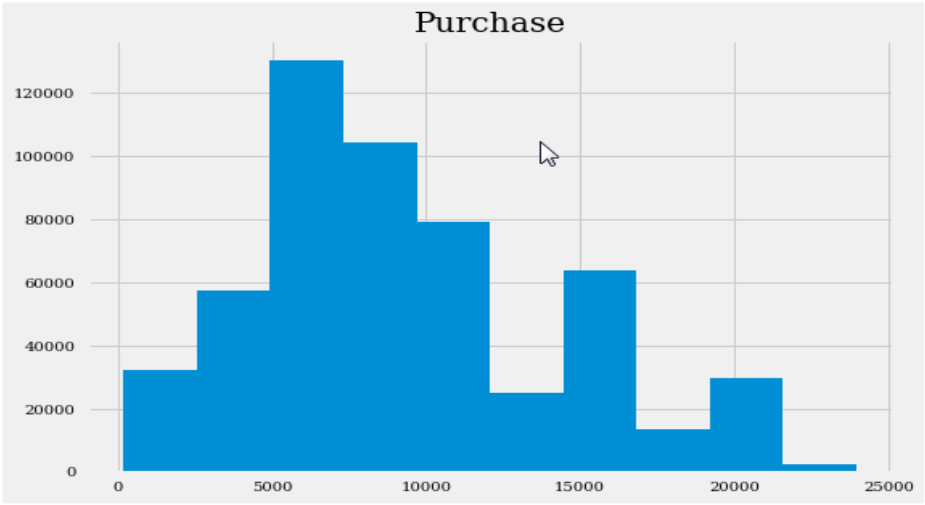
Data Story

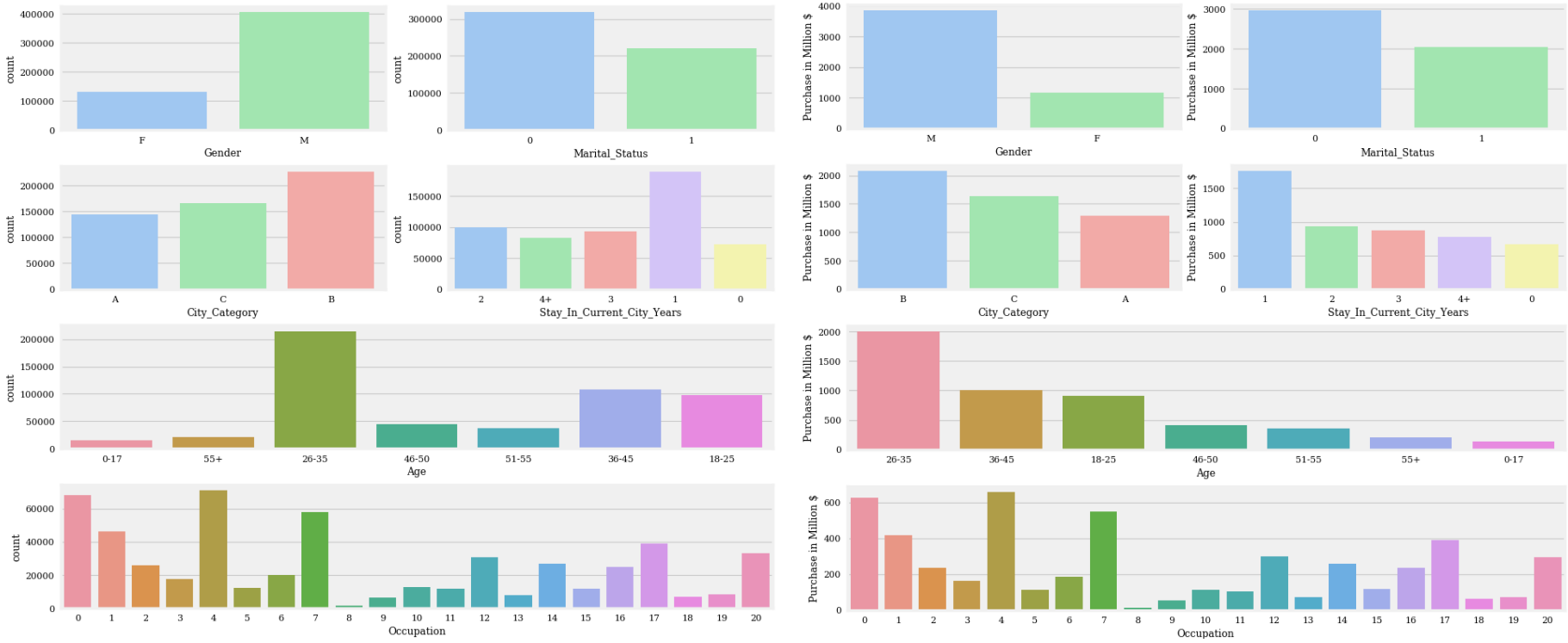
Objective is to find interesting relationships in the data and investigate hypothesis.

Let's look at histogram of purchase amounts.



By looking at the Histogram, it can be inferred that most of the purchases are between $3,000 and $12,000 and around $16,000 mark.

Looking at count plots of number of items purchased and the purchase amount, it can be inferred that they are not very different from each other. We cannot see any variable disproportionately affecting the purchase. So let's form hypothesis and investigate more.



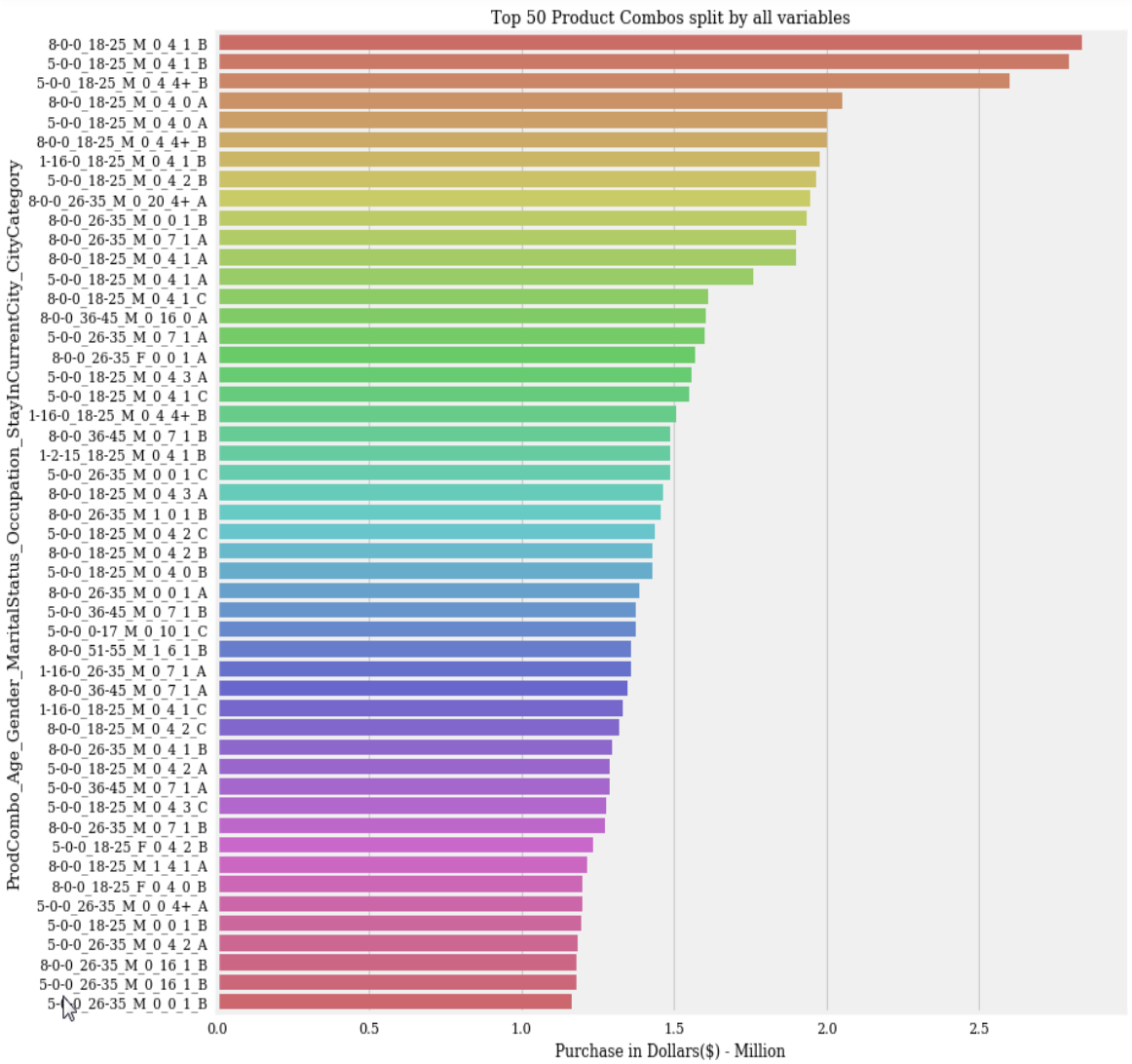
1. **What are the characteristics of my top revenue generators or who is my target audience?**
   Driven by data of top 50 combos, it can be inferred that:

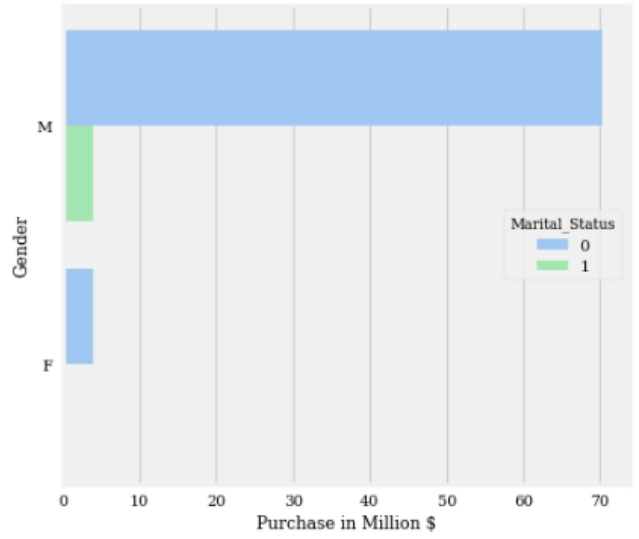
a) Single males buy the most

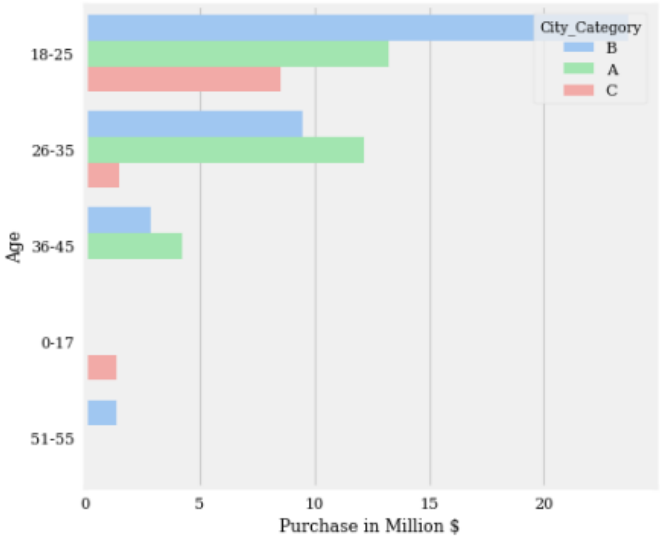
b) These single males are mainly from city B, in age group of 18-35. They are followed by city A in age groups of 18-25 and 25-36.

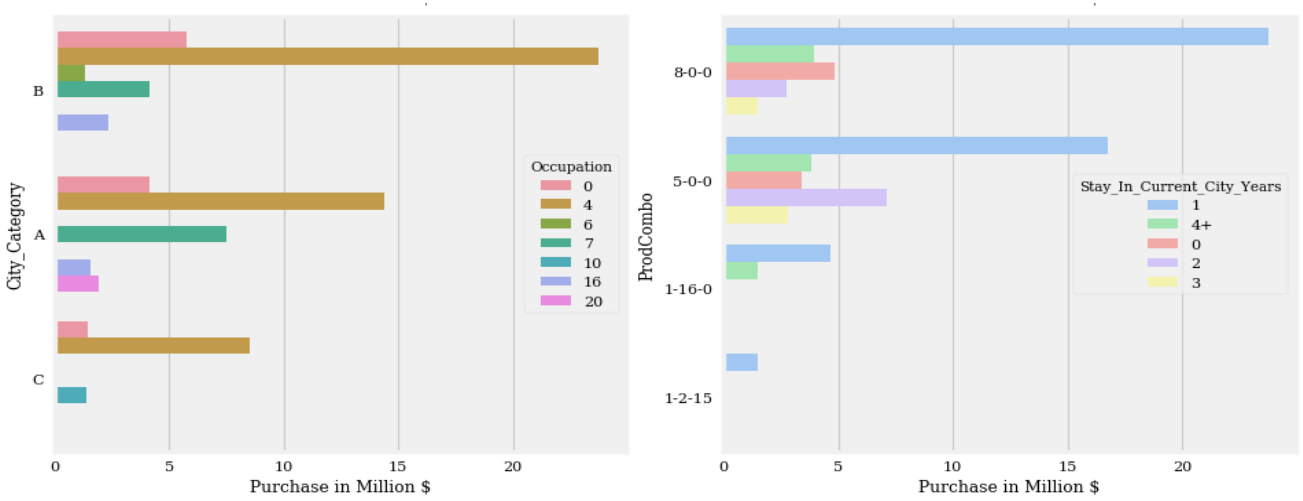
c) A disproportionate number of these people are in Occupation 4, mainly buying from city B and A.

d) People who mainly buy ProdCombos 8-0-0 and 5-0-0 have been living in the city around 1 year.





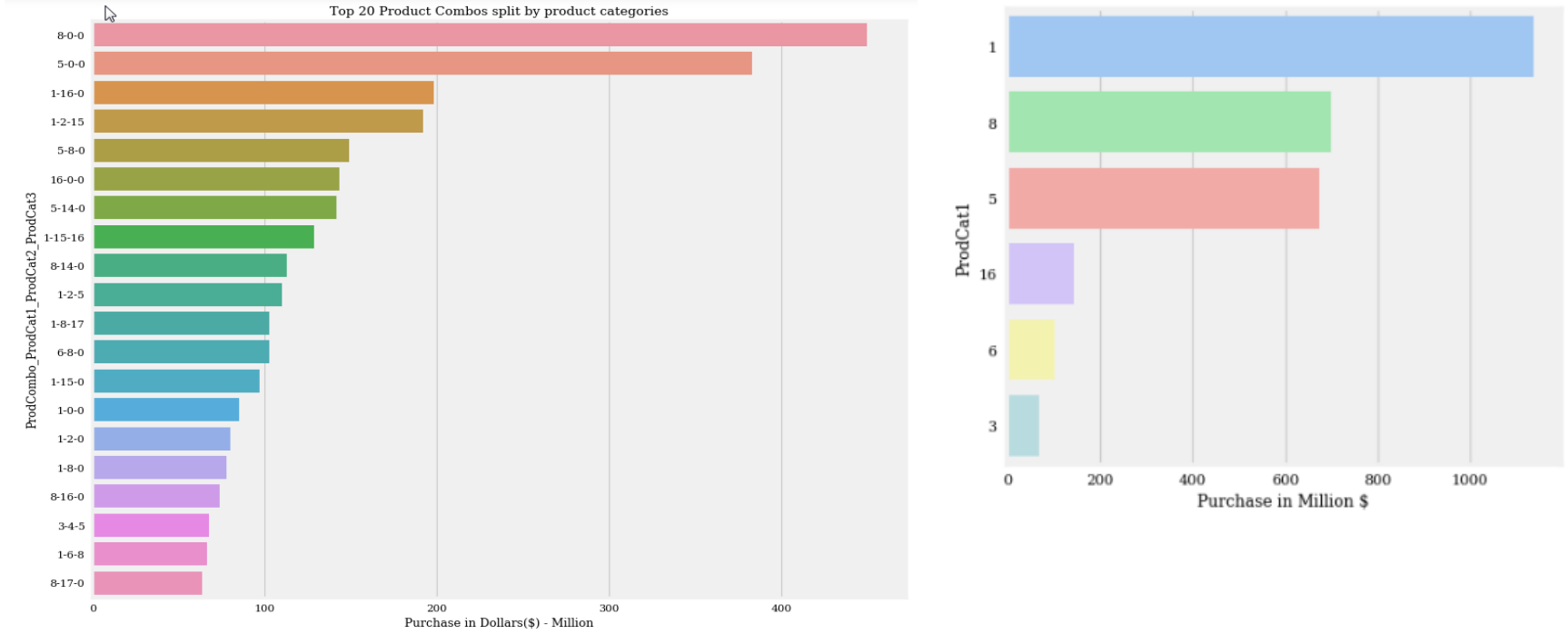


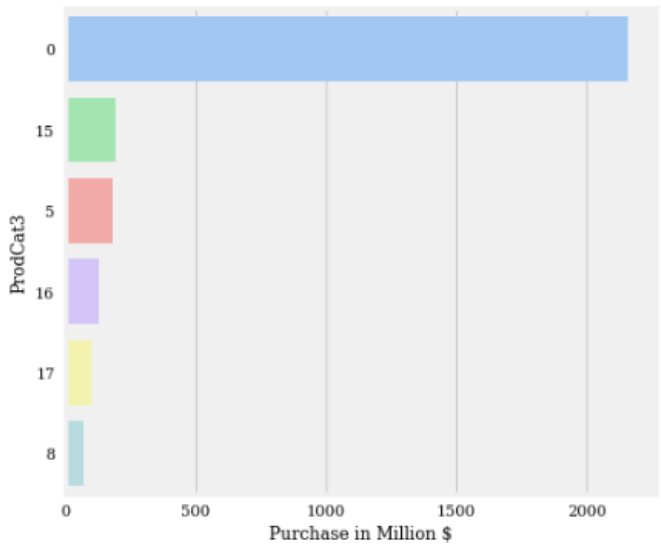


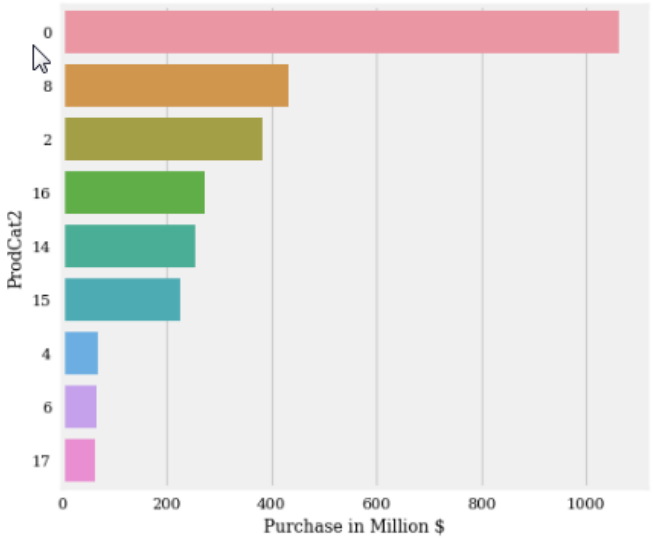
Answer: The target/valuable audience for Black Friday sale are single males from cities A and B in the age group of 18-35 in occupation 4 who have been living in their cities for around 1 year.

2. Who buys the top product combinations?

For this analysis, we restrict the top 50 to further of top 20 product combinations to get some meaningful conclusion. In the top 20 combos, you can now confirm that product 1 in Category 1 still has the most number of combos and generates the highest revenue, followed by 8 and 5. In terms of combos across all three categories, 1-2-15, 1-15-16 and 1-2-5 are frequently bought together.

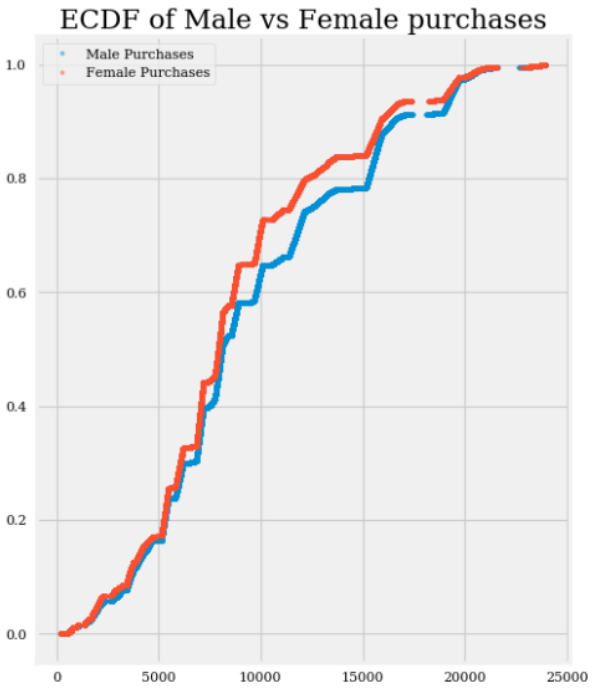


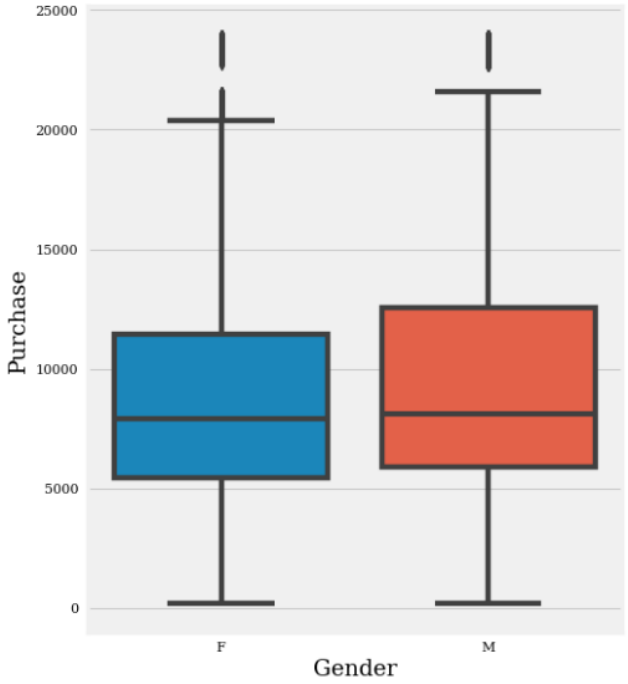




Overall, assuming that these 3 categories have independent products, the top products are:

a) Category 1: 1,8,5
b) Category 2: 8,2,16
c) Category 3: 15,5,16





Question 1: Do men buy more expensive things than women?

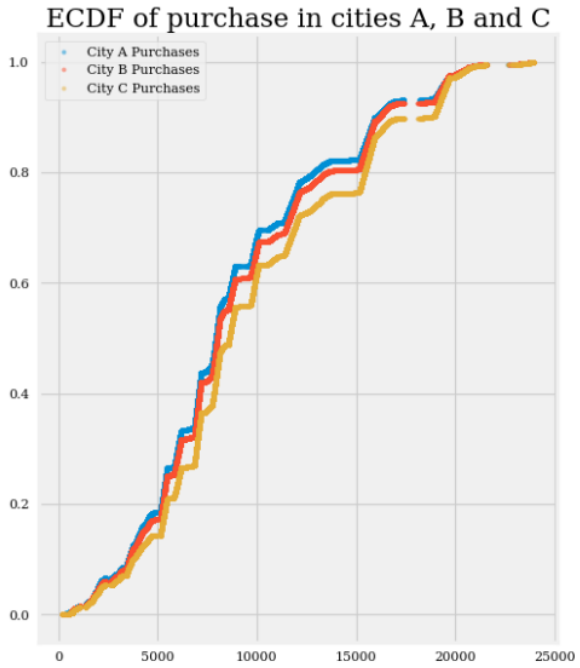
Expensive purchase = > $10,000

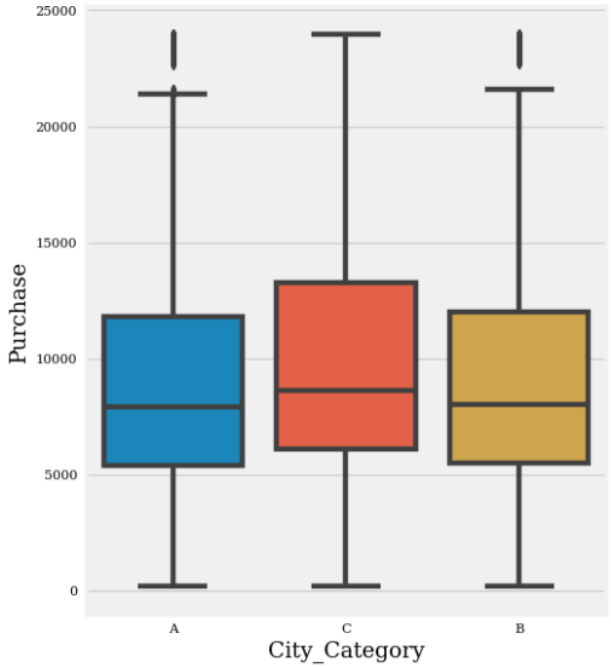
H0: The average purchase price by men is the same as women for expensive items.

H1: The average purchase price by men is more than women for expensive items.

Findings:

- Mean of men's purchase is: 15068.7163
- Mean of women's purchase is: 14931.9739
- Mean of total purchase is: 15082.73146804466
- The probability that average purchases by men is equal to that of women is 0.002
- This negates the null hypothesis. Meaning the average purchase price by men is more than women for expensive items and men do buy more expensive things than women.





Question 2: Do people shop equally in all three cities?
H0: The average purchase price by people in cities A, B and C is the same.
H1: The average purchase price by people in cities A, B and C is NOT the same.
Findings:

- The mean purchase value by people from city B is: 9366.509
- The mean purchase value by people from city C is: 9792.204
- The probability that average purchases from City A are equal to City C are: 0.001
- The probability that average purchases from City A are equal to City B are: 0.141
- The probability that average purchases from City B are equal to City C are: 0.028.

This negates the null hypothesis. Meaning, the average purchase price by people in cities A, B and C is NOT the same and they do not shop equally.