



# SC1015

## Mini Project

Darren Thia Kok Long (U2322414B)  
Esmond Chan Shun Jie (U2322484K)  
Nadathur Ammal Shreya Sudharshan (U2323293D)



# TABLE OF CONTENTS

**01** INTRODUCTION

**02** EXPLORATORY  
DATA ANALYSIS

**03** CORE ANALYSIS

**04** CONCLUSION



# Stress

**Stress** is a word that **resonates deeply** among university students



# Student Satisfaction

## Factors:

- Support system
- Social circle
- Quality of education
- Campus facilities
- Sense of belonging



# Our Objective



Reduce Stress



Increase Satisfaction

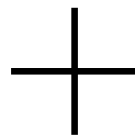


Present ways to  
improve student life



# Exploratory Data Analysis

# Data Preparation



data columns (total 15 columns):

#	Column	Non-Null Count	Dtype
0	Career	2958 non-null	object
1	Citizenship	2958 non-null	object
2	Nationality	2958 non-null	object
3	Year of Study	2958 non-null	int64
4	Primary Programme	2958 non-null	object
5	Gender	2958 non-null	object
6	Department	2958 non-null	object
7	Housing Type	2958 non-null	object
8	Q1-How many events have you Volunteered in ?	2958 non-null	int64
9	Q2-How many events have you Participated in ?	2958 non-null	int64
10	Q3-How many activities are you Interested in ?	2958 non-null	int64
11	Q4-How many activities are you Passionate about ?	2958 non-null	int64
12	Q5-What are your levels of stress ?	2958 non-null	int64
13	Q6-How Satisfied You are with your Student Life ?	2958 non-null	int64
14	Q7-How much effort do you make to interact with others ?	2940 non-null	float64

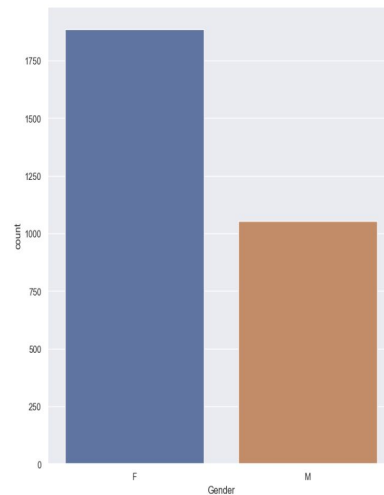
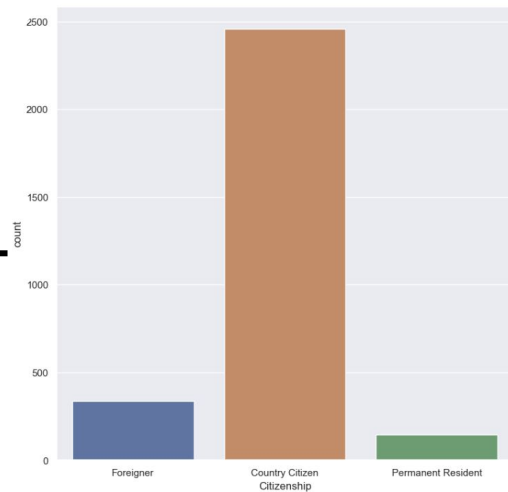
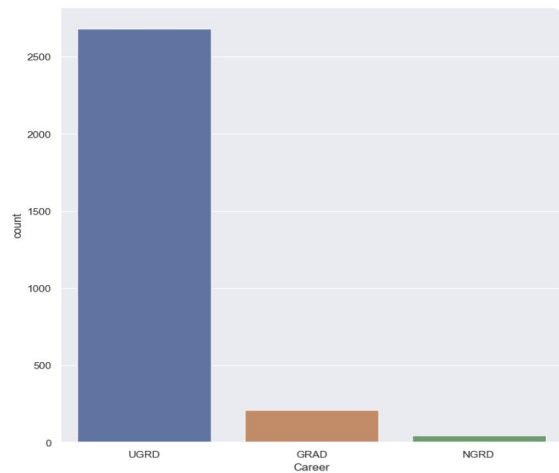
dtypes: float64(1), int64(7), object(7)

The dataset is cleaned by dropping null values and removing columns that do not help predict our problem definition.

	Year of Study	Q1-How many events have you Volunteered in ?	Q2-How many events have you Participated in ?	Q3-How many activities are you Interested in ?	Q4-How many activities are you Passionate about ?	Q5-What are your levels of stress ?	Q6-How Satisfied You are with your Student Life ?	Q7-How much effort do you make to interact with others ?
count	2940.000000	2940.000000	2940.000000	2940.000000	2940.000000	2940.000000	2940.000000	2940.000000
mean	2.209184	1.25102	1.493197	2.728231	3.189456	3.094558	1.923810	1.963265
std	1.098787	1.48243	0.959695	1.264300	1.970303	1.655793	0.562202	0.575096
min	1.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	1.000000	0.000000	1.000000	2.000000	2.000000	2.000000	2.000000	2.000000
50%	2.000000	1.000000	1.000000	3.000000	3.000000	3.000000	2.000000	2.000000
75%	3.000000	2.000000	2.000000	3.000000	4.000000	4.000000	2.000000	2.000000
max	5.000000	11.000000	5.000000	8.000000	11.000000	9.000000	3.000000	3.000000

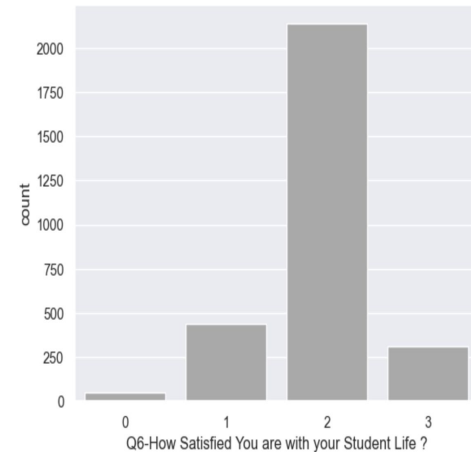
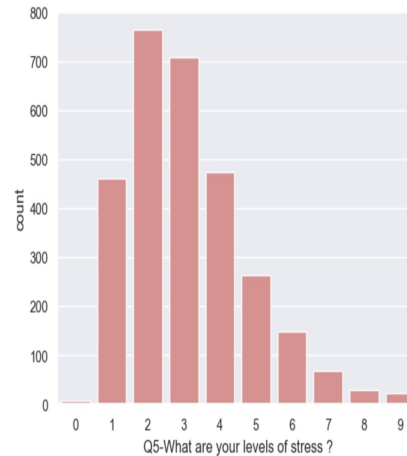
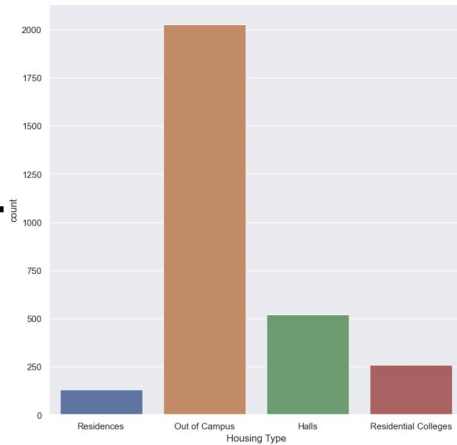
Univariate Statistics

# Univariate Analysis





# Univariate Analysis



# Modifying the Dataset +

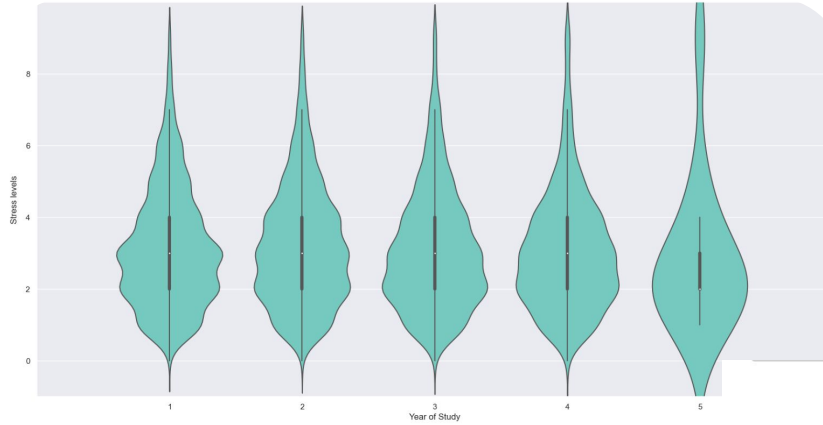
We found that majority of the students were undergraduates.

```
count=0;
a=0;
for x in file1['Career']:
    if x=='UGRD':
        count=count+1
    a+=1
print("Number of undergrads : ", count)
print("Total number of students : ", a)
```

Number of undergrads : 2683  
Total number of students : 2940

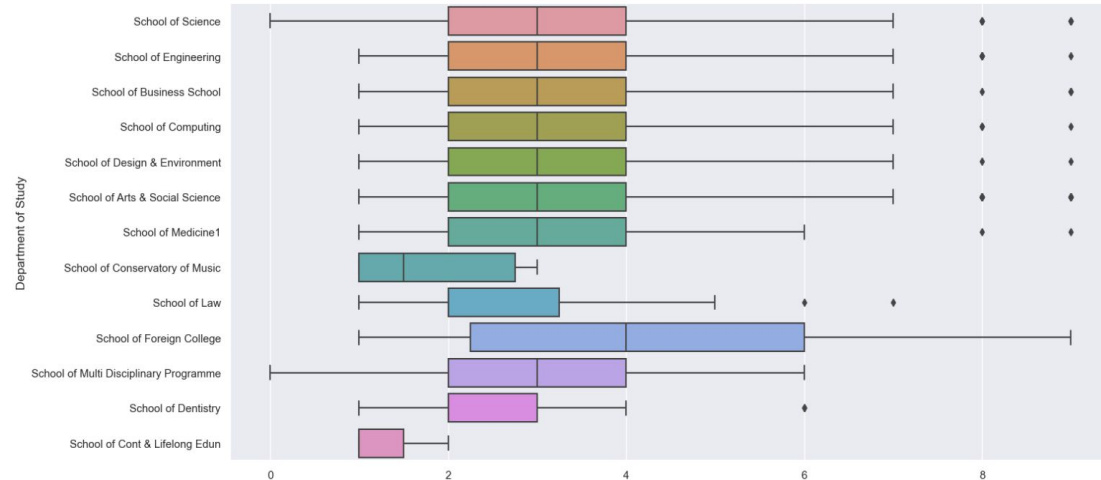
	Year of Study	Q1-How many events have you Volunteered in ?	Q2-How many events have you Participated in ?	Q3-How many activities are you Interested in ?	Q4-How many activities are you Passionate about ?	Q5-What are your levels of stress ?	Q6-How Satisfied You are with your Student Life ?	Q7-How much effort do you make to interact with others ?
count	2683.000000	2683.000000	2683.000000	2683.000000	2683.000000	2683.000000	2683.000000	2683.000000
mean	2.250093	1.316810	1.522177	2.703317	3.221394	3.147223	1.911666	1.948192
std	1.090804	1.498783	0.952924	1.255241	1.976417	1.671229	0.562322	0.577285
min	1.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000
25%	1.000000	0.000000	1.000000	2.000000	2.000000	2.000000	2.000000	2.000000
50%	2.000000	1.000000	1.000000	3.000000	3.000000	3.000000	2.000000	2.000000
75%	3.000000	2.000000	2.000000	3.000000	4.000000	4.000000	2.000000	2.000000
max	5.000000	11.000000	5.000000	8.000000	11.000000	9.000000	3.000000	3.000000

# Factors Affecting Stress



Year of study against stress levels

Department of study against stress levels





# Machine Learning

# What type of ML will we be using?

Question 1: What are the factors affecting student's stress level?

Question 2: What are the factors affecting student life satisfaction level?

**Prediction of a Discrete Class Label**

Target Variable for Qn 1 - Low, Medium & High

Target Variable for Qn 2 - Dissatisfied, Neutral & Satisfied

**Multi Class Classification Problem**



# How can ML solve our objectives?

**Stress Level among Students**

**Student Life Satisfaction**

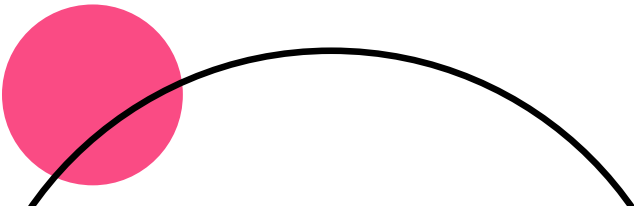
Machine Learning

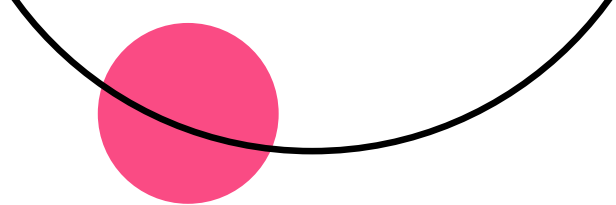
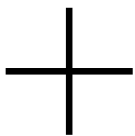


By predicting the factors that have significant effect on stress level/student life satisfaction, we can better focus on those factors.

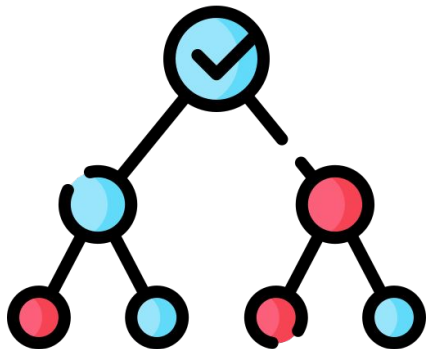
E.g. If 'Housing Type' is a important factor in the prediction model, perhaps we can explore how it exactly helps or worsen stress to help students manage stress better.

E.g. If 'Housing Type' is a important factor in the prediction model, perhaps we can explore how it exactly boost or reduce student life satisfaction to help students have a more enjoyable student life.



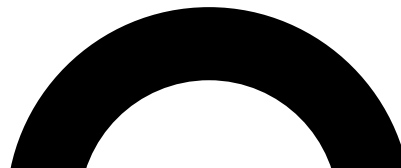
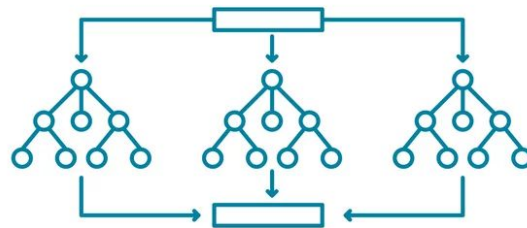


## Decision Tree Classifier



VS

## Random Forest Classifier



# Preparation of Dataset for ML

## Drop Irrelevant Columns & Rows with Missing Values

```
df1.drop(['Year since Matriculation',  
         'Q3-How many activities are you Interested in ?',  
         'Q4-How many activities are you Passionate about ?',  
         'Q8-About How events are you aware about ?',  
         'response_id',  
         'Q9-What is an ideal student life ?'], axis='columns', inplace=True)  
df1 = df1.dropna()
```

## Convert Relevant Columns to Categorical

```
df1['Career'] = df1['Career'].astype("category")  
df1['Citizenship'] = df1['Citizenship'].astype("category")  
df1['Nationality'] = df1['Nationality'].astype("category")  
df1['Primary Programme'] = df1['Primary Programme'].astype("category")  
df1['Gender'] = df1['Gender'].astype("category")  
df1['Department'] = df1['Department'].astype("category")  
df1['Housing Type'] = df1['Housing Type'].astype("category")
```



# Preparation of Dataset for ML

## Apply One Hot Encoding to Categorical Columns

```
prefixes = ['Career', 'Citizenship', 'Nationality', 'Primary Programme', 'Gender', 'Department', 'Housing Type']  
# Convert categorical columns to one-hot encoded variables with prefixes  
df_encoded = pd.get_dummies(df1, prefix=prefixes)  
  
# Concatenate encoded columns to original DataFrame  
df1 = pd.concat([df1, df_encoded], axis=1)  
  
# Drop the original categorical columns  
df1.drop(['Career', 'Citizenship', 'Nationality', 'Primary Programme', 'Gender', 'Department', 'Housing Type'], axis=1, inplace=True)  
  
# Now df contains one-hot encoded columns with specified prefixes
```

## Set the Scale for 'Stress Level' & 'Student Life Satisfaction'

```
df1['Stress Level'] = pd.cut(df1['Q5-What are your levels of stress ?'],  
                             bins=[-1, 3, 7, 9], labels=['Low', 'Medium', 'High'])  
  
df1['Student Life Satisfaction'] = pd.cut(df1['Q6-How Satisfied You are with your Student Life ?'],  
                                           bins=[-1, 1, 2, 3], labels=['Dissatisfied', 'Neutral', 'Satisfied'])
```

# Train Test Split of Dataset

Create 2 sets of dataset - One for **Stress Level Prediction** & Other for **Student Life Satisfaction Prediction**

```
X1 = df1.drop(['Stress Level', 'Student Life Satisfaction', 'Q5-What are your levels of stress ?'], axis=1)
X2 = df1.drop(['Stress Level', 'Student Life Satisfaction', 'Q6-How Satisfied You are with your Student Life ?'], axis=1)

y_stress = df1['Stress Level']
y_satisfaction = df1['Student Life Satisfaction']
```

Choosing the **appropriate** train-test ratio

```
X_train1, X_test1, y_train_stress, y_test_stress = train_test_split(X1, y_stress, test_size=0.2, random_state=42)
X_train2, X_test2, y_train_satisfaction, y_test_satisfaction = train_test_split(X2, y_satisfaction, test_size=0.2, random_state=42)
```

# Fitting of the Models

```
from sklearn.tree import DecisionTreeClassifier
```

```
dt_model1 = DecisionTreeClassifier()  
dt_model1.fit(X_train1, y_train_stress)
```

```
dt_model2 = DecisionTreeClassifier()  
dt_model2.fit(X_train2, y_train_satisfaction)
```

▼ DecisionTreeClassifier ⓘ ⓘ

```
DecisionTreeClassifier()
```

```
from sklearn.ensemble import RandomForestClassifier
```

```
rf_model1 = RandomForestClassifier()  
rf_model1.fit(X_train1, y_train_stress)
```

```
rf_model2 = RandomForestClassifier()  
rf_model2.fit(X_train2, y_train_satisfaction)
```

▼ RandomForestClassifier ⓘ ⓘ

```
RandomForestClassifier()
```

# Choosing the Better Model

	Decision Tree Classifier	Random Forest Classifier
Stress Level Prediction	Accuracy: 0.5714 Precision: 0.5738 Recall: 0.5714 F1-score: 0.5726	Accuracy: 0.6139 Precision: 0.5801 Recall: 0.6139 F1-score: 0.5931
Student Life Satisfaction Prediction	Accuracy: 0.6054 Precision: 0.6133 Recall: 0.6054 F1-score: 0.6092	Accuracy: 0.7160 Precision: 0.6475 Recall: 0.7160 F1-score: 0.6589

# Exploring Other Models

- Logistic Regression (Unable to)
  - Used for **binary** classification task
- Support Vector Classification (SVC)
  - Powerful model that is **effective** in processing **high dimensional data**
- K-Nearest Neighbors (KNN)
  - **Simple & Effective**
  - One of the popular classification model

# Fitting SVC & KNN

```
from sklearn.svm import SVC

svm_model1 = SVC()
svm_model1.fit(X_train1, y_train_stress)

svm_model2 = SVC()
svm_model2.fit(X_train2, y_train_satisfaction)
```

▼ SVC ⓘ ⓘ  
SVC()

```
from sklearn.neighbors import KNeighborsClassifier

knn_model1 = KNeighborsClassifier()
knn_model1.fit(X_train1, y_train_stress)

knn_model2 = KNeighborsClassifier()
knn_model2.fit(X_train2, y_train_satisfaction)
```


▼ KNeighborsClassifier ⓘ ⓘ  
KNeighborsClassifier()




# Comparison to Random Forest Classifier



## SVC

- SVC have higher **accuracy** compared to Random Forest
  - However, its **precision score** is especially **low**
- 

## KNN

- Perform **similarly** to Random Forest in terms of **all the metrics**
- 



# Short Summary

When choosing the **best/right** model for a problem, it is important **keep in mind** the following considerations:

- Type of Problem (e.g. Classification, Regression)
- Data you are working with (e.g. Size of Dataset, Type of Data)
- Compare Performance Metrics (e.g. Accuracy, Precision)
- Consider the Context of the Problem
- Don't limit to just one kind of model, experiment with multiple models.



# Outcome of Project



- Did not directly solve our original problem
- Identified top 5 factors for stress and satisfaction respectively



- Stress and satisfaction are closely linked
- Nature of this relationship is unclear
  - Whether it is positive or negative

# Recommendation



Number of events volunteered is a top factor



Universities could provide opportunities to cater to a broader range of students



Potentially address stress levels and enhance student satisfaction



# Conclusion

- Shed light on the **profound impact** of stress
- Identify **top 5** factors affecting stress and satisfaction
- Sets the path for **future exploration** and **intervention strategies**



# THANKS!

CREDITS: This presentation template was created by  
**Slidesgo**, including icons by **Flaticon**, infographics & images by  
**Freepik** and illustrations by **Storyset**.

# References

[https://www.flaticon.com/free-icon/decision-tree\\_5139787](https://www.flaticon.com/free-icon/decision-tree_5139787)

<https://stock.adobe.com/images/random-forest-line-icon-decision-trees-symbol-machine-learning-technique-that-s-used-to-solve-regression-and-classification-problems-complex-problems-solution-vector-illustration-flat-clip-art/474661732>

<https://www.talk-business.co.uk/2017/09/05/stress-university-cope/>

<https://ed.stanford.edu/news/just-breathe-simple-changes-can-reduce-student-stress-and-improve-learning-say-stanford>

<https://reflectandrefresh.org/2020/10/06/5-ways-to-expand-your-social-circle-in-college/>

<https://www.futurarc.com/project/academic-building-south-in-nanyang-technological-university/>

<https://blogs.worldbank.org/en/education/six-s-quality-education>