Black Friday Sales

Dataset: <https://www.kaggle.com/mehdidag/black-friday>

Data Fields:

User\_ID - User ID

Product\_ID - Product ID

Gender - Sex of User

Age - Age in bins

Occupation - Occupation (Masked)

City\_Category - Category of the City (A,B,C)

Stay\_In\_Current\_City\_Years - Number of years stay in current city

Marital\_Status - Marital Status

Product\_Category\_1 - Product Category (Masked)

Product\_Category\_2 - Product may belongs to other category also (Masked)

Product\_Category\_3 - Product may belongs to other category also (Masked)

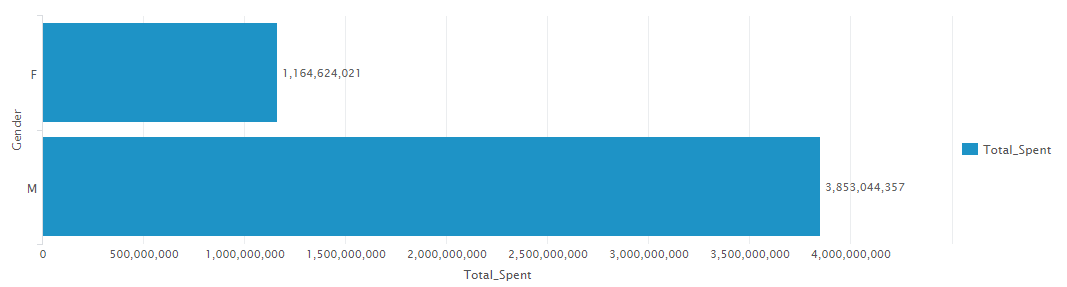
Purchase - Purchase amount in dollars

1. Find out whether Women are more likely to spend in the Black Friday more than Men

Answer: No **men** spent than women as shown below.

| [Gender](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20sum(Purchase)%20as%20Total_Spent%20by%20Gender&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563602600.384) | [Total\_Spent](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20sum(Purchase)%20as%20Total_Spent%20by%20Gender&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563602600.384) |
| --- | --- |
| F | 1164624021 |
| M | 3853044357 |

Search String: source="blackfriday.csv" | stats sum(Purchase) as Total\_Spent by Gender

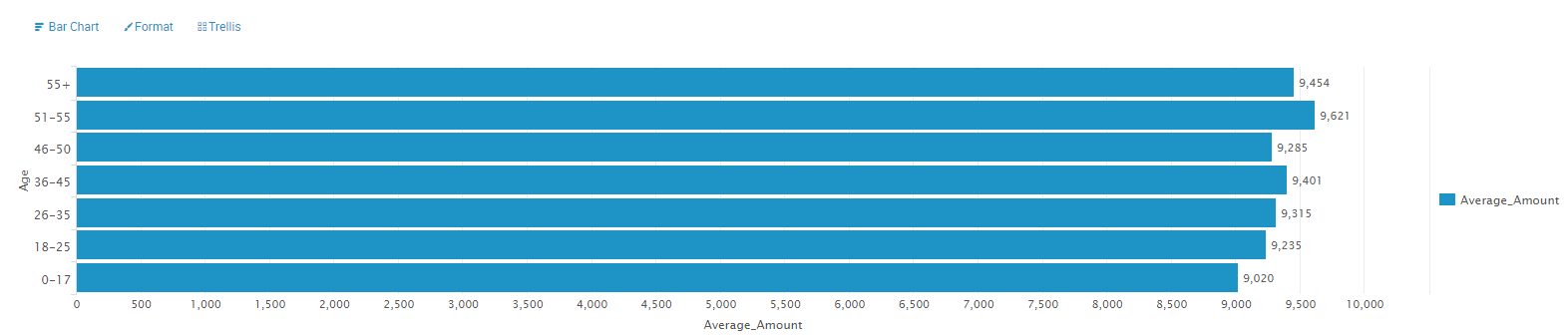


1. Calculate the average amount spent with respect to each age group

Answer: Find below the average per age group

| [Age](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20avg(Purchase)%20as%20Average_Amount%20by%20Age%20%7C%20eval%20Average_Amount%3Dround(Average_Amount)%20%7C%20sort%20-Age&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563602395.383) | [Average\_Amount](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20avg(Purchase)%20as%20Average_Amount%20by%20Age%20%7C%20eval%20Average_Amount%3Dround(Average_Amount)%20%7C%20sort%20-Age&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563602395.383) |
| --- | --- |
| 55+ | 9454 |
| 51-55 | 9621 |
| 46-50 | 9285 |
| 36-45 | 9401 |
| 26-35 | 9315 |
| 18-25 | 9235 |
| 0-17 | 9020 |

Search String: source="blackfriday.csv" | stats avg(Purchase) as Average\_Amount by Age | eval Average\_Amount=round(Average\_Amount) | sort -Age



1. Find out whether Women who are married and have the age between 26-35 are more likely to spend money in Black Friday than the Men who are also married and have the same age's range.

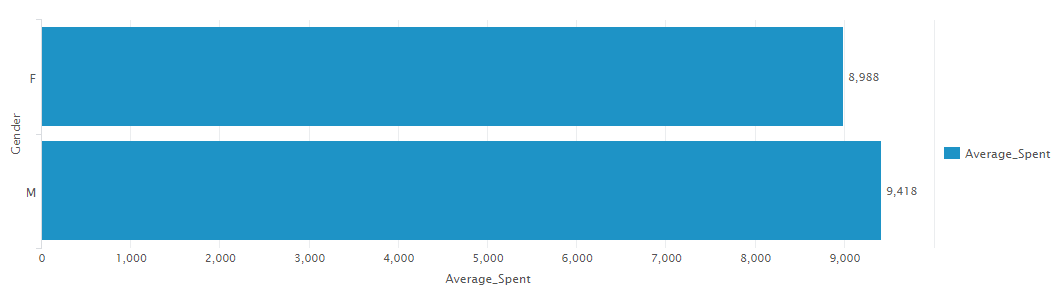
Answer: The likelihood to shop more can be derived by either the tendency to spend more i.e. average amount spent, or overall how many people from each category actually do shopping. On both the accounts it seems that married Men in the age group by 26-35 shop more than women.

So the answer is **Men**.

### By Average Amount Spent:

| [Gender](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20Marital_Status%3D1%20Age%3D26-35%20%7C%20stats%20avg(Purchase)%20as%20Average_Spent%20by%20Gender%20%7C%20eval%20Average_Spent%3Dround(Average_Spent)&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563603198.401) | [Average\_Spent](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20Marital_Status%3D1%20Age%3D26-35%20%7C%20stats%20avg(Purchase)%20as%20Average_Spent%20by%20Gender%20%7C%20eval%20Average_Spent%3Dround(Average_Spent)&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563603198.401) |
| --- | --- |
| F | 8988 |
| M | 9418 |

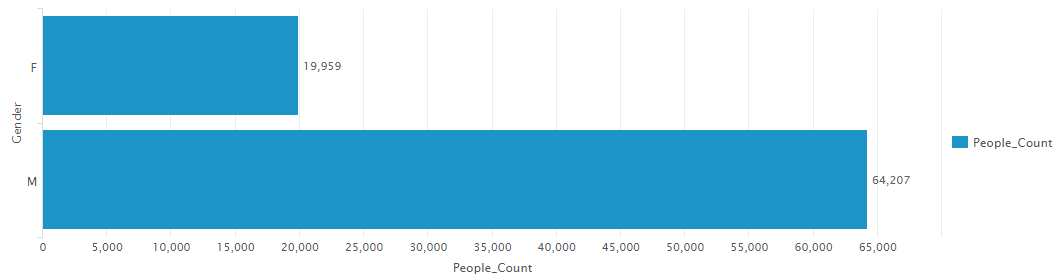
Search String: source="blackfriday.csv" Marital\_Status=1 Age=26-35 | stats avg(Purchase) as Average\_Spent by Gender | eval Average\_Spent=round(Average\_Spent)



### By Number of People:

| [Gender](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20Marital_Status%3D1%20Age%3D26-35%20%7C%20stats%20count(Purchase)%20as%20People_Count%20by%20Gender&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563603324.402) | [People\_Count](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20Marital_Status%3D1%20Age%3D26-35%20%7C%20stats%20count(Purchase)%20as%20People_Count%20by%20Gender&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563603324.402) |
| --- | --- |
| F | 19959 |
| M | 64207 |

Search String: source="blackfriday.csv" Marital\_Status=1 Age=26-35 | stats count(Purchase) as People\_Count by Gender

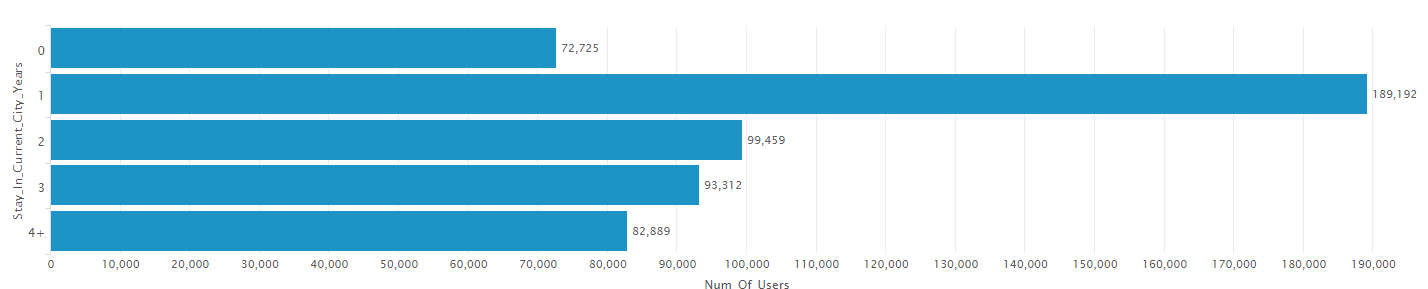


1. Compare the difference between the amount of years that customers lives in that particular city

Answer: Question is a little ambiguous so I have assumed we are looking for the comparison of users durations of stay

| [Stay\_In\_Current\_City\_Years](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20count(User_ID)%20as%20Num_Of_Users%20by%20Stay_In_Current_City_Years&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563604837.420) | [Num\_Of\_Users](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20count(User_ID)%20as%20Num_Of_Users%20by%20Stay_In_Current_City_Years&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563604837.420) |
| --- | --- |
| 0 | 72725 |
| 1 | 189192 |
| 2 | 99459 |
| 3 | 93312 |
| 4+ | 82889 |

Search String: source="blackfriday.csv" | stats count(User\_ID) as Num\_Of\_Users by Stay\_In\_Current\_City\_Years



If we were looking to also get the breakup by the City category because the question mentions ‘particular city’ then the above query will be modified as below.

Search String 2: source="blackfriday.csv" | stats count(User\_ID) as Num\_Of\_Users by Stay\_In\_Current\_City\_Years, City\_Category

| [Stay\_In\_Current\_City\_Years](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20count(User_ID)%20as%20Num_Of_Users%20by%20Stay_In_Current_City_Years%2C%20City_Category&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563605057.421) | [City\_Category](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20count(User_ID)%20as%20Num_Of_Users%20by%20Stay_In_Current_City_Years%2C%20City_Category&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563605057.421) | [Num\_Of\_Users](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20count(User_ID)%20as%20Num_Of_Users%20by%20Stay_In_Current_City_Years%2C%20City_Category&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563605057.421) |
| --- | --- | --- |
| 0 | A | 23700 |
| 0 | B | 28143 |
| 0 | C | 20882 |
| 1 | A | 48160 |
| 1 | B | 81622 |
| 1 | C | 59410 |
| 2 | A | 26548 |
| 2 | B | 40800 |
| 2 | C | 32111 |
| 3 | A | 24389 |
| 3 | B | 41964 |
| 3 | C | 26959 |
| 4+ | A | 21841 |
| 4+ | B | 33964 |
| 4+ | C | 27084 |

1. Which type of products are common among men and among women?

Answer: Product\_Category\_1 is used to determine the product category type. There are 18 category types and the objective was to check if both Male and Female genders have purchased product the categories. As it turns out we can establish that both genders have shopped in all 18 categories.

| [Product\_Category\_1](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20count(User_ID)%20as%20cnt_user%20by%20Product_Category_1%2C%20Gender%20%7C%20stats%20count(cnt_user)%20as%20common%20by%20Product_Category_1%20%7C%20sort%20-Product_Category_1%20%7C%20reverse%20%7C%20where%20common%3D2&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563604183.414) | [common](https://prd-p-rdvv7p5cpdr3.cloud.splunk.com/en-GB/app/search/search?q=search%20source%3D%22blackfriday.csv%22%20%7C%20stats%20count(User_ID)%20as%20cnt_user%20by%20Product_Category_1%2C%20Gender%20%7C%20stats%20count(cnt_user)%20as%20common%20by%20Product_Category_1%20%7C%20sort%20-Product_Category_1%20%7C%20reverse%20%7C%20where%20common%3D2&earliest=0&latest=&display.page.search.mode=smart&dispatch.sample_ratio=1&display.general.type=statistics&display.page.search.tab=statistics&display.visualizations.charting.chart.showDataLabels=all&sid=1563604183.414) |
| --- | --- |
| 1 | 2 |
| 2 | 2 |
| 3 | 2 |
| 4 | 2 |
| 5 | 2 |
| 6 | 2 |
| 7 | 2 |
| 8 | 2 |
| 9 | 2 |
| 10 | 2 |
| 11 | 2 |
| 12 | 2 |
| 13 | 2 |
| 14 | 2 |
| 15 | 2 |
| 16 | 2 |
| 17 | 2 |
| 18 | 2 |

Common columns value of ‘1’ would have indicated that only one gender has shopped in that category.

Search String: source="blackfriday.csv" | stats count(User\_ID) as cnt\_user by Product\_Category\_1, Gender | stats count(cnt\_user) as common by Product\_Category\_1 | sort -Product\_Category\_1 | reverse | where common=2