

Analysis of Consumption Behavior of Campus Cards Based on Python

Chongli Zhong, Zhenyu Cao, Xiaolong Bai

Xi'an Eurasia University, Xi'an City, Shanxi Province, China

Keywords: University student card, Python, Cluster analysis, Consumer behavior

Abstract: The university card is an integrated information system, which includes campus identity-confirmation, consumer finance, and other digital functions. While the system is providing high-quality services for teachers and students, the historical data can help the university to improve quality services. This article uses mathematical models and data mining to explore student life behavior. The results do not only support university improve campus information system but also help leaders make strategic decisions.

1. Introduction

With the development of scientific and technological information and networks, mankind has entered the era of informatization. And colleges and universities are also gradually deepening their digital and information construction. The “University Card” system is highly integrated with various types of consumption, small payments, access control, conference sign-in management, library identity authentication, and borrowing, and student status management. Sharing data management through a common identity authentication mechanism. Therefore, the “campus card” system soon became an important part of the university's informatization construction.

Over time, a large number of consumption and behavior data will be accumulated in the “University Card” system. Through these data, to a certain extent, students can understand the digital trajectory of students on campus, and grasp the laws of student behavior on campus. Resource optimization and allocation are of great significance. For example, the university library provides many intelligent robots. When the student scans the university card on the machine, which can automatically identify the user's background information and help them to borrow books, unlock the computer, and reserve the study room. In university accommodation, students use the card to enter the public bathroom. The system will automatically charge the water bill through the card. The method can help the university save water.

In addition, “poverty alleviation” and “out of poverty” have become the focus of politics in recent years. The poverty living conditions of college students have received more and more attention from government departments. The identification of poor students has a direct impact on poverty alleviation work. The accurate distribution of poverty support is closely related to the learning life of poor students. The consumption data in the campus card plays a key role here. The Education Foundation of Nanjing University of Science and Technology has launched the “Ruihua Heartwarming Card” project on the information obtained from the campus card, which specifically assists poor undergraduates with “substances of food and clothing”. The implementation of this project is to take over 16,000 undergraduates from the whole university to swipe records from mid-September to mid-November, analyze all records, and finally lock more than 500 “quasi-assistance targets”.

This project takes the card transaction flow data of a domestic university in April 2019 (a total of 30 days) as the research object. Through the cleaning, integration, visual analysis and cluster analysis of the original data, the conclusion is drawn to help the school formulate relevant adjustment measures. In order to optimize resources reasonably, at the same time identify “poor students” so that the school can provide certain help. Here, we first conduct preliminary exploration and data preprocessing on all the data; secondly, we analyze the canteen data to obtain a pie chart of the number of meals in each canteen and a graph of meal times on weekdays and non-weekdays; Consumption data and behavior analysis, we get a bar chart of per capita consumption comparison

of different gender student groups in a major; finally, we selected 16 indicators for cluster analysis. After the analysis of the clustering model, it provides reference suggestions for the school's logistics department, appropriately optimizes the cafeteria resources, and provides an effective reference for the school's "precision poverty alleviation".

2. Data Preprocessing

The historical data comes from a Chinese university in April 2019. The first data table contains the student background information, which includes Card Number (CardNo), Sex, and Major fields. The data consists of 4341 records. Figure 1 shows an example of the data table.

	CardNo	Sex	Major
0	180001	男	
1	180002	男	18
2	180003	男	international finance
3	180004	男	
4	180005	男	

Fig.1 Example of Student Background Information Table

The second data table contains the university card usage data, which include:

- Index: consumption number
- CardNo: university card number
- PeoNo: university card authentication number
- Date: consumption date
- Money: amount of consumption
- FundMoney: recharge amount
- Surplus: account balance
- Type: card type
- TermSerNo: consumption tracking number
- conOperNo: product tracking number
- Dept: consumption location

Fig.2 Shows an Example of the Data Table, Which Includes 12 Columns and 519367 Rows.

	Index	CardNo	PeoNo	Date	Money	FundMoney	Surplus	CardCount	Type	TermSerNo	conOperNo
0	117342773	181316	20181316	2019/4/20 20:17	3.0	0.0	186.1	818		NaN	NaN
1	117344786	181316	20181316	2019/4/20 8:47	0.5	0.0	199.5	814	consumption	NaN	NaN
2	117346258	181316	20181316	2019/4/22 7:27	0.5	0.0	183.1	820		NaN	NaN
3	117308066	181317	20181317	2019/4/21 7:46	3.5	0.0	50.2	211		NaN	NaN
4	117309001	181317	20181317	2019/4/19 22:31	2.5	0.0	61.7	209		NaN	NaN

Fig.2 Example of University Card Usage Table

Firstly, according to the project requirements, the research group explored and cleaned data information. Then, the members associate both data tables that can help people easily analyze the behavior of students from different sex and majors. Research group uses the CardNo field to associate the two tables and preprocess the data from the following 5 perspectives:

① Delete the two fields in the table "data2": TermSerNo, conOperNo. Because there are a lot of missing values in these two fields.

② Check whether there is an abnormal value in this field by counting the frequency of Sex in "data1". No outliers were found.

By judging whether the values of the Money, FundMoney, Surplus, and CardCount fields in “data2” are ≥ 0 , detect whether there are abnormal values in these fields.

④ Check whether there are duplicate values in the two data tables. After reviewing, there are no duplicate values, so no duplicate values are processed.

⑤ Change the date of “data2” to “time type” and extract the “year”, “month”, “week”, “day”, “hour”, “minute”, and “second”.

After the above preprocessing, the final combined data table is shown in Figure 3.

CardNo	Sex	Major	Index_y	PeoNo	Date	Money	FundMoney	Surplus	CardCount	Type	Dept	year	month	weekday	day	hour	minute	second
0	180001	男	18国际金融	117331517	20181	2019/4/21 18:30	7.0	0.0	28.4	206		2019	4	7	21	18	30	0
1	180001	男	18国际金融	117341866	20181	2019/4/22 9:40	3.5	0.0	24.9	207		2019	4	1	22	9	40	0
2	180001	男	18国际金融	117154618	20181	2019/4/10 16:42	11.0	0.0	2.7	189	consumption	2019	4	3	10	16	42	0
3	180001	男	18国际金融	117200032	20181	2019/4/15 11:43	7.0	0.0	43.2	192		2019	4	1	15	11	43	0
4	180001	男	18国际金融	117206146	20181	2019/4/15 7:50	2.5	0.0	50.2	191		2019	4	1	15	7	50	0

Fig.3 Preprocessed and Merged Data

3. Dinning Behavior Analysis

We analyze the dining habits, include breakfast, lunch, and dinner. University serves breakfast from 6:00 to 9:00. The lunchtime between 11:00 and 14:00. The dinner time starts from 17:00 to 20:00. The data didn’t consider the festivals. University restaurants provide the meal in 7 days a week. The university has 5 public restaurants and 1 teacher restaurant, which serves traditional Chinese foods. The following pie charts describe the number of diners for breakfast, lunch, and dinner. There are obvious differences in the dining places for students at breakfast, lunch, and dinner.

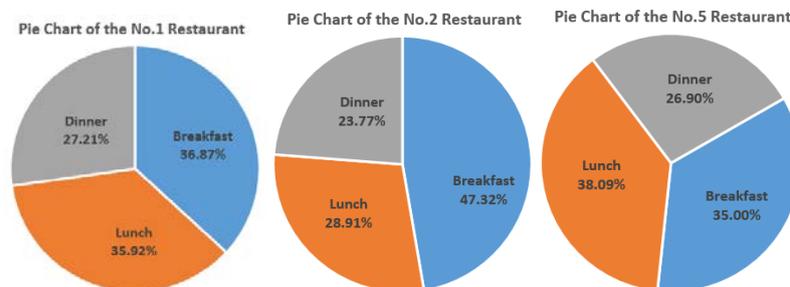


Fig.4 (1) Pie Chart of the No.1/2/5 Restaurant

The No. 1 university restaurant, the No.2 university restaurant, and the No.5 university restaurant have the same trends. But the No.1 and No.2 university restaurants have better sales at breakfast. No. 5 university restaurant has better breakfast sales than dinner.

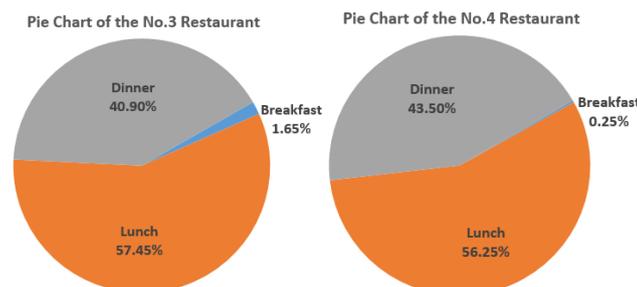


Fig.4 (2) Pie Chart of the No.3/4 Restaurant

No.3 and No.4 university restaurants have the same characteristics. There are few people who

purchase breakfast. But the restaurants have better sales data at lunchtime.

Pie Chart of the Teacher Restaurant

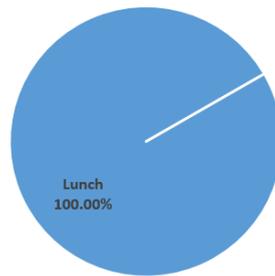


Fig.4 (3) Pie Chart of the Teacher Restaurant

The teacher restaurant is a special case. The dataset only displays lunch sales data. It looks like teachers only choose to have lunch in the workplace. We assume they usually eat breakfast and dinner at home. To avoid waste, the teacher restaurant doesn't serve breakfast and dinner.

The article analyze the data base on the resturant sales data. The following texts describes the consumer behavior of the breakfast, lunch, and dinner.

Pie Chart for Breakfast Consumption in Each of Restaurant

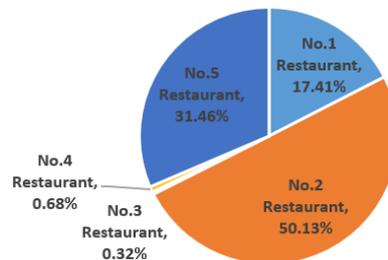


Fig.5 (1) Pie Chart for Breakfast Consumption in Each of Restaurant

The breakfast business can be described as “one-third of the world”. The students have a preference for breakfast in the No.2 restaurant, followed by the No.5 restaurant. The total number of people eating in both restaurants accounted for more than 80%, while the number of people going to the No.1 restaurant for breakfast was less, only 17 %, Few people choose to go to the No.3 and No.4 restaurants for breakfast.

Pie Chart for Lunch Consumption in Each of Restaurant

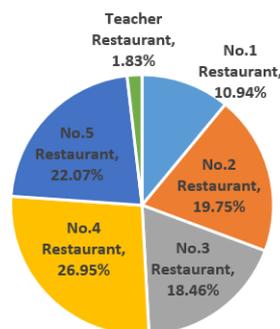


Fig.5 (2) Pie Chart for Lunch Consumption in Each of Restaurant

The lunch business can be described as “sharing the world”. The No.2, No.3, No.4, and No.5 restaurants are not much different, each accounting for about 20%, of which the No.4 restaurant is more prominent, accounting for about 27%. The No.1 restaurant is a bit inferior, only about 10%. Students rarely choose to go to the teacher restaurant for lunch.

Pie Chart for Dinner Consumption in Each of Restaurant

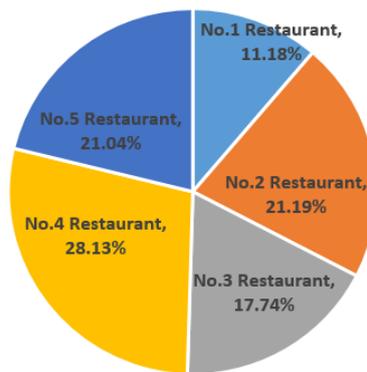


Fig.5 (3) Pie Chart for Dinner Consumption in Each of Restaurant

The dinner is almost in the state of “sharing the world”, and there is not much difference between the canteens. Among them, the No.4 restaurant is the best with 28% market share, the No.2, and No. 5 restaurants are second, and the No.3 restaurant is the first. No.1 restaurant was the worst, accounting for only 11%.

Based on the above analysis of dining behaviors, we can make the following suggestions for the school's logistics group:

(1) No.3 and No.4 restaurants only need to provide lunch and dinner. Canceling the sale of breakfast in these two canteens can save more resources.

(2) The teacher restaurants only need to provide lunch, but the amount is not too large, which is basically enough for teachers' consumption needs.

(3) The sales of the morning, middle, and evening meals in the No.1 restaurant are not particularly good. It is recommended to improve the dining environment or add some new dishes and more promotional activities to increase its appeal to students.

(4) The No.2 and No.5 restaurants sell three meals a day, which indicates that the two canteens are extremely attractive to students. They can increase the investment in these two canteens to make them better.

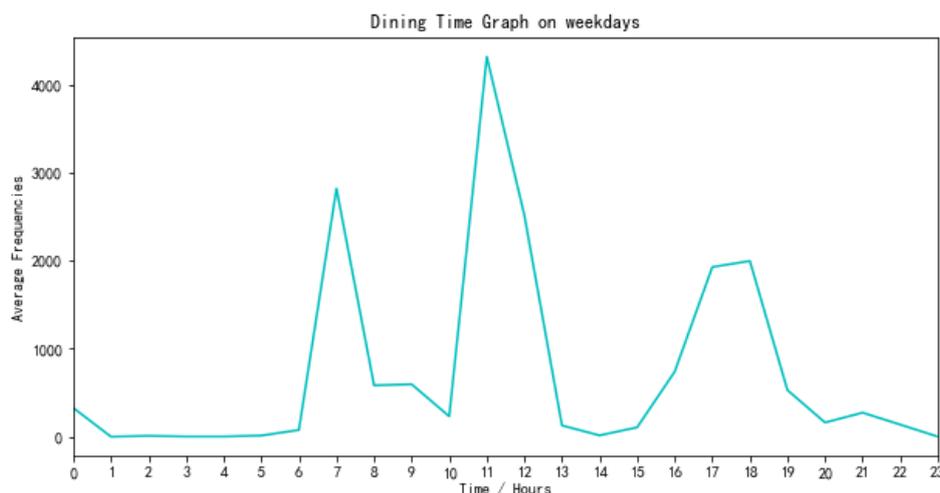


Fig.6 (1) Dining Time Graph on Weekdays

From the mealtime curve chart on weekdays, it can be seen that there are three peak periods in the morning, middle, and evening. The breakfast peak is around 7:00 and the lunch peak is around 11:30. Both peaks are sharp, that is, students The dining time is very concentrated. The dinner peak is relatively flat and lasts a long time, and the peak time is from about 17: 00-18: 30.

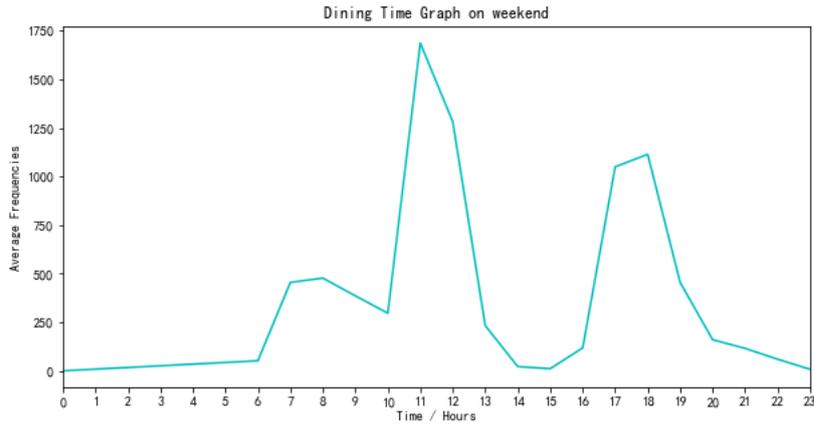


Fig.6 (2) Dining Time Graph on Weekend

From the non-working day dining time curve, it can be seen that there are two obvious peaks, namely lunch and dinner, and breakfast can only be regarded as a small peak, and the duration is longer, the peak is significantly smaller than lunch and dinner. Specifically, the breakfast peak is around 8:00, the lunch peak is around 11:30, and the dinner peak is 17: 00-19: 00. In summary, non-working days, because everyone sleeps lazily, the peak of breakfast is postponed more than usual, and the number of people eating is much lower. It is estimated that most people choose not to eat breakfast during the holiday, and directly combine brunch and lunch. There was no significant difference between lunch and weekdays. The number of people who eat dinner during the holiday will be much larger than usual. It is estimated that everyone is more inclined to engage in some dinner activities during the holidays.

Based on the above analysis, we can make the following suggestions:

- (1) All canteens should be fully prepared before the peak, to ensure that services are provided quickly and easily during peak dining hours.
- (2) During non-working days, the restaurant should reduce the supply of breakfast to avoid wasting resources. In addition, the staff can also enjoy the benefits of the holiday, open an hour later than before. Prepare more ingredients for dinner to meet more dining needs.

4. Consumer Behavior Analysis

Here, we extract all the consumption data of 18-level students for analysis, calculate the per capita card swiping frequency and per capita consumption amount in April, and summarize the consumption characteristics of students of different majors and different genders.

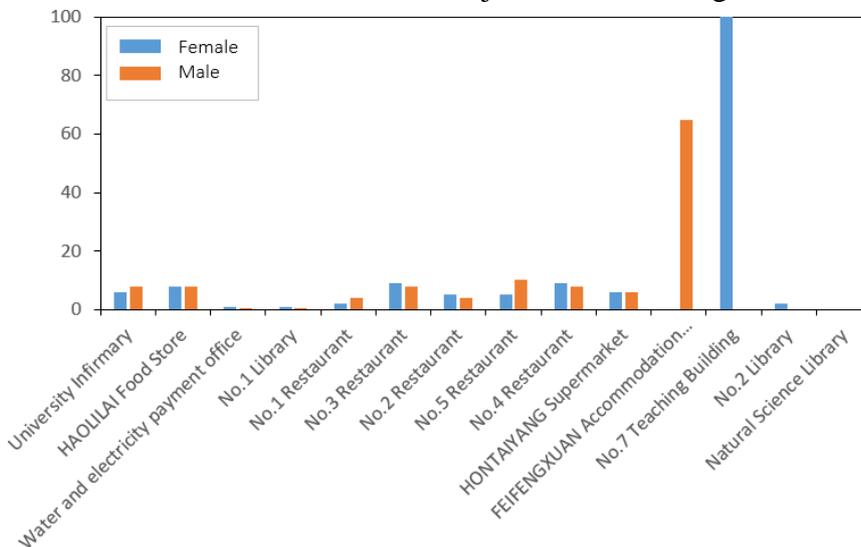


Fig.7 Students Consumption Base on Genders from a Major

We selected the student records of a certain major for analysis, and we can see from Figure 7 that the boys in this major especially like to spend in the seventh teaching building, but this is a place where girls never come; Hollyland Food Store and Red Sun Supermarkets are equally popular shops; boys and girls are very different when it comes to dining. Boys like to eat in the 345 or 535 restaurants.

According to the students' overall campus consumption behavior, select appropriate characteristics, construct a clustering model and analyze the consumption characteristics of each type of student group, so that the school can determine the student's economic status according to the results and take certain assistance measures.

The so-called cluster analysis is a group of statistical analysis techniques that divides research objects into relatively homogeneous clusters. Cluster analysis is an exploratory analysis. In the process of classification, people do not have to give a classification standard in advance. Cluster analysis can start from sample data and automatically perform classification. Different methods used in cluster analysis often lead to different conclusions. The number of clusters obtained by different researchers on the same set of data may not be the same. This paper uses the k-means clustering algorithm, which is an iterative clustering algorithm. The steps are to randomly select K objects as the initial cluster center and then calculate each object and each seed The distance between the cluster centers. Each object is assigned to the cluster center closest to it. The cluster centers and the objects assigned to them represent a cluster. For each sample assigned, the clustering center of the cluster is recalculated based on the existing objects in the cluster. This process is repeated until a certain termination condition is met. The termination condition may be that no (or minimum number) objects are reassigned to different clusters, no (or minimum number) cluster centers change again, and the squared error and local minimum. Here, we constructed 16 indicators such as the average consumption amount of breakfast / middle / dinner, the total consumption amount of each consumption location, and the average mealtime of breakfast / middle / dinner for clustering. Starting from the school's demand goal, we will All students are divided into 3 types and cluster model analysis is performed. The core code is shown in Figure 8:

```

137 # Mean-variance normalized:
138 from sklearn.preprocessing import StandardScaler # Import Means-Variance Normalization Module
139 scaler = StandardScaler()
140 scaler.fit(X)
141 data_prpl=scaler.transform(X)
142
143 # Data Aggregation
144 from sklearn.cluster import KMeans
145 kmeans_model = KMeans(n_clusters= 3, max_iter = 100, random_state=12345)
146 kmeans_model.fit(data_prpl)
147 fit_label = kmeans_model.labels_
148 center=kmeans_model.cluster_centers_
149 Fs=pd.Series(fit_label, index=task3_XL[' CardNo' ])
150 Fs = pd.DataFrame(Fs)
151
152 # Export Data from Center Table
153 center = pd.DataFrame(center)
154 center.to_csv('center.csv', encoding='gbk')
```

Fig.8 K-Means Clustering Core Code

5. Conclusion

After K-means clustering analysis, we get the cluster center table. From this table, we will draw the following conclusions: The student consumption group is divided into three categories: “0” represents the low consumption group, and “1” represents the high consumption group. “2” represents the middle consumer group. The characteristics of these three groups of people are:

Low-consumption people prefer to buy food in the No.3 restaurant, the No.5 restaurant, and HAOLILAI food stores. They consume more on working days and may leave campus to do part-time work on non-working days, so the consumption on non-working days is extremely low;

High-consumption people prefer to consume in the No.1 restaurant, the No.2 restaurant, the No.4 restaurant, the No.5 restaurant, and the HONGTAIYANG supermarket, and their consumption will be lower on weekdays, but not higher on weekdays. The preliminary inference may Consumption is

increased due to the need to party or relax in non-working days;

The No.3 restaurant and HAOLILAI food store, which are more preferred by the middle-consumer population, buy food on the off-weekdays.

Based on the results of the above-mentioned cluster analysis, schools can initially lock in some “quasi-poor students” and give them more assistance measures.

Finally, from the perspective of mealtime, all people prefer to eat breakfast at 7 o'clock, but also at 8 or 9 o'clock, but relatively few. Few people eat breakfast before 7 o'clock, and a small number of low-spending people Eat breakfast; lunchtime is usually at the peak of 11 and 12 o'clock, few people eat lunch after 12:30; some people will choose not to eat dinner, but most still choose to eat at 17 Dining at -18 o'clock, very few after 19 o'clock.

From the above conclusions, we can also explore the habits of students to help them make decisions and optimize the opening hours of the canteen to reduce labor costs.

In summary, the campus card contains a lot of valuable data and information. Through the analysis from different perspectives, many conclusions can be drawn to facilitate teachers to provide better teaching services for students, and also for school policymakers to make timely policies Orientation to optimize resource allocation.

References

- [1] Fan, S. , Li, P. , Liu, T. , & Chen, Y. . (2015). Population Behavior Analysis of Chinese University Students via Digital Campus Cards. 2015 IEEE International Conference on Data Mining Workshop (ICDMW).IEEE.
- [2] Zhang, H. L. , Zhang, Y. , Zhong, W. F. , & Li, M. Y. . (2014). The study of online recharging for campus cards based on the third-party payment platform. *Advanced Materials Research*, 926-930, 2253-2258.
- [3] Liu, Y. , Hu, M. , & Lu, X. . (2016). Social Frequency Analysis of University Students via Digital Campus Cards. *International Conference on Intelligent Human-machine Systems & Cybernetics*. IEEE.
- [4] Huangfu, D. , & Xu, T. . (2011). The design and research of Campus Card SMS platform based on MAS. *International Conference on Artificial Intelligence*. IEEE.