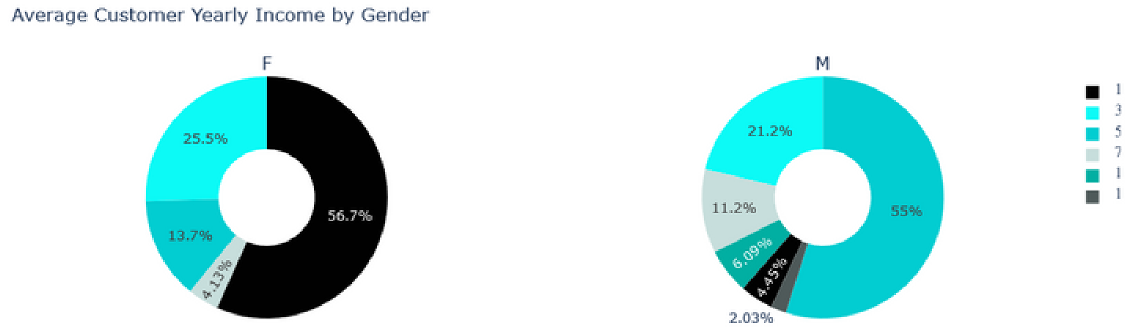


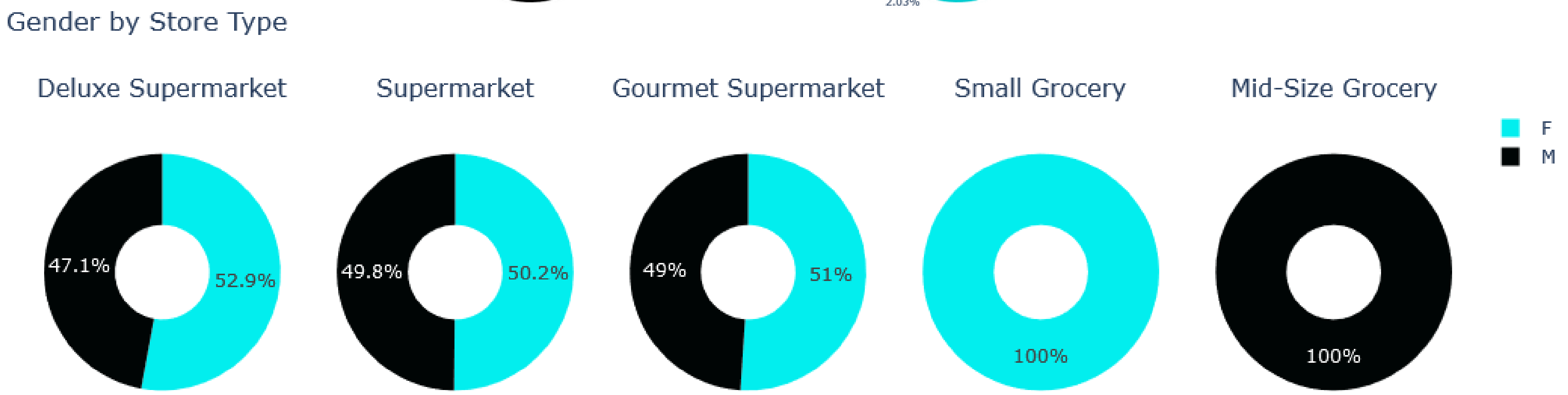
Data Evaluation

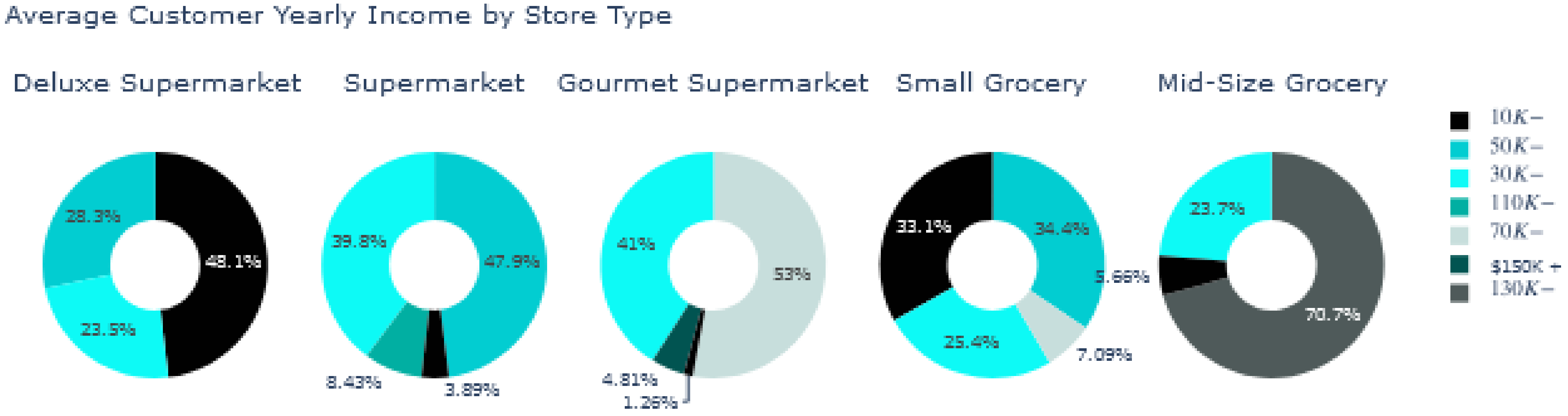




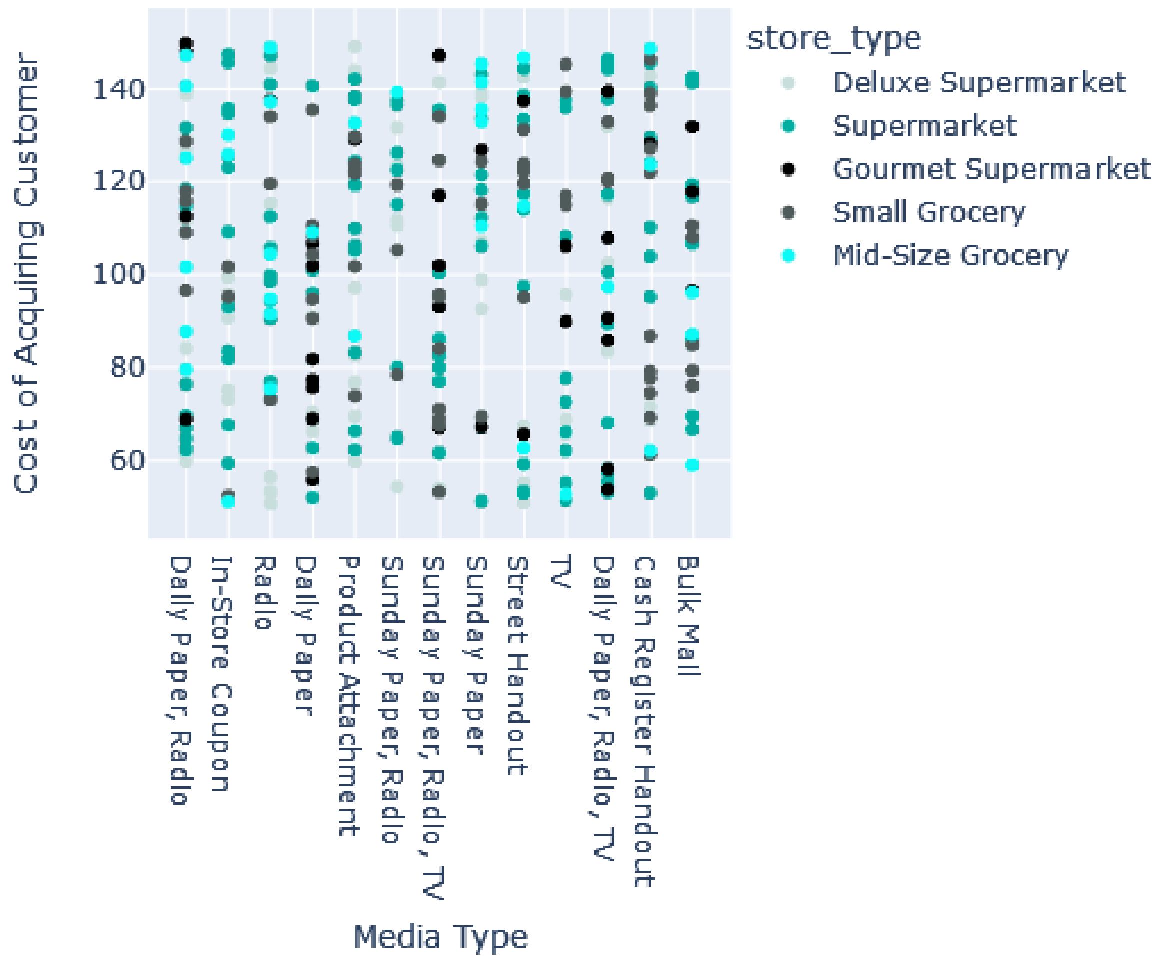
Data Evaluation

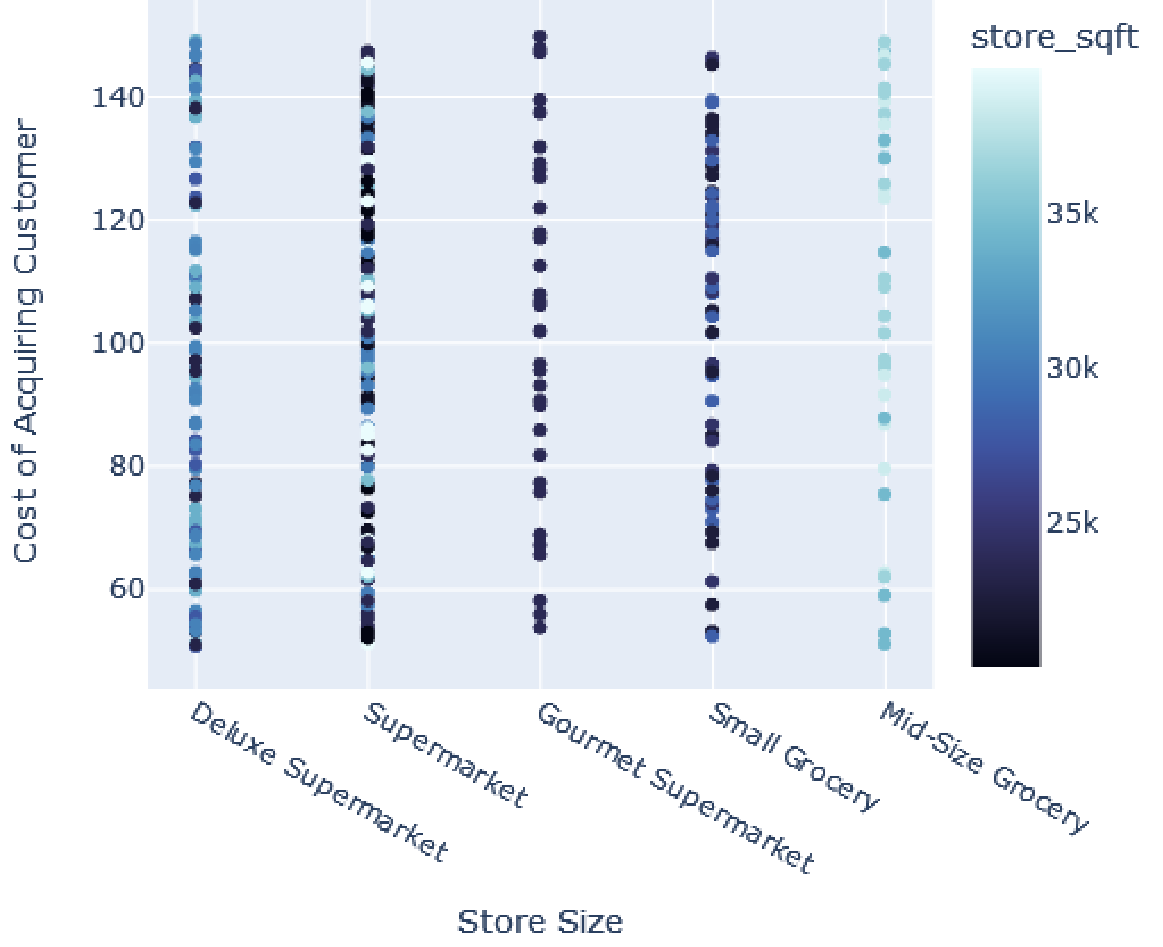


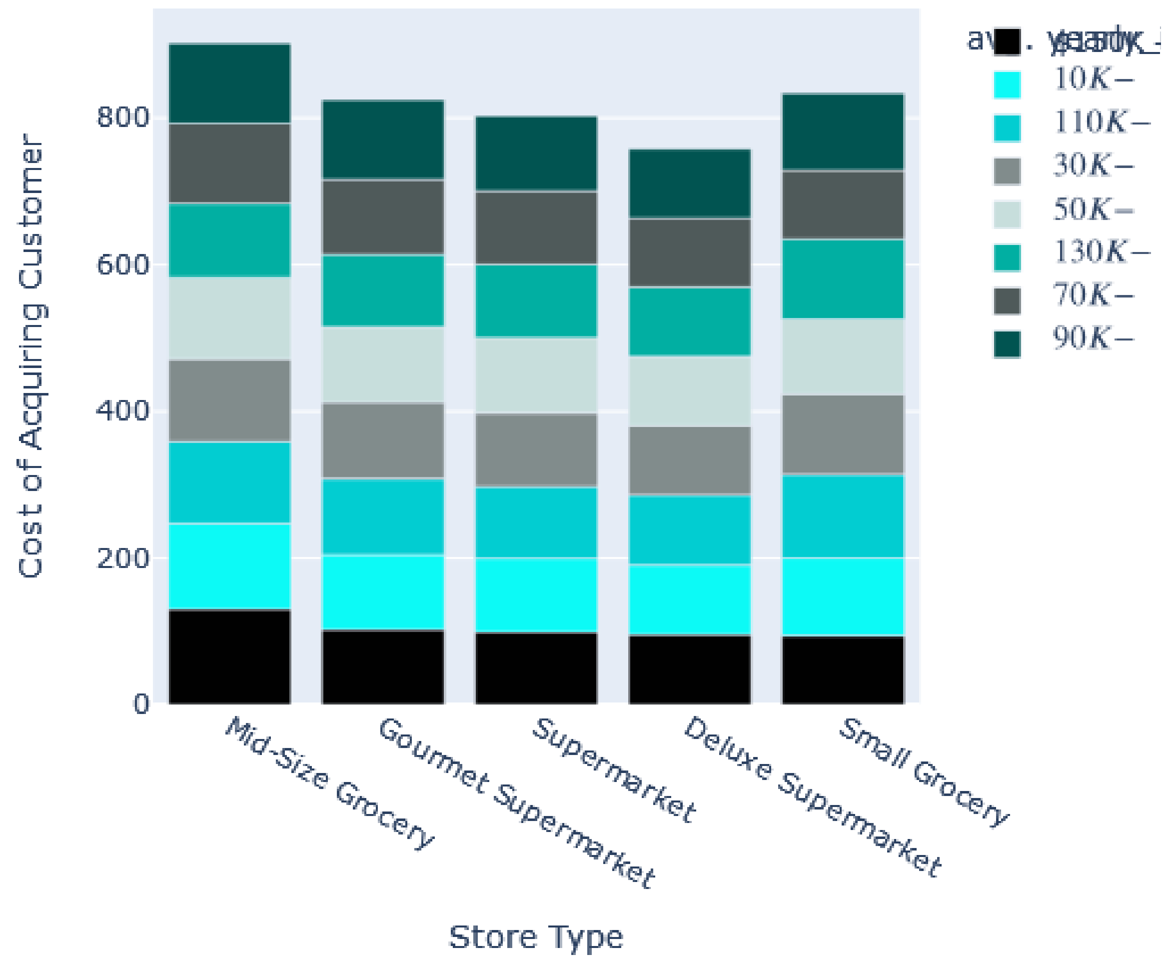




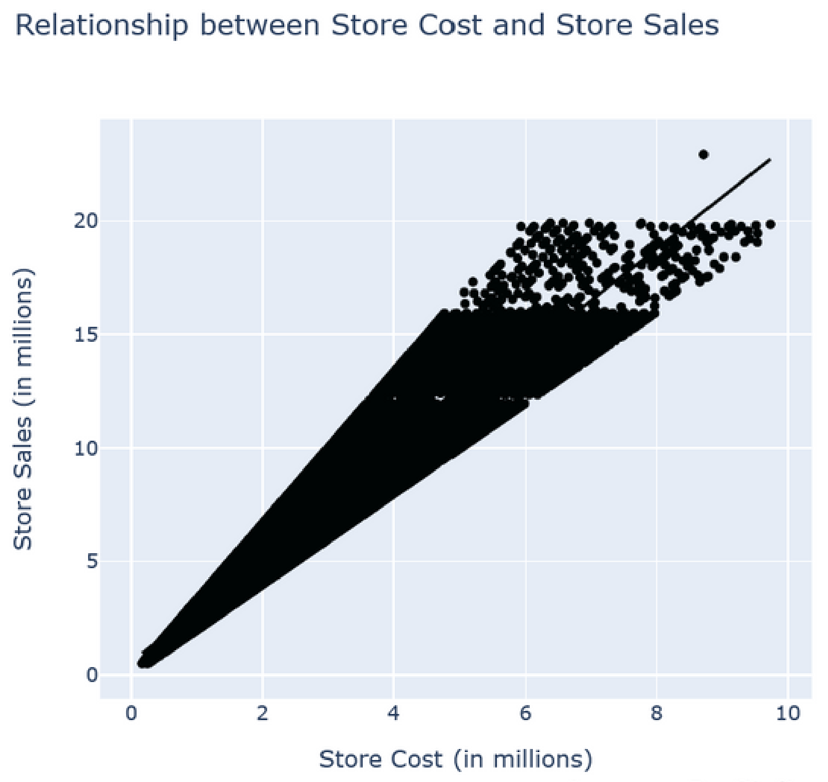
Data Evaluation

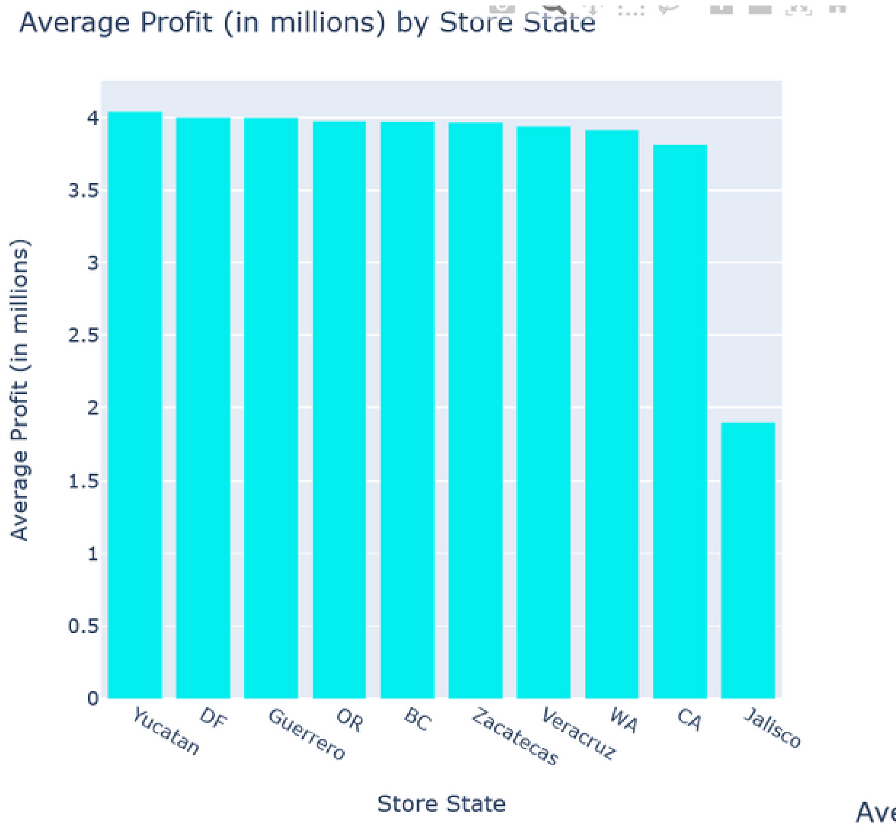


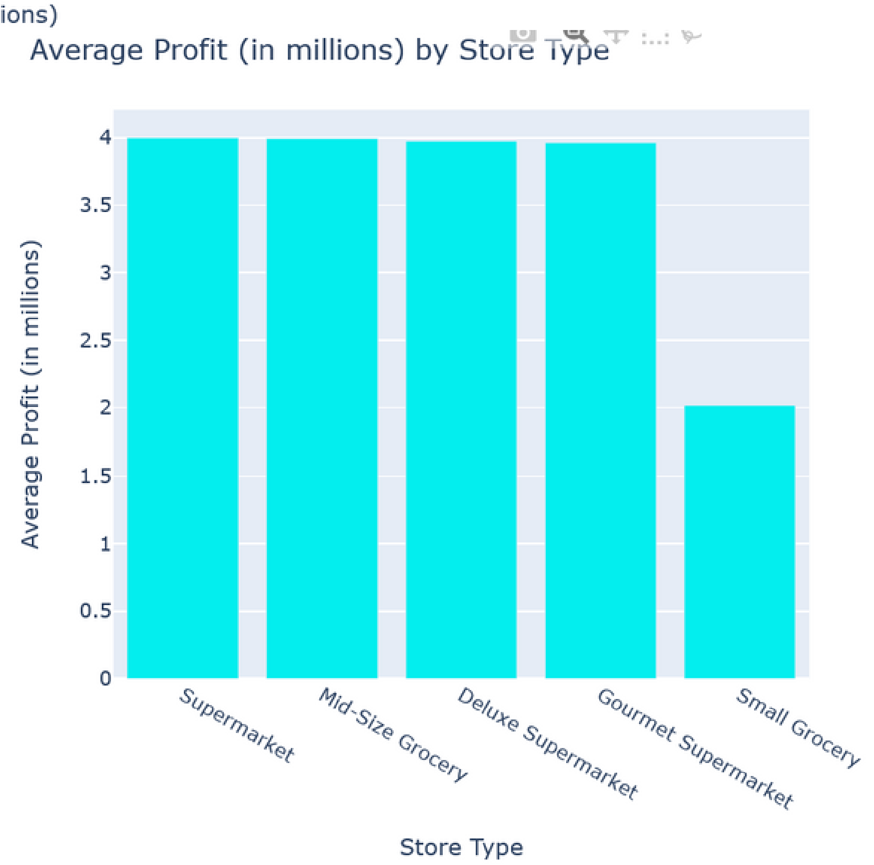


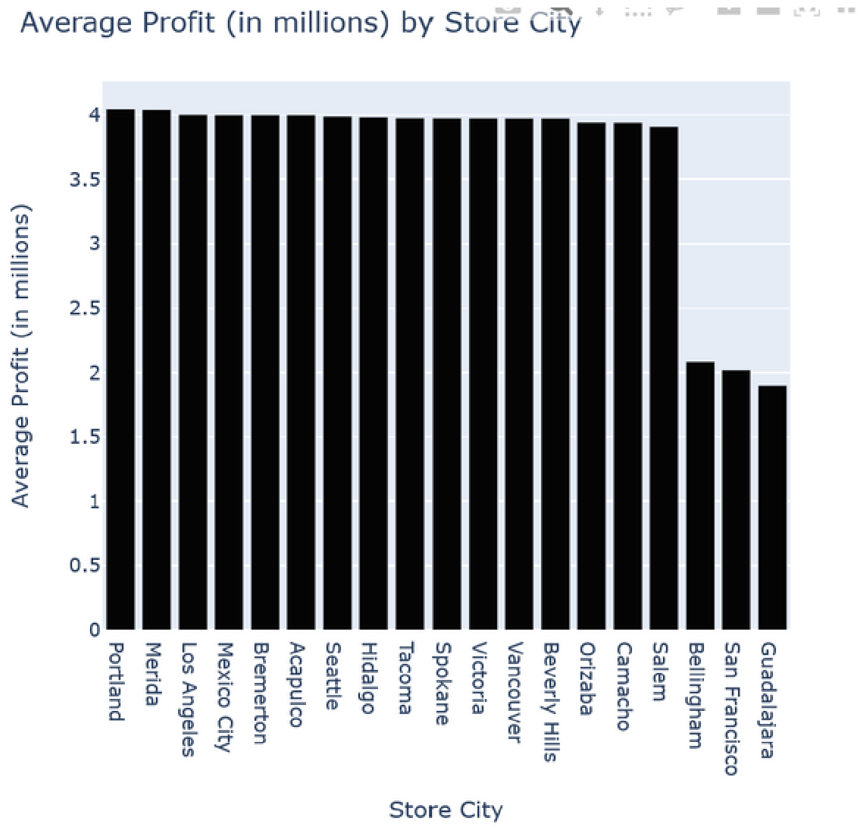


Data Evaluation

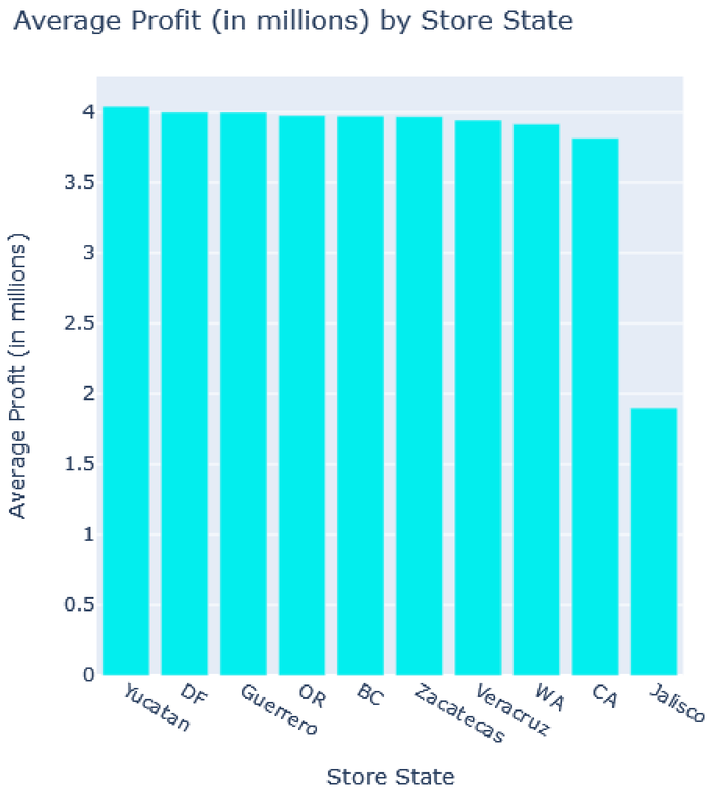


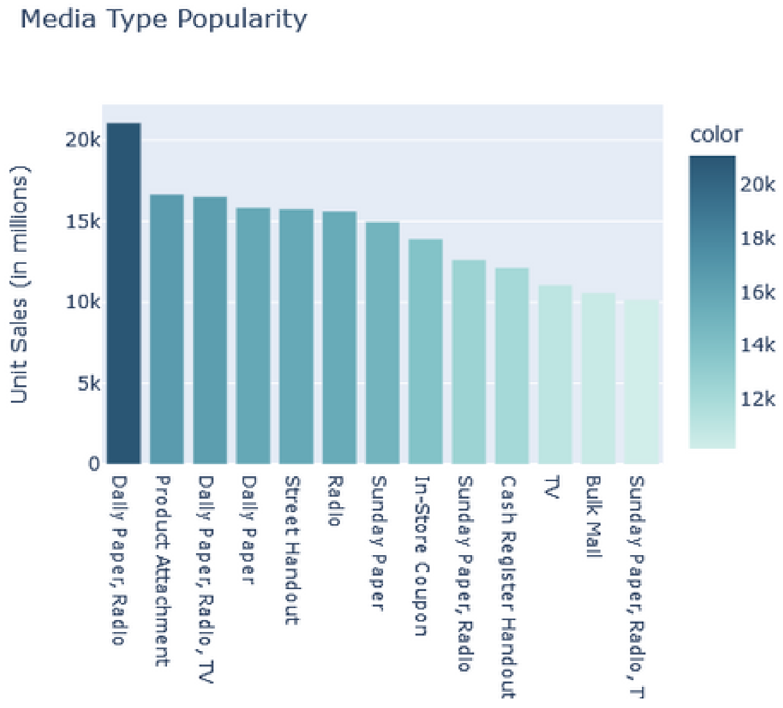


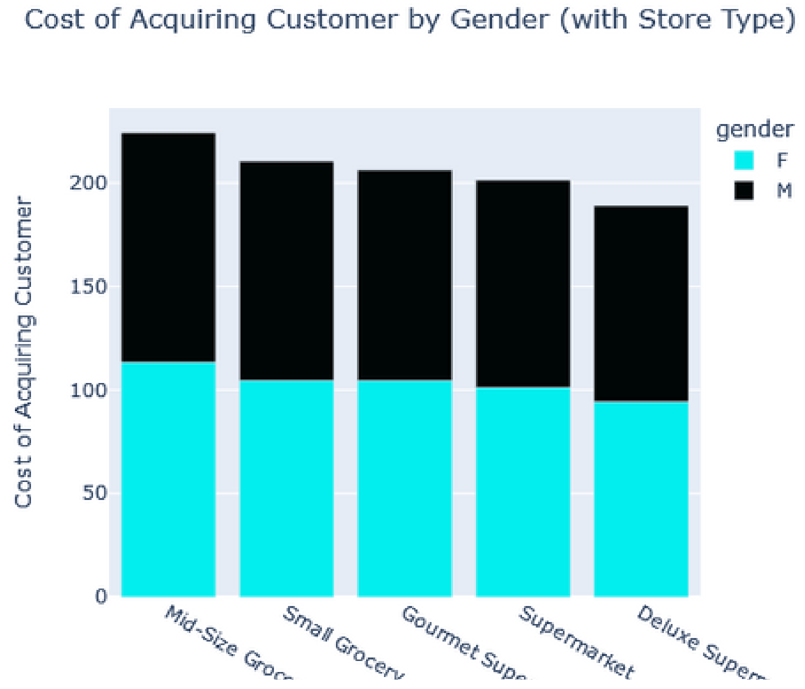


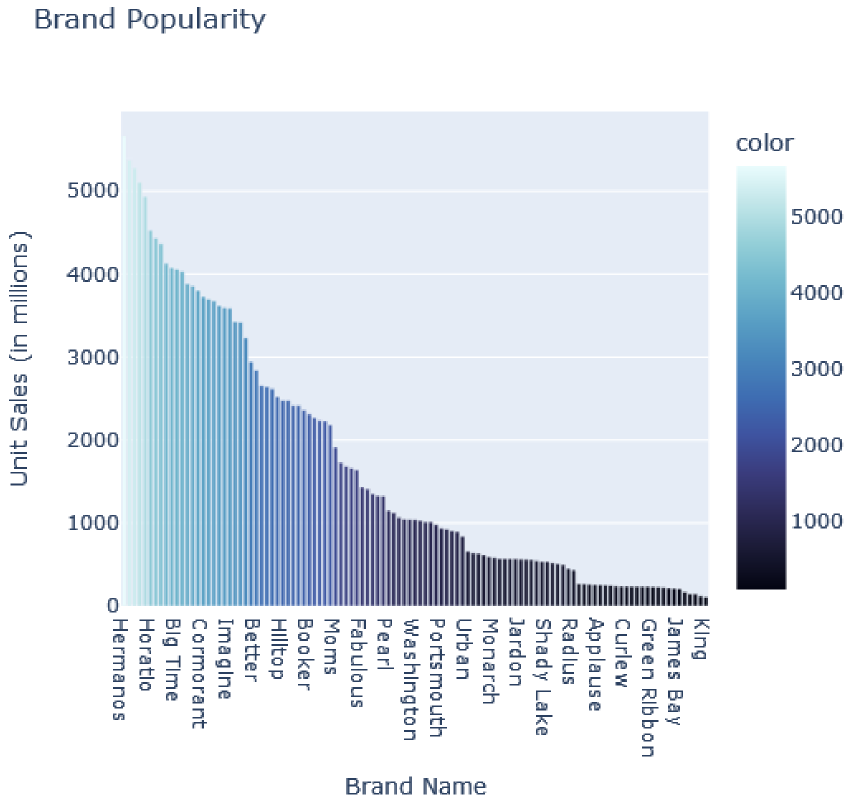


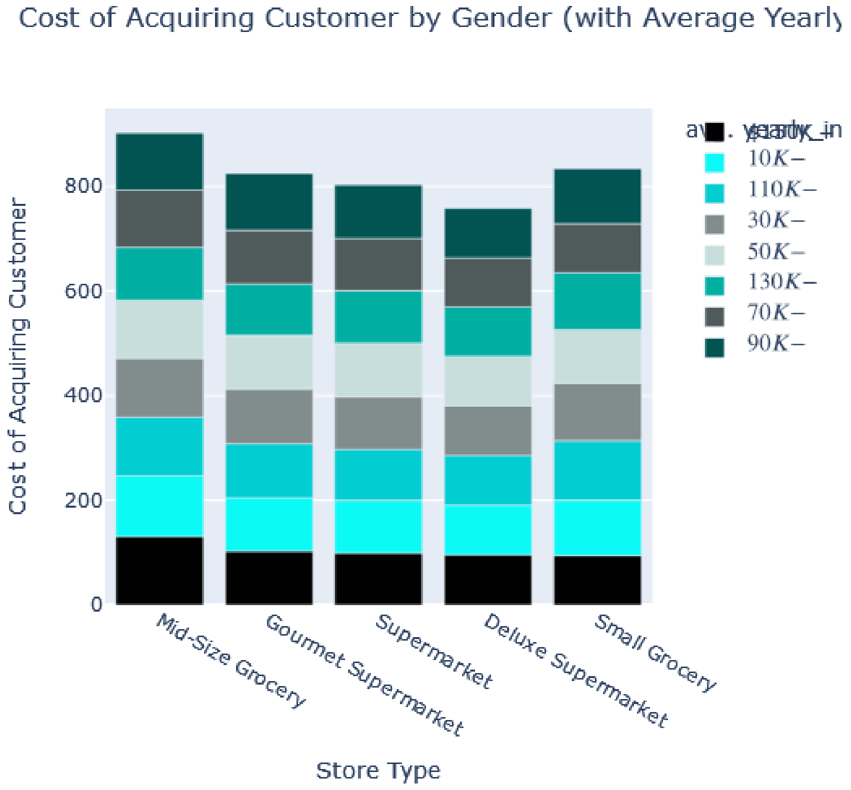
Data Evaluation



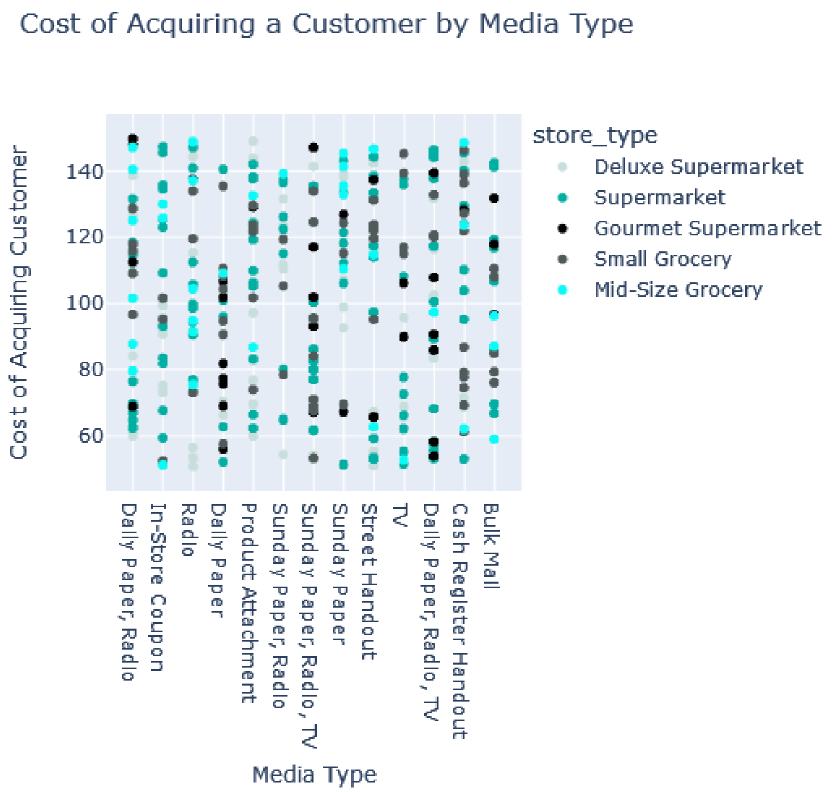


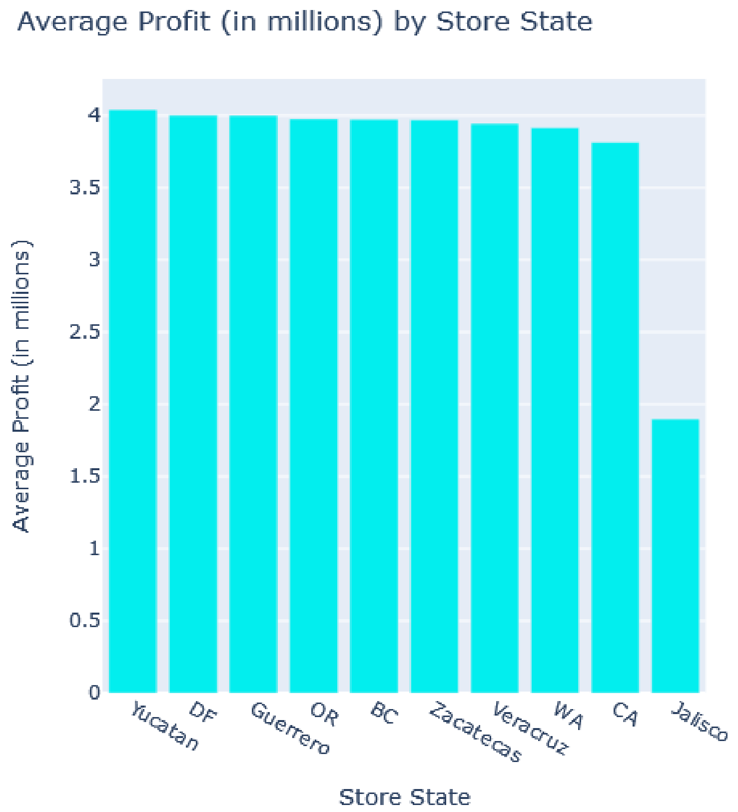


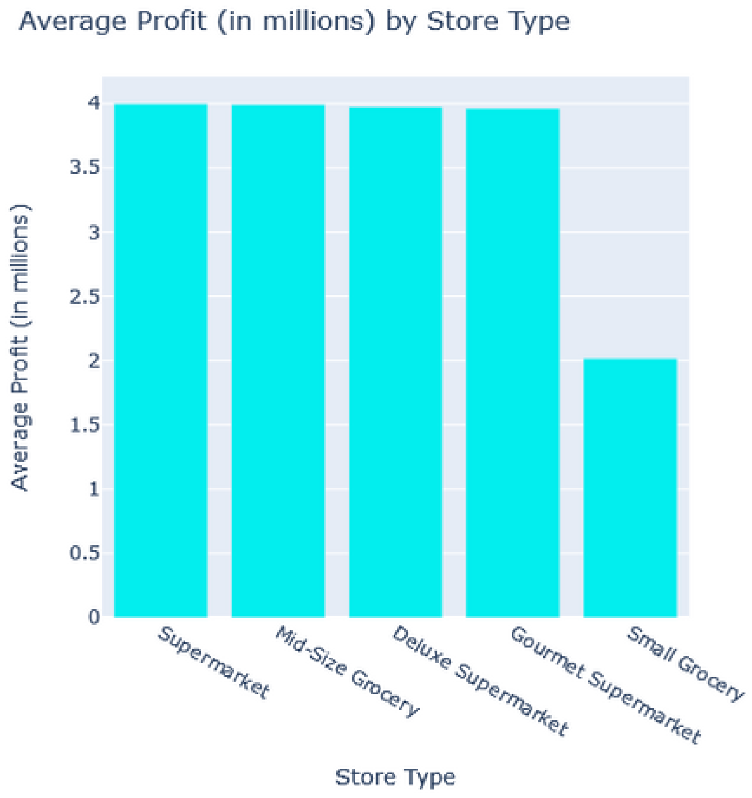


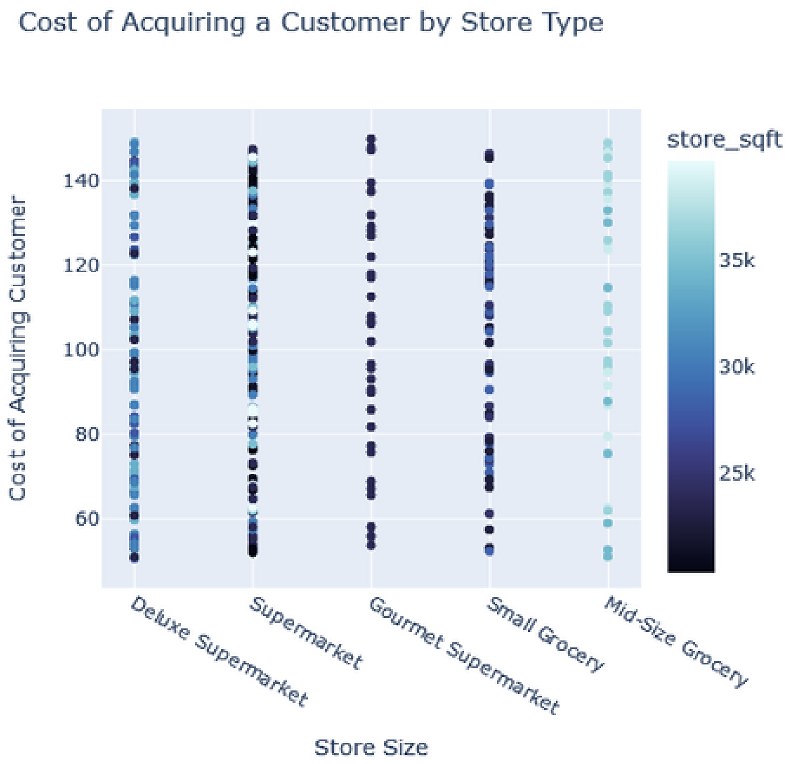


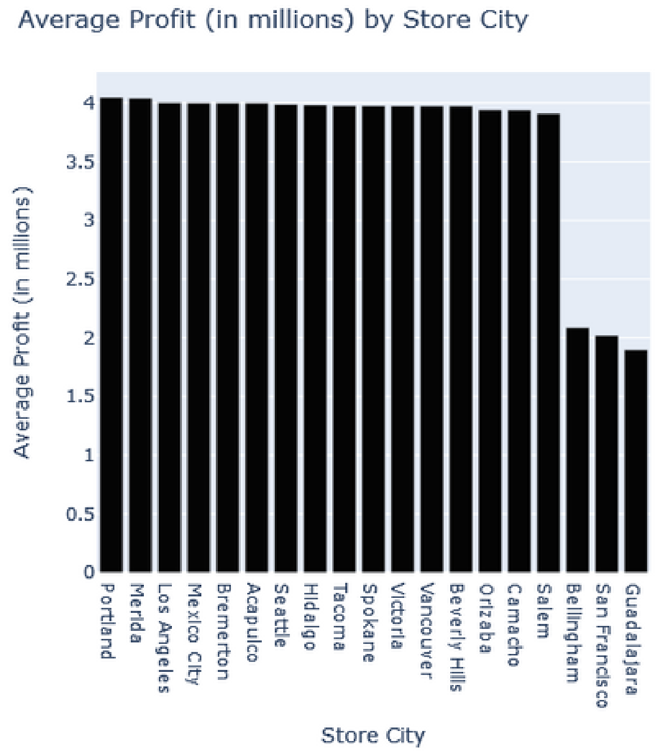
Data Evaluation



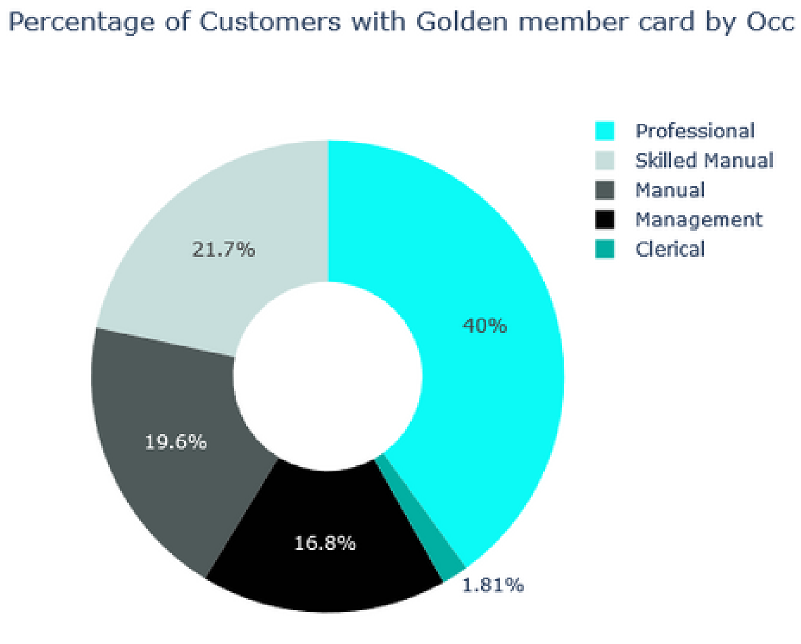


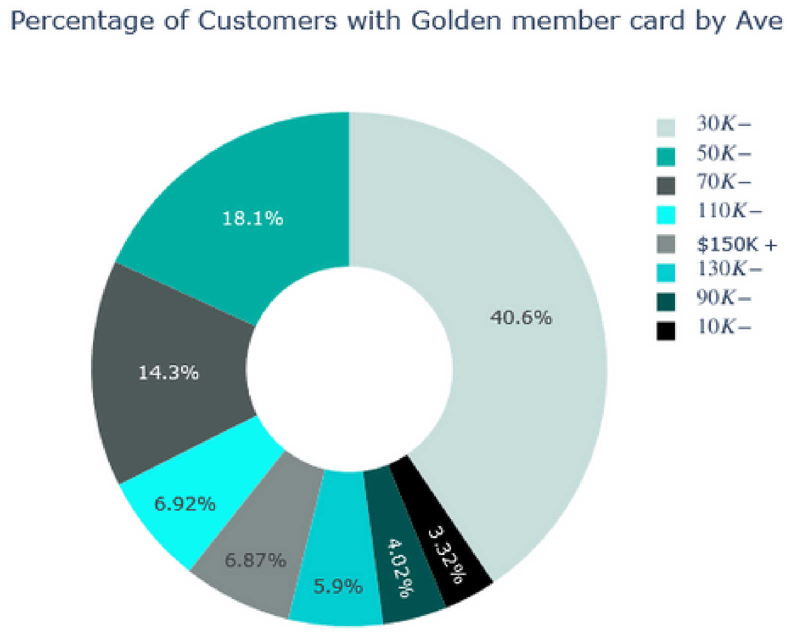


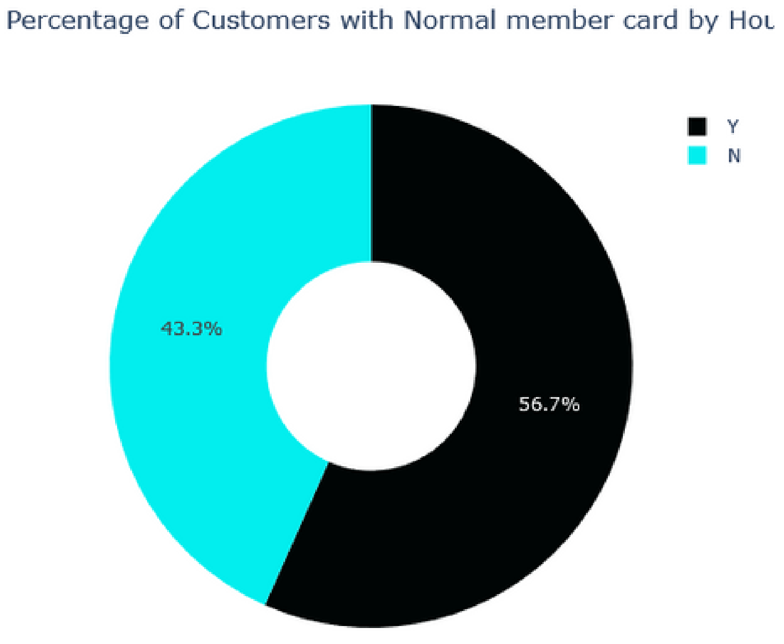


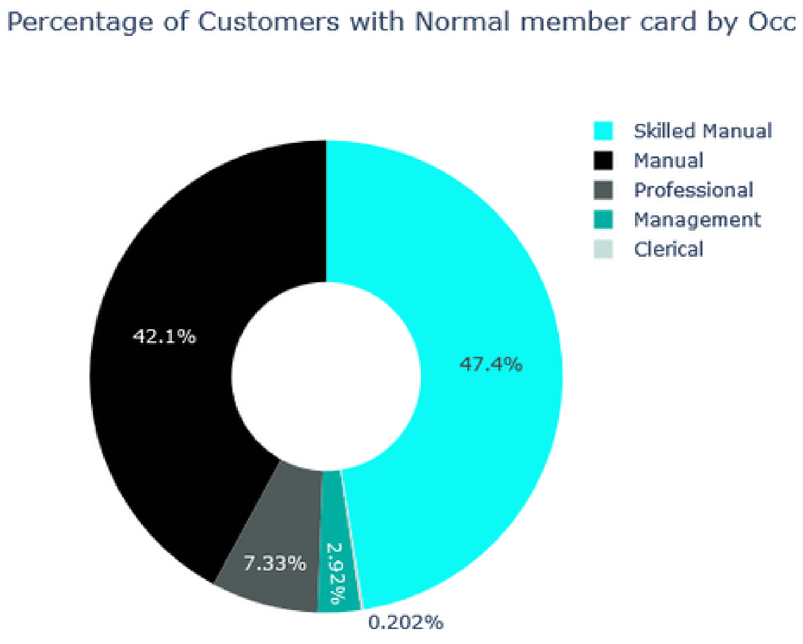


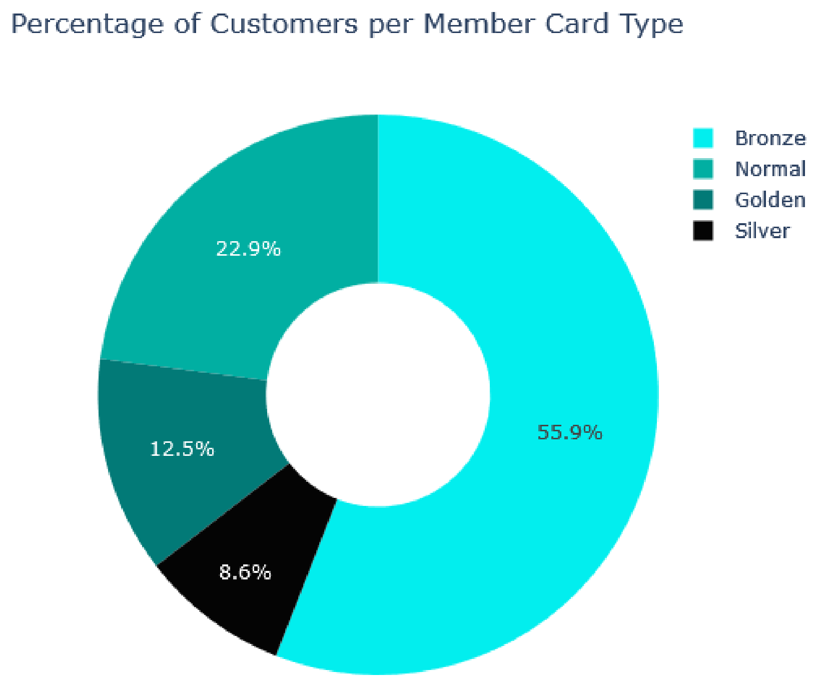
Data Evaluation

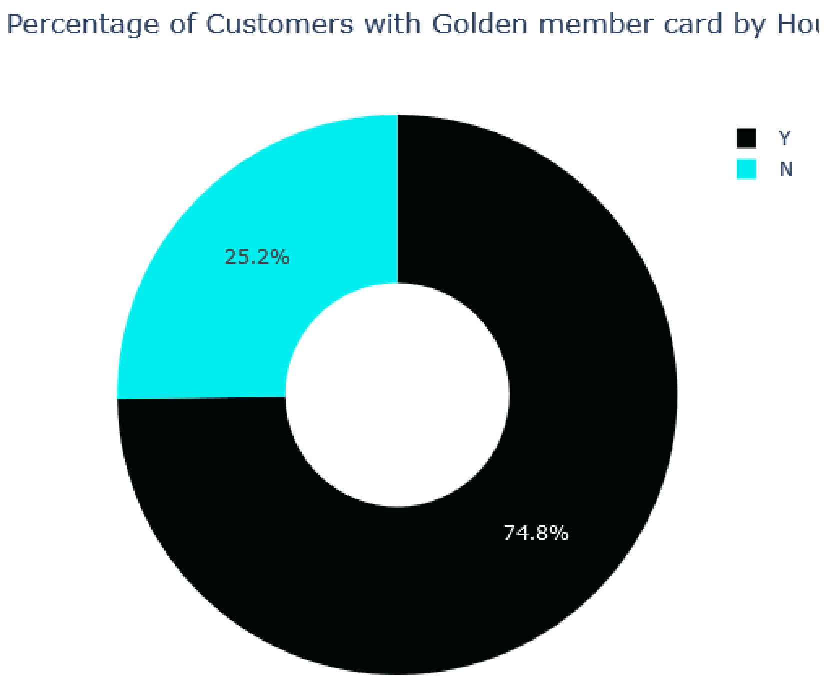


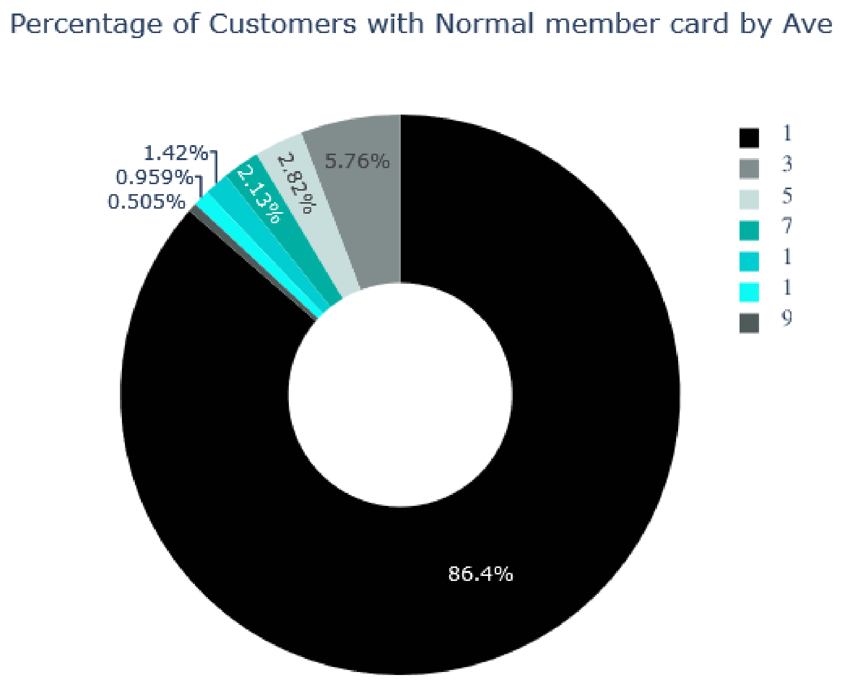




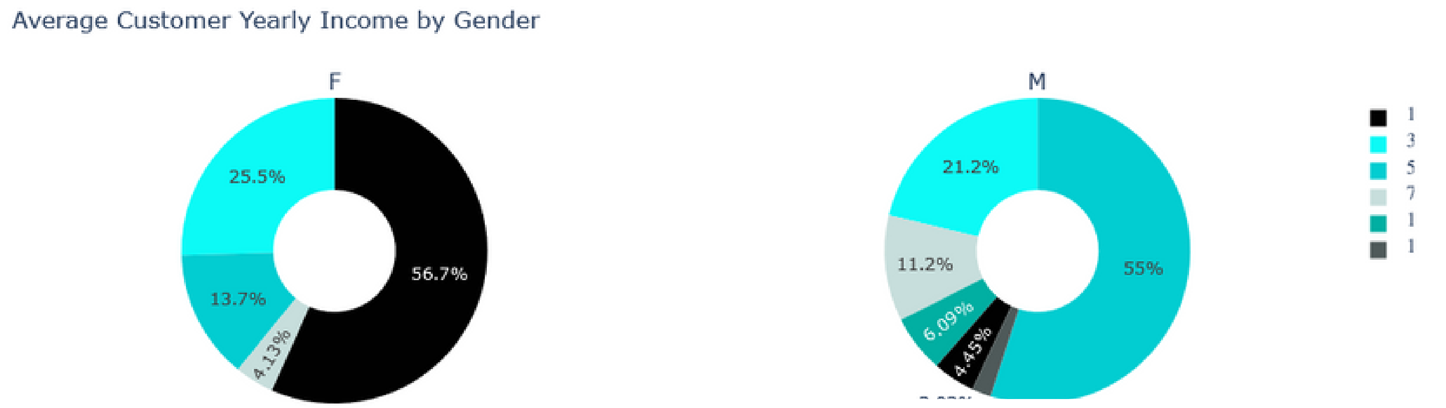


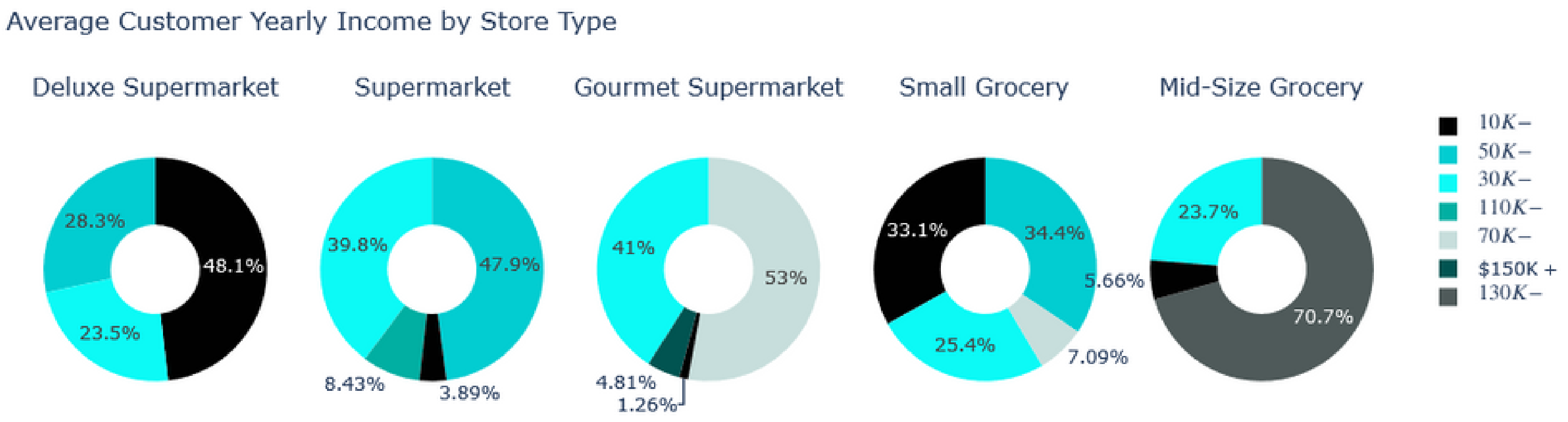


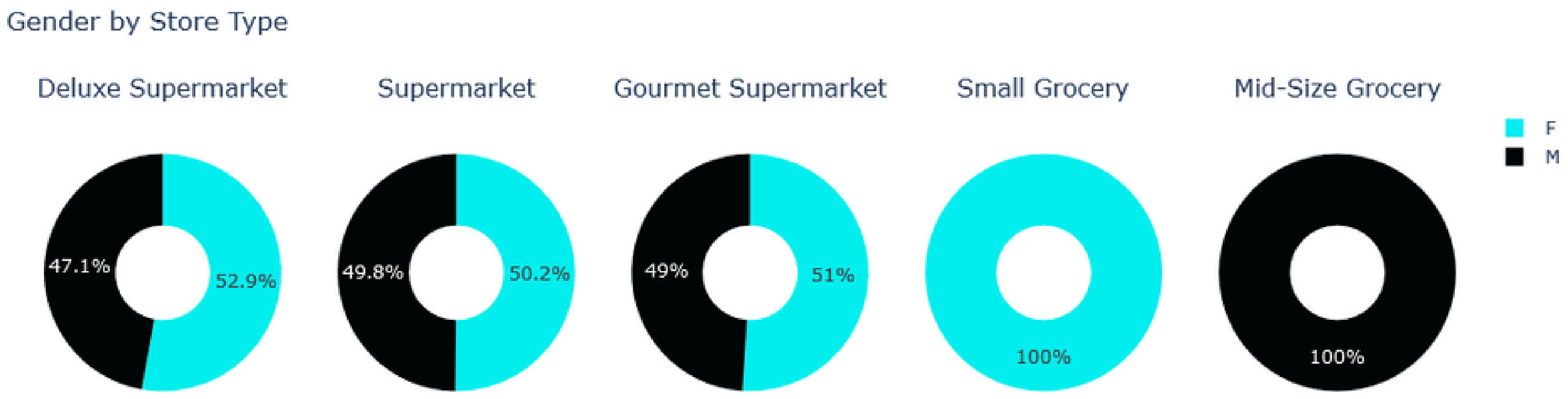


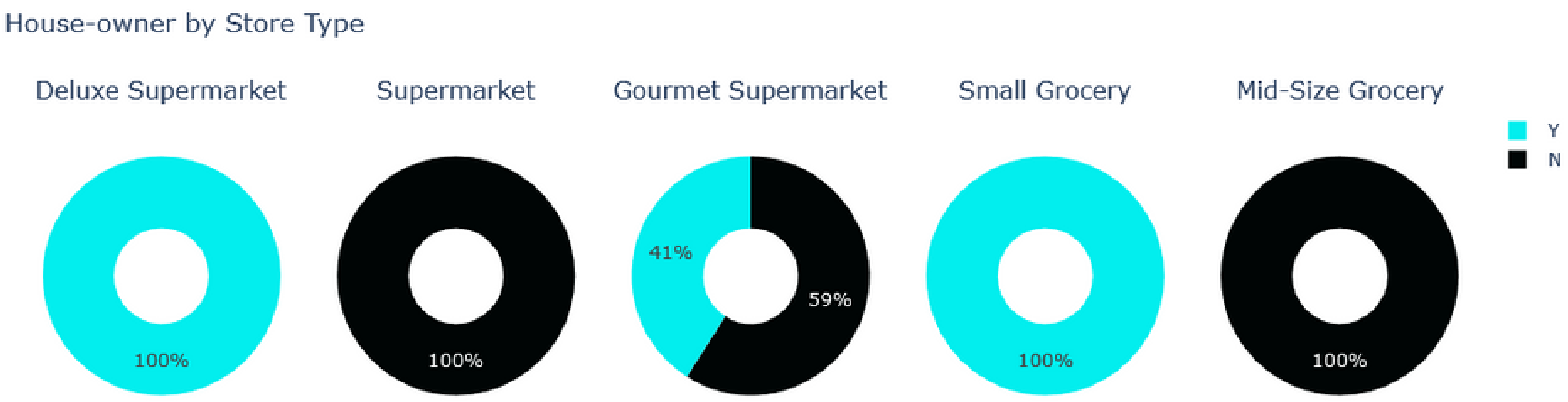


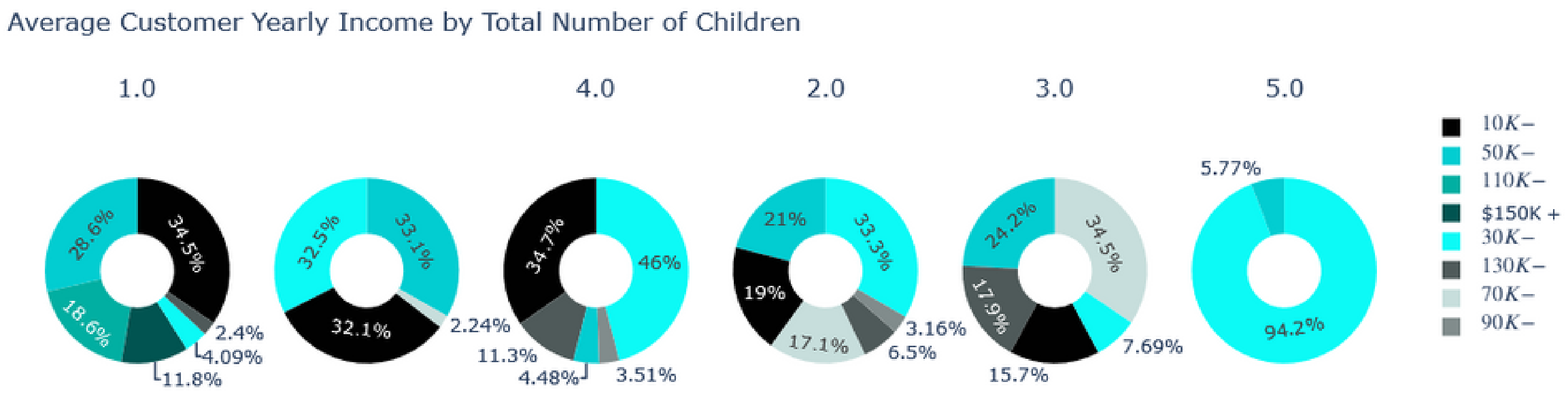
Data Evaluation





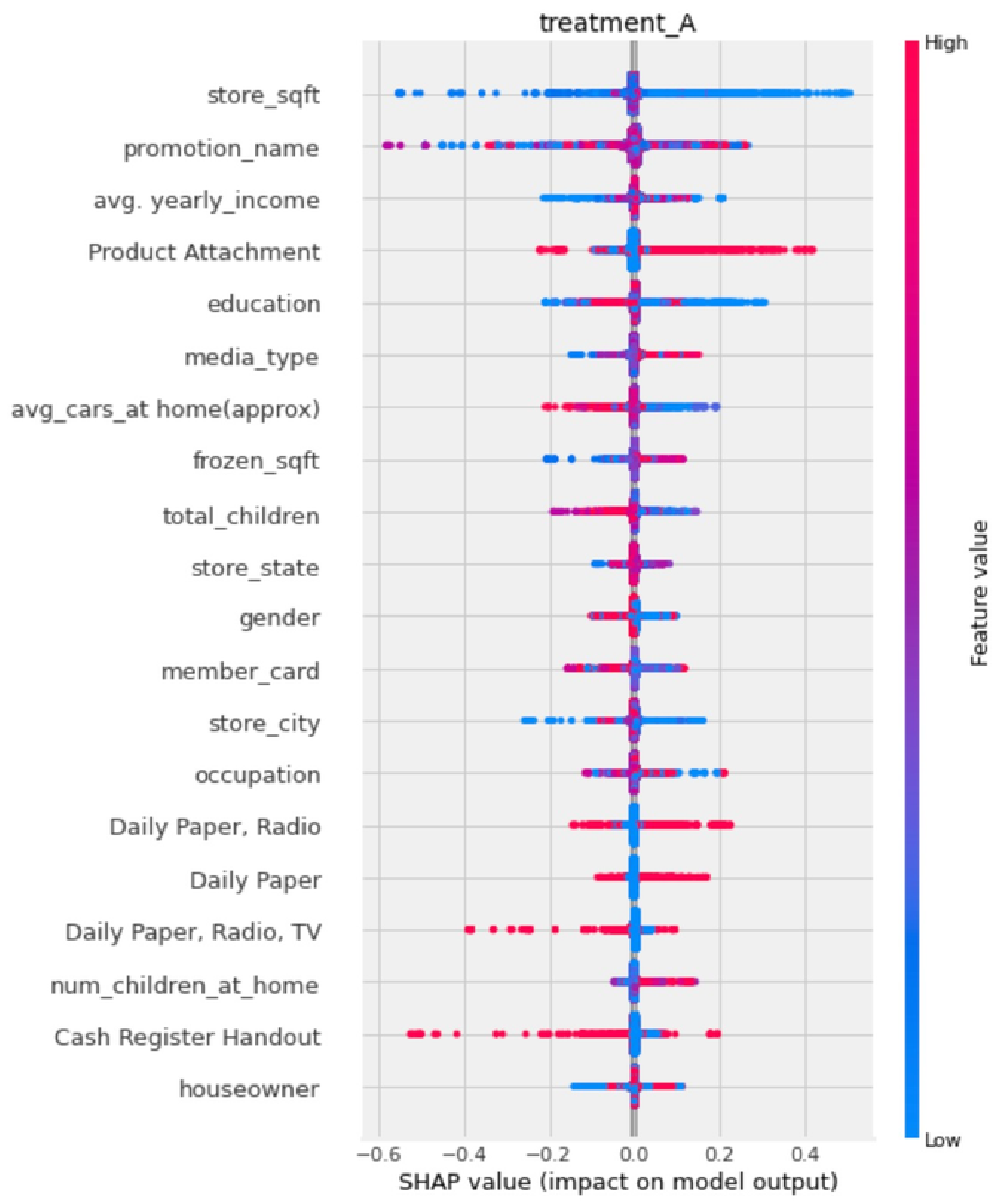




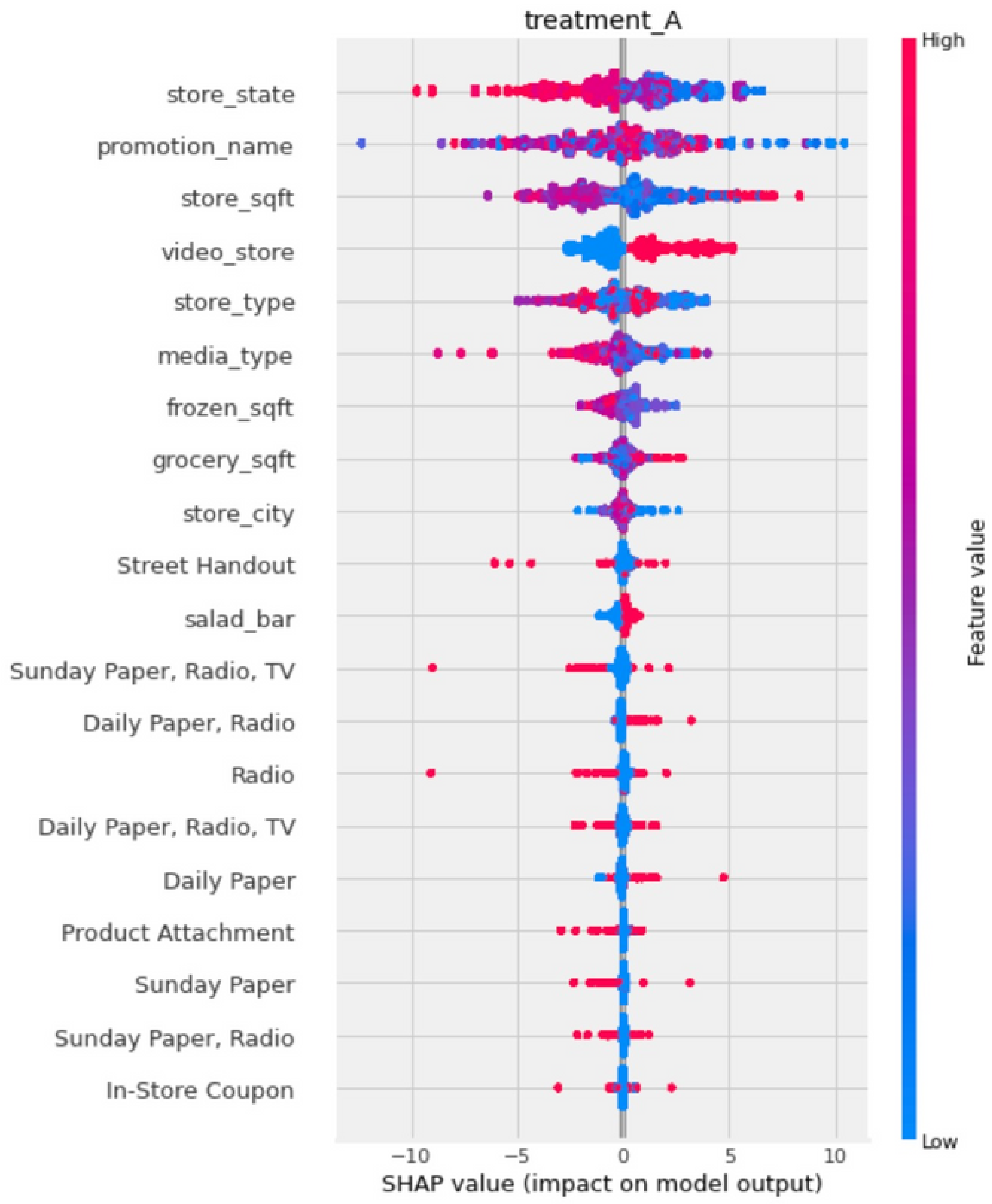


Data Evaluation

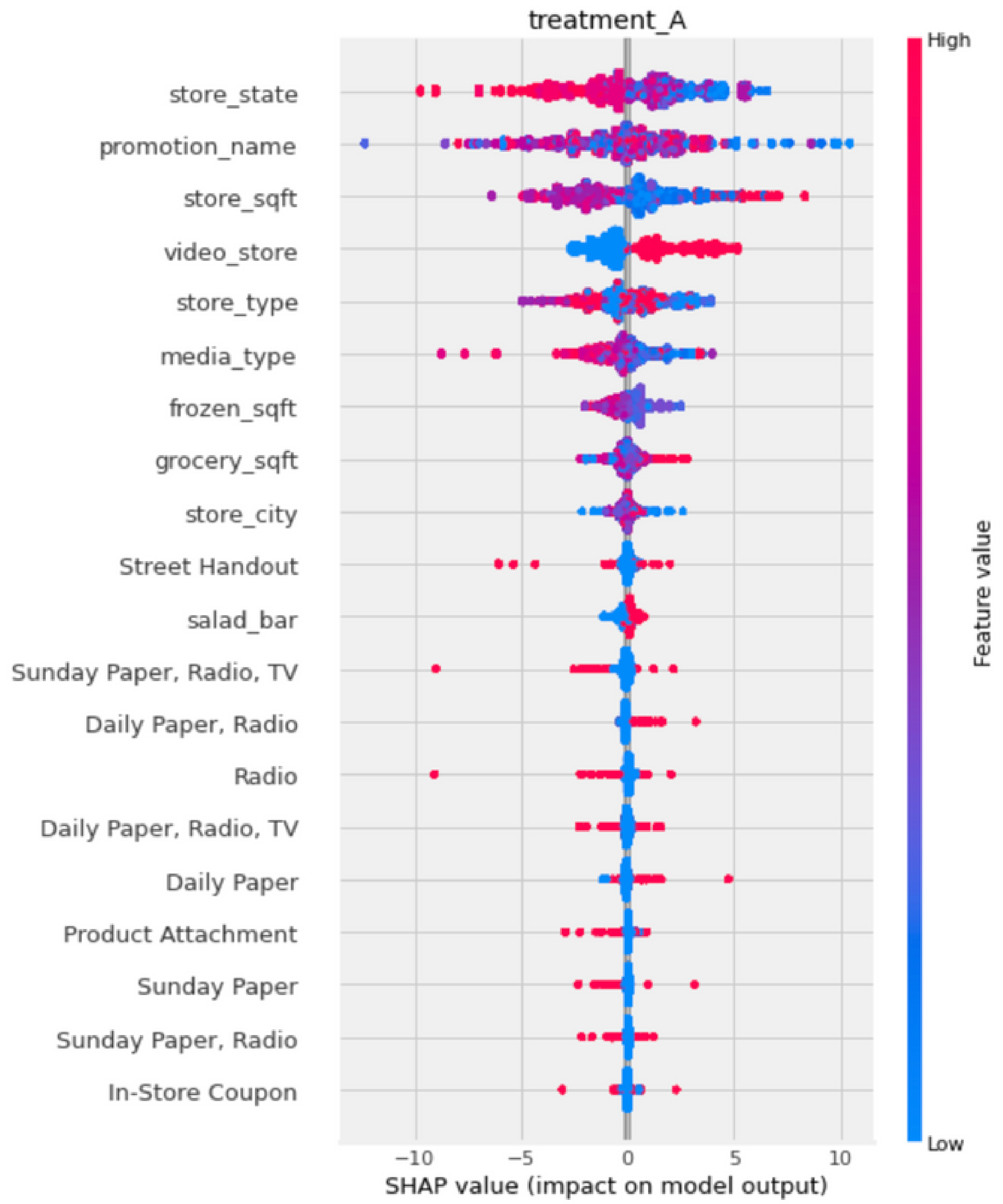
Treatment: Member card
Target: Cost



Treatment: Media Type
Target: Cost



Treatment: Gender
Target: Cost



Team Details

Name	Git ID	Git URL
Raman, Vakil	Ramanvkl	https://github.com/Ramanvkl
Priyanka, Jahul	priyanka-cj	https://github.com/priyanka-cj
Lucie, Peccoux	luciepcx	https://github.com/luciepcx
Julie, Chanzy	Juliechz	https://github.com/Juliechz
Jeongho, Pyo	chongho-pyo	https://github.com/chongho-pyo
Emery, Dittmer	Emery-Dittmer	https://github.com/Emery-Dittmer
Bennett, Fahey	bennettfahey	https://github.com/bennettfahey