Keerthana Sundaresan
Roll No. 2017103058

WEB PROGRAMMING CS7604

Restaurant Data Collection and Analysis using Python Crawler

Objective

To collect information such as name, cuisine, cost, and rating of restaurants as listed in popular food review apps such as **Zomato using Web Scraping in Python and apply ABC Analysis** on this collected data to categorize the restaurants by value for money and profitability.

Algorithm

- **Web Crawling**
    1. Import BeautifulSoup for web crawling.
    2. Construct request.get() by specifying URL for the HTML Parser.
    3. Declare class names and ID to pull information from and store the information as key-value pairs in a DataFrame.
    4. Use Pandas to store the DataFrame contents into a .csv file. This will be used for the analysis
- **ABC Analysis**
    5. Import the libraries pandas, matlablib.pyplot, seaborn and numpy.
    6. Form a sub-dataset with the cost, restaurant name and vote attributes.
    7. Find the additive cost for each restaurant by multiplying the unit cost with number of votes.
    8. Sort the sub-dataset in descending order of additive cost.
    9. Calculate the cumulative cost and percent spend.
    10. Divide the dataset into classes based on the ABC_segmentation() function defined.
    11. Analyse the new data and form conclusions.

Dataset

The dataset is constructed by performing web crawling using BeautifulSoup on the Zomato Bangalore website. This is stored in a .csv file. The various attributes include Cost for Two, Rating, and Number of Votes which are numeric values, as well as the Restaurant Name, Locality and Cuisine which are strings.

Note: The attributes used in the IEEE paper differ from the values in this project die to unavailability of banking transaction information of customers for any business. Therefore, a similar method is applied to a different dataset and accordingly, the output is generated.

Sample Code

- **Web Crawling**

```python
response = requests.get("https://www.zomato.com/bangalore/restaurants?page={0}".format(page_no), headers=headers)
content = response.content
soup = BeautifulSoup(content, "html.parser")
search_list = soup.find_all("div", {'id': 'orig-search-list'})
list_content = search_list[0].find_all("div", {'class': 'content'})
for i in range(0,15):
    res_name = list_content[i].find("a", {'data-result-type': 'ResCard_Name'})
    locality = list_content[i].find("b")

dataframe ={}
dataframe["rest_name"] = res_name.string.replace('\n', ' ')
dataframe["locality"] = locality.string.replace('\n', ' ')


df = pandas.DataFrame(list_restaurants)
df.to_csv("zomato_restaurants.csv")
```

- **ABC Analysis**

```python
def ABC_segmentation(perc):
    '''
    Creates the 3 classes A, B, and C based
    on quantity percentages (A-60%, B-25%, C-15%)
    '''
    if perc > 0 and perc < 0.6:
        return 'A'
    elif perc >= 0.6 and perc < 0.85:
        return 'B'
    elif perc >= 0.85:
        return 'C'
```

```python
data_sub = data[['cost_for_two','rest_name','votes']]
```

```python
data_sub['unit_cost'] = data_sub['cost_for_two']/2
data_sub = data_sub[['unit_cost','rest_name','votes']]
```

```python
data_sub['add_cost'] = data_sub['unit_cost'] * data_sub['votes']
data_sub = data_sub.sort_values(by=['add_cost'], ascending=False)
data_sub['cumulative'] = data_sub['add_cost'].cumsum()
data_sub['total_sum'] = data_sub['add_cost'].sum()
data_sub['run_percent'] =  data_sub['cumulative']/data_sub['total_sum']
data_sub['Class'] = data_sub['run_percent'].apply(ABC_segmentation)
```

Output

Thus, the output is a dataset classified according to the ABC Analysis, which based on value for money as well as popularity of the restaurants. The classification depicts the profitability and success rate of restaurants as listed on the Zomato website.

Plot
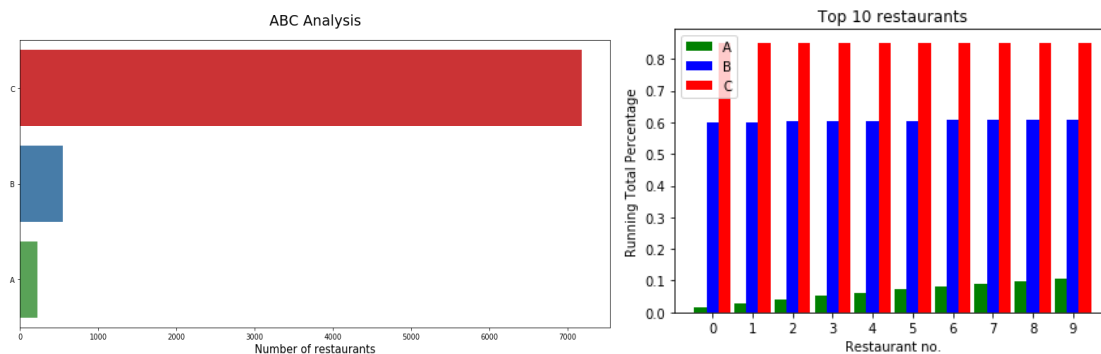
Top 5 restaurants in each class:

rest_name	add_cost	Class
Toit	16262000.0	A
Biergarten	11913600.0	A
AB's - Absolute Barbecues ...	11857600.0	A
Brew and Barbeque - A Microbrewery Pub ...	10787000.0	A
The Black Pearl ...	9060000.0	A

rest_name	add_cost	Class
1Q1	940000.0	B
Oye Amritsar	939950.0	B
Pablo's Gastrobar ...	939750.0	B
Smoor	926100.0	B
Foxtrot	925800.0	B

rest_name	add_cost	Class
Veena Stores	199875.0	C
Magnolia Bakery ...	199150.0	C
Chullah Chandigarh ...	198000.0	C
Urban Solace	198000.0	C
Maa Da Dhaba	197700.0	C

Inference

We find that the maximum number of restaurants with low value fall in the C category, i.e. C represents 90% of the restaurants with 15% of value. B represents 7% of the restaurants with around 25% of value, and A represents 3% of the restaurants with the highest value for money, 60%.

Reference

https://www.kaggle.com/danavg/abc-analysis-of-active-inventory

Customer Value Analysis Based on Python Crawler

Ming Liu, Yurui Du, Xifen Xu

School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China
E-mail: liuming@njust.edu.cn

Abstract: Customer value analysis is an important work in customer relationship management. In this paper, we use ABC classification, RFM (Recency, Frequency, Monetary) model and K-means clustering method to analyze the customer value. First, we use python language to compile a crawler program to collect the data of transaction records from an enterprise's customer information management system. Then, by using different methods, we obtain different classifications of the customers. The test results demonstrate that the proposed methods can provide constructive suggestions to manage the customer value in practice, although they are easy and basic.

Key Words: Python, Customer Value, ABC Classification, RFM Analysis, Clustering

1 INTRODUCTION

Under the background of big data era, the focus of corporate marketing has shifted from product-centered to customer-centered, and companies have gradually recognized the importance of customers and services [1]. Customer relationship management has become the core issue of the enterprise. Customer value classification is one of the key issues in customer relationship management, and it is an important basis for decision makers to optimize marketing resource allocation [2]. Through the results of customer classification, the managers can understand the different value attributes of the customer. Based on the analysis of the current value and potential value of the customer, we can divide the customers into different groups according to their different value. It is a critical way to achieve the ultimate goal of maximizing corporate profits by analyzing the difference value of different groups of customers, formulating personalized marketing plans, and applying customer classification results to enterprise customer relationship management practice [3].

Recently, many scholars have used different classification methods to mine and subdivide customer data sets, and provided several constructive strategies to guide enterprise practice. For example, Mosavi and Afsar (2018) focused on the Tejarat Bank branches in Iran and systematically integrated several data mining techniques and management issues to analyze customer value [4]. Daoud et al. (2015) conducted a case study of applying RFM model and clustering techniques in the sector of electronic commerce with a view to evaluating customers' values so as to achieve maximum benefit and a win-win situation [5]. Lu et al. (2018) used SPSS statistical software direct RFM point model for the customer segmentation and then used the nested classification to distinguish the customer value by RFM score to design different marketing strategies for

online store customers [6]. Ge and Chen (2016) developed an operational indicator system to measure customer value and proposed several suggestions to effectively manage customer relationship [7]. Sarvari et al. (2015) determined the best approach to customer segmentation and extrapolated associated rules based on recency, frequency and monetary (RFM) considerations as well as demographic factors [8]. Chang and Ho (2017) built a two-layer clustering model for mobile telecom customer analysis and used it to combine with the product to assist staff to implement effective marketing [9]. Singh (2015) created a risk-adjusted recency, frequency, monetary value (RARFM) score for each customer to identify the under- lying demographics and behavioral characteristics [10]. Although these scholars have used relevant algorithms to classify and calculate the value of customers, they have not integrated different customer values. Therefore, this paper combines python crawler to obtain relevant data, then applies three customer segmentation algorithms to offer more instructive management strategies.

With the explosive growth of network information, it is difficult for us to get the required information quickly and accurately from the massive information. Under this background, we use python web crawler technology with its powerful ability of automatically extracting web information to program and crawl the required data of customer transaction records. After cleaning the customer transaction records, we classify the customers by using ABC classification method and RFM analysis model, and then apply K-means clustering based on the analysis results of RFM model, which is the main basis for evaluating the value of all kinds of customers. Finally, according to the test results, we propose several directory marketing strategies of customers with different value to realize the precise marketing of enterprises. As far as we know, most of the existing literature only use one related classification method for customer value segmentation. In this study, we use ABC classification, RFM model and K-means clustering to subdivide customer value on the basis of python crawler data. Based on the classification results of RFM model,