Study of Student Satisfaction with University Canteen Services

U Mahima(20112035), Vidhi Saha(20112036), Yash Lucas(20112037)

Data Science Department, Christ (Deemed to be University), Lavasa

Sampling Techniques

Prof. Shine Raju Kappil

Abstract

Topic Details and Key Objectives

Customers want high quality food, affordable pricing and good service. We intend to study and analyze the student satisfaction with the pricing, quality, ambience, variety and service of the four canteens of Christ (Deemed to be University), Lavasa, i.e, Chopsticks, Cafe by valley, Cafe Concerto, MBA canteen.

Population under Study and method of sampling

The population under study are the students of the university and we will be using simple random sampling to select our sample.

Sample Size

Our population size (total number of students of Christ university) is 1250. We are considering our confidence level to be 92%, margin of error to be 8% and precision to be +/-8% and upon calculation considering the above mentioned factors, our sample size is approximately 110 students.

Analytical Methods to be used

For statistical analysis, we will be using the Kruskal Wallis H Test to assess if there are statistically significant variations between two or more groups of an independent variable on an ordinal dependent variable. We will be developing a regression model in order to find relation between the dependent variable, pricing and independent variables, quality and service. To maximize insight into the dataset, its underlying structure and to check correlation between the variables, visualizations using matplotlib , seaborn and pandas will be performed. (*Kruskal-Wallis H Test in SPSS Statistics | Procedure, Output and Interpretation of the Output Using a Relevant Example.*, n.d.)

Method of Data Collection
Data will be collected through Google forms.

Expected Outcome
We are expecting that greater level of satisfaction with quality, ambience, variety and service will lead to greater level of satisfaction with pricing. We assume that overall if the students are provided with good food, good facilities most of the students will be satisfied.

Introduction

The university canteen is a critical component of the seamless progression of teaching and scientific research. It is about students' urgent needs as well as the harmony and stability of university.Students want high-quality meals at a cheap price, as well as prompt service. We want to research and assess student satisfaction with the four canteens of Christ (Deemed to be University), Lavasa, namely Chopsticks, Cafe by Valley, Cafe Concerto, and MBA canteen, in terms of cost, quality, ambience, variety, and service.Canteens should not only provide food, but they should also guarantee that students' diets are nutritious and safe, as well as encourage student health. As a result, the quality of university canteens and student satisfaction with the cafeteria are strongly tied to the university's stability. Investigating and evaluating university students' content with the cafeteria may help canteen staffs better know their customers' perspectives and wants, as well as enhance management and stimulate logistical self-construction and make the necessary modifications to the canteen.

Review of literature

Cardozo was the first one to mix consumer happiness with marketing, and his concept sparked a lot of debate in the academic community. Customer satisfaction, according to Howard (1969), is a type of inner sensation that customers get when they consider the worth of purchasing a certain product. The advantages of purchasing a product are exactly proportional to the cost.Almanza et al. (1994) concluded that the quality of meals, the location of canteens, the sanitation of meals, and the pricing of meals were the most essential aspects determining students' satisfaction with university canteens. Meyer et al. (1998) conducted a survey on a canteen in a middle school in the United States, and found that the aesthetics of the menu design, the presentation of the food, the delicious taste, the quality and features of the canteen staff, the cost of the food, and the great dining environment are the main factors affecting canteen satisfaction.

Methods and methodology

Population: The population under study are the students of the university. The population size is 1250.

Sampling Technique: Simple random sampling was used to select the sample. Samples were chosen randomly from the population, each member of the population has an equal chance of being selected.

Sample size determination: To determine the sample size "Cochran's formula" was used.

The formula developed by William G. Cochran is:
$n_0 = z^2 (pq) / e^2$

where,
n_0 = sample size
z = standard error associated with the chosen level of confidence (here, 1.75)
p = variability/standard deviation
q = 1 - p
e = acceptable sample error/ margin of error

Here we calculated the sample for proportions and found it to be 120 students.

Cochran's formula for finite population correction for proportions is:
$n = n_0 / 1 + [(n_0-1) / N]$

where,
n = reduced sample size
n_0 = sample size
N = population size

Here we calculated the reduced sample according to our population and found it to be 110 students.

Working:

$$n_0 = z^2(pq)/e^2$$

$$= \frac{(1.75)^2 (0.5 \times 0.5)}{(0.08)^2}$$

$$= \frac{3.0625 \times 0.25}{0.0064}$$

$$= \frac{0.765}{0.0064}$$

$$\Rightarrow 120.$$

For 1250 population.

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$

$$= \frac{120}{1 + \frac{(120 - 1)}{1250}}$$

$$= \frac{120}{\frac{1250 + 119}{1250}}$$

$$= \frac{120}{1.0952}$$

$$\Rightarrow 110$$

Statistical tools used:

- SPSS (IBM): To perform Kruskal-Wallis H Test to determine if two or more groups of an independent variable on an ordinal dependent variable have statistically significant differences.
- Sample Size Determination: Using Cochran's formula.
- Logistic Regression: Based on earlier observations of the recorded data set, predict a binary conclusion, i.e. yes or no in terms of customer satisfaction.

Hypothesis (used in Kruskal-Wallis H Test):
n_0 (null hypothesis): There is no statistically significant difference in the medians of the independent and dependent groups.
N_1 (alternate hypothesis): There is statistically significant difference in the medians of the independent and dependent groups.

Relationship between dependent and independent variables: Upon Kruskal-Wallis H Test we found that independent and dependent variables have no statistically significant difference.

Reason to choose dependent and independent variables: We chose price as dependent variable and hygiene, ambience, variety and service as independent variables as these are the main factors that could determine the customer satisfaction with food services. We are trying to establish a positive relationship between student satisfaction with price with satisfaction with the independent variables.

Results and Analysis

EDA(Exploratory Data Analysis)

Step 1: Importing the dataset

```
1 survey=pd.read_csv("final_response.csv")
```

```
1 survey
```

Region	How much would you rate the hygiene of canteens? [Chopstics]	How much would you rate the hygiene of canteens? [Cafe Concerto]	How much would you rate the hygiene of canteens? [MBA Block]	How much would you rate the hygiene of canteens? [Cafe by valley]	...	Do you think the prices are fair for the types of food sold? [Chopstics]	Do you think the prices are fair for the types of food sold? [Cafe Concerto]	Do you think the prices are fair for the types of food sold? [MBA Block]	Do you think the prices are fair for the types of food sold? [Cafe by valley]	Do you have the ability to take a 2nd serving of your choose meal?	Is the food served always hot enough?	Do you get the type of food as per your region?	Do you get place to sit in canteen?	How much average do you spend on food every day?	Ar sati wit o ser
North India	1 (worst)	3 (ok)	3 (ok)	4 (good)	...	4 little expensive	4 little expensive	3 up to mark	5 very expensive	no	yes	yes	yes	300	
South India	3 (ok)	3 (ok)	3 (ok)	4 (good)	...	3 up to mark	3 up to mark	4 little expensive	5 very expensive	yes	yes	yes	yes	150	
South India	3 (ok)	4 (good)	4 (good)	4 (good)	...	4 little expensive	4 little expensive	3 up to mark	4 little expensive	yes	yes	yes	yes	90	
Central India	3 (ok)	4 (good)	5 (excellent)	4 (good)	...	3 up to mark	3 up to mark	2 cheap	3 up to mark	no	yes	no	yes	150	
North India	3 (ok)	4 (good)	4 (good)	4 (good)	...	3 up to mark	5 very expensive	3 up to mark	5 very expensive	yes	yes	no	no	200	

Step2 : Checking the rows and columns

```
1 print(survey.shape)
```

```
(163, 30)
```

This shows that the dataset has 163 rows and 30 columns.

Step3: Checking the datatype of all the columns

```
1 survey.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 163 entries, 0 to 162
Data columns (total 30 columns):
 #    Column                          Non-Null Count   Dtype
---   ------                          --------------   -----
 0    Timestamp                       163 non-null     object
 1    Department                      163 non-null     object
 2    Year                            163 non-null     object
 3    Region                          163 non-null     object
 4    rate_of_hygein_chopstics        163 non-null     object
 5    rate_of_hygein_cafe_concerto    163 non-null     object
 6    rate_of_hygein_mba              163 non-null     object
 7    rate_of_hygein_cafe_by_valley   163 non-null     object
 8    ambience_chopstics              163 non-null     object
 9    ambience_Cafe_Concerto          163 non-null     object
 10   ambience_MBA                    163 non-null     object
 11   ambience_Cafe_by_valley         163 non-null     object
 12   variety_chopstics               163 non-null     object
 13   variety_Cafe_Concerto           163 non-null     object
 14   variety_MBA                     163 non-null     object
 15   variety_Cafe_by_valley          163 non-null     object
 16   service_chopstics               163 non-null     object
 17   service_Cafe_Concerto           163 non-null     object
 18   service_MBA                     163 non-null     object
 19   service_Cafe_by_valley          163 non-null     object
```

Step 4: Summary Statistics of price column

```
1 survey.describe()
```

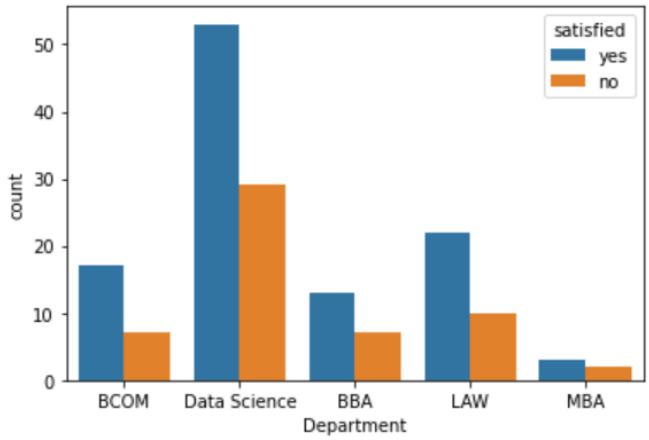
	spend_on_food
count	163.000000
mean	314.319018
std	192.401737
min	50.000000
25%	200.000000
50%	300.000000
75%	400.000000
max	1500.000000

Step 5: Data Visualization

For data visualization , imported the following libraries:

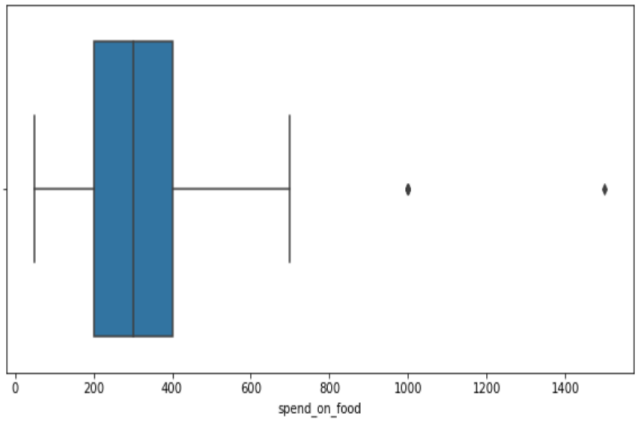
```
1 import matplotlib.pyplot as plt
2 import seaborn as sns
```

Graph 1: Relationship between different departments and their satisfaction with the overall food service of the university.



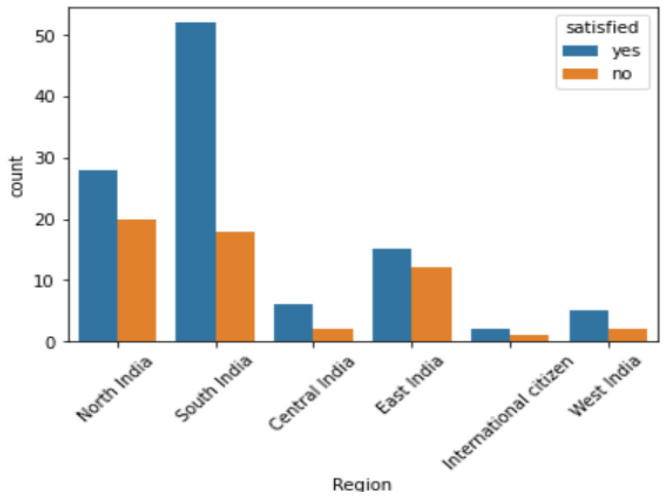
Result : The above graph shows that the department of Data Science is highly satisfied with the overall food service of the university and the department of MBA is least satisfied .

Graph 2: Distribution of average daily expenditure on food.



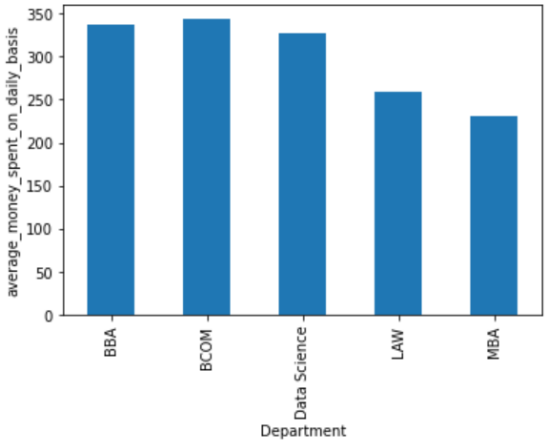
Result : The above boxplot shows that daily Rs. 200-400 is being spent by the students on food.

Graph 3: Relationship between students of different regions and their satisfaction with the overall food service of the university.



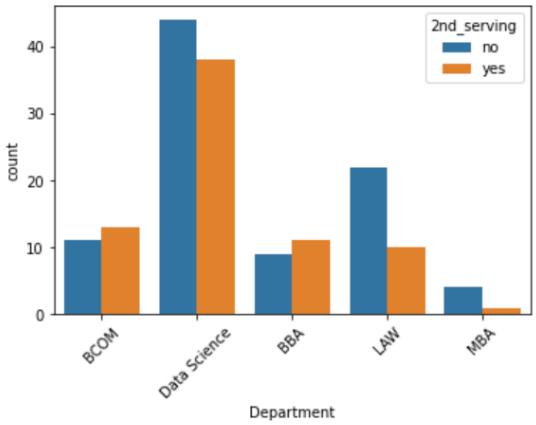
Result : The above bar graph shows that students of South India and students of North India are highly and least satisfied respectively.

Graph 4: Distribution of amount spent on food by students of different departments.



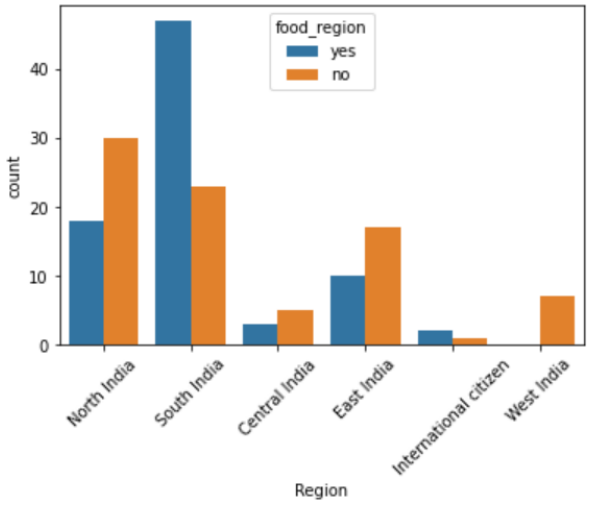
Result : The above graph shows that students of BCOM and BBA spend the most on food on a daily basis.

Graph 5: Relationship between students of different departments and their ability to take a 2nd serving of their choice meal.



Result : The above graph shows that the department of Data Science has the maximum number of people taking a 2nd serving of their choice meal.

Graph 6: Relationship between students of different regions and their satisfaction with the type of food served in the university.



Result : The above graph shows that students of South India are highly satisfied with the type of food served.

Statistical Method

Kruskal-Wallis H Test

The Kruskal-Wallis H test (also known as the "one-way ANOVA on ranks") is a rank-based nonparametric test that may be used to see if two or more groups of an independent variable on a continuous or ordinal dependent variable have statistically significant differences. It's a nonparametric alternative to the one-way ANOVA and an extension of the Mann-Whitney U test that allows you to compare more than two groups.

We have performed the Kruskal-Wallis H test taking the rating of price satisfaction as a dependent variable (grouping variable) and satisfaction rating of hygiene, ambience, variety and service as independent variables.

We used SPSS (Statistical Package for Social Sciences) to perform the test and the following results were obtained:

Cafe by valley

Ranks

	Price	N	Mean Rank
Service	5 very expensive	44	92.73
	4 little expensive	70	82.79
	3 up to mark	33	69.76
	2 cheap	7	76.43
	1 not at all	9	72.67
	Total	163	
Variety	5 very expensive	44	86.22
	4 little expensive	70	86.04
	3 up to mark	33	73.30
	2 cheap	7	67.50
	1 not at all	9	73.17
	Total	163	
Ambience	5 very expensive	44	83.41
	4 little expensive	70	91.41
	3 up to mark	33	64.68
	2 cheap	7	65.93
	1 not at all	9	77.94
	Total	163	
Hygeine	5 very expensive	44	88.51
	4 little expensive	70	86.68
	3 up to mark	33	65.85
	2 cheap	7	77.79
	1 not at all	9	76.28
	Total	163	

Test Statistics[a,b]

	Service	Variety	Ambience	Hygeine
Kruskal-Wallis H	5.503	3.228	9.224	6.204
df	4	4	4	4
Asymp. Sig.	.239	.520	.056	.184

a. Kruskal Wallis Test

b. Grouping Variable: Price

In cafe by valley it is observed that Asymptotic significance for all independent variables is greater than 0.05 so we fail to reject the null hypothesis and can conclude that there is no statistically significant difference between the variables.

Chopsticks

Ranks

	chopstick_pri	N	Mean Rank
Chopstick_hyg	1 not at all	20	43.35
	2 cheap	16	64.63
	3 up to mark	55	93.05
	4 little expensive	56	90.05
	5 very expensive	16	81.50
	Total	163	
chopstick_amb	1 not at all	20	57.45
	2 cheap	16	81.94
	3 up to mark	55	89.89
	4 little expensive	56	85.17
	5 very expensive	16	74.53
	Total	163	
chopstick_var	1 not at all	20	52.90
	2 cheap	16	79.84
	3 up to mark	55	97.45
	4 little expensive	56	82.98
	5 very expensive	16	64.00
	Total	163	
chopstick_ser	1 not at all	20	47.70
	2 cheap	16	80.00
	3 up to mark	55	91.75
	4 little expensive	56	86.57
	5 very expensive	16	77.38
	Total	163	

Test Statistics[a,b]

	Chopstick_hyg	chopstick_amb	chopstick_var	chopstick_ser
Kruskal-Wallis H	22.253	8.688	17.040	14.914
df	4	4	4	4
Asymp. Sig.	<.001	.069	.002	.005

a. Kruskal Wallis Test

b. Grouping Variable: chopstick_pri

In Chopsticks it is observed that Asymptotic significance for all independent variables except ambience is lesser than 0.05 so we reject the null hypothesis As a result, we reject the null hypothesis and conclude that the variables are statistically significant and only ambience is not statistically significant with price.

MBA Canteen

Ranks

	MBA_pri	N	Mean Rank
MBA_hyg	1 not at all	8	86.31
	2 cheap	16	93.75
	3 up to mark	81	80.07
	4 little expensive	46	80.18
	5 very expensive	12	83.42
	Total	163	
MBA_amb	1 not at all	8	104.44
	2 cheap	16	72.81
	3 up to mark	81	79.05
	4 little expensive	46	86.12
	5 very expensive	12	83.42
	Total	163	
MBA_var	1 not at all	8	81.81
	2 cheap	16	101.63
	3 up to mark	81	80.86
	4 little expensive	46	82.47
	5 very expensive	12	61.88
	Total	163	
MBA_ser	1 not at all	8	76.75
	2 cheap	16	91.69
	3 up to mark	81	74.89
	4 little expensive	46	89.26
	5 very expensive	12	92.75
	Total	163	

Test Statistics[a,b]

	MBA_hyg	MBA_amb	MBA_var	MBA_ser
Kruskal-Wallis H	1.420	3.456	5.434	4.732
df	4	4	4	4
Asymp. Sig.	.841	.485	.246	.316

a. Kruskal Wallis Test

b. Grouping Variable: MBA_pri

In the MBA canteen it is observed that Asymptotic significance for all independent variables is greater than 0.05 so we fail to reject the null hypothesis and can conclude that there is no statistically significant difference between the variables.

Concerto

Ranks

	concerto_pri	N	Mean Rank
concerto_hyg	1 not at all	12	66.25
	2 cheap	5	32.10
	3 up to mark	39	76.22
	4 little expensive	74	85.06
	5 very expensive	33	95.26
	Total	163	
concerto_amb	1 not at all	12	63.67
	2 cheap	5	65.70
	3 up to mark	39	65.86
	4 little expensive	74	89.12
	5 very expensive	33	94.24
	Total	163	
concerto_var	1 not at all	12	58.54
	2 cheap	5	48.20
	3 up to mark	39	92.88
	4 little expensive	74	83.29
	5 very expensive	33	79.89
	Total	163	
concerto_ser	1 not at all	12	61.04
	2 cheap	5	60.10
	3 up to mark	39	77.18
	4 little expensive	74	90.33
	5 very expensive	33	79.95
	Total	163	

Test Statistics[a,b]

	concerto_hyg	concerto_amb	concerto_var	concerto_ser
Kruskal-Wallis H	11.982	12.628	8.632	7.189
df	4	4	4	4
Asymp. Sig.	.017	.013	.071	.126

a. Kruskal Wallis Test

b. Grouping Variable: concerto_pri

In cafe concerto it is observed that Asymptotic significance for the independent variables variety and service is greater than 0.05 so we fail to reject the null hypothesis and can conclude that there is no statistically significant difference between the variables on the other hand, hygiene and ambience have Asymptotic significance less than 0.05 so we reject the null hypothesis and conclude that there is statistically significant difference between the variables.

Machine learning- Model (Logistic regression)

Logistic regression

Logistic regression can be termed as a statistical analytic method for predicting a binary outcome,i.e., yes or no of customer satisfaction, based on prior observations of the recorded data set.

Importing Libraries

```
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
```

Label Encoder- Helps in converting labels into numeric data which helps machine to understand it in easy form.

Using label encoder to convert yes or no to 1 or 0.
1 for yes, 0 for no.

```
x1=survey1[["2nd_serving","hot_food","food_region","sitting_place"]]
x1
```

	2nd_serving	hot_food	food_region	sitting_place
0	0	1	1	1
1	1	1	1	1
2	1	1	1	1
3	0	1	0	1
4	1	1	0	0
...	...	...	...	...
158	0	0	0	0
159	1	1	0	0
160	0	1	0	1
161	1	1	0	0
162	1	0	1	0

163 rows × 4 columns

Putting the data for training and testing .
Train size = 70%
Test size= 30%

```
x1_train,x1_test,y_train,y_test=train_test_split(x1,y,test_size=0.3)
```

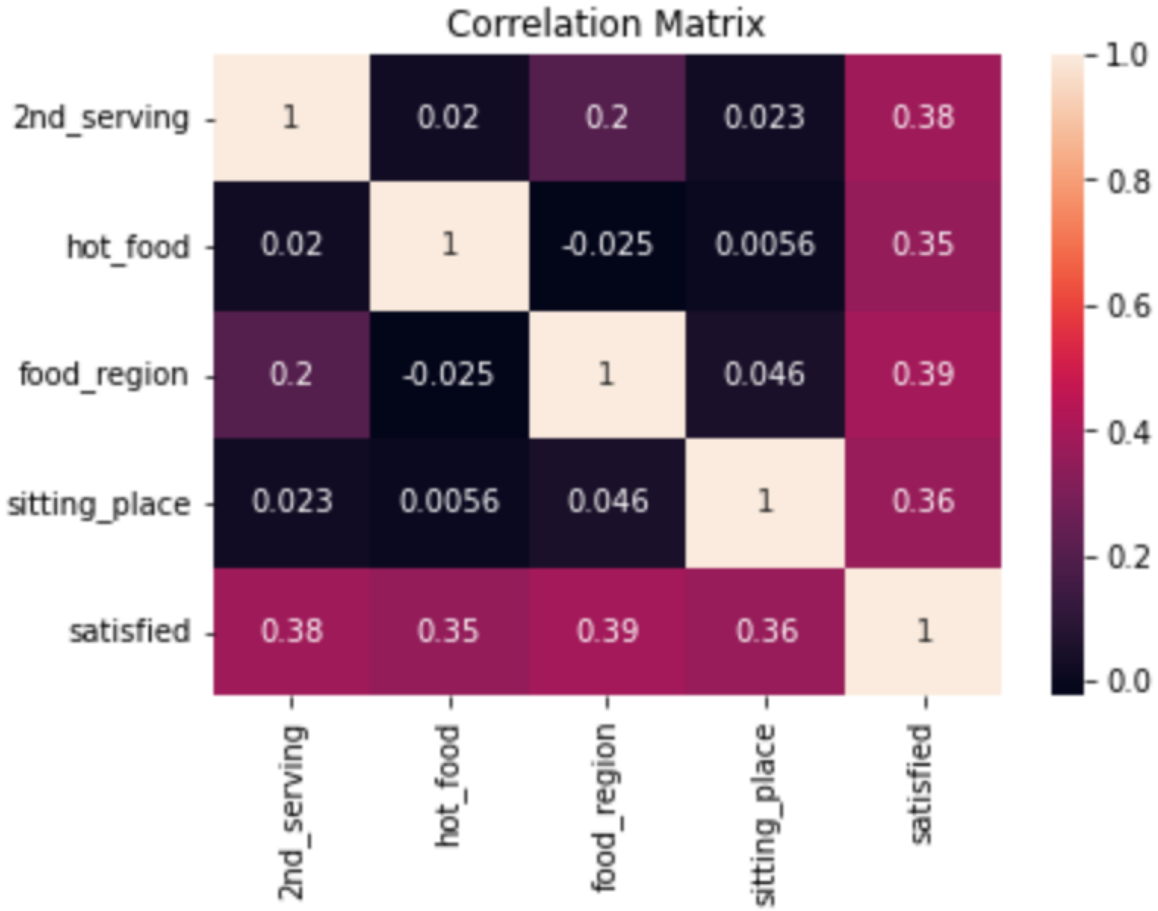
```
Shape of x_train::  (114, 4)

Shape of x_test::  (49, 4)

Shape of y_train::  (114, 1)

Shape of y_test::  (49, 1)
```

Accuracy of the model is 0.85

```
In [22]: print("Accuracy Score of the model:",accuracy_score(y_pred,y_test))
         Accuracy Score of the model: 0.8571428571428571
```

Correlation Matrix- A correlation matrix shows the relationship between two variables. It works well with variables that have a linear relationship with one another.

<u>Correlation Matrix of dependent and independent variables.</u>



‘

Correlation between the dependent and the independent variables.

Independent variables- 2nd serving, hot_food, food_region, sitting_place
Dependent variables- Satisfied

Conclusion: The conclusion which we got from our project is that with the help of independent variables we can predict the customer satisfaction of the canteens in our college. Upon Kruskal-Wallis H Test we found that independent and dependent variables have no statistically significant difference. The graphs above give us an exploratory data analysis based on the survey.

We have used logistic regression as our model which gave us accuracy around 85% which predicts the satisfaction of the students.

One of the major points that came from EDA is about South Indians who are more satisfied with the canteen as compared to North Indians.

Overall if the students are provided with good food with good facilities most of the students will be satisfied.

References:

Harmadi, A. C. (2021, September 25). *10 Things to do when conducting your Exploratory Data Analysis (EDA)*. Medium. Retrieved May 30, 2022, from https://medium.com/data-folks-indonesia/10-things-to-do-when-conducting-your-exploratory-data-analysis-eda-7e3b2dfbf812

Kruskal-Wallis H Test in SPSS Statistics | Procedure, output and interpretation of the output using a relevant example. (n.d.). Laerd Statistics. Retrieved May 30, 2022, from https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php

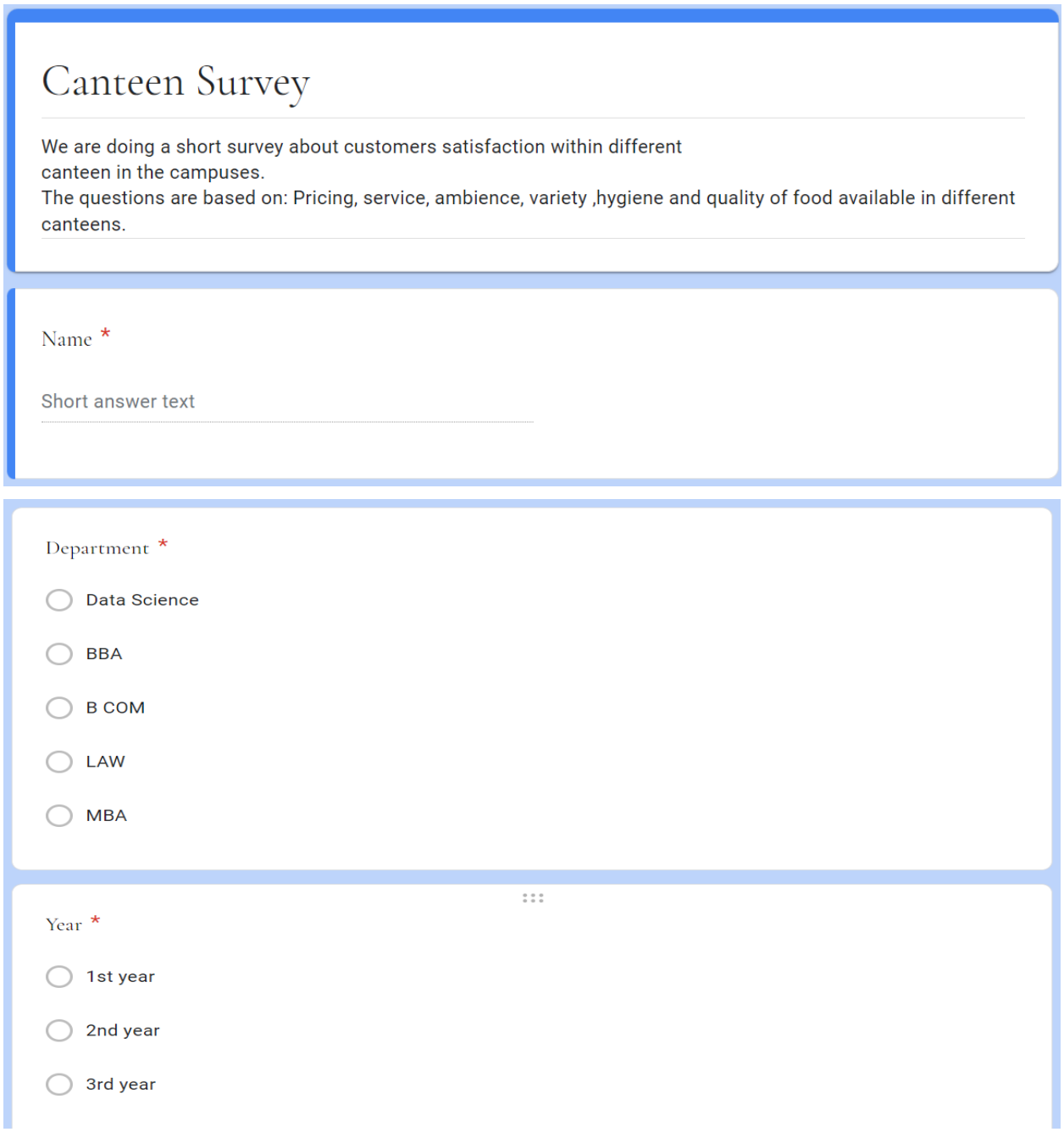
sklearn.linear_model.LogisticRegression — scikit-learn 1.1.1 documentation. (n.d.). Scikit-learn. Retrieved May 30, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Smith, I. (2020, March 23). *Exploratory Data Analysis (EDA) Visualization Using Pandas | by Pararawendy Indarjo*. Towards Data Science. Retrieved May 30, 2022, from

https://towardsdatascience.com/exploratory-data-analysis-eda-visualization-using-pandas-ca5a04271607

Appendix-

Questionnaire:



Canteen Survey

We are doing a short survey about customers satisfaction within different canteen in the campuses.
The questions are based on: Pricing, service, ambience, variety ,hygiene and quality of food available in different canteens.

Name *

Short answer text

Department *

- Data Science
- BBA
- B COM
- LAW
- MBA

Year *

- 1st year
- 2nd year
- 3rd year

E-mail *

Short answer text

Region *

- North India
- South India
- East India
- West India
- Central India
- International citizen

How much would you rate the hygiene of canteens? *

	1	2	3	4	5
Chopstics					
Cafe Concerto					
MBA Block					
Cafe by valley					

How much would you rate the ambience of canteens? *

	1 (worst)	2 (bad)	3 (ok)	4 (good)	5 (excellent)
Chopstics	○	○	○	○	○
Cafe Concerto	○	○	○	○	○
MBA Block	○	○	○	○	○
Cafe by valley	○	○	○	○	○

Is there sufficient variety of meals to choose from? *

	1 (worst)	2 (bad)	3 (ok)	4 (good)	5 (excellent)
Chopstics	○	○	○	○	○
Cafe Concerto	○	○	○	○	○
MBA Block	○	○	○	○	○
Cafe by valley	○	○	○	○	○

How much would rate the overall service facilities? *

	1 (worst)	2 (bad)	3 (ok)	4 (good)	5 (excellent)
Chopstics	○	○	○	○	○
Cafe Concerto	○	○	○	○	○
MBA Block	○	○	○	○	○
Cafe by valley	○	○	○	○	○

Do you think the prices are fair for the types of food sold? *

	1 not at all	2 cheap	3 up to mark	4 little expensive	5 very expensive
Chopstics	○	○	○	○	○
Cafe Concerto	○	○	○	○	○
MBA Block	○	○	○	○	○
Cafe by valley	○	○	○	○	○

Do you have the ability to take a 2nd serving of your choose meal? *

- ○ yes
- ○ no

Is the food served always hot enough? *

- ○ yes
- ○ no

Do you get the type of food as per your region? *

- ○ yes
- ○ no