# CIND 820 – Capstone Project Abstract: In-Vehicle Coupon Recommendation

By: Sucheta Sikdar (ID: 501074722)

Date of Submission: September 15, 2024

Resubmitted after feedback: September 30, 2024

Supervisor: Dr. Ceni Babaoglu

Toronto Metropolitan University

# Abstract

Coupons are a great way for customers to save money on their purchases. It is also a great way for business to attract customers. Coupons create a win-win situation for both companies and customers and hence by offering a correct coupon to users can lead to users to become frequent customers (Niralidedaniya, 2023). If businesses can find the right customers who will use their coupons, then it will help businesses. It will also be interesting to predict what type of coupon will be accepted by a customer based on various attributes about the customer. Luckily, there exists a publicly available dataset called In-Vehicle Coupon Recommendation at UCI Machine Learning Repository from 2020 which describes different driving scenarios of multiple clients such as destination, time, coupon, expiration, gender, age, marital status, whether they have children, education, occupation, income, car, the number of times they go to the bar per month, the number of times they go a coffee shop per month, the number of times that they get take away food per month, if customer’s average expense per person at restaurants is less than 20 dollars a month, if customer’s average expense per person at a restaurant is between 20 dollars to 50 dollars per month, driving distance to the restaurant/bar for using the coupon is greater than 15 minutes, driving distance to the restaurant/bar for using the coupon is greater than 25 minutes, whether the restaurant/bar is in the same direction as destination, whether the restaurant/bar is in the opposite direction as destination, whether the coupon is accepted (*UCI Machine Learning Repository*, 2020).

When printing the info of the dataset in VSCode using Python, the below was observed:

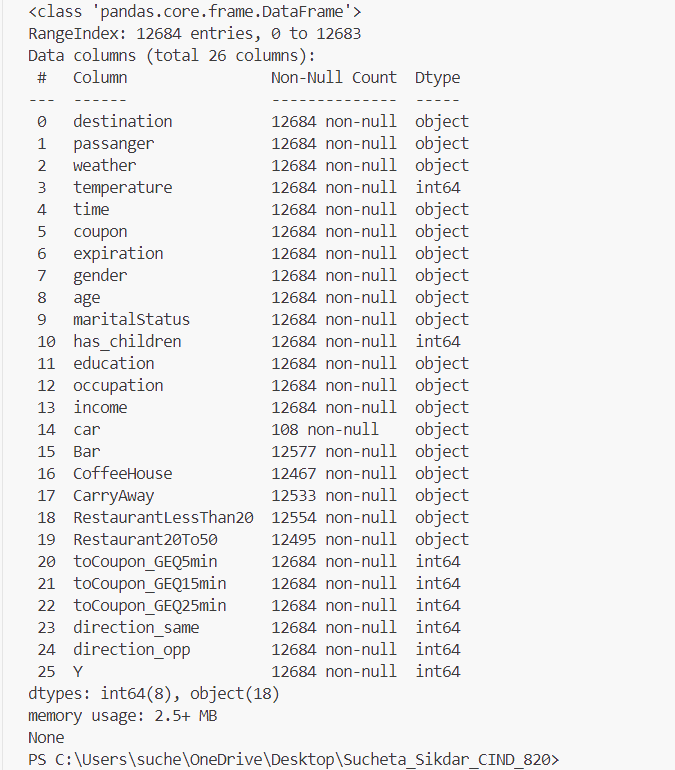


Figure 1: Info on the In-Vehicle Coupon Recommendation Dataset

As seen in Figure 1, the dataset has 12684 records. The dataset has 25 features and 1 target column. Some of the data like marital status and gender are categorical and some of the data like age and temperature are numerical. Using code provided by Niralidedaniya (2023), it was found that the target classes are partially balanced. If the target classes were highly unbalanced, then this dataset could not be used because the results of supervised learning algorithms used to make predictions would skew towards the class with the class with higher percentage of records.

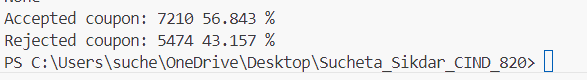


Figure 2: Distribution of Target Classes

It should also be noted that this dataset comes with many missing values. Hence, this dataset requires preprocessing before it can be analyzed with machine learning algorithms. Using the code provided by Niralidedaniya (2023), we can see the features which have missing values.

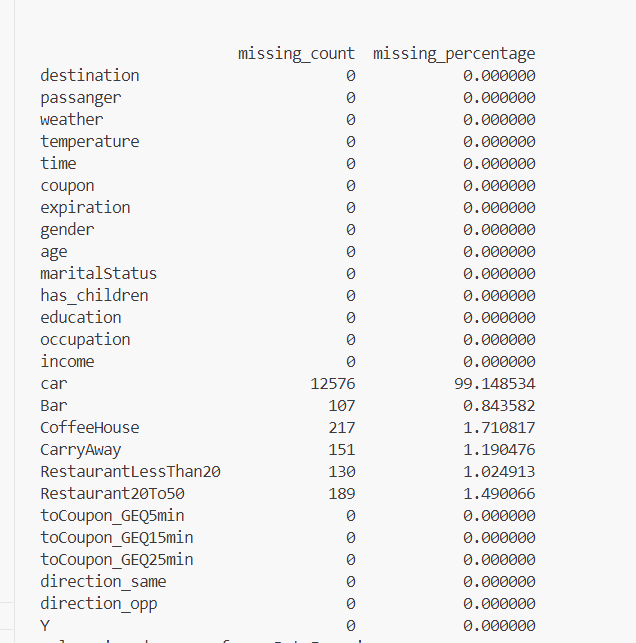


Figure 3: Distribution of missing values in the dataset

The objective of this project will be to:

* Find the best predictive classification algorithm for the In-Vehicle Coupon Recommendation dataset (2020) after evaluation of various supervised learning classification algorithms introduced to us in *CMTH 642 – Data Analytics: Advanced Methods* like Random Forest, Decision Tree, Logistic Regression, Naïve Bayes, k-Nearest Neighbours (k-NN) on the dataset.
* Using a correlation matrix find out which attributes are highly correlated to the target of the customer accepting or rejecting a coupon.
* Find whether we can attain a dataset with fewer dimensions using these 3 methods: Stepwise Regression, Forward Feature Selection and Backward Feature Elimination methods learnt by us in *CMTH 642 – Data Analytics: Advanced Methods*.
* Find the limitations of this dataset.

Data analysis for this project will be done using Python. Pandas and numpy libraries will be used. Seaborn and Matplotlib libraries will be used for visualizations. Supervised learning techniques will require the scikit-learn library (*Supervised Learning*, n.d.). Evaluation of various classification algorithms will be done by comparing evaluation metrics like the Accuracy, Precision and Area under the Curve (AUC) of each algorithm.

# References

Niralidedaniya. (2023, January 19). In Vehicle Coupon Recommendation — a Machine learning classification case study. *Medium*. https://medium.com/@niralidedaniya/in-vehicle-coupon-recommendation-a-machine-learning-classification-case-study-df67e7835703

UCI Machine Learning Repository. (2020). https://archive.ics.uci.edu/dataset/603/in+vehicle+coupon+recommendation

Babaoglu (n.d.). CMTH 642 – Data Analytics: Advanced Methods course [PowerPoint slide printouts]

*Supervised learning*. (n.d.). Scikit-learn. <https://scikit-learn.org/stable/supervised_learning.html>