# Lab Data Preparation

## In-Vehicle Coupon Recommendation



*Préparé par :*

*Mohamed Wahchi*

*Malek Hmem*

*Kalthoum Dridi*

*Selim Sahli*

*Nour Hassine*

*Khouloud Zograni*

Introduction

Les coupons, sous forme de réductions, de bons d'achat ou d'offres spéciales, sont un outil puissant et omniprésent dans le monde du commerce et du marketing. Ils ont une influence significative sur les consommateurs, les entreprises et le marché en général.

Ils favorisent la fidélisation des clients, en influençant la concurrence et les prix, en permettant la collecte de données sur les habitudes d'achat des consommateurs, en impactant le comportement des consommateurs et en permettant le marketing ciblé.

Parmi les différentes approches, les "In-Vehicle Coupons" se distinguent par leur pertinence dans le domaine de la mobilité. En effet, ces coupons sont spécifiquement conçus pour être distribués à l'intérieur des véhicules, que ce soit par le biais de systèmes de divertissementembarqués, d'applications mobiles liées au véhicule ou d'autres moyens de communication à bord tels que les panneaux publicitaires.

1. Business Problem :

**Comment les données recueillies à partir des coupons distribués à l'intérieur des véhicules peuvent-elles être exploitées pour optimiser l'efficacité des campagnes marketing et d'incitation à l'achat, tout en améliorant l'expérience des utilisateurs et en maximisant les résultats commerciaux ?**

Cette problématique pose des questions importantes sur l'utilisation des données provenant des "In-Vehicle Coupons" pour comprendre le comportement des consommateurs, personnaliser les offres, cibler les bons publics et mesurer l'impact de ces coupons sur la croissance de l'entreprise. Elle soulève également des préoccupations quant à la manière dont les entreprises peuvent équilibrer le respect de la vie privée des utilisateurs tout en maximisant les avantages de ces coupons pour leurs activités.

Prédire si un client acceptera ou non un coupon représente un défi complexe, et il n'est pas judicieux de le recommander à tous en raison des coûts associés. Dans cette problématique, notre objectif est de prédire si un client acceptera ou refusera le coupon qui lui est offert, en se basant sur son profil et son historique d'achats.

1. Business Understanding :

Le "business understanding" (compréhension de l'entreprise) est une phase essentielle dans le processus de gestion des données et de l'analyse de données, qui vise à bien comprendre les besoins et les objectifs d'une entreprise avant de collecter, traiter et analyser des données

Ces données ont été recueillies à la suite d'une enquête visant à obtenir des informations approfondies sur le comportement et les préférences des individus en ce qui concerne l'utilisation de coupons dans divers contextes de conduite. Cette enquête revêt une importance significative pour plusieurs raisons :

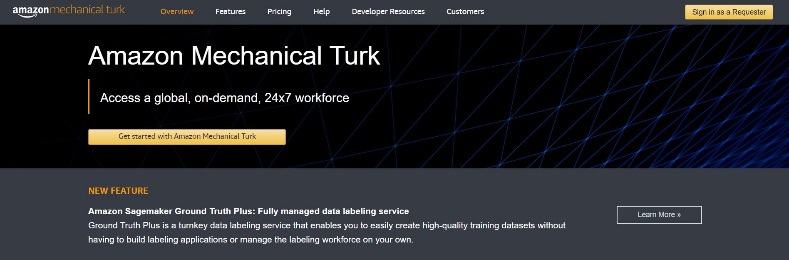
Tout d'abord, elle permet de mieux comprendre comment des facteurs tels que la durée du trajet, les choix de restaurants et l'emplacement géographique influencent la décision d'accepter ou de refuser les coupons. Ces informations sont cruciales pour les entreprises, notamment les restaurants et les bars, car elles leur permettent de cibler plus efficacement leurs efforts promotionnels, d'améliorer l'acquisition de nouveaux clients et de renforcer la fidélisation de la clientèle.

De plus, ces données peuvent être précieuses pour les entreprises cherchant à créer des segments de clientèle en fonction des comportements et des préférences. En segmentant la clientèle de manière plus précise, les entreprises peuvent personnaliser davantage leurs offres pour chaque groupe, ce qui améliore la pertinence des coupons proposés.

Enfin, cette enquête vise à identifier les types de coupons et d'offres les plus susceptibles d'être acceptés en fonction des préférences des clients. Cette connaissance permet aux entreprises de concevoir des offres plus attirantes, augmentant ainsi les chances que les clients les acceptent.

Dans l'ensemble, l'objectif de cette prédiction est d'anticiper si un client acceptera ou refusera un coupon pour un lieu spécifique, en se basant sur des caractéristiques démographiques et contextuelles. Cette classification binaire (Y=1 pour acceptation et Y=0 pour refus) représente un enjeu clé pour les entreprises cherchant à optimiser leur stratégie marketing et à améliorer leur relation avec les clients.

1. Data Understanding :

Ces données ont été obtenues grâce à une enquête publiée en 2017 sur Amazon Mechanical Turk et rendue accessible sur le site de l'UCI.

L'enquête recueille des informations concernant divers scénarios de conduite, notamment la destination de l'utilisateur, l'heure actuelle, les conditions météorologiques, la présence de passagers, les caractéristiques des coupons, les caractéristiques de l'utilisateur, ainsi que les caractéristiques contextuelles. **Ensuite, elle interroge l'utilisateur pour savoir s'il acceptera ou non le coupon.**

Le fichier de jeu de données se nomme "in-vehicle-coupon-recommendation.csv". Il comprend plus de 12 684 points de données et 26 attributs, la plupart étant de nature catégorielle, et quelques-uns sont de type catégoriel numérique.

En ci-dessous, se trouve l’explication de la dataset et ses attributs. En commençant, nous allons explorer les variables d’entrées (les inputs).

1. **Les données sur l’utilisateur :**

|  |  |
| --- | --- |
| Nom de la colonne : | Description : |
| Gender | **Le sexe de l’utilisateur :**  Male / Femelle (Donnée de type string) |
| Age | **L’age moyen de l’utilisateur :**  21/46/26/31/41/50plus/36/below21 (Donnée de type string) |
| Marital Status | **L’état civil de l’utilisateur:**   Unmarried partner, Single, Married partner, Divorced, Widowed (Donnée de type string) |
| Has\_Children | **Variable indiquant si l’utilisateur a des enfants :**  1:has children/0:no children (donnée de type int) |
| Education | **Le niveau d’éducation de l’utilisateur:**  Some college — no degree, Bachelors degree, Associates degree, High School Graduate, Graduate degree (Masters or Doctorate), Some High School (Donnée de type string) |
| Occupation | **L’occupation que l’utilisateur occupe:**  Unemployed, Architecture & Engineering, Student,Education&Training&Library, Healthcare Support,Healthcare Practitioners & Technical, Sales & Related, Management,Arts Design Entertainment Sports & Media, Computer & Mathematical,Life Physical Social Science, Personal Care & Service, Community & Social Services, Office & Administrative Support, Construction & Extraction, Legal, Retired, Installation Maintenance & Repair, Transportation & Material Moving, Business & Financial, Protective Service, Food Preparation & Serving Related, Production Occupations, Building & Grounds Cleaning & Maintenance, Farming Fishing & Forestry (Donnée de type string) |
| Income | **Le salaire annuel de l’utilisateur :**  Less than $12500,$12500 — $24999,$25000 — $37499,$37500 — $49999,$50000 — $62499,$62500 — $74999,$75000 — $87499,$87500 — $99999,$100000 or More (Donnée de type string) |
| Car | **Le véhicule que l’utilisateur conduit :**  Scooter and motorcycle, crossover, Mazda5 (Donnée de type string) |
| Bar | **Combien de fois l’utilisateur fréquente le bar chaque mois :** never, less1, 1~3, 4~8, gt8, nan (Donnée de type string) |
| CoffeeHouse | **Combien de fois l’utilisateur fréquente un salon de thé chaque mois :**  never, less1, 1~3, 4~8, gt8, nan (Donnée de type string) |
| CarryAway | **Combien de fois l’utilisateur prend sa commande à emporter chaque mois :**  never, less1, 1~3, 4~8, gt8, nan (Donnée de type string) |
| RestaurantLessThan20 | **Combien de fois l’utilisateur fréquente un restaurant avec un montant inférieur à 20$ chaque mois :**  never, less1, 1~3, 4~8, gt8, nan (Donnée de type string) |
| Restaurant20To50 | **Combien de fois l’utilisateur fréquente un restaurant avec un montant moyen de 20$ à 50$ chaque mois :**  never, less1, 1~3, 4~8, gt8, nan (Donnée de type string) |

1. **Les données contextuels :**

|  |  |  |
| --- | --- | --- |
| Nom de la colonne : | Description : | |
| Destination | | **La destination de l’utilisateur :**  No Urgent Place, Home, Work (donnée de type string) |
| Passenger | | **Les passagers qui se trouvent avec l’utilisateur dans la voiture:**  Alone, Friend(s), Kid(s), Partner (donnée de type string) |
| Weather | | **L’état météorologique lorsque l’utilisateur est en train de conduire :**  Sunny, Rainy, Snowy (donnée de type string) |
| Temperature | | **La temperature en Fahrenheit lorsque l’utilisateur est en train de conduire :**  55,80,30 (donnée de type int) |
| Time | | **L’heure lorsque l’utilisateur est en train de conduire :**  2PM, 10AM, 6PM, 7AM, 10PM (donnée de type string) |
| toCoupon\_GEQ5min | | **Variable qui indique si la distance en voiture jusqu'au restaurant/bar pour utiliser le coupon est supérieure à 5 minutes :** 0,1 (donnée de type int) |
| toCoupon\_GEQ15min | | **Variable qui indique si la distance en voiture jusqu'au restaurant/bar pour utiliser le coupon est supérieure à 15 minutes :** 0,1 (donnée de type int) |
| toCoupon\_GEQ25min | | **Variable qui indique si la distance en voiture jusqu'au restaurant/bar pour utiliser le coupon est supérieure à 25 minutes :** 0,1 (donnée de type int) |
| Direction\_same | | **Variable qui indique si le restaurant/bar se trouve dans la même direction que la destination actuelle de l'individu :**  0,1 (donnée de type int) |
| Direction\_opp | | **Variable qui indique si le restaurant/bar se trouve dans la même direction que la destination opposée de l'individu :**  0,1 (donnée de type int) |

1. **Les données sur le coupon :**

|  |  |
| --- | --- |
| Nom de la colonne : | Description : |
| Coupon | **Type du coupon offert par l’entreprise :**  Restaurant(<$20), Coffee House, Carry out & Take away, Bar, Restaurant$20-$50 (donnée de type string) |
| Expiration | **Délais d’expiration du coupon:**  1d,2h (donnée de type string) |

1. **L’attribut cible (output) :**

Par la suite, nous avons la variable de sortie (output) qui est la colonne Y qui est une donnée de type où Y indique si le coupon est accepté ou refusé, **1 : accepté, 0 : refusé.**