Please find the link attached for notebook view:

https://drive.google.com/file/d/1q6LOHnv8bv_BOHK8ccmW9ZBLFLuQSTYE/view?usp=sharing

Business Insights based on Non- Graphical and Visual Analysis

Comments on the range of attributes

1.

```
df.shape
```

```
(550068, 10)
```

```
M    414259
F    135809
Name: Gender, dtype: int64
```

2.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 550068 entries, 0 to 550067
Data columns (total 10 columns):
 #   Column                      Non-Null Count   Dtype
---  ------                      --------------   -----
 0   User_ID                     550068 non-null  int64
 1   Product_ID                  550068 non-null  object
 2   Gender                      550068 non-null  object
 3   Age                         550068 non-null  object
 4   Occupation                  550068 non-null  int64
 5   City_Category               550068 non-null  object
 6   Stay_In_Current_City_Years  550068 non-null  object
 7   Marital_Status              550068 non-null  int64
 8   Product_Category            550068 non-null  int64
 9   Purchase                    550068 non-null  int64
dtypes: int64(5), object(5)
memory usage: 42.0+ MB
```

3.

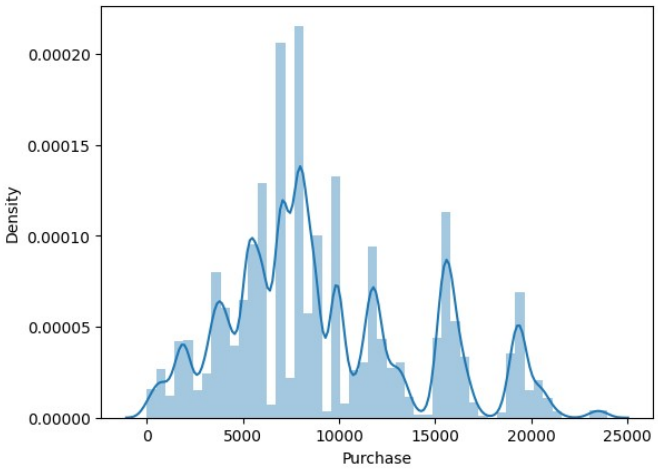
```
User_ID                       0
Product_ID                    0
Gender                        0
Age                           0
Occupation                    0
City_Category                 0
Stay_In_Current_City_Years    0
Marital_Status                0
Product_Category              0
Purchase                      0
dtype: int64
```

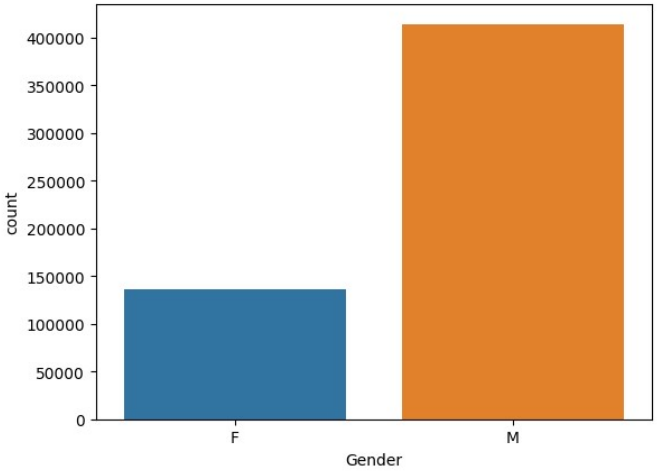
Comments on the distribution of the variables and relationship between them

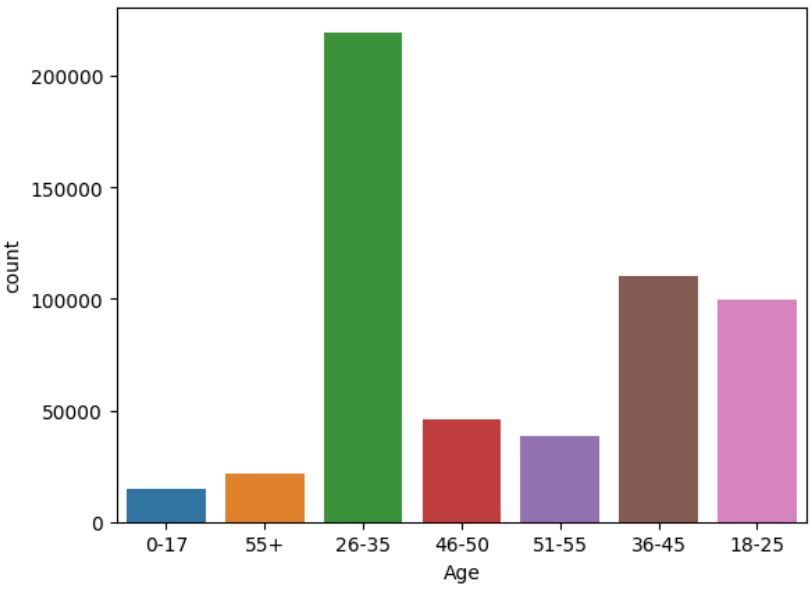
Insights:
--> looks like most of our purchases lie in range 5000 to 9000.

-->Number of products bought by mail is more than female.

-->Customers from age group 26-35 bought more products as compare to any other category.





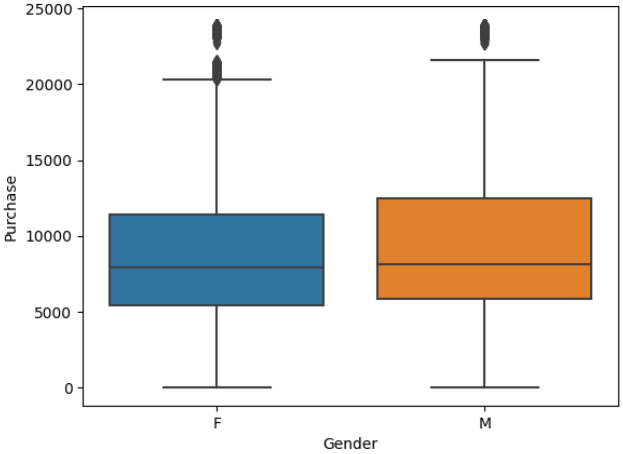


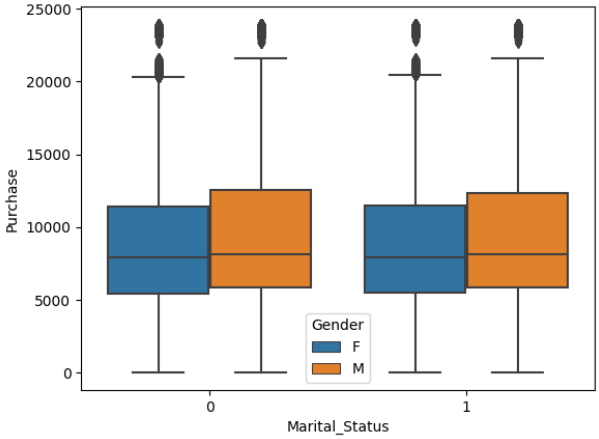
Comments for each univariate and bivariate plot

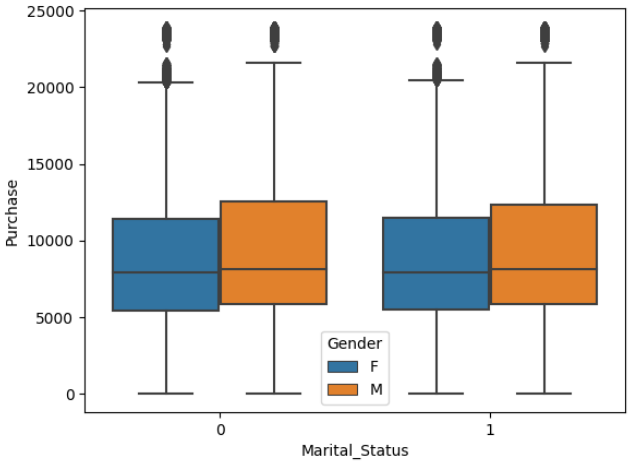
Insights:

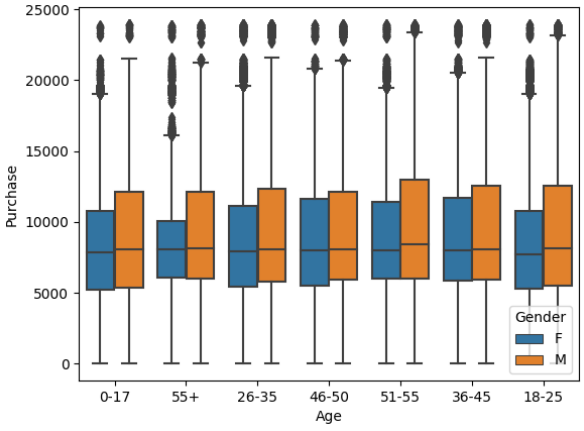
--> Doesn’t look much difference in male and female purchase amount, a little upper value for male on purchase but doesn’t signifies much.

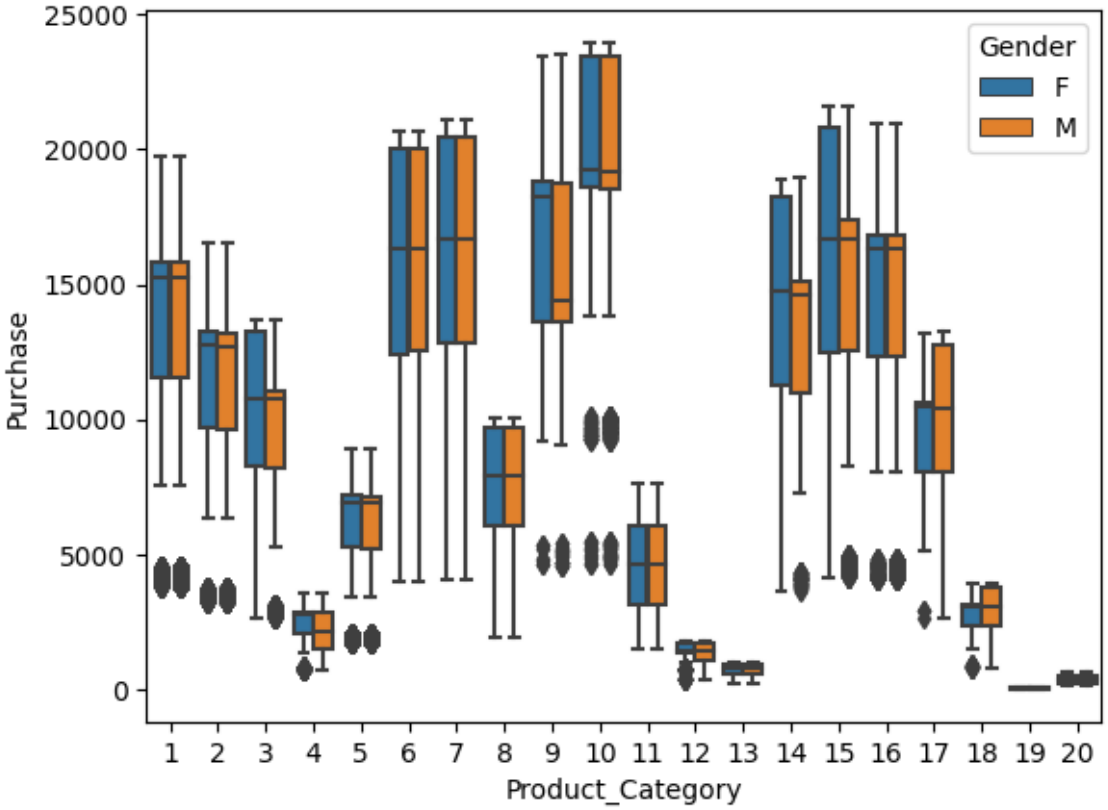
--> For product id 14 and 15 Female tends to buys more as compare to mens. So we specifically focuses on those 2 categories for females.

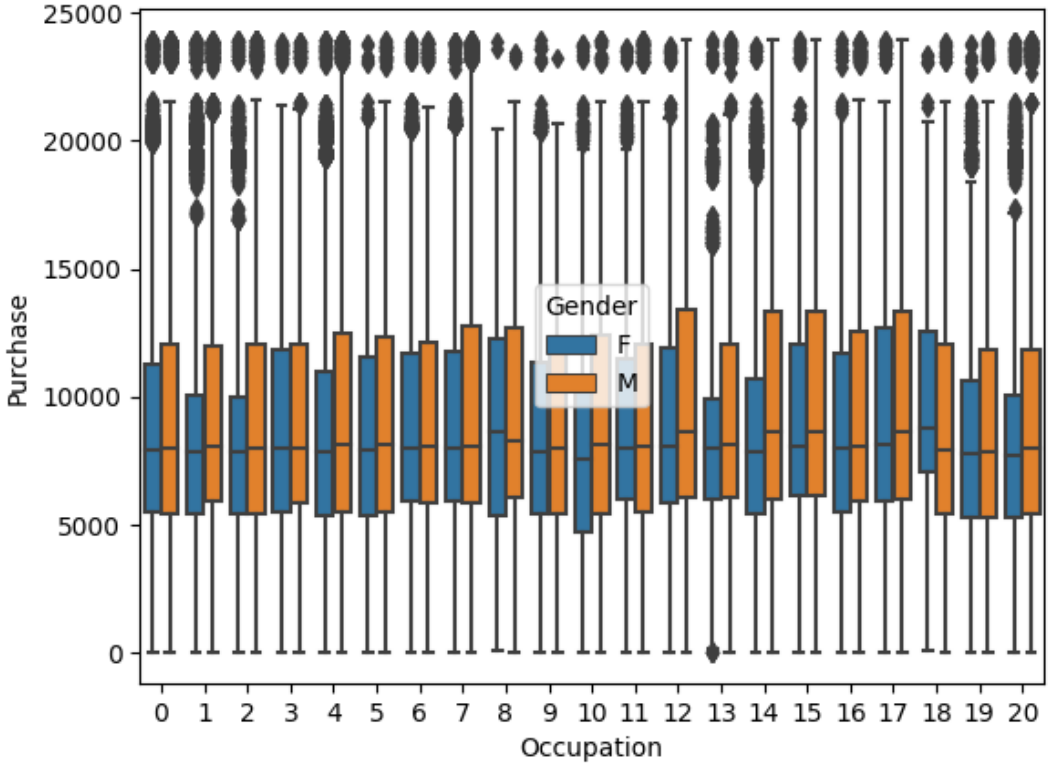









Gender
F
M
Purchase
25000
20000
15000
10000
5000
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Product_Category


Gender
F
M
Purchase
25000
20000
15000
10000
5000
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Occupation

1. Are women spending more money per transaction than men? Why or Why not?

Insights:
--> there is not clear evidence that women spending more than men as per data we received.

Gender	count	mean	std	min	25%	50%	75%	max
F	135809.0	8734.565765	4767.233289	12.0	5433.0	7914.0	11400.0	23959.0
M	414259.0	9437.526040	5092.186210	12.0	5863.0	8098.0	12454.0	23961.0

Gender	Marital_Status	count	mean	std	min	25%	50%	75%	max
F	0	78821.0	8679.845815	4740.048367	12.0	5417.00	7895.0	11370.0	23955.0
	1	56988.0	8810.249789	4803.594163	12.0	5456.75	7939.0	11451.0	23959.0
M	0	245910.0	9453.756740	5101.803346	12.0	5854.00	8101.0	12543.0	23961.0
	1	168349.0	9413.817605	5078.027482	12.0	5874.00	8094.0	12312.0	23961.0

2. Confidence intervals and distribution of the mean of the expenses by female and male customers

Insights:
-With sample size of 5000, it clearly indicates that male spends more while purchasing as compare to females.

```
----- 90 %-------
 Male confidence interval: (9125.58195554966, 9649.28644445034)
 Female confidence interval: (8492.580271941353, 8991.467728058646)
----- 95 %-------
 Male confidence interval: (9075.41800598471, 9699.45039401529)
 Female confidence interval: (8444.793465225628, 9039.25453477437)
----- 99 %-------
 Male confidence interval: (8977.375412151921, 9797.492987848078)
 Female confidence interval: (8351.396862246238, 9132.65113775376)
```

3. Are confidence intervals of average male and female spending overlapping? How can Walmart leverage this conclusion to make changes or improvements?

Insights:
As per the results received if we fetch the confidence intervals of male and female, then they don’t overlap and states that there is difference in spends by male and female.

4. Results when the same activity is performed for Married vs Unmarried.

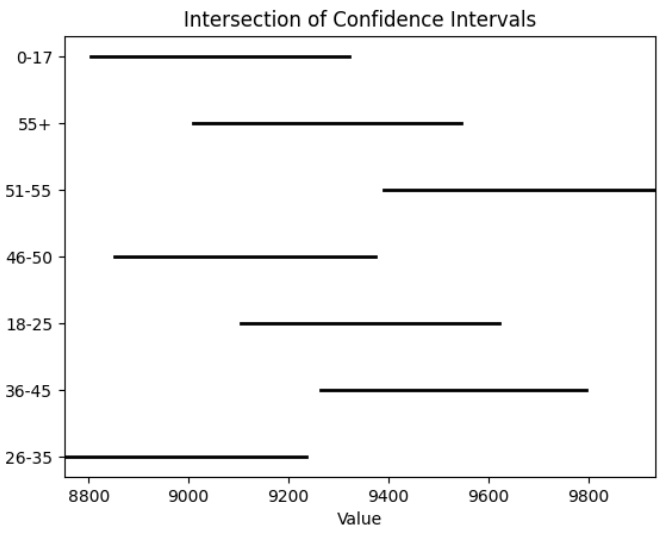
Insights:
Looks like an overlapping in Single Vs Married, hence there is no suffiecient proof to say that there is any purchase disparity.

```
----- 90 %-------
 Single confidence interval: (9004.40605056871, 9526.398749431291)
 Married confidence interval: (9126.084208461514, 9654.566191538484)
----- 95 %-------
 Single confidence interval: (8954.406067803659, 9576.398732196343)
 Married confidence interval: (9075.462638268928, 9705.187761731071)
----- 99 %-------
 Single confidence interval: (8856.683937780117, 9674.120862219885)
 Married confidence interval: (8976.525650878339, 9804.12474912166)
```

5. Results when the same activity is performed for Age.

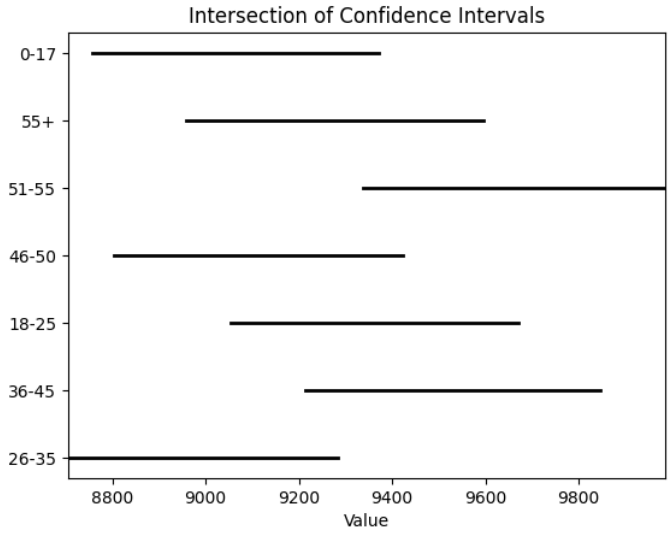
Insights:
--> when we take a little close look, there is a difference between 51-55 Vs (0-17, 26-35) and also 36-45 Vs 26-35 spend pattern with 90% significance value.

```
----- 90 %-------
26-35 confidence interval: 8751.521389002177 9236.452610997821
36-45 confidence interval: 9264.992229647985 9795.049770352016
18-25 confidence interval: 9104.379141438014 9621.422858561986
46-50 confidence interval: 8852.999383766888 9374.08461623311
51-55 confidence interval: 9390.18689980946 9933.16710019054
55+ confidence interval: 9009.171814696234 9544.888185303767
0-17 confidence interval: 8805.334479961088 9320.763520038914
```

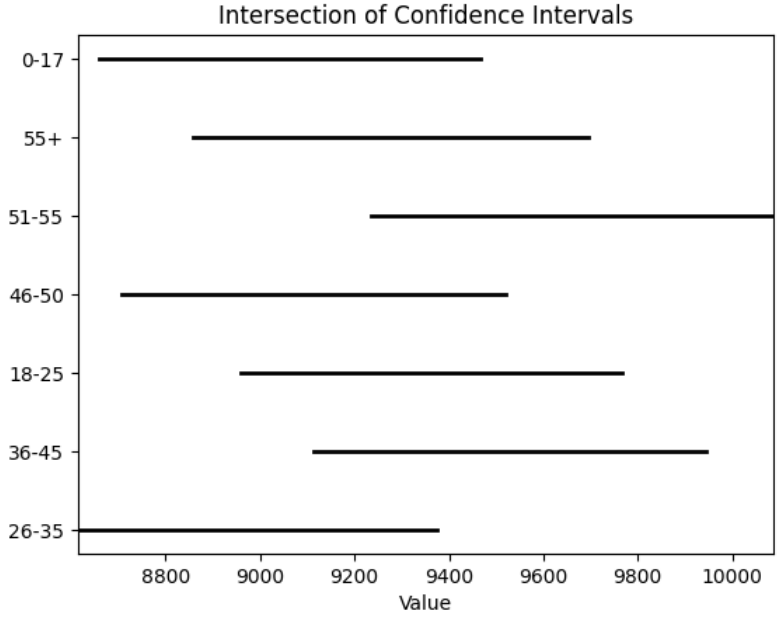


Insights:
--> with 95% significance value there is a different in purchasing pattern between age group 26-35 Vs 51-55
--> with 99% significance value no purchase dispairity seen.

```
----- 95 %-------
26-35 confidence interval: 8705.071404558234 9282.902595441765
36-45 confidence interval: 9214.219741915038 9845.822258084963
18-25 confidence interval: 9054.853205536403 9670.948794463597
46-50 confidence interval: 8803.086324256272 9423.997675743727
51-55 confidence interval: 9338.176592538925 9985.177407461075
55+ confidence interval: 8957.857286053018 9596.202713946983
0-17 confidence interval: 8755.963208721749 9370.134791278253
```




```
----- 99 %-------
26-35 confidence interval: 8614.287544876686 9373.686455123312
36-45 confidence interval: 9114.987794752738 9945.054205247263
18-25 confidence interval: 8958.057573216753 9767.744426783247
46-50 confidence interval: 8705.534080802734 9521.549919197265
51-55 confidence interval: 9236.525397306752 10086.828602693247
55+ confidence interval: 8857.565950679318 9696.494049320683
0-17 confidence interval: 8659.469859711027 9466.628140288974
```



Recommendations

- •Actionable items for business. No technical jargon. No complications. Simple action items that everyone can understand

1. Purchase Trends:

a. There is difference in purchase between male and female, male tends to have more purchase value as compare to females. We have to do market analysis and focus more on products that females tends to buys so that we can improve our female purchase as well.

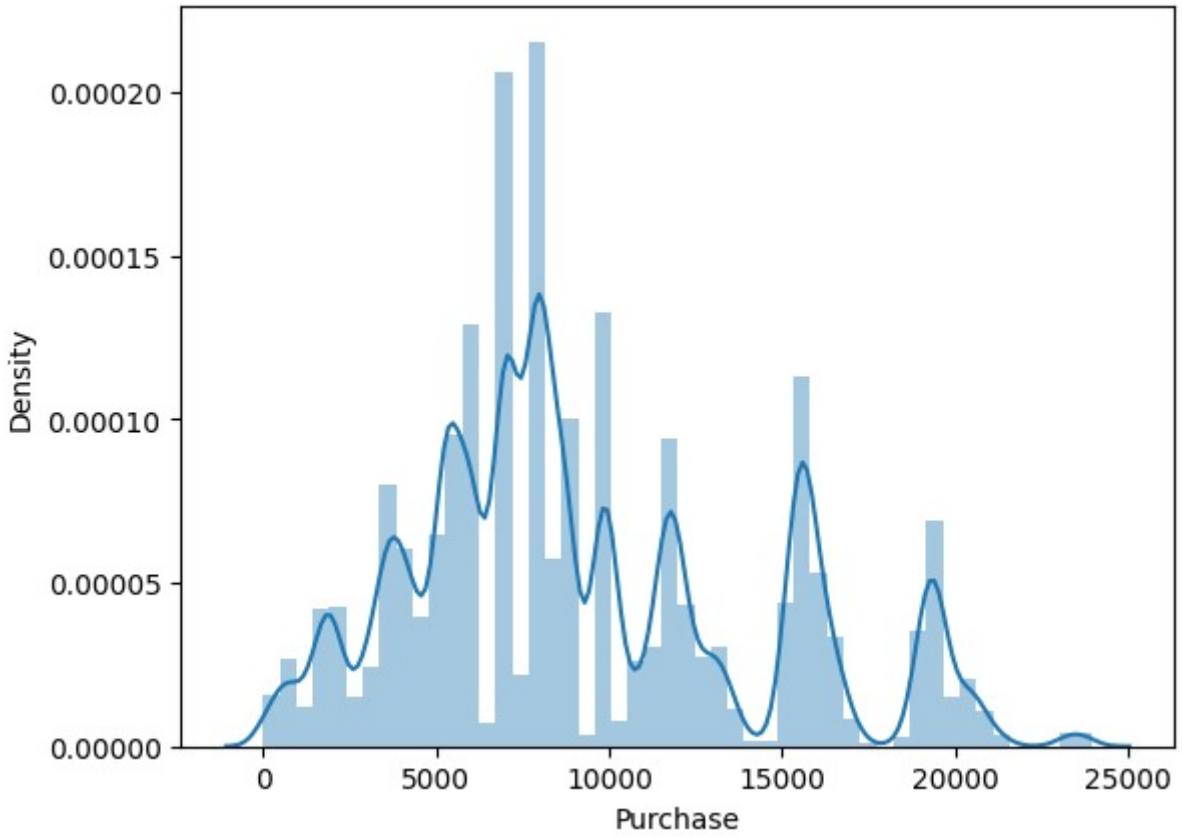
b. No difference based on Marital Status, that means our bussiness is unbiased and no impact on marital status of our customers

c. No difference or disparity in purchase with different age group, which is positive for our bussiness. We are catering to all king of generations and keep focusing on building our bussiness for all generations.
But yes few of the age groups have purchase disparity which we have to understand and make our inventory available for that age group as well especifically for age group 26-35.

3. For few of the product categories attached below, we see females tends to buy more which give us an important input on their liking. We should keep focusing on such kind of item and make our inventory more strong to improve female interection with our brand.

Gender	F	M	All
Product_Category			
1	0.182838	0.278925	0.255201
2	0.041661	0.043948	0.043384
3	0.044224	0.034295	0.036746
4	0.026795	0.019587	0.021366
5	0.308971	0.263053	0.274390
6	0.033569	0.038399	0.037206
7	0.006944	0.006706	0.006765
8	0.247097	0.194002	0.207111
9	0.000515	0.000821	0.000745
10	0.008556	0.009566	0.009317
11	0.034895	0.047188	0.044153
12	0.011281	0.005830	0.007175
13	0.010765	0.009866	0.010088
14	0.004587	0.002173	0.002769
15	0.007702	0.012659	0.011435
16	0.017687	0.017926	0.017867
17	0.000457	0.001246	0.001051
18	0.002813	0.006621	0.005681
19	0.003321	0.002781	0.002914
20	0.005324	0.004410	0.004636

4. Our main purchase sales ranges from 6000-9000, we have to make sure to cater this range of product to our users and keep the inventory in good shape for the given range.



END