

Rapport Challenge HelloZack

Salut Tim, voici le détaillé de ma résolution du 'challenge' HelloZack; code python commenté et .csv des valeurs d'Iphone 6S 64GB sur les 3 derniers mois en PJ. Ma réponse en 3 étapes:

I) Inscription sur Ebay et découverte des outils (en particulier de la Finding API)

➔ Les requêtes utilisant les APIs Ebay nécessitent l'obtention de credentials (*appid*, *devid*, *certid*, *token*). En ce qui concerne 'Finding', seule l'*appid* est requise pour effectuer des recherches. On peut l'obtenir via la création d'un compte sur l'annexe developer du site d'Ebay.

➔ L'*appid* ainsi récupérée est utilisée pour la requête 'Finding' faite à Ebay. La 'Finding API' comporte différents fields qui permettent de déterminer le type de recherche voulu par le programmeur (par keyword, catégorie, recherche avancée...). Chacun de ces fields est accompagné d'une généalogie de subfields permettant de préciser la requête. C'est ainsi qu'il est possible de filtrer la recherche avec des mots tels que 'Iphone', '6S', ou de gérer le nombre de pages d'outputs voulu (et tant d'autres paramètres!).

II) Cibler notre recherche de façon optimale (exérimental)

➔ Afin de faire augmenter la popularité de leur article dans les recherches des acheteurs, les e-vendeurs ajoutent des mots clés au titre de leur article. Il est donc crucial de parvenir à filtrer correctement une recherche (afin de ne pas se retrouver avec toutes les chaussettes pour Iphone dans notre base de donnée).

Dans un 1e temps, ceci est possible par l'observation des attributs des items de la dataset: quels titres correspondent effectivement aux articles qui nous intéressent? "*iphone – 6S – 64GB*"

Dans un 2e temps, il s'agit d'éviter des recherches trop larges: le maximum d'articles visible lors d'une recherche Finding est de 10 000 (100pages à 100 entrées). Une requête trop vaste implique donc une perte d'informations.

➔ Si certains articles passent malgré cela l'étape du filtrage par Keyword (outliers), il est possible de les filtrer par prix par exemple.

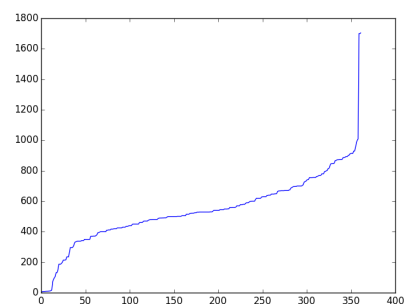


Figure 1 : Prix de toutes les ventes iPhones 6S - 64GB par ordre croissant

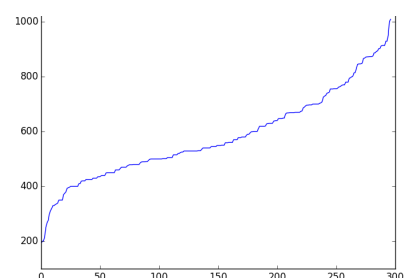


Figure 2 : Prix de toutes les ventes iPhones 6S - 64GB par ordre croissant
FILTRE A PRIX

****cette étape a son équivalent dans le filtrage par Keyword; cependant, il est important de mettre en place du type 'filtre à prix' indépendamment pour permettre un focus sur certains types de budgets par exemple*

III) Utilisation de pandas pour stocker et manipuler les données → obtention de la moyenne

On stocke ensuite les données dans des DataFrames pandas (on commence par placer les attributs intéressants – prix, disponibilité, vente réussie ... – dans des Series, que l'on concatène ensuite en DataFrames). Celles-ci vont nous permettre une manipulation efficace et rapide des informations (des bibliothèques telles que *scikit learn* sont alors très utiles pour implémenter des algos de machine learning dans le code).

→ Dans le cadre d'une analyse temporelle (évolution du prix moyen de l'Iphone 6S 64GB sur 3mois), *pandas* propose un type 'datetime' permettant de gérer les time series. Ainsi, il nous est possible de découper nos 3mois en intervalles de temps (pertinence de l'intervalle à déterminer en fonction du besoin), sur lesquels il est alors possible de simplement moyenner le prix d'une vente.

On trouve avec ceci la moyenne du prix de vente d'un Iphone 6S - 64GB sur les 3derniers mois.

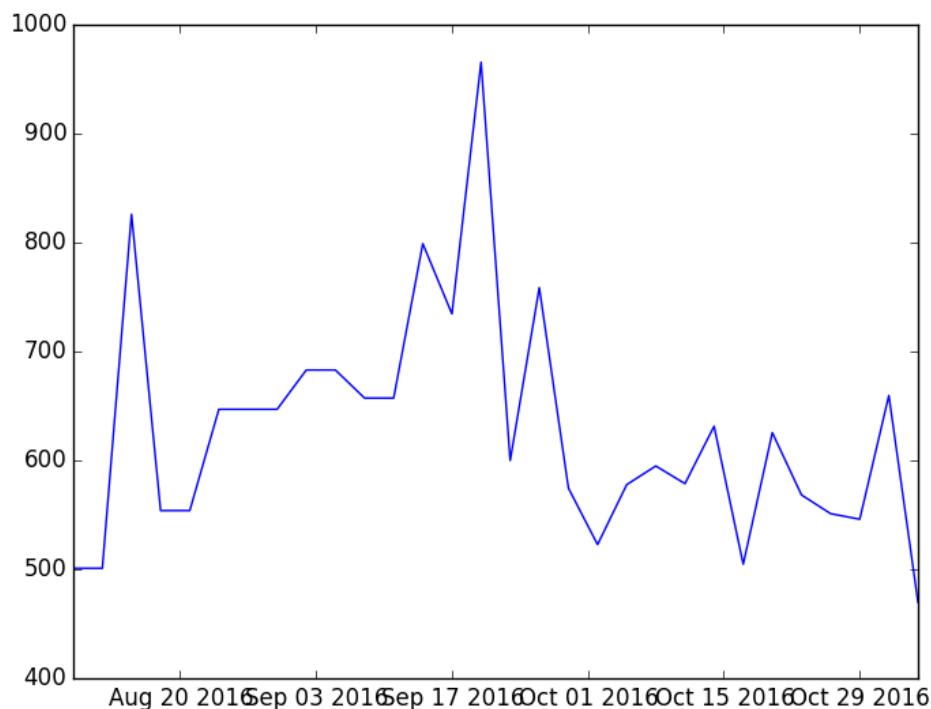


Figure 3 : Evolution du prix de vente d'un Iphone 6S - 64GB (sensibilité = 3 jours)