WIH3001 DATA SCIENCE PROJECT
2023/2024

ONLINE ASSESSMENT CLASSIFICATION BASED ON PERSONALISATION

BY JASMINE CHONG SEE YAN (S2132419)

SUPERVISED BY AP. DR. NOR LIYANA BT MOHD SHUIB

OVERVIEW

- Introduction
- Problem Statement
- Objectives
- Data Science Methodology
  - Analysis & Modelling Work
  - Tools & Coding
- Apps Demonstration

01.
INTRODUCTION

INTRODUCTION

01

ONLINE ASSESSMENT

Multiple-choice questions, essays, oral presentations and project-based cases. However, **different students have different preferences for the type of assessment they wish to undertake.**

INTRODUCTION

PERSONALISATION

Personalisation in education **raises the motivation and interests** of the learners, which are **critical success** factors in the learning as well as in the assessment process. (Hosny & El-Korany, 2022)

01

02

ONLINE ASSESSMENT

Multiple-choice questions, essays, oral presentations and project-based cases. However, **different students have different preferences for the type of assessment they wish to undertake.**

INTRODUCTION

PERSONALISATION

Personalisation in education **raises the motivation and interests** of the learners, which are **critical success** factors in the learning as well as in the assessment process. (Hosny & El-Korany, 2022)

01

ONLINE ASSESSMENT

Multiple-choice questions, essays, oral presentations and project-based cases. However, **different students have different preferences for the type of assessment they wish to undertake.**

02

03

LEARNING STYLE

Learning style influences how students prefer to receive and process information.

INTRODUCTION

01

ONLINE ASSESSMENT

Multiple-choice questions, essays, oral presentations and project-based cases. However, **different students have different preferences for the type of assessment they wish to undertake.**

PERSONALISATION

Personalisation in education **raises the motivation and interests** of the learners, which are **critical success** factors in the learning as well as in the assessment process. (Hosny & El-Korany, 2022)

02

03

LEARNING STYLE

Learning style influences how students prefer to receive and process information.

VAK LEARNING STYLE MODEL

Research by Hosny and El-Korany, 2022 has established that the VAK Learning Style Model is the **most widely used** among the learning style assessment instruments

04

INTRODUCTION

01

ONLINE ASSESSMENT

Multiple-choice questions, essays, oral presentations and project-based cases. However, **different students have different preferences for the type of assessment they wish to undertake.**

02

PERSONALISATION

Personalisation in education **raises the motivation and interests** of the learners, which are **critical success** factors in the learning as well as in the assessment process. (Hosny & El-Korany, 2022)

03

LEARNING STYLE

Learning style influences how students prefer to receive and process information.

04

VAK LEARNING STYLE MODEL

Research by Hosny and El-Korany, 2022 has established that the VAK Learning Style Model is the **most widely used** among the learning style assessment instruments

05

Assessments should be conveyed in various forms appropriate for each student. (Maya et al., 2021)

The VAK Learning Style Model

- Developed by psychologists in the 1920s to **classify the most common ways that people learn**.
- According to the model, most of us prefer to learn in **one of three ways: visual, auditory or kinesthetic**
- 30 scenarios (questions) to answer to determine your dominant learning style

The VAK Learning Style Model

- Developed by psychologists in the 1920s to **classify the most common ways that people learn**.
- According to the model, most of us prefer to learn in **one of three ways: visual, auditory or kinesthetic**
- 30 scenarios (questions) to answer to determine your dominant learning style

What kind of learner are you?

Learner	Preference
Visual	Seeing and reading
Auditory	Listening and speaking
Kinesthetic	Touching and doing

Table 1: Learners characteristics

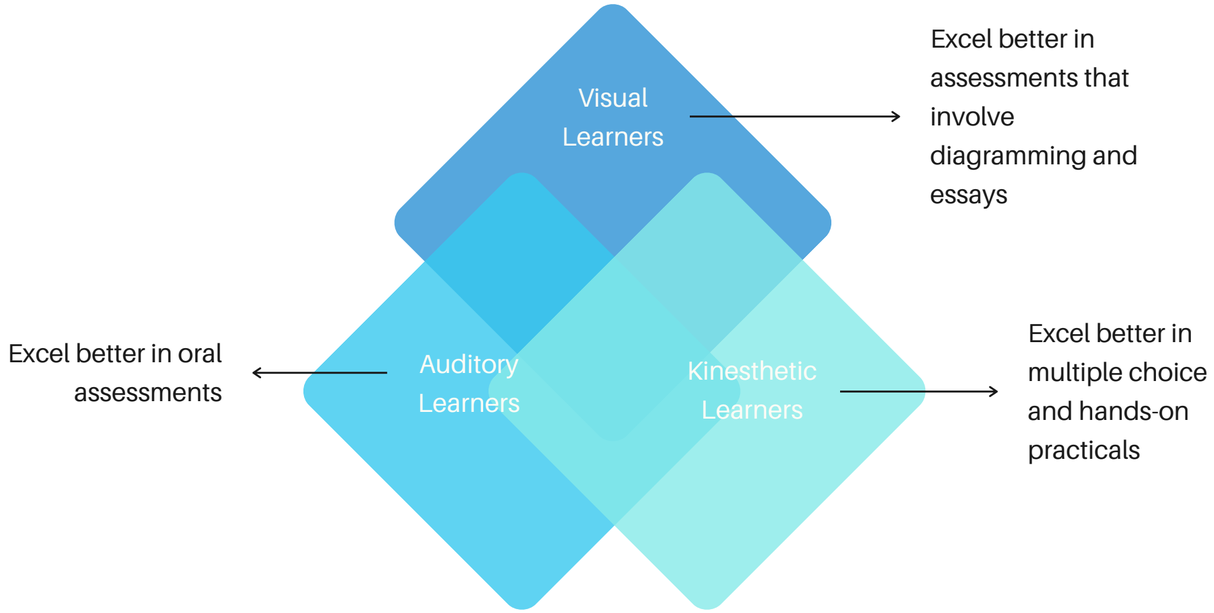


Figure 1: Example assessments based on the VAK Learning Style Model. (Wickramasinghe & Hettiarachchi , 2017)

VAK Question Example

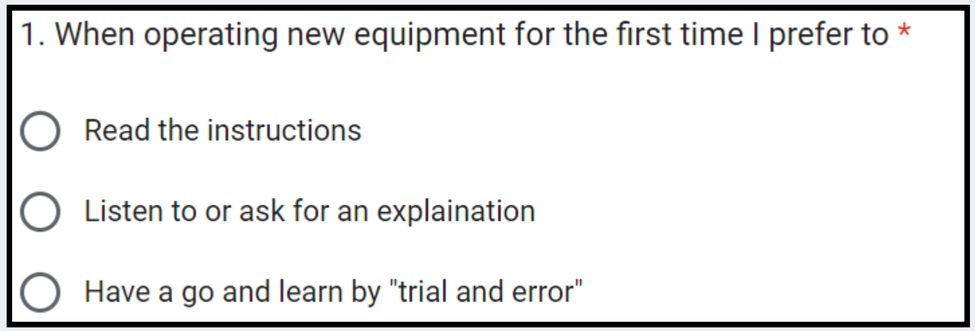


Figure 2: Sample VAK question

- A **visual** learner would most likely prefer to **‘Read the instructions’** (Option A)
- An **auditory** learner would most likely prefer to **‘Listen to or ask for an explanation’** (Option B)
- A **kinesthetic** learner would most likely prefer to **‘Have a go and learn by trial and error’** (Option C)

VAK Question Example

1. When operating new equipment for the first time I prefer to *

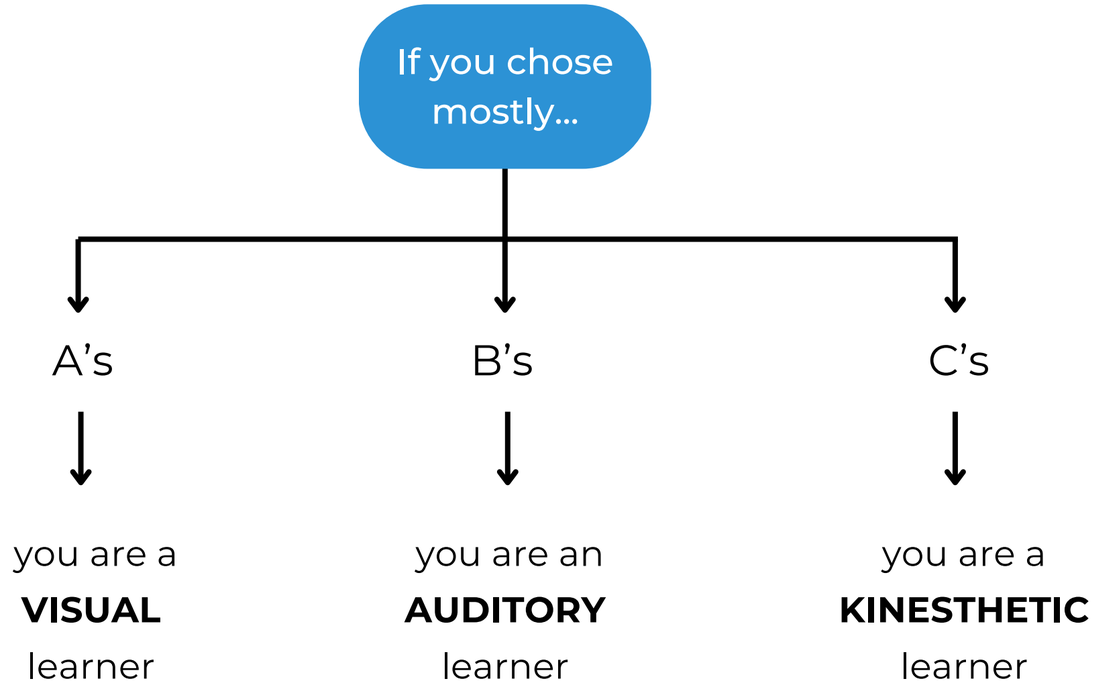
- ◯ Read the instructions
- ◯ Listen to or ask for an explaination
- ◯ Have a go and learn by "trial and error"

Figure 2: Sample VAK question

- A **visual** learner would most likely prefer to **‘Read the instructions’** (Option A)
- An **auditory** learner would most likely prefer to **‘Listen to or ask for an explanation’** (Option B)
- A **kinesthetic** learner would most likely prefer to **‘Have a go and learn by trial and error’** (Option C)

How does the VAK model determine your dominant learning style?

- Each option (A, B & C) represents a learning style
- The responses are calculated by totalling every As, Bs and Cs.



02.
PROBLEM STATEMENT

LITERATURE ANALYSIS

References	Objectives	Methods	Results	Limitations
Hosny and El-Korany (2022)	• **Detection of students' learning styles from learning activities** and recommend suitable assessment methods	• Clustering techniques (KMeans, DBScan & Expectation–maximization)	• Proposed a **learning style identifier and recommend suitable assessment** methods system with KMeans exhibits the highest accuracy of 95%	• **Assessment methods recommendation** for each learning style is solely **based on the characteristics of that learning style** • **Do not** take student's assessment methods preference into account
Maya et al. (2021)	• Explore the **influence of the different learning styles** on academic performance according to the **assessment methods**	• Descriptive analysis using SPSS	• Using a **variety of assessment methods** can help students with **different learning profiles** to demonstrate their competencies effectively.	
Wickramsinghe and Hettiarachchi (2017)	• Identify **students' learning styles and their relationship with assessment methods**	• Descriptive analysis with qualitative research methodology	• Students **perform better** in assessment methods that align with their learning styles	

Table 2: Learning style and assessment methods literature review

LITERATURE ANALYSIS

References	Title	Models	Evaluation	Limitations
Agarwal et al. (2021)	• Classification model for accuracy and intrusion detection using machine learning approach	• **KNN**, Naïve Bayes, **SVM**	• SVM has the highest accuracy of 95% • Naive Bayes has the lowest accuracy of 92%	• Results can be **improved** by comparing **with Random Forest**
Santana et al. (2021)	• Classification Models for COVID-19 Test Prioritization in Brazil: Machine Learning Approach	• **Decision Tree**, Logistic Regression, MLP, **Random Forest, SVM, XGBoost**	• All **except logistic regression** exhibit high performances with similar accuracy (~94%)	• Not specified
Zhang et al. (2020)	• Mapping Rice Paddy Based on Machine Learning with Sentinel-2 Multi-Temporal Data: Model Comparison and Transferability	• MLP, **Random Forest, SVM, XGBoost**	• XGBoost has the highest accuracy of 89.73% in Banan District • SVM has the highest accuracy of 88.57% in Zhongxian County • XGBoost and random forest has the highest model transferability	• Performance of **MLP** was **unstable**

Table 3: Classification models literature review

PROBLEM STATEMENT

Problem 1

- Most research focuses on recommending assessment solely based on learning style characteristics without taking learners' preferences into account (Hosny and El-Korany, 2022; Maya et al., 2021; Wickramsinghe and Hettiarachchi, 2017)

Problem 2

- Currently, there is **limited study** in using a classification model for online assessment based on personalisation of learner's learning styles and preferences

03.
OBJECTIVES

OBJECTIVES

Objective 1

To **develop** an online assessment **classification model** based on personalisation

Objective 2

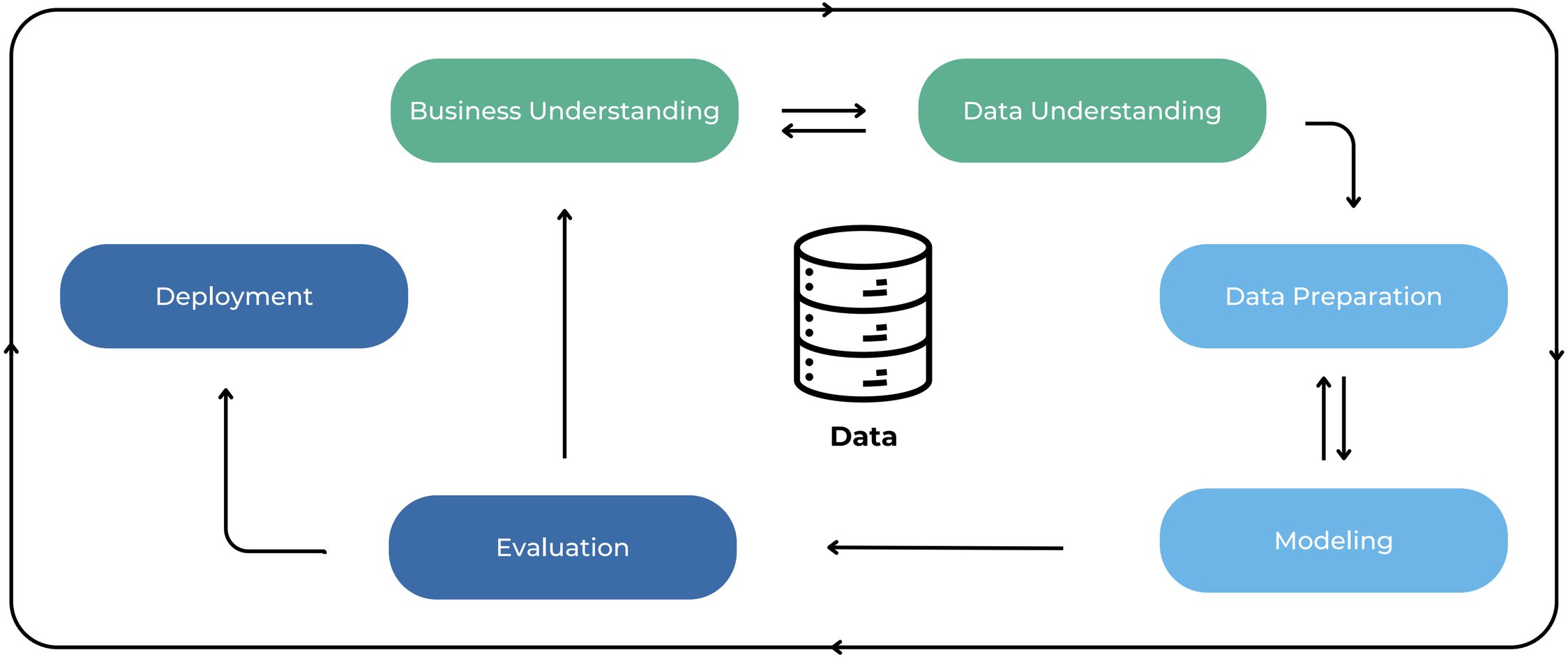
To **evaluate** the online assessment **classification model** based on personalisation

Objective 3

To **develop a data product** that functions as an online assessment methods recommendation

04.
DATA SCIENCE (DS) METHODOLOGY

CRISP-DM


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

Business Understanding

1. **Understand the topic,** particularly on the learning style, assessments and personalisation through **literature analysis**

Data Understanding

a. Data was collected via a **Google Form survey** from **2021 to 2023**

b. The survey questions include demographics, preferred learning objects, preferred online instructional strategies and assessment methods, learning style awareness and the VAK Learning Style questions

c. Data consists of **1052 rows and 104 columns**

d. **Understand the data** to identify the relevant columns using **SAS Enterprise Miner**

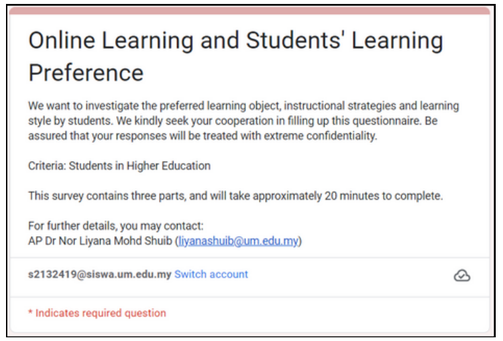


Figure 3: Google form survey

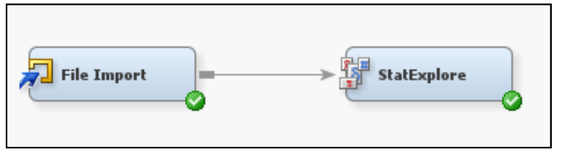


Figure 4: Process flow for data exploration (SAS)

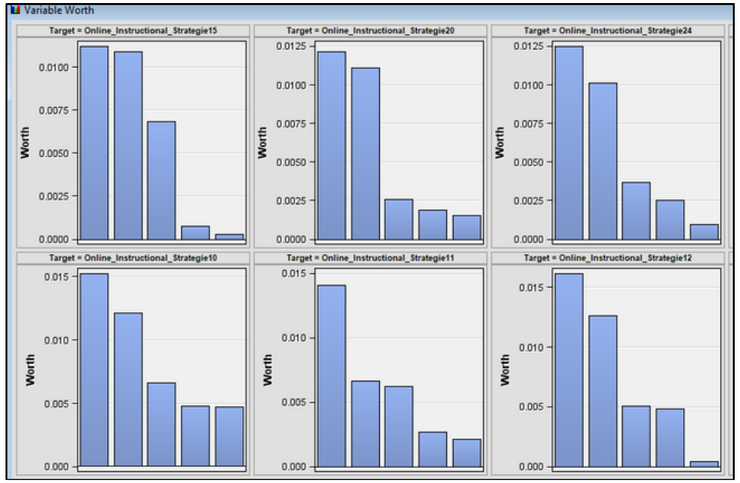


Figure 5: Input variable worth against target (SAS)

Data Preparation - Data Profiling and Feature Selection

a. Drop the test data row

---

b. Check for **null values**

- The dataset does not contain any null values

---

c. **Feature selection**

- Drop irrelevant columns (e.g., timestamp, household income, preferred learning mode, preferred social media platform, learning objects and preferred online instructional strategies)
- The final columns:
  - Gender
  - Level of Study
  - Preferred Communication Platform
  - VAK Questions

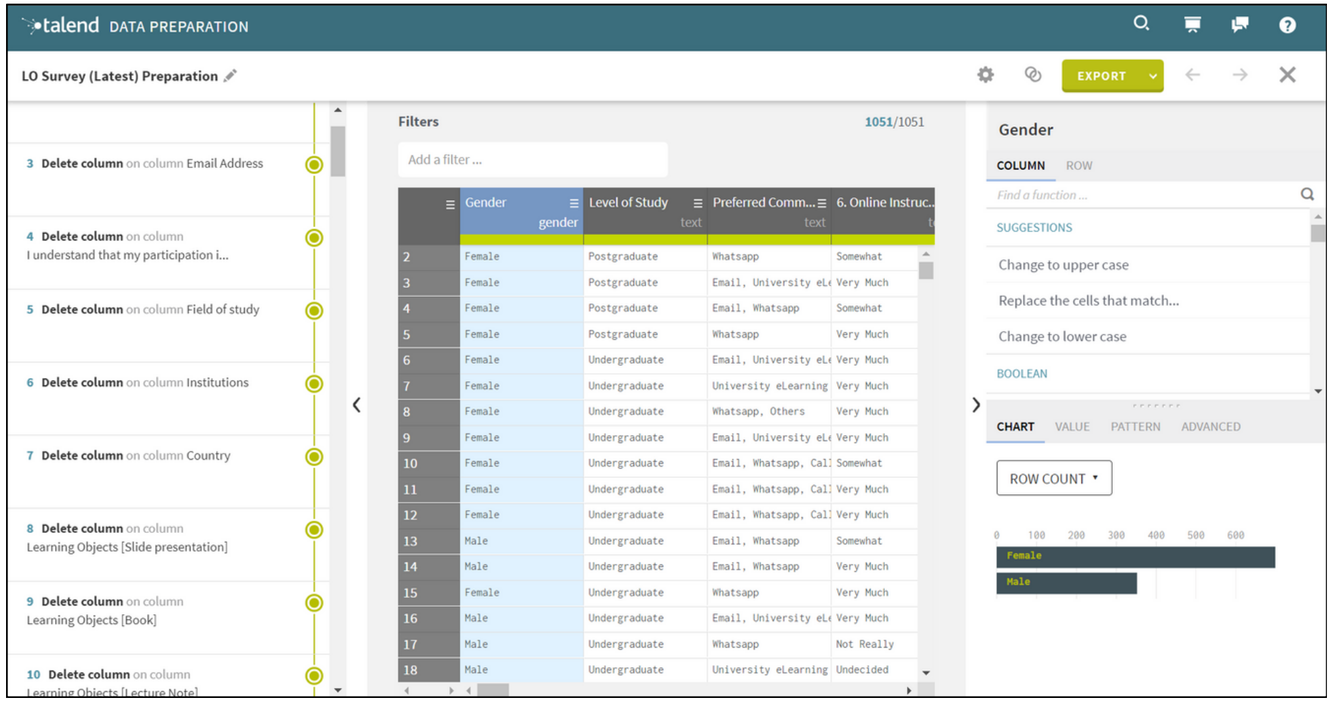


Figure 6: Talend Data Preparation work space

Data Preparation - Data Standardisation (Part I)

a. Categories **'Master' and 'PhD' as Postgraduate**

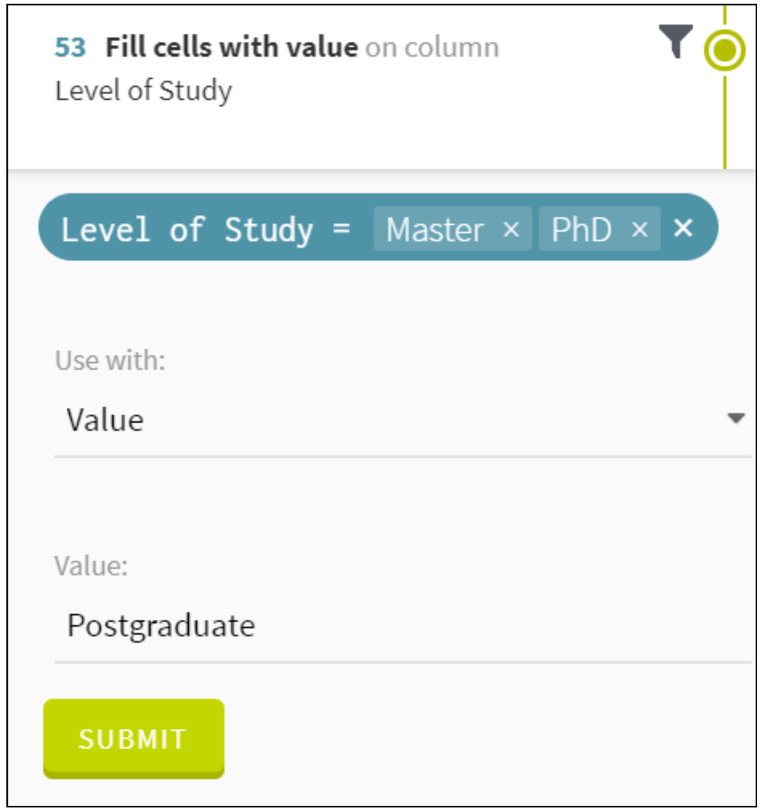


Figure 7: Rename to 'Postgraduate' using the 'Fill cells with value' function

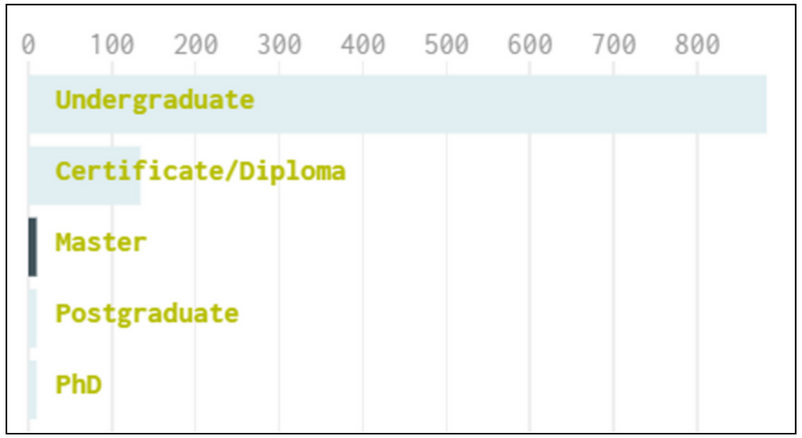


Figure 8: Data before standardisation

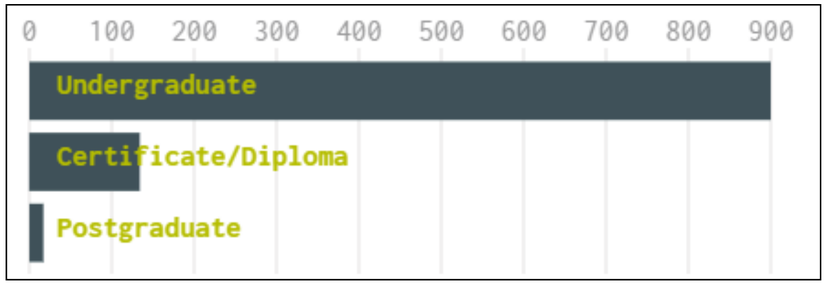


Figure 9: Data after standardisation

Data Preparation - Data Standardisation (Part 2)

b. Categories the **minority responses of 'Preferred Communication Platform'** column as **'Others'**

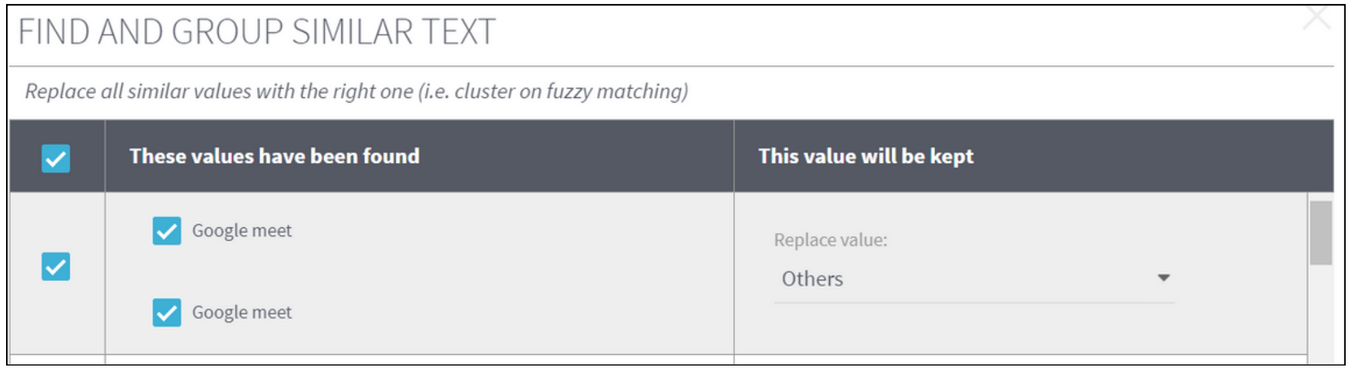


Figure 10: Rename to 'Others' using 'Find and group similar text' function

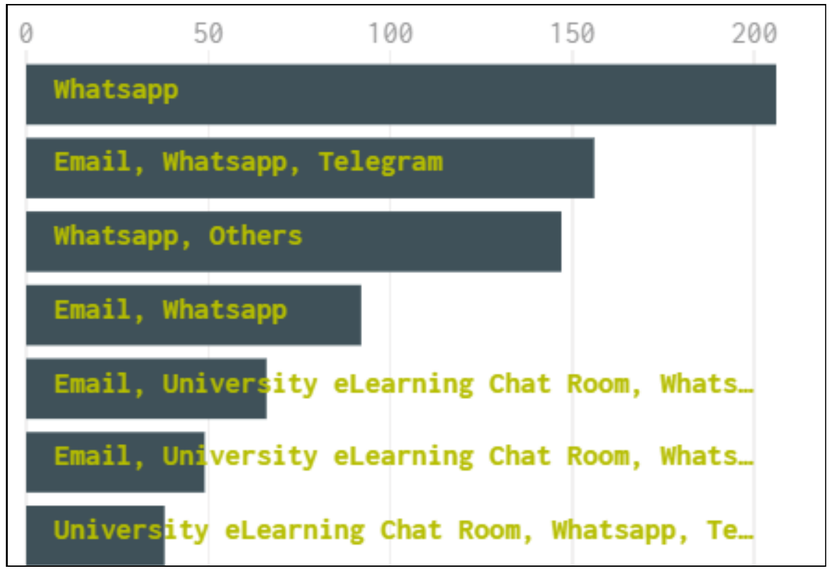


Figure 11: Data after standardisation

talend
A Qlik Company

Data Preparation - Data Standardisation (Part 3)

c. **Correct spelling mistake** in the survey responses ‘Hoe they make me’ to ‘How they make me’

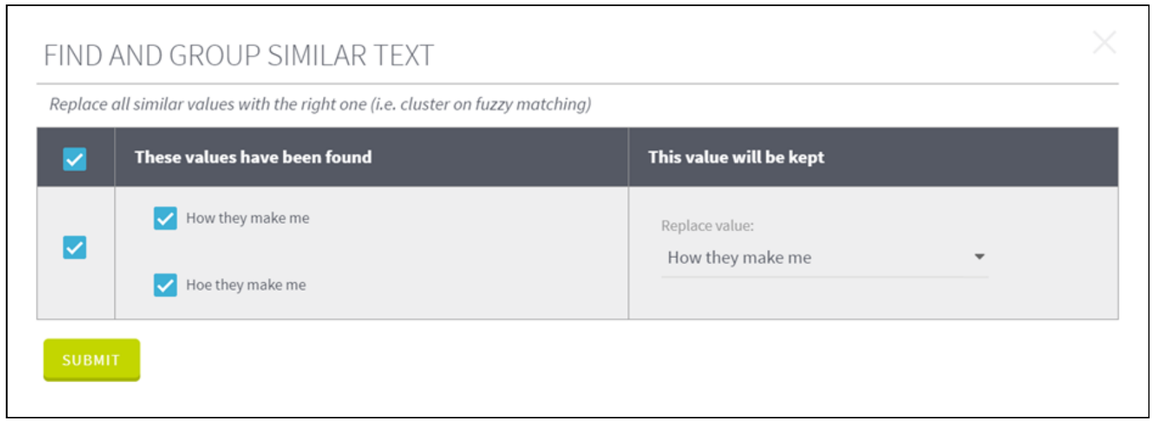


Figure 12: Correct the spelling error using ‘Find and group similar text’ function

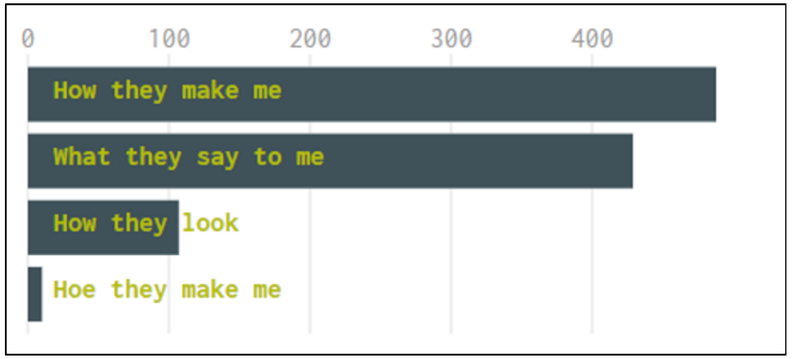


Figure 13: Data before standardisation

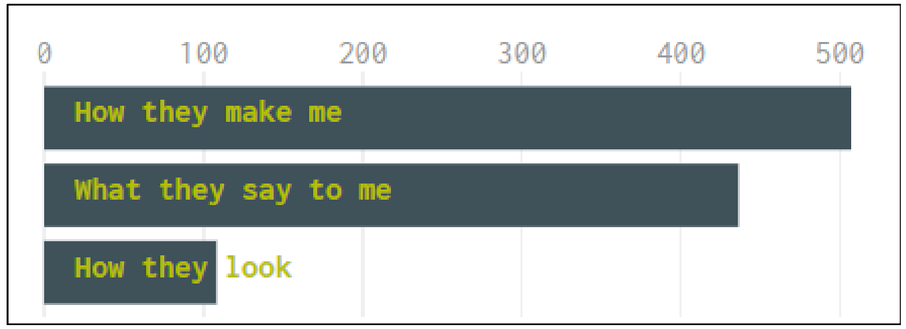


Figure 14: Data after standardisation

Data Preparation - Identify Dominant Learning Style (Part I)

a. Define the **answers** for each learning style

```
# Define answers options
visual_keywords = ["Read the instructions",
                   "Look at a map",
                   "Follow a recipe",
                   "Write Instructions",
```

```
auditory_keywords = ["Listen to or ask for an explaination",
                     "Ask for spoken directions",
                     "Call a friend for explaination",
                     "Explain verbally",
```

```
kinesthetic_keywords = ["Have a go and learn by \"trial and error\"",
                        "Follow my nose or maybe use a compass",
                        "Follow my instinct, tasting as I cook",
                        "Demonstrate and let them have a go",
```

Figure 15: Sample answer for visual, auditory and kinesthetic learners

Data Preparation - Identify Dominant Learning Style (Part 2)

b. **Calculate the sum** of the responses for each learning style selected

```python
# Iterate through the columns (each column is a question)
for column in vak_df.columns:
    response = row[column].lower()

    # compare answer with the response
    for answer in visual_keywords:
        if answer.lower() in response:
            visual_count += 1

    for answer in auditory_keywords:
        if answer.lower() in response:
            auditory_count += 1

    for answer in kinesthetic_keywords:
        if answer.lower() in response:
            kinesthetic_count += 1

# Determine the dominant VAK preference for this respondent
preferences = {
    "Visual": visual_count,
    "Auditory": auditory_count,
    "Kinesthetic": kinesthetic_count
}

dominant_preference = max(preferences, key=preferences.get)
dominant_learning_style.append(dominant_preference)
```

Figure 16: Determine the dominant learning style's Python code

c. **Append** to the dataset as a **new column** called Dominant_VAK

```python
# Add the list of dominant learning style as a new column in the DataFrame
vak_df['Dominant_VAK'] = dominant_learning_style
```

Figure 17: Add a new column to dataset

Dominant_VAK
Visual
Visual
Visual
Auditory

Figure 18: Sample output of dominant VAK column

python™

Data Preparation - Exploratory Data Analysis (Part I)

a. Dominant learning style

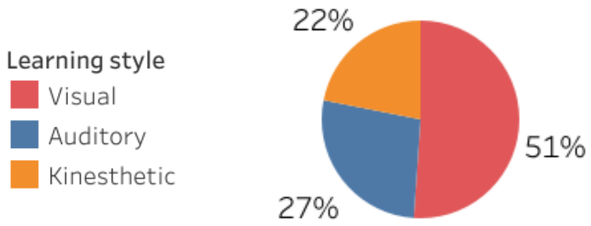


Figure 19: Proportion of visual, auditory and kinesthetic learners

b. Gender

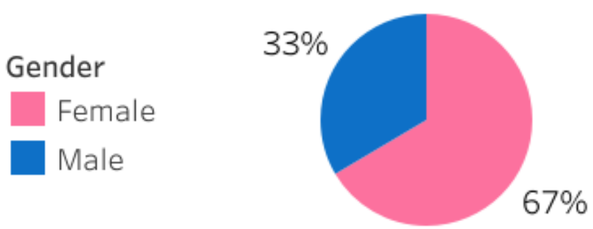


Figure 20: Proportion of female and male

c. Level of study

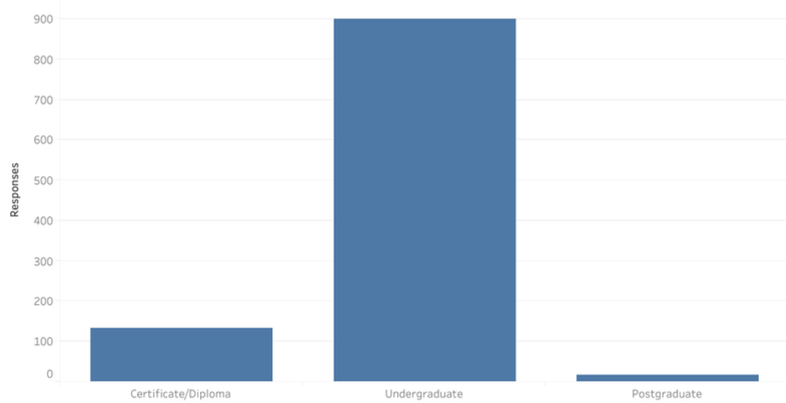


Figure 21: Proportion of certificate/diploma, undergraduate and postgraduate

Data Preparation - Exploratory Data Analysis (Part 2)

d. Top 10 preferred communication platform

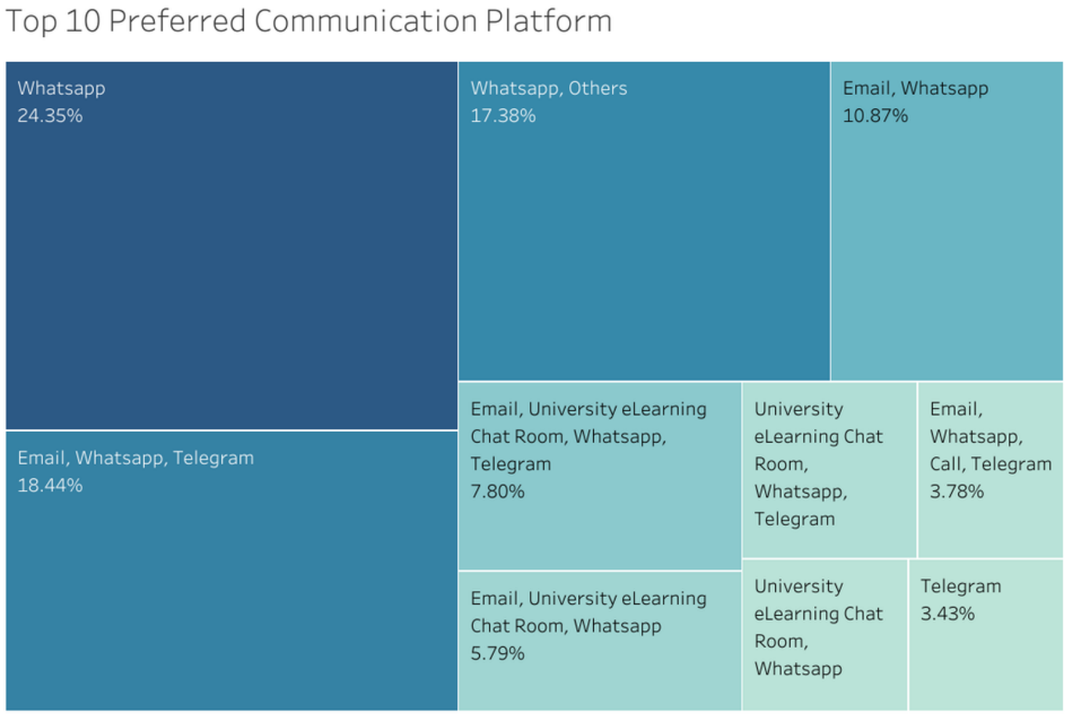


Figure 22: Preferred communication platform

e. Learning style importance and awareness

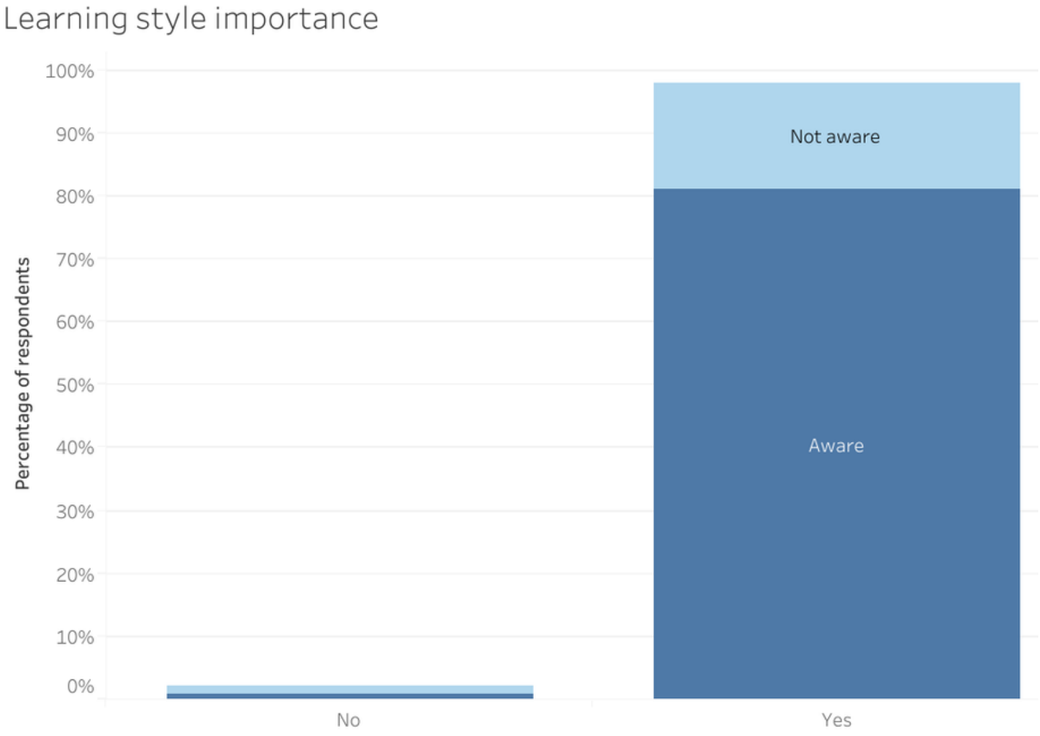


Figure 23: Proportion of whether knowing one's learning style is important in improving their learning ability in comparison with the awareness of one's learning style

Data Preparation - Exploratory Data Analysis (Part 3)

f. **Visual** learners' online assessment preference

- Based on the VAK model, visual learners should prefer seeing and reading

Surprisingly, **'Demonstration'** which mostly involves 'touching' and 'doing' is the **most preferred** online assessment method

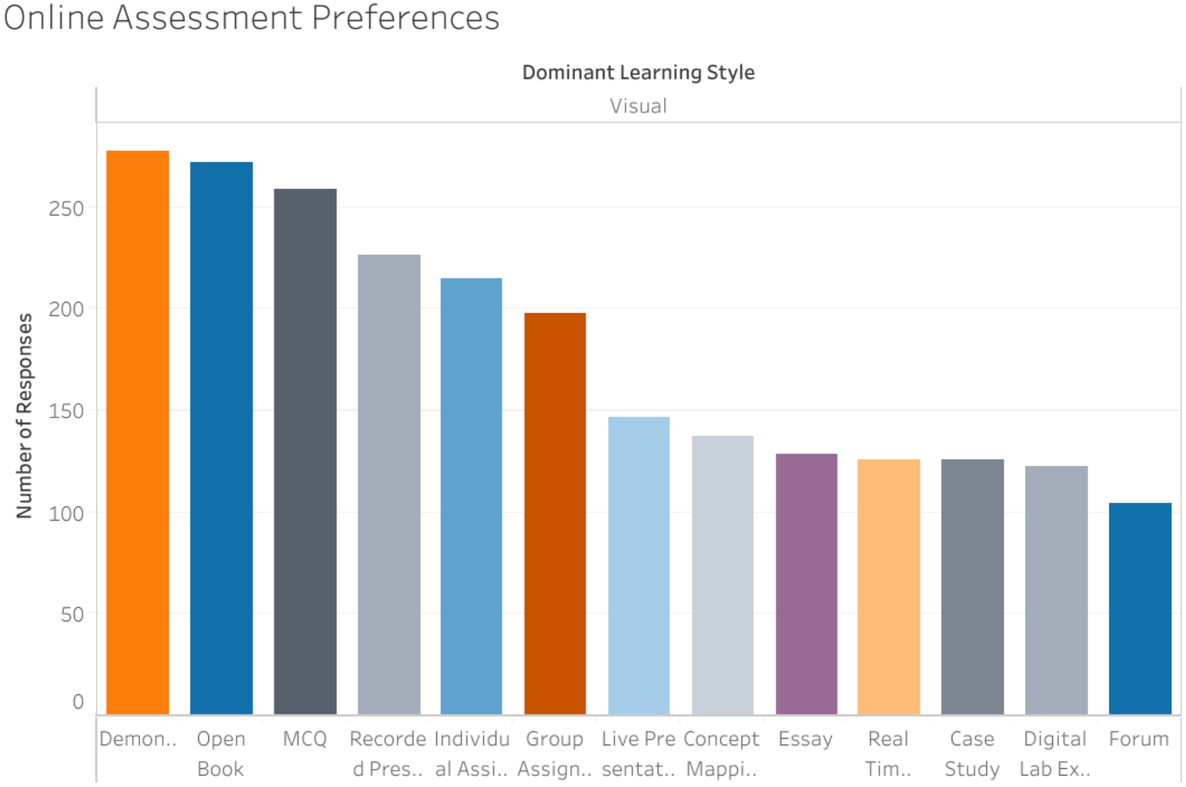


Figure 24: Online assessment methods preference for visual learners

Data Preparation - Exploratory Data Analysis (Part 4)

g. **Auditory** learners' online assessment preference

- Based on the VAK model, auditory learners should prefer listening and speaking

Surprisingly, **'Open book'** which involves 'seeing' and 'reading' is the **most preferred online assessment method** instead of recorded or live presentation which mostly involves 'listening and speaking'.

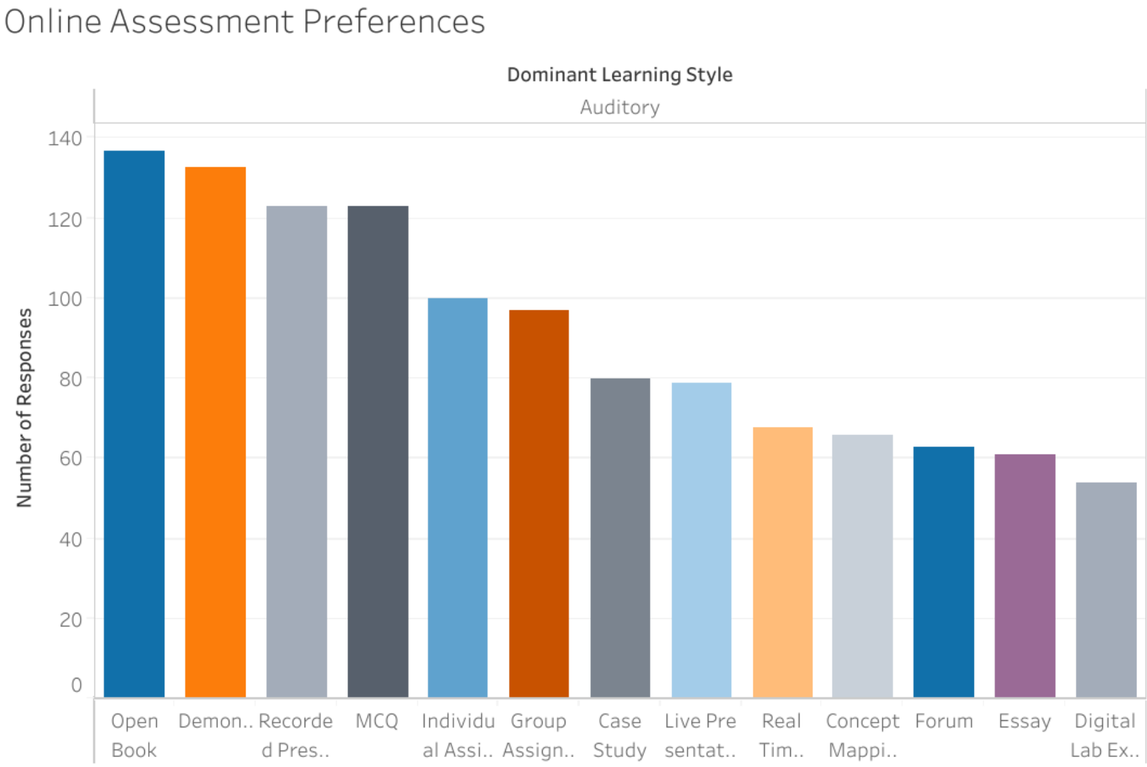


Figure 25: Online assessment methods preference for auditory learners

tableau

Data Preparation - Exploratory Data Analysis (Part 5)

h. **Kinesthetic** learners' online assessment preference

- Based on the VAK model, kinesthetic learners should prefer touching and doing

It is not a surprise that **'Demonstration'** is the **most preferred** online assessment method. However, one thing to highlight is 'Recorded presentation' is preferred over 'Digital lab experiment' which involve 'touching' and 'doing'.

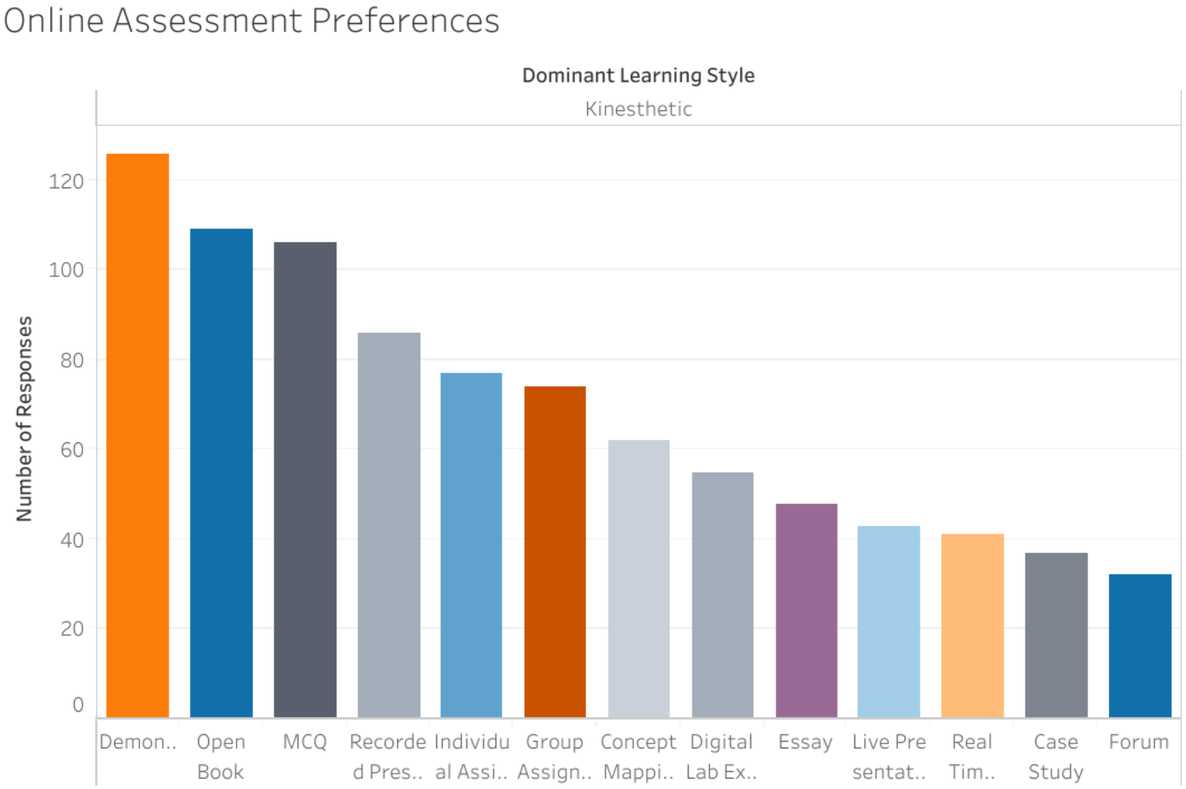


Figure 26: Online assessment methods preference for kinesthetic learners

tableau

Data Preparation - Exploratory Data Analysis (Part 5)

h. **Kinesthetic** learners' online assessment preference

- Based on the VAK model, kinesthetic learners should prefer touching and doing

It is not a surprise that **'Demonstration'** is the **most preferred** online assessment method. However, one thing to highlight is 'Recorded presentation' is preferred over 'Digital lab experiment' which involve 'doing'.

This emphasises the need to consider visual, auditory and kinesthetic learners' preferences instead of assuming their preferences based on the learning style characteristics.

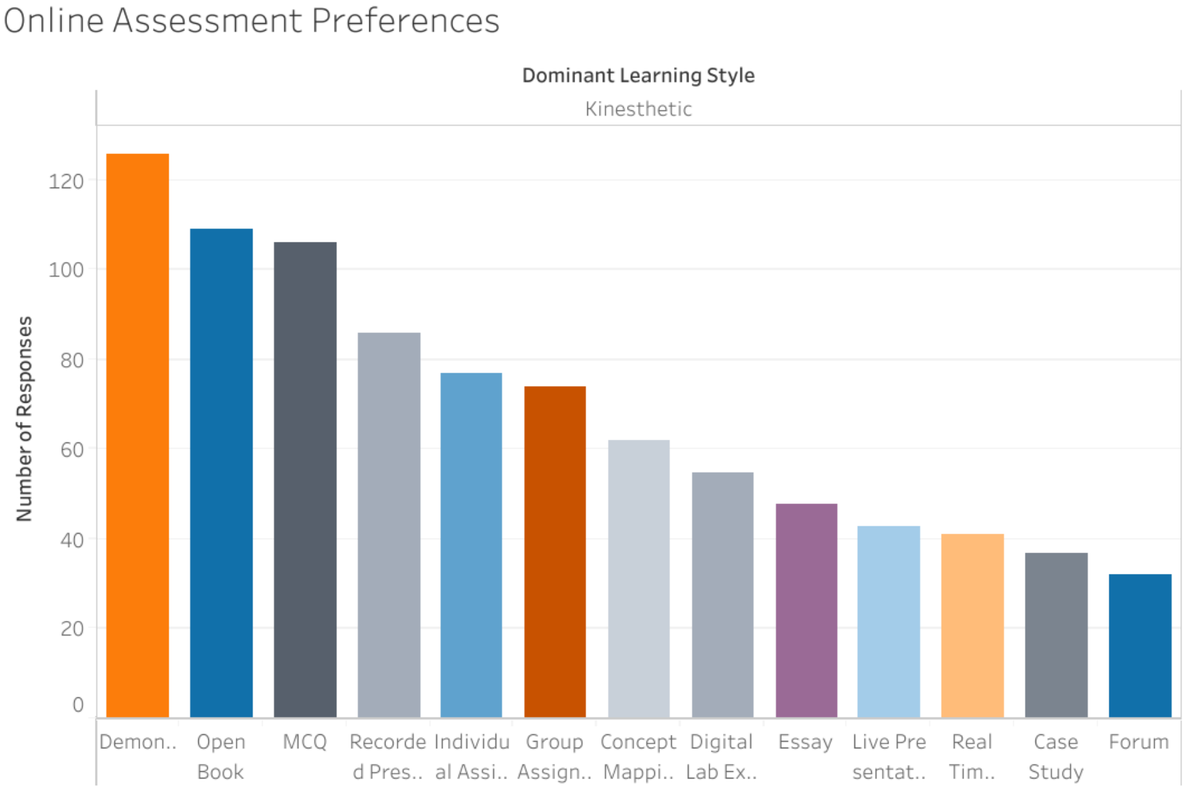


Figure 26: Online assessment methods preference for kinesthetic learners

Data Preparation - Data Exploding

a. Split **multiple selection answers** question ('Preferred Communication Platform') **into individual rows**

```
df['Preferred Communication Platform'] = df['Preferred Communication Platform'].str.split(', ')
df = df.explode('Preferred Communication Platform')

# Reset index after exploding
df = df.reset_index(drop=True)

df.head()
```

Figure 27: Python code for data exploding

1	Female	Postgraduate	Email, University eLearning Chat Room, Whatsap...

1	Female	Postgraduate	Email
2	Female	Postgraduate	University eLearning Chat Room
3	Female	Postgraduate	Whatsapp
4	Female	Postgraduate	Call

Figure 28: Sample output after data exploding

python™

Data Preparation - Data Encoding (Part I)

a. **Ordinal encoding** for ordinal categorical data

```
preference_mapping = {
    'Not at All': 0,
    'Not Really': 0,
    'Undecided': 0,
    'Somewhat': 0,
    'Very Much': 1
}
```

```
for column in online_asssessment_columns.columns:
    df[column] = df[column].map(preference_mapping)

df.head()
```

Figure 29: Python code for ordinal encoding

	Gender	Level of Study	6. Online Instructional Strategies/Assessment [Demonstration]	6. Online Instructional Strategies/Assessment [Digital Lab Experiments]	6. Online Instructional Strategies/Assessment [Forum]	6. Online Instructional Strategies/Assessment [Case Study]
0	0	3	0	0	0	0
1	0	3	1	1	0	0
2	0	3	1	1	0	0
3	0	3	1	1	0	0

Figure 30: Sample output for ordinal encoding

python™

Data Preparation - Data Encoding (Part 2)

b. **One-hot encoding** for nominal categorical data

```
df = pd.get_dummies(df, columns=columns_to_encode, prefix=columns_to_encode)
```

Figure 31: Python code for one-hot encoding

Preferred Communication Platform_Call	Preferred Communication Platform_Email	Preferred Communication Platform_Others	Preferred Communication Platform_Telegram	Preferred Communication Platform_University eLearning Chat Room	Preferred Communication Platform_Whatsapp
False	False	False	False	False	True
False	True	False	False	False	False

Figure 32: Sample output for one-hot encoding

Data Preparation - Reduce Noisy Data

- Remove rows where respondents **did not select any** online assessment methods as their preferred tools

6. Online Instructional Strategies/Assessment [Demonstration]	6. Online Instructional Strategies/Assessment [Digital Lab Experiments]	6. Online Instructional Strategies/Assessment [Forum]	6. Online Instructional Strategies/Assessment [Case Study]	6. Online Instructional Strategies/Assessment [Concept Mapping]
0	0	0	0	0

Figure 33: Sample rows where none of the online assessment tools are select as preferred

```python
columns_to_check = [
    '6. Online Instructional Strategies/Assessment [Demonstration]',
    '6. Online Instructional Strategies/Assessment [Digital Lab Experiments]',
    '6. Online Instructional Strategies/Assessment [Forum]',
    '6. Online Instructional Strategies/Assessment [Case Study]',
    '6. Online Instructional Strategies/Assessment [Concept Mapping]',
    '6. Online Instructional Strategies/Assessment [Real Time Online Exam]',
    '6. Online Instructional Strategies/Assessment [Individual Project/Assignment]',
    '6. Online Instructional Strategies/Assessment [Group Project/Assignment]',
    '6. Online Instructional Strategies/Assessment [Online Quiz/Test - MCQ]',
    '6. Online Instructional Strategies/Assessment [Online Quiz/Test - Essay]',
    '6. Online Instructional Strategies/Assessment [Online Quiz/Test - Open Book]',
    '6. Online Instructional Strategies/Assessment [Peer Review Assessment Live Presentation]',
    '6. Online Instructional Strategies/Assessment [Recorded Presentation]'
]

# Check if all values in the specified columns are equal to 0
rows_to_drop = df[df[columns_to_check].eq(0).all(axis=1)]

# Drop the rows
df = df.drop(rows_to_drop.index)

df.head()
```

Figure 34: Python code to remove rows where preference was not selected

python™

Modelling (Part 1)

a. **Extract the target** columns

```
# Extract online assessment columns as target
target_columns = df.iloc[:, 2:15]
target_columns.columns
```

```
Index(['6. Online Instructional Strategies/Assessment [Demonstration]',
       '6. Online Instructional Strategies/Assessment [Digital Lab Experiments]',
       '6. Online Instructional Strategies/Assessment [Forum]',
       '6. Online Instructional Strategies/Assessment [Case Study]',
       '6. Online Instructional Strategies/Assessment [Concept Mapping]',
       '6. Online Instructional Strategies/Assessment [Real Time Online Exam]',
       '6. Online Instructional Strategies/Assessment [Individual Project/Assignment]',
       '6. Online Instructional Strategies/Assessment [Group Project/Assignment]',
       '6. Online Instructional Strategies/Assessment [Online Quiz/Test - MCQ]',
       '6. Online Instructional Strategies/Assessment [Online Quiz/Test - Essay]',
       '6. Online Instructional Strategies/Assessment [Online Quiz/Test - Open Book]',
       '6. Online Instructional Strategies/Assessment [Peer Review Assessment Live Presentation]',
       '6. Online Instructional Strategies/Assessment [Recorded Presentation]'],
      dtype='object')
```

Figure 35: Python code to extract target columns

b. Split dataset to 70% training set and 30% testing set using train_test_split from sklearn

```
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(df.drop(target_columns.columns, axis=1), target_columns,
                                                    test_size=0.3, random_state=42)
```

Figure 36: Python code to split dataset

python™

Modelling (Part 2)

c. Train and fit **5 classification models** to classify each online assessment tool (13 target variables)

- Decision Tree
- Extreme Gradient Boost (XGBoost)
- K-Nearest Neighbour (KNN)
- Random Forest
- Support Vector Machine (SVM)

```python
# Initialise an empty dictionary to store the trained model
model_dict = {}

# Train a Random Forest Classifier
for column in target_columns.columns:
    model = RandomForestClassifier(random_state=42)
    model.fit(X_train, y_train[column])
    model_dict[column] = model
```

```python
for column in target_columns.columns:
    model = KNeighborsClassifier()
    model.fit(X_train, y_train[column])
    model_dict[column] = model
```

```python
for column in target_columns.columns:
    model = SVC(kernel='rbf', random_state=42)
    model.fit(X_train, y_train[column])
    model_dict[column] = model
```

```python
for column in target_columns.columns:
    model = DecisionTreeClassifier(random_state=42)
    model.fit(X_train, y_train[column])
    model_dict[column] = model
```

```python
for column in target_columns.columns:
    model = XGBClassifier(random_state=42)
    model.fit(X_train, y_train[column])
    model_dict[column] = model
```

Figure 37: Python code to build, train and fit the classification model

python™

Evaluation (Part I)

- **Evaluate every model's performance** using classification_report and accuracy_score from sklearn

```python
# Make predictions on the test set
y_pred = pd.DataFrame({col: model.predict(X_test) for col, model in model_dict.items()})

# Classification Report
print("Classification Report:")
print(classification_report(y_test, y_pred))
```

```
Classification Report:
              precision    recall  f1-score   support

           0       0.91      0.98      0.95       389
           1       0.98      0.77      0.86       163
           2       1.00      0.83      0.91       140
           3       0.96      0.86      0.91       173
           4       0.99      0.74      0.85       179
           5       1.00      0.78      0.88       153
           6       0.93      0.85      0.89       265
           7       0.92      0.83      0.87       252
           8       0.88      0.95      0.91       344
           9       0.98      0.81      0.89       161
          10       0.87      0.97      0.92       359
          11       0.97      0.84      0.90       176
          12       0.89      0.89      0.89       312

   micro avg       0.93      0.88      0.90      3066
   macro avg       0.95      0.85      0.89      3066
weighted avg       0.93      0.88      0.90      3066
 samples avg       0.91      0.89      0.88      3066
```

Figure 38: Python code of to obtain the precision, recall and f1-score using classification report

```python
# Initialize a dictionary to store accuracy scores
accuracy_scores = {}

# Loop through each column and calculate accuracy score
for col in y_test.columns:
    accuracy = accuracy_score(y_test[col], y_pred[col])
    accuracy_scores[col] = accuracy
    print(f"Accuracy for {col}: {accuracy}")

# Overall accuracy score
overall_accuracy = accuracy_score(y_test.values.flatten(), y_pred.values.flatten())
print(f"\nOverall Accuracy: {overall_accuracy}")
```

```
Accuracy for 6. Online Instructional Strategies/Assessment [Demonstration]: 0.9294117647058824
Accuracy for 6. Online Instructional Strategies/Assessment [Digital Lab Experiments]: 0.9327731092436975
Accuracy for 6. Online Instructional Strategies/Assessment [Forum]: 0.9596638655462185
Accuracy for 6. Online Instructional Strategies/Assessment [Case Study]: 0.9478991596638655
Accuracy for 6. Online Instructional Strategies/Assessment [Concept Mapping]: 0.9210084033613445
Accuracy for 6. Online Instructional Strategies/Assessment [Real Time Online Exam]: 0.9428571428571428
Accuracy for 6. Online Instructional Strategies/Assessment [Individual Project/Assignment]: 0.9058823529411765
Accuracy for 6. Online Instructional Strategies/Assessment [Group Project/Assignment]: 0.8957983193277311
Accuracy for 6. Online Instructional Strategies/Assessment [Online Quiz/Test - MCQ]: 0.8957983193277311
Accuracy for 6. Online Instructional Strategies/Assessment [Online Quiz/Test - Essay]: 0.9445378151260504
Accuracy for 6. Online Instructional Strategies/Assessment [Online Quiz/Test - Open Book]: 0.8957983193277311
Accuracy for 6. Online Instructional Strategies/Assessment [Peer Review Assessment Live Presentation]: 0.946218487394958
Accuracy for 6. Online Instructional Strategies/Assessment [Recorded Presentation]: 0.8873949579831932

Overall Accuracy: 0.9234647705235941
```

Figure 39: Python code to obtain the overall accuracy using accuracy_score function

Evaluation (Part 2)

- Compare, select and **save the best-performing model** using joblib

Model	Accuracy (%)	Precision (%)	Recall (%)	FI-score (%)
Decision Tree	89.46348	86	85	86
XGBoost	92.11377	92	87	89
KNN	75.16484	67	62	64
Random Forest	**92.34648**	**95**	**85**	**89**
SVM	82.99935	86	67	73

Table 4: Evaluation of models

```
joblib.dump(model_dict, 'Model/rf.joblib')
```

```
['Model/rf.joblib']
```

Figure 40: Python code to save the model

Deployment

- Create a **web application** (Assessifier) using **Python's Streamlit**
- Deploy the data product

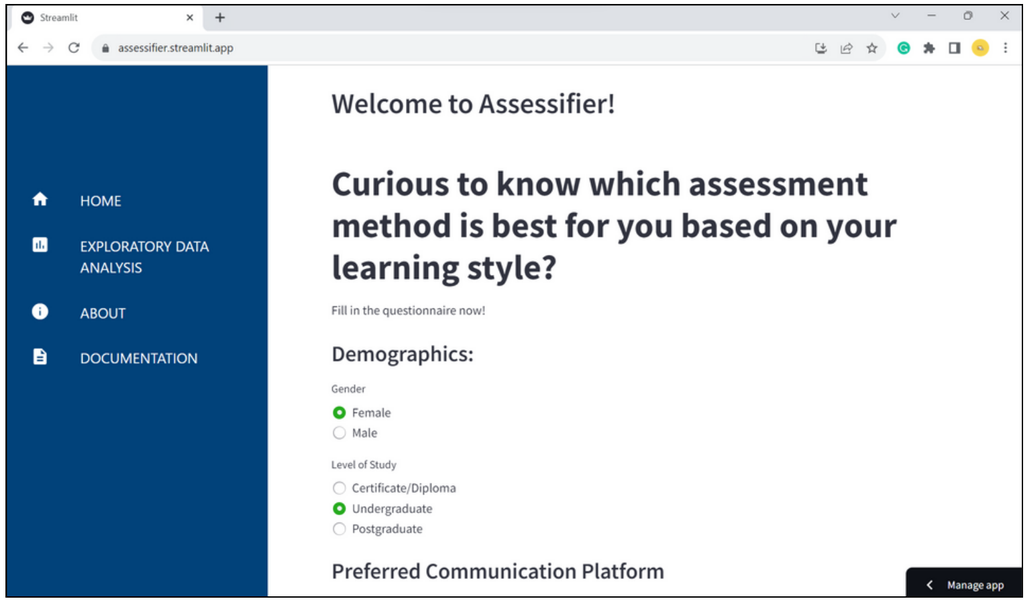


Figure 41: Data product (Web application)

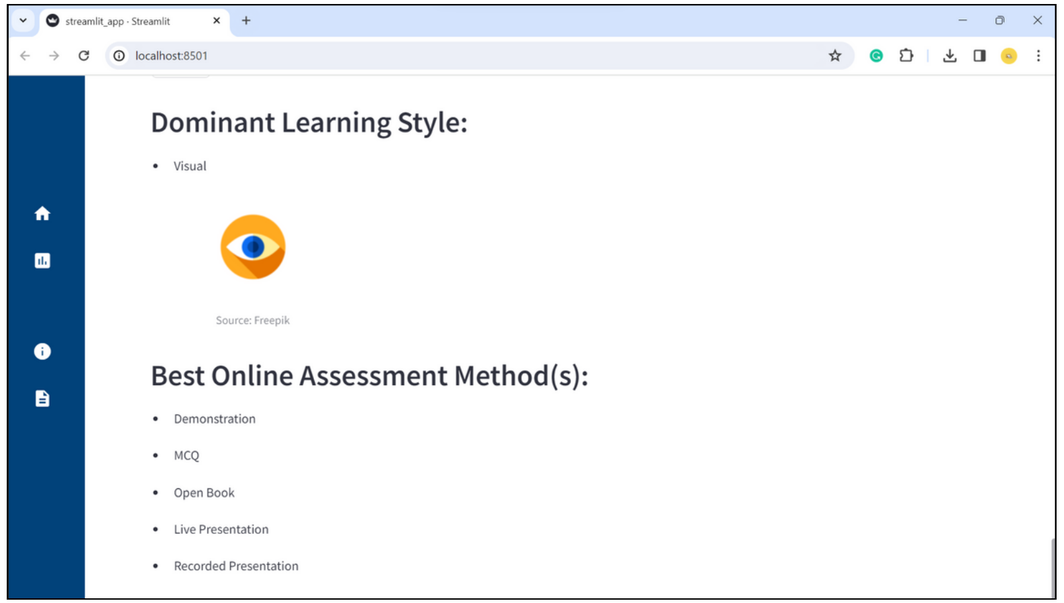


Figure 42: Sample output of the data product

Streamlit

DS METHODOLOGY - SUMMARY

1 - Business Understanding

- Problem understanding
- Objectives definition

2 - Data Understanding

- Data collection
- Data analysis

3 - Data Preparation

- Feature selection
- Data standardisation
- Exploratory data analysis
- Data encoding

4 - Modelling

- Decision Tree
- XGBoost
- KNN
- Random Forest
- SVM

5 - Evaluation

- Accuracy score
- Precision
- Recall
- F1-score

6 - Deployment

- Web application on Streamlit

TOOLS USED

- SAS Enterprise Miner
- Talend Data Preparation
- Tableau
- Python
- Streamlit

05.
APP DEMONSTRATION

Objectives	Output	Achievement
To **develop** an online assessment **classification model** based on personalisation	• Decision Tree • XGBoost • KNN • Random Forest • SVM	**ACHIEVED**
To **evaluate** the online assessment classification **model** based on personalisation	• Accuracy, precision, recall & f1-score • The best-performing model: ◦ Random Forest	**ACHIEVED**
To **develop a data product** that functions as an online assessment methods recommendation	• Web application	**ACHIEVED**

Table 5 Objectives achievement summary

THANK YOU

REFERENCES

Agarwal, A., Sharma, P., Alshehri, M., Mohamed, A. A., & Alfarraj, O. (2021). Classification model for accuracy and intrusion detection using machine learning approach. PeerJ, 7, e437. https://doi.org/10.7717/peerj-cs.437

Dhawan, S. (2020). Online Learning: A Panacea in the Time of COVID-19 Crisis. Journal of Educational Technology Systems, 49(1), 5-22. https://doi.org/10.1177/0047239520934018

Hosny, K., & El-Korany, A. (2022). Applying adaptive learning by integrating semantic and machine learning in proposing student assessment model. International Journal of Power Electronics and Drive Systems, 12(2), 2014. https://doi.org/10.11591/ijece.v12i2.pp2014-2025

Maya, J., Luesia, J. F., & Pérez-Padilla, J. (2021). The Relationship between Learning Styles and Academic Performance: Consistency among Multiple Assessment Methods in Psychology and Education Students. Sustainability, 13(6), 3341. https://doi.org/10.3390/su13063341

Santana, Í. V. D. S., Da Silveira, A. C., Sobrinho, Á., Silva, L. C. E., Da Silva, L. D., Santos, D. F. S., Gurjão, E. C., & Perkusich, Â. (2021). Classification models for COVID-19 test prioritization in Brazil: Machine learning approach. Journal of Medical Internet Research, 23(4), e27293. https://doi.org/10.2196/27293

Zhang, W., Liu, H., Wu, W., Zhan, L., & Wei, J. (2020). Mapping Rice Paddy Based on Machine Learning with Sentinel-2 Multi-Temporal Data: Model Comparison and Transferability. Remote Sensing, 12(10), 1620. https://doi.org/10.3390/rs12101620