**Introduction**

I used dataset from a small online store (E-Commerce products) with customer purchase data in November 2018 (20.11-29.11) for the analysis and to find hidden patterns and specific user behaviors within this dataset. I analyzed the data using python.

The data include columns:

[**date**, **customer\_id**, **product\_category**, **payment\_method**, **value**, **time\_on\_site**, **clicks\_in\_site**]

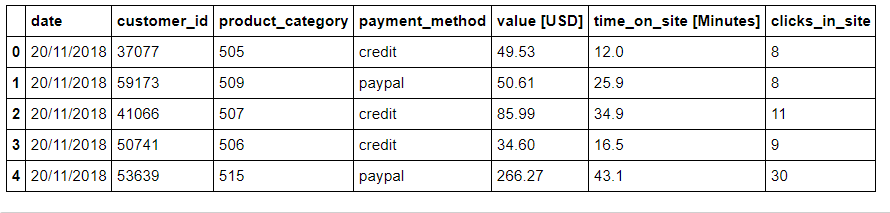
Notes:

* Product\_category is an id for a category of products (for example clothing, gadgets…).
* Each row is a cart (1-many items) for a customer. assume customers purchase from a single category each time.
* Payment\_method can be credit card or Paypal
* Value is the total value for the cart (can include any number of items) time\_on\_site is in minutes

**Summary**

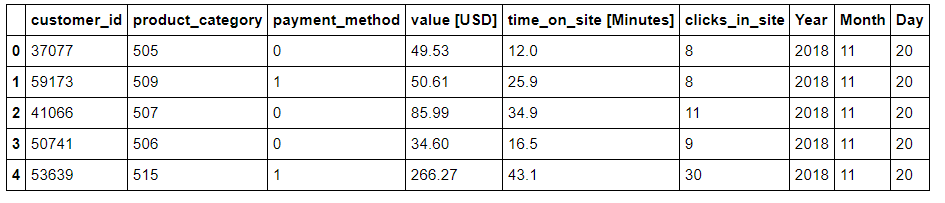
1. **Pre-processing**

First lets look at the dataset that I used in the analysis.



**Fig1: E-commerce Dataset**

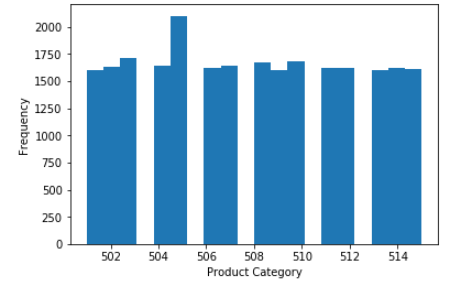
As it is seen in the figure1 that there are columns that need to be pre-processed before analyzing. The date column is not in proper form and I also have to convert the payment\_method column into numeric variables’ column. So, after pre-processing the dataset, I got this shape of dataset:



**Fig2: Modified Dataset**

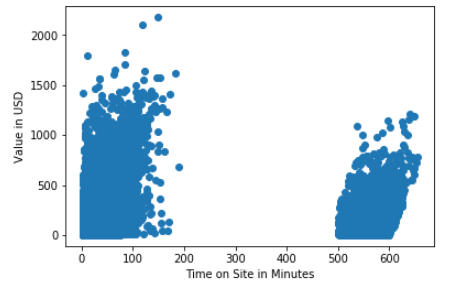
1. **Exploratory Data Analysis**

First, look at individual column’s behavior to find some insights:

****

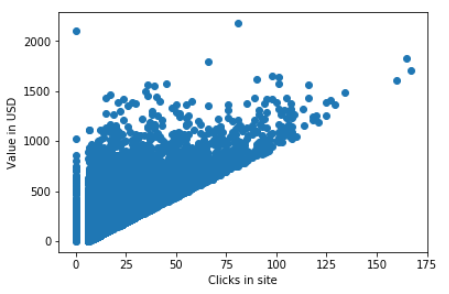
**Fig3: Product Category**

We can see that 505 id of product category has more sales than that of other categories and all the other categories have almost same sales, only 505 is sold more than 2000 times.

****

**Fig4: (Time on Site in minutes) VS. (Value in USD)**

Following figure reveals that there is a highly positive correlation between clicks on site and value in USD. So, we inferred from here that more number of clicks may increase the sales and less number of clicks may decrease the sales. That’s why, on day 24, there was least sales and day 23, there was the most sales occurred.

****

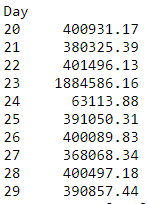
**Fig5: Clicks on site VS. Value in USD**

We can see that payment method doesn’t have any impact on sales since payment method is not impacting on the sales on e-commerce. As 0 represents Paypal method and 1 represents Credit Card method.

**C:\Users\Rao\Desktop\5.PNG**

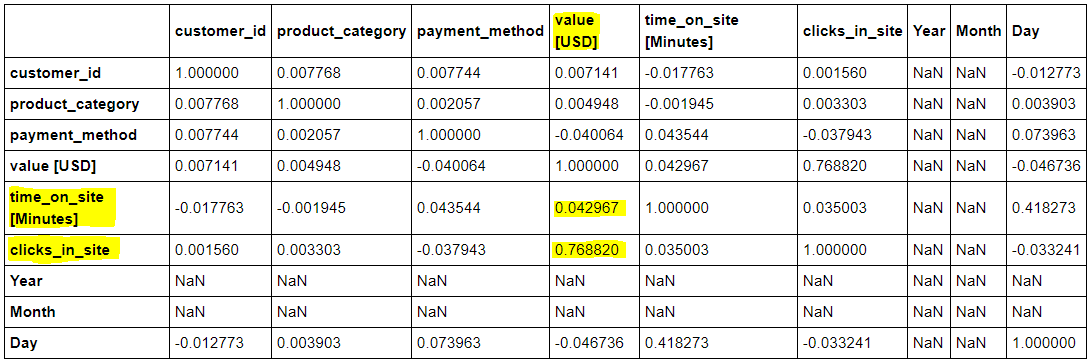
**Fig6: Payment method Vs. Values in USD**

As you can see in the following figure, day 23 has the most sales that are 1884586.16 USD and day24 has least sales that are 63113.88 USD. And that is because of less number of customers visited the site because on the day 24, there are less number of clicks on sites and less time consumption on site.

****

**Fig7: Day VS. USD**

The following figure demonstrates the correlation among all the features but I just more interested to see the correlation of all the features with the Value [USD]. So, We can see that there is highly positive relation of time on site and clicks in site with the value[USD]. So, we inferred from here that more number of clicks and more time consuming on site may increase the sales and less number of clicks and less time consuming on site may decrease the sales. That’s why, on day 24, there was least sales and day 23, there was the most sales occurred.

****

**Fig8: Correlation Table**

**Conclusion:**

The analysis indicated that the time consuming on site and more number of clicks on site are positively correlated to value in USD. So, the value in USD is highly dependent on these two factors from the clients. More number of clicks and more time consumption on site may increase the sales of your e-commerce website. Product category is also positively correlated but not so important to predict the sales of the site. All other factors are not so important for the sales of your site.