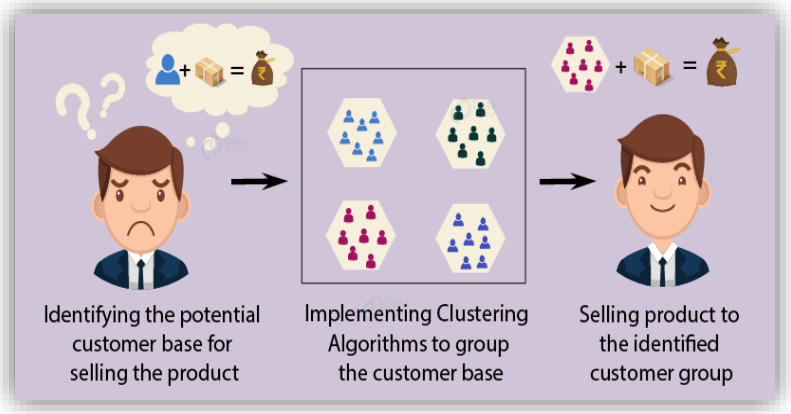
Customer Segmentation using K-Means algorithm

Sameer Sekhar Bashistha

Date – 16/12/2021

1. Introduction

Customer segmentation is the method of distributing a customer base into collections of people based on mutual characteristics so organizations can market to **group** efficiently and competently individ2ually.



The purpose of segmenting customers is to determine how to **correlate** to customers in multiple segments to **maximize** customer benefits. Perfectly done customer segmentation empowers marketers to interact with every customer in the best efficient approach. The fundamental goal is to identify groups, segments or clusters of customers (the terms are interchangeable) that from a *marketing perspective* are meaningfully different from each other.

There are different methodologies for customer segmentation, and they depend on four types of parameters:

- geographic,
- demographic,
- behavioural,
- psychological.

Geographic customer segmentation is very simple, it's all about the user's location. This can be implemented in various ways. You can group by country, state, city, or zip code.

Demographic segmentation is related to the structure, size, and movements of customers over space and time. Many companies use gender differences to create and market products. Parental status is another important feature. You can obtain data like this from customer surveys.

Behavioural customer segmentation is based on past observed behaviours of customers that can be used to predict future actions. For example, brands that customers purchase, or moments when they buy the most. The behavioural aspect of customer segmentation not only tries to understand reasons for purchase but also how those reasons change throughout the year.

Psychological segmentation of customers generally deals with things like personality traits, attitudes, or beliefs. This data is obtained using customer surveys, and it can be used to gauge customer sentiment.

2. Problem Statement

To analyse and create a model that estimates credit card customer segmentation to help the company to define its marketing strategy by using the K-means algorithm with the K value determined by silhouette score. I also used PCA for dimension reduction and better visualization**.**

3. Market/Customer/Business need Assessment

Over the years, the commercial world has become more competitive, as organizations such as these have to meet the needs and desires of their customers, attract new customers, and thus improve their businesses. Intuitively, we know that the customer base, be it 50,000 or 5 million, is not homogeneous. In other words, at any point in time it will contain some good customers, some bad ones; some new customers, some old ones; some young, some old; some rich, some less rich, some poor; some price sensitive, some not; some extremely loyal, some not loyal at all. The list goes on and on.

And, because of this diversity, it makes little apparent sense to market to all customers the same way, i.e., the same level of marketing effort, the same offer, the same copy, the same creative. Yet this is what direct marketers do, perhaps without realizing it, when they search for the best control package in direct mail (the package that works best across all customer files) or the one best script in outbound telemarketing.

So, by segmenting users, **marketers** can obtain the most maximum of their operations budgets by targeting the appropriate audiences. Businesses can converse straight to customers who are most assuring to transform without spending money on impressions or users who aren't inclined to purchase the following product.

Segmentation promotes a corporation in the following ways:

- Design and deliver targeted marketing advice that will resonate with particular customer associations but not with others (who will accept notifications according to their requirements and importance, preferably).
- Decide the most reliable communication course for the segment, anything from email, social media posts, radio advertising, or a different procedure, depending on the feature.
- Distinguish methods to promote products or new merchandise or assistance opportunities.
- Build more trustworthy consumer relations to enhance customer assistance.
- Analysis of pricing selections to concentrate on the most influential customers.

4. Target Specification and Characterization

The goal of this analysis report is to discover the Customer Segmentation of a bank, by looking through their behaviour/profile while using Credit Card. Hopefully, we can get a clear segmentation of the customer, so we can deploy effective marketing campaign or sales promotion to the targeted costumer. The sample Dataset summarizes the usage behaviour of about 9000 active credit card holders during the last 6 months. The file is at a customer level with 18 behavioural variables.

5. External Search (Information Sources/References)

https://medium.com/analytics-vidhya/credit-card-customers-segmentation-bc3c5c87ddc

https://survicate.com/customer-segmentation/customer-segmentation-examples/

https://medium.com/analytics-vidhya/credit-card-customers-segmentation-bc3c5c87ddc

The dataset is available on Kaggle:

https://www.kaggle.com/kashyaprahul/credit-card-segmentation

In [3]:
```
cc_df
```

Out[3]:

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQ
0	C10001	40.900749	0.818182	95.40	0.00	95.40	0.000000	0.
1	C10002	3202.467416	0.909091	0.00	0.00	0.00	6442.945483	0.
2	C10003	2495.148862	1.000000	773.17	773.17	0.00	0.000000	1.
3	C10004	1666.670542	0.636364	1499.00	1499.00	0.00	205.788017	0.
4	C10005	817.714335	1.000000	16.00	16.00	0.00	0.000000	0.
...	...	...	...	...	...	...	...	...
8945	C19186	28.493517	1.000000	291.12	0.00	291.12	0.000000	1.
8946	C19187	19.183215	1.000000	300.00	0.00	300.00	0.000000	1.
8947	C19188	23.398673	0.833333	144.40	0.00	144.40	0.000000	0.
8948	C19189	13.457564	0.833333	0.00	0.00	0.00	36.558778	0.
8949	C19190	372.708075	0.666667	1093.25	1093.25	0.00	127.040008	0.

8950 rows × 18 columns

In [4]:
```
cc_df.shape
```

Out[4]:
```
(8950, 18)
```

6. Bench marking alternate products

In today's world, almost all companies do customer segmentation for their target customers in one way or another (sometime without even explicitly realising it). But not all are successful as they fail to fully understand & explain they product proposition to their target markets. For this, an in-depth understanding of customer segmentation is the basic requirement coupled with dynamic refinement to product & strategy to counter ever changing needs of customers.

Policy Bazaar: Indian Fintech Unicorn Policy Bazaar is the largest Insurance aggregator with major market share. It uses Geo-Demographic (Location, Age, Income) & Firmographic (Company profile, Work type) segmentation aided with Psychographic (Social status, lifestyle etc) & Behavioural (Usage, benefits sought) data to target their potential customers. The moto for the company is to provide the right product to the right customer. Company also uses Big data to use the customer buying to cross sell next product thus creating a huge loyalty marketing base.

H&M: The Swedish multinational clothing company H&M uses psycho-demographic & behaviour segmentations like Urbanity, Age, Gender, AIO variables to attract the right customers, who are spending more in their store.

The online wine club **Vinomofo,** for example, segmented web traffic into new visitors, returning visitors and returning customers. New visitors got a **$15** incentive to join the Vinomofo community. Returning visitors who hadn't purchased anything were reminded they had the **$15** to spend. Regular customers saw a completely different pitch.

The luxury goods manufacturer **Montblanc** uses market segmentation to present different offers to different groups of website visitors. Shoppers looking for corporate gifts had a different online experience, with different offers and incentives, then shoppers looking for upscale Father's Day gifts.

7. Applicable patents

US20080082386A1, Systems and methods for customer segmentation by Stuart Alan Cunningham, William Thomas Franey.

https://patents.google.com/patent/US20080082386

8. Applicable Constraints

- To segment customers, companies need a way to track product usage for each user which is not easy.
- Understanding that people can belong to multiple segments.
- Privacy and safety of the users/customers.
- Continuous data collection and maintenance.

9. Business opportunity

Customer segmentation is often used as a top marketing priority for banks and many other commercial companies, for good reason. Because segmentation solutions help them to group customers by behaviour, they can offer more tailored products and services. Moreover, by better understanding customer preferences, marketers can maximize cross and up-selling opportunities and encourage customers to explore related services. Other ways that segmentation can help companies are:

- Decide on the right kind of promotional content to send each customer.
- Select the right marketing channels for the right audience.
- Identify new and profitable segments and launch innovative products and services.

Nowadays customer segmentation technique is also used by small business to better understand their customers and grow efficiently. It is very much important for small businesses to understand and determine their **target markets**, which are basically the segments of customers whom they should target in order to receive the most sales from.

It is used to ensure effective marketing to those who want or need it most. Without customer segmentation, it is like marketing to everyone and in most to almost all cases, not everyone needs the product or service.

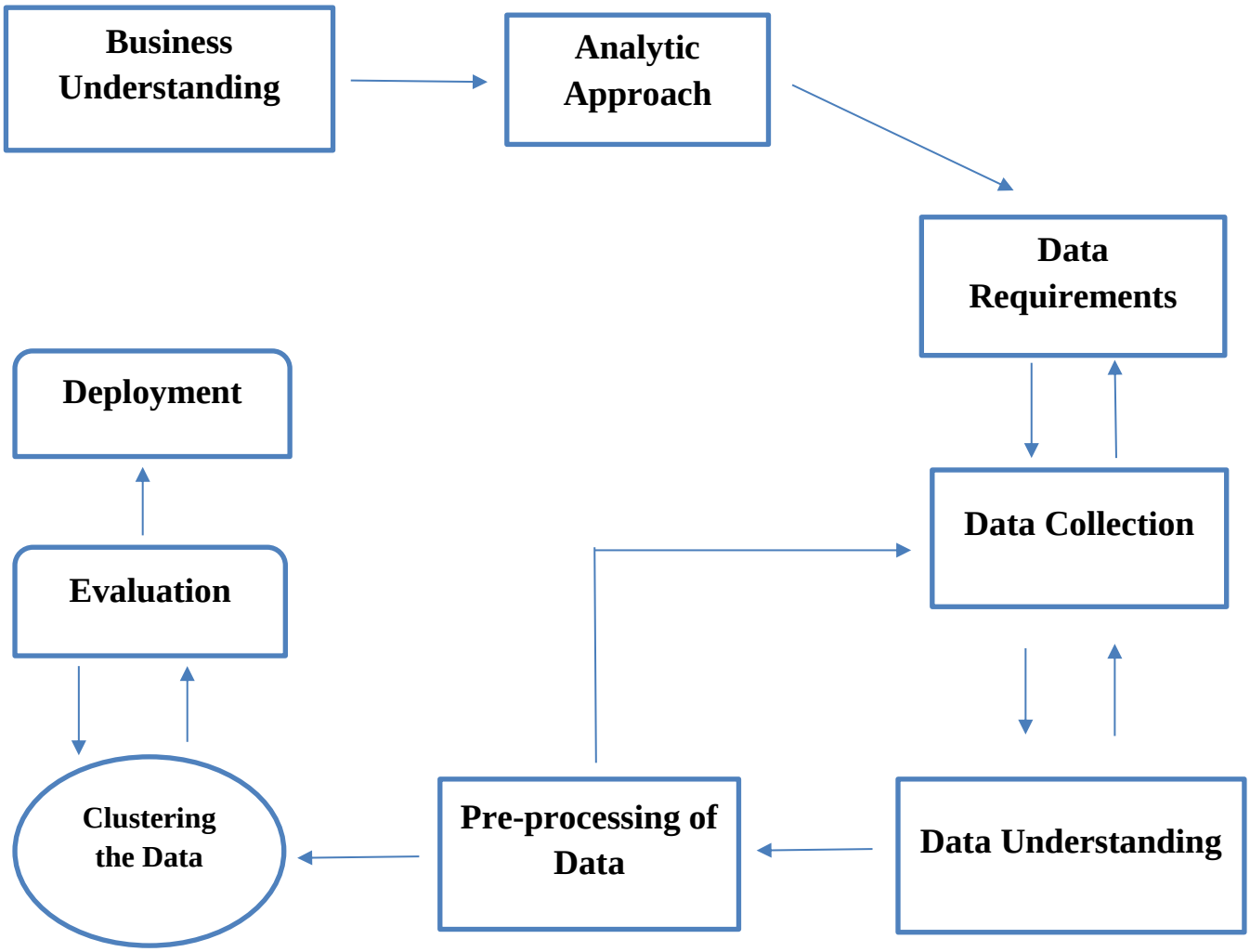
10. Concept Generation

The sole purpose of this concept is to make a model for segmenting user on the basis of their behaviour (for example: amount of purchase made from their account, how frequently they are purchasing things etc), so that the bank could define its marketing strategy.

11. Concept Development

On the Credit Card customer segmentation data, we will firstly do the pre-processing so that we can fit it in our model. We will do some basic feature comparison and visualization to understand the data better in a better way. This is a clustering problem therefore we will use K-Means clustering model and the number of clusters to be created will be decided by using elbow method and silhouette score method.

12. Final Product Prototype (abstract) with Schematic Diagram



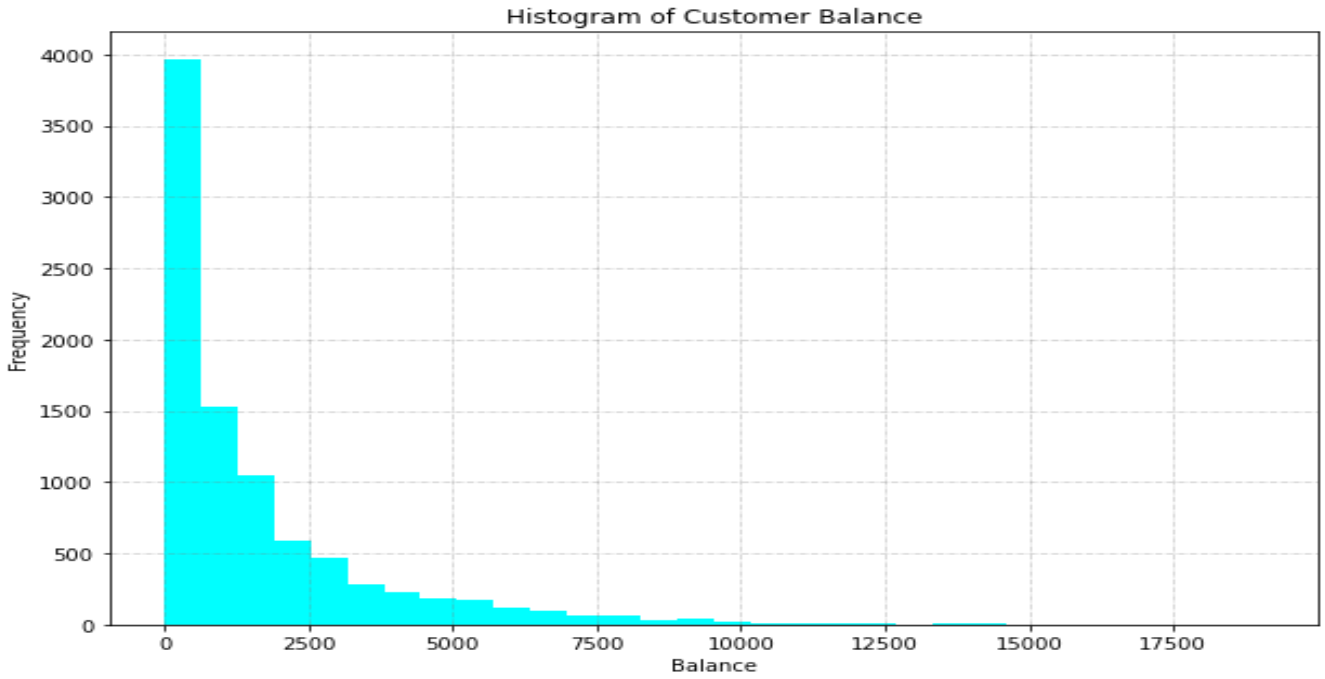
13. Code Implementation

GitHub Link: https://github.com/sameer06032000/Credit-card-customer-segmentation.git

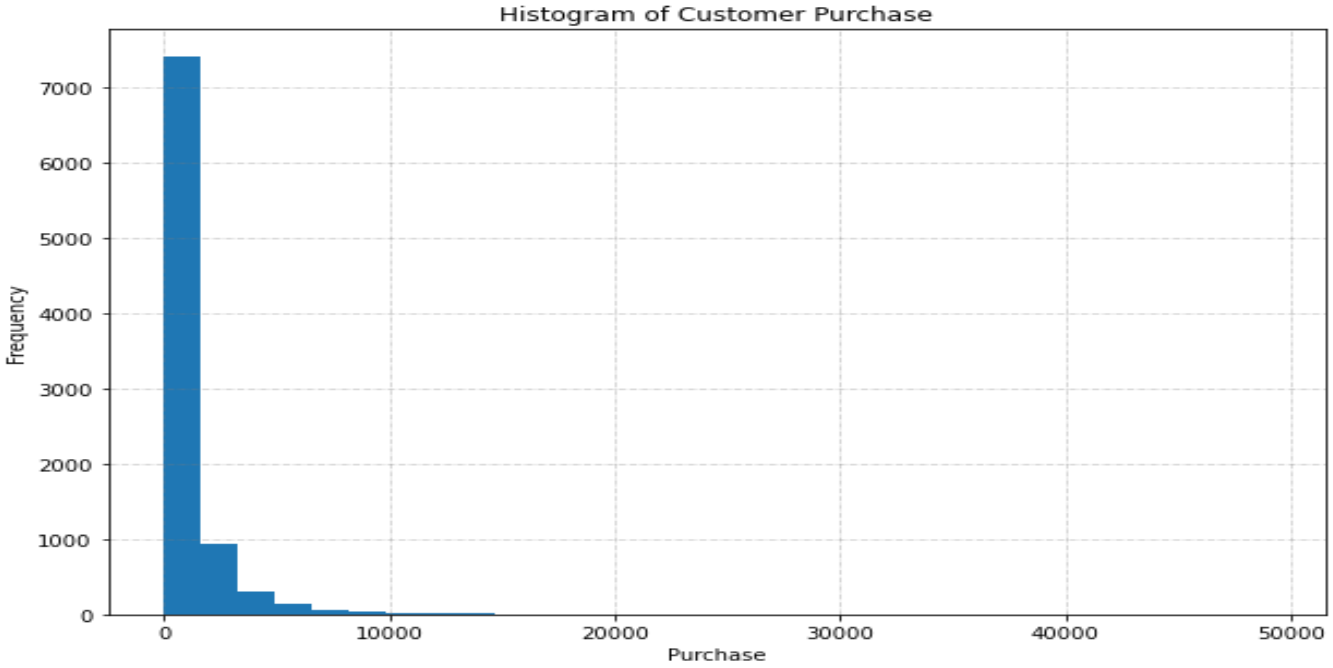
Environment and tools: -

- **Scikit-learn**
- **Seaborn**
- **NumPy**
- **Pandas**
- **Matplotlib**

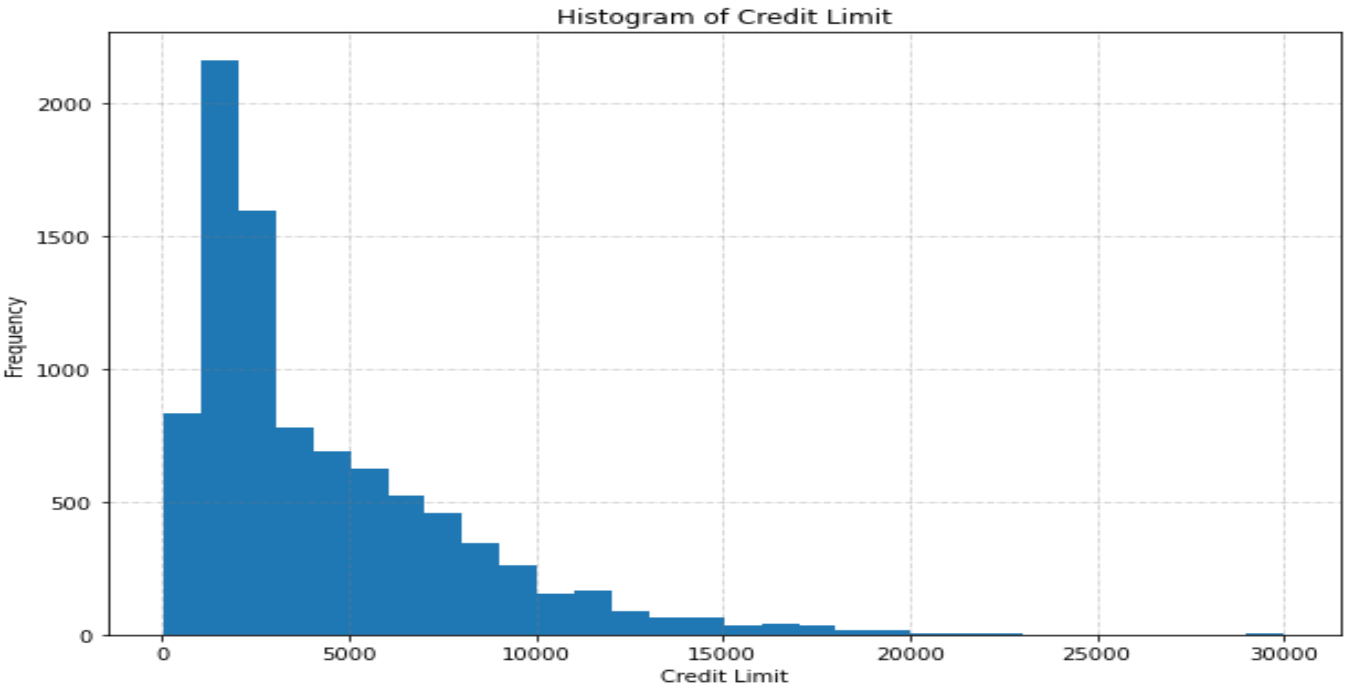
Some Basic Visualisations:



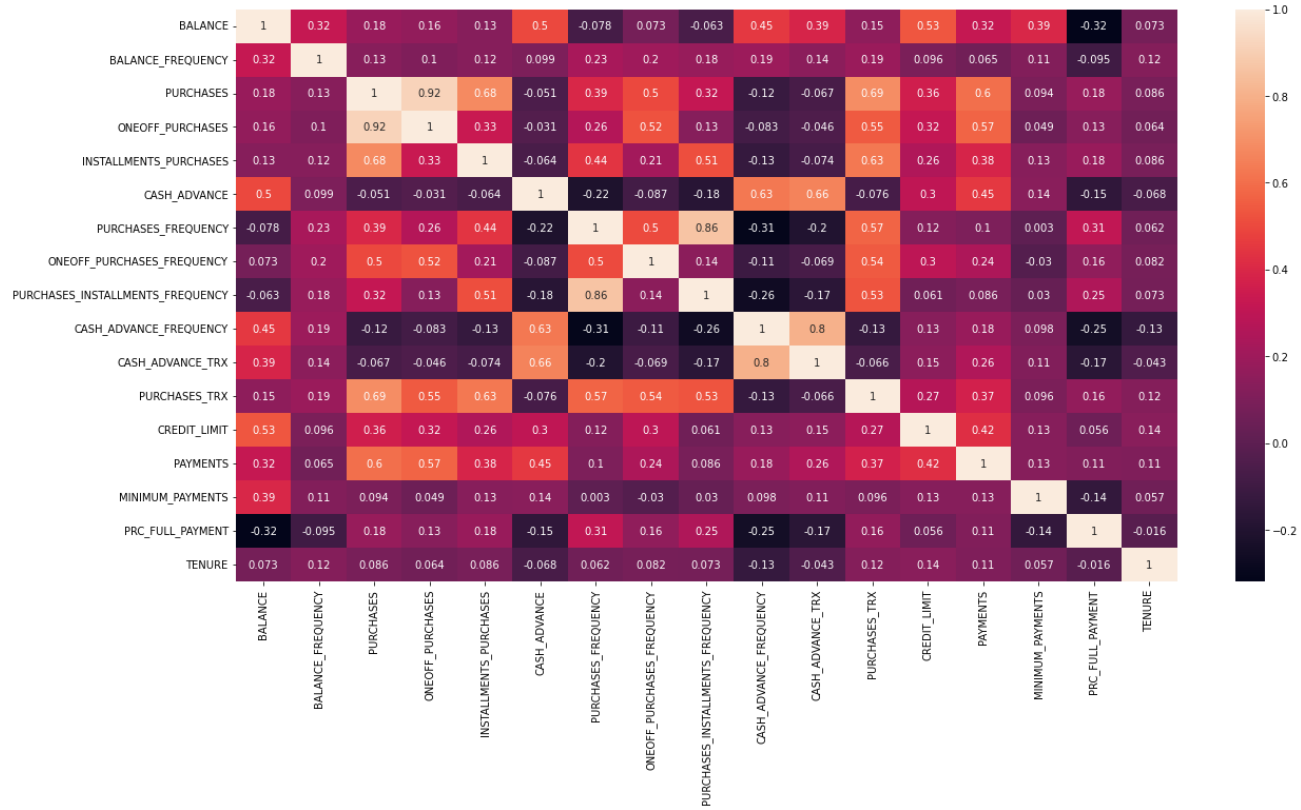
It seems most of the customers who use their Credit Card would maximize the usage of their credit balance, until it reaches 0.



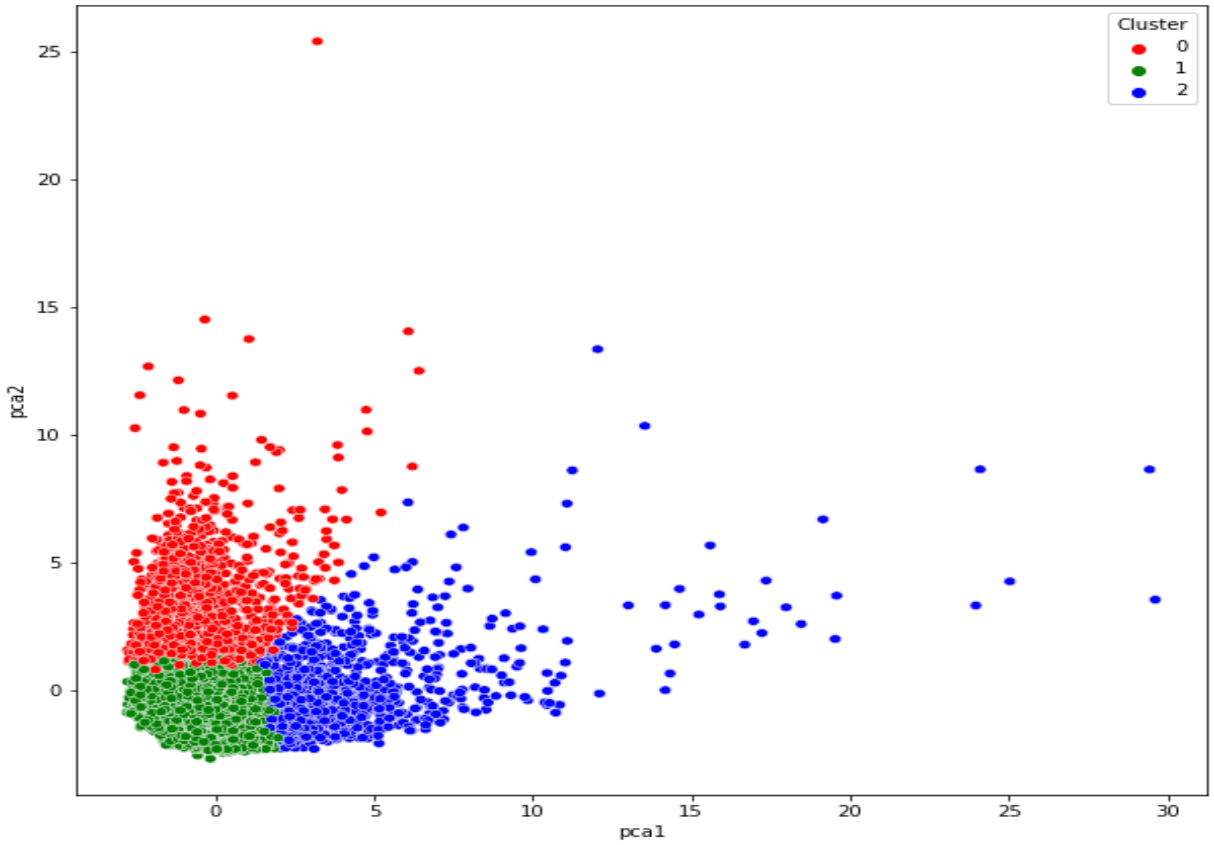
Most of the customers don't use their Credit Card to purchase something.



Most of the customers have Credit Limit nearly around 2000.



Some of the features are highly correlated but we are not going to delete those features because it doesn't affect our model.



Cluster 1(Green): This customer group indicates a small group of customers who have low balances, small spenders (low purchase) with the lowest credit limit.

Cluster 2(Blue): This customer group indicates a large group of customers who have medium balances, spenders (high purchase) with the highest credit limit.

Cluster 0(Red): This customer group indicates a small group of customers who have high balances and cash advances, low purchase frequency with high credit limit. We can assume that this customer segment uses their credit cards as a loan.

14. Conclusion

K means clustering is one of the most popular clustering algorithms and usually the first thing practitioners apply when solving clustering tasks to get an idea of the structure of the dataset. The goal of K means is to group data points into distinct non-overlapping subgroups. One of the major applications of K means clustering is segmentation of customers to get a better understanding of them which in turn could be used to increase the revenue of the company.