Manipulating a Synthetic Dataset

CONSUMER BEHAVIOR AND SHOPPING HABITS

By: Holly, Brooke, Kylie, and Nicole

OUTLINE

OUR DATASET

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

Variables

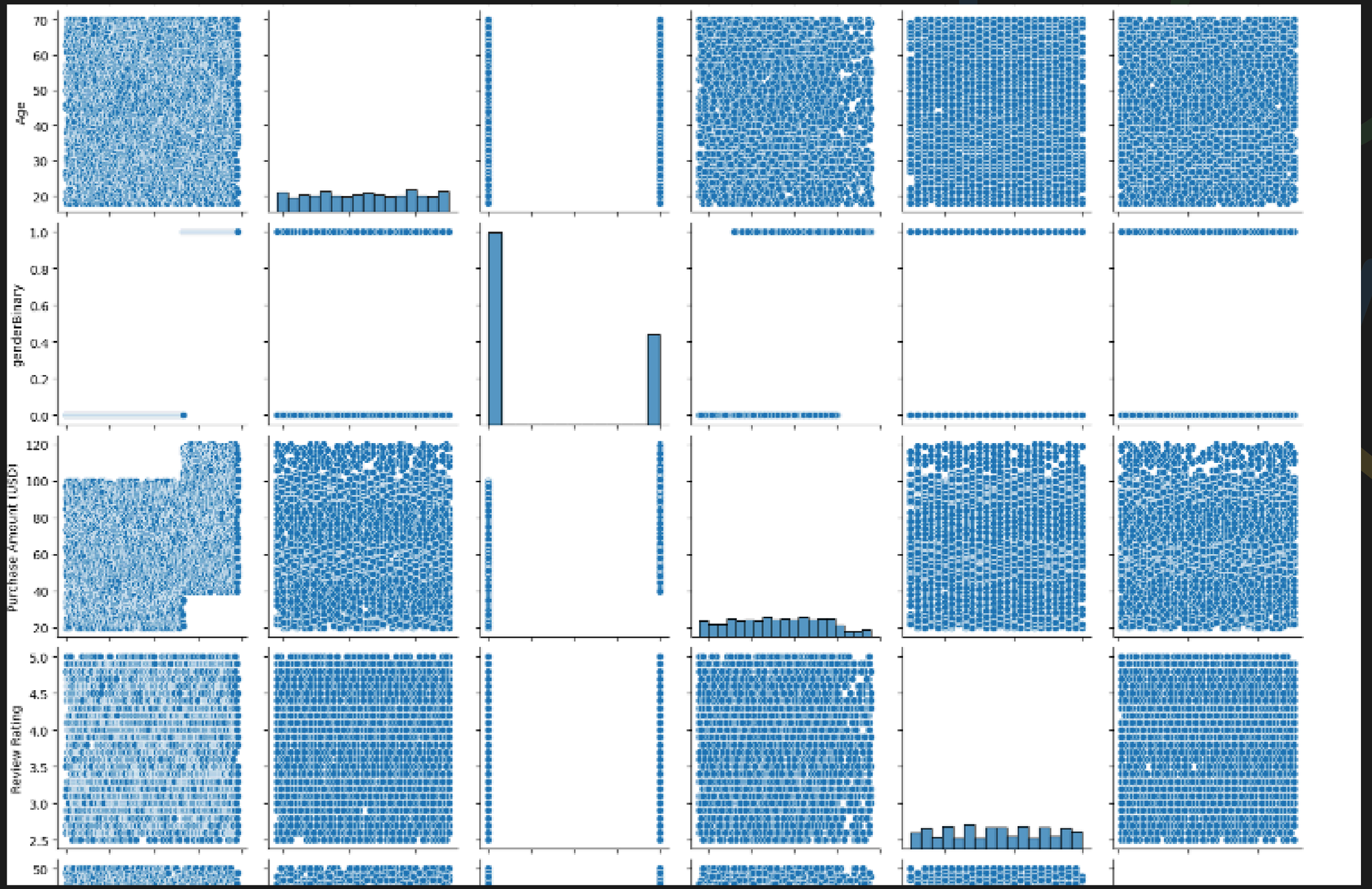
- Demographic information
- Purchase history
- Purchase preferences
- Shopping frequency
- Online and offline shopping behavior

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

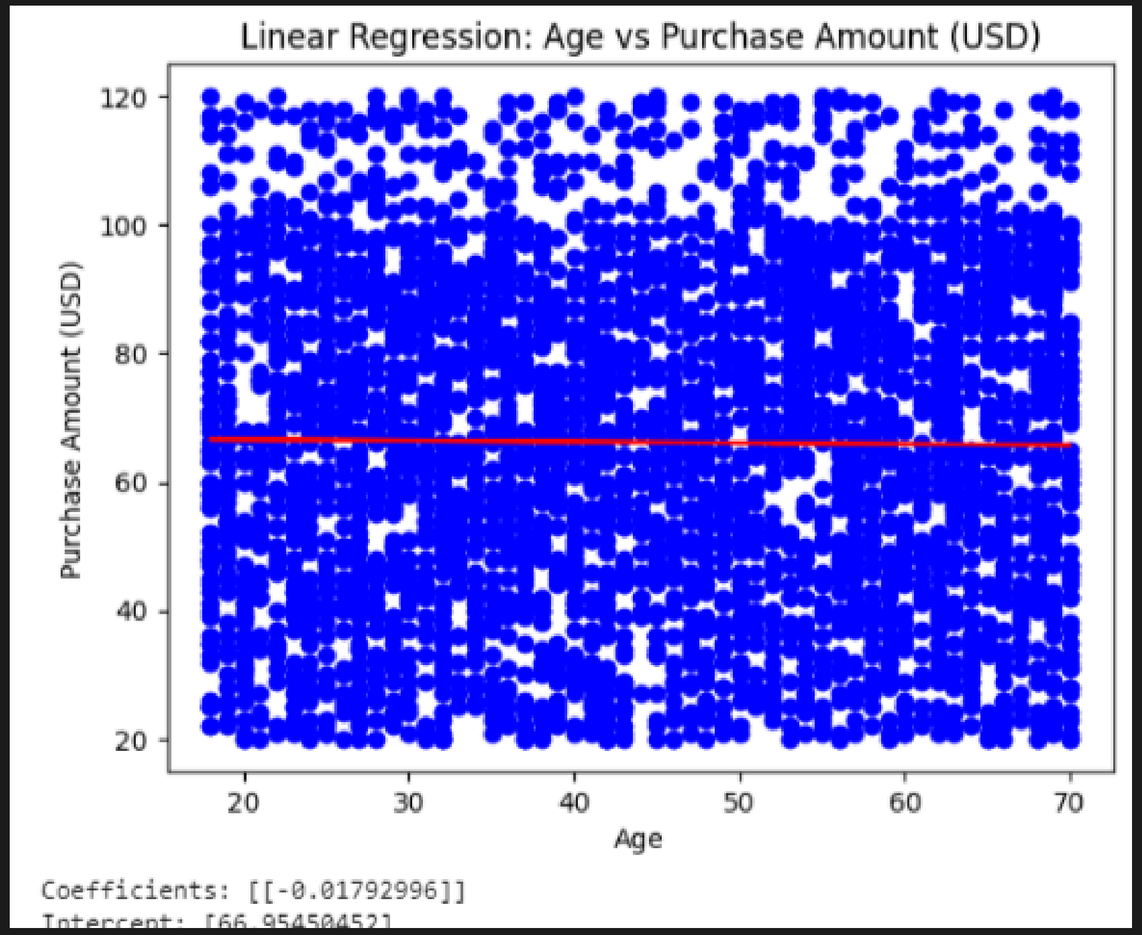
- Age vs Purchase Amount or Predicted Purchase Amount

A/B Testing

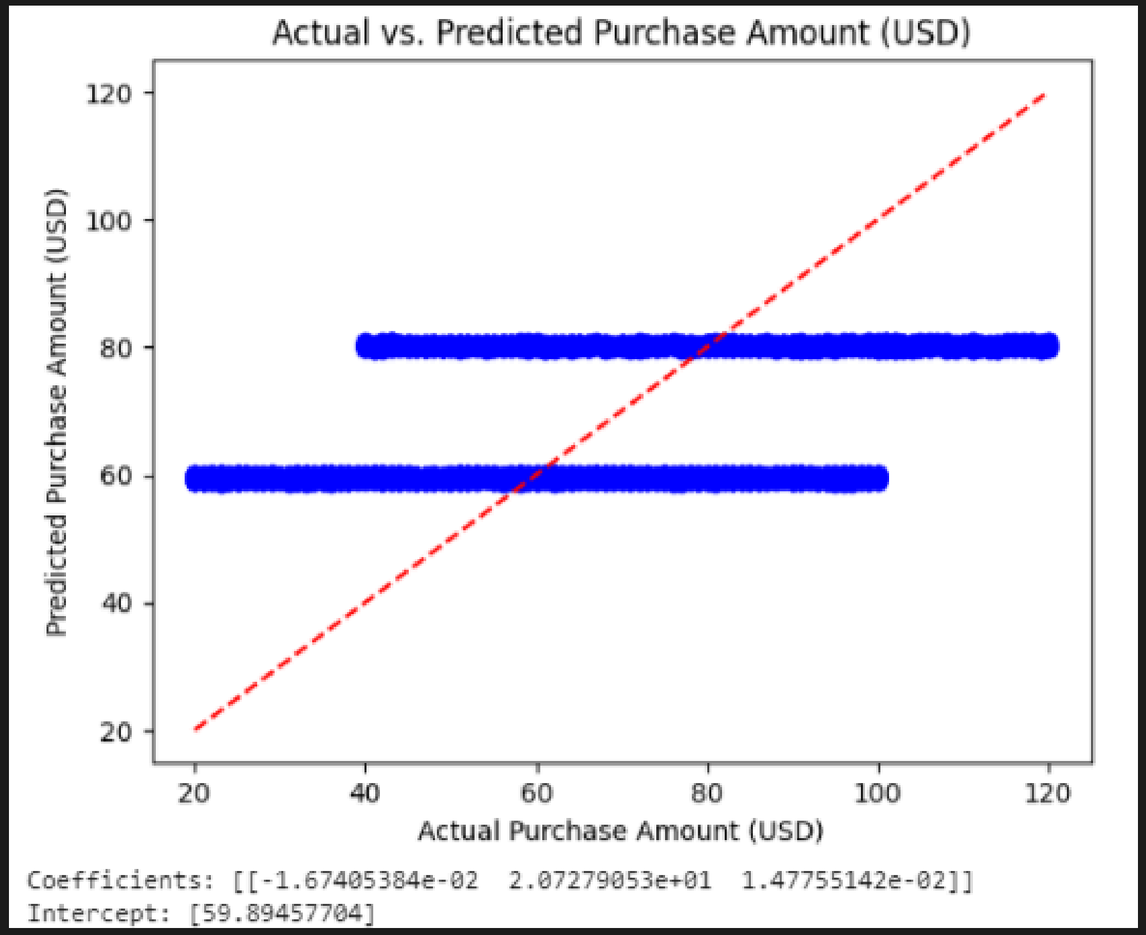
Compares two or more versions of something to determine which performs better

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

LINEAR REGRESSION



Coefficients: [[-0.01792996]]
Intercept: [66.95450452]



Coefficients: [[-1.67405384e-02 2.07279053e+01 1.47755142e-02]]
Intercept: [59.89457704]

A/B TESTING

Gender vs Purchase Amount (USD)

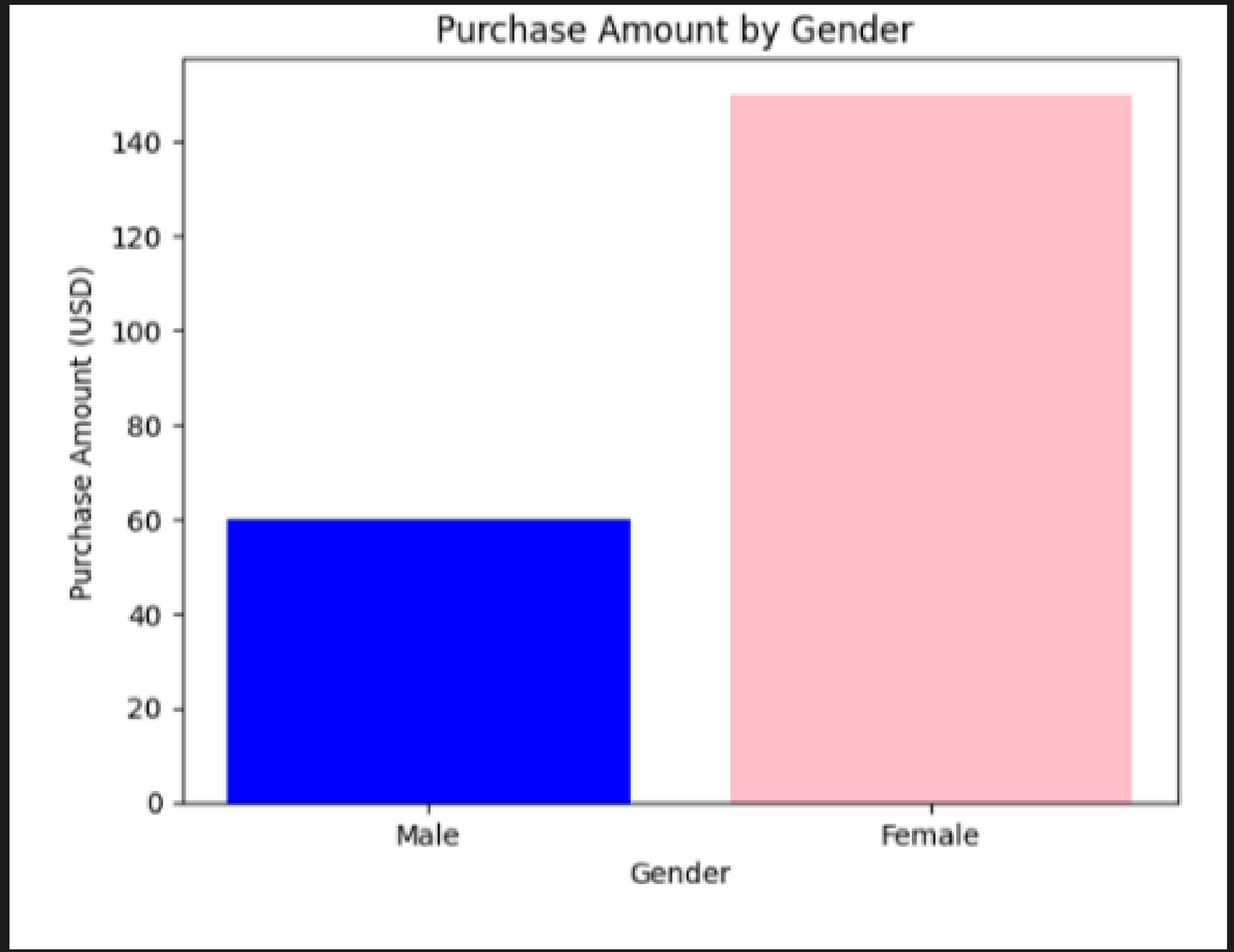
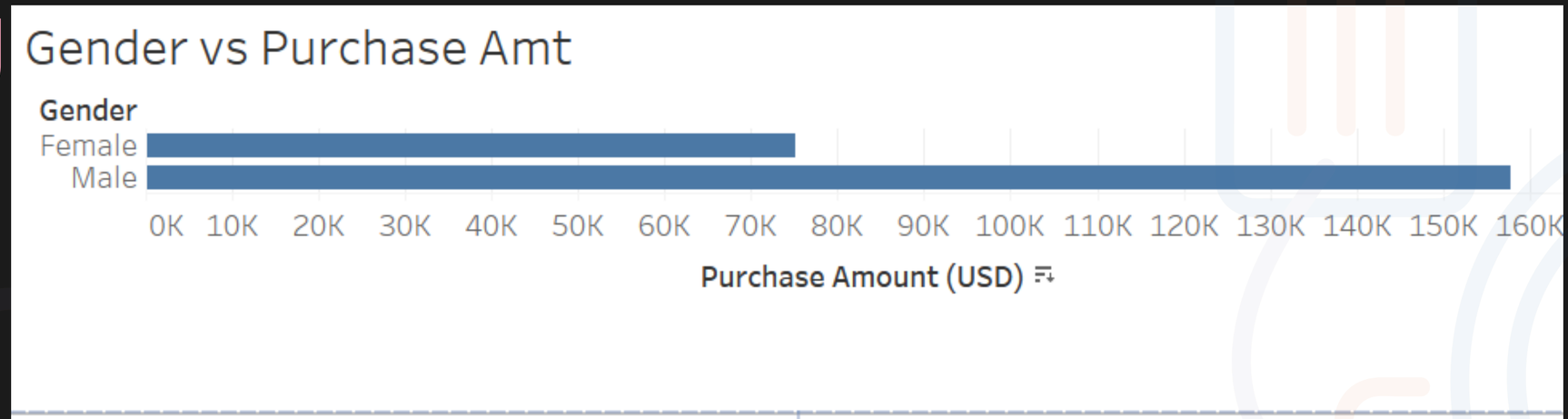
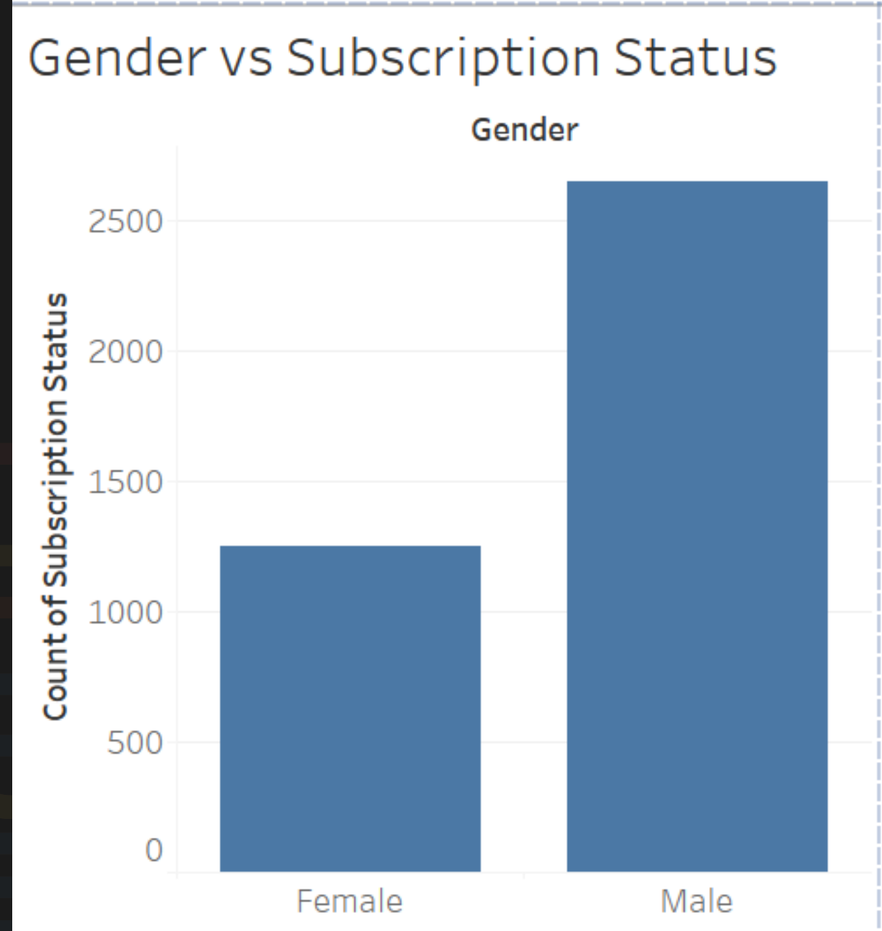
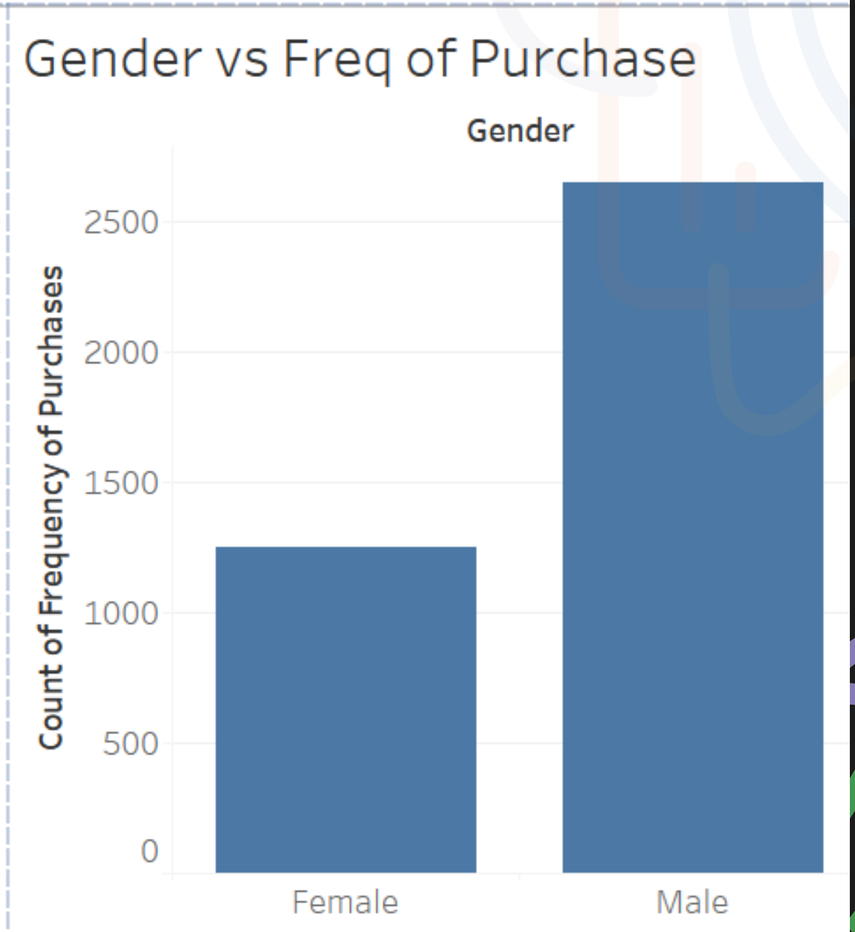


TABLEAU DASH-BOARDS

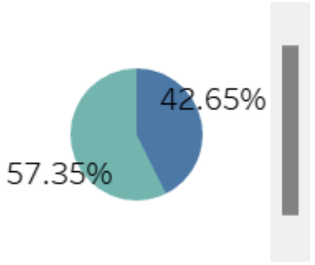




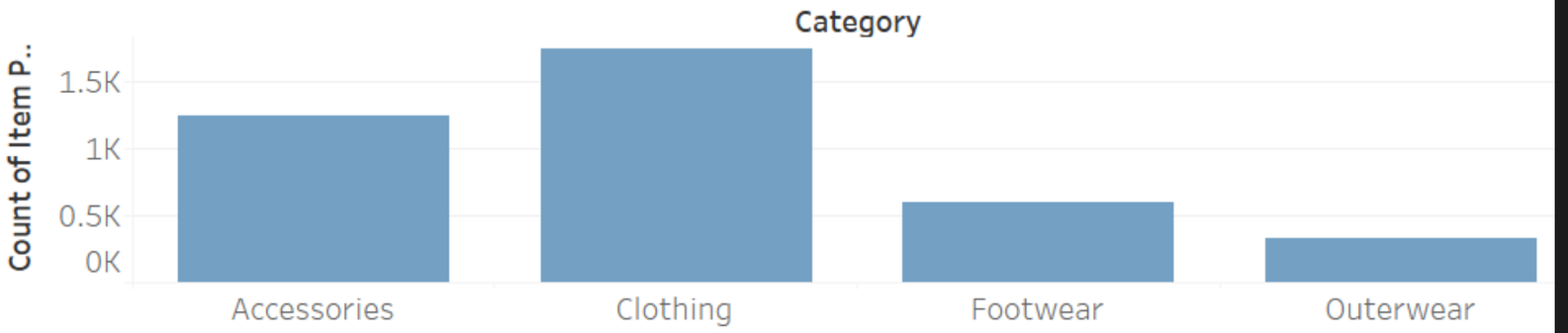


Shopping Behaviors Dashboard

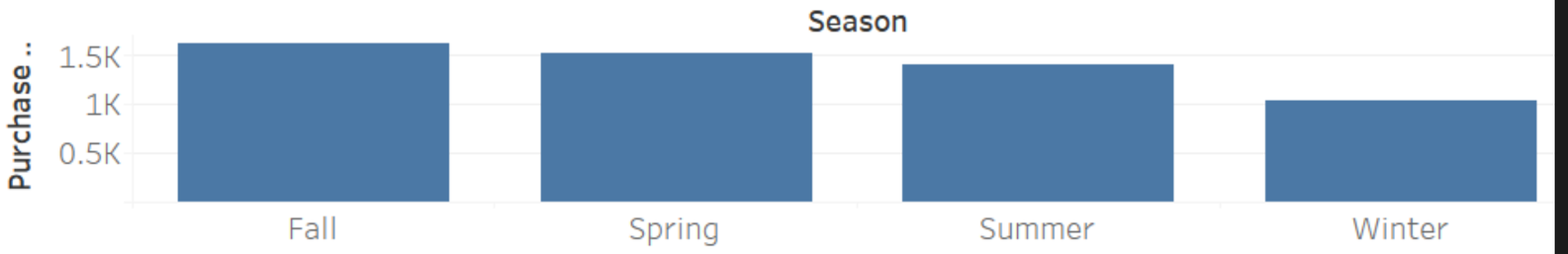
Promo Code Used



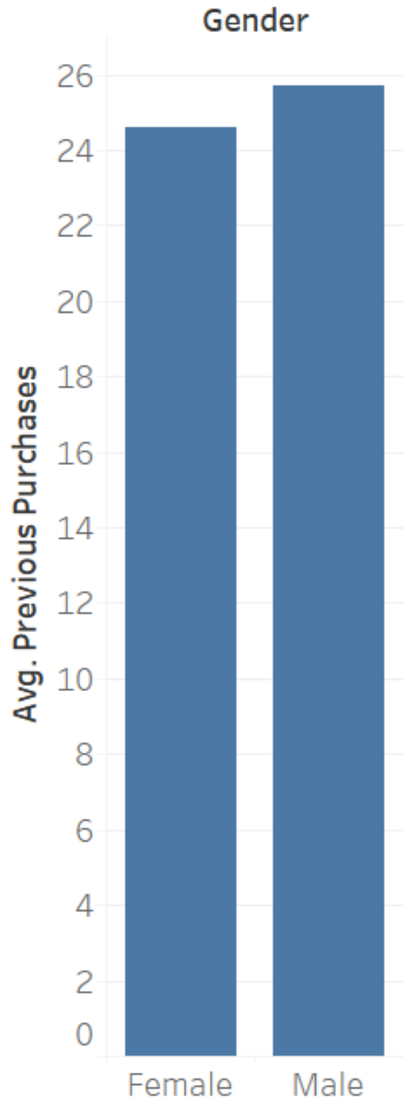
Item Purchased vs Category



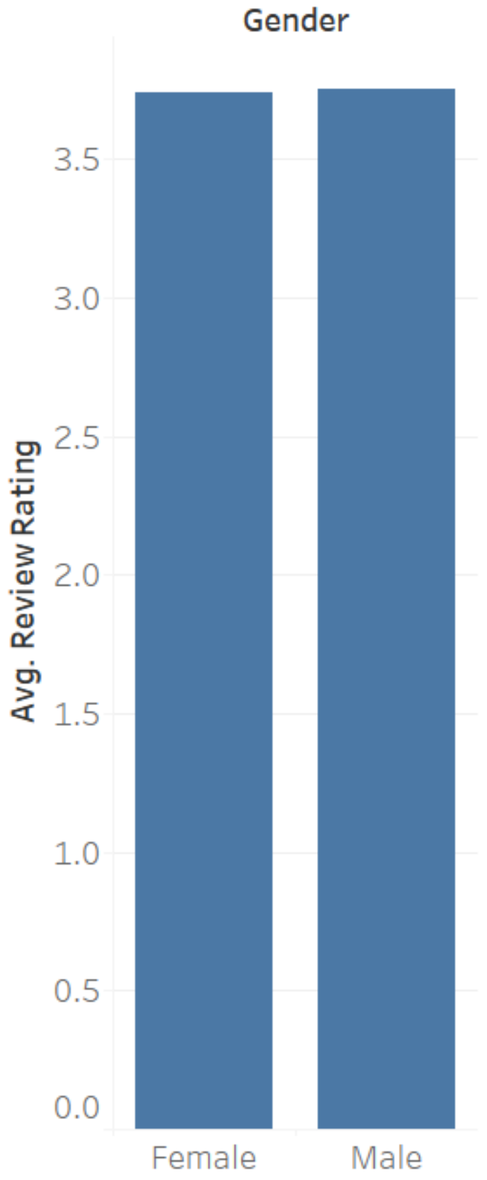
Season vs Purchase Amt



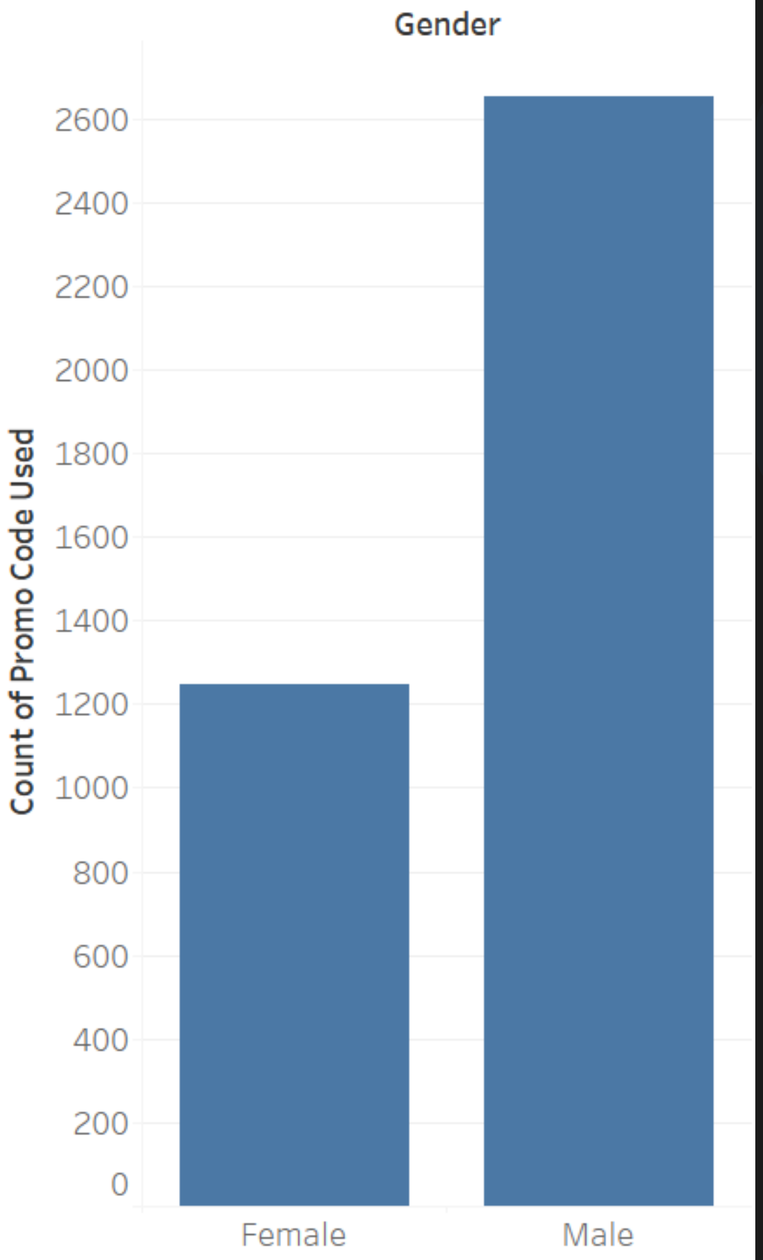
Gender vs Avg Previous Purchases



Gender vs Review Rating

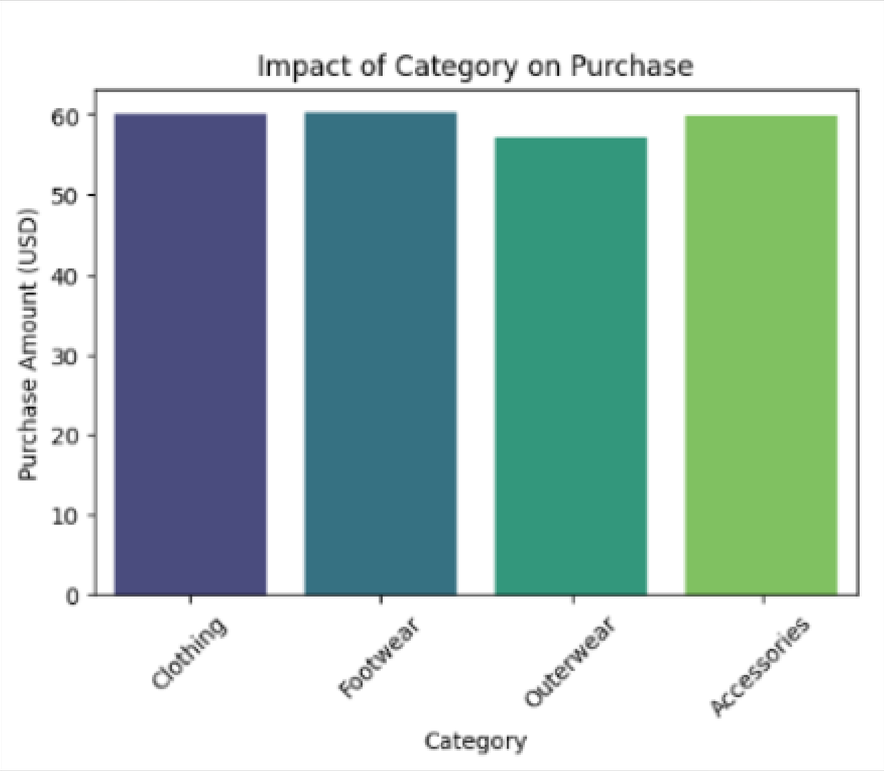


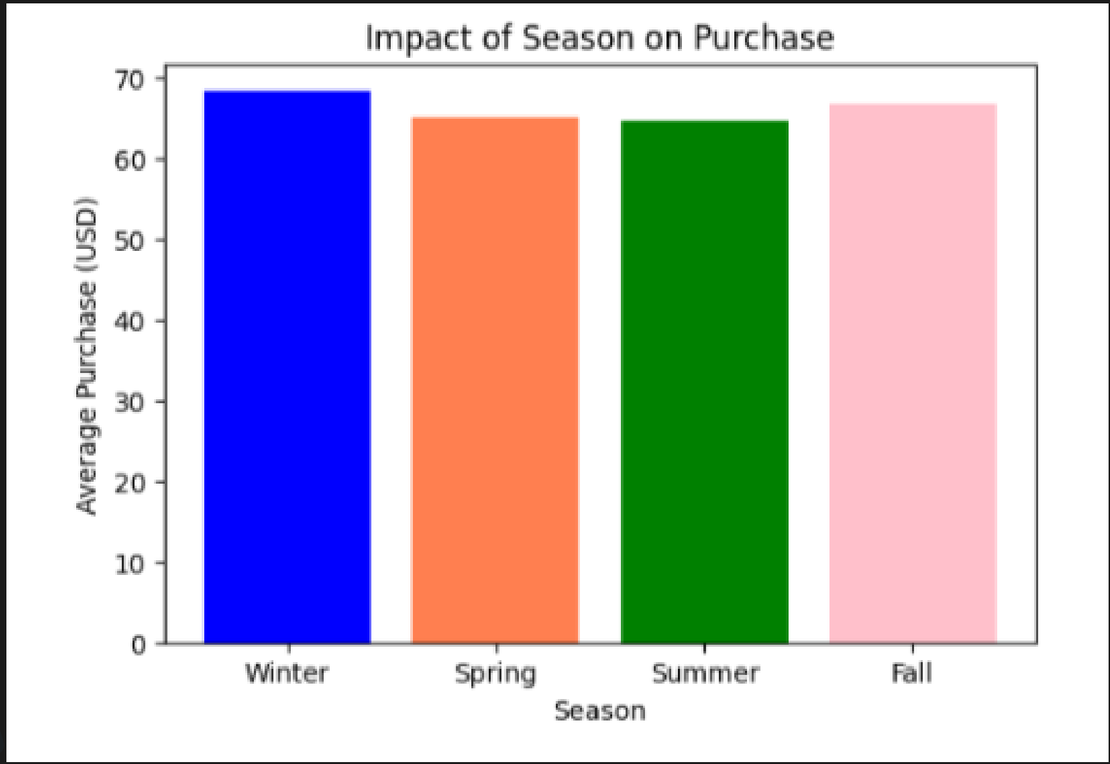
Gender vs Promo Code Used

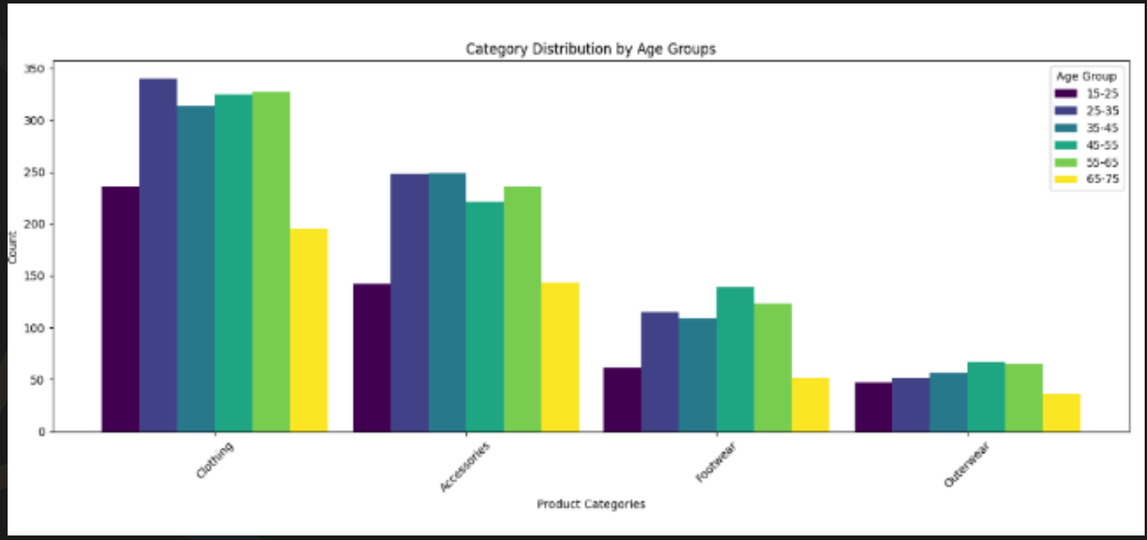


EDA FROM KAGGLE

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

THANK YOU FOR LISTENING!