**Project Report**

****

**Customer Analysis**

**Sukrit Arora**

**Sunakshi Dutt**

**Varun Ahuja**

**15CSU239**

**15CSU241**

**15CSU254**

**Date:14 Nov ,2018**

**THE NORTHCAP UNIVERSITY**

**GURUGRAM**

**TABLE OF CONTENT**

|  |  |  |
| --- | --- | --- |
| S.no | Title | Page Number |
| 1. | List of Figures | 3 |
| 2. | Abstract | 4 |
| 3. | Title of the Project | 5 |
| 4. | Introduction | 6-10 |
| 5. | Description | 11-20 |
| 6. | Analysis | 21 |
| 7. | Conclusion | 22 |
| 8. | Bibliography | 23 |
| 9. | Appendices | 24-27 |

**LIST OF FIGURES**

|  |  |
| --- | --- |
| **Title** | **Page No.** |
| Project Title Image | 5 |
| Data Flow Diagram (DFD) | 14 |
| Quartiles Distribution Diagram | 18 |
| RFM Quartile Table | 18 |
| Segment Profile Table | 19 |

# ABSTRACT

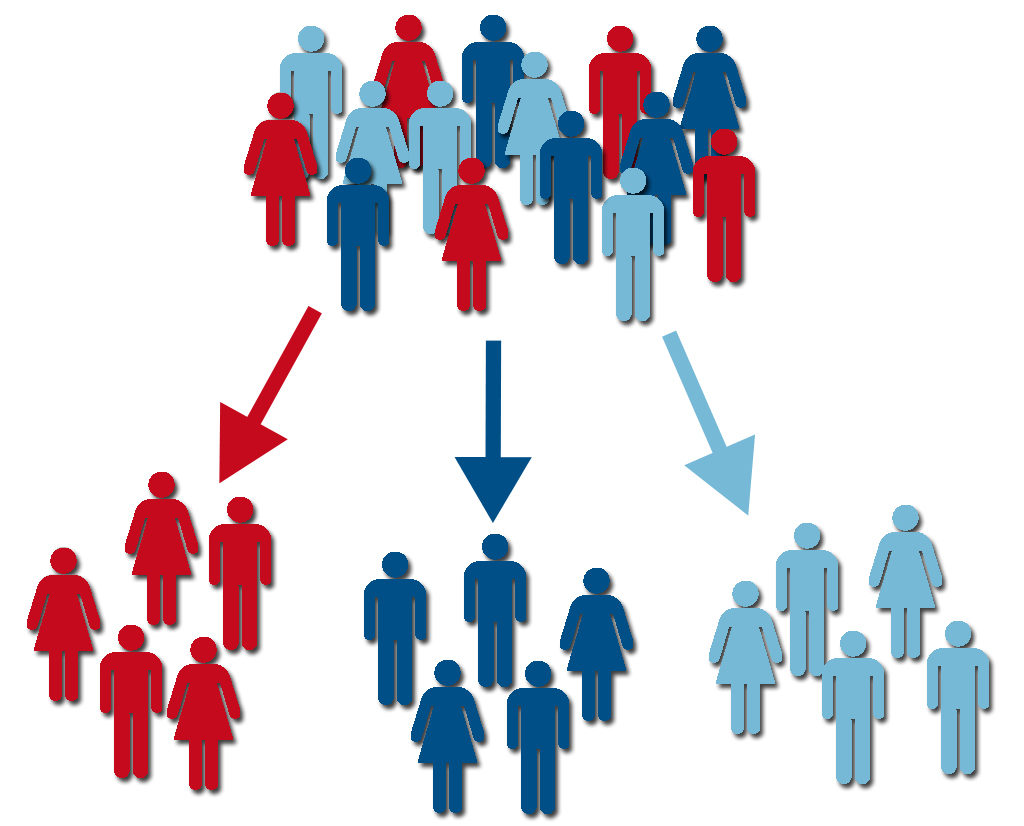
Under the data science field, initially we researched the different sectors in which we can apply data analytics skills with R language and Python and then planned out the most feasible and in demand usage area with my mentor.

Later I designed out the model for achieving customer analysis using the segmentation process.

The report presents the two main tasks ,which are listed below:

* Conceptual learning of R language and Python (Spider IDE)
* Working on project – Customer Analysis

**Customer Analysis**



**INTRODUCTION**

With the help of my mentor we were able to come up with the idea of Customer Analysis in the field of Data Sciences. Then we researched in this field to understand how it can be useful for different organization.

After understanding it, we brushed up our concepts on the technology used for the corresponding project. This led to practice of technical skills which were new to me.

Under the constant guidance and support of our mentor, We were successfully able to complete our project.

**DESCRIPTION**

**Research of the Project and Practice of Technical Skills:** -

For the Research of the Project, I was recommended to study on various platforms. Understanding the requirement of the organization.

After choosing the field of data sciences, my task was to implement it for the benefit of the company.

Since the company deals with a lot of product manufacture and services, it acquires various customers.

Customer analysis is a critical component of any business plan in all stages of growth. When you analyze your customers, you define who your target market is, and decide how you'll reach them. A recent article in Forbes stated that 81% of enterprises rely on analytics to improve their understanding of customers

Understanding the importance of Customer Analysis, I decided to work on this project.

Later came the task to identify the technical skills used for this project.

The two main languages which can be used for the same was R programming and Python.

I began my practice by understanding the basic functioning of various commands and syntax in both of the programing languages.

Initially I brushed up my concepts of R language, since I already had an idea about it.

Then, I started studying Python, which was a new language for me.

I began by installing the python 3.6version.

Then I installed and worked on the following:-

Anaconda

Anaconda is professional data science platform and python IDE's where we can use as console as well as GUI .

Spyder

Spyder is a powerful interactive development environment for the Python language with advanced editing, interactive testing, debugging and introspection features. Additionally, Spyder is a numerical computing environment thanks to the support of IPython and popular Python libraries such as NumPy, SciPy, or matplotlib.

Pandas

Pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labeled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real world data analysis in Python.

**MAIN PROJECT**

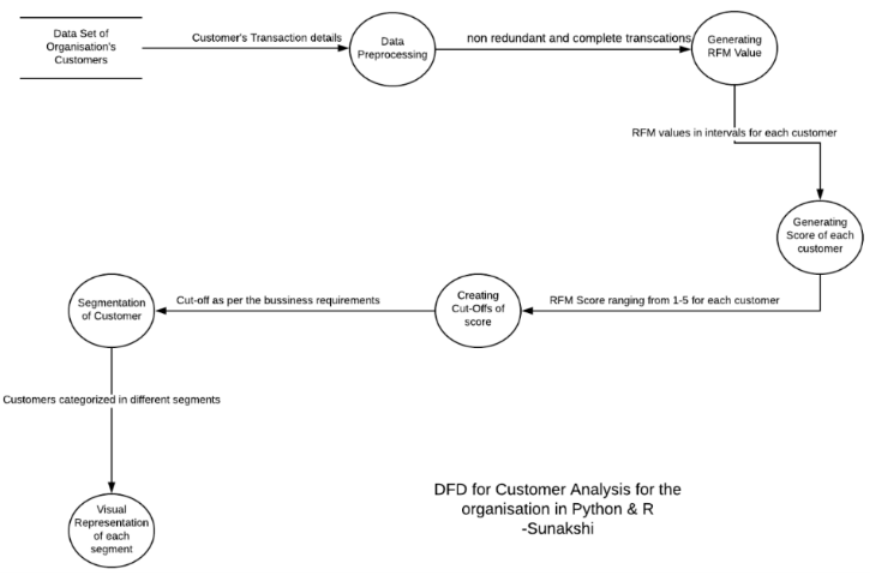
BACKGROUND

Understanding customer purchasing decisions is the key to increasing sales. Use customer analysis to identify factors that have both a positive and negative impact on sales. Customers play a key role in any organisation. The process of Customer Analysis begins with the various data of customers, which is used to later segment the customer. These Segments are given a profile according to which the management decisions can be undertaken.

OBJECTIVE

The purpose of undertaking customer analysis is to examine the consumers most likely to purchase your product or service in-depth. Organization can establish different groups of customers and the needs of those customers. By understanding what motivates them to make a purchase, and can build their business around providing solutions to those needs.

DATA FLOW DIAGRAM



|  |
| --- |
|  |
| ADVANTAGES |
|  |
| |  |  | | --- | --- | |  | 1. Marketing Efficiency  Focusing on the individual customer takes your marketing analysis beyond just knowing your spend and the eyeballs you received in return. Knowing which marketing channels bring the highest value customers in terms of order size, retention rate and profitability allows you to either cut marketing costs or expand your reach more efficiently.  2. Customer Retention  Customer acquisition is expensive, so it's important to understand what causes customers to leave. Customer analysis can help you identify common denominators among lost customers and give you an early warning that existing customers may be in danger of leaving if you don't take corrective action.  3. Increased Sales  Understanding customer purchasing decisions is the key to increasing sales. Use customer analysis to identify factors that have both a positive and negative impact on sales. This could include shipping times, how customer service interactions are handled, whether you have a minimum order or bundled discount, or the customer's location or income.  4. Improved Profit Margins  Not all customers are equal. Some customers are more profitable than others, and some may even cost you money. Factors that affect customer profitability include order size, cost of handling the order, time spent servicing the account and returns. | |  | 5. Fully customizable and scalable setup  According to the business needs and goals, the profiling of customer can be easily modified. One modification results in overall change. The amount of data can be huge. | |  | 6. System is accessible anytime and anywhere  The setup once put up in a system is easily accessible anytime. The new data can be uploaded to it for new results anytime and anywhere. | |  | 7. Updates to date according to the data loaded  The Recency factor for calculation is updated according to the current date and data loaded. | |  | 8. Reduces workload of manual calculation  Earlier the calculation was done manually and wasted a lot of time. Hence the project saves time and human labour. | |  | 9. Eliminate paper work. Go green | |

All the paper work done during manual calculations are saved.

Becoming an ecofriendly option.

SECTIONS COVERED

* **SECTION 1:** DATA PRE-PROCESSING:

In this section the data of the customers are entered.

This is in raw form with various unwanted data. The following were the tasks performed on the data: -

* Removing duplicate Data
* Dealing with Null Value in Data
* Removing Outliers if any

After performing the basic tasks of pre-processing.

In this section, the data which was not required for analysis according to geographic sense was also removed. (Customer analysis of only one country)

(See Appendix B.1)

Lastly, the datatypes of all the required field were manipulated according to the formulas that would be applied on them Later.

At the end of this section, we had the relevant and crisp data for usage. (See Appendix B.2)

* **SECTION 2:** GENERATING RFM VALUE:

In this Section the RFM value is calculated using RFM analysis. It is one such popular customer segmentation technique that can help retailers maximize the return on their marketing investments.

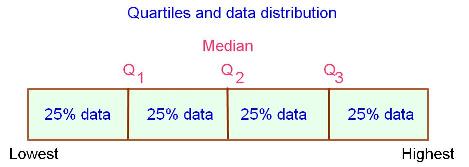
Under RFM analysis, each customer is scored based on three factors, namely Recency, Frequency, and Monetary value. (See Appendix B.3)

* **Recency:** Recency is the most important predictor of who is more likely to respond to an offer. Customers who have purchased recently from you are more likely to purchase again from you compared to those who did not purchase recently.
* **Frequency:**The second most important factor is how frequently these customers purchase from you. The higher the frequency, the higher is the chances of them responding to your offers.
* **Monetary:** The third factor is the amount of money these customers have spent on purchases. Customers who have spent higher are more likely to purchase based on the offer compared to those who have spent less.

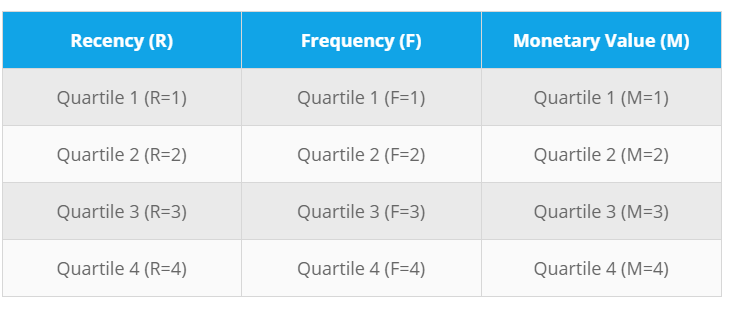
The above factors were calculated by analyzing the customer’s data like date of purchase, Number of items purchased and the amount associated with it.

* **SECTION 3 :**GENERATING SCORE OF CUSTOMER:

The RFM score is analyzed using the concept of Quartiles. The specific column data had division in four parts. Later each entry was given a specific number from 1-4 on basis of which quartile range it falls in. The diagram below explains the concept of Quartile distribution.



According to the analysis of the above concept the following score is calculated individually and then combined to attain a specific RFM score for each customer. (See Appendix B.4)



* **SECTION 4 :** USING CUT-OFFS FOR SEGMENTATION:

By combining these three scores, a final RFM score is calculated. The customers with the highest RFM score are considered to be the ones that are most likely to respond to their offers. In our example, the customers score will range from 111 to 444 (add up all these scores). Customers with a score of 111 are the best customers. (See Appendix B.5)

These cut-offs of these scores then help in profiling the customer to different segment. The cut-offs can change according to the business plans and goals.

I have used the following profile according to corresponding score

These segmentation portrays the real position of customer and management initiatives which can help them. Such analysis leads to customer satisfaction.

* **SECTION 5 :** VISUAL REPRESENTATION OF EACH PROFILE:

The total number of the customers are eventually now segmented into different segments for analysis. The representation can be in different forms.

Like bar graphs, pie charts, histograms, whatever is required by the management to take required decision for analysis.

Pie Chart

In the pie chart created (See Appendix B.6), The total customers, can now be seen in form of different segments.

TECHNOLOGIES USED

* Programming Language: Python & R language
* Operating System: Windows 8
* IDEs: Anaconda(Spyder) , R studio

**ANALYSIS**

During our initial days, we got the opportunity to research in various spheres within my fields of Data Science. It actually enhanced my knowledge on the upcoming advancements in this field.

Then very swiftly I moved on to understand new concepts in programming languages like Python. Which I later used for my project.

Initially I was directed to concentrate on the project but as I was progressing well in it, I understood various future prospects which I was not able to work on due to time constraint. But these ideas will eventually prove to be beneficial when I would extend my project in future.

I feel my primary strength for completing my project was my prior knowledge of R language. It helped me to quickly grasp the concepts of Panda Package in Python. I would certainly like to enhance my knowledge in much more such packages.

Now coming to skills, a person never thinks of that is adapting to corporate culture, working in a team etc. While working here I inculcated all of it quite successfully and got quite comfortable with the professional routine such as reporting my work on regular basis.

One of the strength that I developed during this training period was being optimistic about unfamiliar work within the given deadline.

It was a great experience for me as I learned a lot of new things during my internship. And made me realize that practical knowledge is very much important to learn anything.

**CONCLUSION**

Customer Analysis model allows for the effective allocation of marketing resources and the maximization of selling opportunities in any organization. Customer segmentation can also improve customer service and assist in customer loyalty and retention.

In the past years, technology has changed the nature of conducting analysis especially for a diverse and large data as of customers. In the old days it was done manually. Now that the computers and technology have penetrated the industry, automation has become the competitive advantage in today's world. We are now completely reliable on technology for every work.

This summer internship had helped me a lot to know about the working of IT companies in today's world and how to they withstand the high level competitive environment. The real meaning of team work, punctuality, work compilation, authority, responsibility was known to me during my project.

I had a great opportunity to build great skills of working and coordinating with team members and giving my best to the work allotted to me. I not only gained technical skills but also soft skills which are very important in every domain of life.

**BIBLIOGRAPHY**

Books

# Murtaza Haider .Getting Started with Data Science 1 Edition. Publisher:PEARSON,2016

Sandip Rakshit. R Programming for beginners. Publisher: McGraw Hill Education,2017

Sumita Arora.Computer Science With Python Textbook & Practical Book. Publisher: Dhanpat Rai,2018

Internet Sites

www.blastam.com/blog/rfm-analysis-boosts-sales,June 2018

www.eightleaves.com/2011/01/using-rfm-to-identify-your-best-customers,June 2018

www.w3schools.com/python,June 2018

www.python.org/downloads/windows,June 2018

www.conda.io/docs/download.html,July 2018

Encyclopedia Articles on Internet Data Base

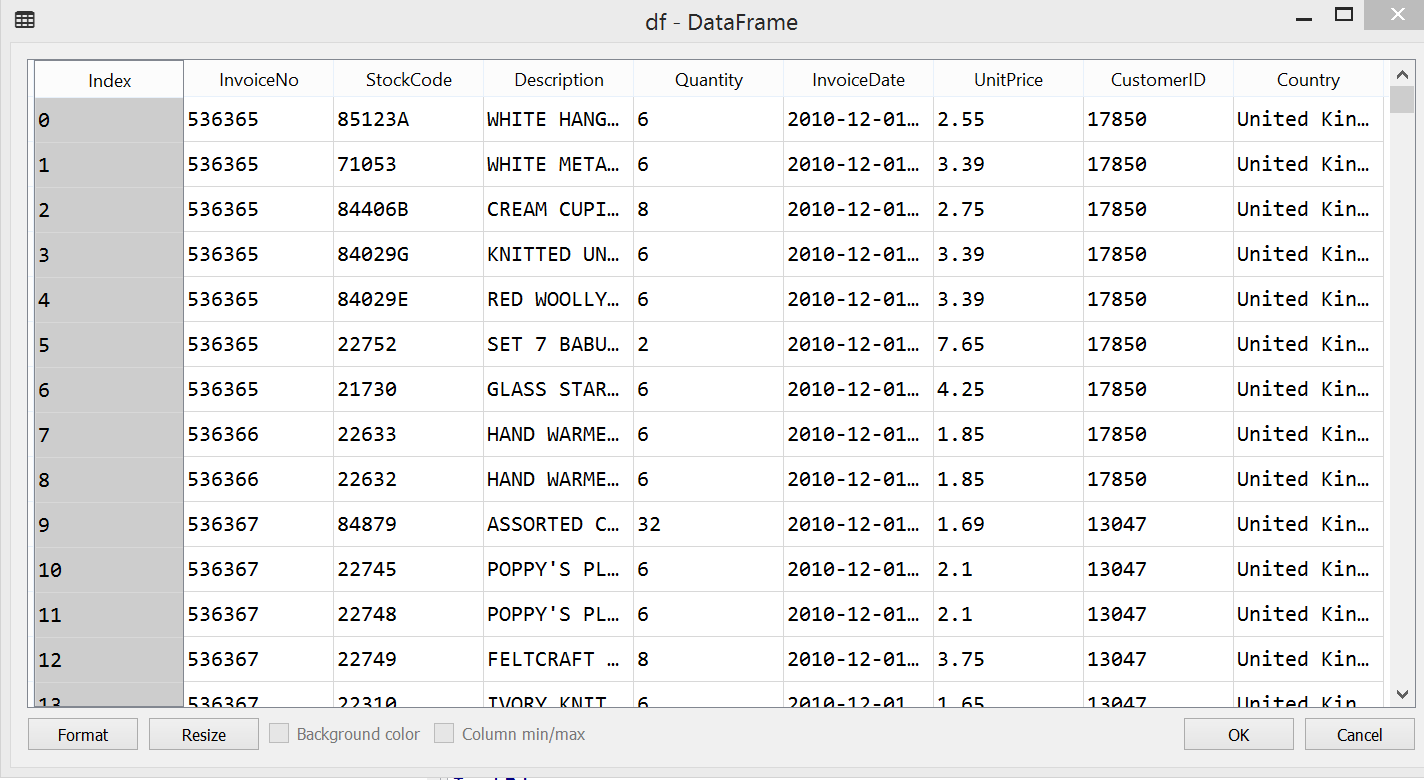
Hayashi, Chikio (1998-01-01). "What is Data Science? Fundamental Concepts and a Heuristic Example". In Hayashi, Chikio; Yajima, Keiji; Bock, Hans-Hermann;

International Council for Science : Committee on Data for Science and Technology. (2012, April). CODATA, The Committee on Data for Science and Technology. Retrieved from International Council for Science : Committee on Data for Science and Technology: http://www.codata.org/

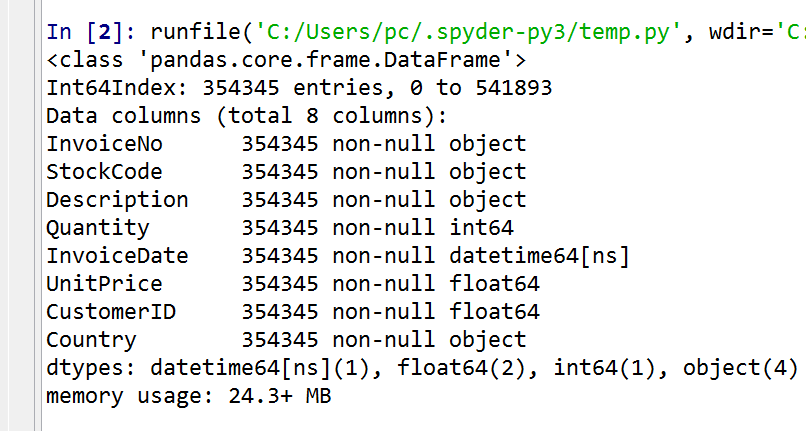
**APPENDICES**

**Appendix B: Screenshots of Different Sections**

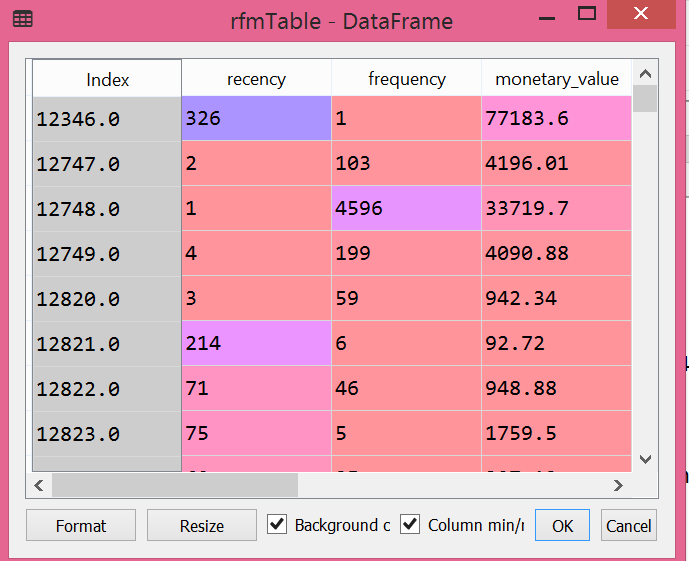
B.1 Data frame in Spyder after pre-processing

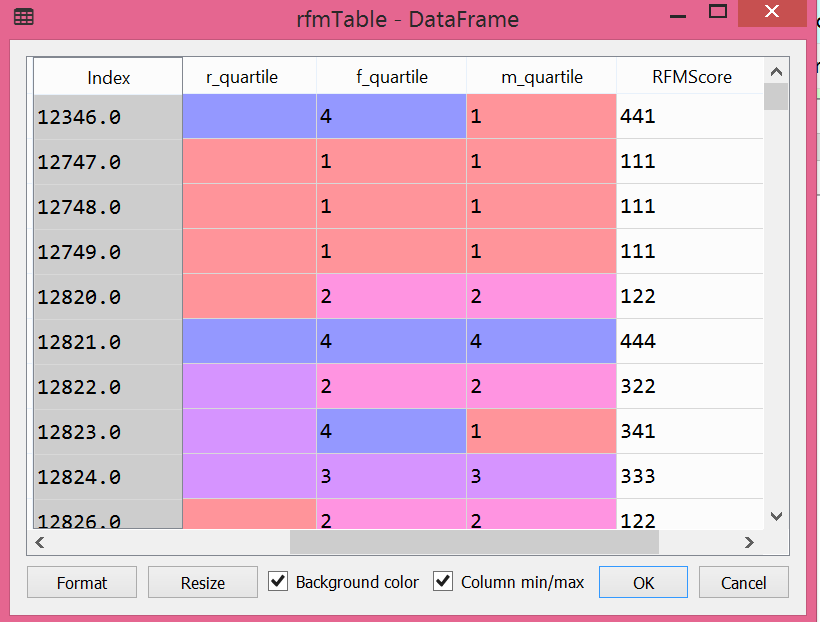


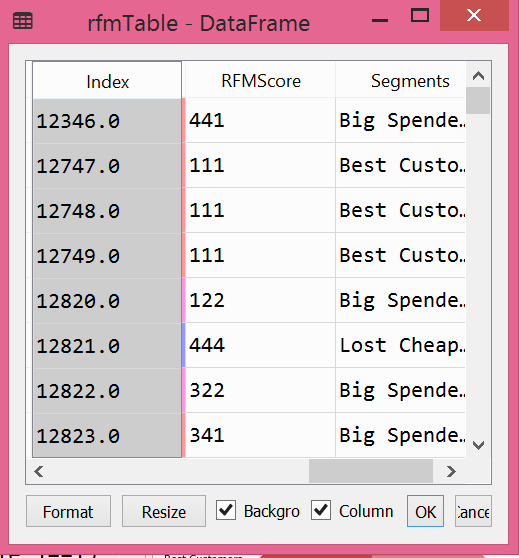
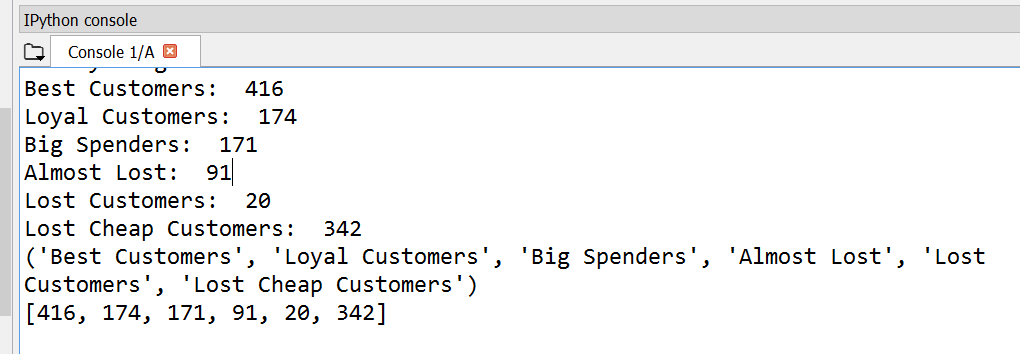
B.2 Data frame information on console



B.3 RFM Table



B.4 RFM Score for each customer

B.5 Segments according to Score

B.6 Pie chart of Analysis