

STUDENT SPENDING ANALYSIS

FCEE: Team 10

Socratees Harshia (U2320903C)

Sundar Subramanian Ayngara Sudha (U2321482C)

MOTIVATION:

WHY STUDENT SPENDING ANALYSIS?

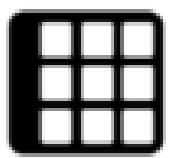
AS STUDENTS ARE BUSY WITH ACADEMICS AND OTHER HOBBIES THAT CAN OCCUPY THEIR TIME, THIS DIVERTS THEIR FOCUS FROM MANAGING FINANCES.

Student Spending habits dataset- Kaggle

PROJECT OBJECTIVE:

Our goal is to analyze and predict student spending behaviors based on various demographic factors such as Year in School(University), Major and Preferred Payment Methods.





DATA PREPARATION

```
: habits.dtypes
```

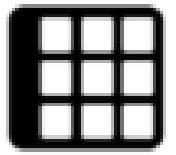
```
: Unnamed: 0          int64
age            int64
gender         object
year_in_school object
major           object
monthly_income int64
financial_aid  int64
tuition        int64
housing         int64
food            int64
transportation int64
books_supplies int64
entertainment  int64
personal_care   int64
technology     int64
health_wellness int64
miscellaneous   int64
preferred_payment_method object
dtype: object
```

- Too many columns to consider separately as many are important factors in analysing students spending
- Possible to combine similar columns to better organise Exploratory Analysis

DATA PREPARATION

Problem

FORMULATION



DATA PREPARATION

Personal_expenses

Personal_care &
Health_wellness

Education_expenses

Tuition &
book_supplies

Discretionary_expenses

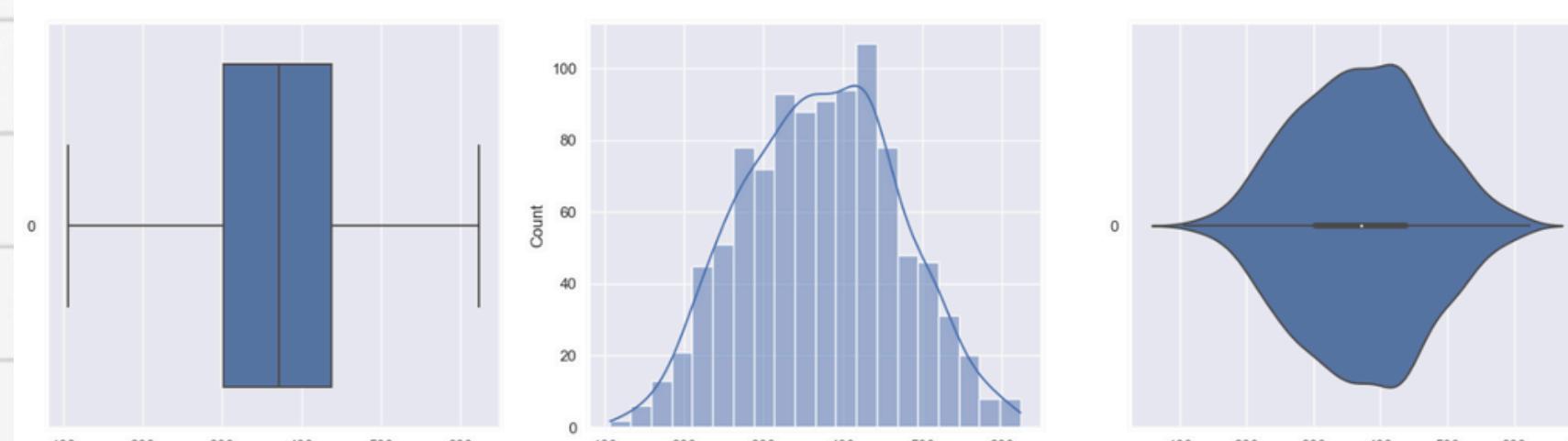
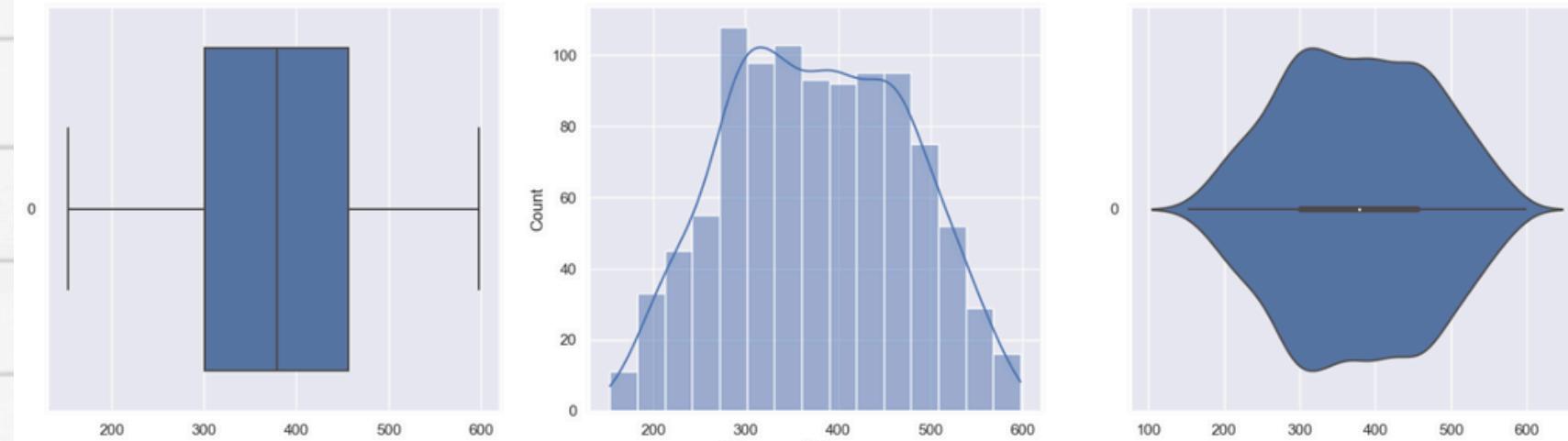
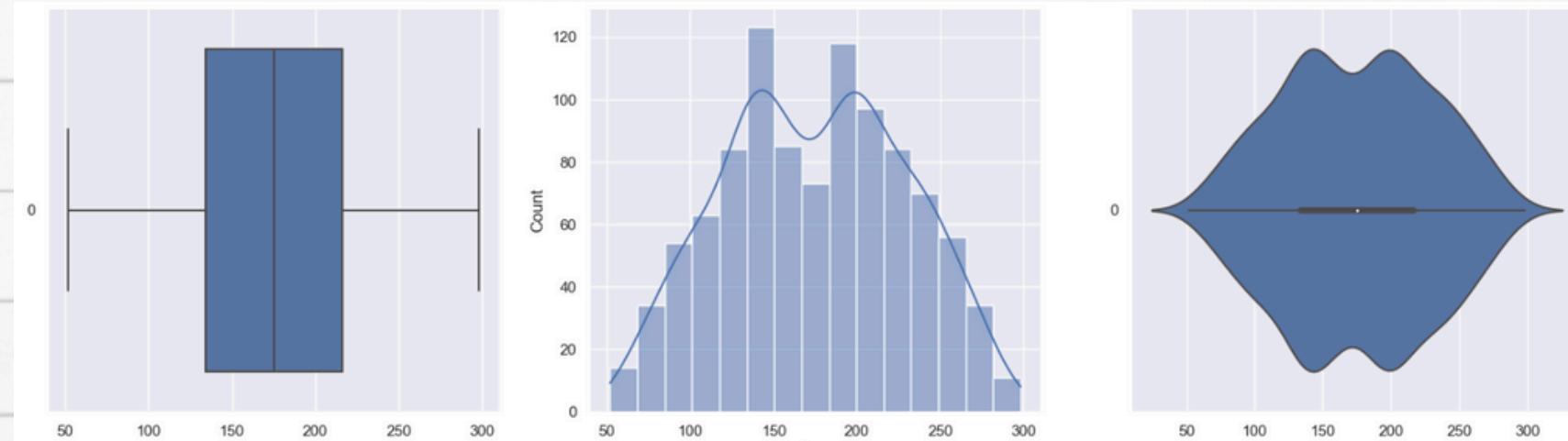
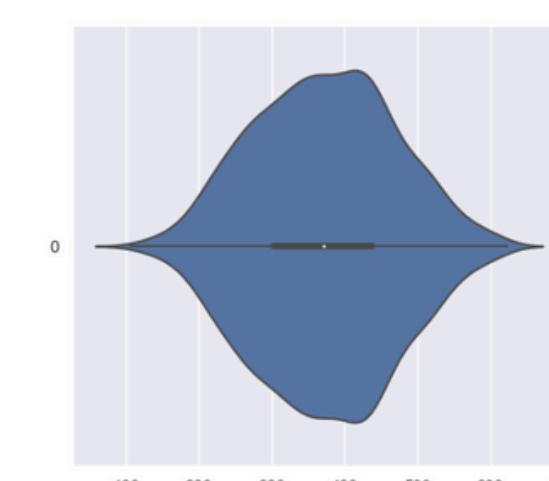
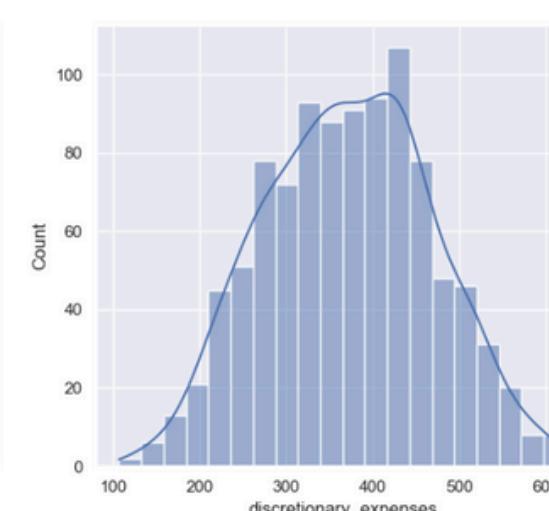
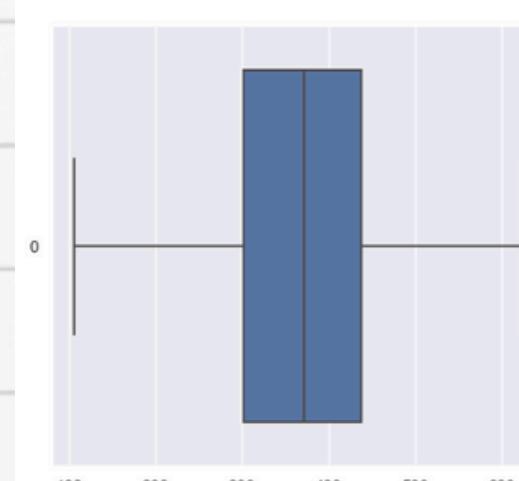
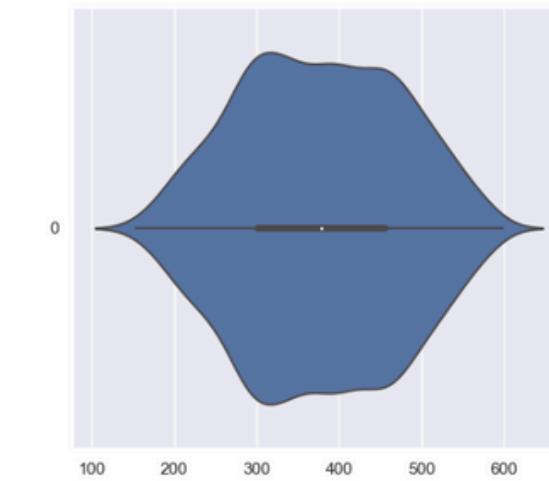
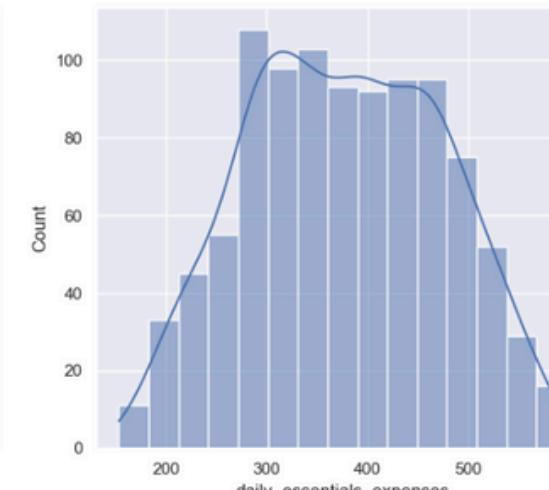
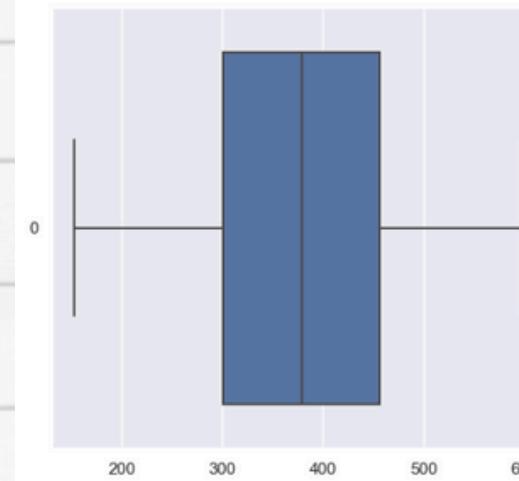
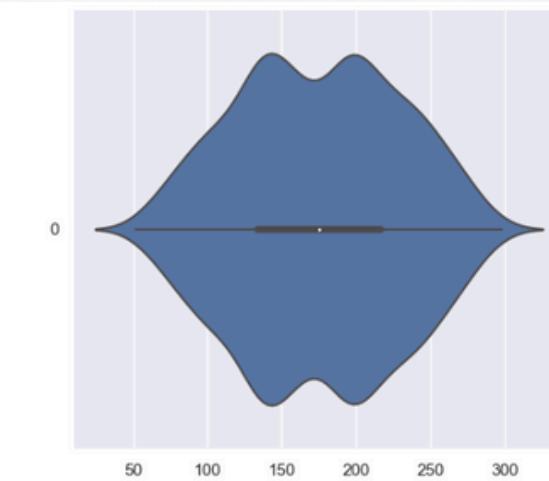
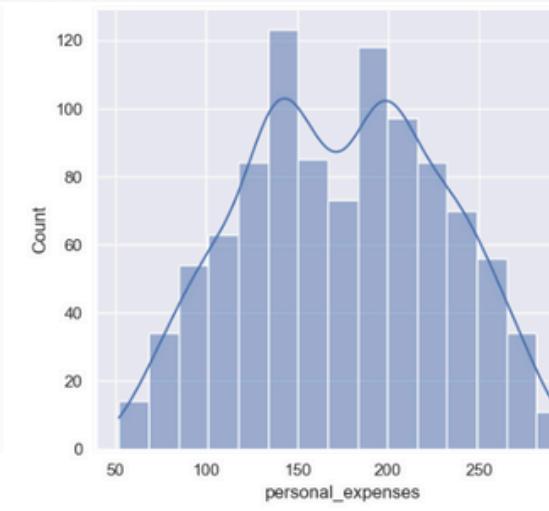
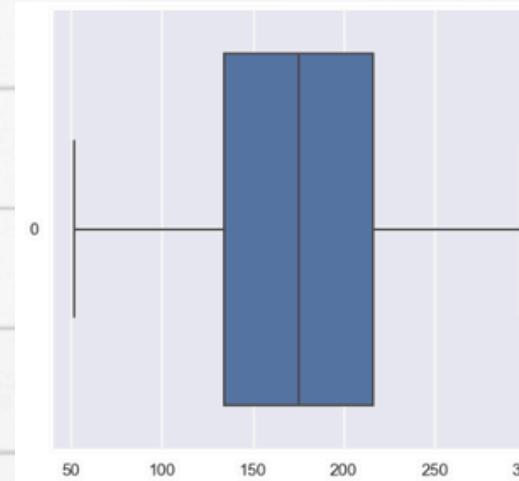
Entertainment &
Miscellaneous &
Technology

Daily_Essential_Expenses

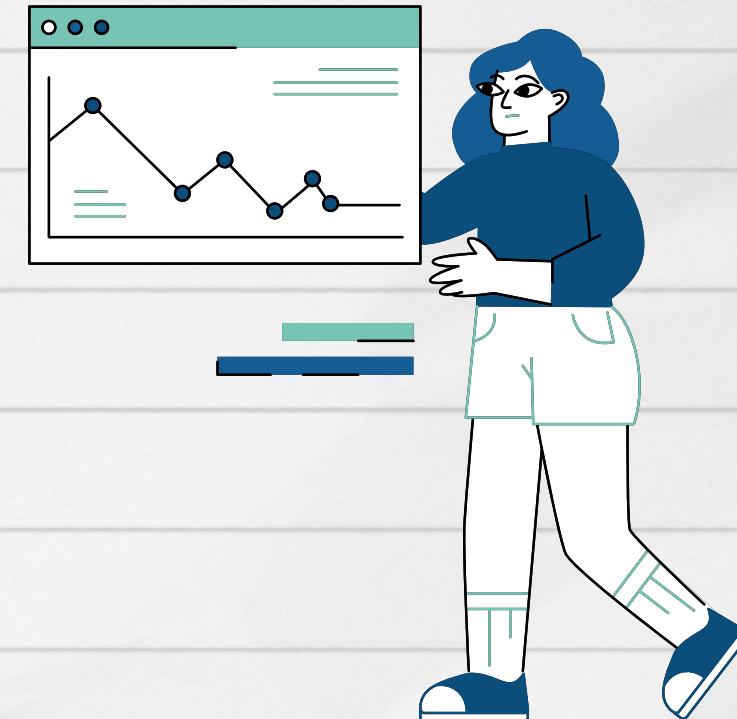
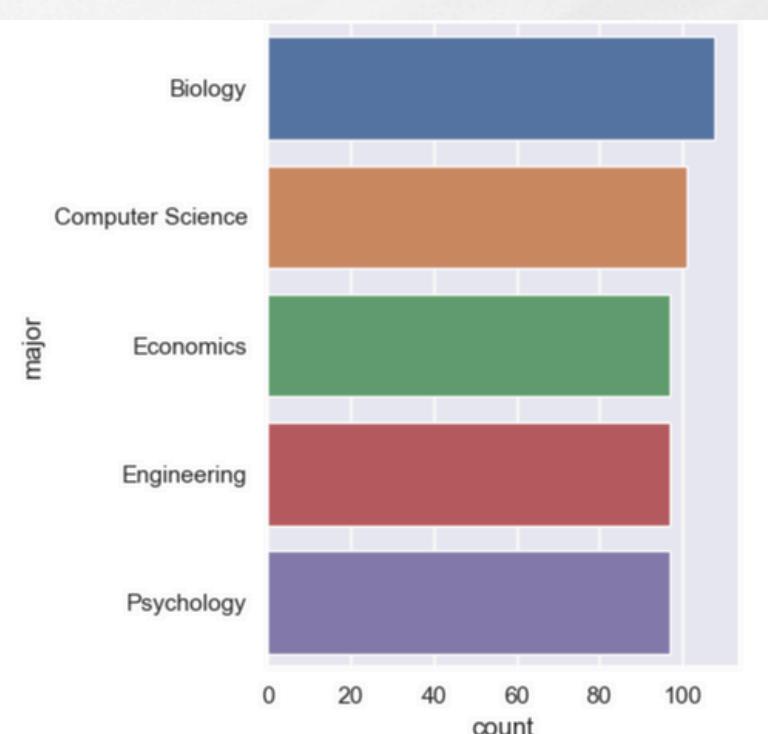
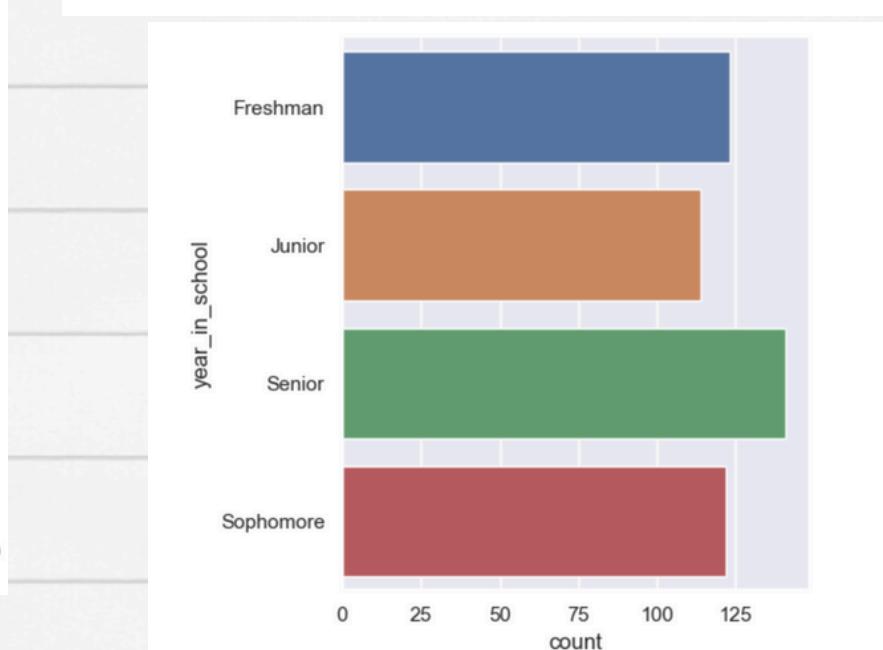
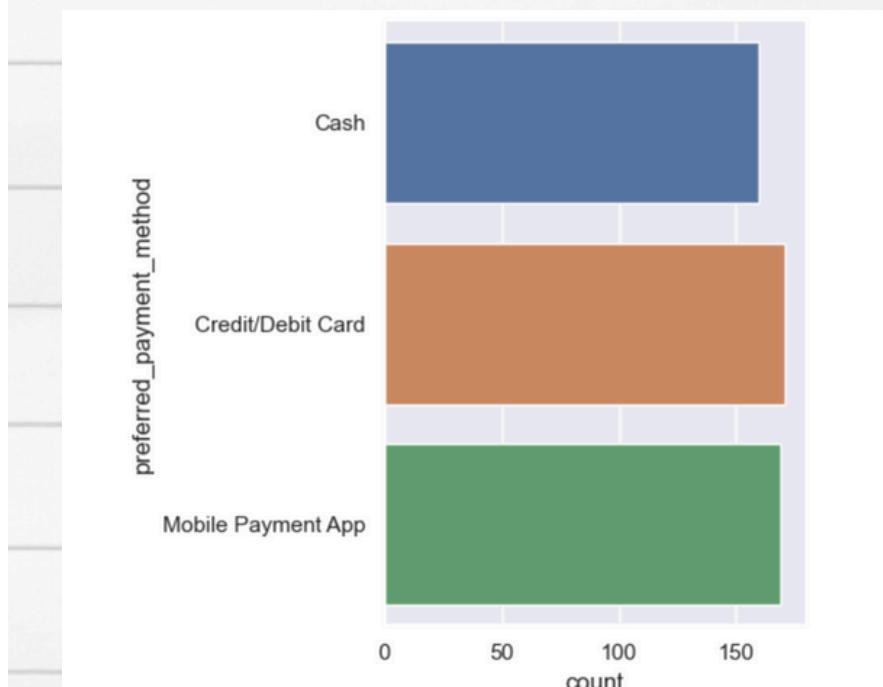
Food &
Transportation

Exploratory ANALYSIS

Statistical DESCRIPTION

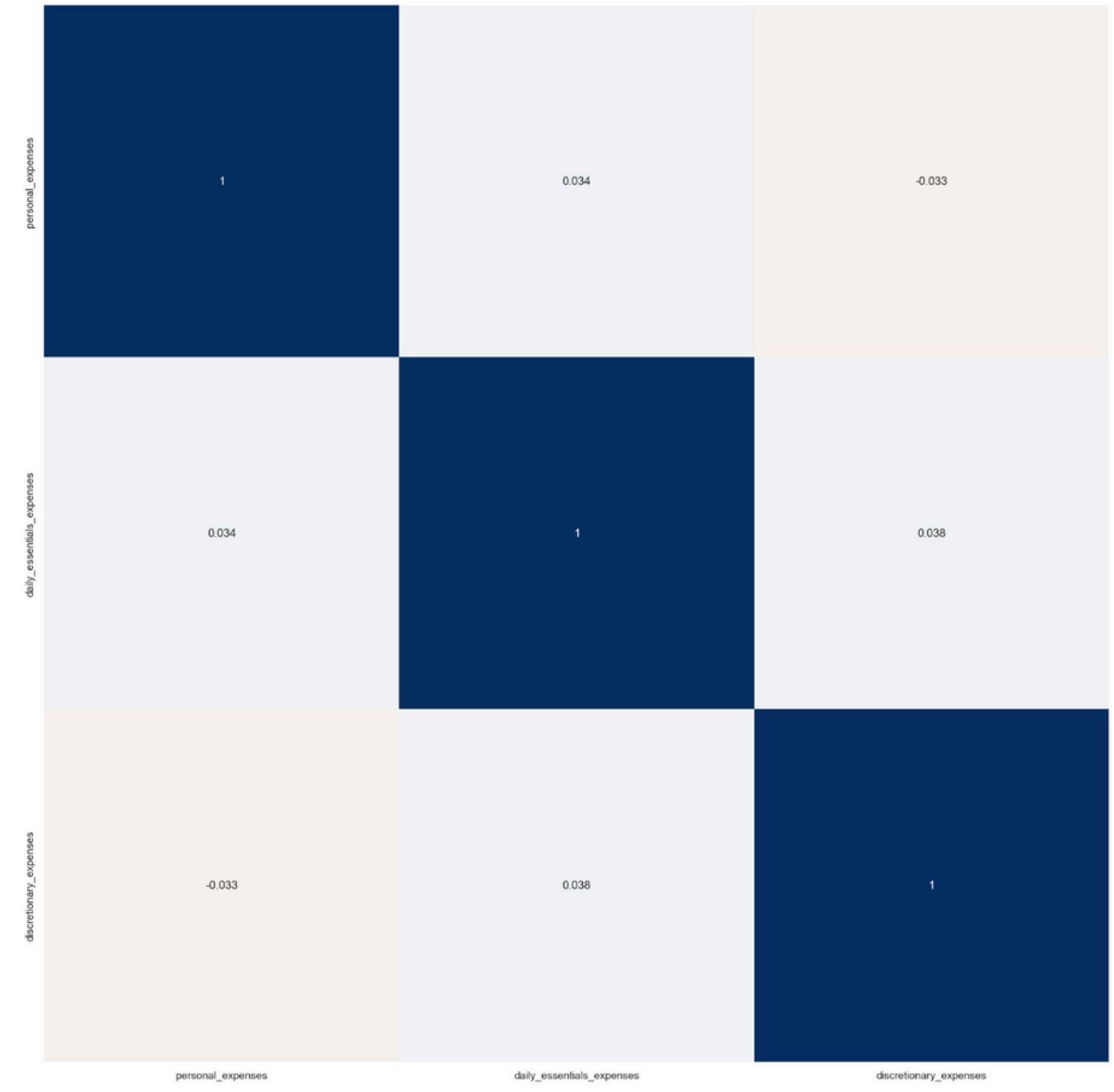


CATEGORICAL VARIABLES:



Exploratory ANALYSIS

Statistical DESCRIPTION



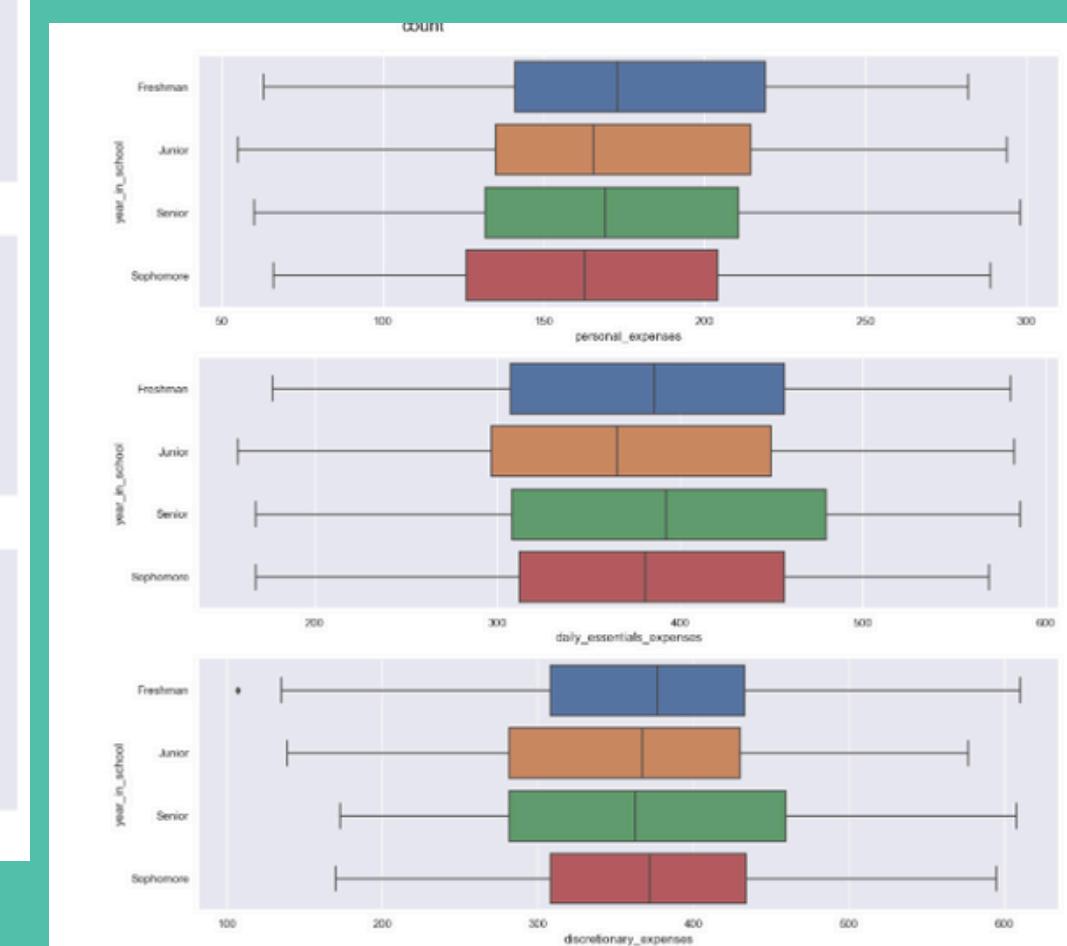
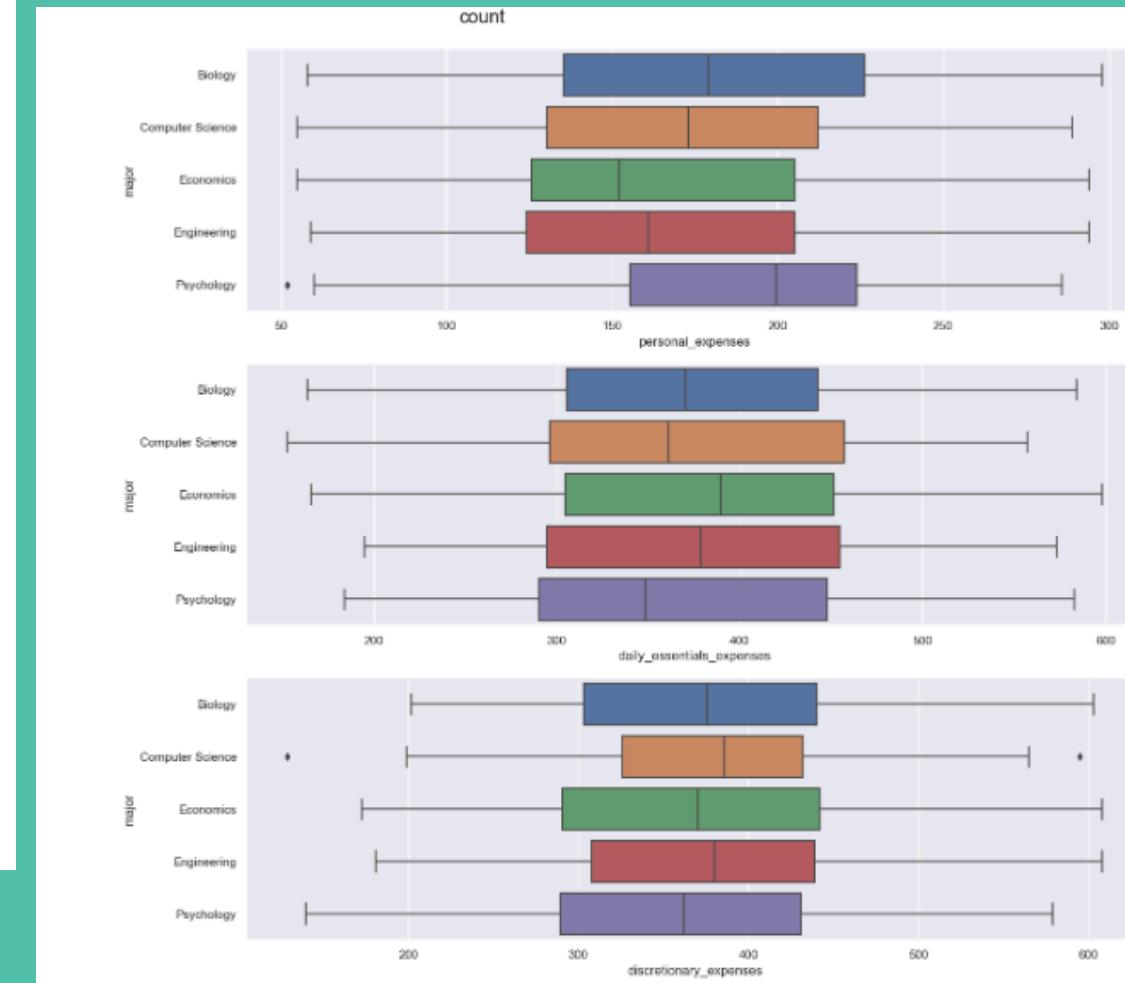
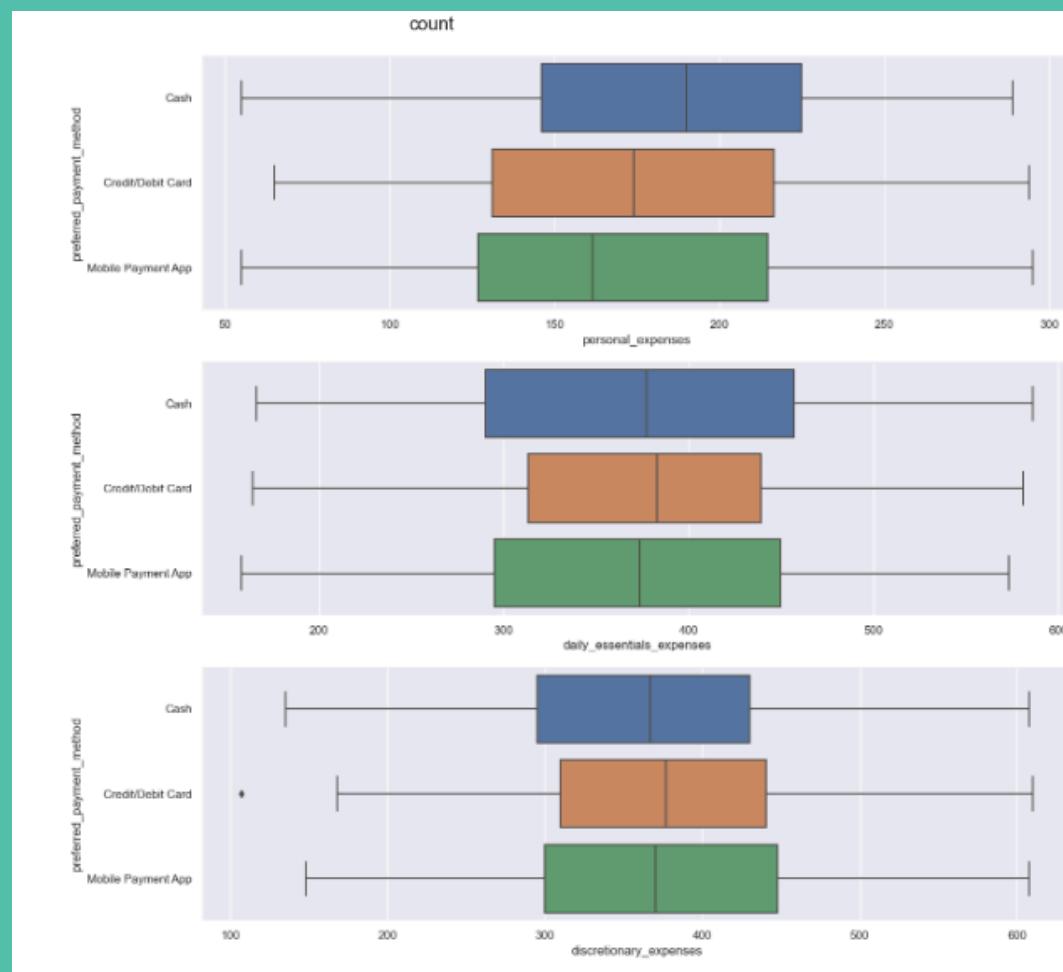
CORRELATION MATRIX:

- There is a moderate positive correlation between personal expenses and daily essentials expenses.
- The correlation between personal expenses and Discretionary Expenses is close to zero, indicating a very weak, almost negligible negative relationship.
- There is a moderate positive correlation between daily essentials expenses and discretionary expenses.



MULTI-CLASS CLASSIFICATION

The test datasets seem to be less accurate compared to the train data which showed us that the model is not generalizing as well to new data, as it indicates overfitting or memorisation of the training data.



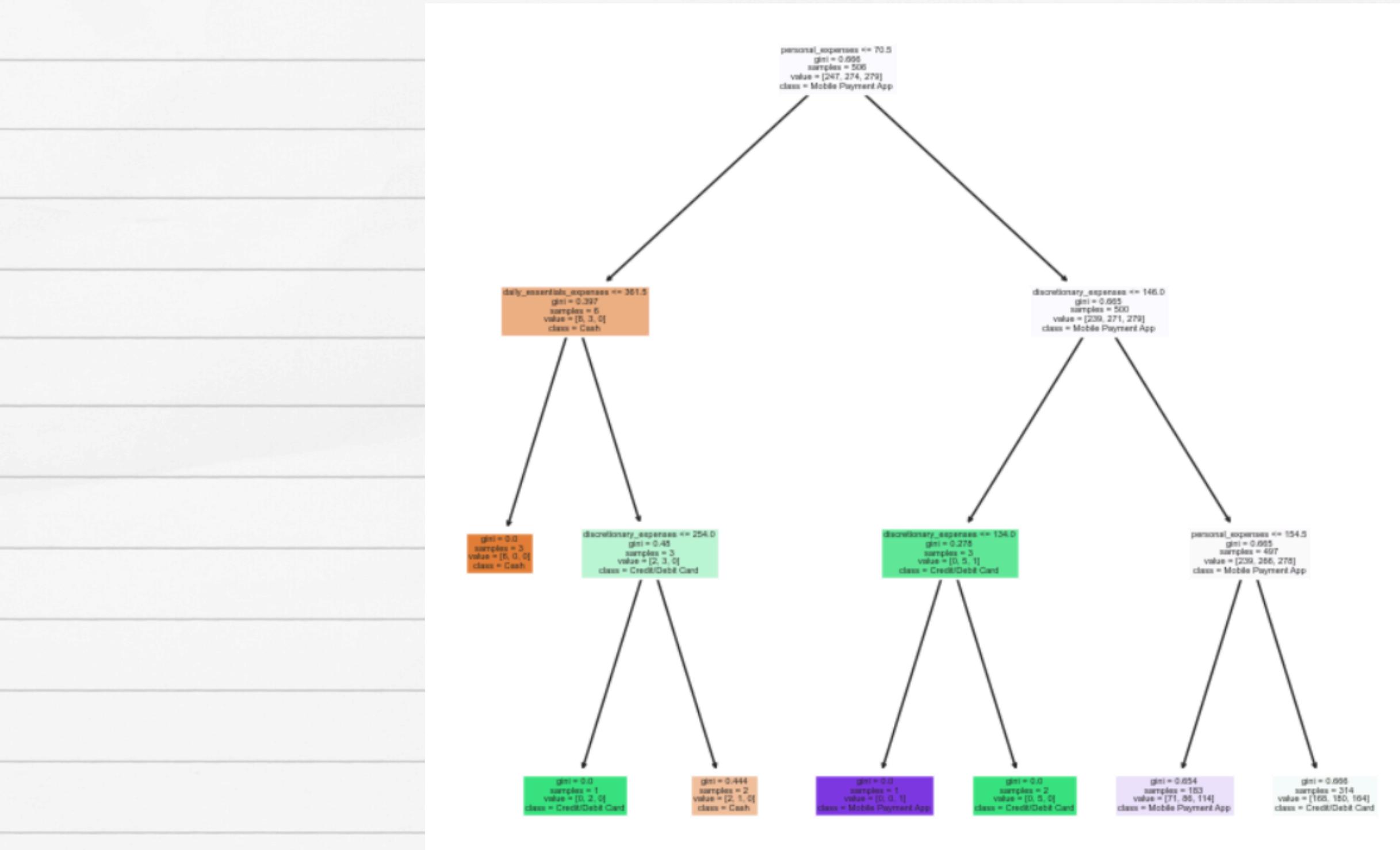


RANDOM FOREST

Random Forest seems better than a single multi-class classification tree primarily due to its ability to mitigate over-fitting. By aggregating predictions from multiple decision trees trained on random subsets of the data and features, Random Forest reduces the risk of overfitting and improves generalization. Resulting in higher accuracy and robustness across various classification tasks.



RANDOM FOREST



Information PRESENTATION



Statistical INFERENCE

Accuracy: 0.315

Classification Report:

	precision	recall	f1-score	support
Cash	0.25	0.29	0.27	59
Credit/Debit Card	0.37	0.34	0.35	74
Mobile Payment App	0.33	0.31	0.32	67
accuracy			0.32	200
macro avg	0.32	0.31	0.31	200
weighted avg	0.32	0.32	0.32	200

Accuracy: 0.245

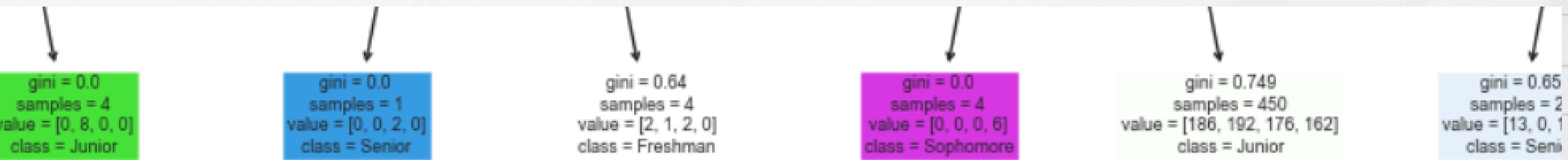
Classification Report:

	precision	recall	f1-score	support
Freshman	0.23	0.30	0.26	46
Junior	0.25	0.17	0.20	59
Senior	0.26	0.28	0.27	50
Sophomore	0.24	0.24	0.24	45
accuracy			0.24	200
macro avg	0.25	0.25	0.24	200
weighted avg	0.25	0.24	0.24	200

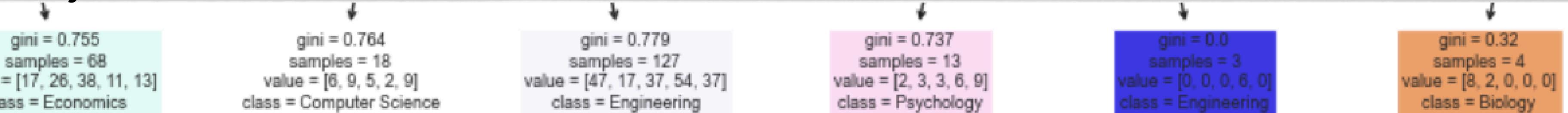
Preferred payment method



Year in school



Major





CONCLUSION

- Random Forest is a better model than a Multi-class classification tree as it mitigates over-fitting and provides more accuracy.
- Recommendations:
 - > Hyperparameter tuning
 - > Cross validation

**THANK
YOU VERY
MUCH!**