

Projet Datamining : Analyse des comportements et des leviers opérationnels

fabrice.simon@aramisauto.com

2023/2024

Contexte :

La marque « Reine des maracas » est soumise à une concurrence de plus en plus forte sur le marché de la lingerie. Une de ses 1ères actions fut de développer une gamme plus large de produits complémentaires avec des maillots de bain et des vêtements homewear.

En parallèle un programme de fidélisation a été développé afin de créer un lien avec les clientes et les récompenser.

Néanmoins afin d'aller plus loin dans l'optique de différenciation par rapport à ses concurrents, « Reine des maracas » souhaite mettre en place une segmentation RFM afin de pouvoir communiquer de façon différenciée avec ses clientes et pouvoir allouer de façon optimale ses investissements marketing.

La marque de lingerie vous mandate donc pour réaliser **une analyse RFM de sa clientèle sur les deux dernières années (extraction au 31 août 2016).**

Données :

Les données à votre disposition sont les tickets de caisse (Table Tickets), les caractéristiques des clients (Table Individus) ainsi que les tables de référence sur les magasins et les produits (Table Magasin, Référentiel, Typologie Produit).

Grandes étapes pour réaliser le projet :

1. Constitution de la matrice de travail

Pour mener à bien votre analyse, la réunion de lancement du projet a permis de faire ressortir la liste des variables intéressantes à étudier par segments de la segmentation RFM :

DIMENSION	AGGREGAT	DEFINITION
CLIENT	AGE	Différence entre la date d'extraction et la date de naissance
CLIENT	SEXE	1 : Homme / 2 : Femme
CLIENT	ANCIENNETE	Différence entre la date d'extraction et la date de création

		de la carte
CLIENT	LOCALISATION DU MAGASIN GESTIONNAIRE	Région (Paris / Province) et Localisation (Centre ville / Commercial)
ACHAT	NOMBRE DE VISITES	Une visite est définie par un numéro de ticket dans un magasin, un jour, à une heure donnée.
ACHAT	PANIER PAR VISITE	Montant des achats d'une visite rapporté au nombre de produits
ACHAT	NOMBRE MOYEN DE PRODUITS PAR VISITE	Moyenne du nombre de produits achetés lors des différentes visites
ACHAT	MONTANT MOYEN PAR VISITE	Moyenne des dépenses réalisées lors des différentes visites
ACHAT	PANIER MOYEN	Moyenne des différents Paniers par visite
COMPORTEMENT	RECENCE	Différence entre la date d'extraction et la date du dernier achat
COMPORTEMENT	PART ACHAT MAGASIN GESTIONNAIRE	Rapport entre le montant réalisé dans le magasin gestionnaire et le montant total des dépenses
TYPE_PRODUIT	NOMBRE DE MAGASINS	Nombre de magasins distincts où le client a réalisé au moins un achat
TYPE_PRODUIT	NB FAMILLES DIFFERENTES	Nombre de familles distinctes où le client a réalisé au moins un achat
TYPE_PRODUIT	NB LIGNES DIFFERENTES	Nombre de lignes distinctes où le client a réalisé au moins un achat
PROGRAMME	NOMBRE DE CADEAUX	Nombre de produits ayant un code modèle MODELE = "FAVO" "FAVORI"

2. Analyse des comportements clients

Le programme d'animation commerciale de « Reine des maracas » est essentiellement basé sur des critères de Récence, Fréquence et Montant.

Aussi une fois que vous aurez découpé le portefeuille clients selon ces 3 critères, vous comparerez les 9 segments clients selon les critères de Sensibilité programme, Eclectisme (nombre de lignes différentes) et Diversification (nombre de familles différentes).

Puis vous choisirez à votre convenance une population regroupée sur laquelle **vous analyserez le profil via les agrégats calculés précédemment**. Le but étant de faire émerger des leviers opérationnels pour optimiser l'animation commerciale.

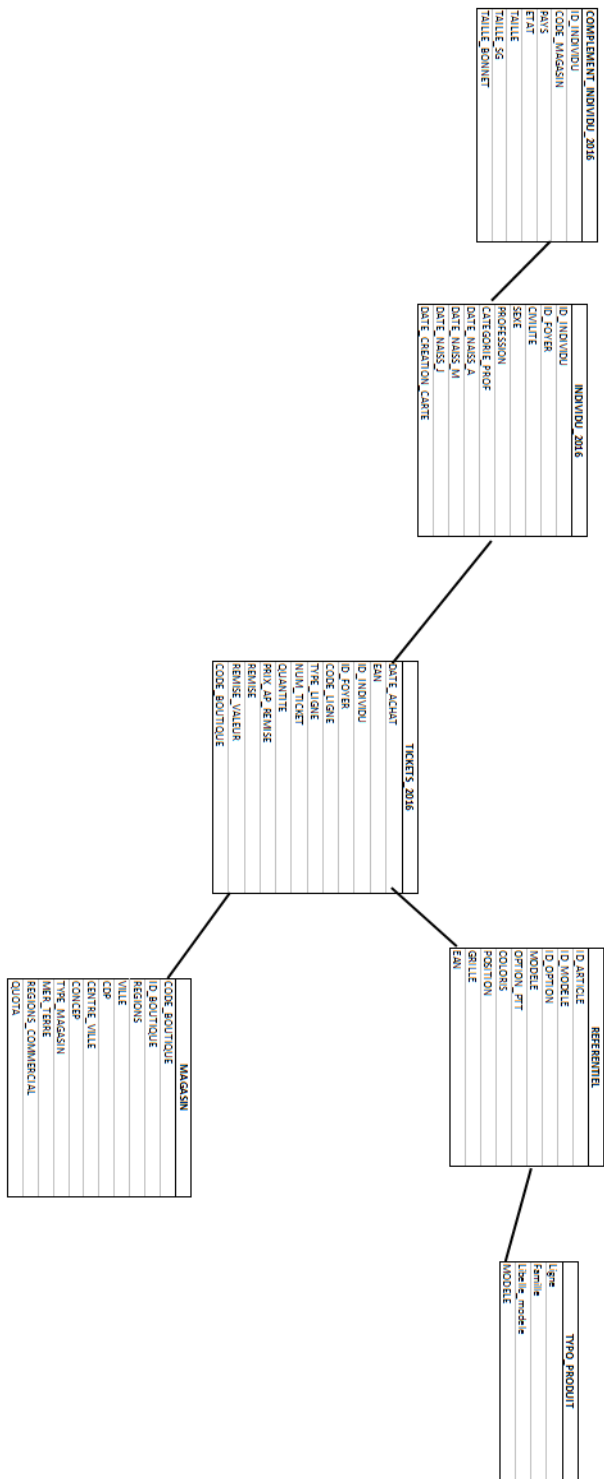
3 : Comité de pilotage – Evaluation du projet

Vous présenterez les résultats de votre projet sous la forme d'un rapport écrit.

Le rapport devra à minima contenir les parties suivantes :

- 1- Contexte & Objectifs de la réunion
- 2- Présentation des résultats de l'analyse RFM
- 3- Analyse du profil d'un groupe de clients
- 4- Conclusion et recommandations

Modèles des données



Guide pratique pour réaliser le projet par groupe :

Ce document a pour objectif de vous aider dans la mise en place du projet sous Python.

0) Initialisation du projet sous VSCODE / Google Collab...







Etape 1 – Installation des packages nécessaires

Pour réaliser votre projet, nous utiliserons les packages recommandés ci-dessous pour le traitement des données :

- Python : Pandas / Numpy / datetime

Etape 2 – Téléchargement des données initiales

Les données sont disponibles sur le lien Drive : http://bit.ly/marketing_rennes2

 R_COMPLEMENT_INDIVIDU_2016
 R_INDIVIDU_2016
 R_MAGASIN
 R_REFERENTIEL
 R_TICKETS_2016
 R_TYPO_PRODUIT

Etape 3 – Import des fichiers csv

Importer les fichiers CSV (délimiteur « ; ») en utilisant les noms d'objets suivants :

- df_complement_individu_2016
- df_individu_2016
- df_magasin
- df_referentiel
- df_tickets_2016
- df_typo_produit

1) Construction du dataframe permettant l'analyse des clients

L'objectif de cette étape est de construire la table client:

- Une table de travail regroupant les individus à l'aide de DF_INDIVIDU_2016 et DF_COMPLEMENT_INDIVIDU_2016)

Etape 1 – Construction de la table au niveau Individus

Pour cette étape, nous devons rapprocher la table DF_INDIVIDU_2016 et DF_COMPLEMENT_INDIVIDU_2016.

Pour réussir cette étape, vous devez vous poser les questions suivantes :

- Combien d'individus je dois retrouver dans ma table finale ?
- Est-ce que je dois conserver les individus de la table DF_INDIVIDU_2016 n'ayant pas de complément ?
- Quelle est la clé de jointure entre mes 2 tables ?
- Quel est le filtre que je dois appliquer pour conserver uniquement les clients utiles sur l'analyse ?

Je peux maintenant écrire le code pour faire la jointure des 2 tables. Est-ce que je retrouve bien le bon volume ?

Etape 2 – Calcul Age / Ancienneté sur la table « Individus »

On cherche maintenant à calculer l'âge des clients à l'aide des informations disponibles dans la table précédente.

Attention, nous ne souhaitons pas l'âge du client aujourd'hui mais à la fin de la période d'extraction des données.

On réalise le même travail pour calculer l'ancienneté des clients.

Pour réussir cette étape, vous devez vous poser les questions suivantes :

- Quelles unités je dois utiliser pour l'âge et l'ancienneté ?
- Quelles sont les données que je dois considérer comme aberrantes pour l'âge et l'ancienneté ? Quelles sont les actions correctrices à réaliser ?

En sortie j'appelle la table de sortie : DF_INDIVIDU

2) Construction du dataframe permettant l'analyse des tickets

- Une table de travail agrégée des tickets à l'aide des tables DF_TICKETS_2016 / DF_REFERENTIEL / DF_TYPO_PRODUIT / DF_MAGASIN

Etape 1 – Travail sur la table des Tickets : Sélection du bon périmètre sur la table ticket

Nous avons vu sur la feuille projet que nous devons sélectionner 2 ans d'historique sur la table DF_TICKETS_2016.

La table en output a 215 577 lignes et 12 colonnes.

Etape 2 – Ajout des données des tables REFERENTIEL / MAGASIN / TYPO PRODUIT dans la table des TICKETS

Avec l'annexe 1 du document, vous pouvez connaître les clés de jointure à utiliser. Il est préférable de faire les jointures 2 à 2 pour valider que le volume de lignes reste constant.

COLUMN COUNT	FILE COUNT	SIZE	RECORD COUNT
34	1	5.6 MB	215 577
2020-01-24 10:37	2020-01-24 10:37	2020-01-24 10:37	2020-01-24 10:37

Pour réussir cette étape, vous devez vous poser les questions suivantes :

- Quelles sont les clés de jointure entre mes tables ?
- Quel est le volume final que je dois obtenir ?

Vous pourrez nommer le dataframe final df_matrice_travail.

Pour réduire la taille des objets sur R, vous pouvez conserver uniquement les colonnes suivantes :

- "DATE_ACHAT" / "EAN" / "ID_INDIVIDU" / "ID_FOYER" / "CODE_LIGNE" / "TYPE_LIGNE" / "NUM_TICKET" / "QUANTITE" / "PRIX_AP_REMISE" / "REMISE" / "REMISE_VALEUR" / "CODE_BOUTIQUE" / "REGIONS" / "CENTRE_VILLE" / "TYPE_MAGASIN" / "REGIONS_COMMERCIAL" / "MODELE" / "Ligne" / "Famille"

3) Auditer la table df_Matrice_travail et df_individu

Etape 1 – Analyse de la table df_Matrice_travail

L'objectif de cette partie est de contrôler la cohérence des données sur les variables suivantes :

- Régions
- Centre_ville
- Type_magasin
- Régions_commercial
- MODELE
- Ligne
- Famille
- EAN
- MODELE
- PRIX_AP_REMISE
- QUANTITE
- REMISE_VALEUR

Analyser les données pour en déduire des incohérences sur certaines colonnes ?

Vous appliquerez les modifications nécessaires pour créer en sortie la table Matrice_travail_OK

4) Calculer avec la table **Matrice_travail_OK** contenant les modifications précédentes

Etape 1 – Déterminer la règle permettant d'identifier un ticket d'achat unique = une visite en magasin

En utilisant la table Ticket_174591, il s'agit de l'extraction de l'individu 174591, vous pouvez déduire la règle permettant d'identifier un ticket unique.

Quelles sont les colonnes à retenir pour identifier un ticket d'un client ?

Le numéro de ticket est important car si on regarde les achats du 16/06/2016, on remarque qu'il y a en fait 3 tickets différents.

Etape 2 – Calculer par individu les indicateurs suivants

Construire une table regroupant les indicateurs suivants, tous les indicateurs ne sont pas calculables en une fois, vous pouvez le faire en autant d'étapes que vous le souhaitez :

- Calculer **par visite** (en utilisant la règle précédente) :
 - Chiffre d'affaires (CA_VISITE)
 - Nb Produits (NB_PRODUITS)
 - Prix Moyen par Produit (hors modèle non retenu) (PRIX_MOYEN)
- Calculer par individu, à partir de la table précédente par visite :
 - Nb Visites (NB_VISITES)
 - Chiffre d'affaires moyen (CA_MOY_VISITE)
 - Chiffre d'affaires cumulé (MONTANT_CUMULE)
 - Nb Produits moyen (NB_PRDT_MOY_VISITE)
- Calculer par individu :
 - Récence du dernier achat (en jours) (RECENCE)
 - Nb Mag différents (NB_MAG_DIFF)
 - Nb lignes de produits différentes (NB_LIGNES_DIFF)
 - Nb familles de produits différents (NB_FAM_DIFF)
 - Nb Cadeaux reçus (NB_CADEAUX)
 - Part des visites dans le magasin gestionnaire (PART_VIST_MAG_GEST = Nb visites MAG_GESTIONNAIRE / Nb visites)

Vous devez ensuite regrouper les indicateurs dans une table finale. Cette table doit contenir 26 871 individus (nommée df_Matrice_finale).

5) Constitution de la RFM

Etape 1 – Constructions de la segmentation RFM

Découper les dimensions calculées précédemment sur les percentiles 33 / 66.

- Montant
- Fréquence
- Récence

Vérifier les seuils

Créer les segments, ne pas oublier les segments complémentaires

6) Grands constats

Etape 1 – Calculer les variables suivantes, elles permettront d'observer les différents groupes

Analyser les dimensions Eclectisme / Réactivité / Diversification / Attachement avec les variables adéquates. Pour cela regrouper en classe ces variables selon votre analyse.

DIMENSION	AGGREGAT
CLIENTS	AGE
CLIENTS	SEXE
CLIENTS	ANCIENNETE
ACHAT	MONTANT DES ACHATS CUMULE
ACHAT	PANIER MOYEN
ACHAT	NOMBRE DE VISITE
COMPOTEMENT	RECENCE
COMPOTEMENT	LOCALISATION MAGASIN GESTIONNAIRE
COMPOTEMENT	PART ACHAT MAGASIN GESTIONNAIRE
TYPE_PRODUITS	NOMBRE DE MAGASINS
TYPE_PRODUITS	NB DE FAMILLES DIFFERENTES
TYPE_PRODUITS	NB DE LIGNES DIFFERENTES
PROGRAMME	NOMBRE DE CADEAUX

Choisir un segment et l'étudier précisément afin de répondre à la problématique de la marque. Réaliser des recommandations marketing.