Data Analysis: Shopping Behaviors

Kelby Palmer

Data: What's in My Basket?

- Age
- Gender
- Location
- Item purchased
- Category
- Size
- Color
- Season
- Amount spent (in USD)
- Payment method

Exploring U.S. shopping behavior using Kaggle data

Data: Dimensions

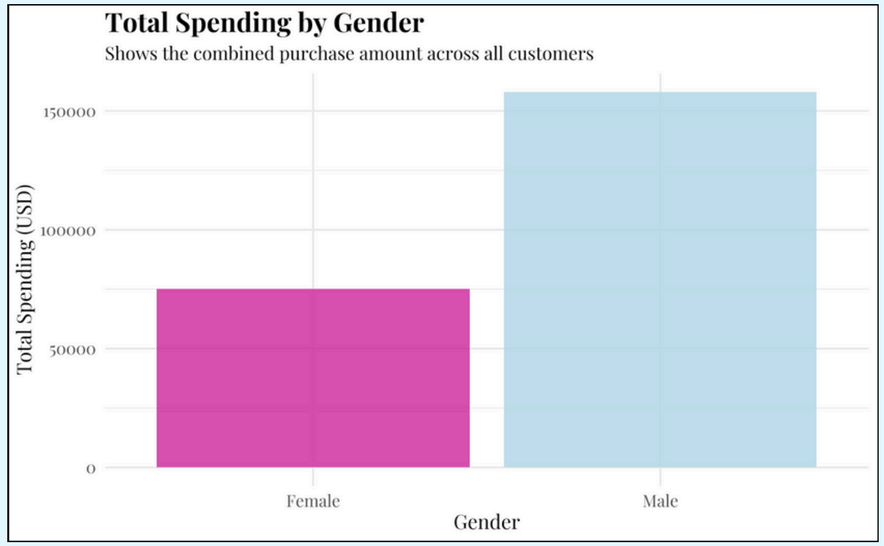
- 3,900 shopper records from the U.S.
- 10 variables used (out of 18 in the original dataset)
- Mix of categorical and numerical features
- No major cleaning required
- Data was already tidy, so no major restructuring needed

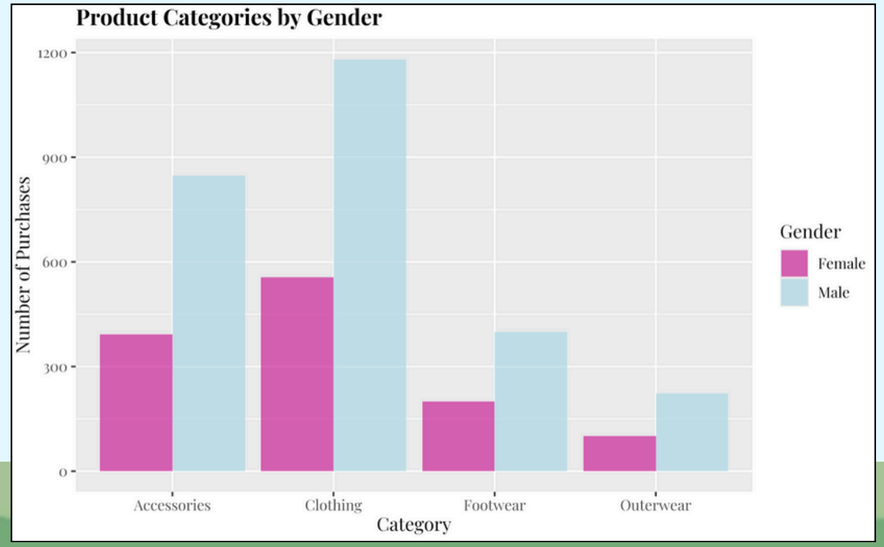
Limitations/Things to Note:

- Dataset is male-heavy → affects totals and spending comparisons
- Another limitation: Only seasonal labels (not specific dates), so timing patterns can't be explored in detail

** I focused on spending patterns, color trends, gender differences, and seasonal/regional activity.

Interesting Stats: Gender Imbalance

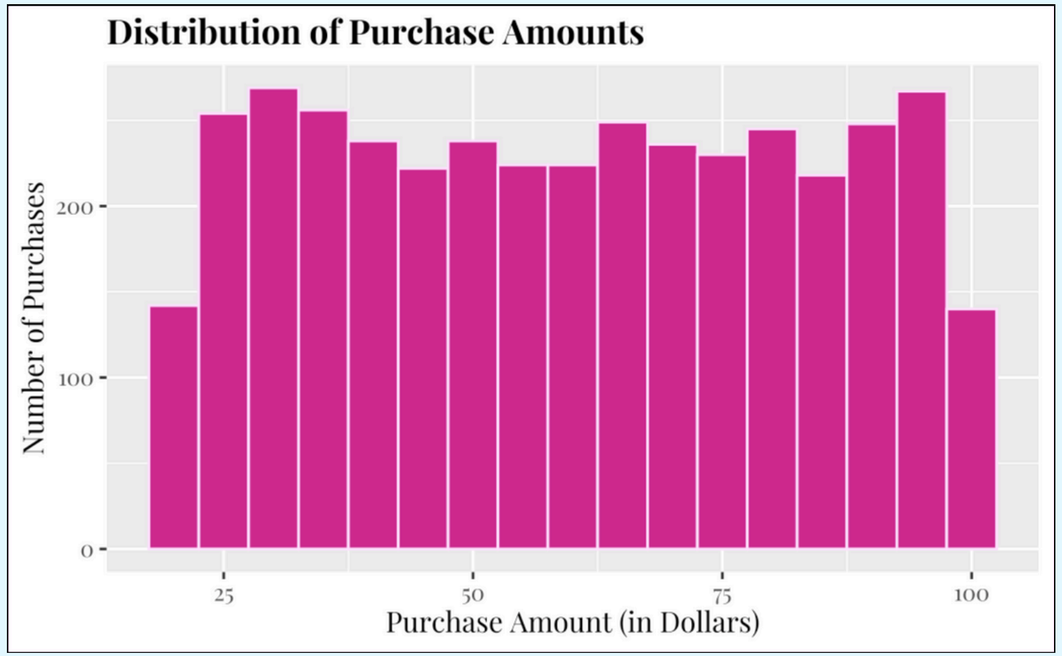


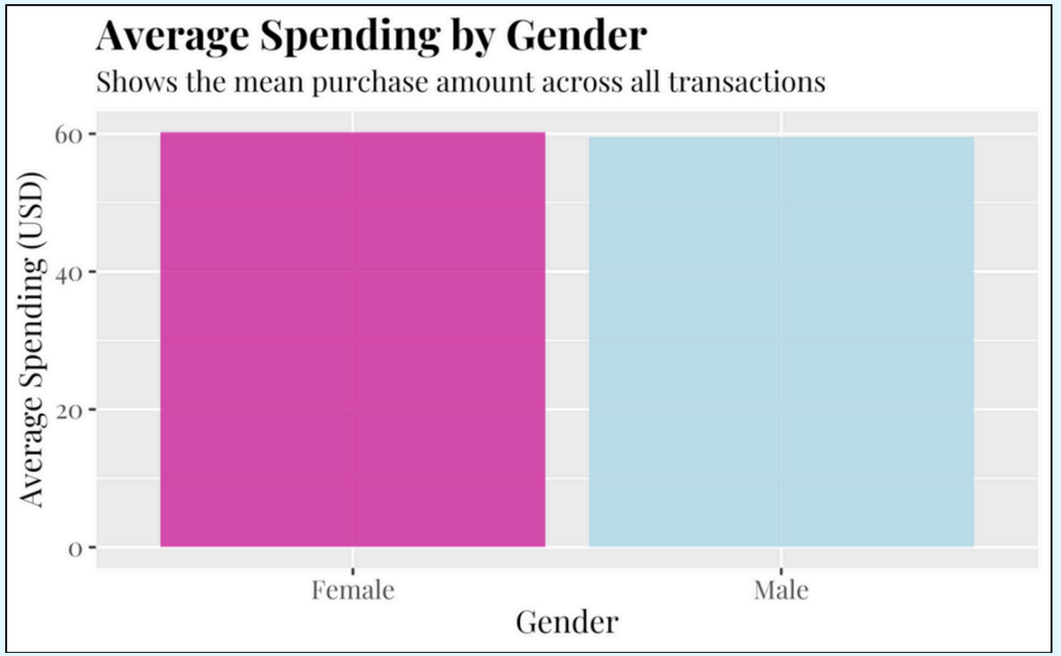


- Dataset is more male-heavy (more than a 2:1 ratio)
- Skewed category counts, color trends, and total spending
- Important to keep in mind when interpreting results

Gender <chr>	n <int>
Female	1248
Male	2652

Interesting Stats: Average Spending



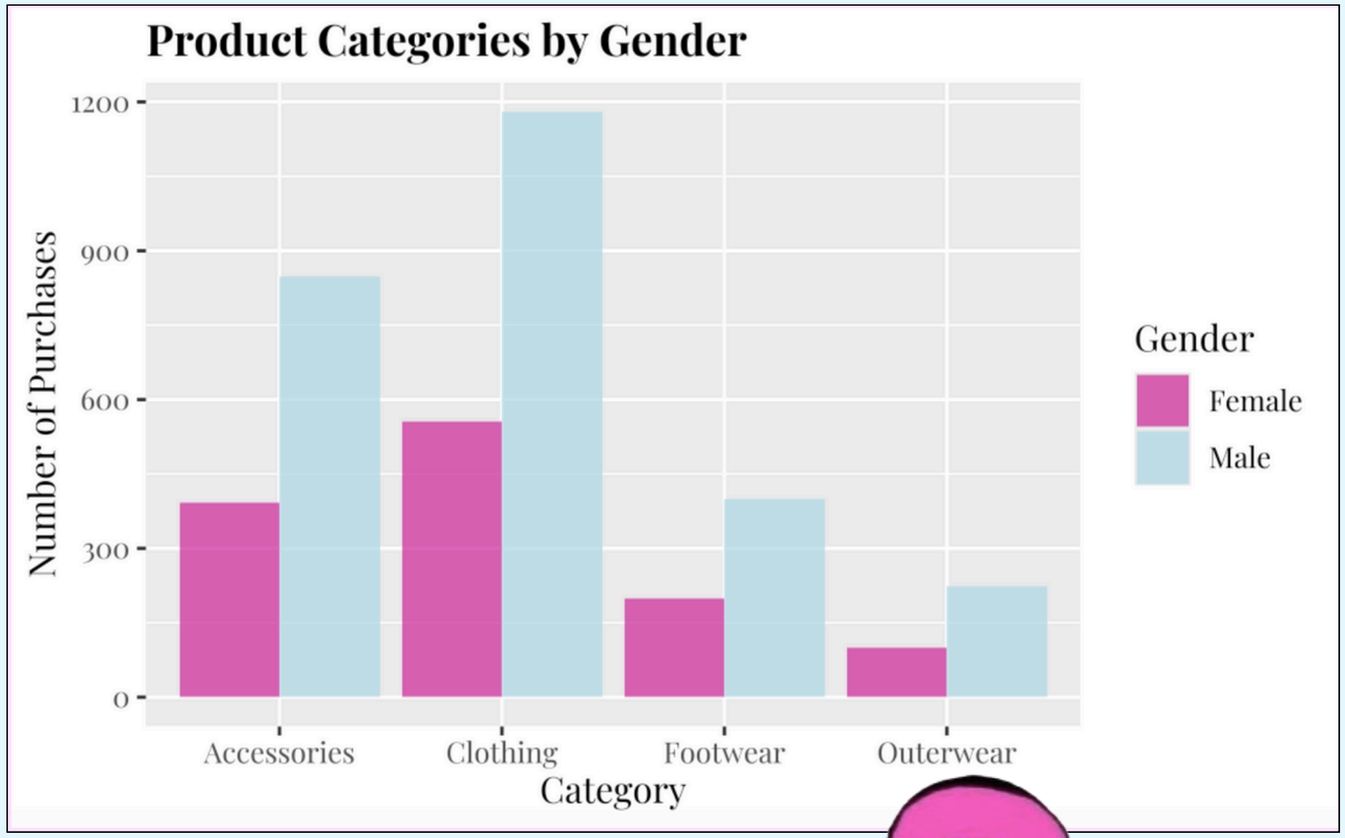


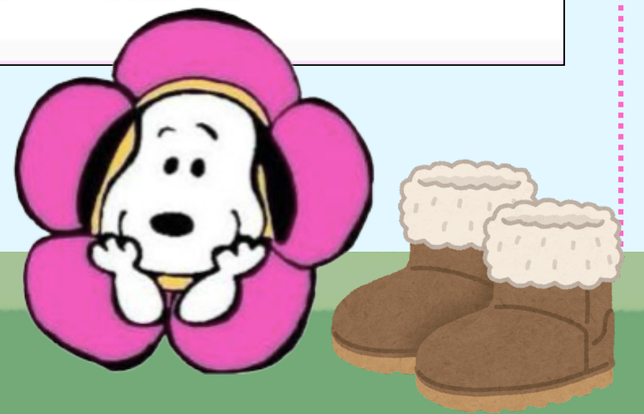
- Spending isn't totally random, there are a couple clear peaks:
  - One cluster is around $30–35, and another is closer to $90–95.
  - So people tend to shop on the lower or higher end rather than in the middle.
  - This suggests two types of shoppers: quick, cheaper buys vs. bigger planned purchases.

- Average purchase amount was about the same for both genders (around $59–$60)
- Women spent slightly more on average (keep gender skew in mind)
- Men appear to spend more overall only because there are way more men in the dataset
- Averages are the better comparison (totals are misleading here)

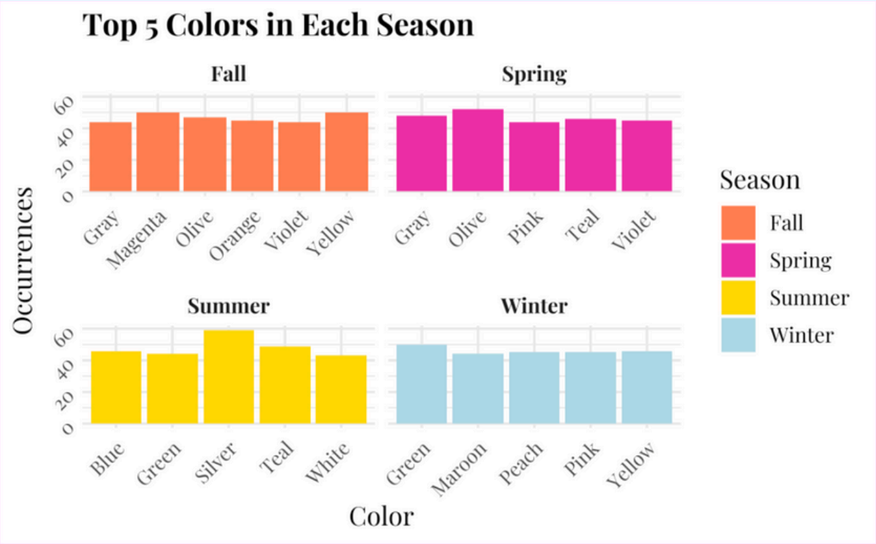
Interesting Stats: Popular Categories

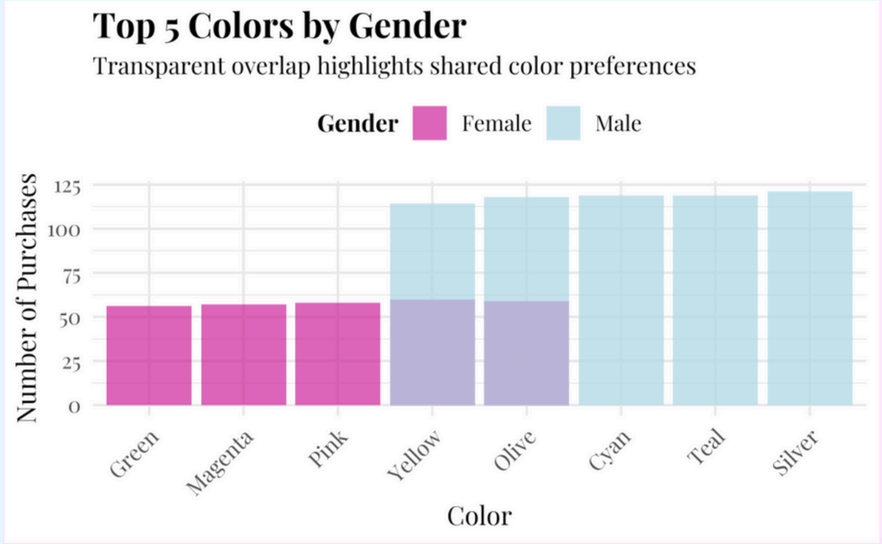
- Clothing was the most purchased category overall (see male skew)
- Accessories were also high
- Footwear & outerwear were bought much less often
- This pattern was consistent across genders
- Suggests shoppers mainly prioritize everyday/wearable items
- **** Again, likely reflects “everyday” vs. “occasional” shopping patterns

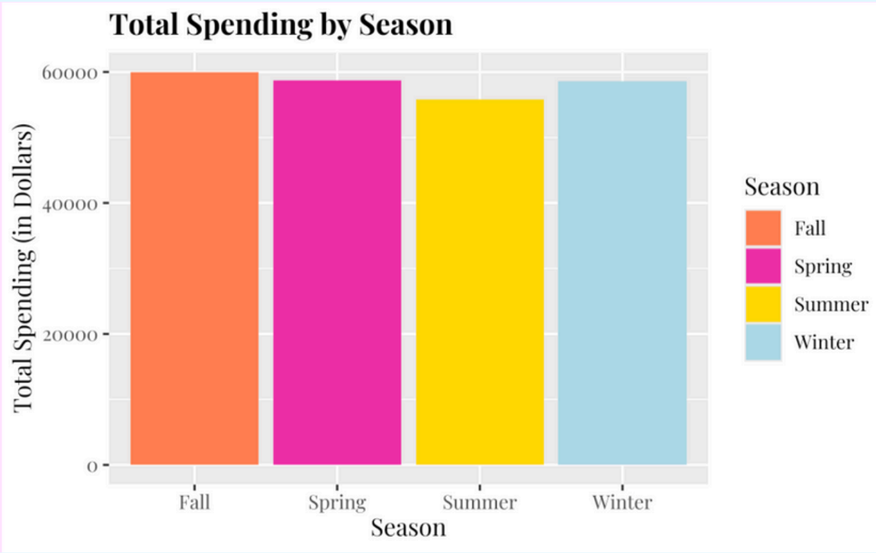


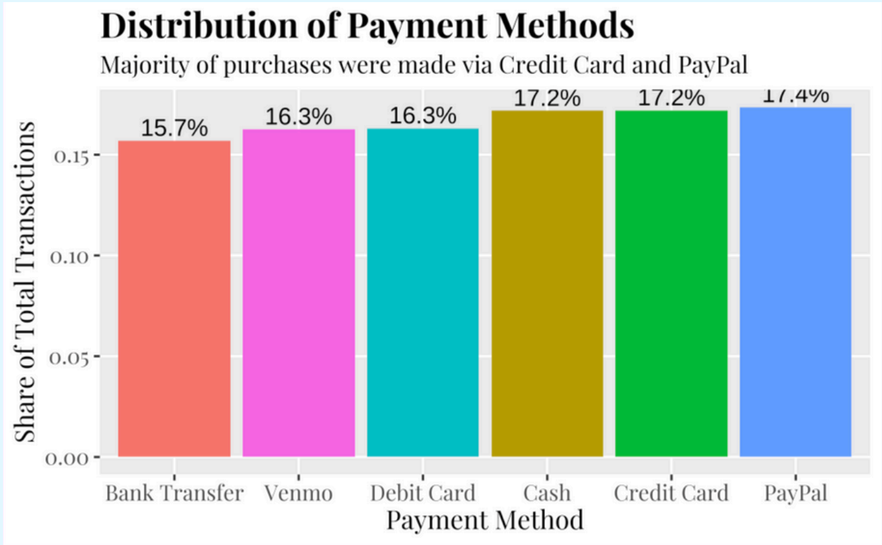


Other Interesting Visuals









Summary of Analysis:

- Gender imbalance heavily influenced totals!!
- Spending amounts peaked in two clusters ($30s and $90s)
- Average spending nearly identical for men + women
- Clothing + accessories dominated all purchases
- Patterns were consistent across genders and seasons

Future Work:

I'd like to further explore:

- Explore color trends (more detail)
- Compare states/regions using a map
- Look at age subgroups (not just overall averages)
- Analyze review ratings + previous purchases
- See if payment method affects spending patterns
- Explore size trends across seasons/categories
- Use more detailed dates instead of just seasons
- Compare with a messier but similar dataset to see if the trends I found are still consistent there.
- Look at data/trends across several years

Thank you!

Kelby Palmer ~ COP 2073