

Data Against Food Waste

ENSAE x Too Good To Go | Décembre 2020

1. L'entreprise

A. La mission

Too Good To Go est une application mobile lancée en 2016 afin de **réduire le gaspillage alimentaire**. Elle permet de connecter des commerçants souhaitant écouler leur invendus du jour et des consommateurs à la recherche de repas accessibles.

B. En chiffres

Aujourd'hui Too Good To Go en France c'est:

- 7 millions d'utilisateurs partout sur le territoire
- 18 millions de repas sauvés de la poubelle
- 24 000 tonnes de CO2 économisées.
- 15 000 commerçants partenaires

C. Complémentarité avec le don

Too Good To Go est une solution complémentaire au don, en aucun cas elle s'y substitue. Il s'agit de sauver des produits que les associations ne récupèrent pas:

- Produits interdits au don
 - Viande hachée, volailles, sandwiches et snacking, boucherie, poissonnerie, traiteur, boulangerie, viennoiserie, pâtisserie.
- Produits à date limite de consommation trop proche (J0)
 - Ils ne peuvent être redistribués à temps
- Produits difficiles à traiter
 - Certains produits bruts tels que les fruit et légumes sont souvent abîmés lors du transport
- Produits ne correspondant pas aux besoins des bénéficiaires des associations.
 - Pas halal ou casher, pas assez nourrissants.

2. Le service

A. Avantages pour les commerçants

Des supermarchés aux boulangeries de quartier, des restaurants aux hôtels, tous plébiscitent Too Good To Go pour sa simplicité et son efficacité. Les avantages pour les commerçants sont réels:

- **Un geste pour la planète**
10 millions de repas sains ont été sauvés de la poubelle
- **Un revenu supplémentaire**
95% des paniers sont vendus en moins de 5 heures
- **Un nouveaux flux de clients**
70% des utilisateurs reviennent en clients traditionnels
- **Une image engagée**
Too Good To Go est une marque reconnue qui sert de label aux commerçants pour montrer leur engagement éco-responsable. Les clients y sont aujourd'hui sensibles.
- **Pas de référencement, ni d'étiquetage**
Les clients achètent des "paniers surprise" avec vos invendus du jour. Ils ne connaissent donc pas le contenu de ces paniers à l'avance: c'est le principe même de l'application.

B. Coût du service

Too Good To Go ne se rémunère que **sur la valeur créée**. Ainsi, le commerçant ne nous verse jamais d'argent : nous nous contentons de prendre 25% de commission sur chaque panier sauvé et de verser les 75% restants au commerçant.

Par exemple, sur un panier vendu à 4€, nous récupérons 1€ et nous versons 3€ au commerçant. C'est 3€ de chiffre d'affaires sur des produits qui allaient être jetés et perdus.

C. Modalités d'inscriptions

L'inscription se fait en 5 minutes sur [le site de Too Good To Go](#).

C'est simple, facile et flexible car il n'y a aucun engagement (ni de volume, ni de durée) et aucun paiement préalable.

3. Les partenaires

A. Localisation

L'intégralité de la France métropolitaine.

Régions, villes, villages et hameaux de toutes tailles sont concernés.

B. Entreprises concernées

4 critères sont à prendre en compte pour qu'un prospect soit considéré comme qualifié:

- Le commerce est alimentaire
→ Il vend directement de la nourriture au consommateur final
- Le commerce est déjà ouvert depuis plus d'1 mois
→ Pas d'ouverture prochaine et autres projets non-finalisés
- Le commerce est situé sur un lieu fixe et permanent
→ Pas de food trucks et autres marchés
- Le volume d'inventus journaliers **est au moins de 20€ par jour ou 100€ par semaine**
→ En prix de vente TTC.

C. Segments adressés

Nous avons 4 segments prioritaires:

1. Boulangeries et pâtisseries

- Type de commerce: boulangeries et pâtisseries.
- Enseignes: Paul, Eric Kayser, La Mie Câline, La Croissanterie, Brioche Dorée, Lenôtre, Fauchon, Dalloyau, ...

2. Restauration rapide et snacking

- Type de commerces: stations-services, sushis, buffets à volonté asiatiques, sandwicheries, salades bar, plats à emporter.
- Enseignes: Pomme de Pain, Subway, Bagelstein, Starbucks, Sushi Shop, Planet Sushi, Columbus Café, Exki, Flunch, Jour, Cojean, Marks & Spencers, BP, Total, ...

3. Supermarchés

- Type de commerce: alimentation générale, épiceries, supérettes, supermarchés, hypermarchés.
- Enseignes: U Express, Utile, E. Leclerc, Auchan, Proxy, Coccinelle, Carrefour City, Carrefour Express, Carrefour Contact, À 2 pas, Spar, 8 à Huit, Casino Shop, Aldi, Cora, G20, Diagonal, ...

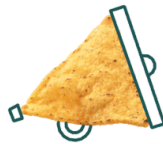
4. Hotels

- Type de commerce: budget, milieu de gamme, luxe, boutique & business.
- Enseignes: Ibis, Mercure, Pullman, Sofitel, Hyatt, Holiday Inn, Regent, Marriott, Partouche, Radisson, Intercontinental, Crowne Plaza, Novotel...

D. Segments exclus

Certains commerces appartiennent à nos segments prioritaires mais ne sont pas considérés comme qualifiés en raison de leur très faible quantité d'inventus journaliers:

- Kebabs
- Tacos
- Cavistes & domaines viticoles
- Pizzerias (Speed Rabbit, Dominos, Pizza Hut, ...)
- Fast-foods (Mcdonald's, Burger King, Quick, KFC, Speed Burger, ...)
- Coffee shops (cookies, cupcakes, brownies, muesli, ...)
- Restauration traditionnelle / à table / à la carte / en brasserie.



4. La problématique 🤔

A. Contexte

L'un de nos principaux enjeux est de maintenir nos commerçants actifs sur notre plateforme afin qu'ils sauvent un maximum de panier-repas de la poubelle. Un commerçant est considéré actif lorsqu'il propose des repas à la vente sur l'application.

B. Objectif

L'objectif est ici de construire un modèle qui permettrait de **déterminer la probabilité qu'un store devienne inactif**, c'est-à-dire qu'il arrête de mettre en vente des paniers sur notre plateforme.

Les différents facteurs de corrélation peuvent être cherchés dans le comportement des commerçants sur notre plateforme (last meal saved date, average rating, store cancellation, ...) mais aussi ailleurs (saisonnalité, emplacement, vacances, ...).

In fine, nous souhaiterions être capable, à partir des données dont nous disposons, de générer chaque jour ou chaque semaine une liste de magasins avec une forte probabilité de churn afin de mener des actions préventives.

Par “*churn*” nous entendons ici : un magasin qui avait l'habitude de mettre des paniers en ligne sur l'application (au moins 1 par semaine) et qui cesse de le faire pendant plus de 30 jours consécutifs. La limite de 30 jours a été fixée de façon empirique et peut tout à fait être remise en question, **l'enjeu réel étant de conserver les magasins actifs sur l'application.**

Pour information, les magasins suivent en général un rythme hebdomadaire assez marqué et vont avoir des comportements similaires d'une semaine sur l'autre. La limite des 30 jours permet de prendre une marge de 4 périodes et d'éviter de s'intéresser aux magasins qui seraient simplement fermés en période de vacances (été, Noël, Pâques)

C. Livrable attendu

Le format du livrable est libre.

Nous vous suggérons néanmoins de fournir :

- Un script clair et commenté (txt, ipynb, ...)

- Une documentation (PPT, Excel, Word ou équivalent) expliquant les choix de modélisation
- Un guide d'utilisation si le résultat s'y prête.
- Un document détaillant les principales conclusions (PPT, Excel, Word ou équivalent)



5. Les données

A. Sources

Vous êtes libres d'utiliser les données qui vous semblent utiles, quelque soit leur source.

- **Base de donnée interne Too Good To Go**

→ Voir glossaire, volumétrie et colonnes ci-dessous.

- **Bases de données publiques**

→ [OpenDataSoft](#), [API Gouv](#), [Pappers](#), [Google Data Set Search](#), [Kaggle](#), ...

- **Bases de données privées**

→ [Manageo](#), [SocieteInfo](#), [Verif](#), ...

B. Glossaire

- Consumer (utilisateur)

Un utilisateur de notre application mobile, achetant un ou plusieurs paniers avec son compte.

- Store (commerçant)

Un commerçant alimentaire ayant souscrit au service Too Good To Go.

- Meal (panier)

Un ensemble d'inventures que le commerçant regroupe pour les vendre à travers l'application Too Good To Go en tant que "repas antigaspi".

- Item

Un type de panier mis en ligne par le store. Les stores peuvent en effet avoir différents types de produits et différents prix pour un panier vendu. Les différences principales entre les items sont le prix, le type de produits (végétarien, sans gluten, ...) et le type de rayons pour les supermarchés (fruits & légumes, un peu de tout, ...)

- Chain (chaîne)

Un regroupement de commerçants (Monoprix, Franprix, Bio c bon, ...)

- Rating (note)

Une note de 0 à 5 affectée par les utilisateurs après chaque achat de panier.

- Churn (perte de clientèle)
Correspond à un statut affecté à un commerçant dès lors qu'il ne met plus en vente aucun panier sur notre plateforme durant 30 jours consécutifs.

C. Volumétrie

Approximativement 40 colonnes et 4 000 000 de lignes consolidées en format csv :

- 1 ligne par item par jour (il faut ici bien différencier le store de l'item : un store peut avoir un ou plusieurs items disponibles sur l'application)
- Environ 15 000 magasins, chaque magasin pouvant avoir plusieurs items.
- 17 mois d'historique (janvier 2019 à Mai 2020)

D. Principales colonnes

Dimensions

Données qualitatives liées aux items

- **Before Price** : prix de vente avant mise en ligne sur l'application (le cas le plus fréquent étant un before price de 12 euros et un prix de 4 euros sur l'application)
- **Chain Name** : nom de la chaîne auquel appartient le store (Carrefour / Auchan / Paul etc.)
- City
- Item ID
- Item Name
- **Item Price** : prix de vente de l'item sur l'application
- No Unique Consumers
- Pickup Time End Time of Day : heure à partir de laquelle les consommateurs peuvent venir chercher le panier qu'ils ont commandé
- Pickup Time Start Time of Day : heure à partir de laquelle les consommateurs ne peuvent plus venir chercher le panier qu'ils ont commandé
- Store Address
- **Store First Enabled Date** : date à laquelle le store est en ligne sur l'application et peut y proposer des paniers
- **Store First Meal Saved Date** : date à laquelle le store a mis son premier panier en ligne (cette donnée concerne le store et pas l'item. Cela implique que tous les items d'un store auront la même date pour cette colonne)
- Store ID
- **Store Last Meal Saved Date** : date du dernier panier vendu par le store sur l'application
- Store Name
- Store Postal Code
- Store Region

- **Store Segment** : le segment est la principale dimension qui distingue un store d'un autre. Nous avons 23 segments différents, les principaux étant les supermarchés et les boulangeries.
- **Time Date** : jour de la mise en ligne du panier

Mesures

Données numériques liées à l'activité des magasins

- Avg Rating Overall : moyenne des notes données par les consommateurs pour l'item
- Consumer Cancellation : nombre d'annulations des consommateurs (qui peuvent annuler jusqu'à 3 heures avant l'heure de collecte)
- Count Ratings
- Declared Supply : nombre de paniers programmés sur l'application (c'est un nombre de paniers que le store peut programmer à l'avance pour un jour de la semaine donnée. Exemple un store peut programmer 5 paniers tous les lundi et jeudi)
- Item View : Nombre de fois où l'item a été vu sur l'application
- Manual Added Supply : le jour de collecte, le store peut ajouter manuellement des paniers sur l'application pour s'adapter à son nombre réel d'inventus.
- Manual Removed Supply : le jour de collecte, le store peut supprimer manuellement des paniers sur l'application pour s'adapter à son nombre réel d'inventus.
- Meals Refunded : nombre de paniers remboursés. Le store ne perçoit donc pas d'argent.
- **Meals Saved** : nombre de paniers vendus et remis à un consommateur
- Store Cancellation
- **Total Supply** : nombre de paniers mis à disposition par le store sur l'application (Total Supply = Declared Supply + Manual Added Supply + Manual Removed Supply - Store Cancellation)

Nous vous fournissons également un document excel avec une description des données présentes dans les fichiers csv.

On espère avoir été suffisamment clairs et on vous remercie pour votre engagement!

Force & Honneur,

[Michael](#) et [Georges](#) de Too Good To Go.



PS. [DO NOT CLICK HERE](#)

PPS. [NOR HERE](#)

PPPS. [ALSO NOT HERE](#)