Homework 1

Decker Mecham

Introduction

We are using some clothing store data to help the company predict how much their customers will spend with them per year. The data will be useful to clothing stores and will help them make a bigger profit. It will help them target customers better too. The following factors will help us find the yearly spent.

gender: self-disclosed gender identity, male, female, nonbinary or other age: current age of customer height_cm: self-reported height converted to centimeters waist_size_cm: self-reported waist size converted to centimeters inseam_cm: self-reported inseam (measurement from crotch of pants to floor) converted to centimeters test_group: whether or not the customer is in an experimental test group that gets special coupons once a month. 0 for no, 1 for yes. salary_self_report_in_k: self-reported salary of customer, in thousands months_active: number of months customer has been part of the clothing store's preferred rewards program num_purchases: the number of purchases the customer has made (a purchase is a single transaction that could include multiple items)

variables above help solve for below.

amount_spent_annual: the average amount the customer has spent at the store per year

Methods

We performed a linear regression as well as a polynomial regression on the data.

This helped us find the average amount spent annually by each customer. And allowed us to evaluate each variable for its impact on the resulting conclusion / money customer spent at store per year.

Pre processing included checking the null inputs as well as dropping missing / null values.

Results

For Linear Regression our model was unfit as the R^2 was 0.430579278748477

For polynomial Regression our model was fit as the R^2 was 0.7966808675876813

And overfit model would be closer to 1.00 for r^2.

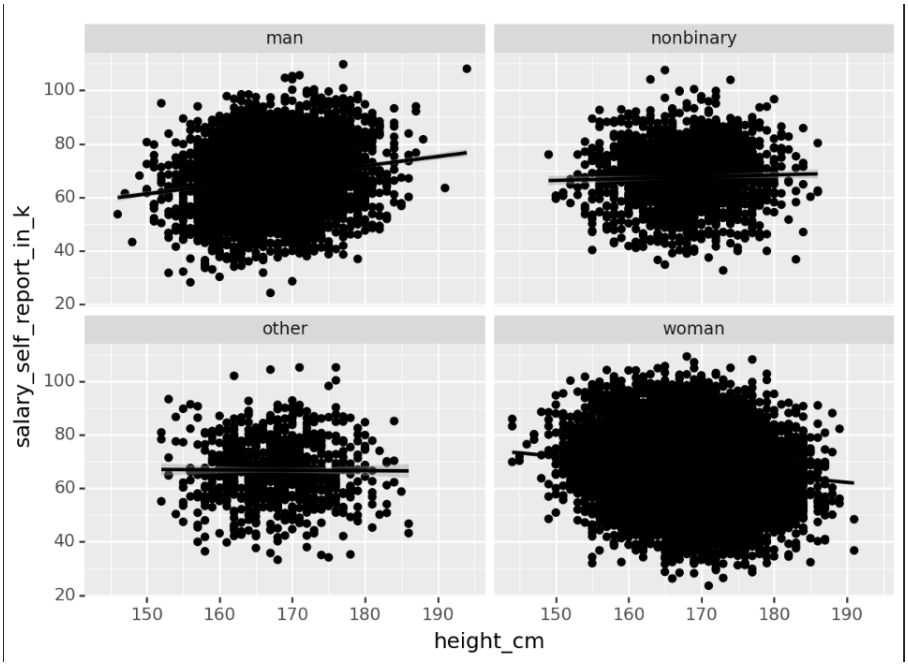

Figure 1: As we can see in our first model. The taller the male is the more he is paid, by looking at the line of fit. For Girls it is the opposite, the taller they are, the less they are paid. We see this because of the negative slope of the line of fit. For non binary an other, the data shows it does not matter much how tall they are. They get paid the same no matter what.

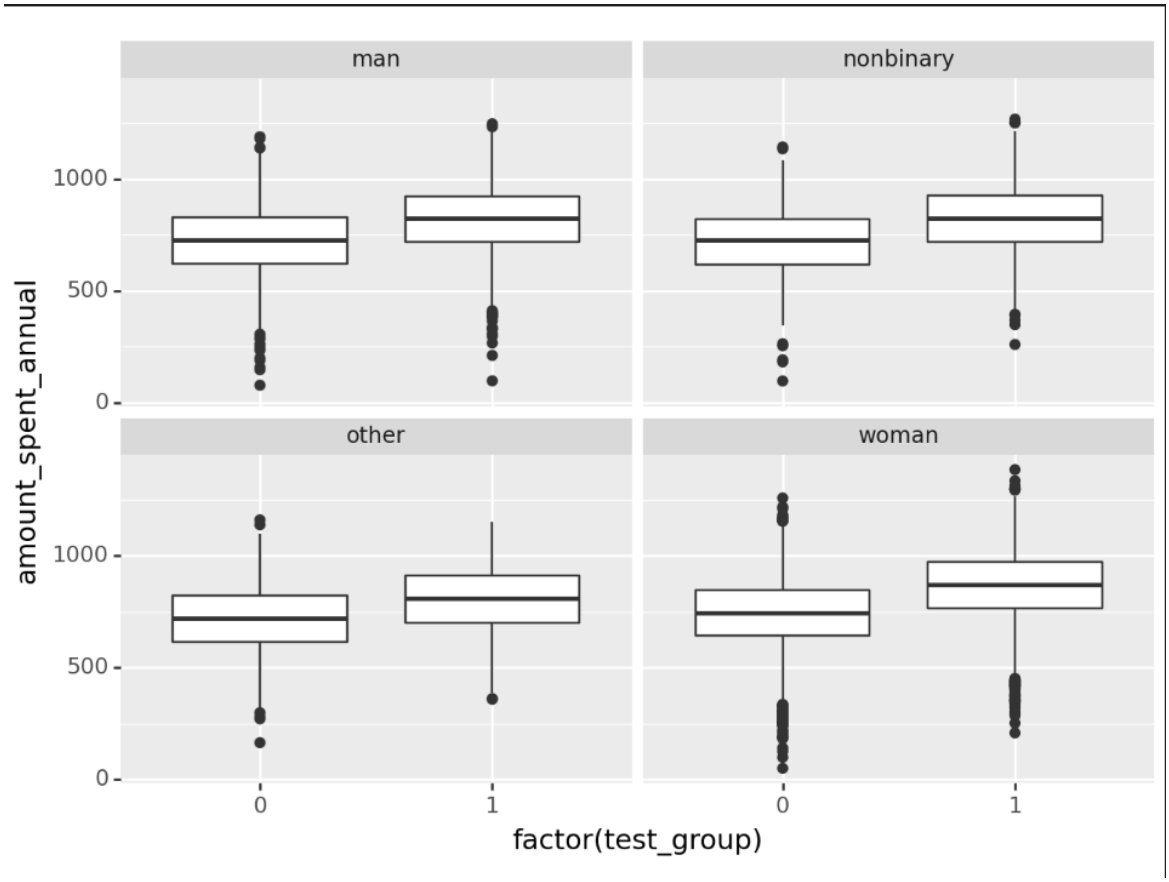

Figure 2: Woman spent the most annualy on average, followed by men, non binary, and lastly other.

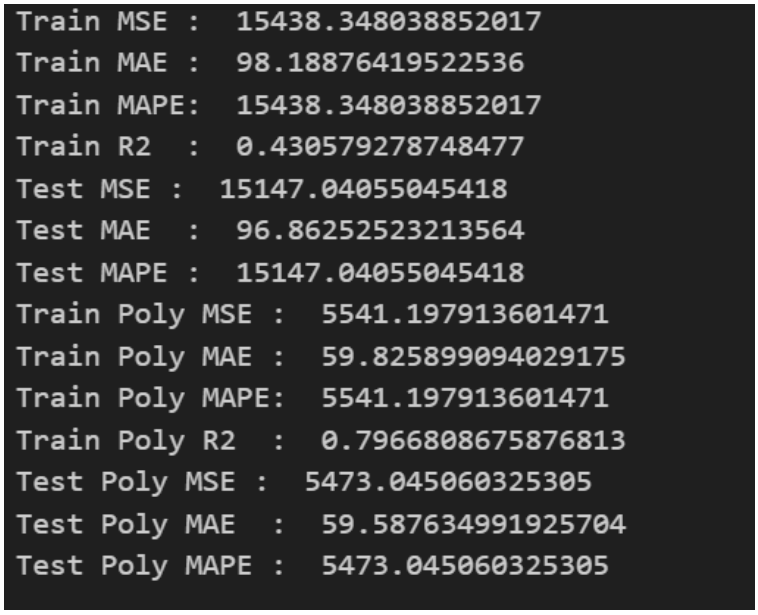

Figure 3: Regression Values.