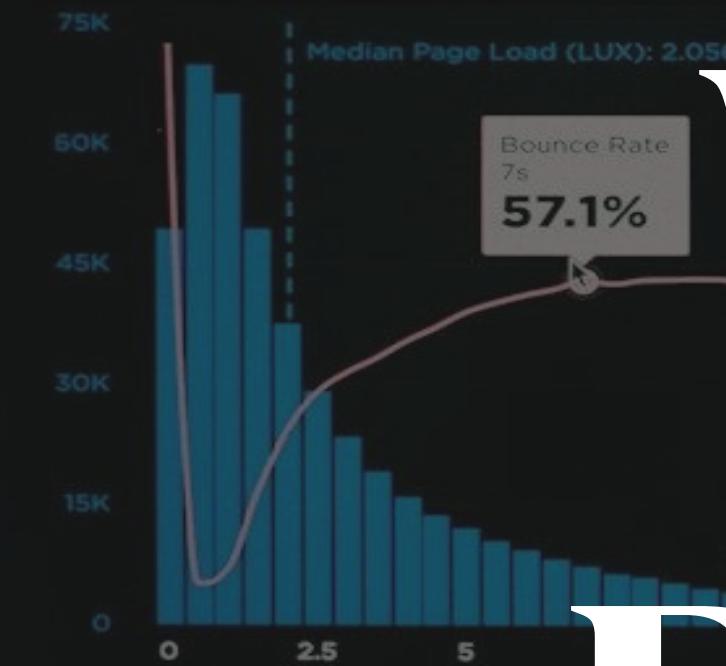


WEEKLY REPORT 2

Kelompok 3 Intern Data Science BCC FILKOM UB 2024

USERS: LAST 7 DAYS USING MEDIAN ▾

LOAD TIME VS BOUNCE RATE



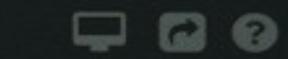
OPTIONS

100 %

START RENDER VS BOUNCE RATE

40K

Median Start Render (LUX): 1.031s



OPTIONS

100 %



PAGE VIEWS VS ONLOAD

Page Load (LUX)

0.7s

Page Views (LUX)

2.7Mpv

OPTIONS

Sessions (LUX)

40.6%

OPTIONS

Sessions (LUX)

479K

OPTIONS

PVs Per Session (LUX)

17min

2pvs

100K 40 min

80K 32 min

60K 24 min

40K

1s

0.8s

0.6s

0.4s

400K 80%

300K 60%

200K 40%

3.2 pvs

2.4 pvs

1.6 pvs

4

Our Team



**Johanes Paulus
Bernard Purek**
Sistem Informasi '22



**Pieter Christy
Yan Yudhistira**
Teknik Informatika '23

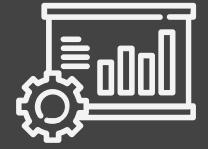
Contents



Theme
and
Dataset



Business Insight
and Analysis



EDA and
Analysis

Theme!

Consumer Segmentation
Based on Behavior and Shopping
Habits

Theme

Dataset

Dataset

3900 Rows
18 Columns

Acknowledgements
Sir Sourav Banerjee
Associate Data Scientist
at CogniTensor Kolkata,
West Bengal, India.

https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and-shopping-habits-dataset?select=shopping_behavior_updated.csv

shopping_behavior_updated.csv (416.61 kB)																		
Detail		Compact		Column														
Customer ID	# Age	Gender	Item Purch...	Category	Purchase ...	Location	Size	Color	Season	Review	Rating							
1	55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1								
2	19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1								
3	50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1								
4	21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5								
5	45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7								
6	46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	2.9								
7	63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2								
8	27	Male	Shorts	Clothing	34	Louisiana	L	Charcoal	Winter	3.2								
9	26	Male	Coat	Outerwear	97	West Virginia	L	Silver	Summer	2.6								
10	57	Male	Handbag	Accessories	31	Missouri	M	Pink	Spring	4.8								
11	53	Male	Shoes	Footwear	34	Arkansas	L	Purple	Fall	4.1								
12	30	Male	Shorts	Clothing	68	Hawaii	S	Olive	Winter	4.9								
13	61	Male	Coat	Outerwear	72	Delaware	M	Gold	Winter	4.5								
14	65	Male	Dress	Clothing	51	New Hampshire	M	Violet	Spring	4.7								
15	64	Male	Coat	Outerwear	53	New York	L	Teal	Winter	4.7								
16	64	Male	Skirt	Clothing	81	Rhode Island	M	Teal	Winter	2.8								
17	25	Male	Sunglasses	Accessories	36	Alabama	S	Gray	Spring	4.1								
18	53	Male	Dress	Clothing	38	Mississippi	XL	Lavender	Winter	4.7								

Theme

Dataset

https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and-shopping-habits-dataset?select=shopping_behavior_updated.csv

Dataset Column

- Customer ID
- **Age**
- Gender
- **Item Purchased**
- Category
- **Purchase Amount (USD)**
- **Location (USA)**
- Size
- Color
- Season
- **Review Rating**
- Subscription Status
- Shipping Type
- Discount Applied
- Promo Code Used
- **Previous Purchases**
- Payment Method
- **Frequency of Purchases**

Bagaimana cara perusahaan mampu memaksimalkan sales, profit, dan quantity melihat track record pembelian dari customernya?

Bagaimana perusahaan dapat mengidentifikasi faktor krusial yang membuat customer mau berbelanja kembali lagi?

Kapan customer tertarik untuk membeli produk perusahaan pada suatu periode tertentu?

WHAT MAKE CONSUMER BEHAVIOR AND THEIR SHOPPING HABITS HARD TO PREDICT?

S.M.A.R.T ANALYSIS

Specific

Measurable

Action
Oriented

Relevant

Time-Bound

Question

Apa tujuan yang ingin dicapai dari penggerjaan Project ini?

Bagaimana cara menentukan bahwa Project ini dapat dicapai sesuai dengan rencana yang telah ditetapkan?

Apakah tujuan penggerjaan Project dapat dicapai secara maksimal? Apa yang perlu dilakukan agar mencapai hal tersebut?

Apakah penggerjaan Project ini selaras dengan kehidupan atau karir yang akan dijalani ke depan? Seberapa relevan penggerjaan Project ini?

Kapan Batas Waktu Maksimal untuk Mencapai Tujuan Project ini?

Answer

Tujuan yang ingin dicapai dari project ini adalah untuk memprediksi segmentasi customer dalam kebiasaan berbelanja dengan melakukan beberapa teknik untuk melatih model Machine Learning sehingga mampu memprediksi secara akurat.

Project ini dapat dicapai apabila telah melakukan pemodelan untuk setiap segmentasi customer sehingga akan didapatkan hasil metrik yang akan disesuaikan dengan teknik ML.

Dapat dicapai secara maksimal dengan tahapan awal untuk memahami isi dari dataset, mengeksplorasi setiap fitur, melakukan tahap preprocessing, dan membuat model klasifikasi yang baik.

Pengerjaan project ini selaras dengan kemampuan hard skill dalam bidang Data Science serta dapat membantu dalam memahami aspek bisnis mengenai perilaku seorang customer dalam kebiasaan belanja-nya.

Project ini bertujuan untuk dicapai secara maksimal pada tanggal 23 Maret 2024.

Background

Overview

Value Count

Regional

Correlation

Customer Habits

Dataset Overview

Customer ID	0
Age	0
Gender	0
Item Purchased	0
Category	0
Purchase Amount (USD)	0
Location	0
Size	0
Color	0
Season	0
Review Rating	0
Subscription Status	0
Shipping Type	0
Discount Applied	0
Promo Code Used	0
Previous Purchases	0
Payment Method	0
Frequency of Purchases	0
dtype:	int64

**Any
Missing
Value?**

Any Duplicated Value?

```
[19] data.duplicated().sum()
```

```
0
```

Even When Removed the Customer ID?

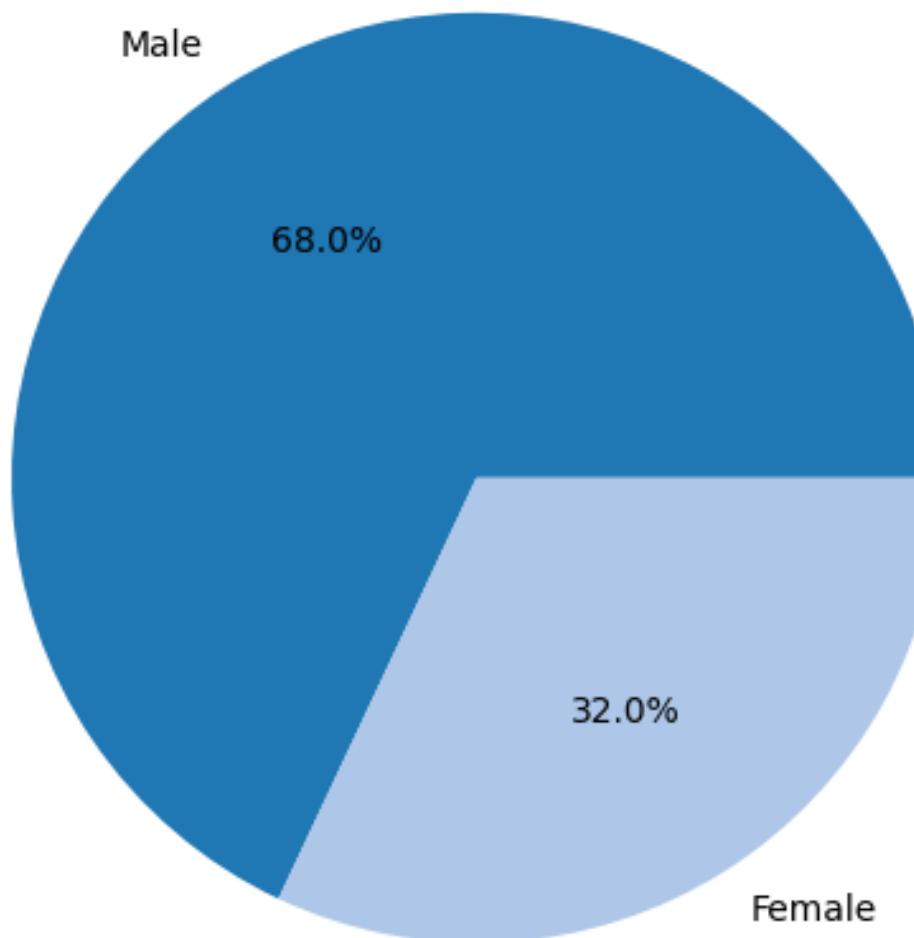
```
[20] duplicate_without_id = (data.drop(columns=["Customer ID"]))  
duplicate_without_id.duplicated().sum()
```

```
0
```

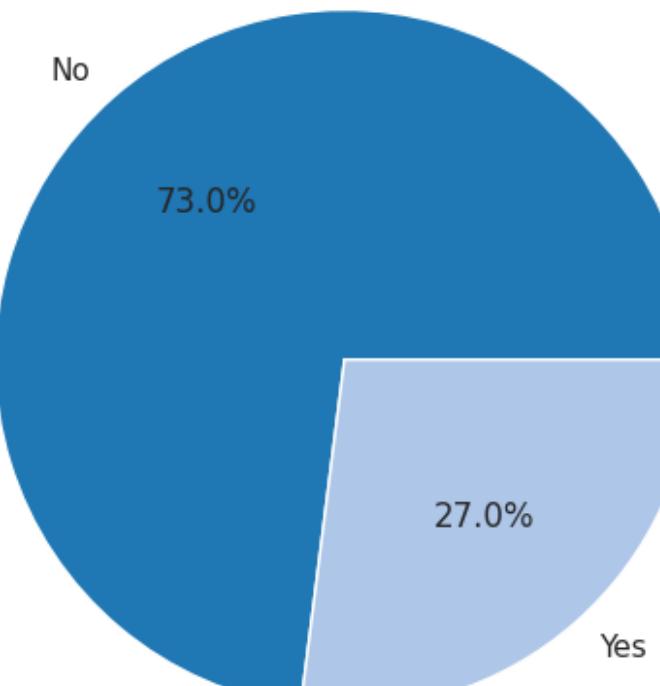
Value Counts

Kolom Yes/No

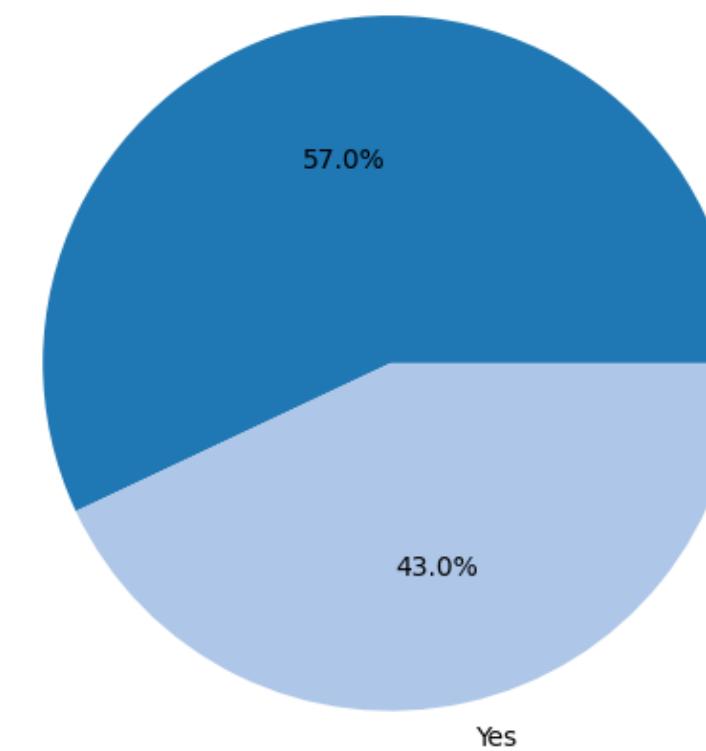
Komposisi Kolom Gender Dengan 2 Unique Values



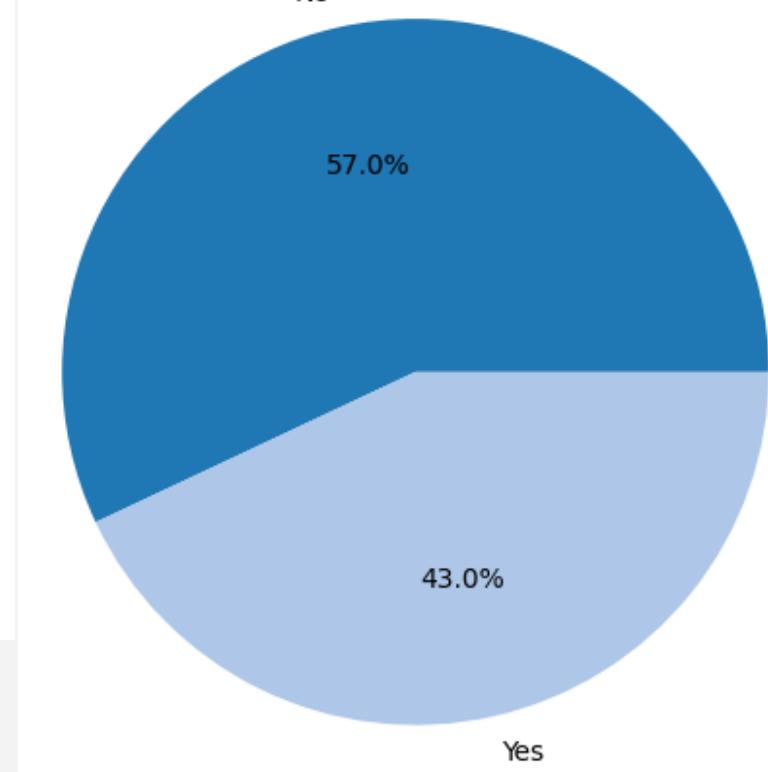
Komposisi Kolom Subscription Status Dengan 2 Unique Values



Komposisi Kolom Promo Code Used Dengan 2 Unique Values



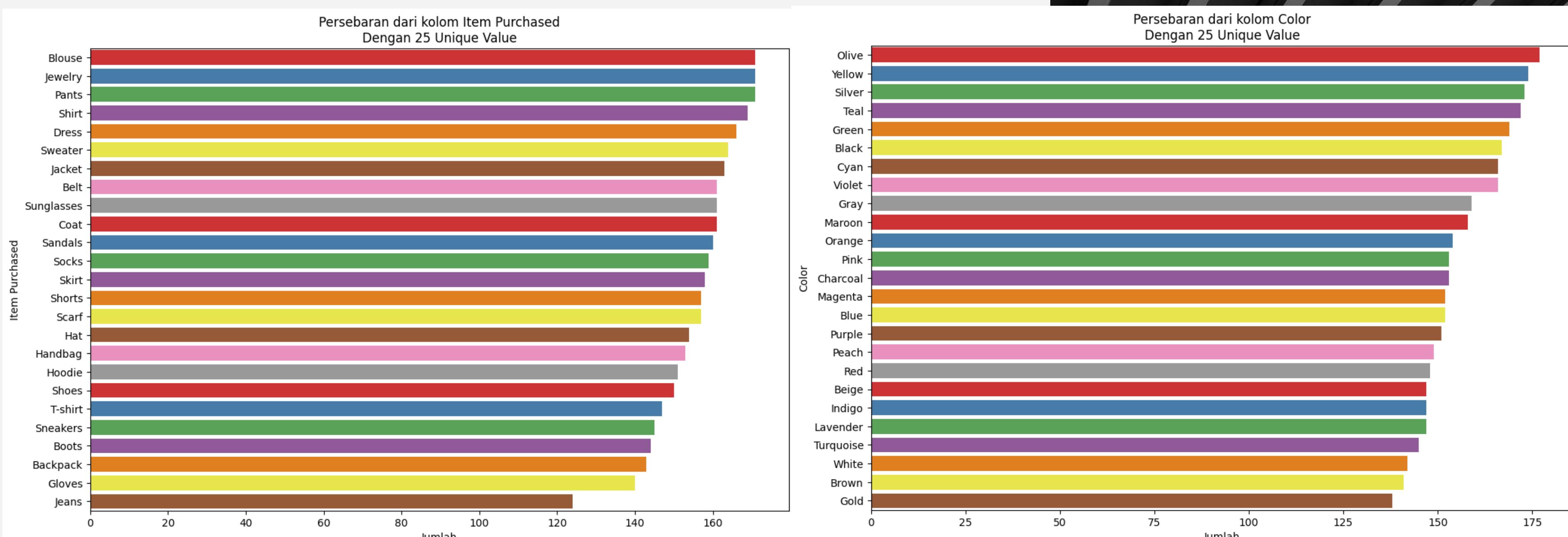
Komposisi Kolom Discount Applied Dengan 2 Unique Values



Kolom Promo Code Used dan Discount Applied Memiliki Persentase yang sama

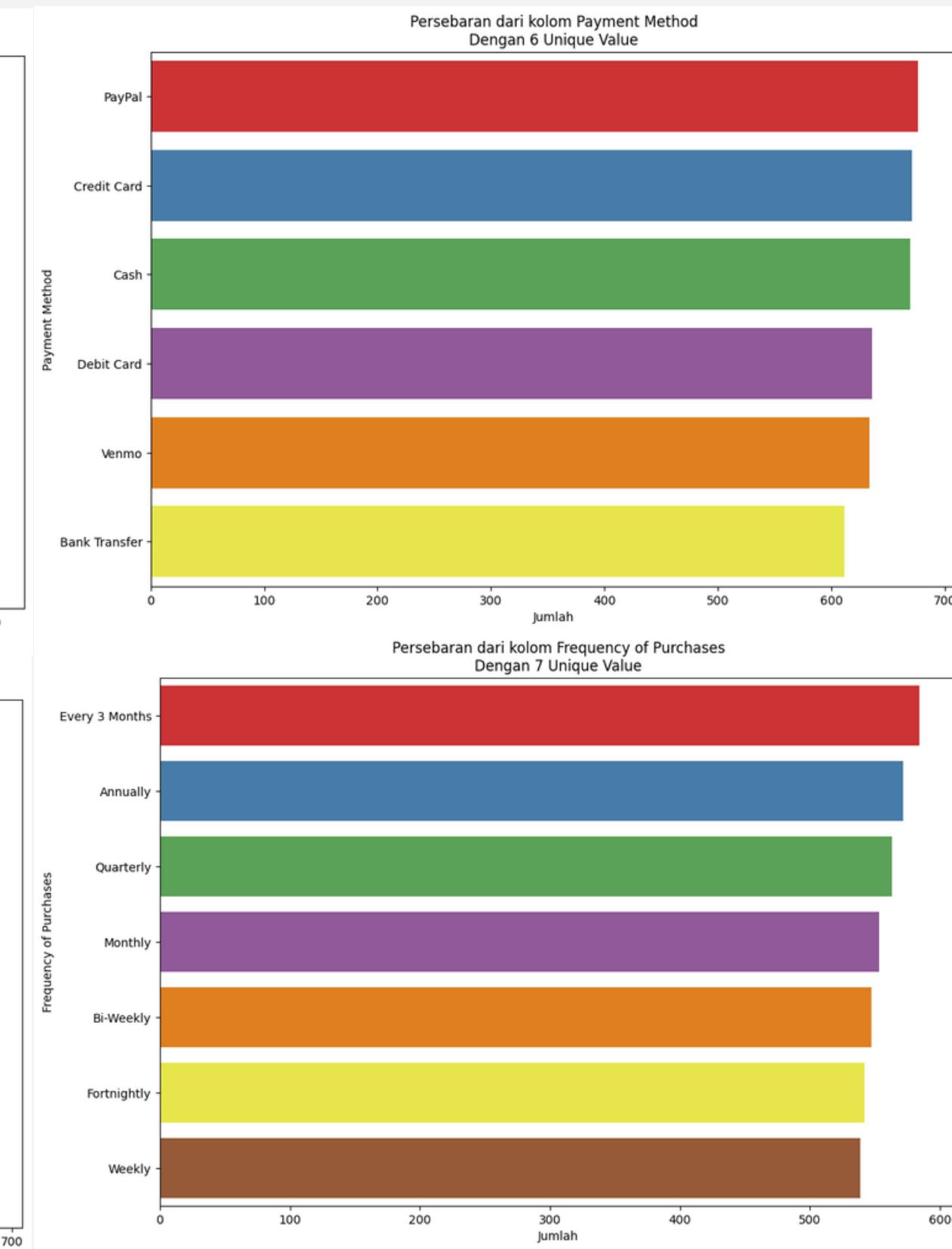
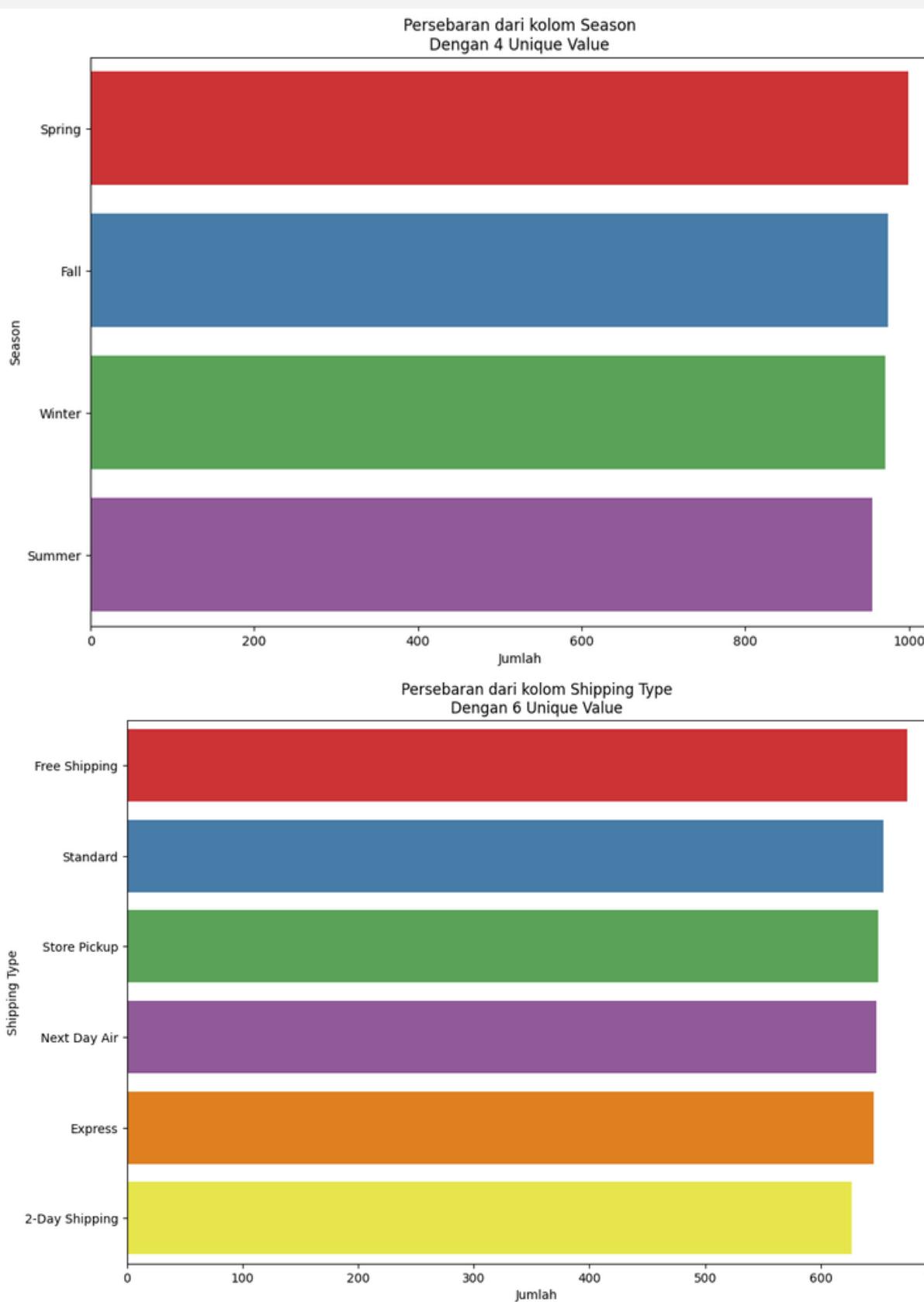
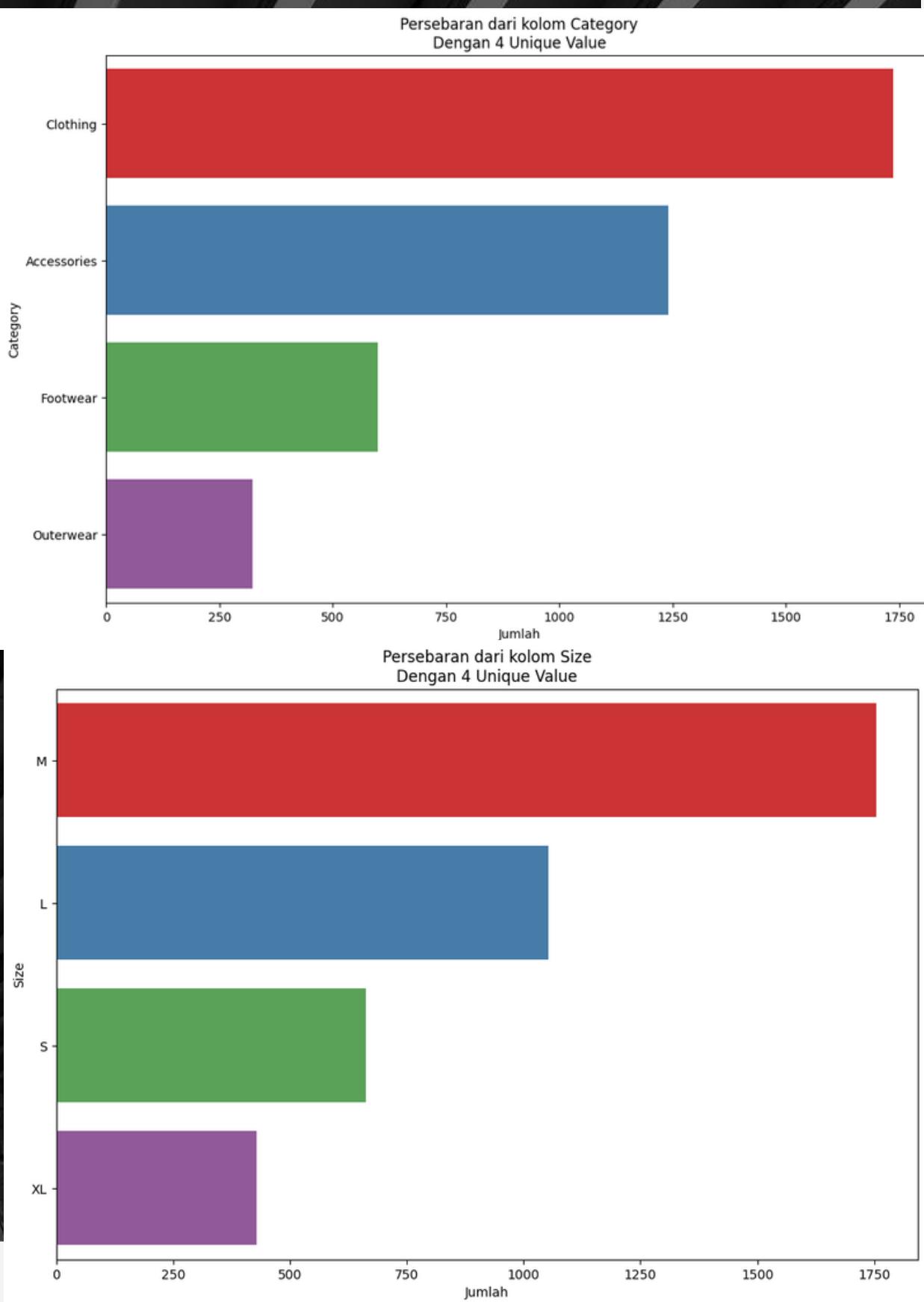
Apa Hubungannya?

Value Counts with Lot Unique Values

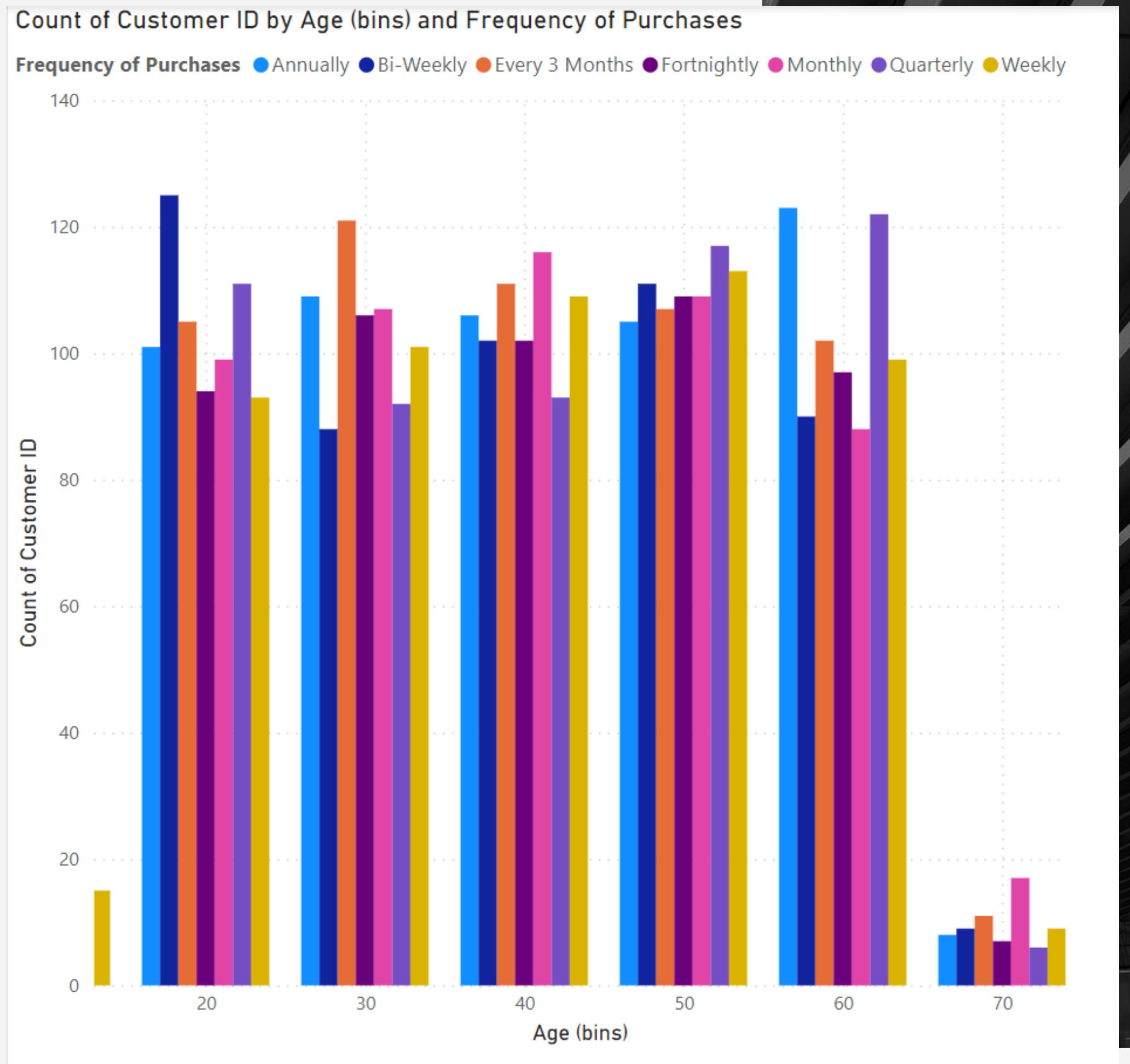


Value Counts

Kolom Preference



Kelompok umur mana yang sering membeli produk berdasarkan frekuensi pembeliannya?



Background

Overview

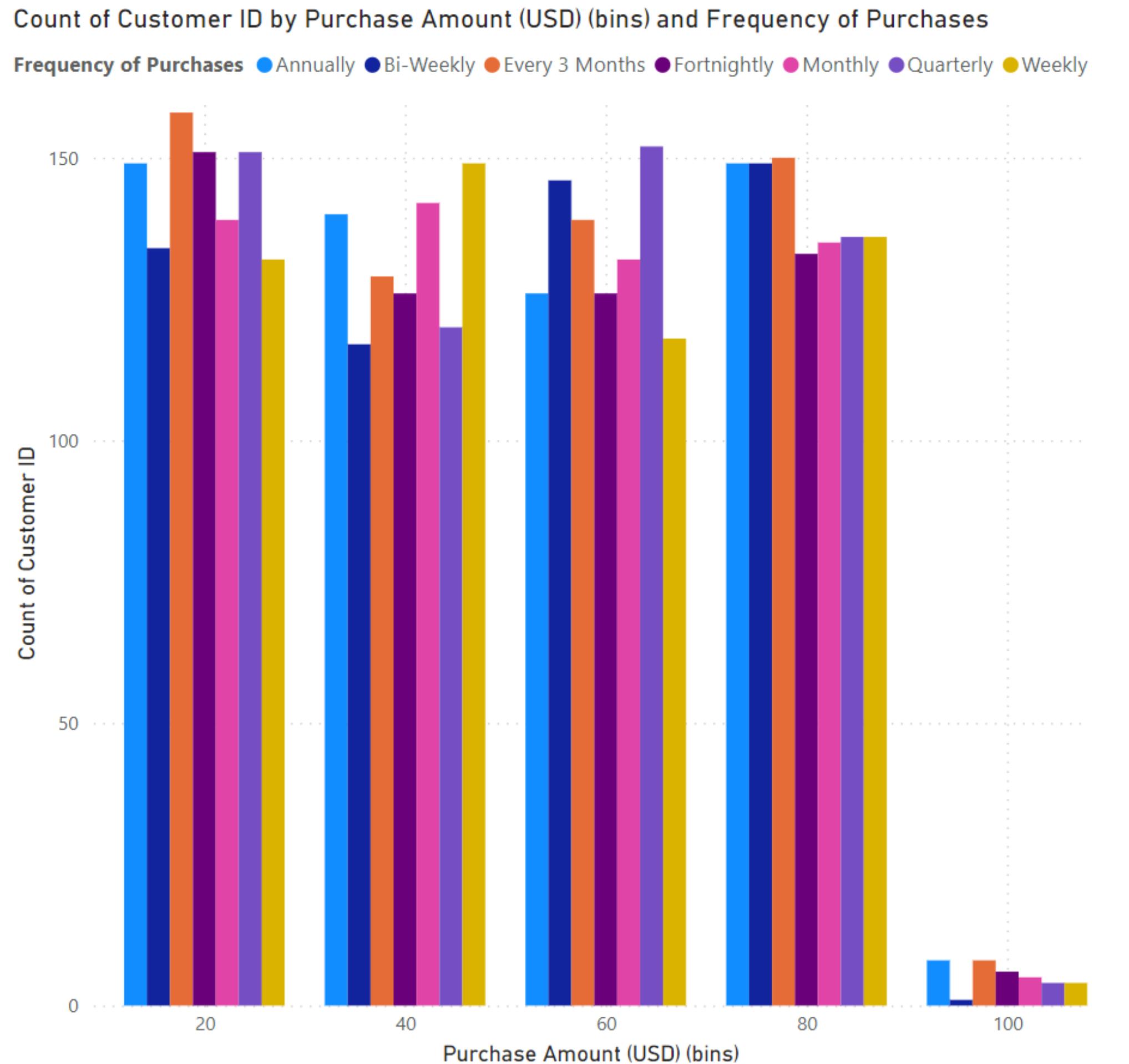
Value Count

Regional

Correlation

Customer Habits

Bagaimana average dari besar pengeluaran berdasarkan frekuensi pembeliannya?



Background

Overview

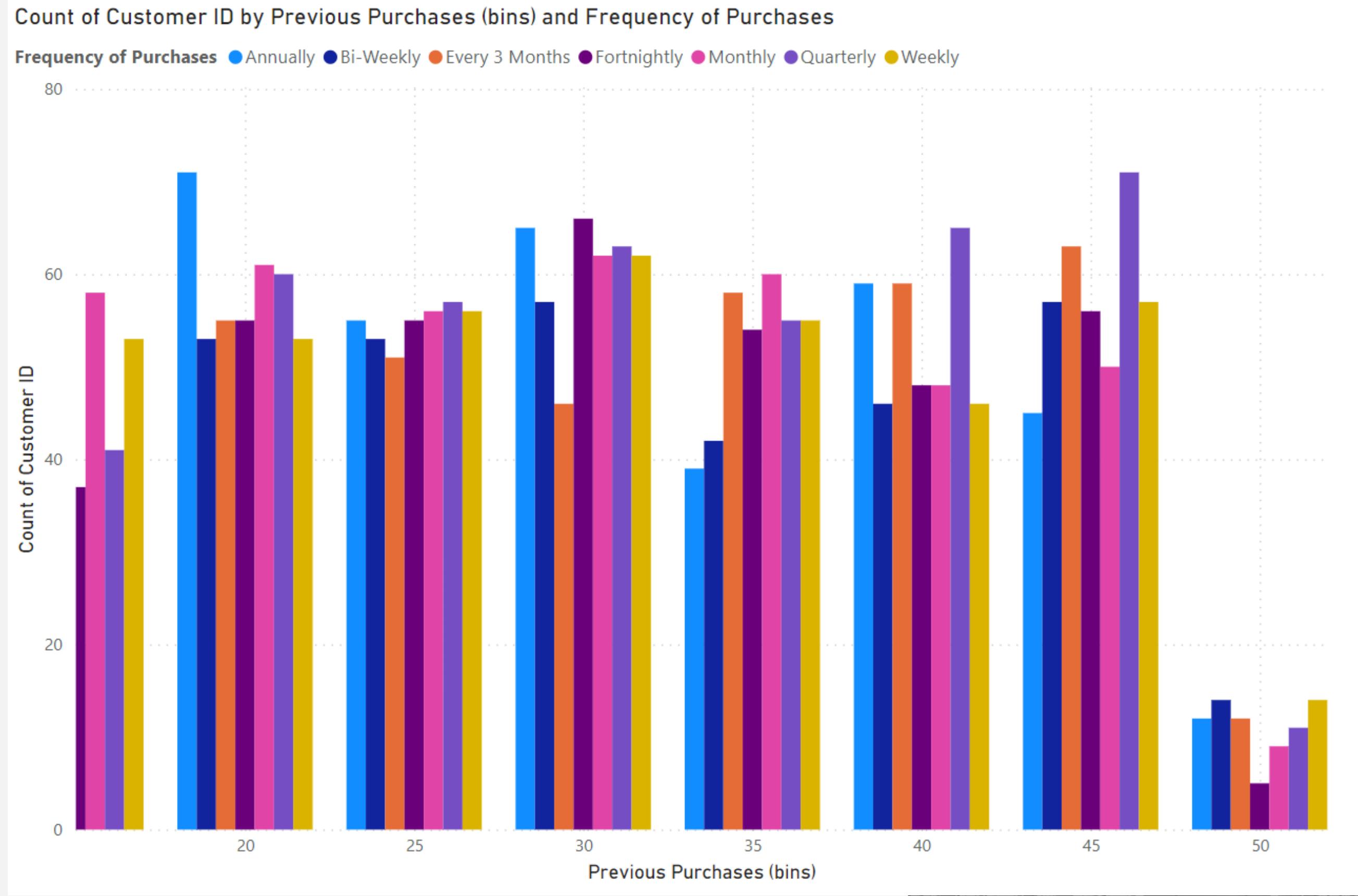
Value Count

Regional

Correlation

Customer Habits

Bagaimana besar Pengeluaran sebelumnya berdasarkan frekuensi pembeliannya?



Background

Overview

Value Count

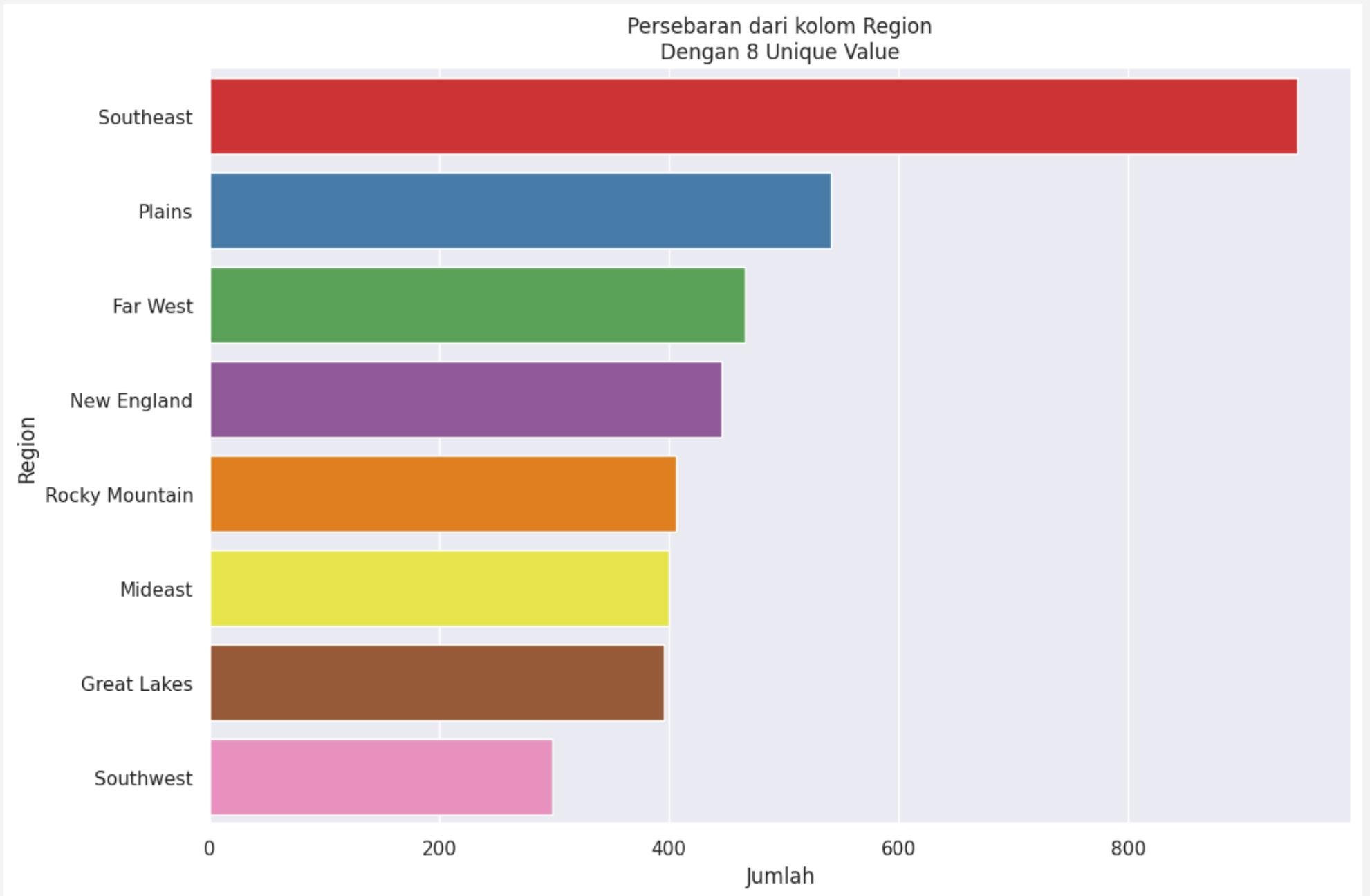
Regional

Correlation

Customer Habits

Regional Value Count

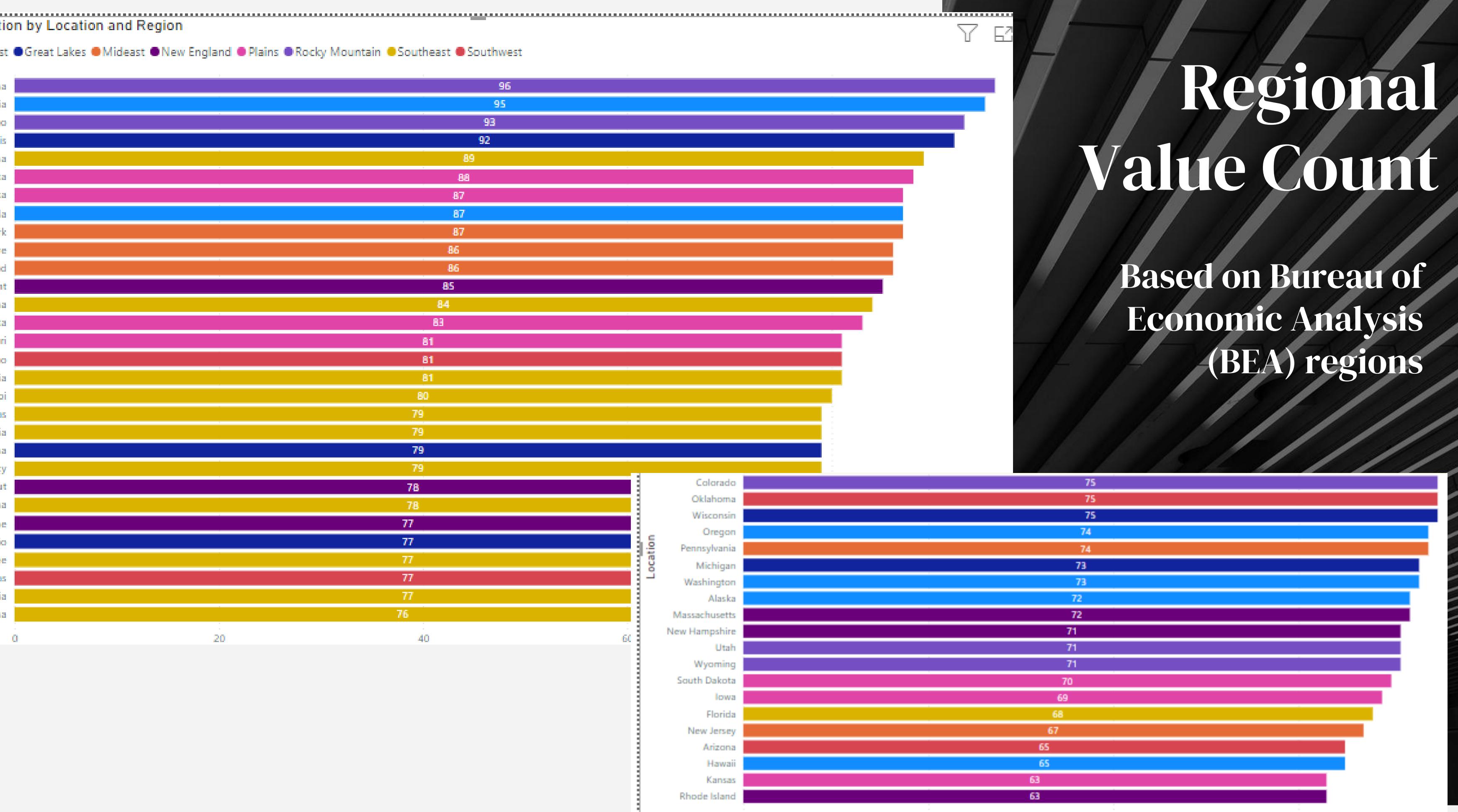
Based on Bureau of Economic Analysis (BEA) regions



```
regions = {
    'New England': ['Connecticut', 'Maine', 'Massachusetts', 'New Hampshire', 'Rhode Island', 'Vermont'],
    'Mideast': ['Delaware', 'District of Columbia', 'Maryland', 'New Jersey', 'New York', 'Pennsylvania'],
    'Great Lakes': ['Illinois', 'Indiana', 'Michigan', 'Ohio', 'Wisconsin'],
    'Plains': ['Iowa', 'Kansas', 'Minnesota', 'Missouri', 'Nebraska', 'North Dakota', 'South Dakota'],
    'Southeast': ['Alabama', 'Arkansas', 'Florida', 'Georgia', 'Kentucky', 'Louisiana', 'Mississippi', 'North Carolina'],
    'Southwest': ['Arizona', 'New Mexico', 'Oklahoma', 'Texas'],
    'Rocky Mountain': ['Colorado', 'Idaho', 'Montana', 'Utah', 'Wyoming'],
    'Far West': ['Alaska', 'California', 'Hawaii', 'Nevada', 'Oregon', 'Washington']
}
```

East Coast vs. West Coast: The impact of the Panama Canal's expansion on the routing of Asian imports into the United States

Preference for ethnic specialty produce by the Hispanics in the east coast of the USA



Regional Value Count

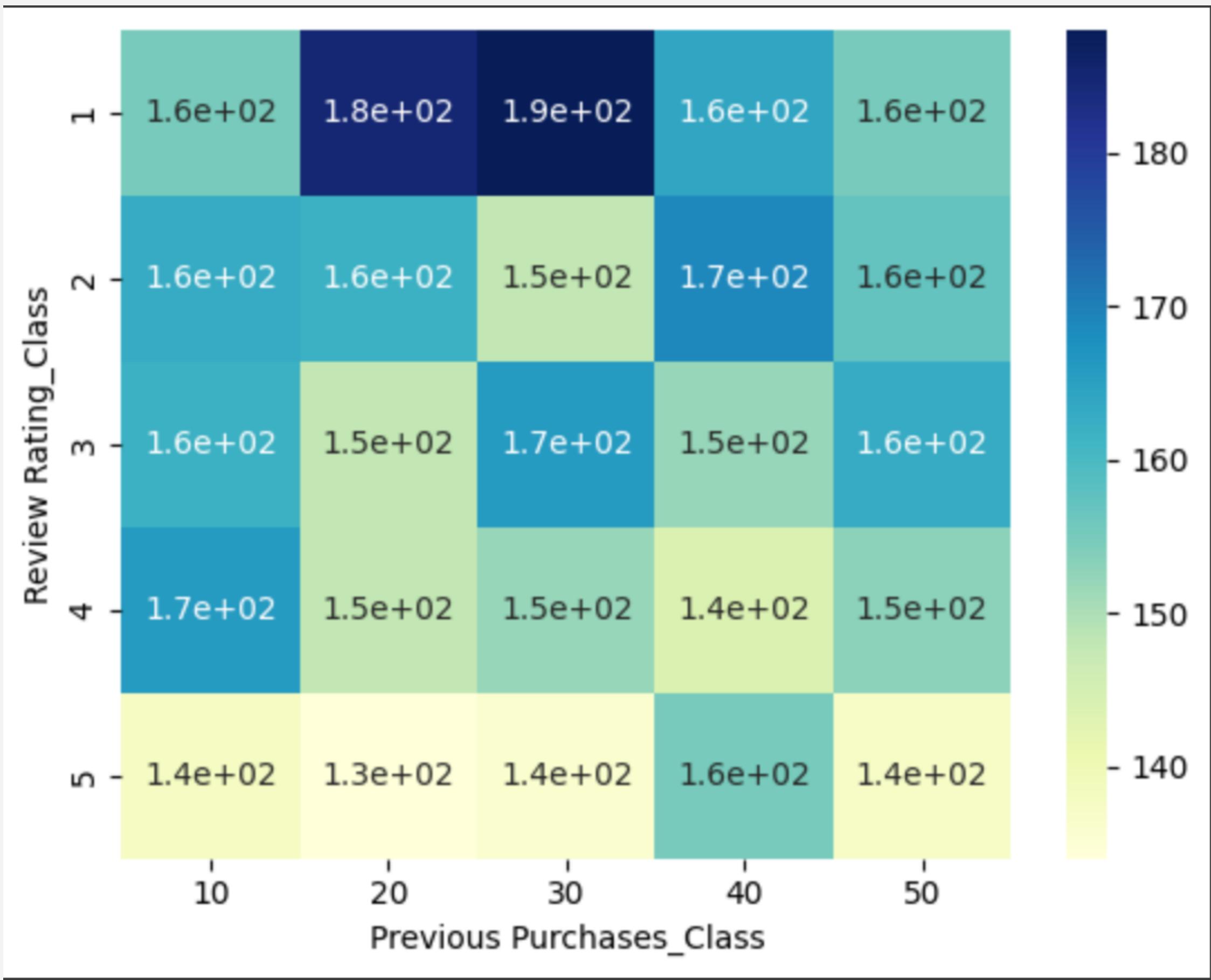
Based on Bureau of Economic Analysis (BEA) regions

Male (1),
Female
(0),
Yes (1),
No (0)

Num Column Correlation



Rating & Previous Purchases



Background

Overview

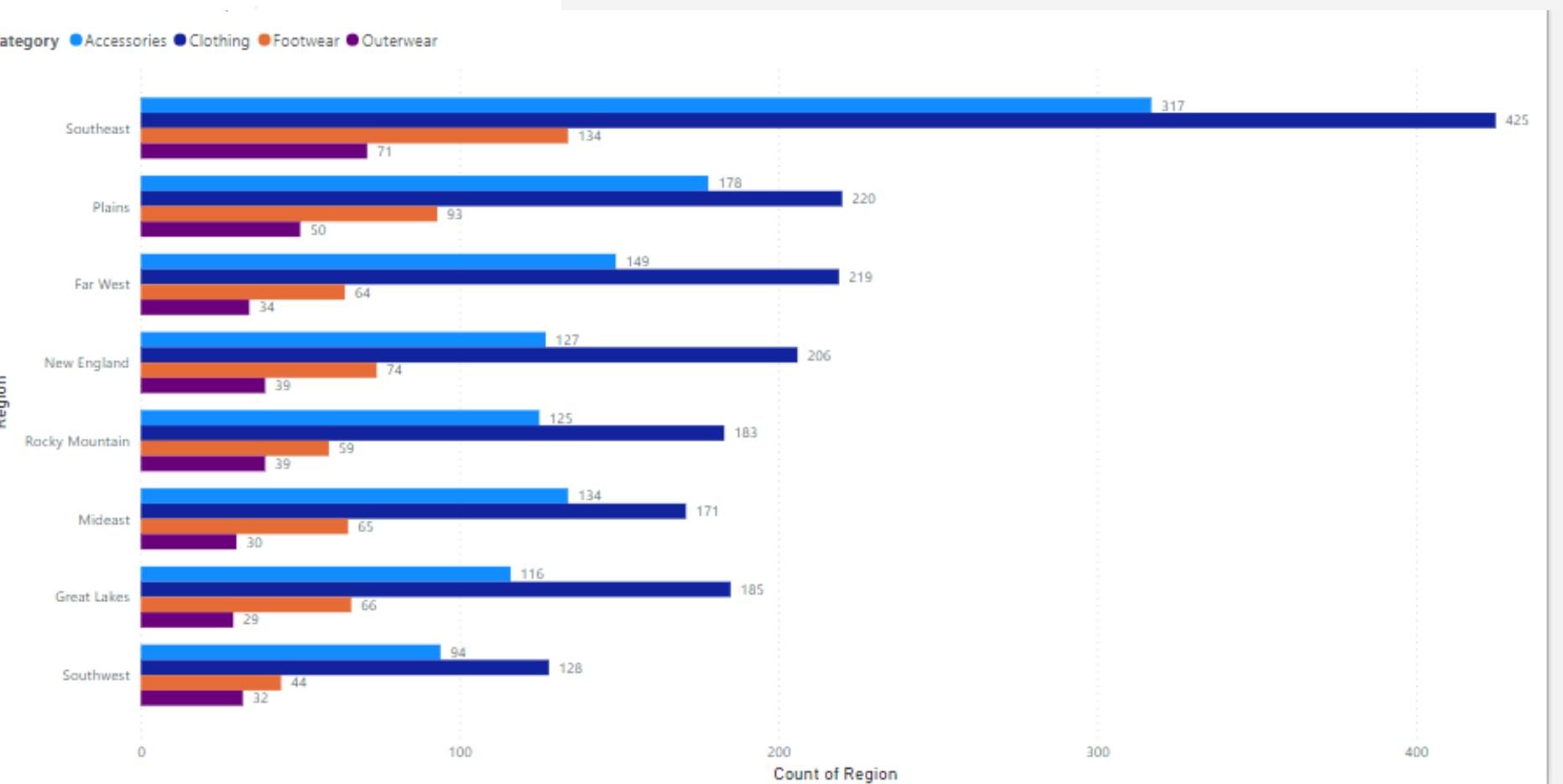
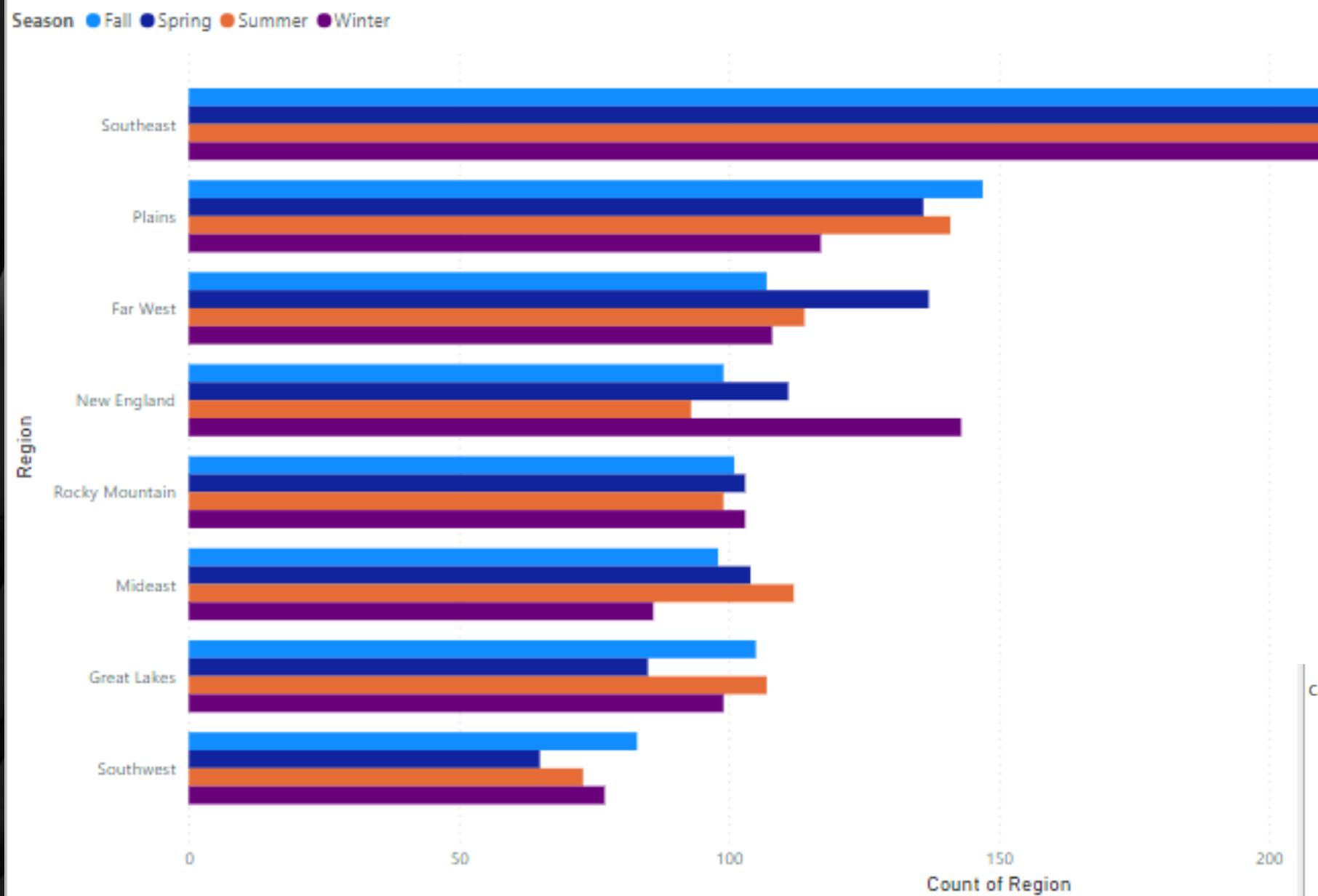
Value Count

Regional

Correlation

Customer Habits

Customer Habits Based Season or Category



Background

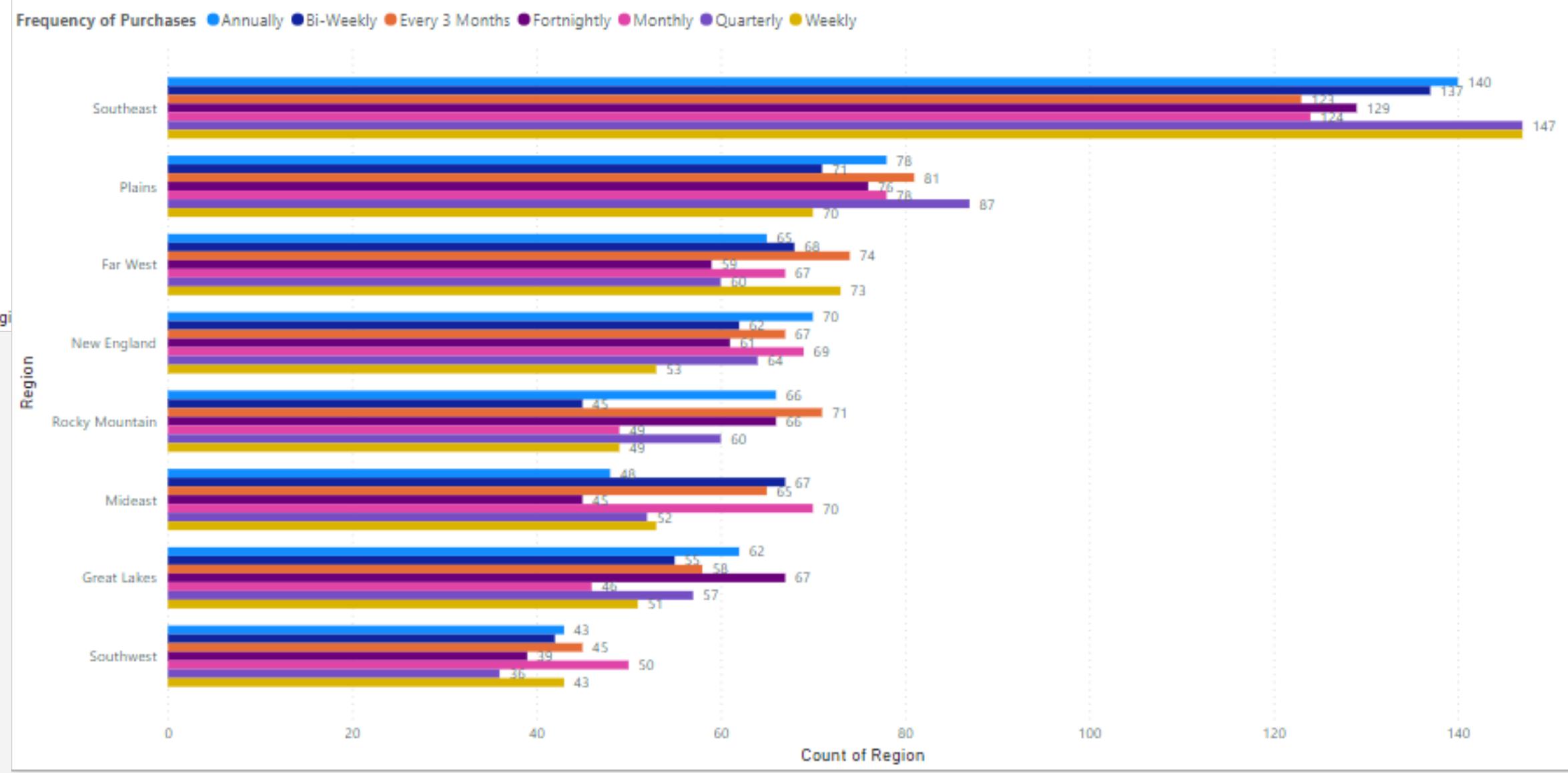
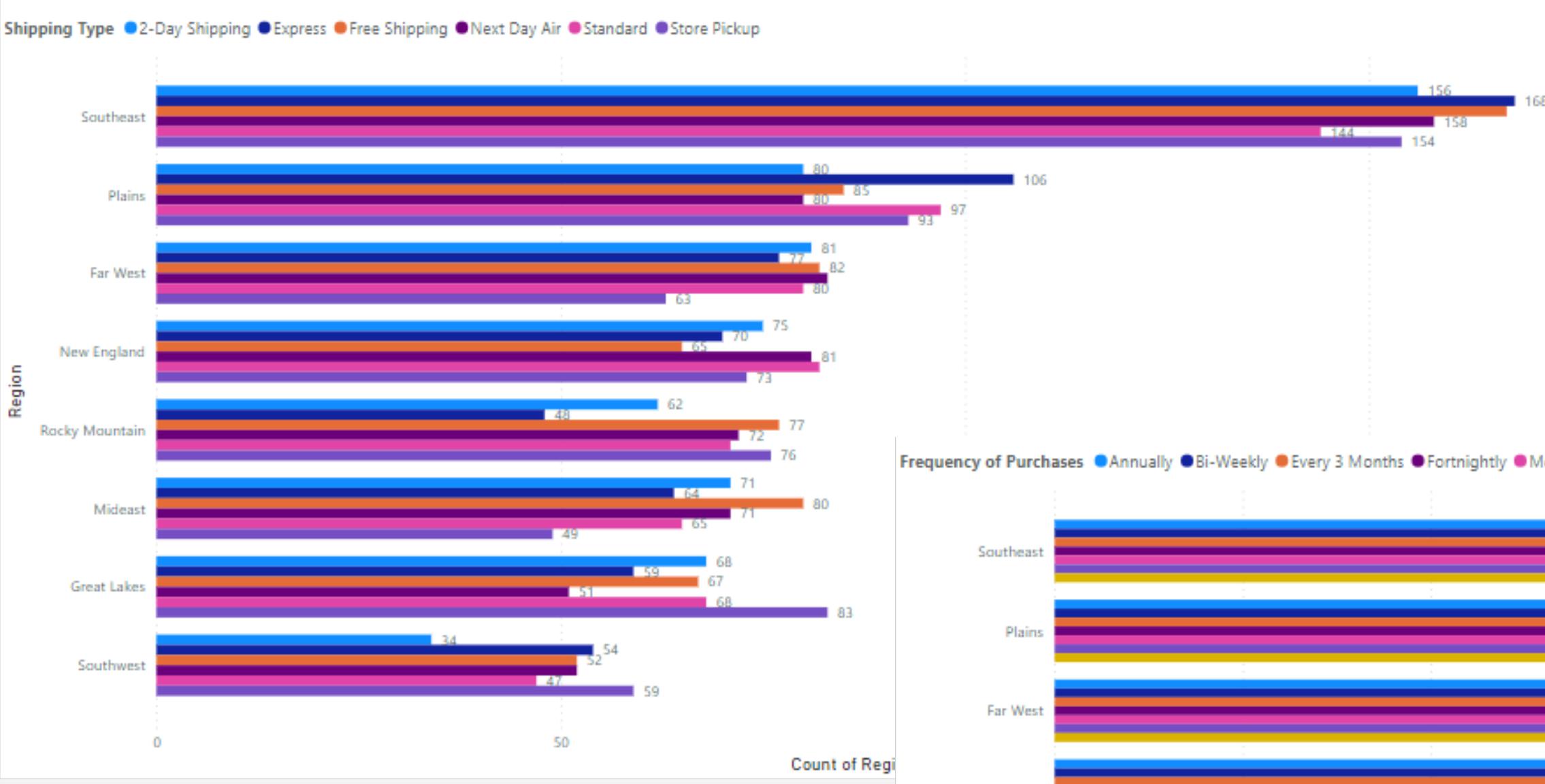
Overview

Value Count

Regional

Correlation

Customer Habits



Shopping Type and Freq

Background

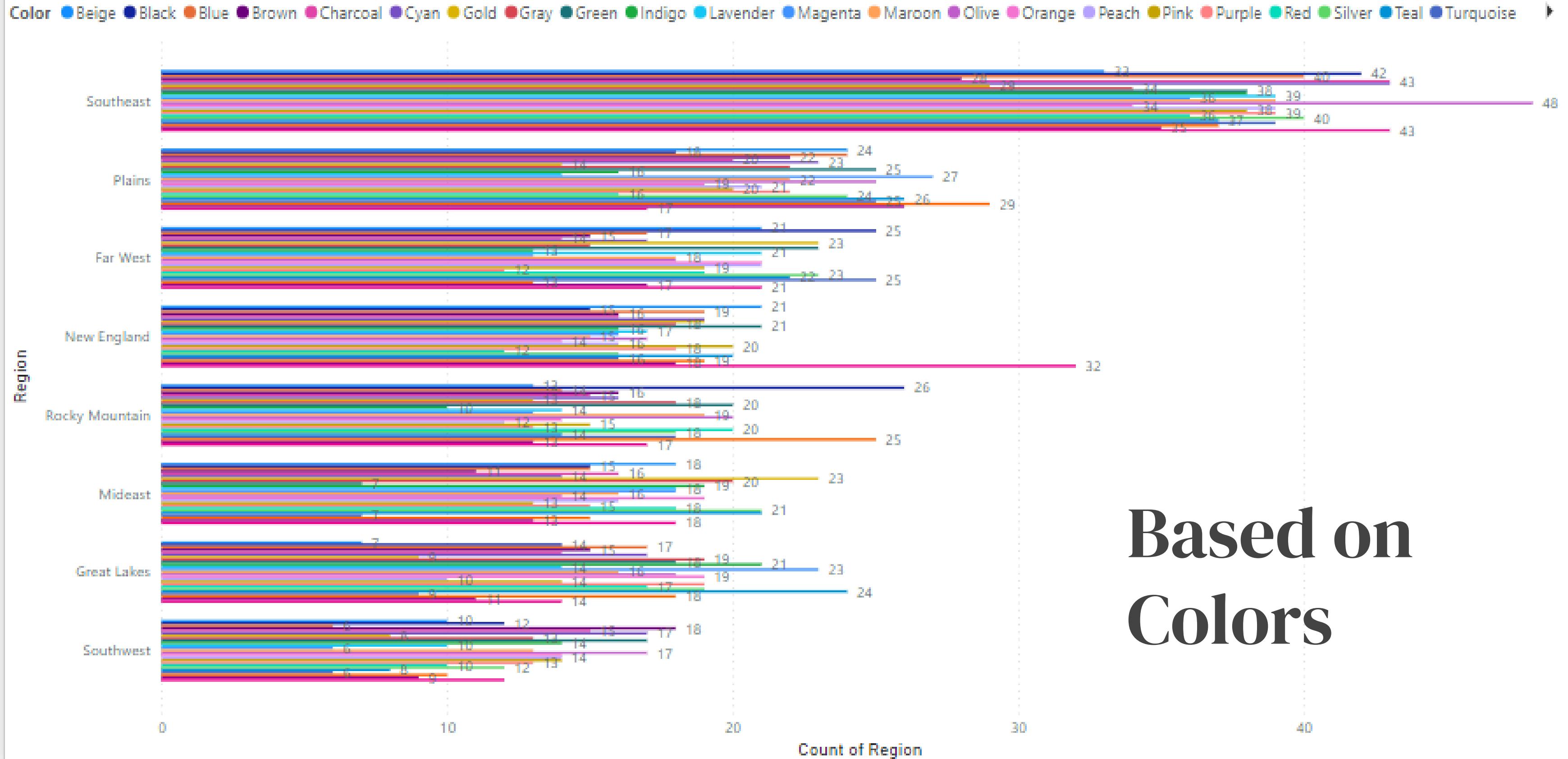
Overview

Value Count

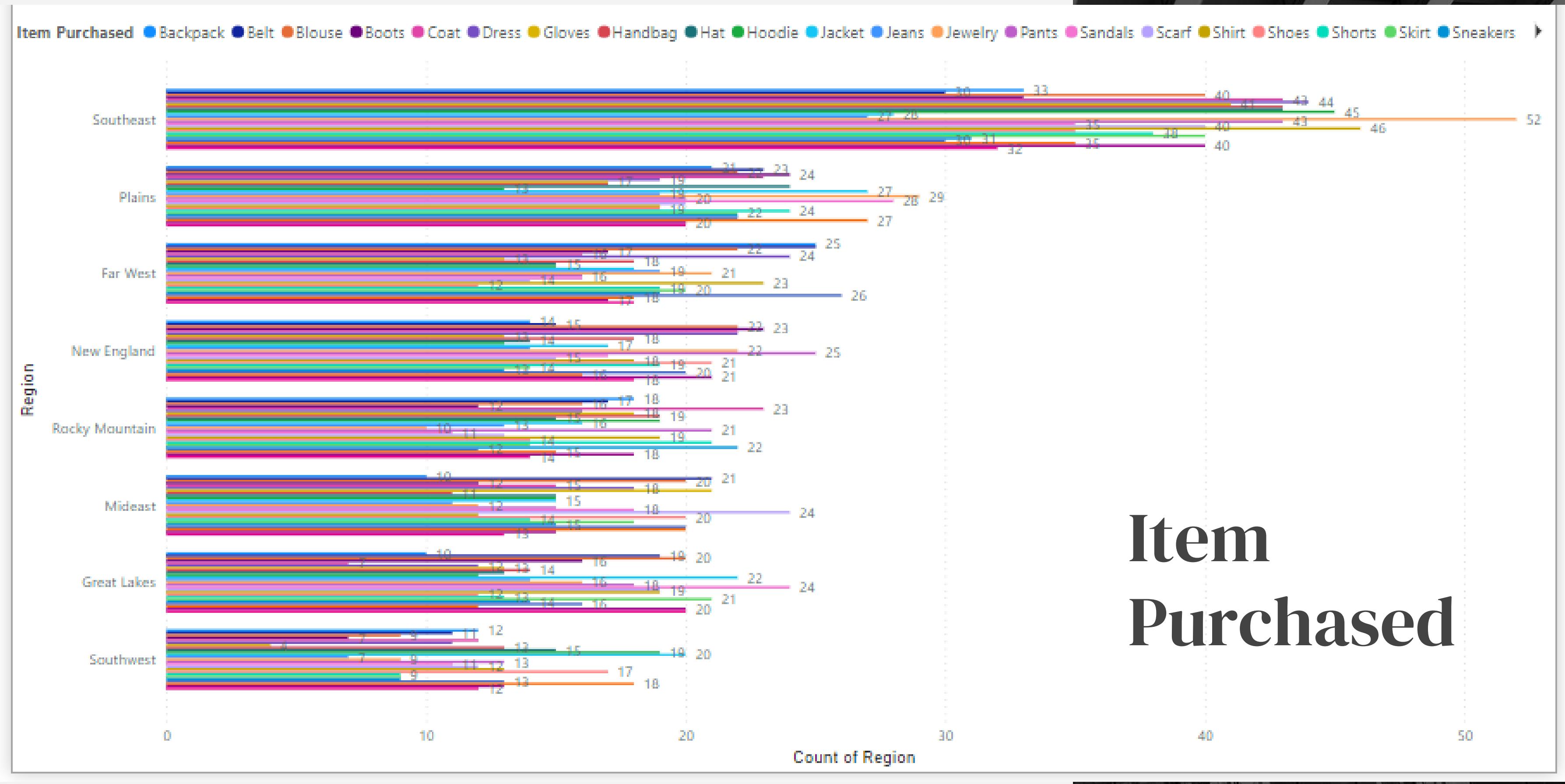
Regional

Correlation

Customer Habits

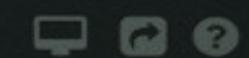


Based on
Colors

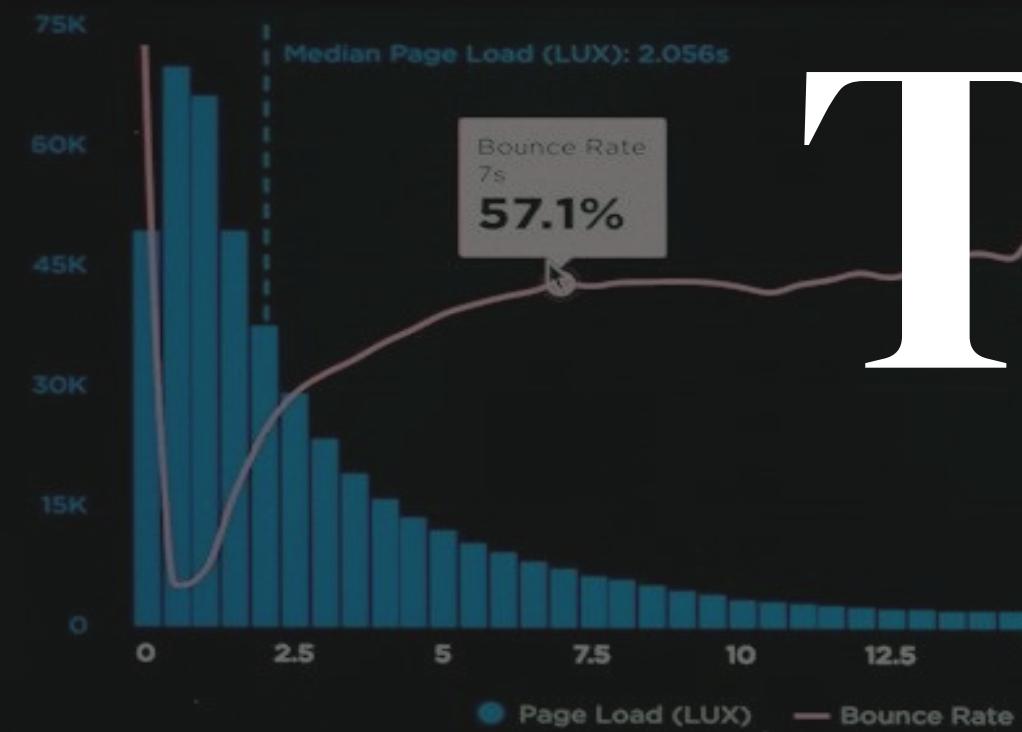


Item Purchased

USERS: LAST 7 DAYS USING MEDIAN ▾



LOAD TIME VS BOUNCE RATE



OPTIONS

START RENDER VS BOUNCE RATE



OPTIONS

THANK YOU

PAGE VIEWS VS ONLOAD

Page Load (LUX)

0.7s

Page Views (LUX)

2.7MpvS

Bounce Rate (LUX)

40.6%

OPTION

SESSIONS

SESSION LENGTH

PVs PER SESSION

OPTIONS

Sessions (LUX)

479K

Session Length (LUX)

17min

PVs Per Session (LUX)

2pvs

OPTIONS

1s

0.8s

0.6s

0.4s

Page Load (LUX)

2.7MpvS

40.6%

Bounce Rate (LUX)

40.6%

Sessions (LUX)

479K

Session Length (LUX)

17min

PVs Per Session (LUX)

2pvs

OPTIONS

SESSION LENGTH

PVs PER SESSION

OPTIONS

SESSION LENGTH

PVs PER SESSION

OPTIONS

SESSION LENGTH

PVs PER SESSION

OPTIONS

SESSION LENGTH

PVs PER SESSION

Feel free if you have any questions