

An abstract graphic on the left side of the slide features a dark background with various colorful lines and shapes. These include a blue line forming a large circle, a yellow line forming a square, and several smaller lines in red, green, and purple. Some lines are straight, while others are curved or wavy. There are also some small circles and squares scattered around. The overall effect is a dynamic and colorful composition.

Manipulating a Synthetic Dataset

CONSUMER BEHAVIOR AND SHOPPING HABITS

By: Holly, Brooke, Kylie, and Nicole



OUTLINE

- 01** Our Data Set
- 02** Restrictions
- 03** Our Intended Outcomes
- 04** Our Findings Using ML
- 05** Tableau Dashboard
- 06** EDA
- 06** Conclusions

OUR DATASET

Consumer Behavior and Shopping Habits Dataset provides comprehensive insights into consumer preferences, tendencies and patterns during their shopping experiences

Variables

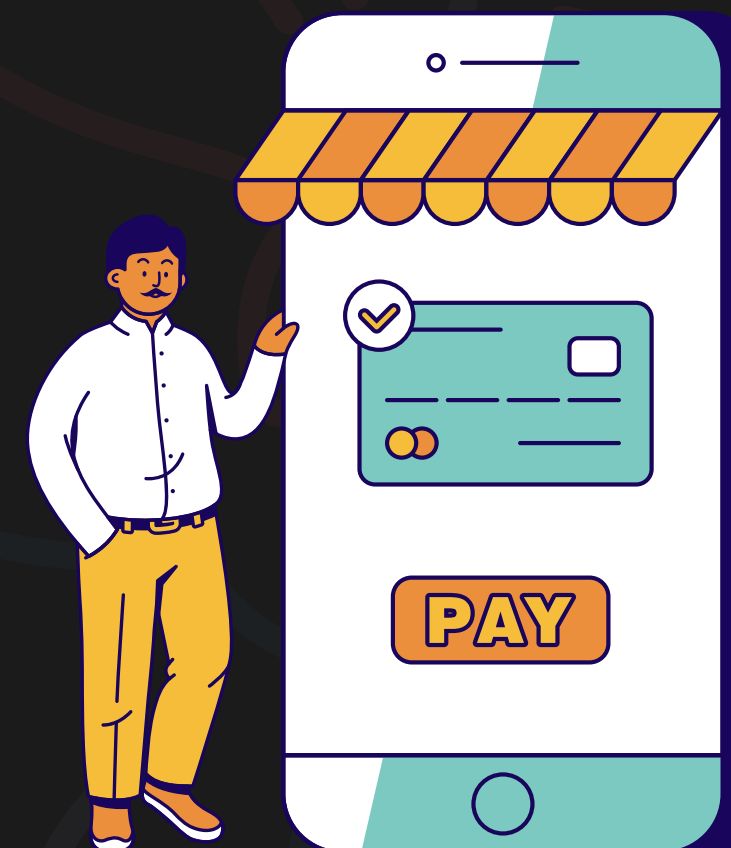
- Demographic information
- Purchase history
- Purchase preferences
- Shopping frequency
- Online and offline shopping behavior

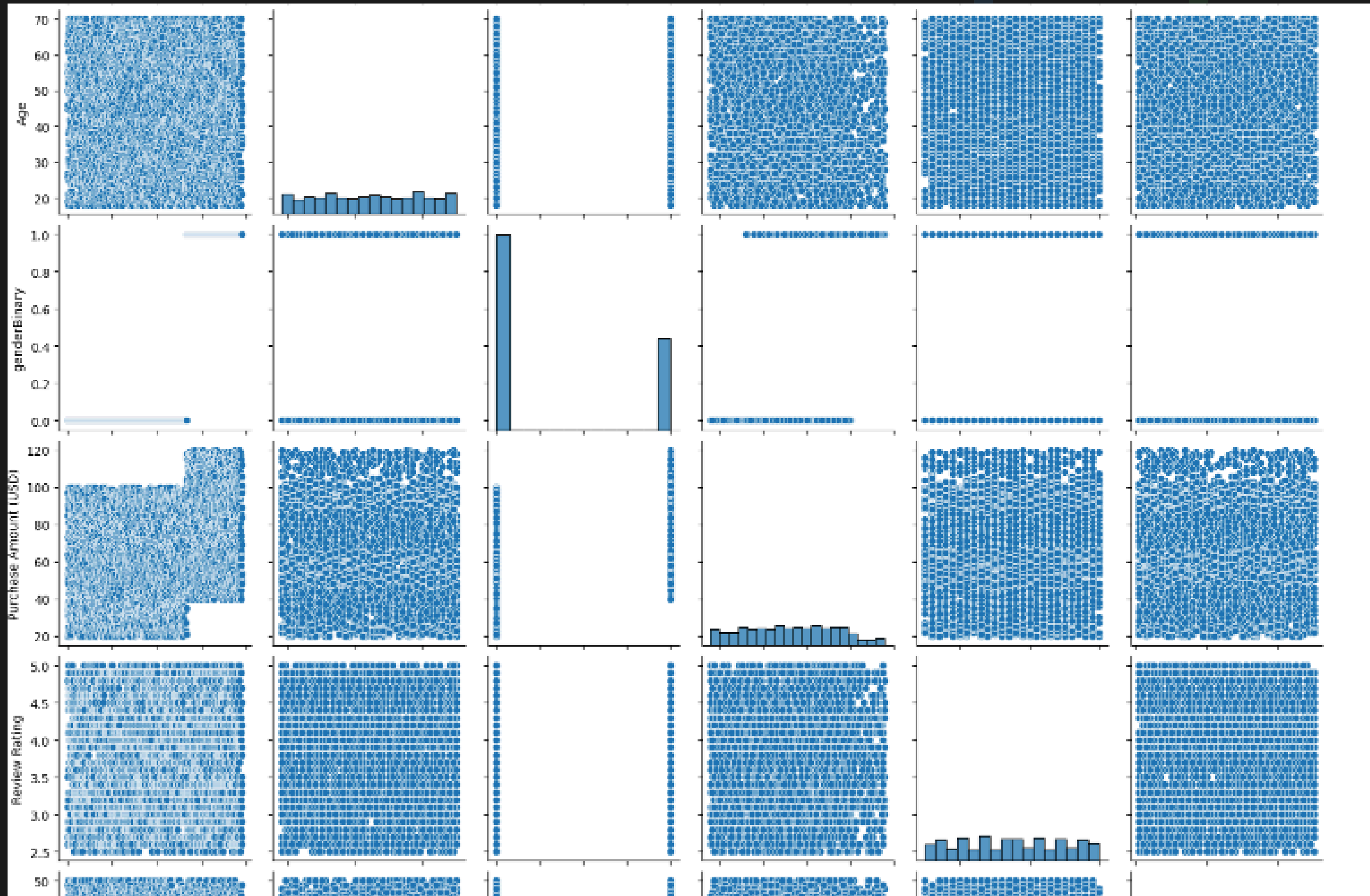
Intended Use

Provides an overview of consumer preferences and purchasing behavior to be used for businesses aiming to tailor their strategies to meet customer needs and enhance their shopping experience, driving sales and loyalty

Synthetic Data

Artificially generated to mimic real data for testing algorithms and models when real data is limited or sensitive. Used to increase dataset size, and test scenarios that may not be accessible in real data.





Pairplot to display no relationship between any variables in the dataset

OUR INTENDED OUTCOME

Linear Regression

Show relationships between independent variable and its impact of the dependent variable, model fitting to find a best-fitting line that predicts relationship

- Age vs Purchase Amount or Predicted Purchase Amount

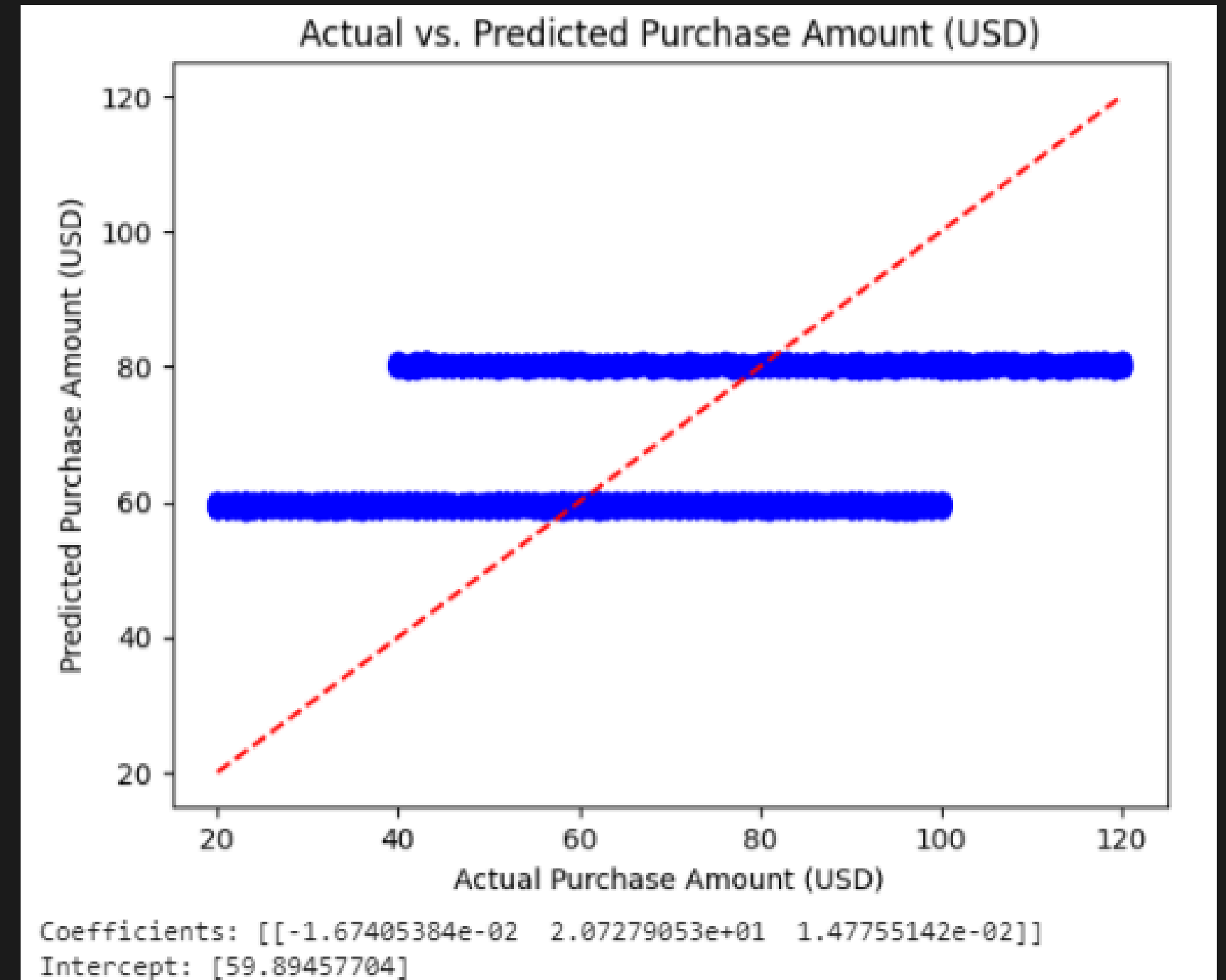
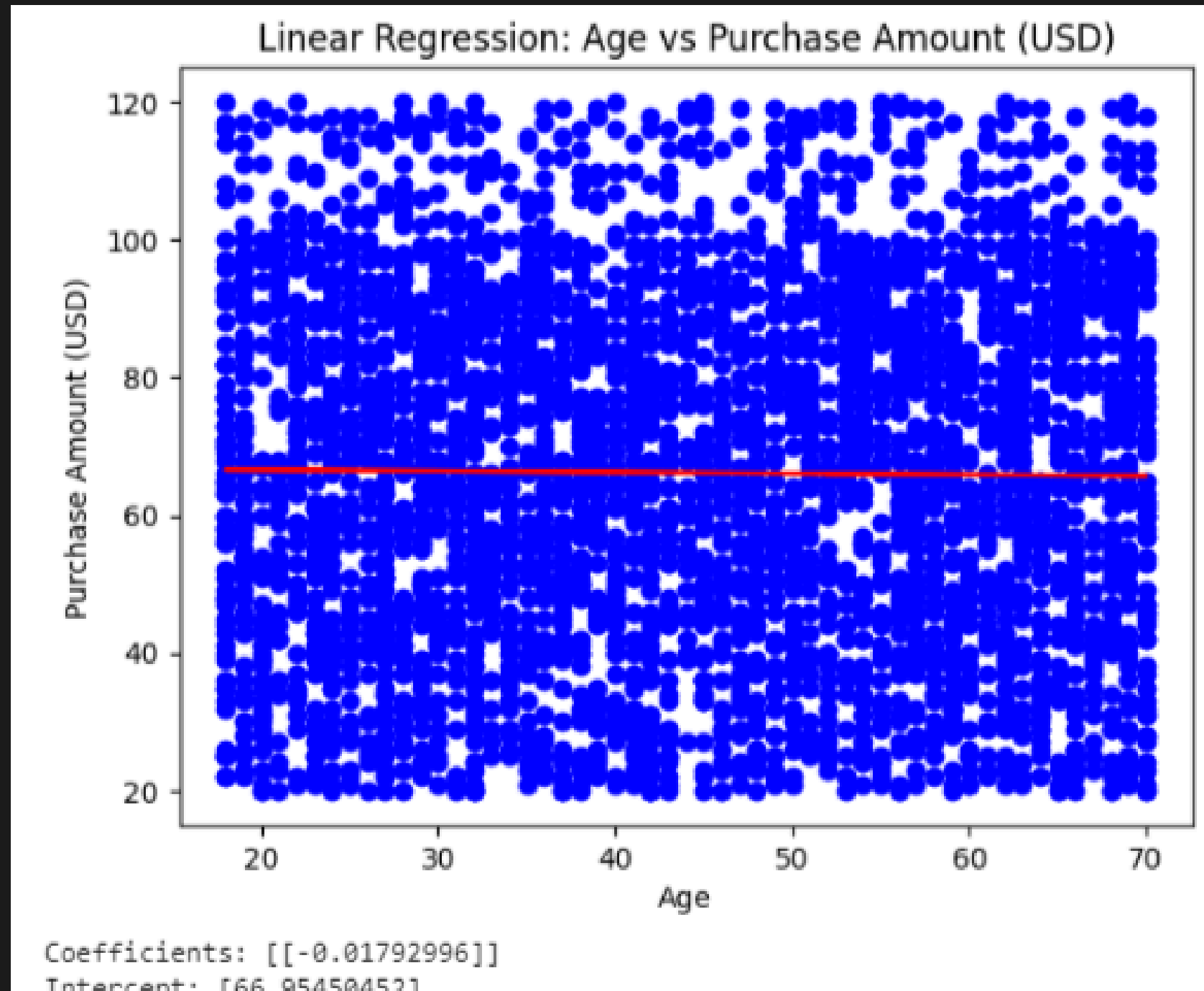
A/B Testing

Compares two or more versions of something to determine which performs better

- We wanted to see if Gender (Male or Female) impacts Purchase Amount (USD)



LINEAR REGRESSION



A/B TESTING

Gender vs Purchase Amount
(USD)

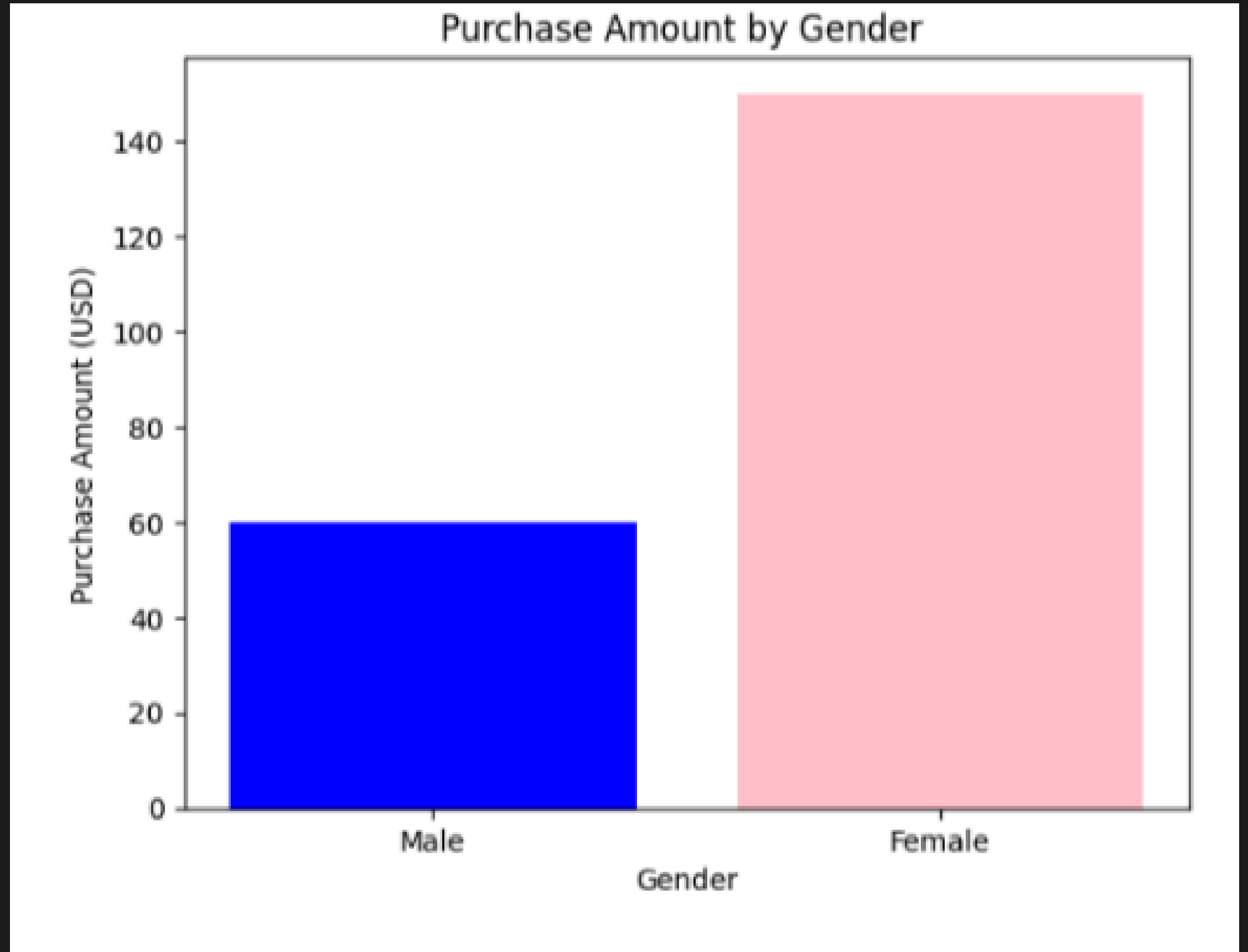
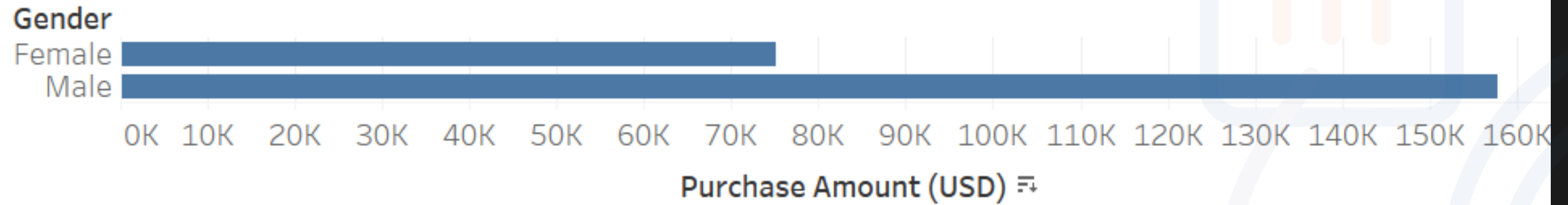
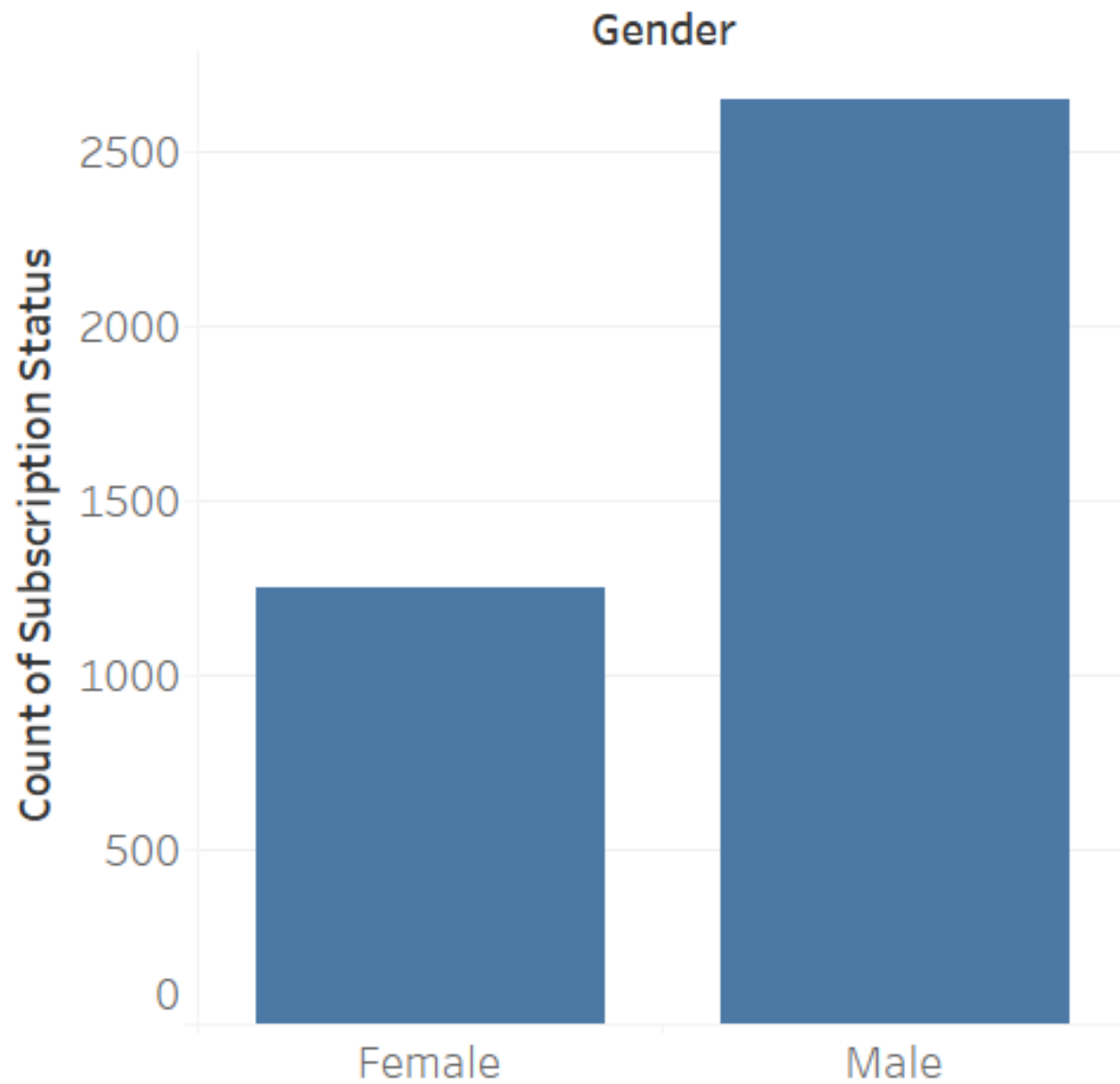


TABLEAU DASH- BOARDS

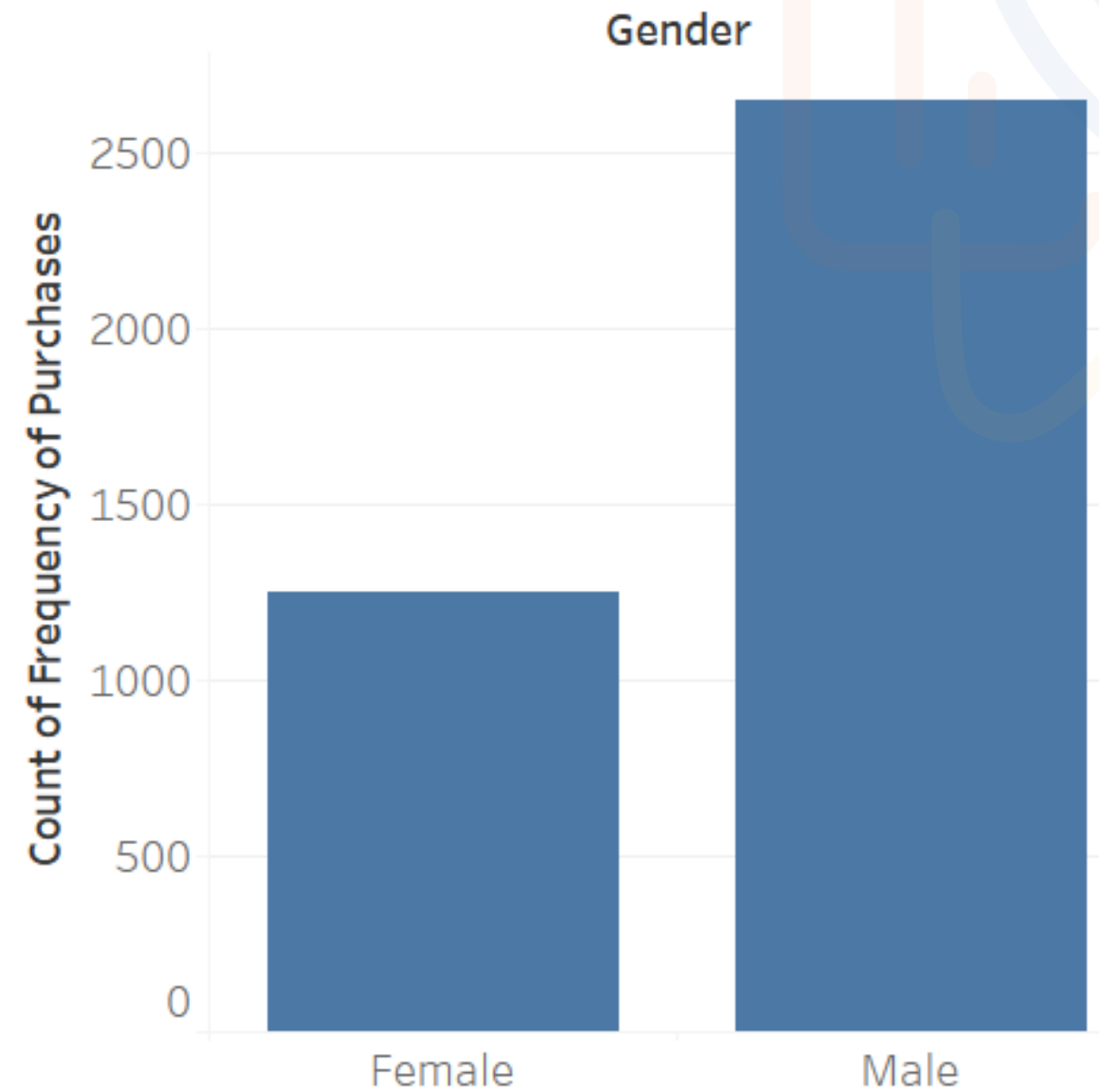
Gender vs Purchase Amt



Gender vs Subscription Status

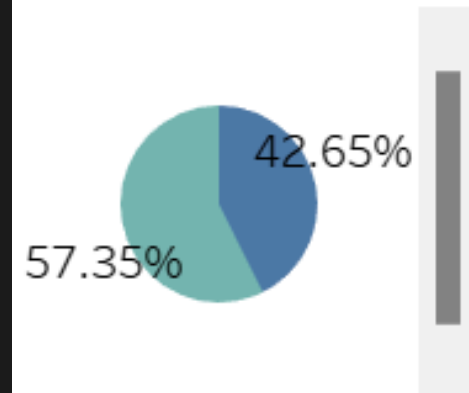


Gender vs Freq of Purchase

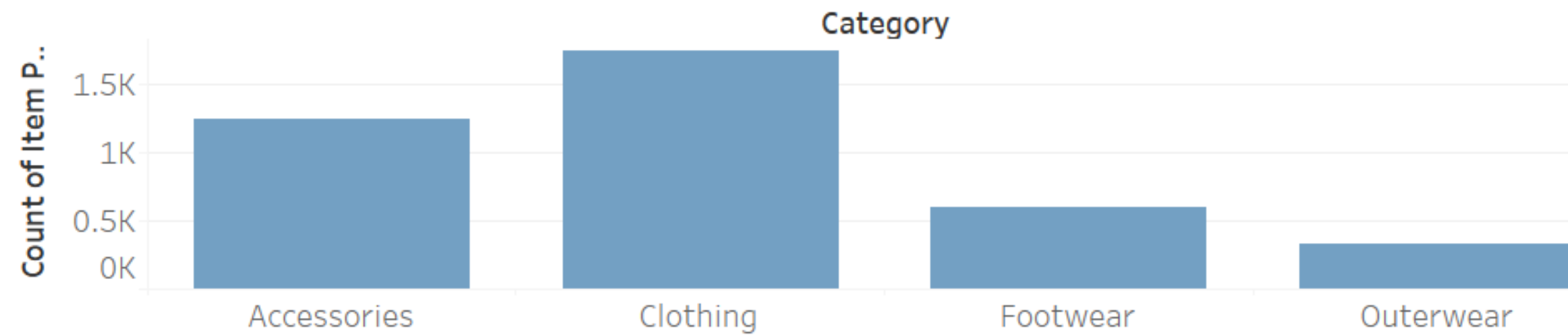


Shopping Behaviors Dashboard

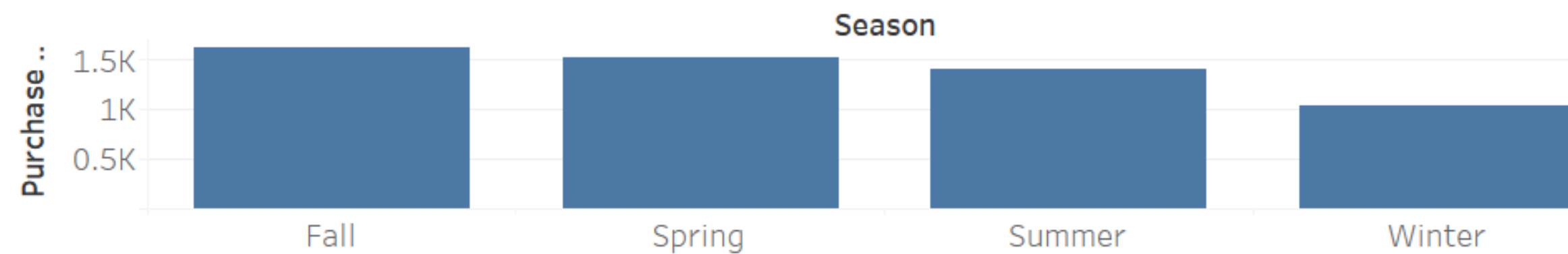
Promo
Code Used



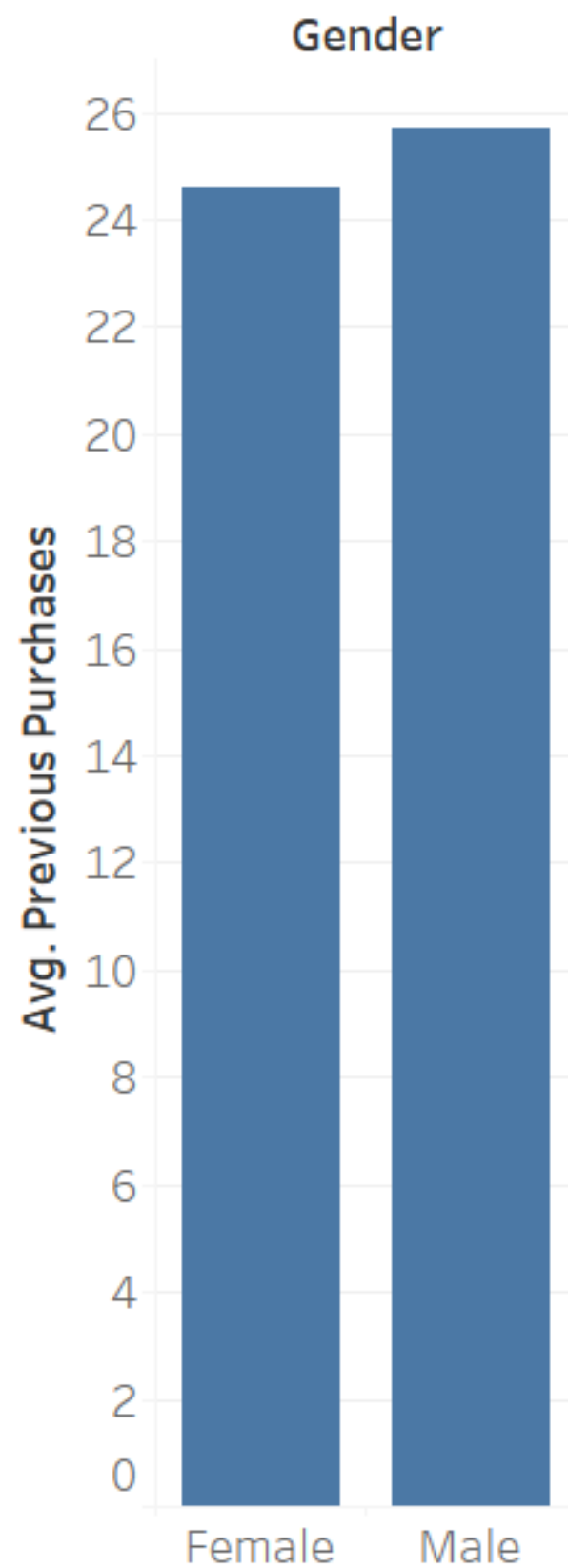
Item Purchased vs Category



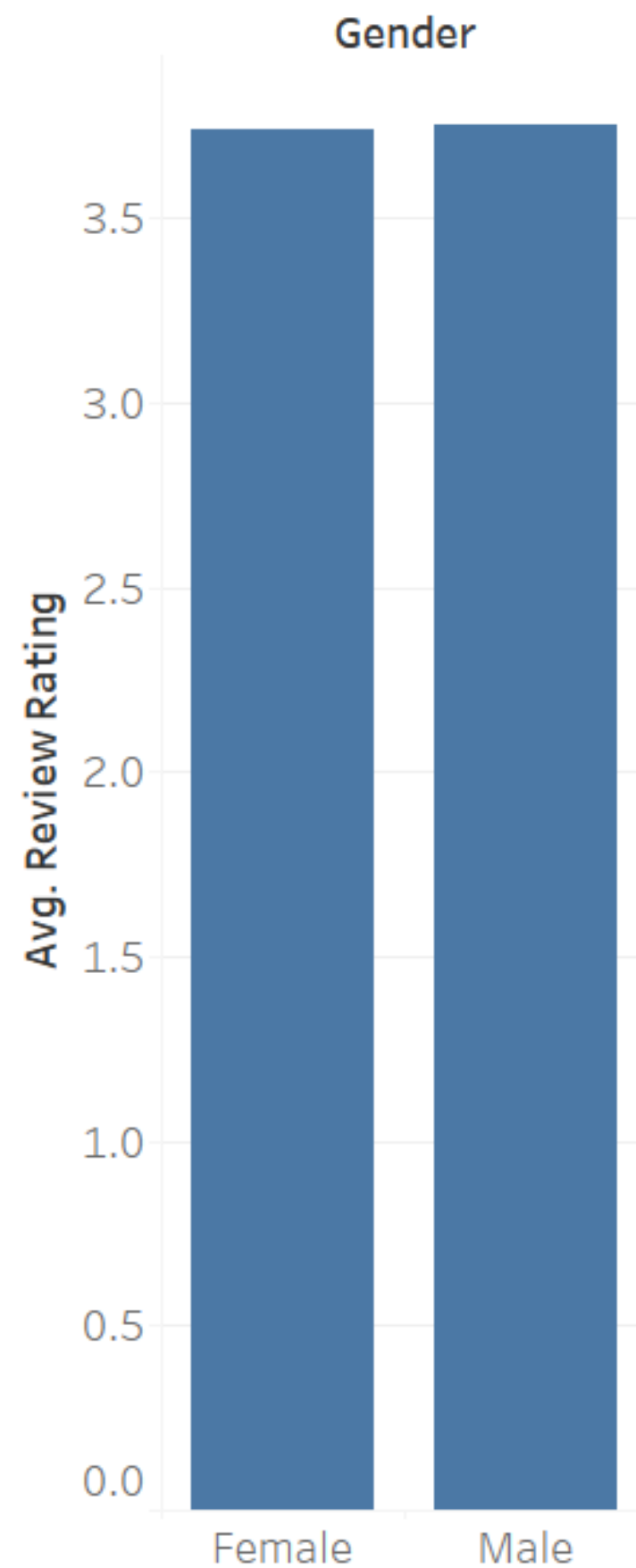
Season vs Purchase Amt



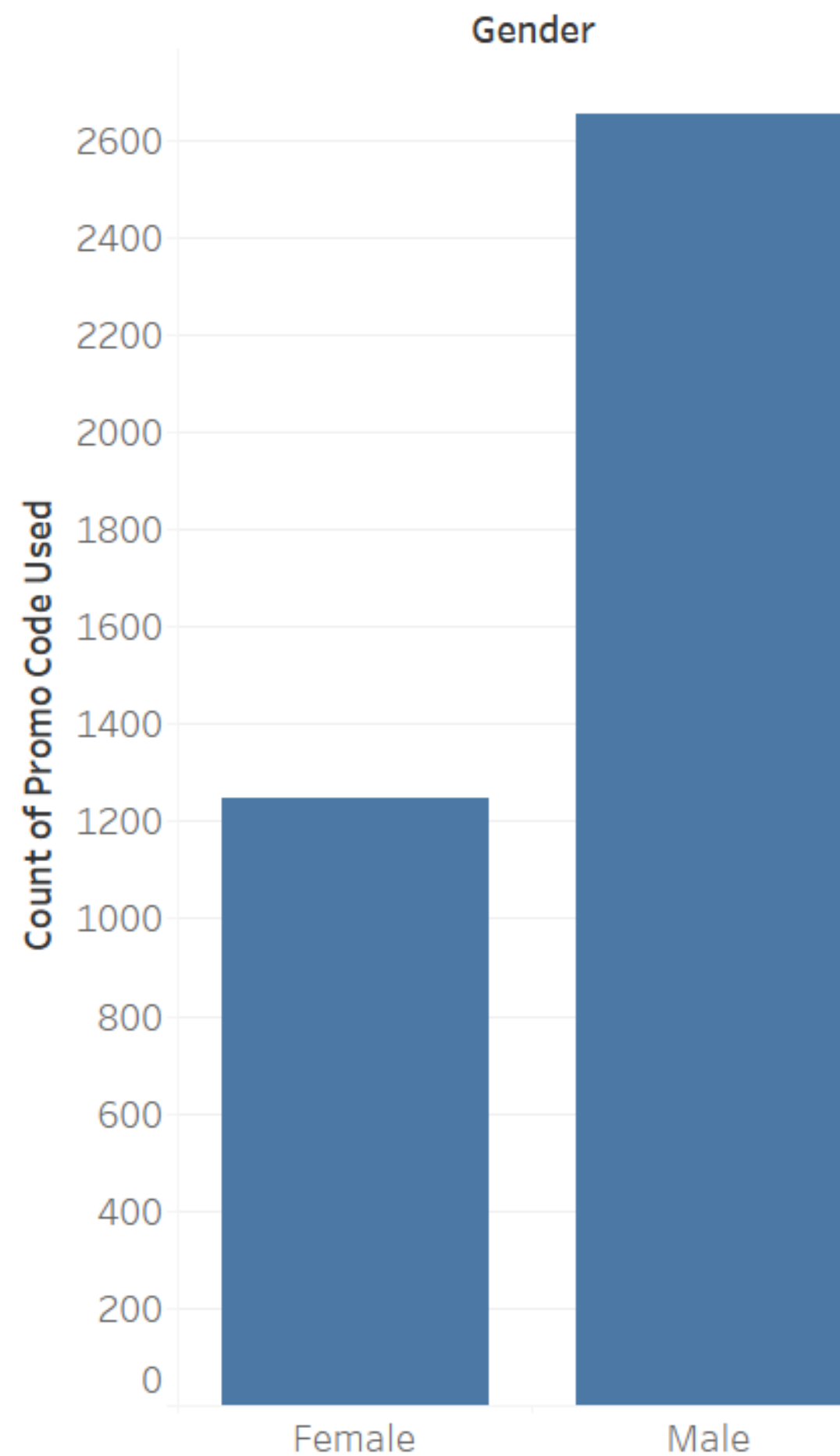
Gender vs Avg
Previous
Purchases



Gender vs Review
Rating

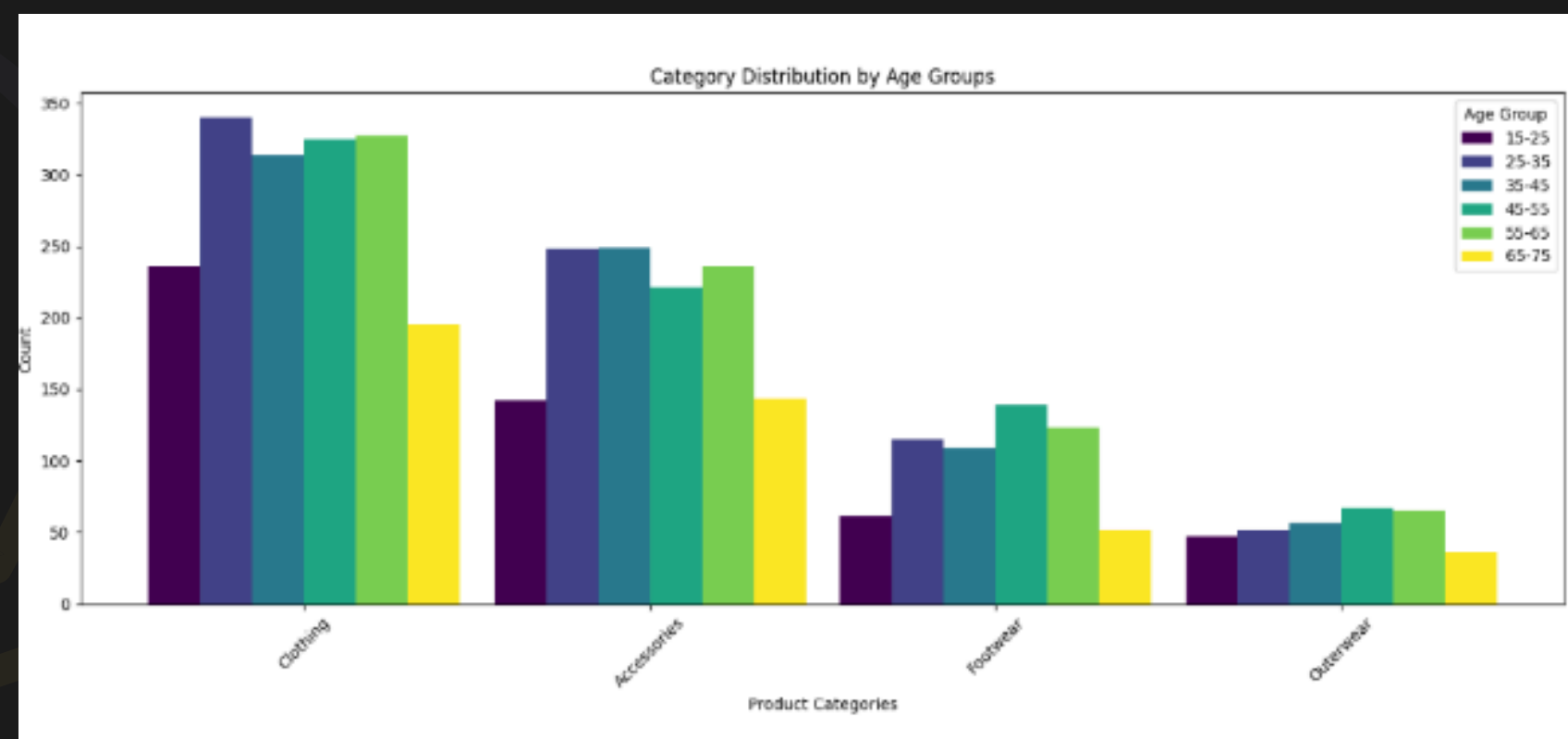
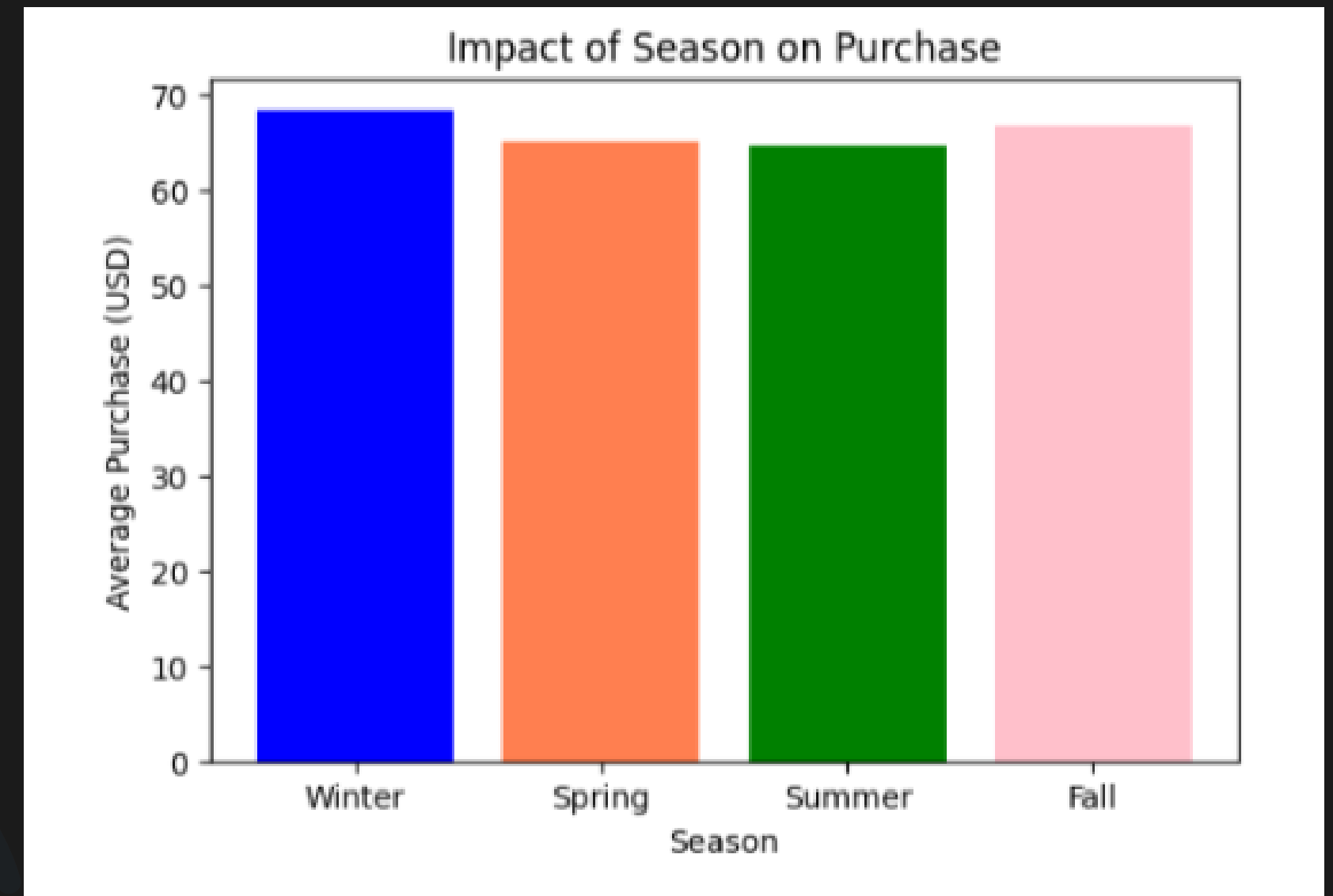
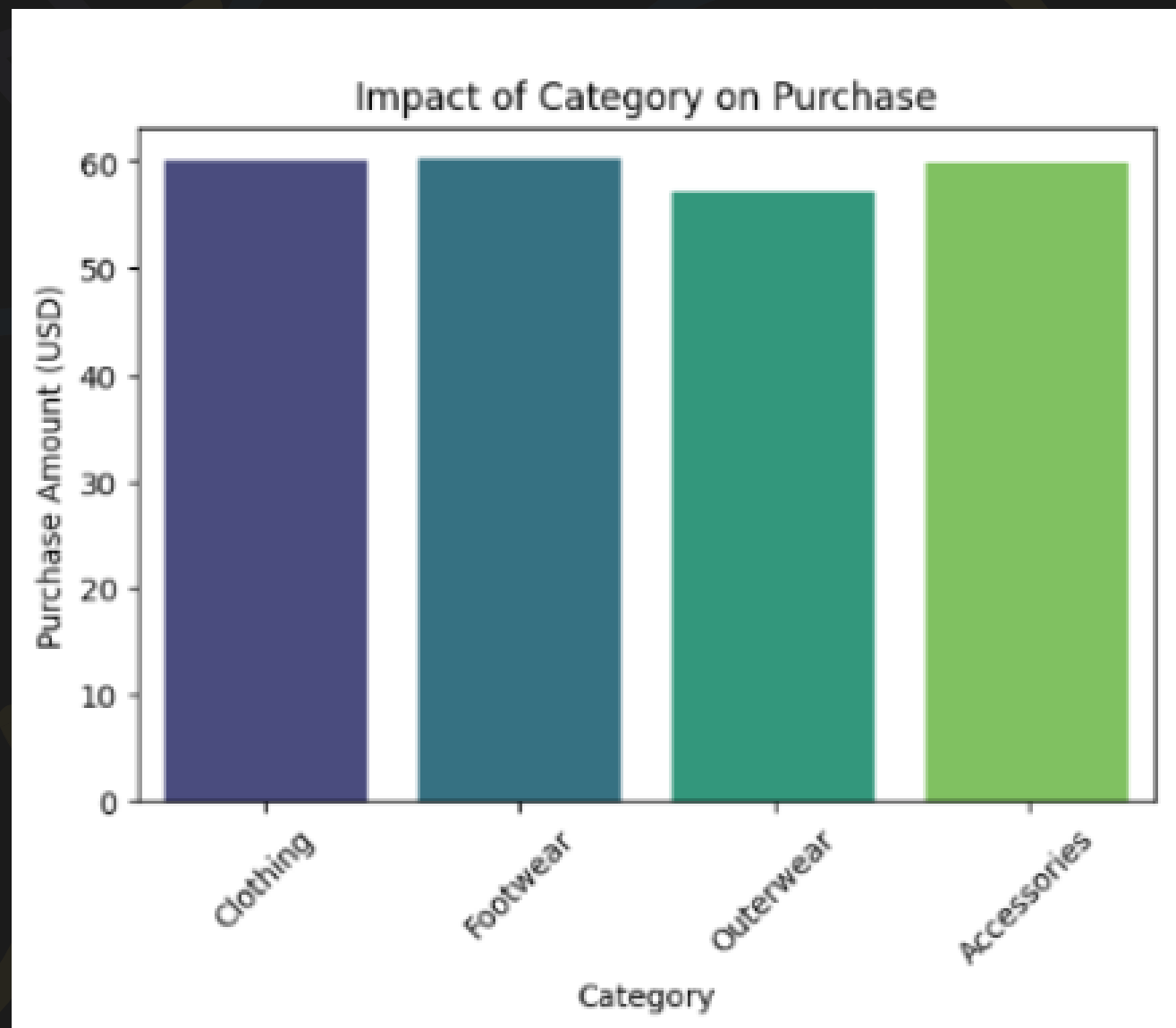


Gender vs Promo Code Used



EDA FROM KAGGLE

If we were to do this project again, we now know the different signs that a dataset is synthetic



CONCLUSIONS

- We learned how to manipulate synthetic data
- We went through many ML models before we found that the data was the issue, not our models
- Don't always trust related workbooks on Kaggle

The background is a dark charcoal grey. It is filled with various abstract, hand-drawn style elements in bright colors: yellow, light blue, purple, orange, green, and pink. These include loops, swirls, straight lines, and geometric shapes like rectangles and circles. Some shapes contain horizontal lines, resembling simplified icons for a menu or a list. The overall effect is a busy, celebratory, and modern aesthetic.

**THANK YOU
FOR
LISTENING!**