

Recent Customer Shopping Trends

Ava Cascario, Sydney Holt, and Kacie Rohn

2024-12-03

```
qti codeAppend, ref.label=knitr::all_labels(), echo=TRUE, eval=FALSE}
```

Research Topic: Latest Customer Shopping Trends

This research focuses on recent customer shopping trends, a topic that is both familiar and increasingly important to each member of our team. The rise of online shopping, particularly after the Covid-19 pandemic, combined with rapid technological advancements, has made this subject more relevant than ever. Our research will revolve around gaining a better understanding on what customers tend to purchase pertaining to gender, age, geographical location, season, item, price, and other features. Through a series of research questions and data visualizations, we will explore these relationships to uncover insights and draw connections between key attributes. Our goal is to contribute new knowledge to the reader and deepen our understanding of modern consumer behavior.

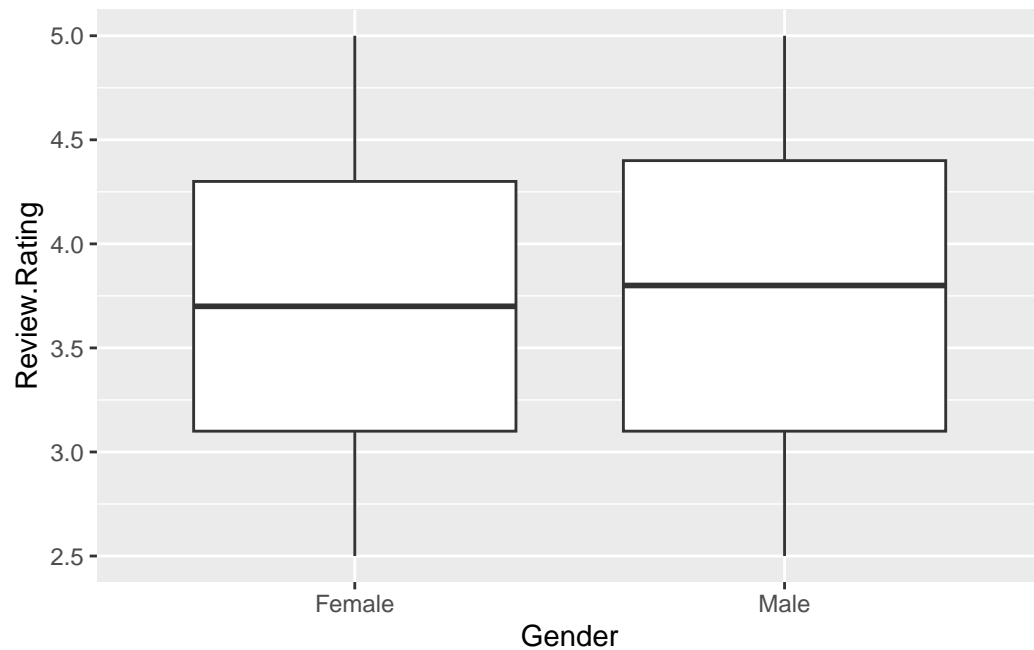
Research Questions

The first research question we will explore is, how do different demographics such as age, gender, location, and price affect the shopping trends of customers. We will create different visuals to present our findings and explain the correlation between each one of these features and customer shopping behavior. We are also curious to know, does gender have an affect on how much the customer spends, what item(s) they buy, and what reviews they left on the product. We predict that there will be large differences between the shopping trends of males versus the shopping trends of females, and intend to explore this further using multiple types of visualizations and tables. We must be aware of bias, as we are all females who experience the female shopping trends ourselves, and we cannot allow this to alter the conclusions we make. We are also curious on if the time of year (season) and geographical location of a customer changes what specific item they purchase. For example, does someone who is experiencing summer in Florida tend to buy something different from a customer experiencing winter in Maine? Overall, this is not an exhaustive list on what we intend to explore, as there are many different combinations of features that allow for different discoveries.

Provenance Of Our Data

We are utilizing a data set that we found on Kaggle. Kaggle is a website focused towards data scientists with a goal in helping others learn about data.

****FAIR and CARE Principles***



Call:

```
lm(formula = Purchase.Amount..USD. ~ Category, data = shopping_trends_raw)
```

Residuals:

Min	1Q	Median	3Q	Max
-40.255	-21.025	-0.025	20.975	42.827

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	59.8387	0.6725	88.979	<2e-16 ***
CategoryClothing	0.1866	0.8804	0.212	0.8321
CategoryFootwear	0.4167	1.1783	0.354	0.7236
CategoryOuterwear	-2.6659	1.4775	-1.804	0.0713 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.68 on 3896 degrees of freedom

Multiple R-squared: 0.001118, Adjusted R-squared: 0.0003489

F-statistic: 1.454 on 3 and 3896 DF, p-value: 0.2252

Research Topic: Field of Study vs Occupation

Our focus in conducting our study is to look at the association between field of study and actual job occupation. That is, to know if what an individual studied in college would have an affect on the job they get post-graduation. We will conduct our research by looking at different research questions and creating visualizations to represent the data that correlates with each question. We will then explore our outcomes and summarize our findings and how they connect back to our topic.

Research Questions

The first question we will explore is, how does your field of study correlate with your job occupation? We want to know what kind of association is found between the two factors, and if what an individual studied has an affect on the job they pursue. We would also like to explore, what the most commonly chosen occupations are based on the field of study? For example, if an individual was a software engineering major, what is the most commonly chosen job after graduation. Finally, we would like to explore the question, is there an association between gender and field of occupation? Not only will our research focus on field of study and occupation, but we would also like to analyze whether gender plays a role in this as well.

Provenance Of Our Data

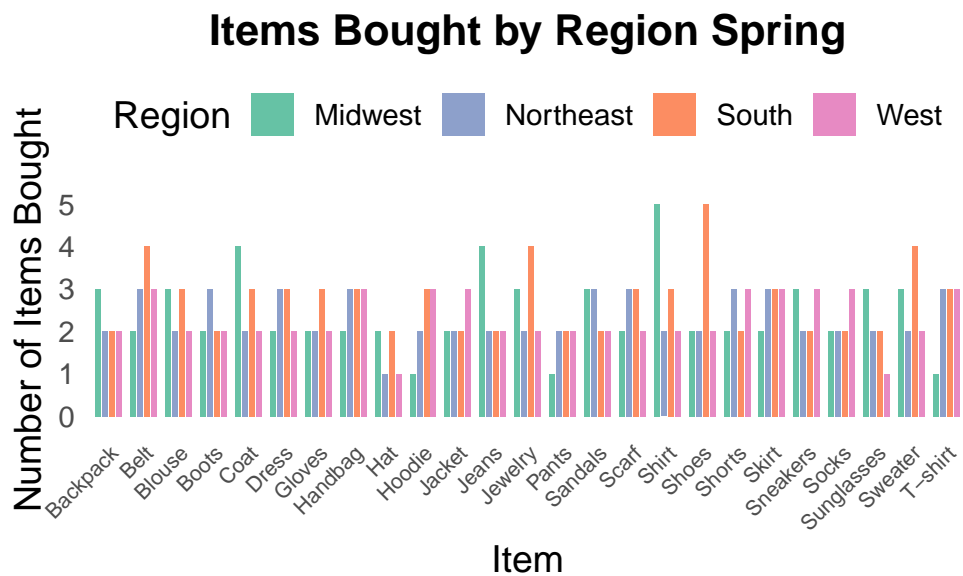
«««< Updated upstream We are utilizing a data set that we found on Kaggle. Kaggle is a website focused towards data scientists with a goal in helping others learn about data. The data was collected by Jahnvi Paliwal, a data science masters student at the University of San Francisco. The data has a usability score of 10.0 and is annually updated by Jahnvi, with the latest update being one month ago. This means the data is collected from a reliable source, and updated consistently to remain relevant. The data set is designed to help you predict whether individuals are likely to change their occupation based on their academic background, job experience, and other demographic factors. This can help you answer questions based on numerous aspects of the job industry including human resources, income, industry analysis, job markets, and job availability. This data set contains over 30,000 records each with 22 attributes. This data set constitutes a case as a single individual. ===== We are utilizing a data set that we found on Kaggle. Kaggle is a website focused towards data scientists with a goal in helping others learn about data. The author of the data is Bhadra Mohit, and they describe it as offering a comprehensive view of consumer shopping trends, aiming to uncover patterns and behaviors in retail purchasing. It contains detailed transactional data across various product categories, customer demographics, and purchase channels. This data set was last updated 20 days ago, and is expected to be updated 4 times a year. This ensures that the data remains relevant and is as accurate as possible. In this data set, case is an individual transaction. This includes the attributes, customer ID, age, gender, item purchased, category, purchase amount USD, location, size, color, season, review rating, subscription status, payment method, shipping type, discount applies, promo code used, previous purchases, preferred payment

method, and frequency of purchases. We intend to focus on age, gender, item purchased, location, season, review rating, and previous purchases. All of the attributes come from the data set, but we also created the attribute of region, which groups all fifty states into four regions: Northeast, South, Midwest, and West. This is based upon the national recognized regions in the United States.

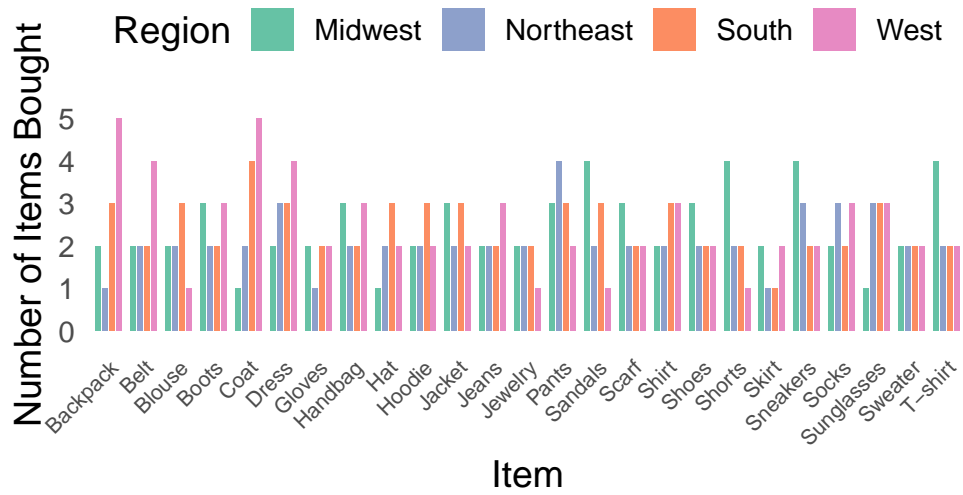
FAIR Principles

The data we are utilizing meets the FAIR principles. The data is **findable**, and includes unique identifiers as well rich and substantial metadata. Each case is given an ID number, and there are numerous attributes that they are defined by. The data is **accessible** and can be found in our public repository. We also downloaded the data from Kaggle, which is public and we were able to easily locate it and access it. The data is **interoperable** and uses the R language which is widely known and accepted. This way our data can be exchanged between collaborators and allows for open communication. By citing our sources and explaining the provenance of our data, this ensure our data is also **reusable**. By meeting the FAIR principles we are ensuring that our data is universal, and can be easily understood. »»»> Stashed changes

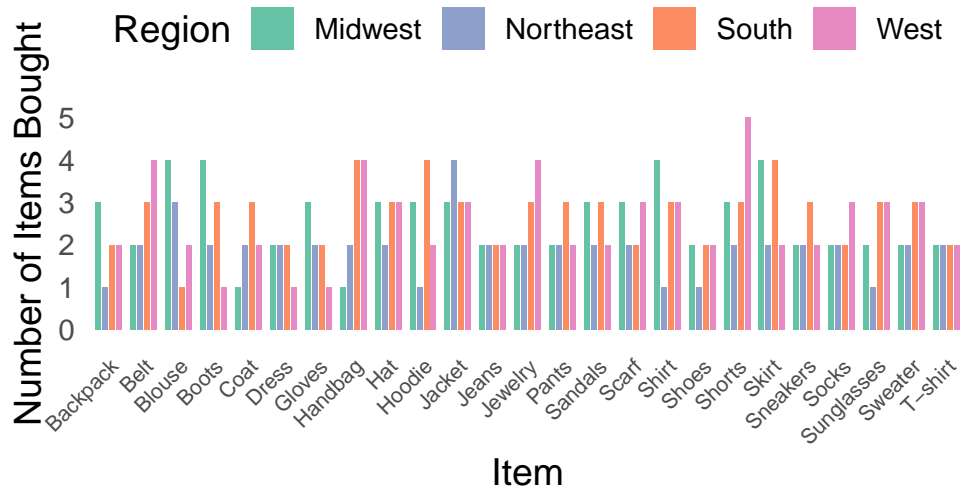
FAIR and CARE Principles



Items Bought by Region Summer



Items Bought by Region Fall



Items Bought by Region Winter

