Insight into an 84.51 Kroger Dataset

December 12th, 2018

GROUP 6

Siddharth Ahuja

Zhe Huang

Krti Tallam

Jiewen Wu

Background on 84.51

- Subsidiary of Kroger

- Utilize data to provide customer insights to their clients

- Creative thinking to utilize data and science to deliver game-changing customer experiences

- Take on client partners who are ready to challenge convention

Purpose of Project

'Kroger operates in a competitive marketplace, and it is crucial that we understand customer needs going forward. This project is about understanding the current customer landscape as well as predicting future behaviour.'

Primary Goal: How can Kroger increase future sales?

Dataset Overview

- Merged 3 separate datasets: **households**, **products**, and **transitions**.
- Households:
  - 5000 households
  - Demographic information
- Products:
  - 150,000 products across 43 categories
- Transactions:
  - More than 1 million transactions
  - January 2016 to December 2017 (a dataset over a period of two years)
  - Contained Household, Product, Basket numbers, and store regions

Project Overview

- **What are the current sales trends and how can we increase future sales?**
    - Demographic and geographic factors
    - Product performance
    - Time series model
    - Outliers and anomalies

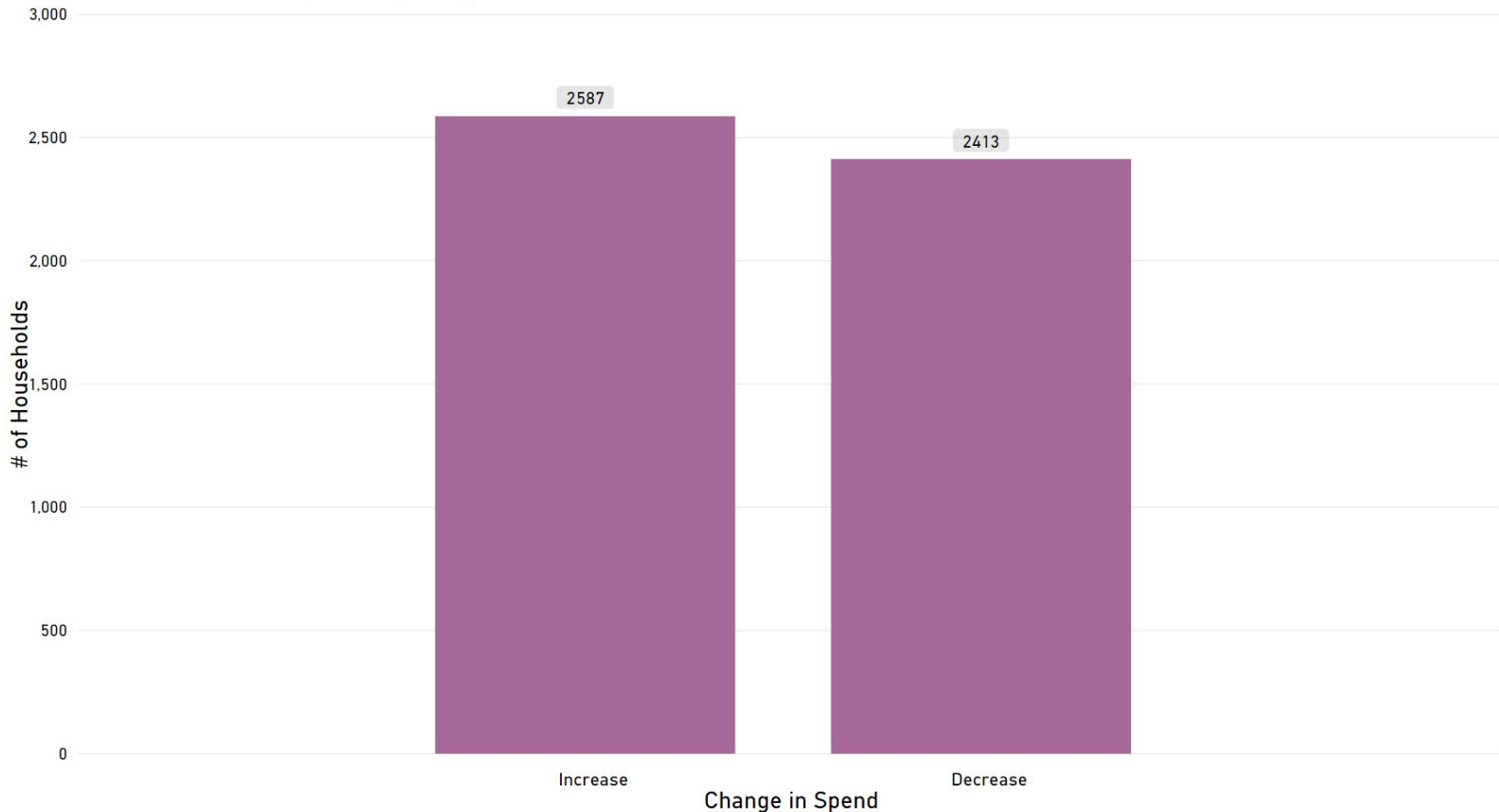- **What are the limitations of the data currently available?**

Are customers spending more or fewer dollars over time?
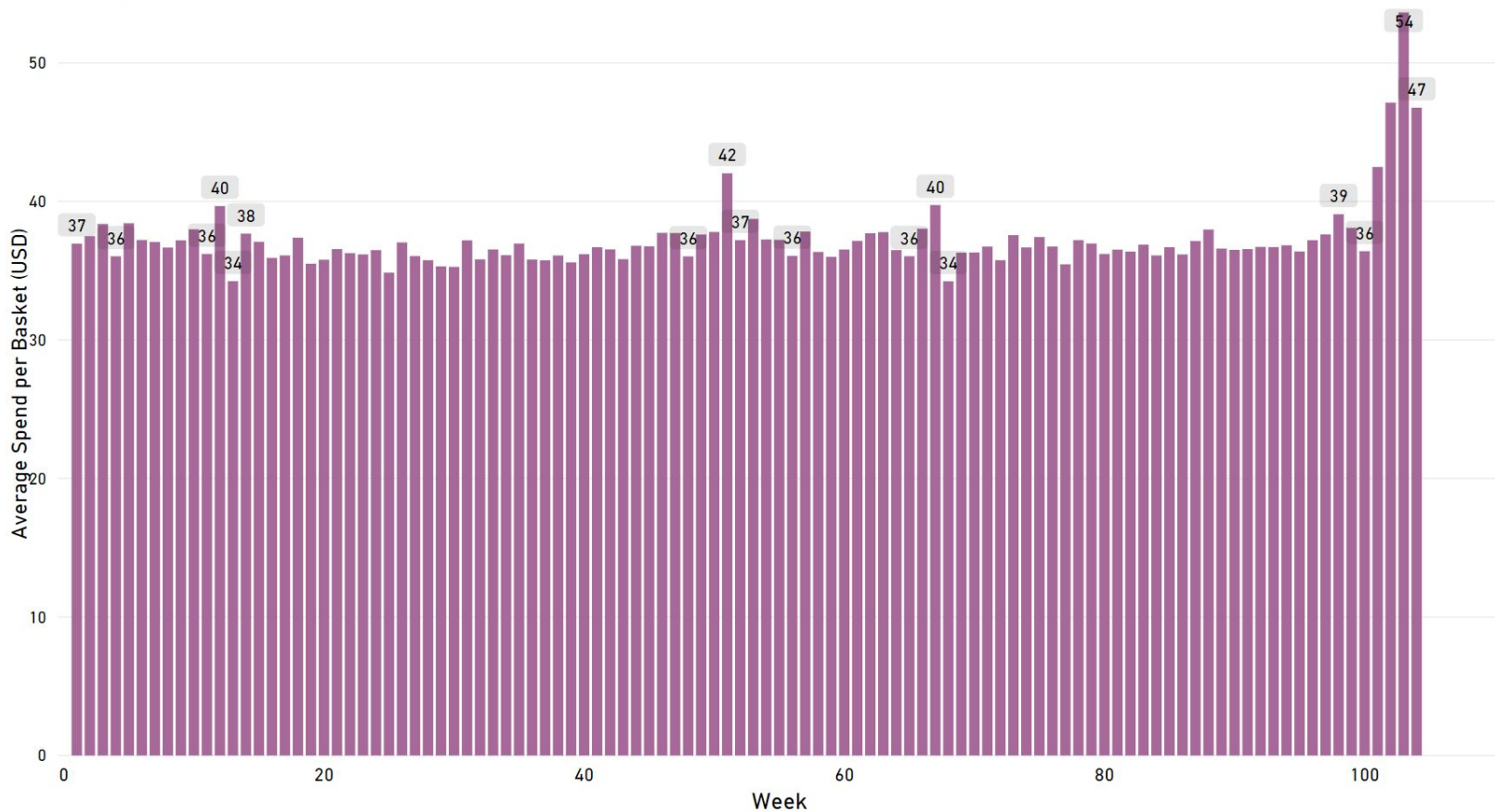

Total Spend per Week
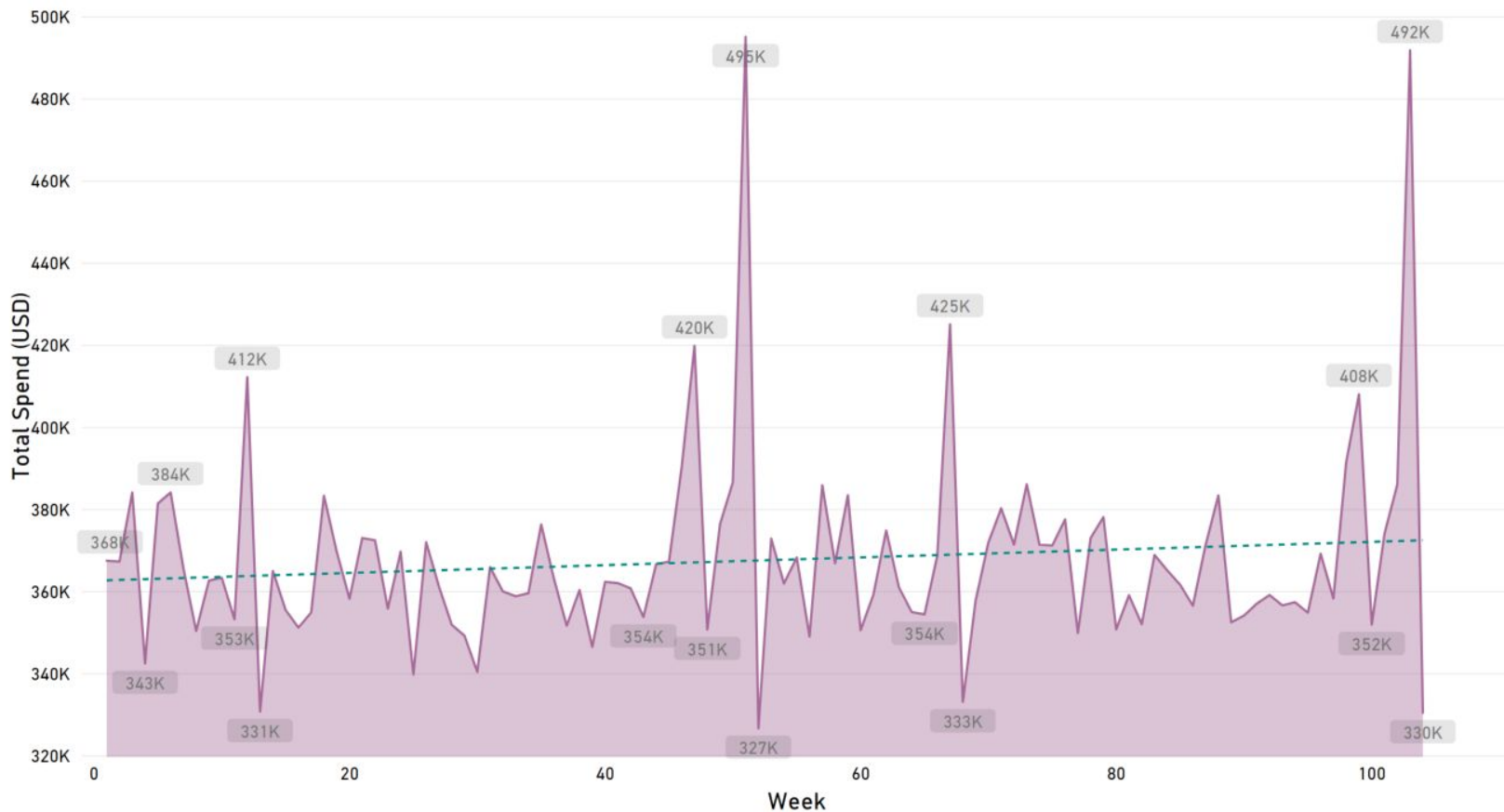
Are customers spending more or fewer dollars over time?
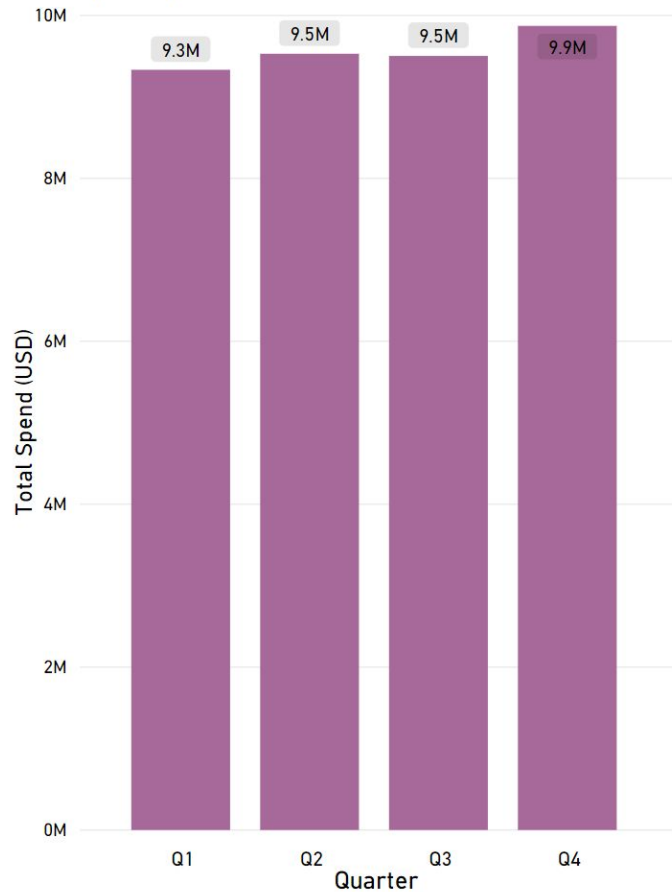

Count of # of Household by Change in Spend

Are customers spending more or fewer dollars over time?
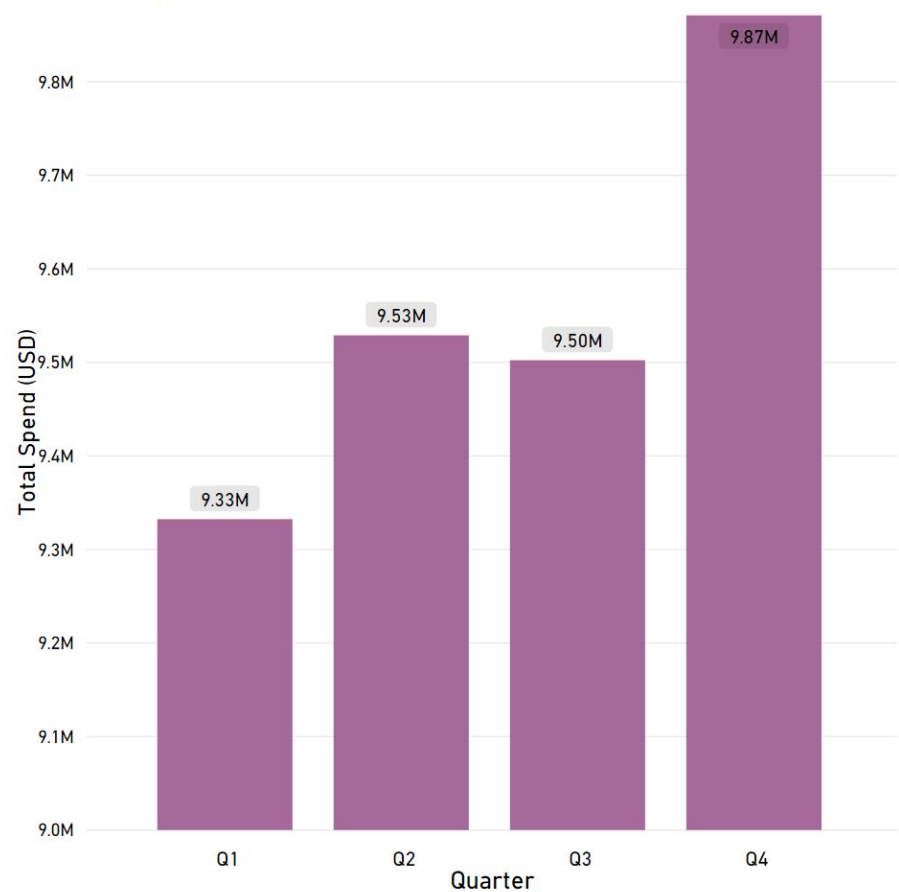

Average Spend per Basket by Week
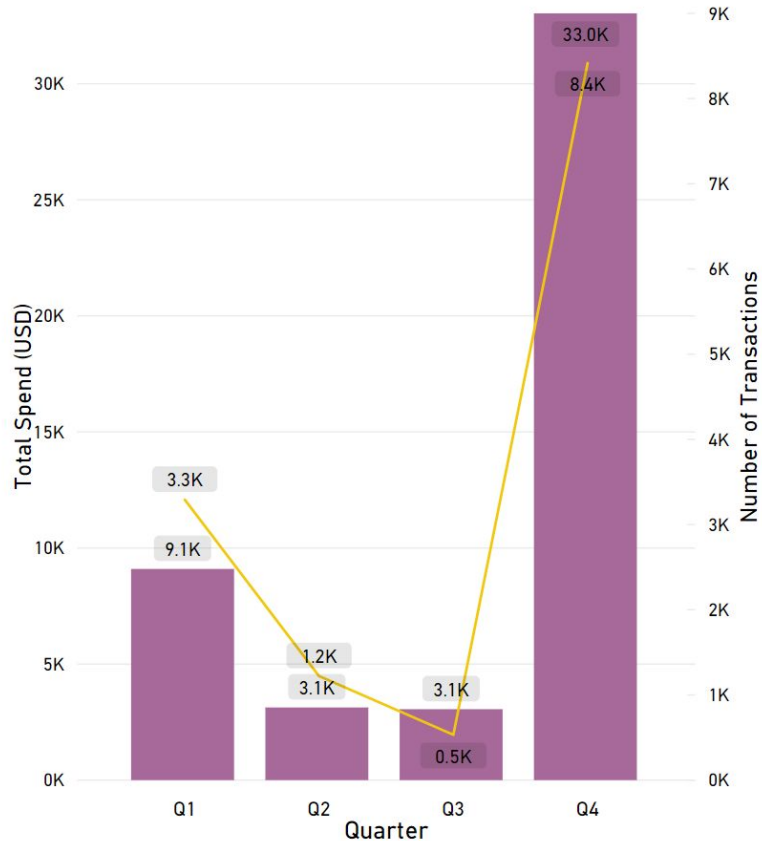
Seasonality in Data
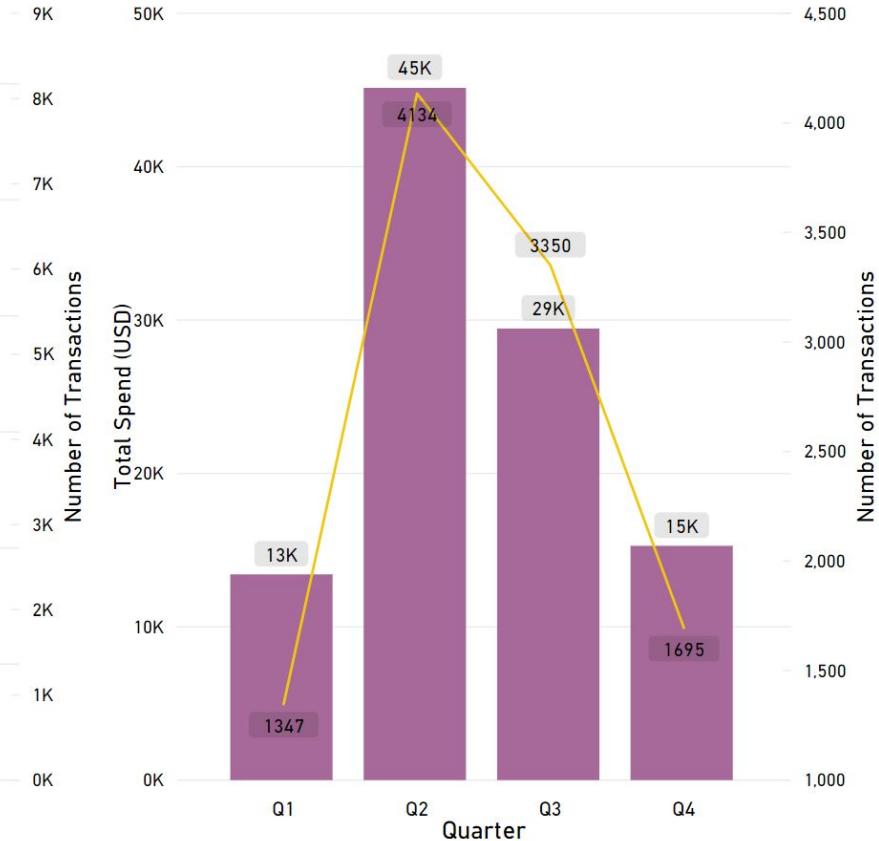

Total Spend per Quarter

Seasonality in Data

Total Spend per Quarter on Holiday Items

● Spend ● Number of Transactions

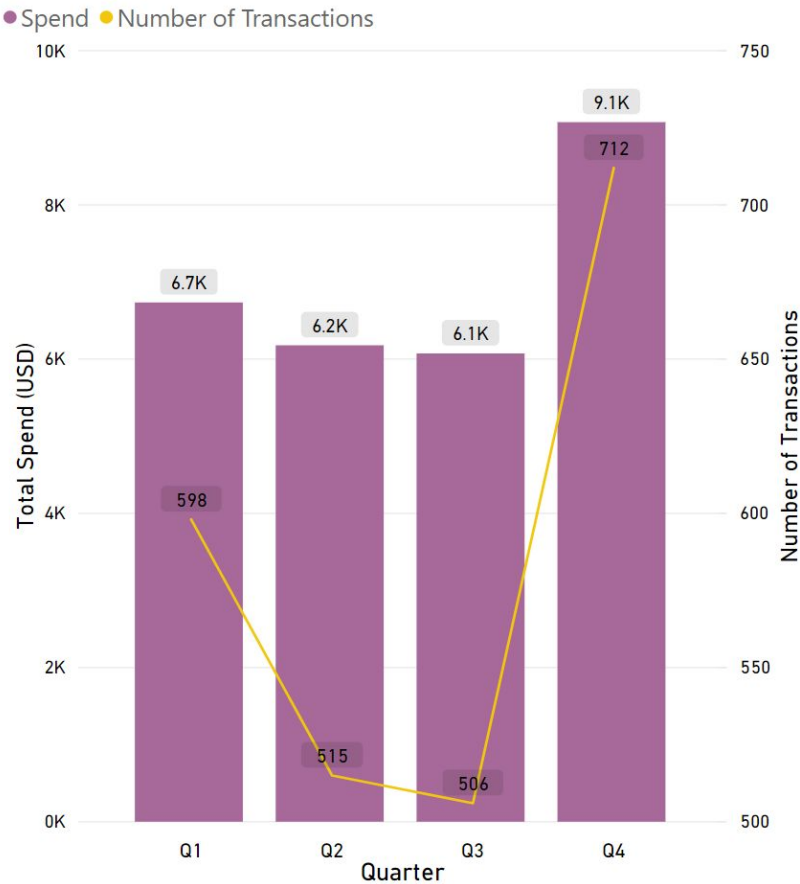| Quarter	Spend	Number of Transactions
Q1	9.1K	3.3K
Q2	3.1K	1.2K
Q3	3.1K	0.5K
Q4	33.0K	8.4K

Total Spend per Quarter on Activity / Outdoor Items

● Spend ● Number of Transactions

Quarter	Spend	Number of Transactions
Q1	13K	1347
Q2	45K	4134
Q3	29K	3350
Q4	15K	1695

Seasonality in Data

Total Spend per Quarter on Electronics

● Spend ● Number of Transactions

Total Spend (USD) — Number of Transactions

- Q1: 6.7K, 598
- Q2: 6.2K, 515
- Q3: 6.1K, 506
- Q4: 9.1K, 712

Total Spend per Quarter on Floral Items

● Spend ● Number of Transactions

Total Spend (USD) — Number of Transactions

- Q1: 78K, 7594
- Q2: 86K, 8540
- Q3: 50K, 5891
- Q4: 63K, 6758
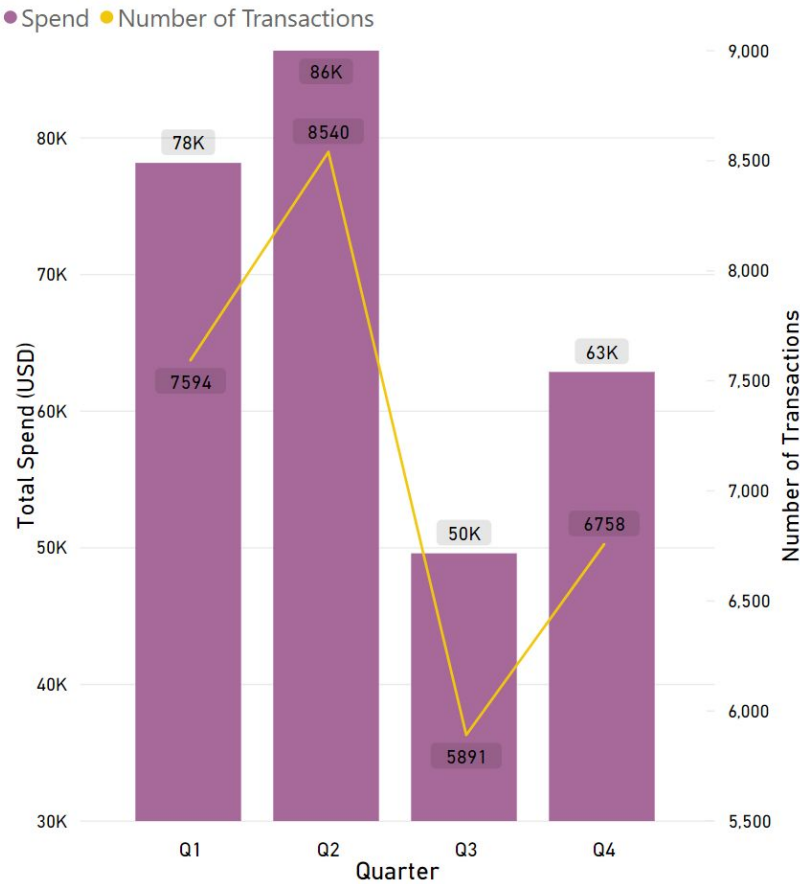
Geographic and Demographic Factors - Region

Total Spend by Region

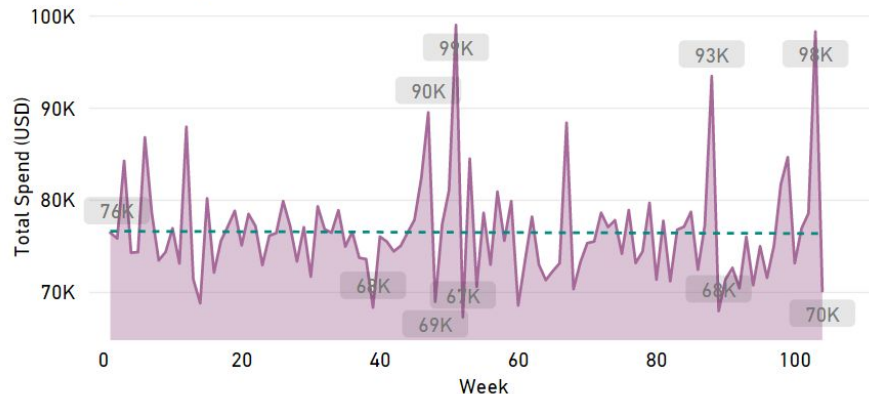● Spend ● Number of Transactions
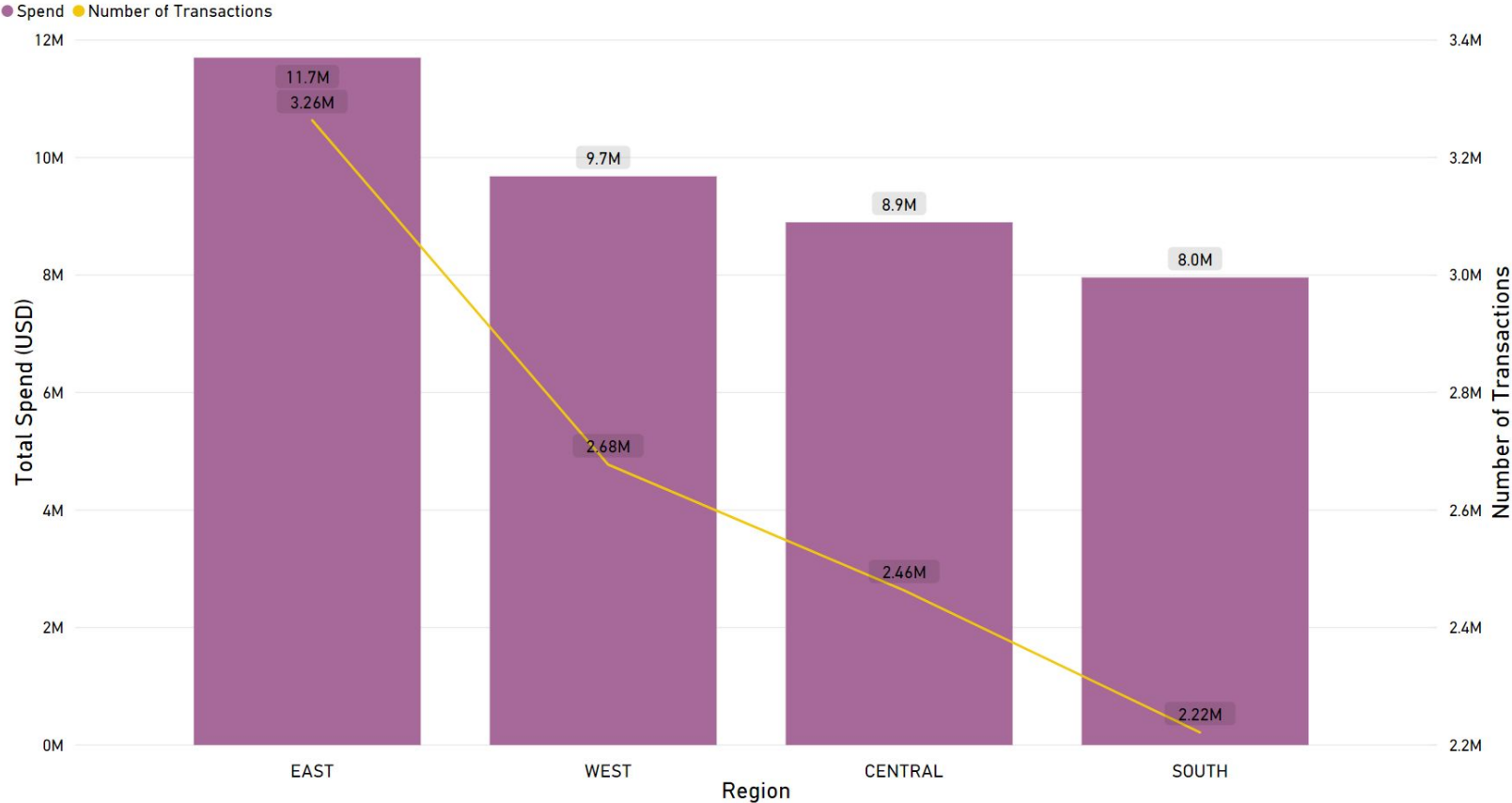
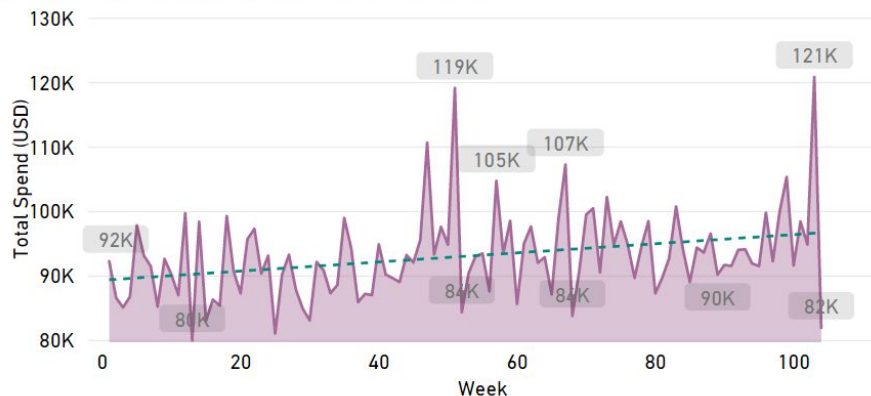# Geographic and Demographic Factors - Region

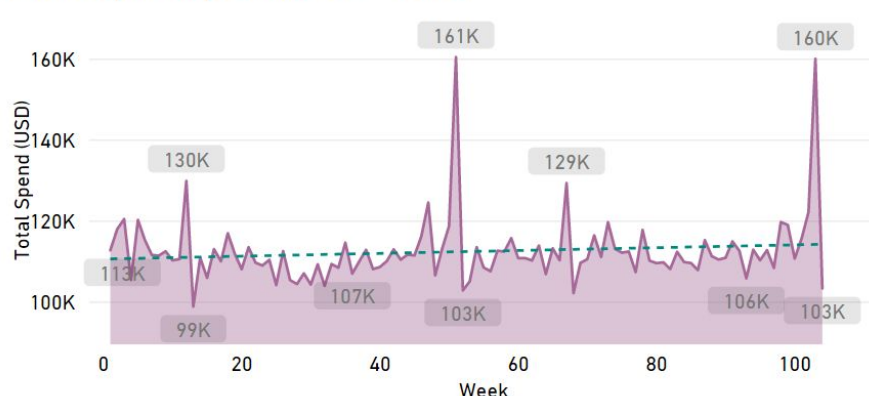**Total Spend per Week – Central**
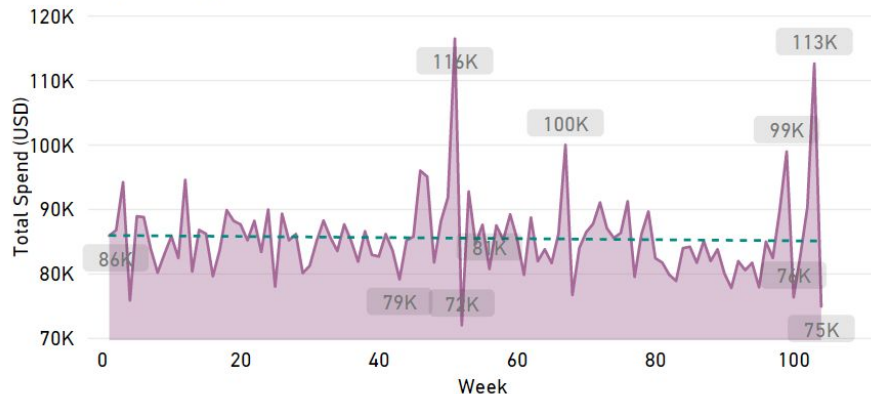
Total Spend per Week – South
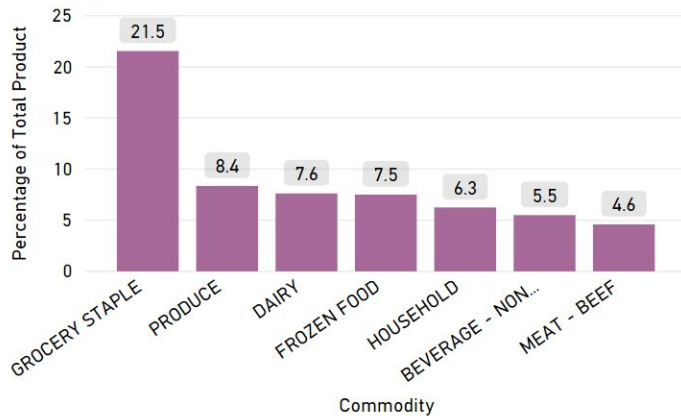
Total Spend per Week – West
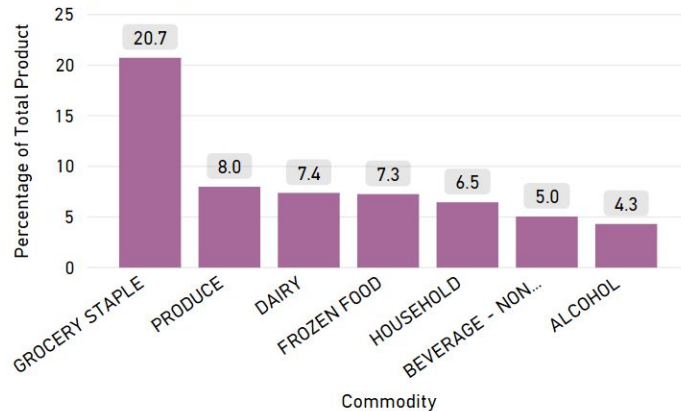
Total Spend per Week – East
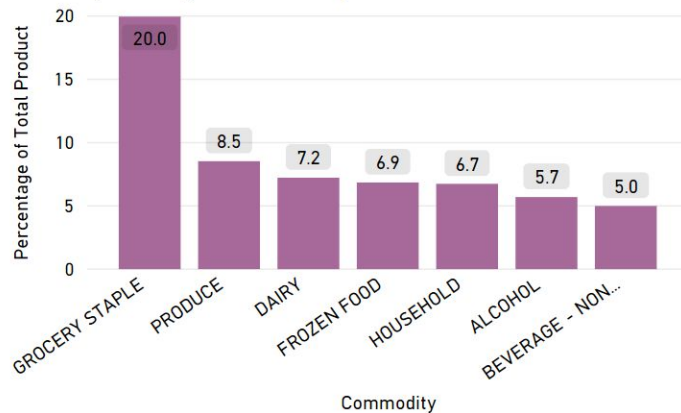
Geographic and Demographic Factors - Region
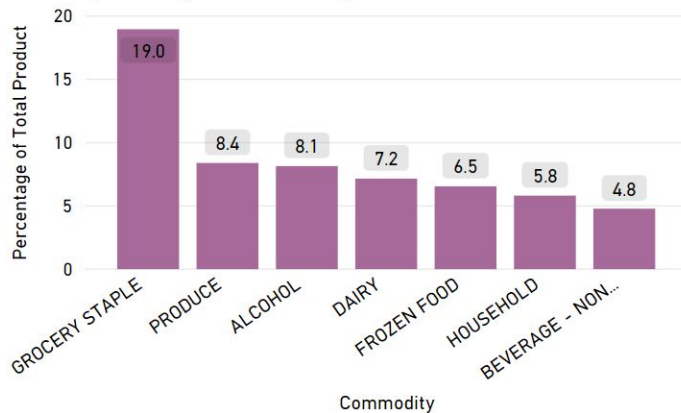

Total Spend by Commodity - Central

Total Spend by Commodity - East

Total Spend by Commodity - South
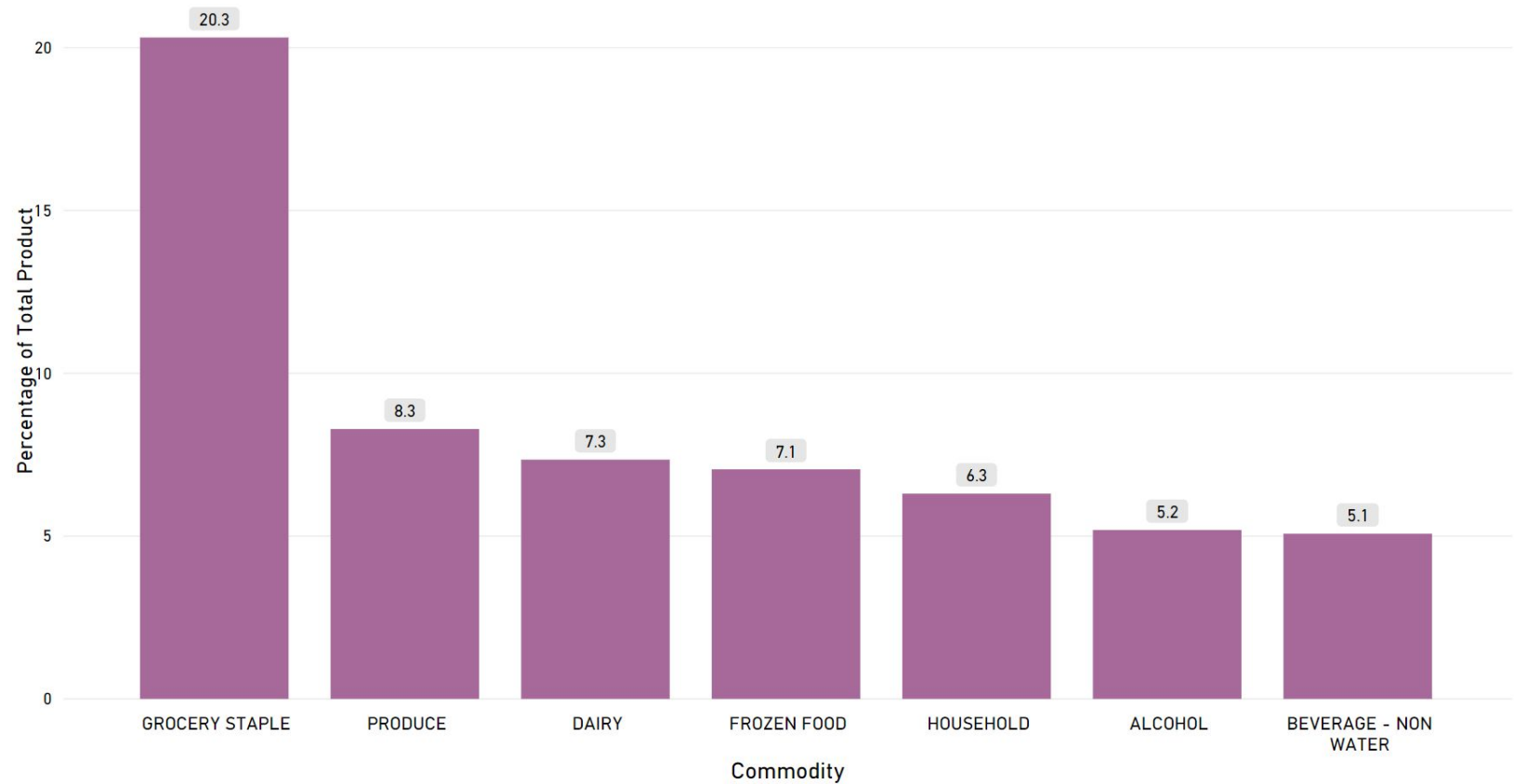
Total Spend by Commodity - West
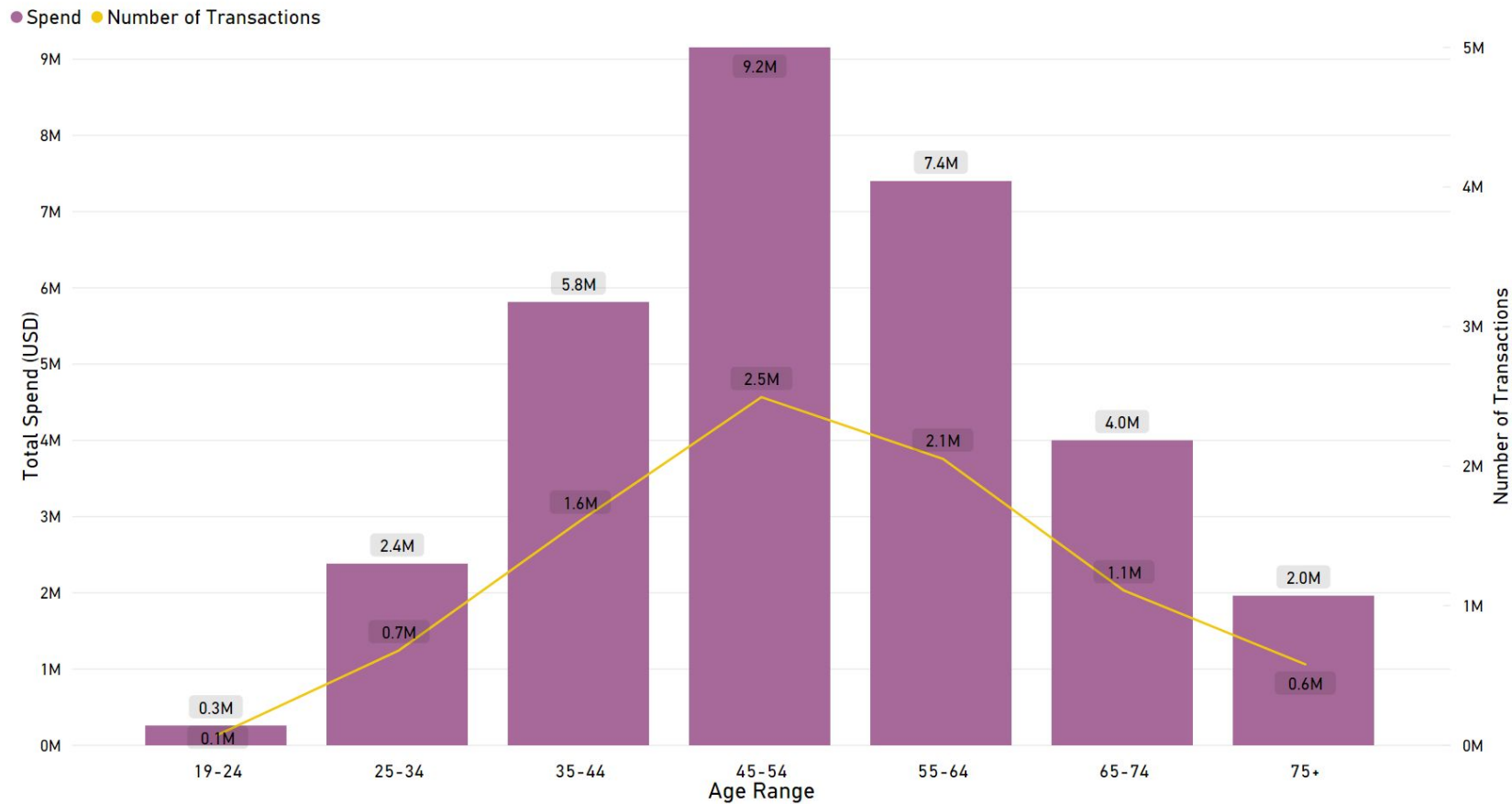
Geographic and Demographic Factors

Total Spend by Commodity (Top 7 Commodities)

Geographic and Demographic Factors - Age

Total Spend by Age Range

● Spend ● Number of Transactions

Geographic and Demographic Factors - Age

Total Spend per Week - Age 19 -24
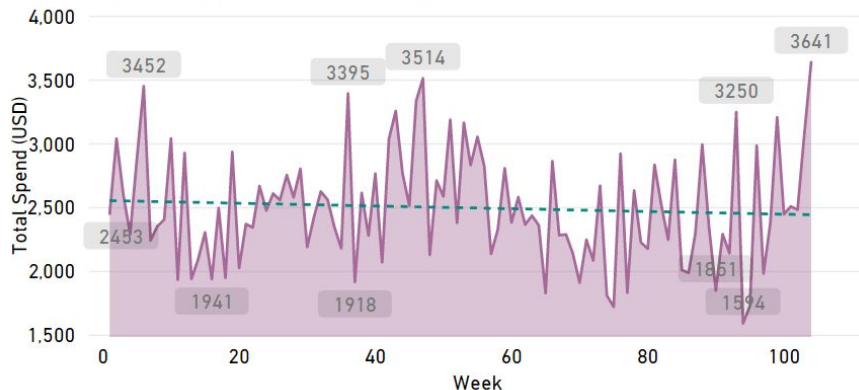
Total Spend per Week - Age 25 - 34

Total Spend per Week - Age 35 - 44

Total Spend per Week - Age 45 - 54

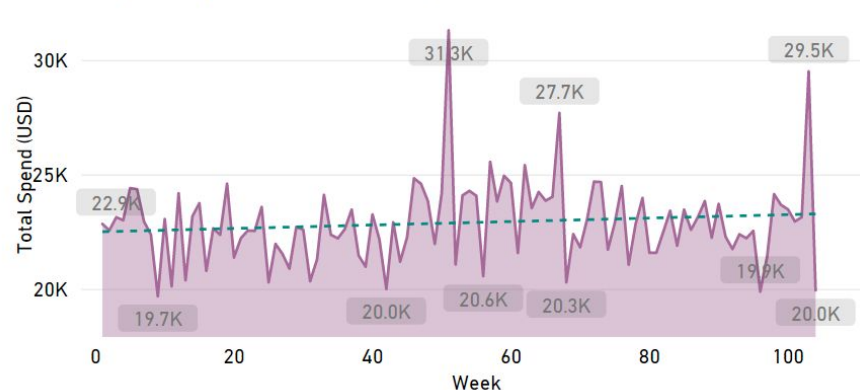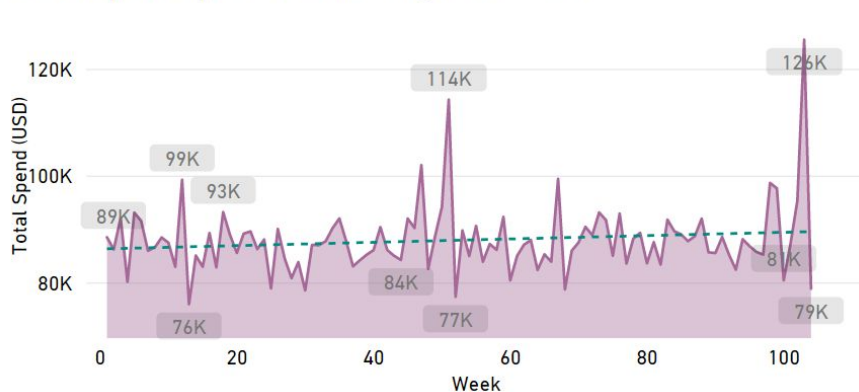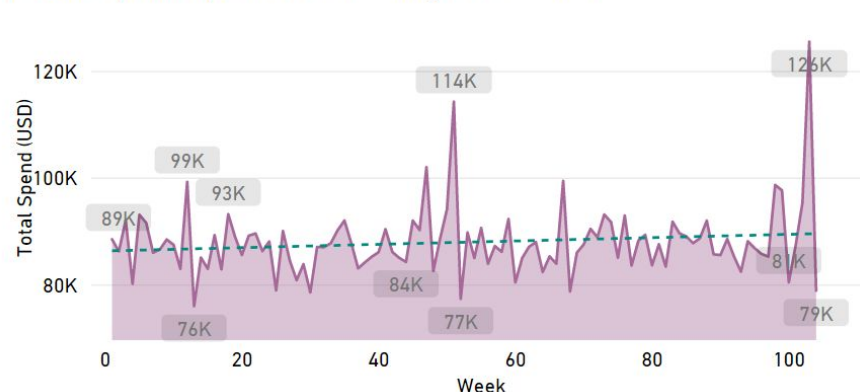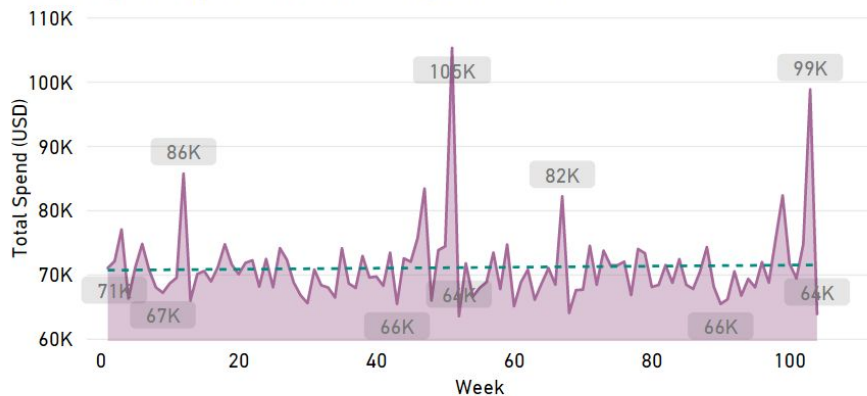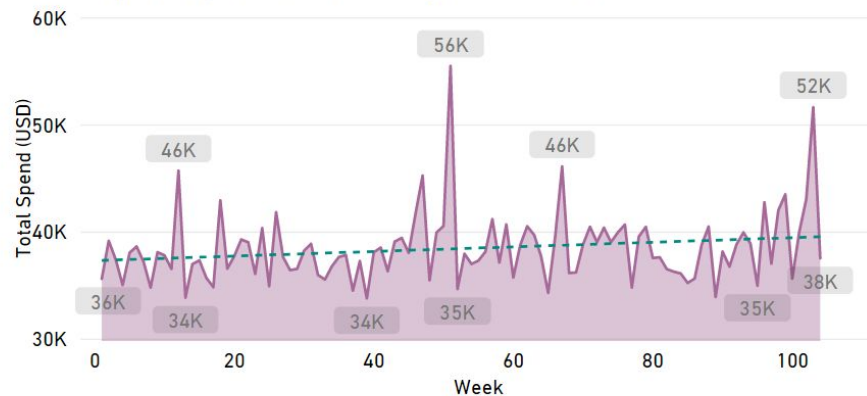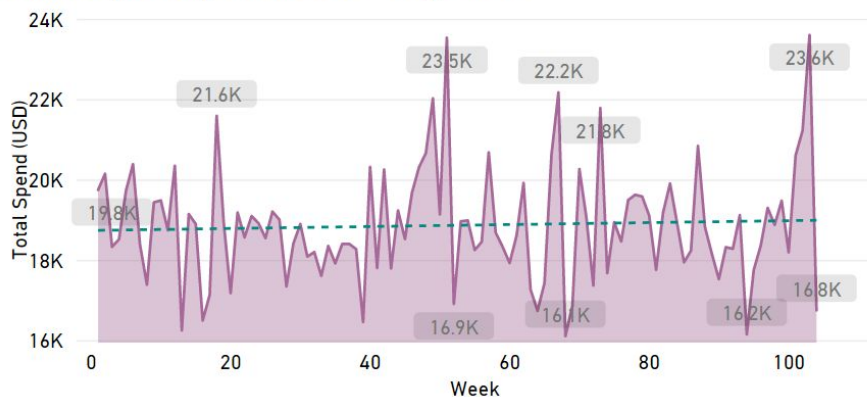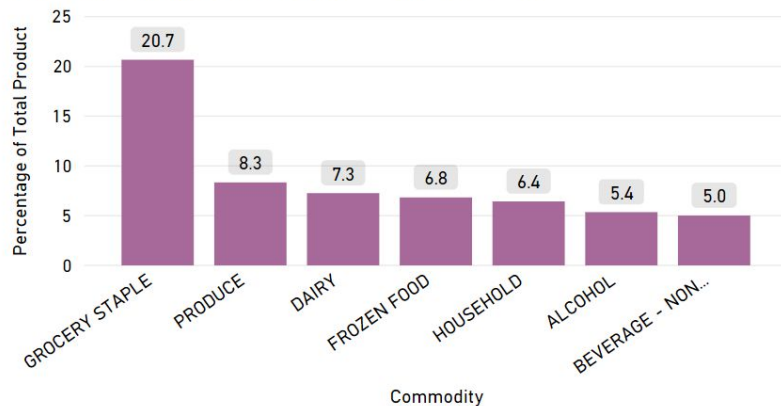# Geographic and Demographic Factors - Age

Total Spend per Week - Age 55 - 64
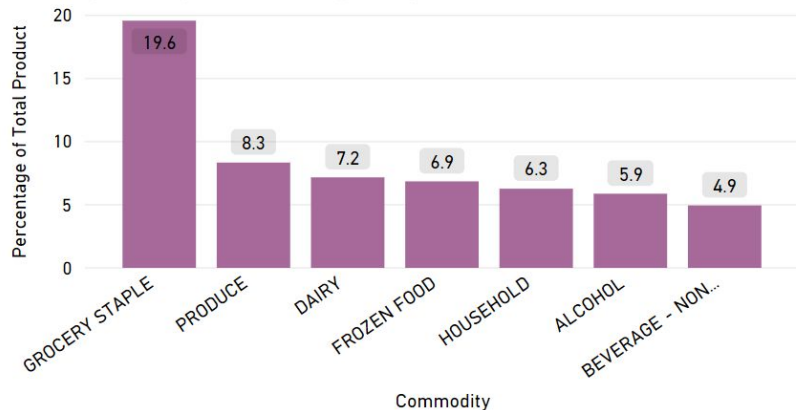
Total Spend per Week - Age 65 - 74

Total Spend per Week - Age 75+

Geographic and Demographic Factors - Age

Total Spend by Commodity – Age 19 – 24

Commodity	Percentage of Total Product
GROCERY STAPLE	21.5
PRODUCE	8.0
DAIRY	7.2
FROZEN FOOD	7.1
HOUSEHOLD	5.8
BEVERAGE – NON...	5.0
ALCOHOL	4.5

Total Spend by Commodity – Age 25 – 34

Commodity	Percentage of Total Product
GROCERY STAPLE	21.0
PRODUCE	8.0
DAIRY	7.6
FROZEN FOOD	7.3
HOUSEHOLD	6.1
BEVERAGE – NON...	5.5
ALCOHOL	4.5

Total Spend by Commodity – Age 35 – 44

Commodity	Percentage of Total Product
GROCERY STAPLE	21.5
PRODUCE	8.4
DAIRY	8.0
FROZEN FOOD	7.2
HOUSEHOLD	6.1
BEVERAGE – NON...	5.3
ALCOHOL	4.1

Total Spend by Commodity – Age 45 – 54

Commodity	Percentage of Total Product
GROCERY STAPLE	20.7
PRODUCE	8.3
DAIRY	7.3
FROZEN FOOD	6.8
HOUSEHOLD	6.4
ALCOHOL	5.4
BEVERAGE – NON...	5.0

Geographic and Demographic Factors - Age

Total Spend by Commodity – Age 55 – 64

Commodity	Percentage of Total Product
GROCERY STAPLE	19.6
PRODUCE	8.3
DAIRY	7.2
FROZEN FOOD	6.9
HOUSEHOLD	6.3
ALCOHOL	5.9
BEVERAGE - NON...	4.9

Total Spend by Commodity – Age 65 – 74

Commodity	Percentage of Total Product
GROCERY STAPLE	19.6
PRODUCE	8.6
DAIRY	7.3
FROZEN FOOD	7.1
HOUSEHOLD	6.4
ALCOHOL	5.5
BEVERAGE - NON...	4.5

Total Spend by Commodity – Age 75+

Commodity	Percentage of Total Product
GROCERY STAPLE	19.8
PRODUCE	8.4
FROZEN FOOD	7.8
HOUSEHOLD	7.0
DAIRY	6.9
BAKERY	5.2
BEVERAGE - NON...	4.9

Geographic and Demographic Factors - Marital Status

Total Spend by Marital Status

● Spend ● Number of Transactions

Total Spend (USD) — left axis
Number of Transactions — right axis

Married: 18.7M, 5.1M
Single: 10.6M, 3.0M

Marital Status

Geographic and Demographic Factors - Marital Status

Total Spend per Week - Single
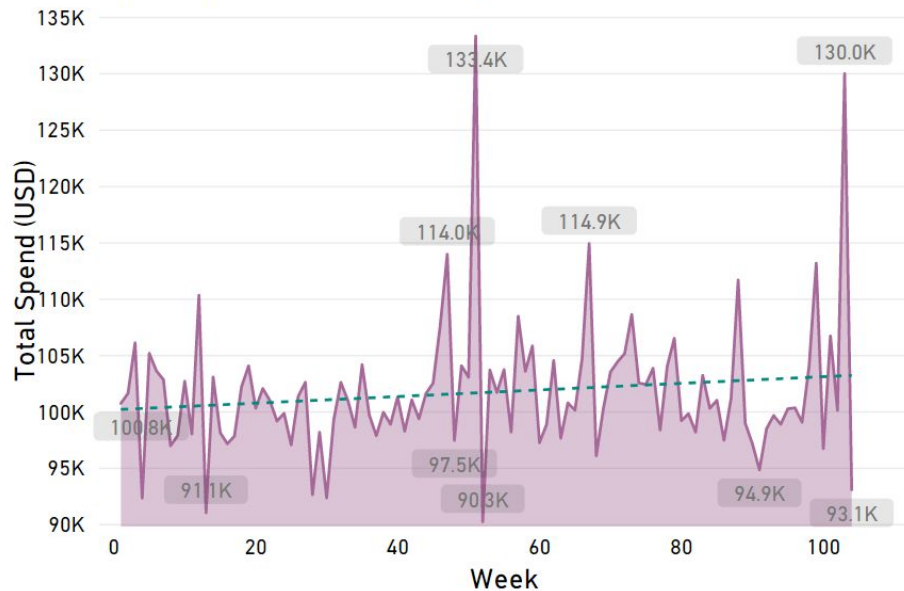
Total Spend per Week - Married

Geographic and Demographic Factors - Marital Status
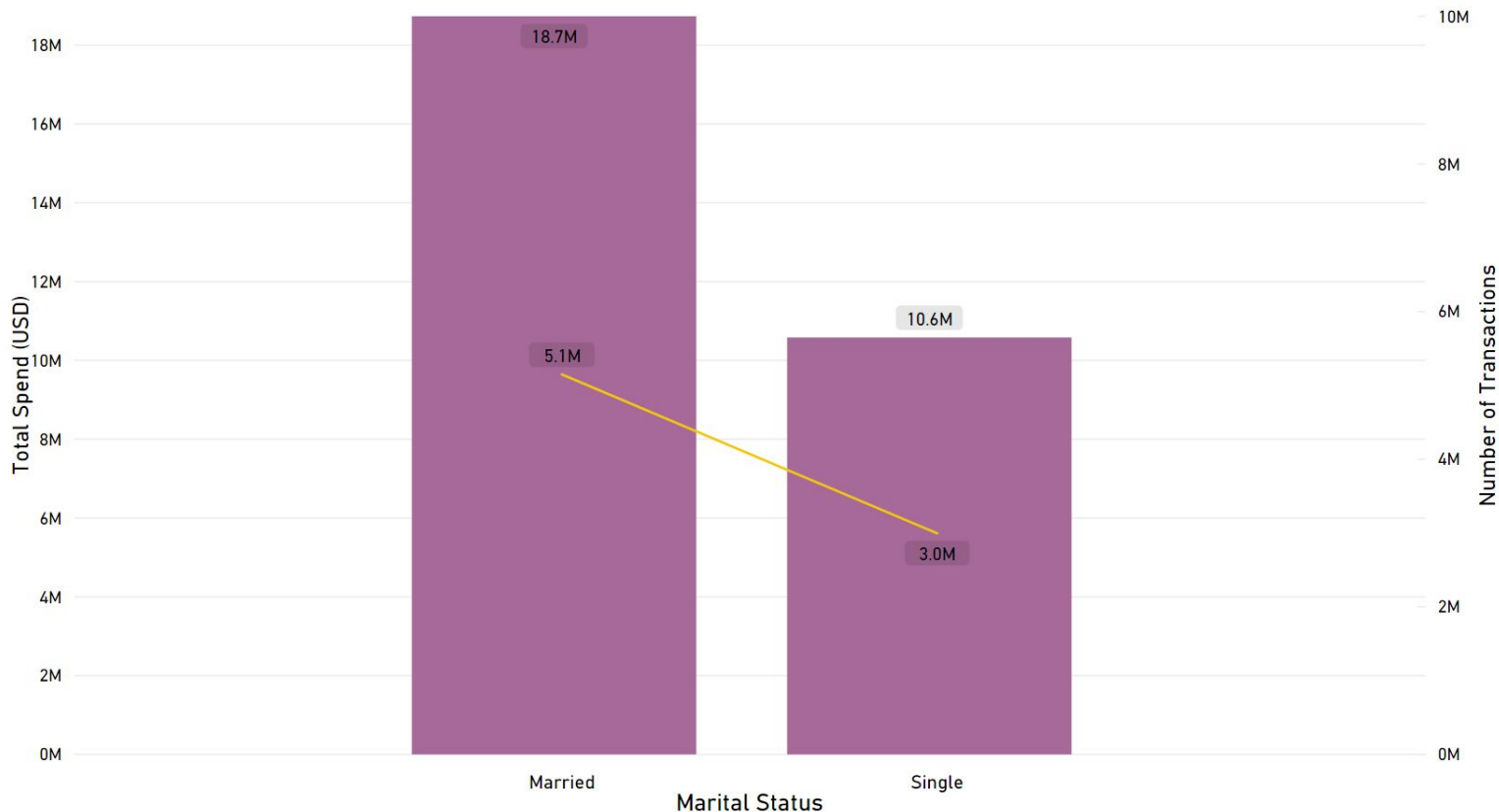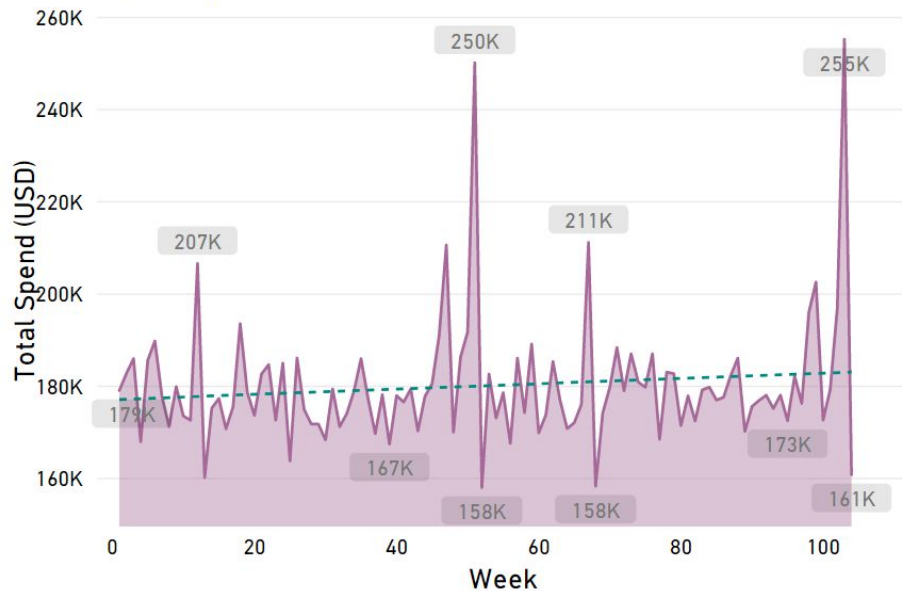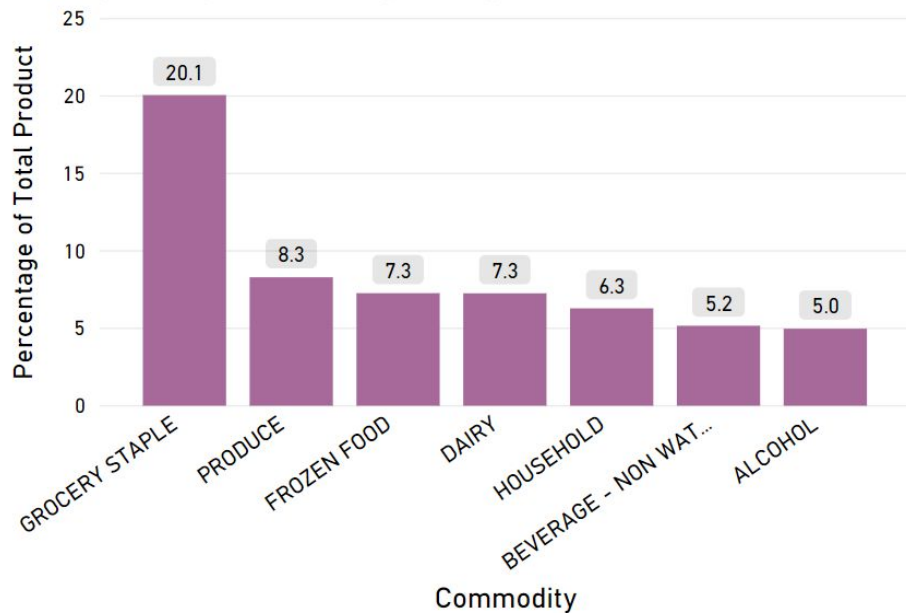

Total Spend by Commodity - Single

Total Spend by Commodity - Married
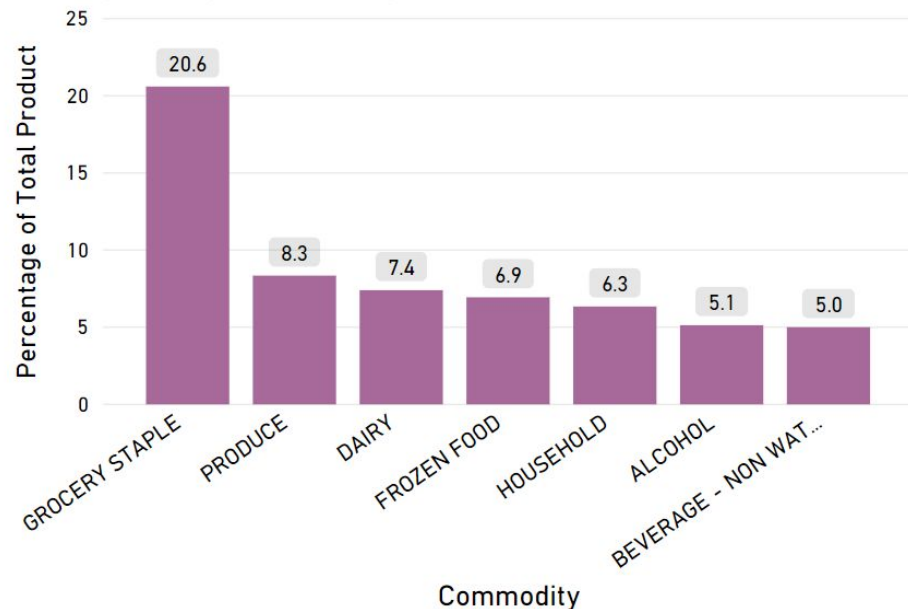
Geographic and Demographic Factors - Home Status

Total Spend by Home Status

● Spend ● Number of Transactions

24.5M

6.8M

3.5M

1.0M

Total Spend (USD)

Number of Transactions

Homeowner

Renter

Home Status

Geographic and Demographic Factors - Home Status

Total Spend per Week – Homeowner

Total Spend per Week – Renter

Geographic and Demographic Factors - Home Status

Total Spend by Commodity - Homeowner

Percentage of Total Product vs Commodity

- GROCERY STAPLE: 20.5
- PRODUCE: 8.5
- DAIRY: 7.4
- FROZEN FOOD: 6.9
- HOUSEHOLD: 6.3
- ALCOHOL: 5.2
- BEVERAGE - NON WAT...: 4.9
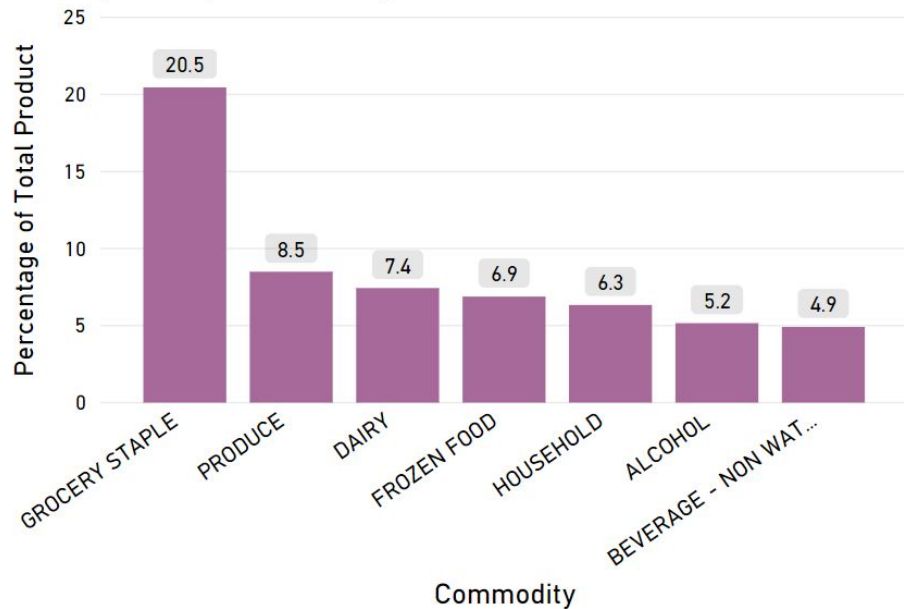
Total Spend by Commodity - Renter

Percentage of Total Product vs Commodity

- GROCERY STAPLE: 20.0
- FROZEN FOOD: 7.9
- PRODUCE: 7.6
- DAIRY: 7.1
- HOUSEHOLD: 6.2
- BEVERAGE - NON WAT...: 5.6
- ALCOHOL: 4.2
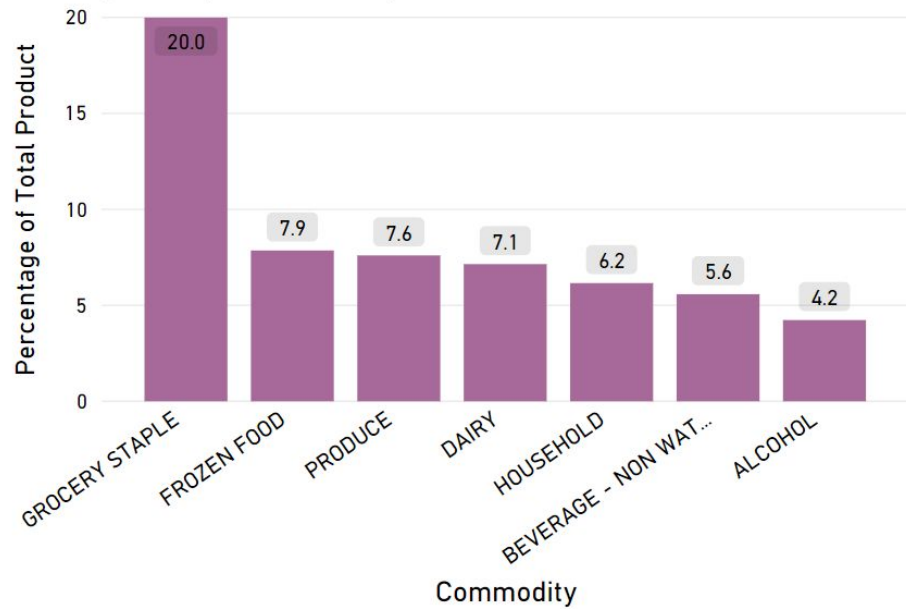
Geographic and Demographic Factors - Income Level

Total Spend by Income Range

● Spend ● Number of Transactions
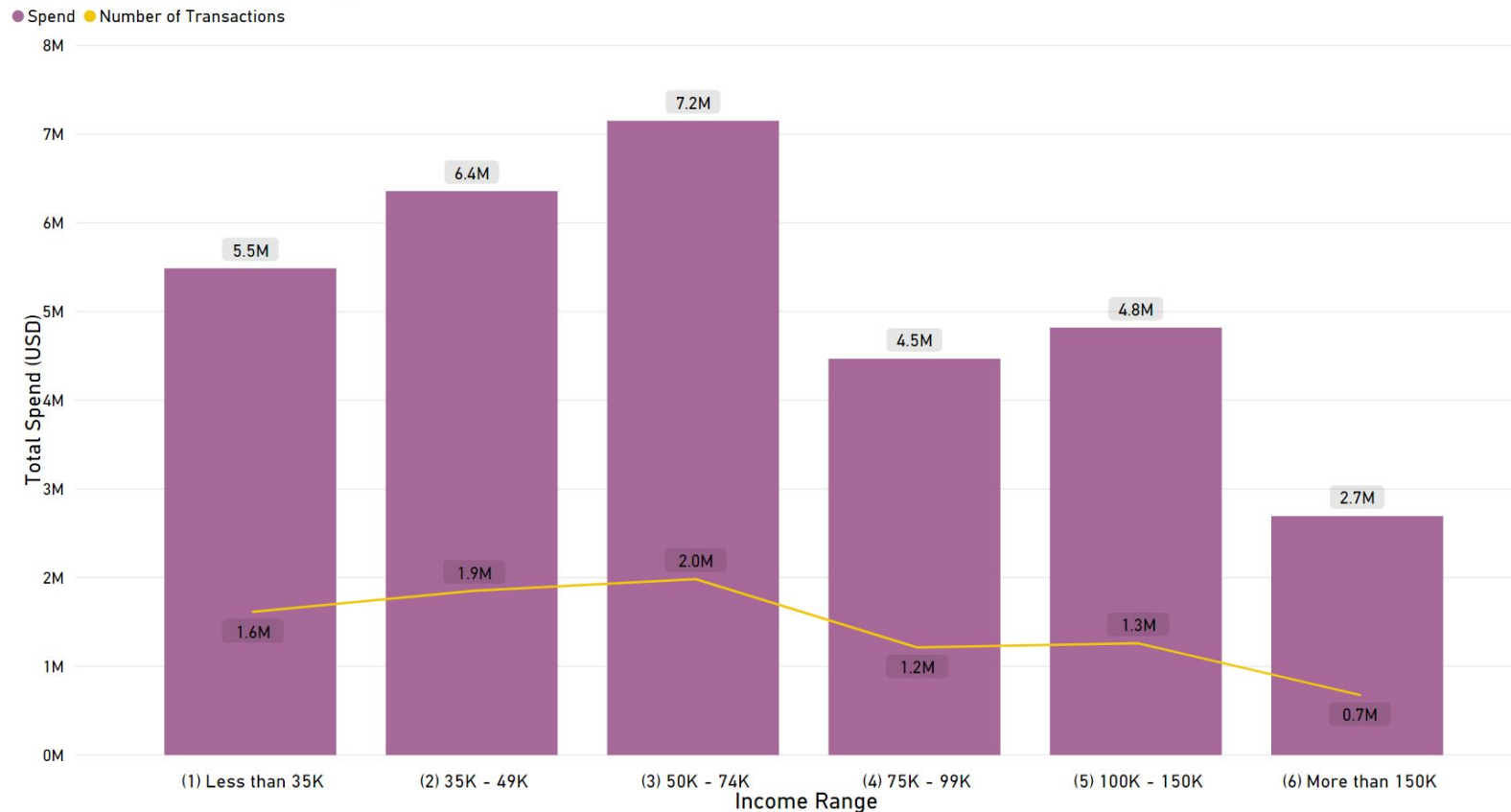
Geographic and Demographic Factors - Income Level

Total Spend per Week – Less than 35K
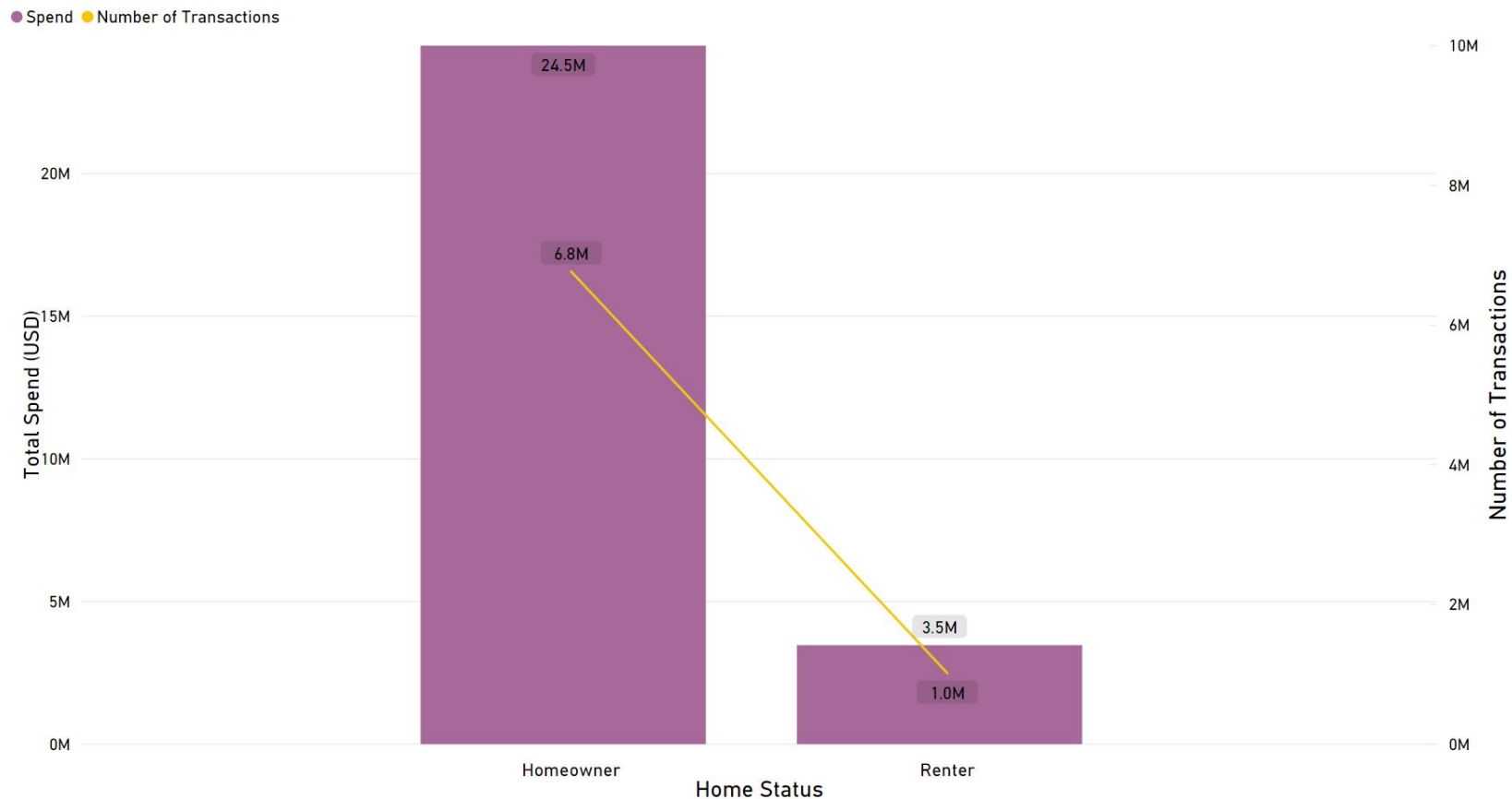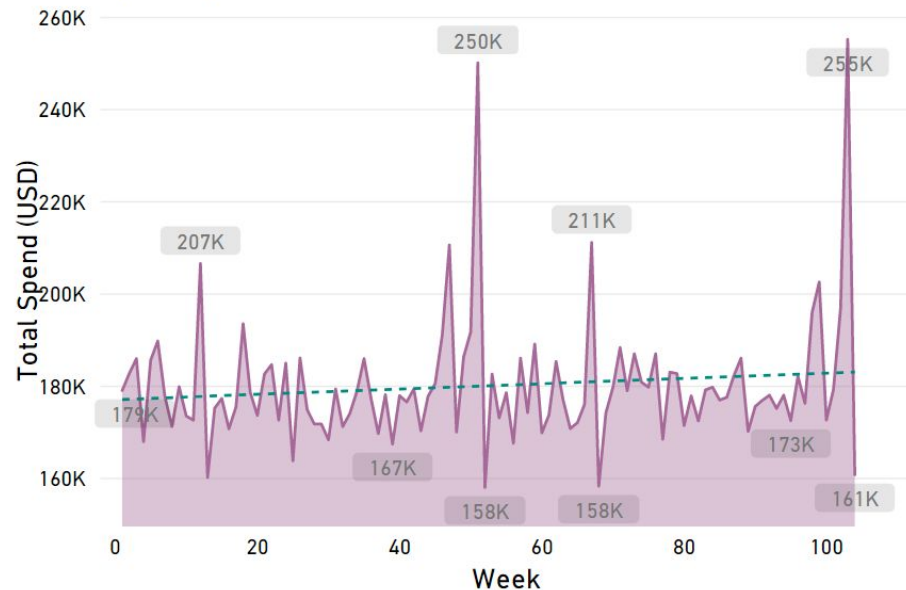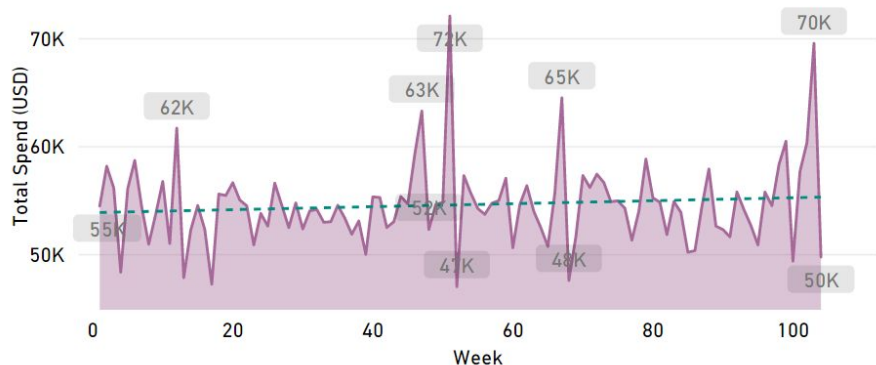
Total Spend per Week – 35K–49K
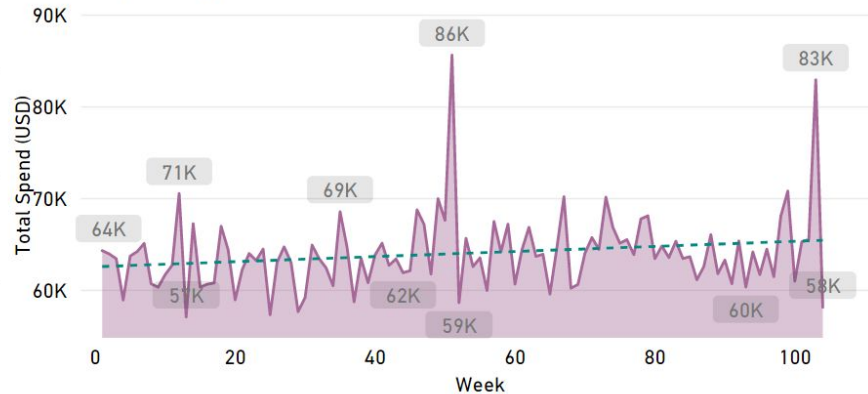
Total Spend per Week – 50K–74K
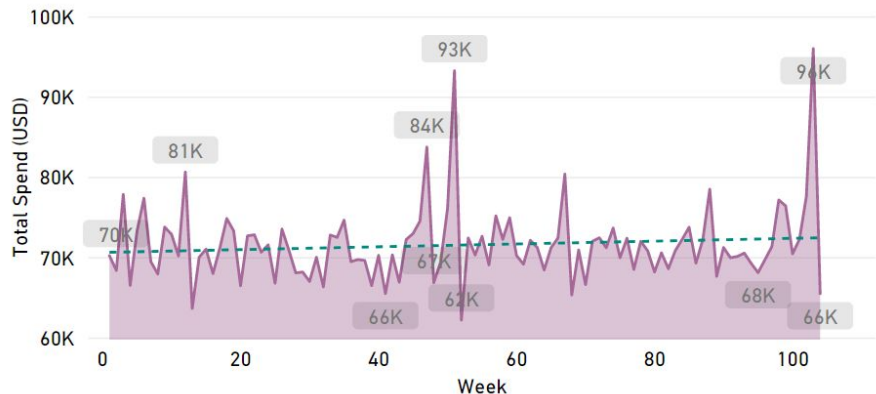
Total Spend per Week – 75K–99K
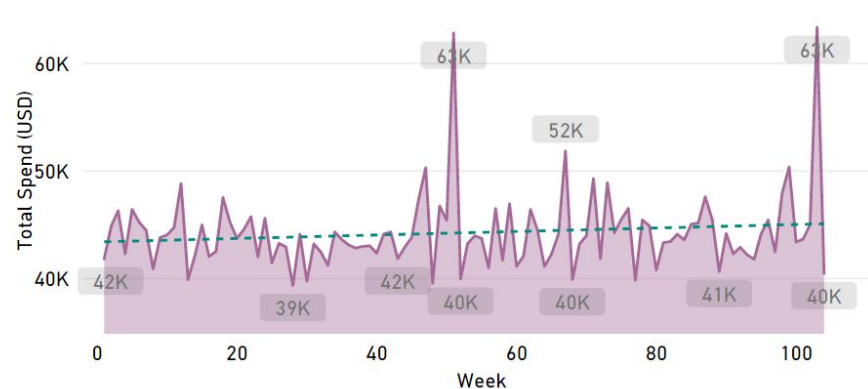
Geographic and Demographic Factors - Income Level

Total Spend per Week – 100K-150K

Total Spend (USD) vs Week

- 47.3K
- 54.2K
- 44.9K
- 42.7K
- 69.2K
- 55.6K
- 41.2K
- 42.3K
- 69.9K
- 41.6K

Total Spend per Week – More than 150K

Total Spend (USD) vs Week

- 27.7K
- 31.5K
- 24.4K
- 22.6K
- 37.5K
- 22.9K
- 33.0K
- 22.0K
- 37.3K
- 21.6K
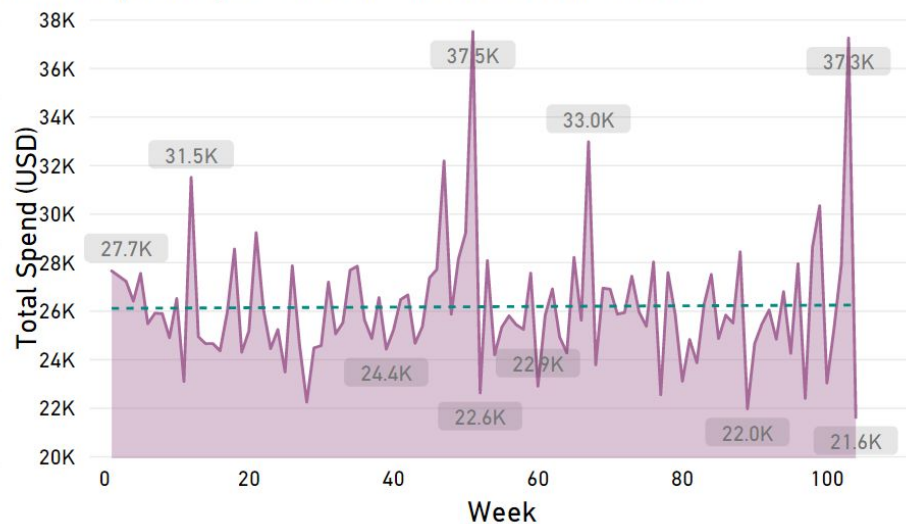
Geographic and Demographic Factors - Income Level

Total Spend by Commodity - Less than 35K

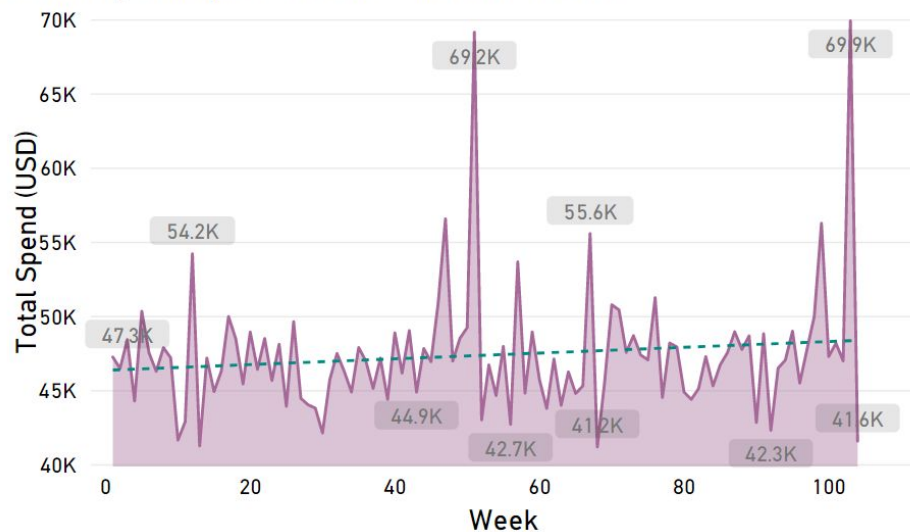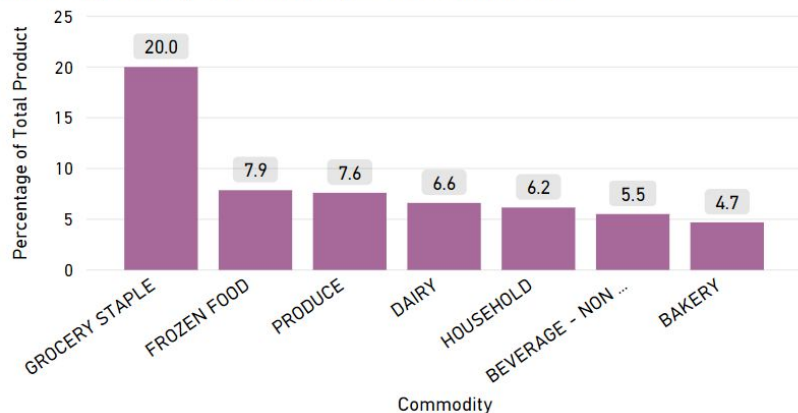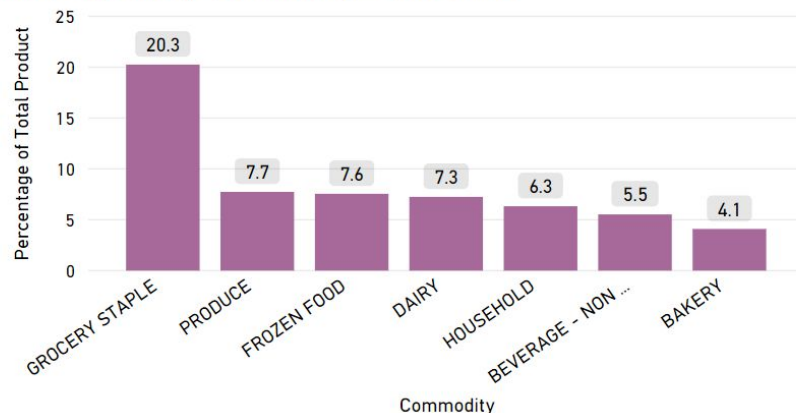| Commodity	Percentage of Total Product
GROCERY STAPLE	20.0
FROZEN FOOD	7.9
PRODUCE	7.6
DAIRY	6.6
HOUSEHOLD	6.2
BEVERAGE - NON ...	5.5
BAKERY	4.7

Total Spend by Commodity - 35K-49K

Commodity	Percentage of Total Product
GROCERY STAPLE	20.3
PRODUCE	7.7
FROZEN FOOD	7.6
DAIRY	7.3
HOUSEHOLD	6.3
BEVERAGE - NON ...	5.5
BAKERY	4.1

Total Spend by Commodity - 50K-74K

Commodity	Percentage of Total Product
GROCERY STAPLE	20.5
PRODUCE	7.9
DAIRY	7.3
FROZEN FOOD	7.1
HOUSEHOLD	6.4
BEVERAGE - NON ...	5.1
ALCOHOL	5.1
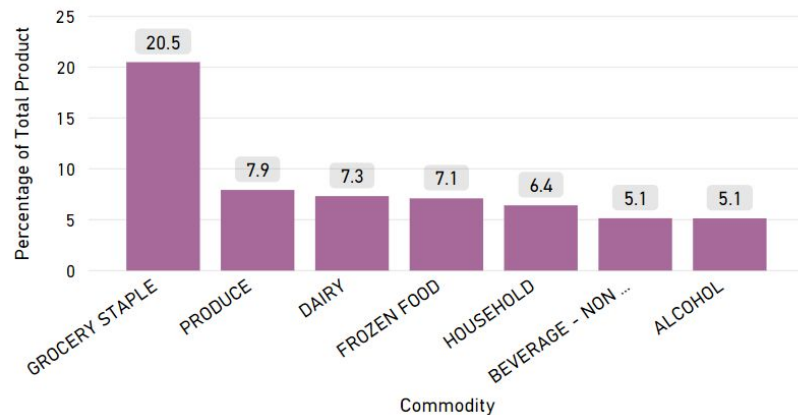
Total Spend by Commodity - 75K-99K

Commodity	Percentage of Total Product
GROCERY STAPLE	20.5
PRODUCE	8.8
DAIRY	7.8
FROZEN FOOD	6.8
ALCOHOL	6.1
HOUSEHOLD	6.1
BEVERAGE - NON ...	4.7
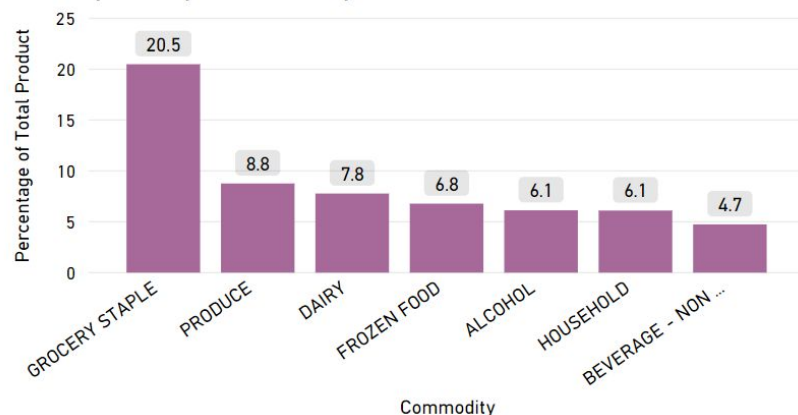
Geographic and Demographic Factors - Income Level

Total Spend by Commodity - 100K-150K

Commodity	Percentage of Total Product
GROCERY STAPLE	20.6
PRODUCE	9.2
DAIRY	7.7
HOUSEHOLD	6.4
FROZEN FOOD	6.4
ALCOHOL	6.1
BEVERAGE - NON WATER	4.4
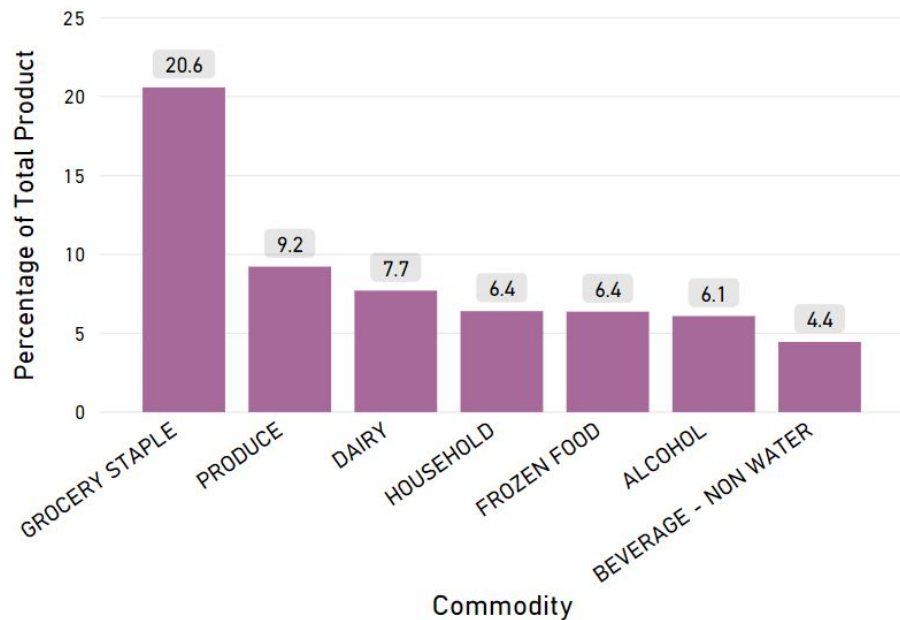
Total Spend by Commodity - More than 150K

Commodity	Percentage of Total Product
GROCERY STAPLE	20.8
PRODUCE	10.2
DAIRY	8.2
ALCOHOL	6.9
HOUSEHOLD	6.5
FROZEN FOOD	5.5
BEVERAGE - NON WATER	3.9
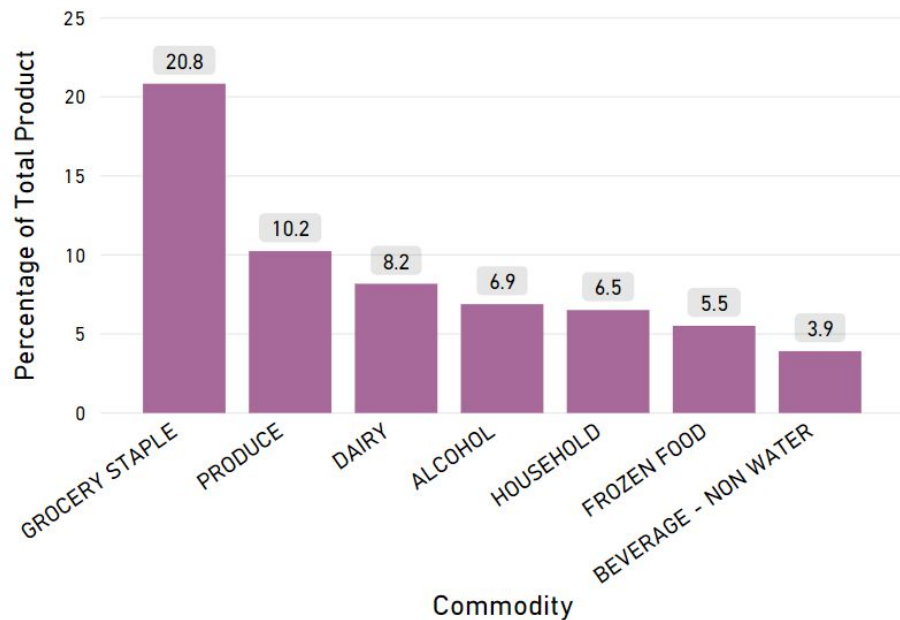
Forecasting Model 1

- **Cross-validation**
- **Elastic net regularization**
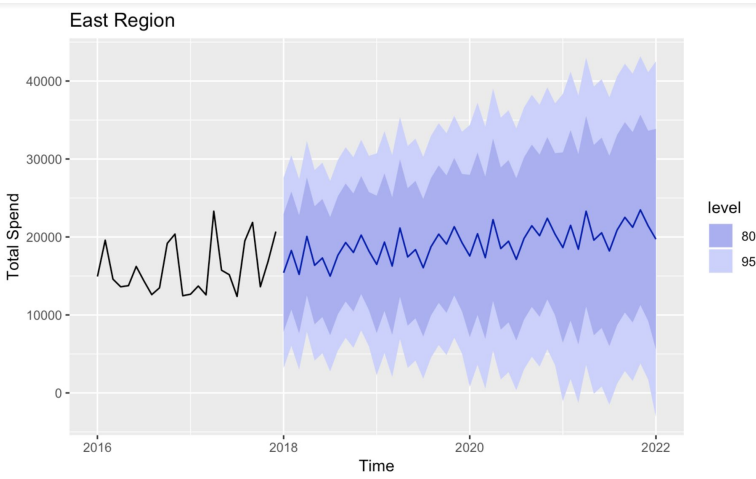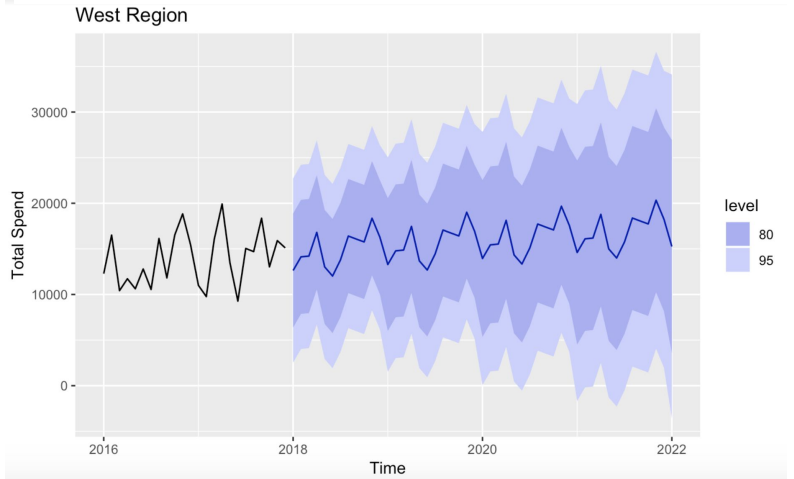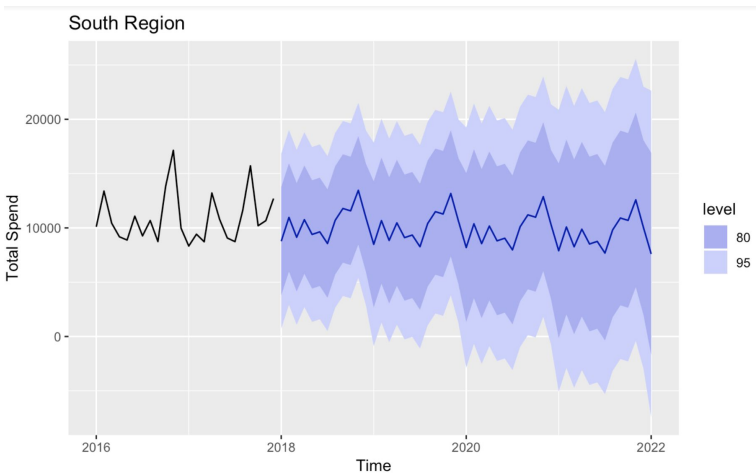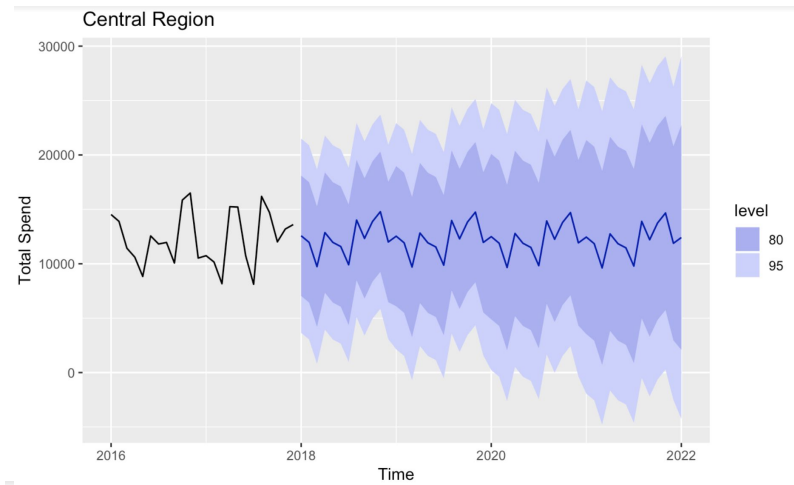 - **Root-mean-square error (RMSE)**
  - **Multiple R-squared**

alpha	lambda	RMSE	Rsquared	MAE	RMSESD	RsquaredSD	MAESD
0.9	6.16	1453.37	0.94	1039.37	63.48	0	41.57

Forecasting Model 2
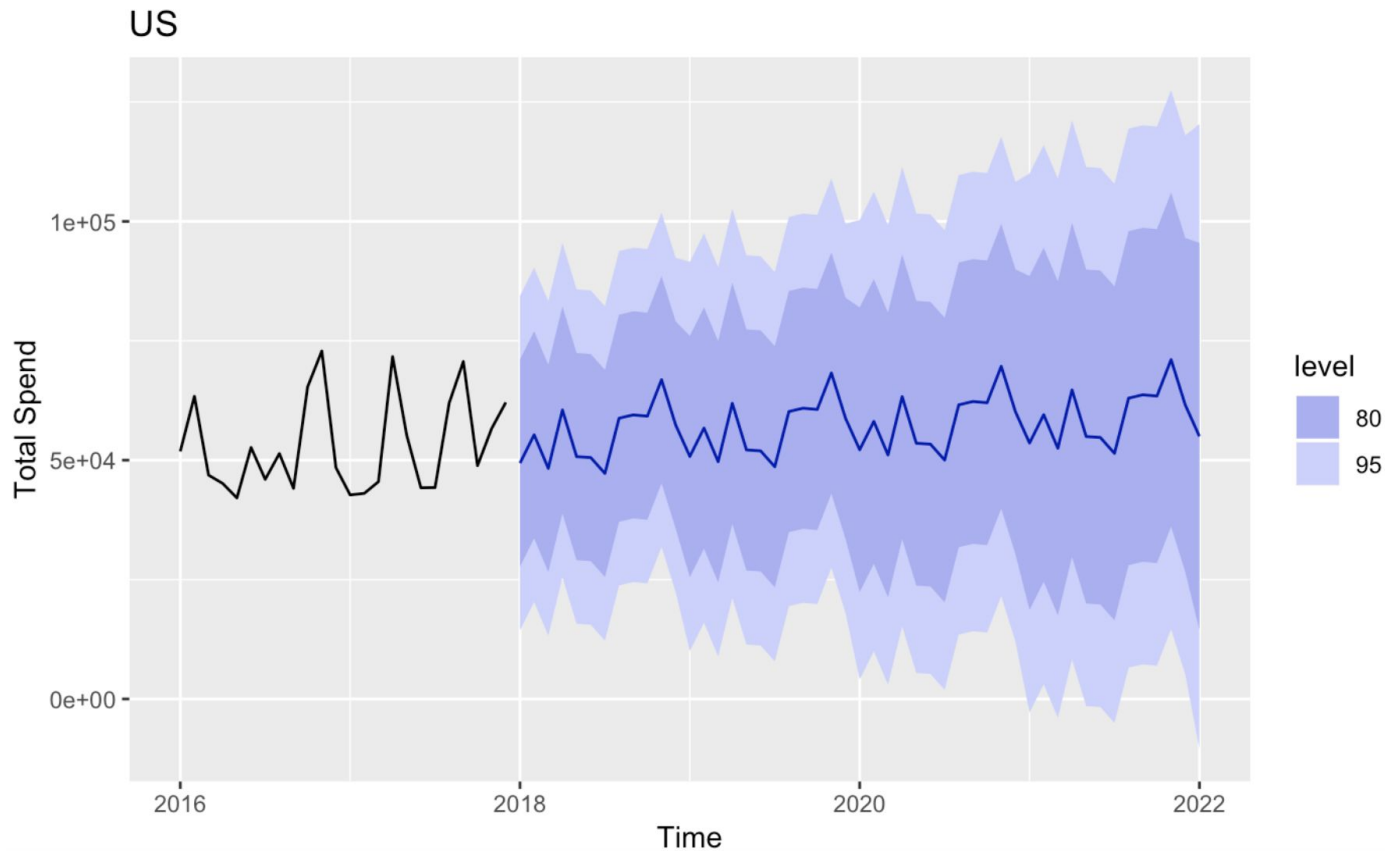
Forecasting Model 2

Anomalies and Outliers

Total Spend per Week

Y-axis: Total Spend (USD) — 300K, 350K, 400K, 450K, 500K

X-axis: Week — 0, 20, 40, 60, 80, 100

Data labels: 368K, 384K, 343K, 353K, 331K, 412K, 354K, 351K, 420K, 495K, 327K, 425K, 354K, 333K, 408K, 352K, 492K, 330K
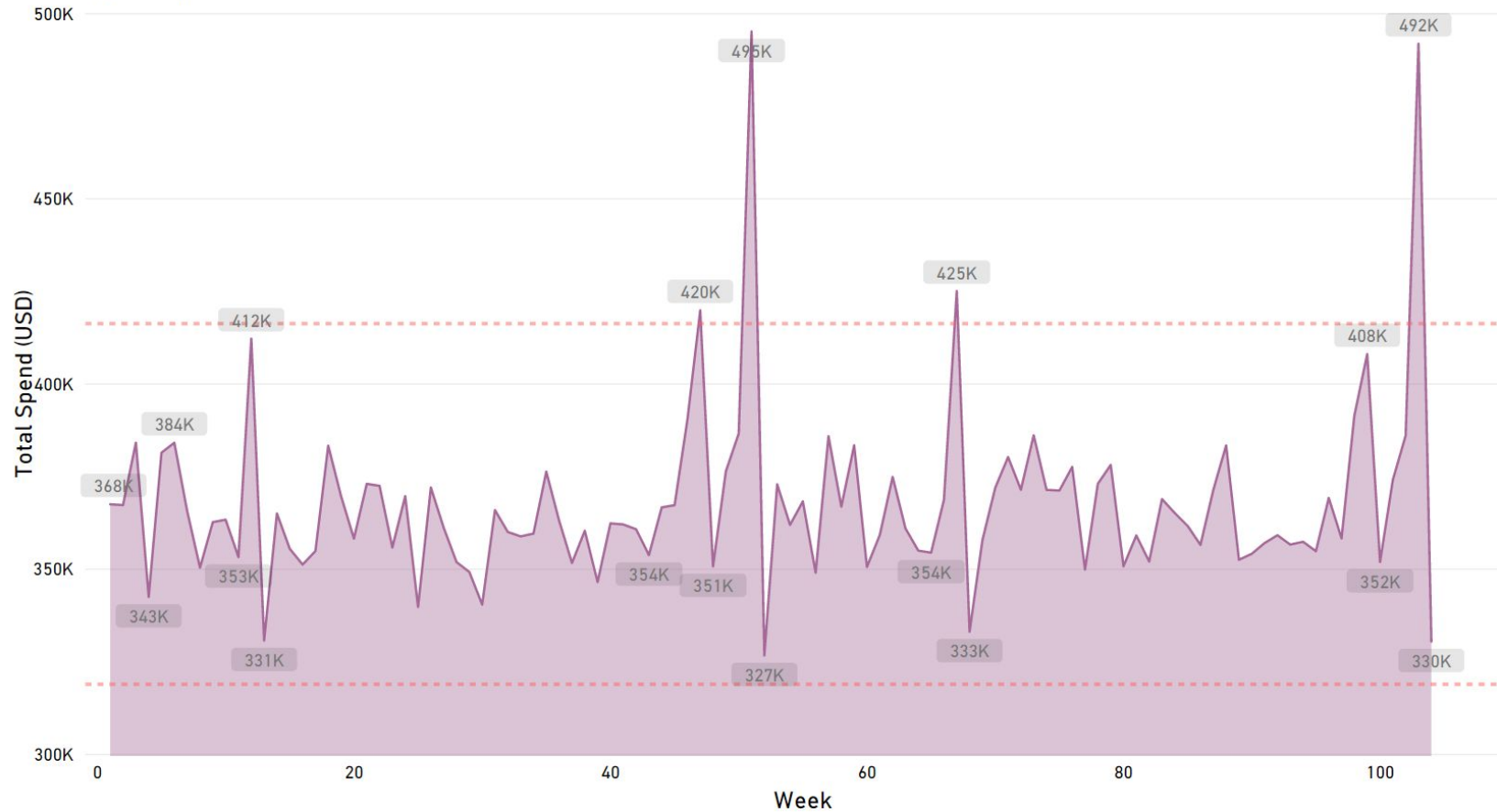
Anomalies and Outliers

Greatest Change by Commodity – Easter Week

Bar chart — Percentage Change vs Commodity:
- HOLIDAY: 322
- TOYS: 122
- OUTDOOR: 73
- SEASONAL PRODU...: 69
- MEAT – OTHER: 65
- FLORAL: 61
- ELECTRONICS: 52
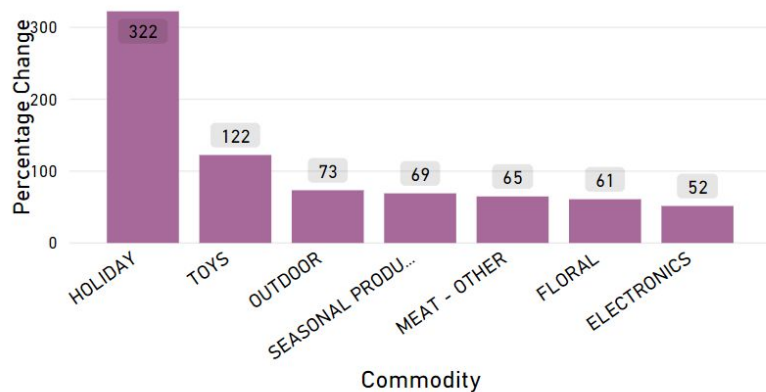
Greatest Change by Commodity – Thanksgiving Week

Bar chart — Percentage Change vs Commodity:
- MEAT – TURKEY: 450
- ELECTRONICS: 145
- HOLIDAY: 121
- TOYS: 82
- CANNED GOODS: 73
- MEAT – OTHER: 60
- DRY GOODS: 47
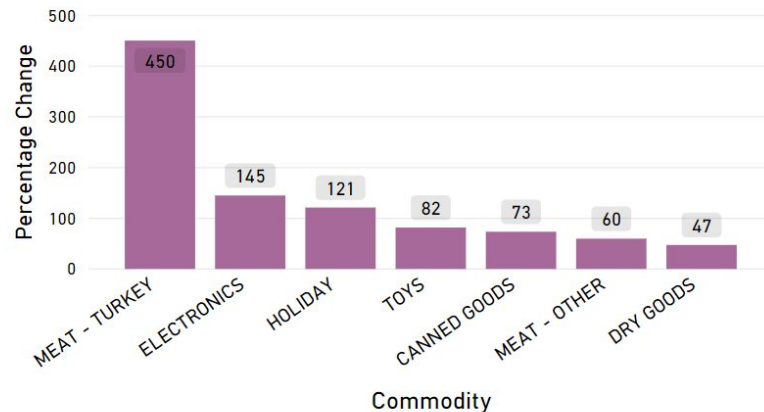
Greatest Change by Commodity – Christmas Week

Bar chart — Percentage Change vs Commodity:
- HOLIDAY: 517
- TOYS: 250
- GIFT: 159
- CLOTHING: 140
- ELECTRONICS: 115
- MEAT – TURKEY: 114
- SEASONAL PRODU...: 102
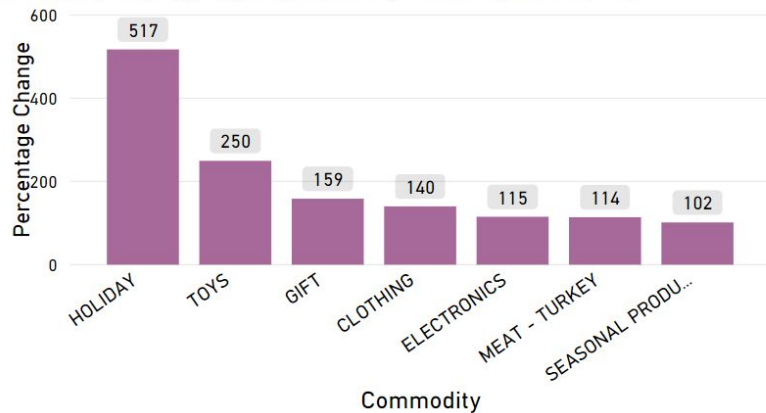
Anomalies and Outliers

Average Spend per Basket by Week
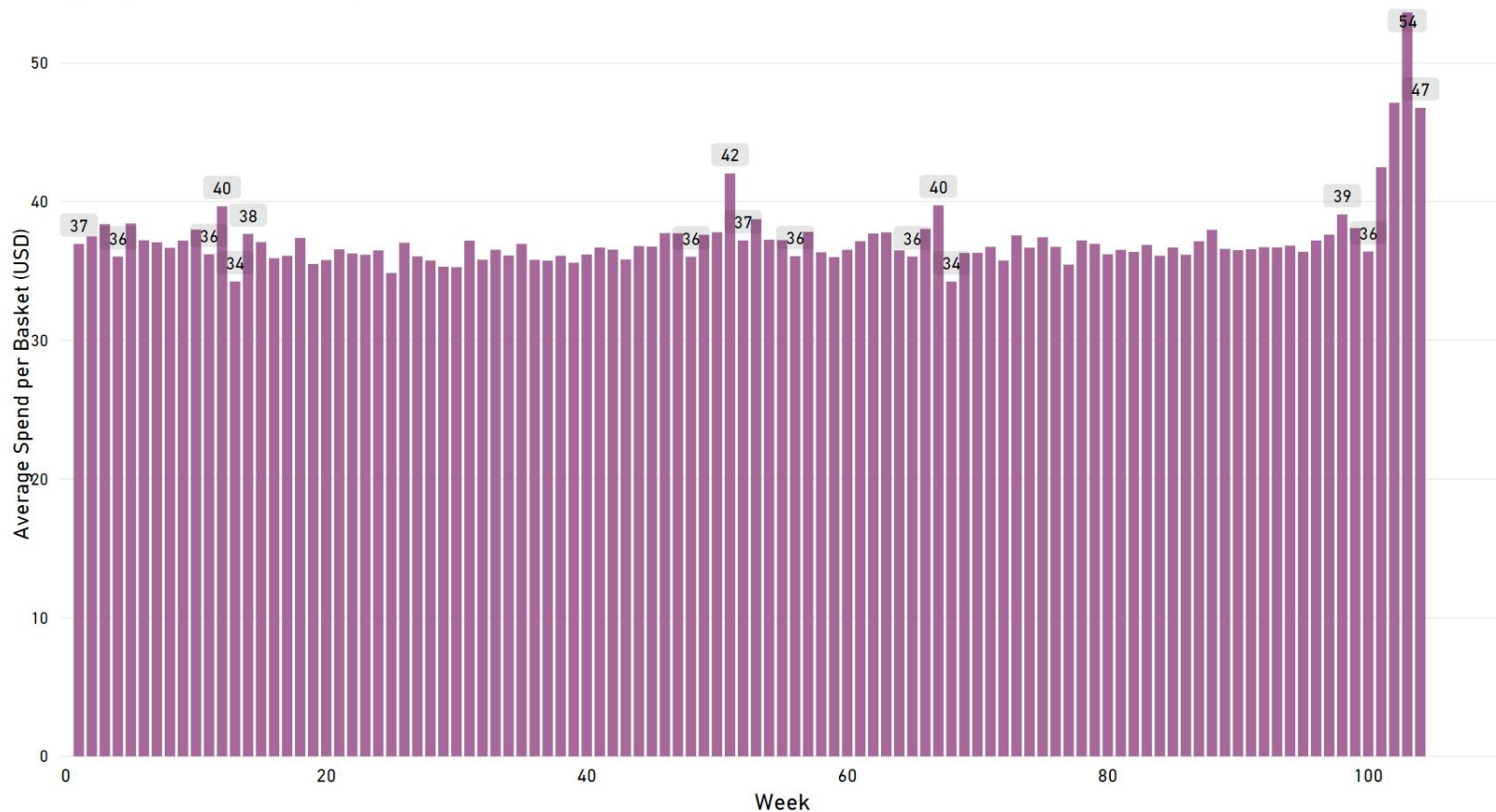
Customers to Target Summary

- Region
  - Central
  - South

- Income Range
  - Less than 35K
  - More than 150K

- Age
  - 19 - 24
  - 55+

Limitations in Data

- Only 2 years of data
- No children and missing information not differentiated
- No data available on discounted items
- No information by Store, only by region

Total Spend per Quarter by # of Children

# of Children ● (Blank) ● 1 ● 2 ● 3