PROJECT REPORT

DATA MINING

TEACHER: AZAMAT KIBEKBAEV

COM-16:

MEDERBEK ABDYLDAEV

GULZAT UMETALIEVA

PROJECT NAME: **BLACK FRIDAY** DATA SET

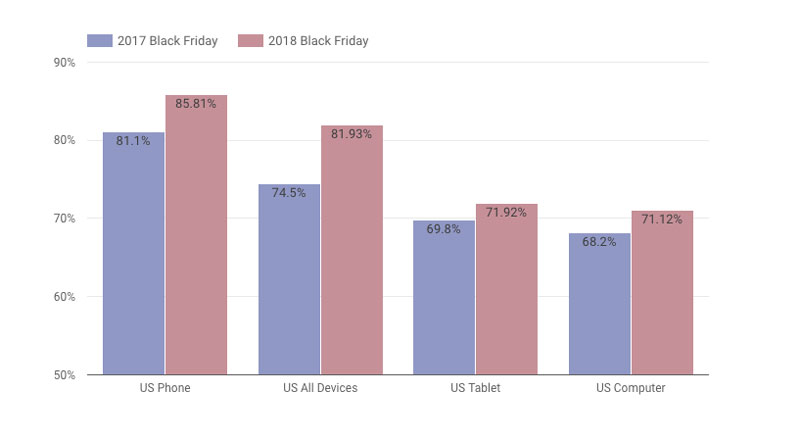
## 

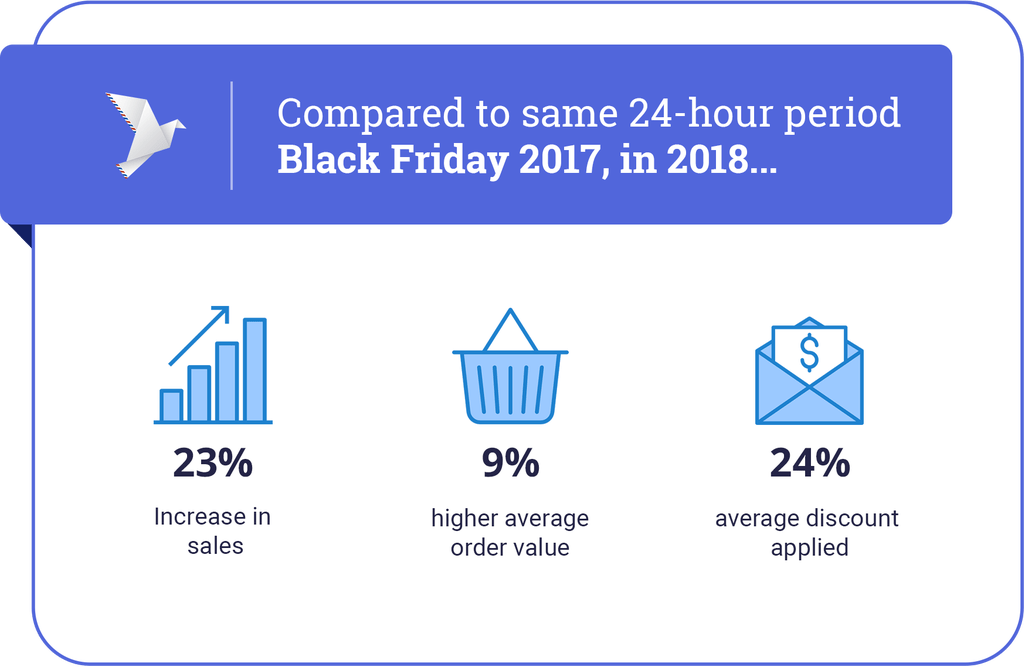
# CONTENT

* Our team is using as dataset details about Black Friday. The main purpose of this dataset is to show category of people buying in Black Friday. Work of our prediction with this dataset: predicting if person is married or not, regarding to his city category, age, occupation, purchase etc. We made a website to show the result, so it is easy for users just go to website, enter every detail and get the result.
* Our project made with the help of programming languages as Python, Django, also HTML, CSS to present it on a web site.

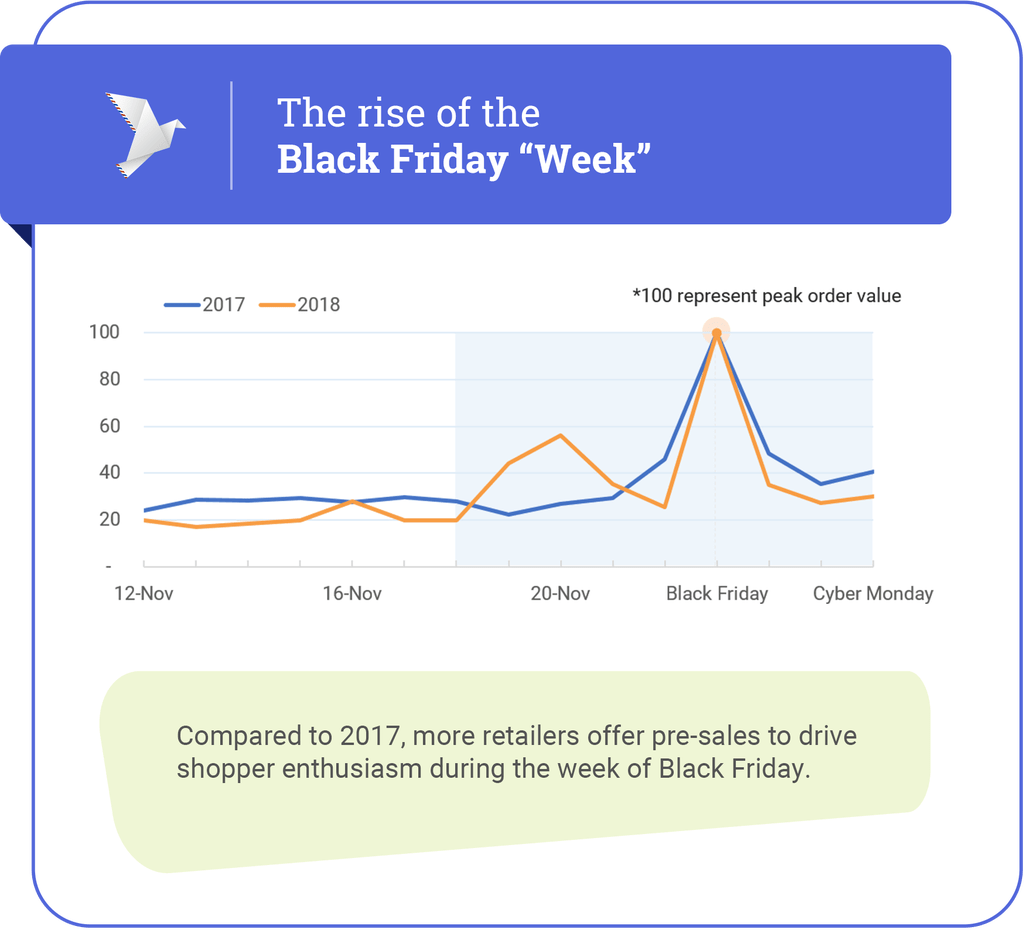
# LITERATURE REVIEW

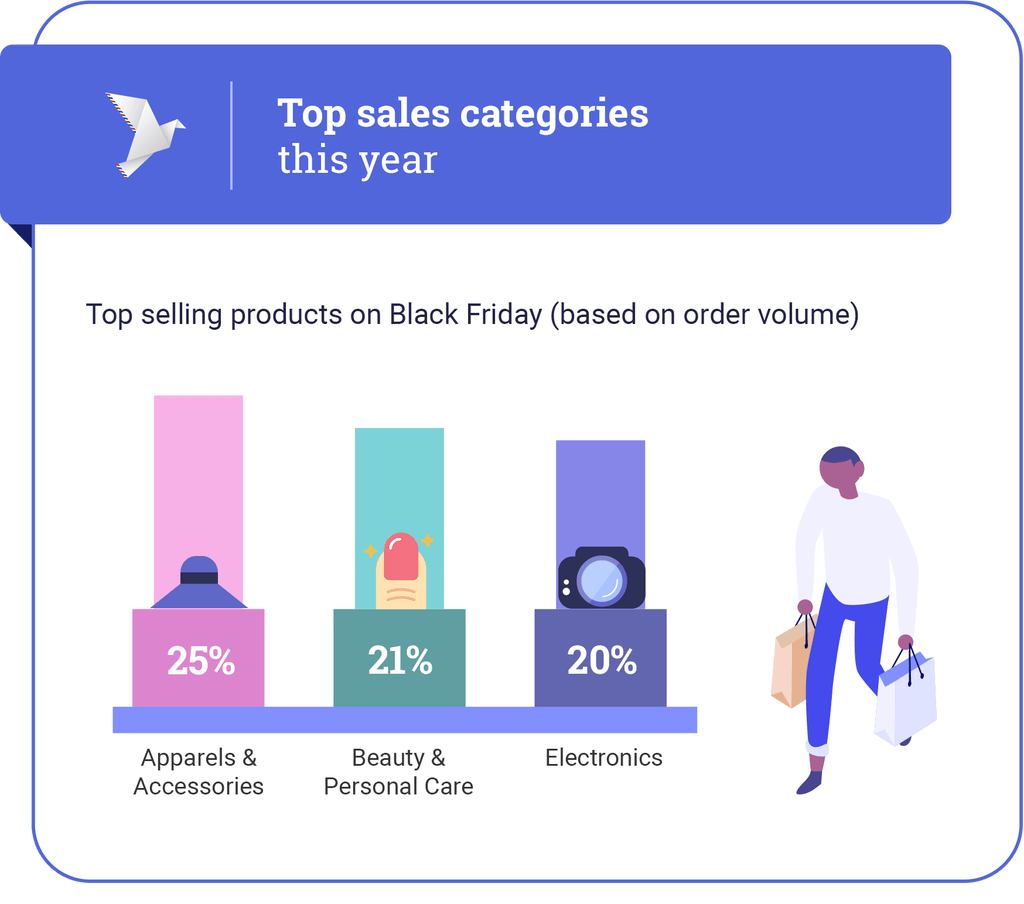
## How US Black Friday Statistics in 2018 compare to 2017:





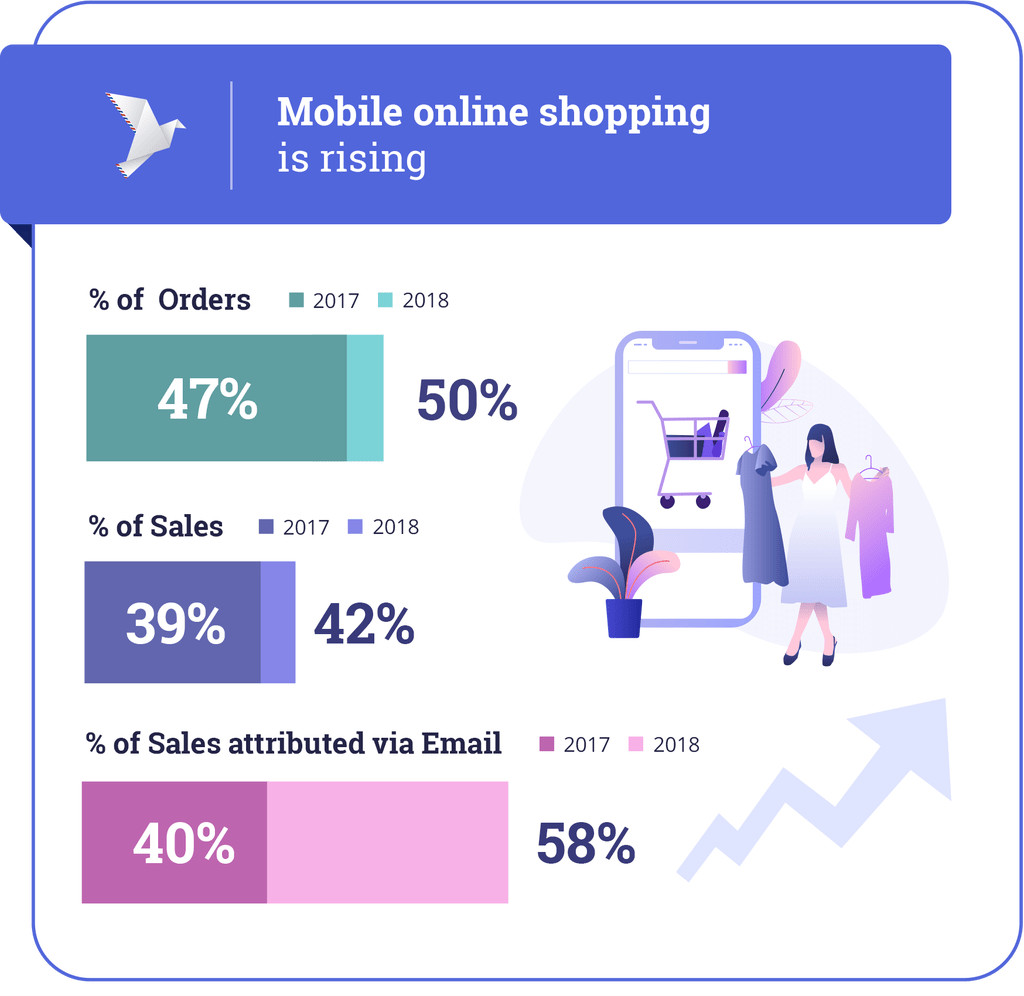
* Black Friday 2018 saw another 23% increase in total sales across online retailers.
* The 9% increase in average order value for Black Friday deals comparing to Black Friday 2017.This may be driven by growing consumer willingness to purchase more and higher value deal items this season – with popular items such as fashion and electronics.





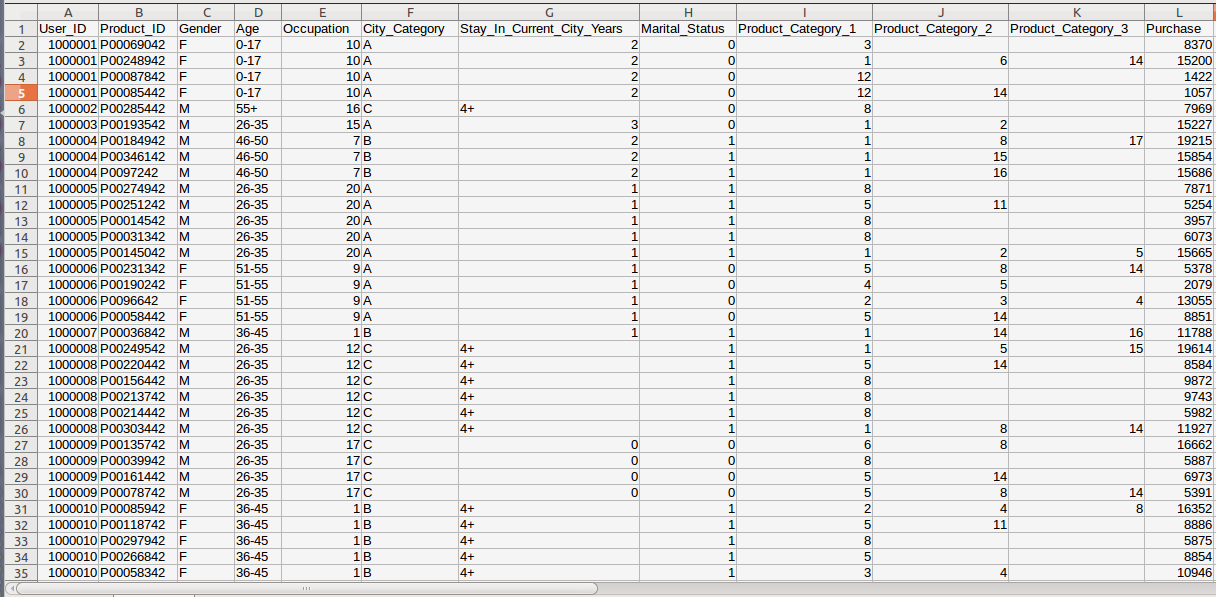
* It seems that fashion, beauty, and electronics takes up the most popular purchases for holiday gifting in 2018.



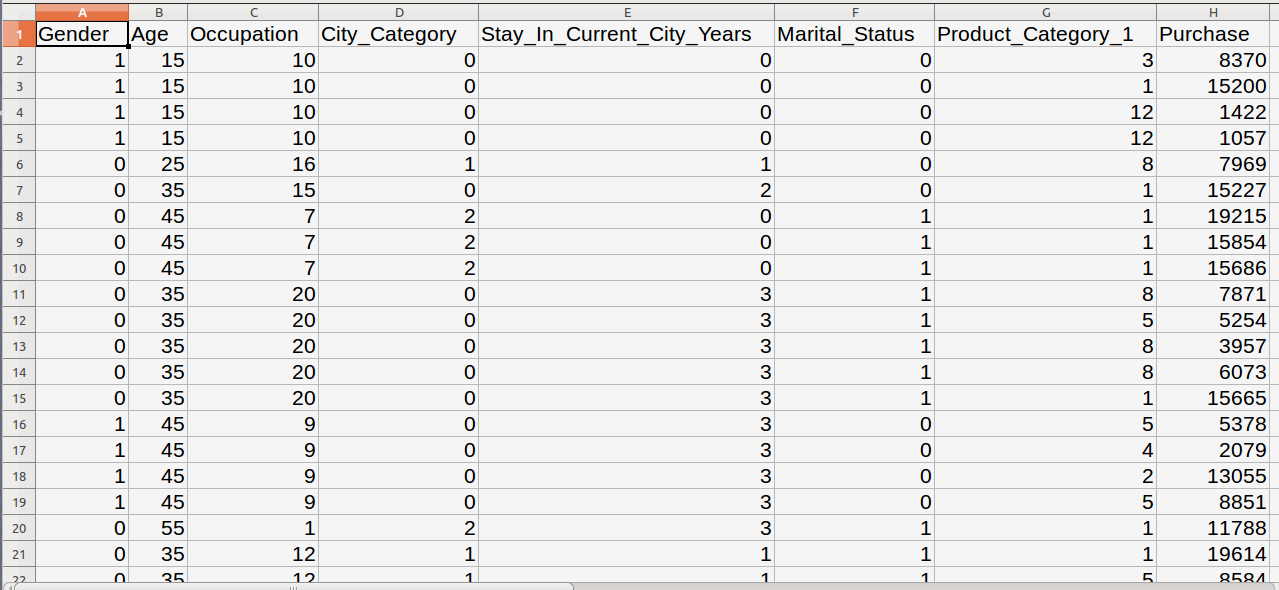


# METHODOLOGY

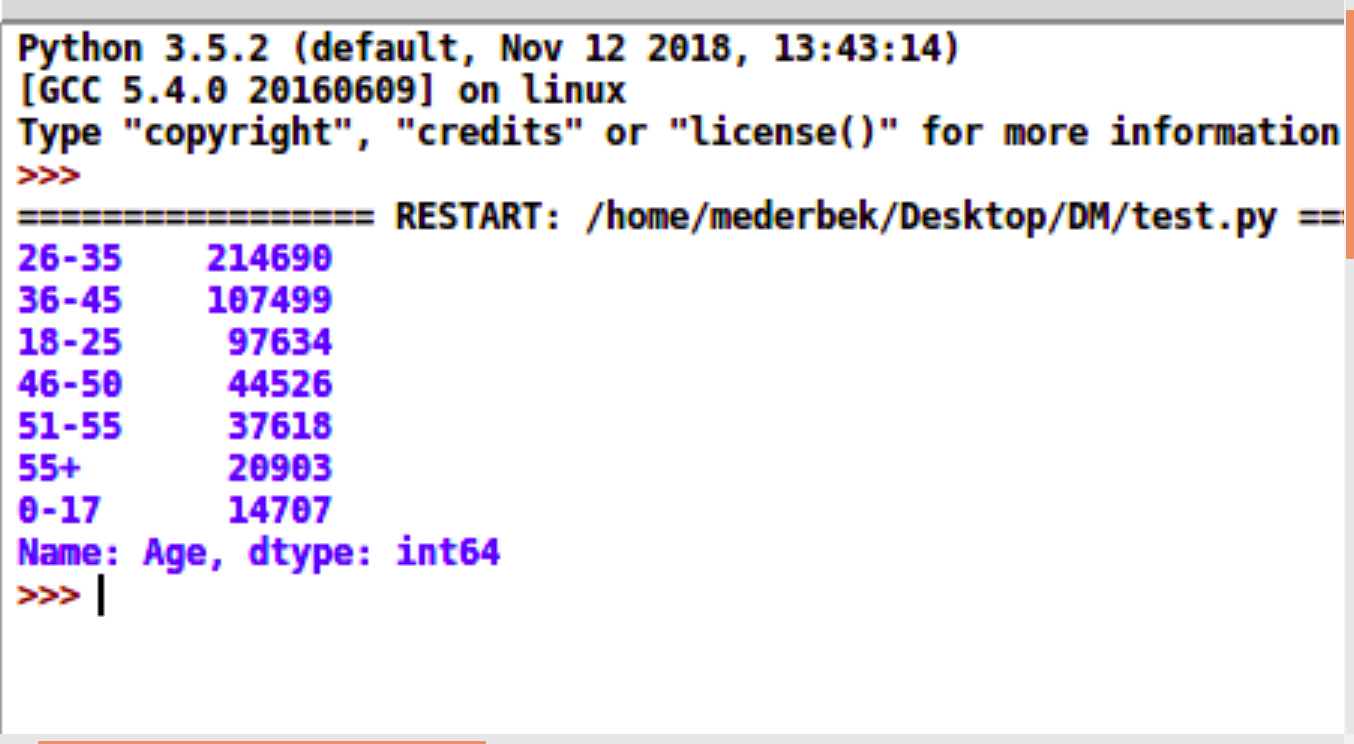
* Deleting User\_Id, Product\_Category\_2 and Product\_Category\_3
* Converting strings to int in Gender, Stay\_Years and City\_Category
* >>> dataset.head(35)

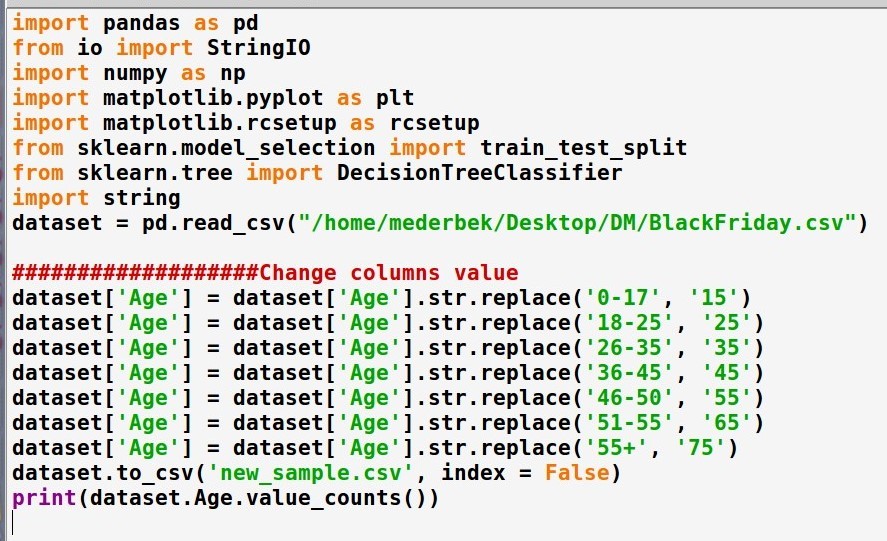


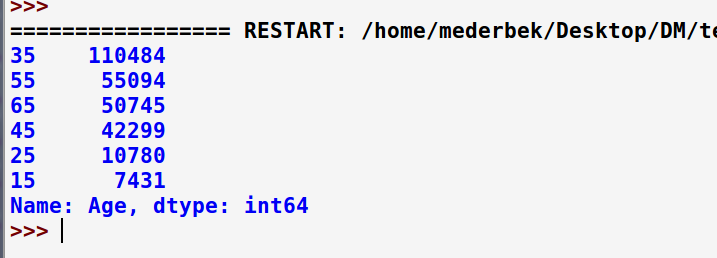
* >>> dataset.drop(['Product\_Id\_2', 'Product\_Id\_3', ‘User\_Id’], axis=1)
* >>> dataset['City\_Category'],\_ = pd.factorize(dataset['City\_Category'])
* >>> dataset['Stay\_In\_Current\_City\_Years'],\_ = pd.factorize(dataset['Stay\_In\_Current\_City\_Years'])
* >>> dataset[Gender],\_ = pd.factorize(dataset[Gender])
* >>> dataset.to\_csv('new\_sample.csv', index = False)
* >>> dataset.head(5)



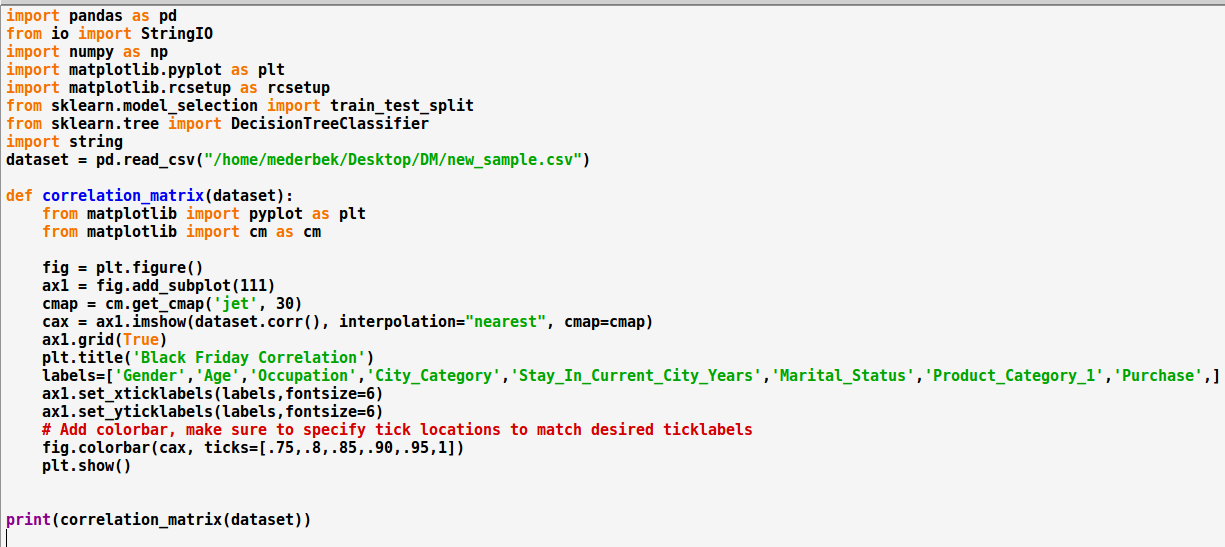
* >>>dataset=pd.read\_csv("/home/mederbek/Desktop/DM/BlackFriday.csv")
* >>> print(dataset.Age.value\_counts())



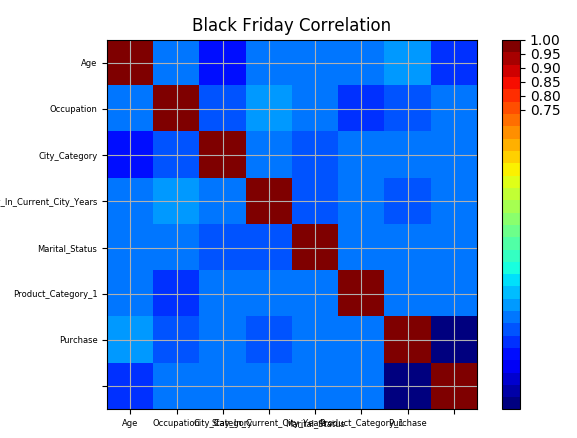




* **CORRELATION**
* [Correlation](https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient) gives an indication of how related the changes are between two variables. If two variables change in the same direction they are positively correlated. If the change in opposite directions together (one goes up, one goes down), then they are negatively correlated.

****

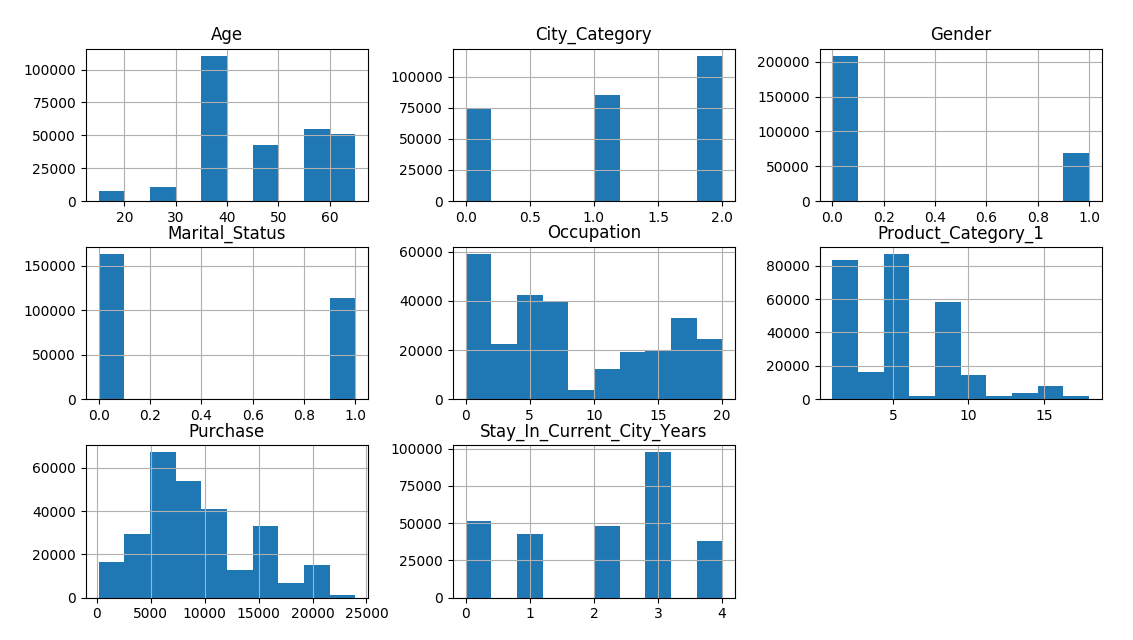
* >>> correlation\_matrix(dataset)



* **HISTOGRAMS**

Histograms group data into bins and provide you a count of the number of observations in each bin.

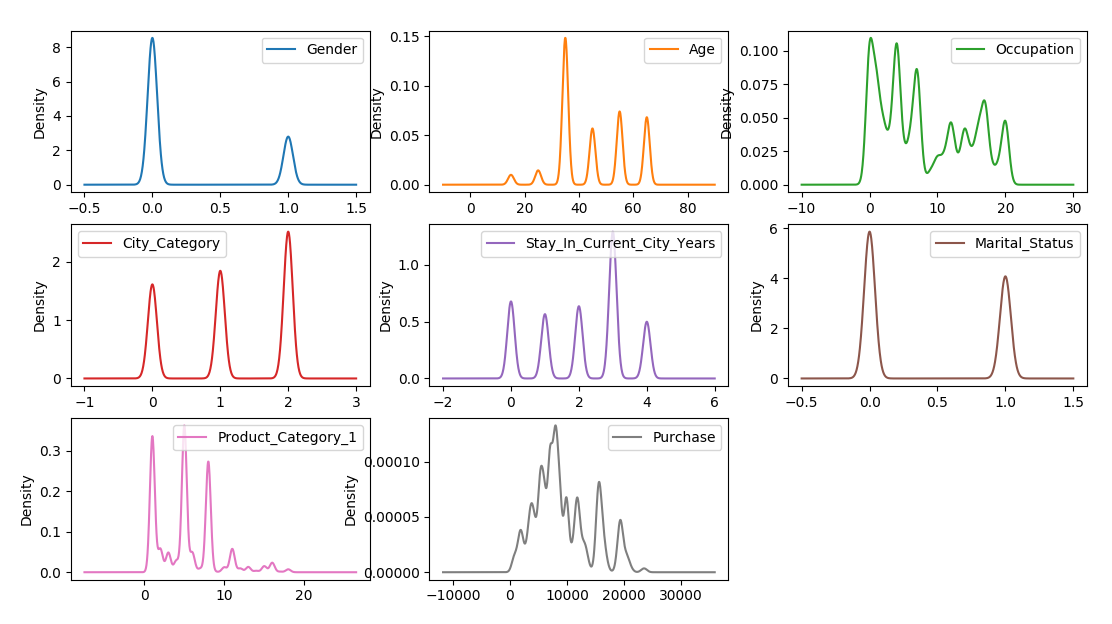
* >>> dataset.hist()
* >>> plt.show()



* **DENSITY**

Density plots are another way of getting a quick idea of the distribution of each attribute.

* >>> dataset.plot(kind='density', subplots=True, layout=(3,3), sharex=False)
* >>> print(plt.show())



# RESULT

* We get a well-designed web site, with menus (Project proposal, Prediction form, Result), everything is easy to understand for new users, in first page they can read detailed information about our project, on the second- form to fill in order to get the prediction and on the last page- result of prediction with the information that user give.
* We get an experience working with datasets with the help of data mining. We understood the process of working prediction.

# REFERENCES

* <https://machinelearningmastery.com/visualize-machine-learning-data-python-pandas/>
* <https://blog.contactpigeon.com/black-friday-report-2018/>