Complex Computing Problem Report

Introduction to Data Science (CSC 487)

BS(CS)-6(A)

Assignment 03

Group Members

Name	Enrollment
1. IMAAN SHAHID	02-134201-009
2. AWAIS ARSHAD	02-134201-038
3. MANAHIL SHAMIM	02-134201-055

Submitted to:

Dr. Saifullah Tumrani

BAHRIA UNIVERSITY KARACHI CAMPUS

Table of Contents

CHAPTER 1: INTRODUCTION AND PROBLEM STATEMENT

Introduction

Bahria university has a great strength of students, thus, to serve these students with snacks and food items there are two canteens, Wow fries and the cafeteria. Both canteens have multiple food options to select from, however none of them have a proper breakfast menu which students can enjoy. In this complex computing problem, we will be discussing whether students want a breakfast menu at university.

Problem Statement

The problem statement we will be working on is:

Perform classification and interpret which features influence whether a student wishes to have breakfast menu included in canteen or not?

Studying the above problem statement will enable us to know if we can add a breakfast menu or not and if yes then in what way people take items from the menu. The reason behind choosing this problem is that most students come from distant places so they are not able to do proper breakfast in the morning and because of that they aren't active which makes them lose concentration during the morning classes.

CHAPTER 2: DATASET AND DATA COLLECTION

Dataset

For this assignment we collected a dataset of 226 entries from Bahria university. The dataset included various individuals from different departments and semesters. The questions asked to collect were a few personal details such as Name, email, enrolment, semester, department. Then a few details related to our problem statement these included the following questions:

- Do you live near university or far?
- Are you satisfied by the services provided by cafeteria/wow fries?
- Do you have breakfast in university or at home?
- If in University, then why not at home?
- Do you think it will be nice to have a proper breakfast menu in cafeteria/wow fries?
- What should be available in the breakfast menu if offered by cafeteria/wow fries?
- How much should price vary for breakfast?
- what should be the timings for breakfast?

Data collection form

The screenshots below display google form used to collect data

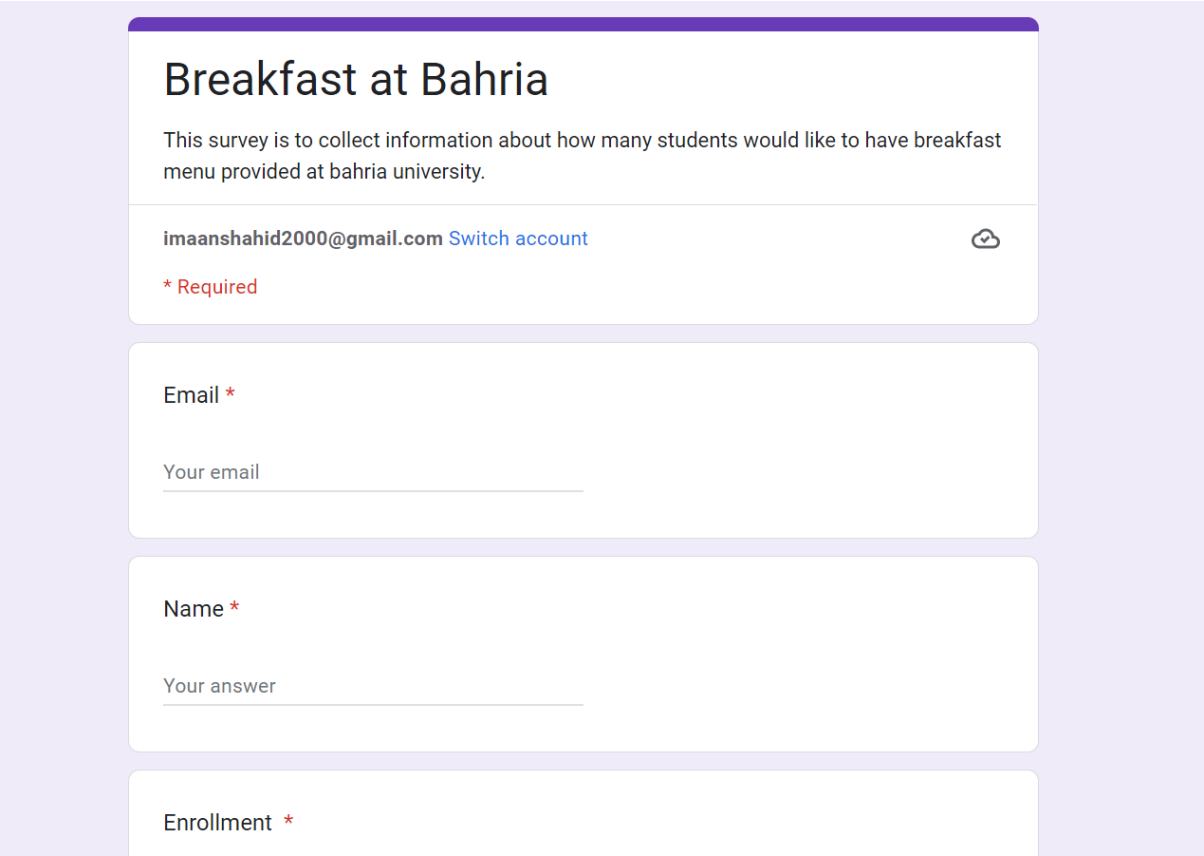

Fig 2.1

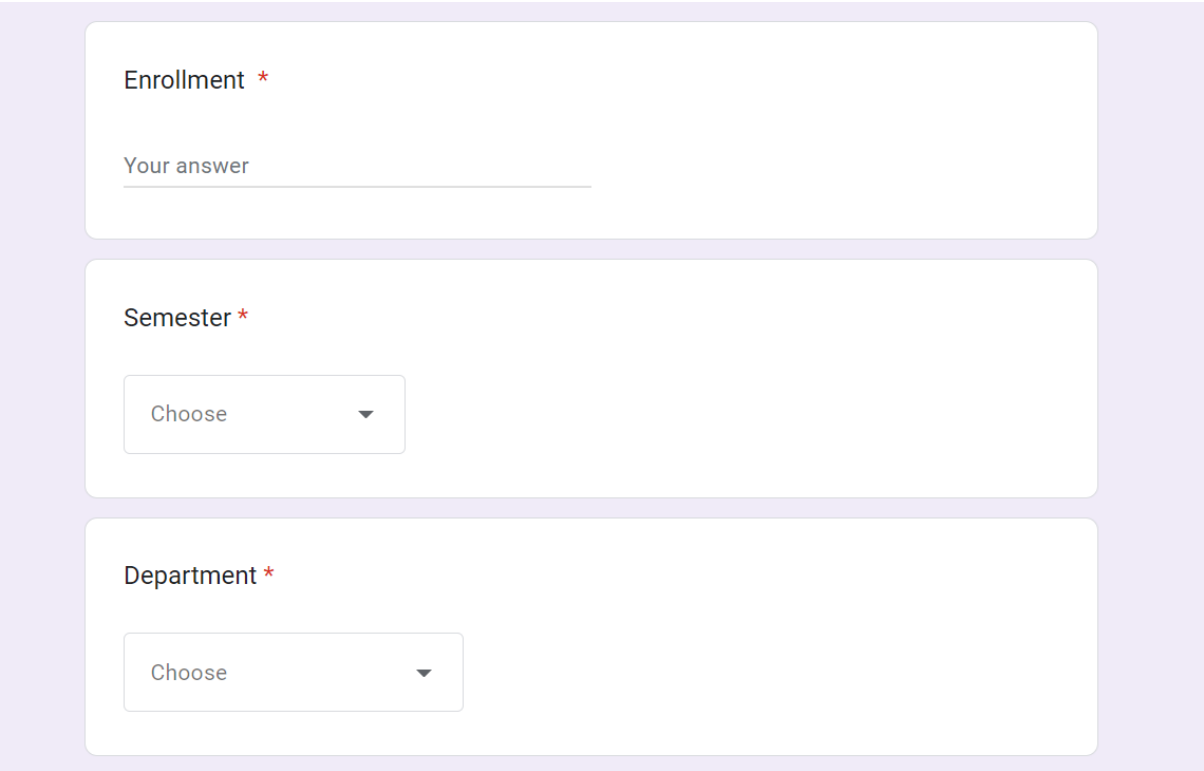
Enrollment *
Your answer
Semester *
Choose
Department *
Choose

Fig 2.2

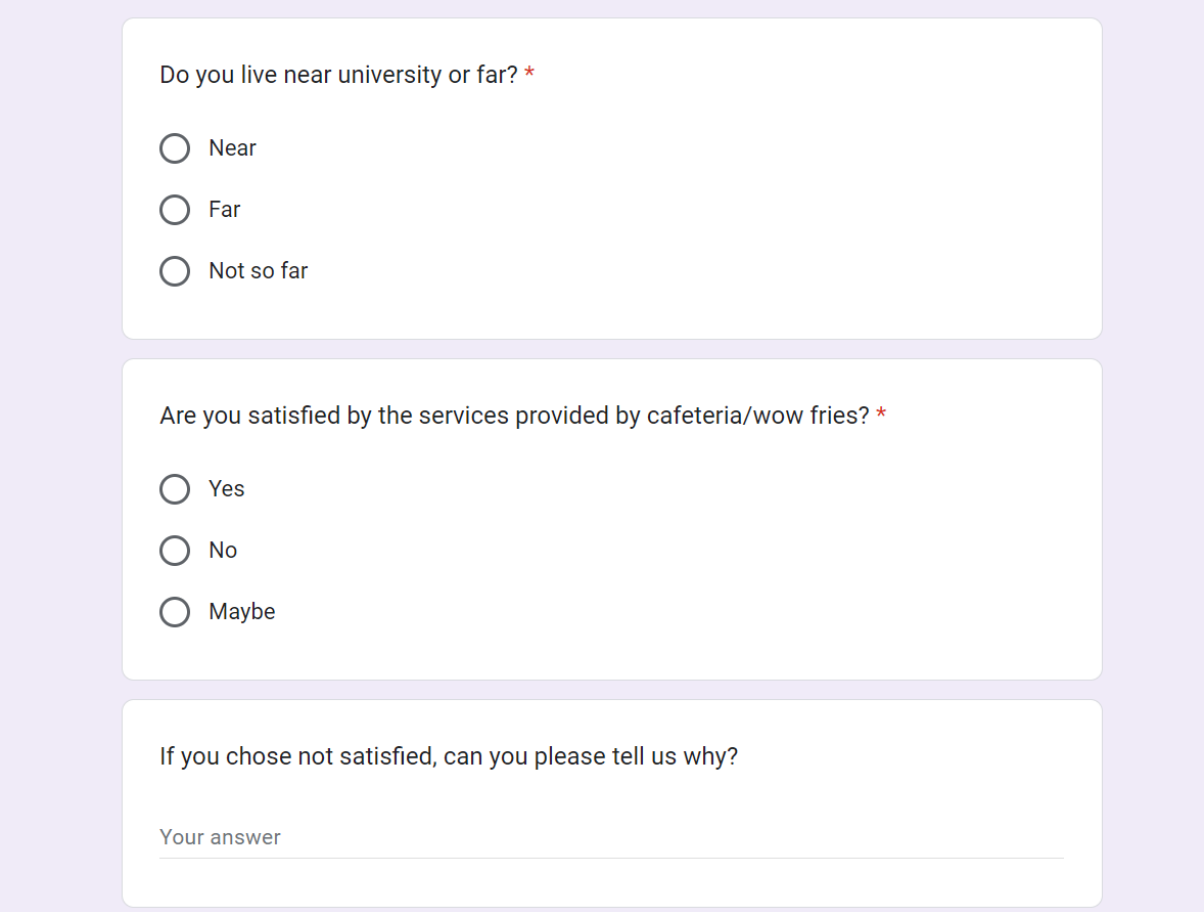
Do you live near university or far? *
Near
Far
Not so far
Are you satisfied by the services provided by cafeteria/wow fries? *
Yes
No
Maybe
If you chose not satisfied, can you please tell us why?
Your answer

Fig 2.3

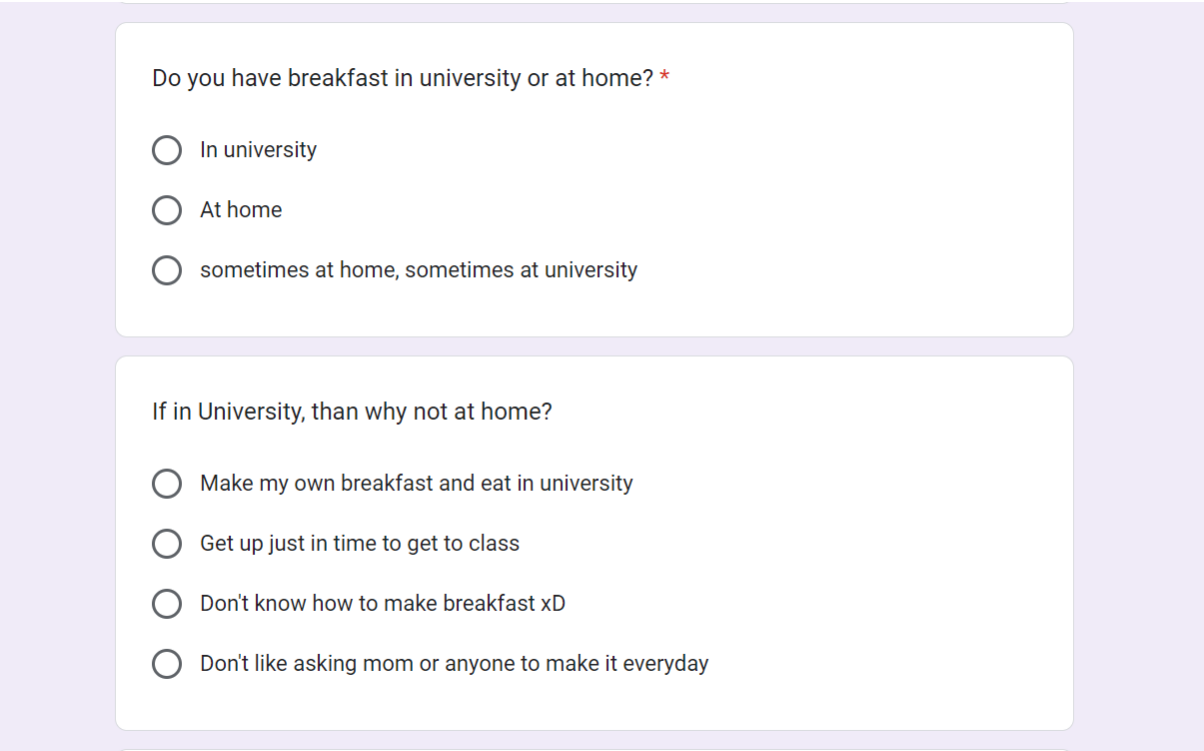
Do you have breakfast in university or at home? *
In university
At home
sometimes at home, sometimes at university
If in University, than why not at home?
Make my own breakfast and eat in university
Get up just in time to get to class
Don't know how to make breakfast xD
Don't like asking mom or anyone to make it everyday

Fig 2.4

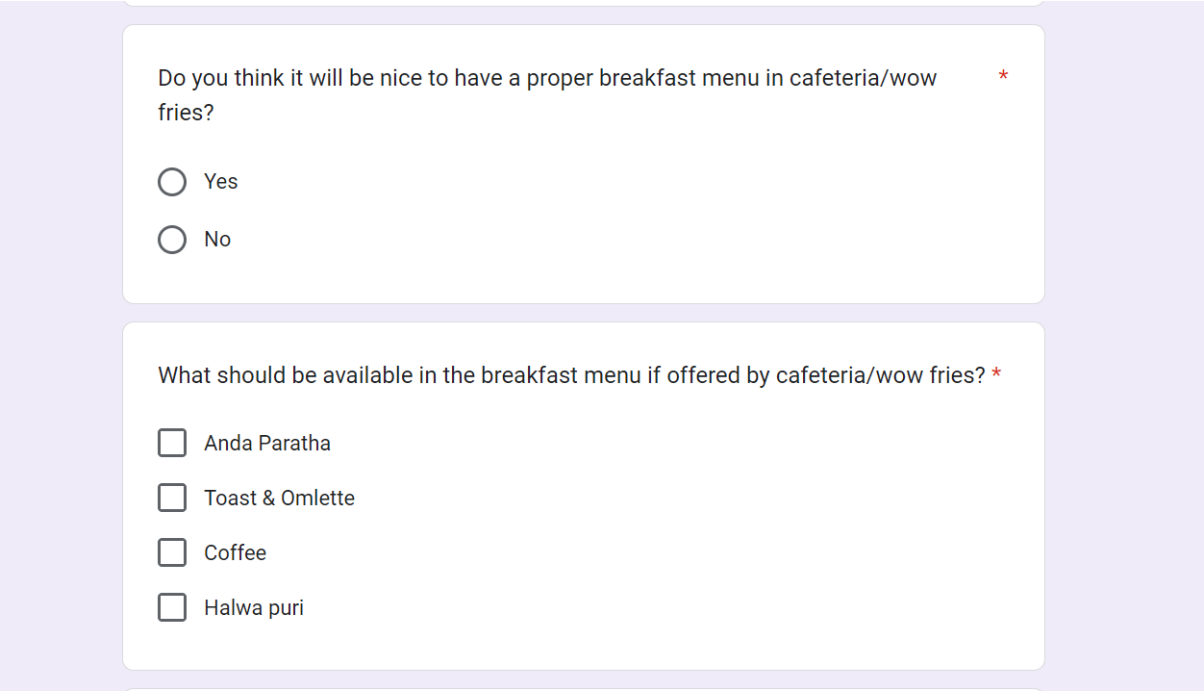
Do you think it will be nice to have a proper breakfast menu in cafeteria/wow fries? *
Yes
No
What should be available in the breakfast menu if offered by cafeteria/wow fries? *
Anda Paratha
Toast & Omlette
Coffee
Halwa puri

Fig 2.5

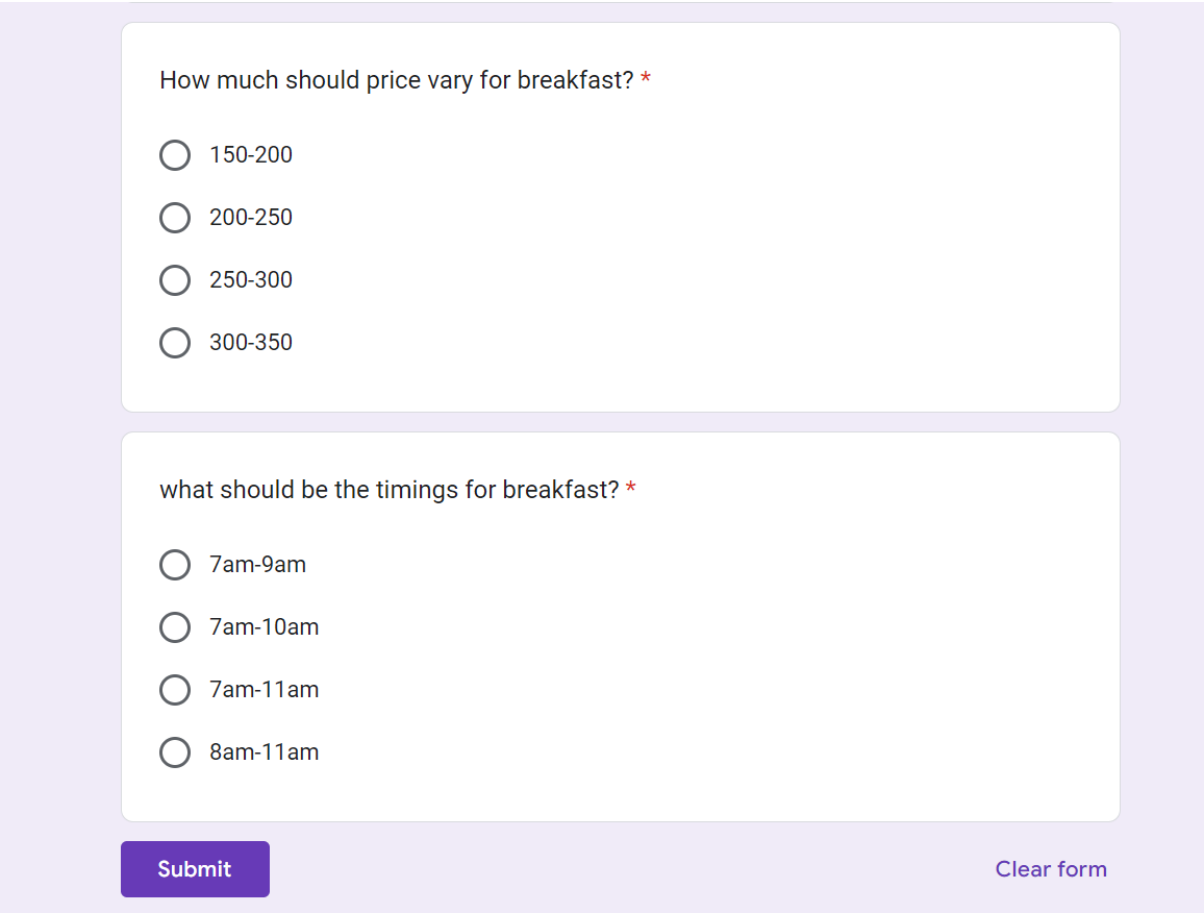

Fig 2.6

Google Forms analysis

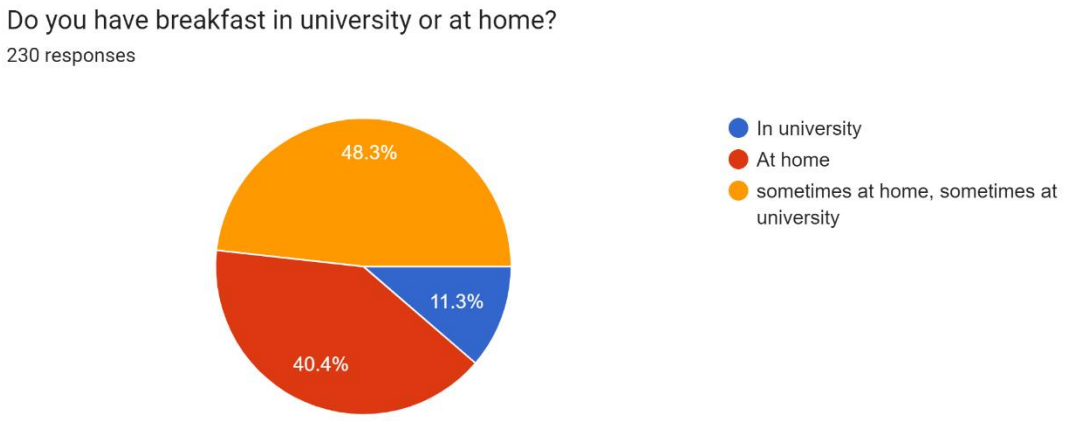

Fig 2.7

Do you think it will be nice to have a proper breakfast menu in cafeteria/wow fries?

230 responses

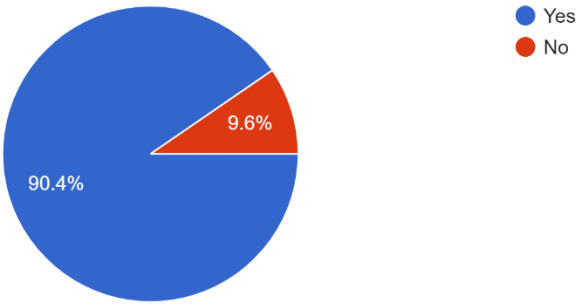

Fig 2.8

How much should price vary for breakfast?

230 responses

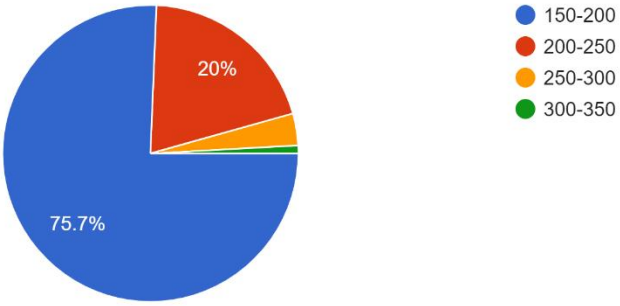

Fig 2.9

If in University, than why not at home?

154 responses

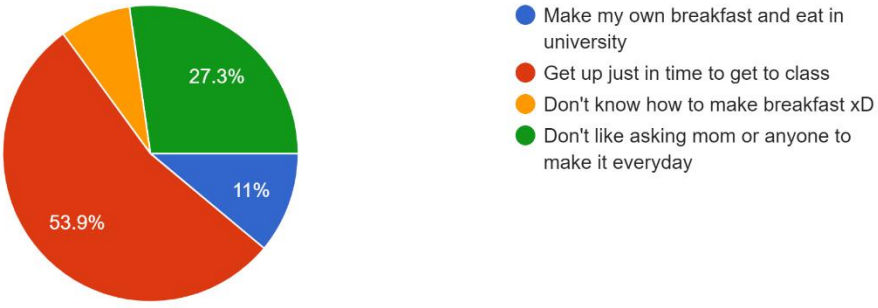

Fig 2.10

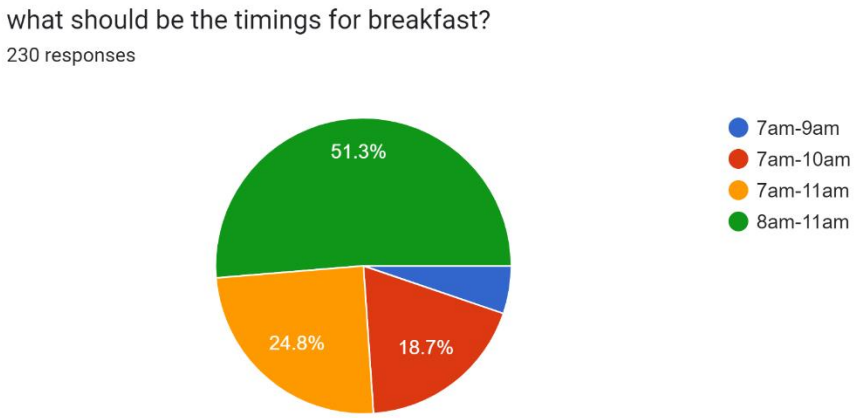

Fig 2.11

CHAPTER 3: ANALYSIS

To analyze the problem of which number of individuals would like breakfast at university or not we will need to focus based on which individuals say yes or no. To interpret that we have implemented Random Forest algorithm. The ability to precisely classify observations is extremely valuable for various applications, in our case we need to classify whether a student wants to have breakfast menu at university or not. Random forest, like its name implies, consists of many individual decision trees that operate as an ensemble. Thus, in our case the questions related to the distance the student has to travel every morning, the time, satisfaction with the current menu available, and whether they have breakfast at home, all these queries built up individual decision trees.

Again, referring to our initial problem statement our focus is not only to classify but extract which features have the highest importance in answering ‘Yes’ or ‘No’ to breakfast menu. Thus, the below figures display the implementation of our algorithm and feature identification of the most crucial features for our problem.

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, accuracy_score, confusion_matrix
from sklearn.preprocessing import LabelEncoder
from sklearn.feature_selection import chi2


RFC = RandomForestClassifier(n_estimators=500)

x = df.drop(columns=['Name', 'Semester','Reason','Enrollment','Menu','BreakfastMenu_at_cafe'])
y = df['BreakfastMenu_at_cafe']

X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2, stratify=y)

RFC.fit(X_train, y_train)

y_pred=RFC.predict(X_test)
```

Fig 3.1 Random Forest applied on the required columns of dataset

```
print('Confusion Matrix:\n', confusion_matrix(y_test, y_pred), '\n')
print('Accuracy Score:\n', accuracy_score(y_test, y_pred), '\n')
print('Classification Report:\n', classification_report(y_test, y_pred))
```

```
Confusion Matrix:
 [[ 0  4]
 [ 0 41]]

Accuracy Score:
 0.9111111111111111

Classification Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00         4
           1       0.91      1.00      0.95        41

    accuracy                           0.91        45
   macro avg       0.46      0.50      0.48        45
weighted avg       0.83      0.91      0.87        45
```

Fig 3.2 Accuracy of the algorithm

```
feature_importance = np.array(RFC.feature_importances_)
feature_names = np.array(x.columns)

#Create a DataFrame using a Dictionary
data={'feature_names':feature_names,'feature_importance':feature_importance}
fi_df = pd.DataFrame(data)

#Sort the DataFrame in order decreasing feature importance
fi_df.sort_values(by=['feature_importance'], ascending=False,inplace=True)
sns.barplot(data=fi_df[:20], y='feature_names', x='feature_importance')
plt.show()
```

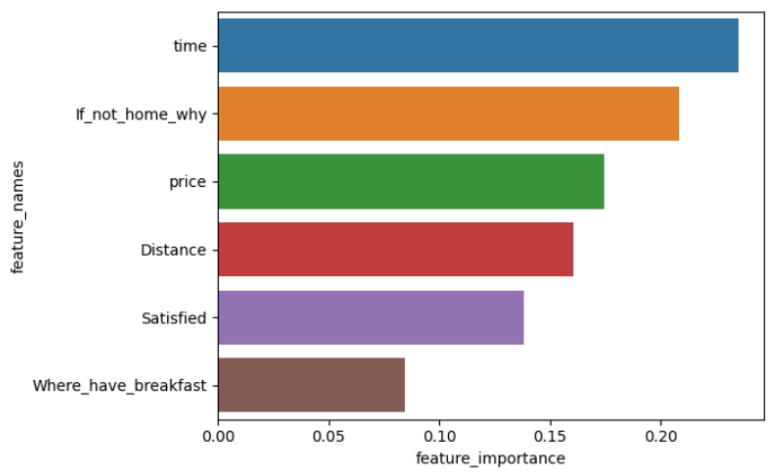

Fig 3.3 Feature importance insight gained after application of algorithm

As you can see in Fig 3.3 the top 3 relevant features from our specified columns are Time, reason of not doing breakfast at home, and Distance. These three indicate whether a person wants breakfast at university or not. If Bahria universities catering services sell breakfast menu at a mutually preferable price rate and time, they will gain profit because students who aren't able to do breakfast at home is one of the important features and so are the preferred time and price range given by students.

CHAPTER 4: CONCLUSION

To conclude our analysis related to this problem we must first acknowledge the fact that our chosen problem is valid, as the survey result shows that people aren't currently satisfied

with the menu and would like breakfast menu to be added. As well as after applying the random forest algorithm we analyzed that most important features that have resulted in the classification of ‘Yes’ and ‘No’ of the dataset are those features which can be used to solve the problem we are facing of not having a breakfast menu.

CHAPTER 5: REFERENCES

sklearn.ensemble.RandomForestClassifier — scikit-learn 1.2.0 documentation

Exploratory data analysis of food preference | Kaggle