**Introduction**

Nous sommes 3 étudiants du bachelier complémentaire en Business Data Analysis, ce bachelier est réalisé à la Haute Ecole l’Ephec. Lors de notre dernière année, il nous est demandé de réaliser un travail de fin d’étude. La problématique de notre travail de fin d’études est : « Quelle sont les facteurs qui influences la réussite d'une application et déterminer ce que signifie la réussite une application ? » Afin de répondre à cette question, nous avons récupéré une base de données sur kaggle que nous avons nettoyer et mise en forme afin de pouvoir produire ce travail. Il est important de préciser que la base de données dates de 2019 et qu’elle ne reprend uniquement les applications téléchargées sur le google play store.

Bien sûr la définition de réussite est très subjective. C’est pourquoi nous avons pris la décision de faire une enquête afin de ne pas définir la réussite d’une application uniquement sur nos critères personnels mais sur un échantillon de population plus large.

Nous nous sommes penchés sur cette problématique car le développement ainsi que le marketing derrière le développement sont des éléments qui nous intéressent. Probablement dû à notre background (2 Diplômés en Technologie de l’informatique ainsi qu’un Diplômé en marketing option digital marketing).

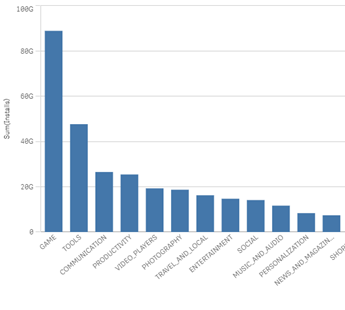
         Lors de ce travaille nous allons choisir des critères pertinents qui peuvent avoir un impact sur la qualité ou la popularité d’une application. Il y a pour cela 2 approches ; Il faudra aussi déterminer si l’une des 2 approches est plus efficace que l’autre.

L’approche plus technique avec la technique comme le contenu voire même le code source de l’application, le poids de stockage de l’application, les mises à jour ainsi que d’autres critères qui seront traité plus en détails dans ce travail de fin d’étude.

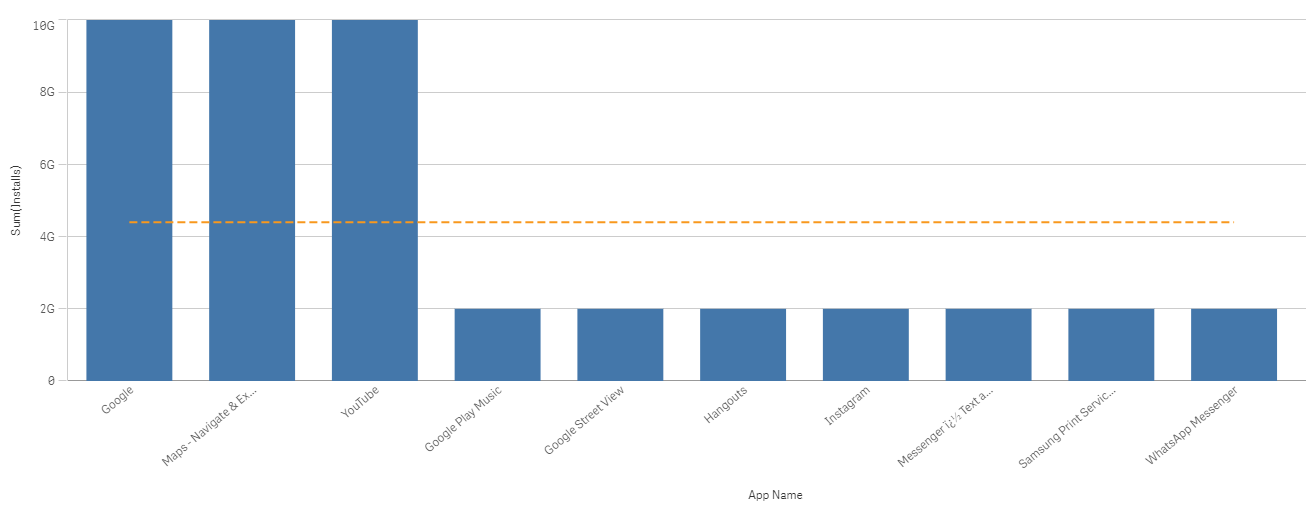
Deuxièmement une approche plus marketing la manière dont l’application se fait connaître ainsi que les pubs qui sont réalisés dans le but de créer un engouement autour de cette application. Dans le cadre de cette deuxième approche, nous avons réalisé une enquête sur les habitudes qu’on les utilisateurs des stores d’applications. Les résultats de cette enquête seront exposés dans ce travail.

**La problématique**

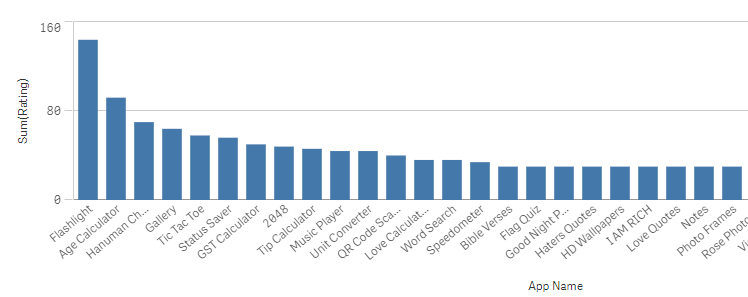
Afin de présenter notre problématique nous avons choisi de d’abord montrer quelques graphiques que nous trouvons intéressants. Pour démontrer le choix de notre problématique.



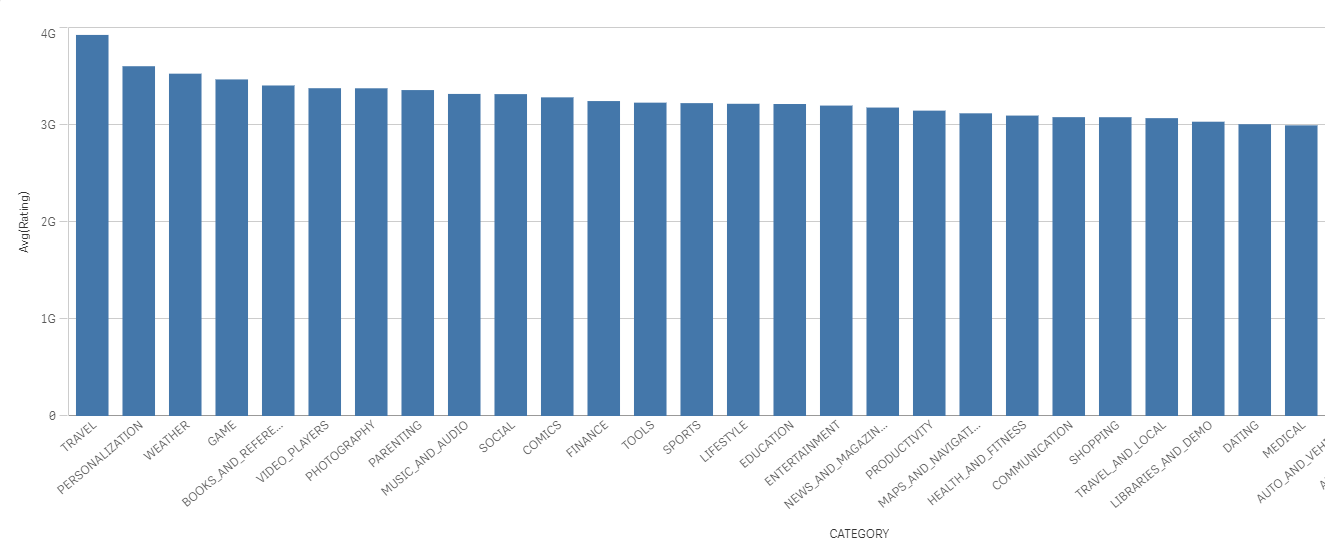
On peut voir sur ce graphique qui reprend la somme des nombres de téléchargements des applications selon leurs catégories. On voit très clairement que (la catégorie (“GAME”) les jeux sont nettement les plus télécharger mais également que les outils sont aussi énormément téléchargés, loin devant le reste.



Sur ce graphique on voit pourtant que dans le top 10 des applications les plus téléchargées, aucune ne fait partie de la catégorie “GAME”. Mais on peut aussi voir que les 3 applications qui ont largement la tête sont des applications qui sont premièrement a google eu même mais surtout que ces 3 applications sont très souvent installé de base, dès le premier démarrage sur vos smartphone.



Nous avons ici repris la note moyenne que les applications ont reçu de leur utilisateurs. La question qui se pose est alors, quel est l’impacte que la note moyenne à sur le nombre de téléchargements.



Sur ce graphique, nous avons voulu comparer les ratings moyens en fonction des catégories et on voit par exemple que les application de voyage (TRAVEL) ont une note données par leurs utilisateurs plus élevé que les autres catégories. Or les applications de voyage ne se trouve pas dans le top des applications les plus télécharger.

Avec ces quelques exemples on peut alors se demander quels sont les facteurs qui font qu’une application est de nombreuse fois téléchargée. Mais une application qui a de nombreux téléchargement signifie-t-elle que c’est une applications réussie ? C’est également ce que traite notre problématique.

**La réussite**

Comme introduit dans l’introduction, la réussite d’une application est une impression qui est personnels et difficilement quantifiable. C’est pourquoi nous définirons la signification de réussite dans le cadre de ce travail.

Selon le dictionnaire Larousse : *“1. Succès, résultat favorable. La réussite d’une entreprise. 2. Entreprise, action, oeuvre qui connaît le succès. Notre voyage n’a pas été une réussite. 3. Jeu de cartes au cours duquel un loueur solitaire s’efforce de placer ou d’employer toutes les cartes selon certaines règles dans une combinaison déterminée par le hasard.”*

Nous allons donc définir dans le cadre de ce travail, de manière homonyme à la définition 1, le succès sera comparé aux nombres de téléchargements et donc indirectement la popularité de l’application. Et nous allons associer le résultat favorable à l’appréciation que les utilisateurs ont de cette application, c’est-à-dire à la valeur « rating ».

**Etude quantitative**

Lors de réflexion qui précède ce travail, nous avons pensé que réaliser une enquête pourrait nous aiguiller ou confirmé les sujets auquel nous devrons plus particulièrement prêter attention dans le cadre de notre problématique. De prime abord, le nombres de critères qui détermine si quelqu’un va décider de télécharger une application plutôt qu’une autre sont extrêmement nombres et personnel. C’est pourquoi nous avons réalisé cette enquête, qui a l’heure où j’écris ces lignes, a déjà recueilli plus de 270 réponses.

Premièrement, je pense qu’il faut décrire l'échantillon qui a répondu à notre enquête. Sans rentrer dans les détails et afin d’éviter de bombarder de chiffre, les graphiques et les chiffres seront disponible en annexe. Il est donc important de préciser que notre échantillon est composé de personnes de tout âge, de moins de 18 ans à plus de 50 ans. Il va de soit que la tranche de la population la plus âgé n’est pas ou très peu représenté dans cet échantillon, puisque cette enquête ne l’est sûrement que peu parvenu et que pour certaines personnes complétée cette enquête peut s'avérer compliqué. D’un point de vue des proportions d’âges, il y a une large majorité de personnes entre 18 et 30 ans. Au niveau des genres, nous avons plus d’hommes que de femmes mais sans grande disparité.

Nous pensons qu’il est également important de préciser qu’une très large majorité de notre échantillon a un diplôme d’études supérieur. Seulement 15% de l'échantillon ne possède pas de diplôme supérieur.

Lors de commencement de cette étude, nous avions un doute sur la quantité de personnes qui peuvent potentiellement être touchés par le marché qu’est le marché des applications. Dans notre échantillon qui est assez varié comme décrit précédemment, on se rend compte que 99,6% de notre échantillon possède un smartphone.

Malheureusement dans le cadre de ce travail, 49% des personnes qui ont répondu à l’enquête possède un smartphone sous Android. Notre dataset étant basé sur le google play store qui n’est accessible que pour les utilisateurs Android. Nous avons donc décidé d’élargir notre analyse réalisée sur le google play store ou App store IOs car nous pouvons voir sur les graphiques que le comportement des utilisateurs des 2 ont des comportements et réponses similaire.

**Analyse des reviews (Text mining)**

**Nettoyage de la base de données**

Afin d’analyser les retours des utilisateurs, nous avons décidé de prendre une base de données de commentaire d’applications du Play Store. Avant tout travail d’analyse, il a fallu faire un gros travail de préparation des données. Le fichier que nous avons récupéré était en format csv avec toutes les données séparées par des virgules. Il a fallu donc dans un premier temps mettre les données sous forme de tableau ainsi que nettoyer toutes les données. Il est important de préciser que pour chaque ligne de données, il était précisé si c’était un avis, commentaire positif négatif ou neutre.

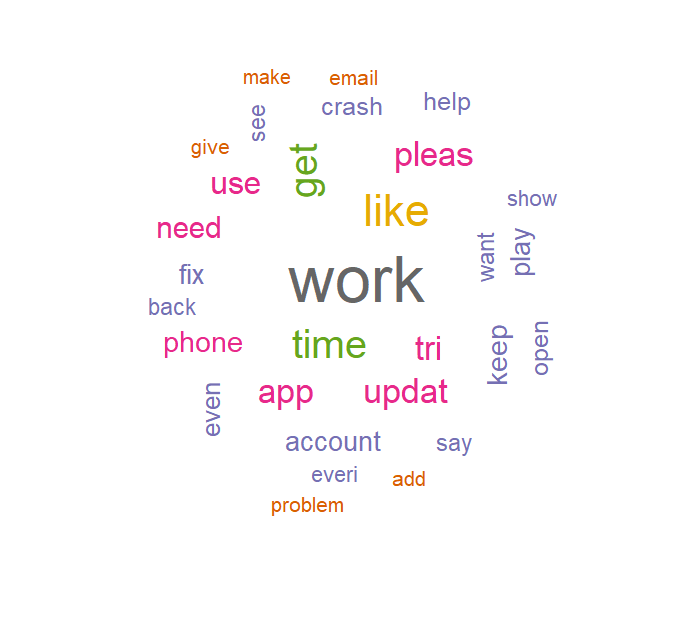
Afin d’analyser au mieux et de manière visuel les commentaires, la décision a été prise de réutiliser une méthode d’affichage des mots les plus utilisés dans un texte, méthode vue durant les cours. Nous avons donc mis tous les avis neutre, négatif et positif chacun dans un fichier txt afin de pouvoir les traiter dans Rstudio. Voici les résultats des 30 mots les plus utilisés selon les 3 catégories (Positif, Neutre et Négatif).

**Positif**

****

Comme on peut le vrai rapidement les mots qui sont principalement utiliser dans les commentaires positifs des utilisateurs sont sans surprises des mots tel que : “love, great, good, like, best”. Ce qui correspond bien à l’avis positif qu’ils ont laissé.

**Neutre**

****

Nous pouvons voir dans les commentaires neutres que les mots les plus utilisés sont à la fois positif et négatif. Comme par exemple like et crash. On peut aussi voir que les utilisateurs qui mettent des avis neutres semblent également plus avide de donner ce qu’on soupçonne comme des “conseils”. Avec les mots update, fix, add et want par exemple. On peut donc penser que les utilisateurs qui prennent la peine de faire un report sur une application considéré comme moyen essayent que celle-ci s’améliore.

**Négatif**



Lors de l’analyse de ce schéma le mot qui saute aux yeux et qui est donc le plus utilisé est “game”. Ce qui signifie que les jeux ont souvent des commentaires négatifs. Dans le reste de ce travail, le jeu domine le rating positif. Or, ici on voit que dans le négatif il est également fort présent.

On voit également beaucoup de mots négatifs telle que “hate, fix, update” ce qui est logique.