Domaine du temps:

1 -  ~~Grouper les achats/retours par mois et vérifier s’il y a un periode du mois que des gens retourne plus (justificatif: quelqu’un a dependu plus d’argent qu’il avais dans le début du mois et a décidé retourner pour sauver d’argent)~~

2 - ~~Grouper les achats/retours par jour de la semaine (verifier si proche du weekend des gens achète plus et retourne dans lundi/mard, fête)~~

3 - ~~Grouper les achats/retours par heure (vérifier si dans le nuit des gens achete et regrette la matin prochain)~~

~~4 - Vérifier pour le type de chaussures~~

~~5 - Vérifier type product~~

~~6 - Vérifier si la première achat dans le site a plus retour que des clients fidélisés (quelqu’un n’est pas aimé le service)~~

~~7 - Vérifier si un personne qui achat un chassure pour un genre différent retourne plus~~

~~8 - Intégrer productcolorid (parce que couleur “flash” difficile à estimer sur une photo),~~ Renata: j’essaie et ça change rien

~~9 - Intégrer SeasonLabel (manteau plus difficile à estimer sur internet)~~

~~10 - Intégrer SizeAdviceDescription. Ultra important pour le choix de la taille~~

11- Idées de pré traitement : beaucoup de features présentent des corrélations. On peut d’abord essayer de choisir les features qui corrèlent le moins. (version 8 pour matrice de corrélation)

12- On peut aussi sélectionner que les features qui ont un score >0.5 pour chaque classifier (les scores ont été calculés version 8) - Sélection faite pour la linear regression version 9 , en revanche, il faut ré-implémenter l’algorithme de sélection de features ou sinon procéder manuellement.