

Data Challenge : analyser le comportement des consommateurs de services en ligne

Woohyun Shin

Phuong Nguyen

Julien Guerin

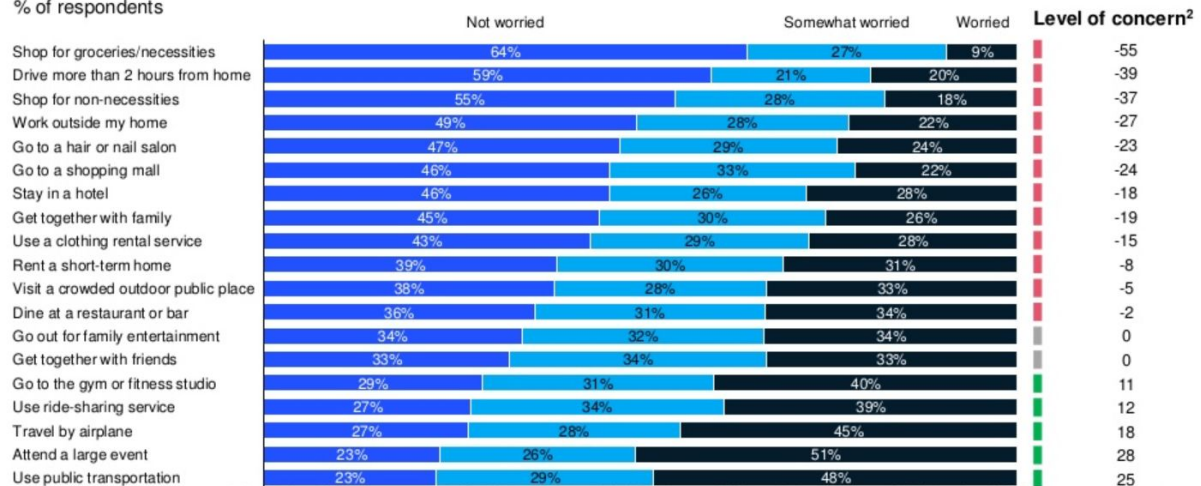
Introduction

Depuis le COVID-19, les comportements des consommateurs ont beaucoup évolué, autant dans le monde qu'en France. Les tendances longues comme l'adoption du e-commerce, davantage de réservations en ligne, livraison à domicile de plats et produits se sont accélérées.

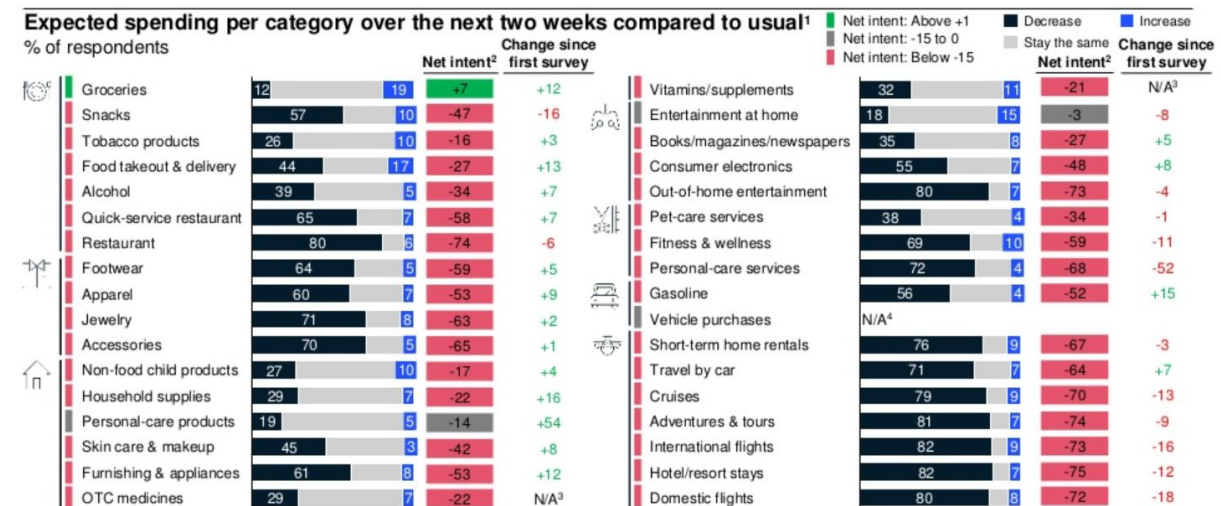
Un sondage de McKinsey montre que ce qui inquiète le plus les français est l'utilisation des transports publics, assister à un événement public, et voyager en avion. On peut donc s'attendre à ce que ces secteurs soient touchés, et à ce que le comportement des consommateurs soit modifié sur le long-terme.

Consumers' level of concern undertaking various activities¹

% of respondents



Les secteurs les plus impactés sont le voyage et le tourisme —croisières, vols internationaux et domestiques, réservations d’hôtels— qui étaient attendus en baisse de 80%.



Nous avons choisi d’étudier le comportement des consommateurs pour les secteurs du tourisme et des voyages, très pertinents pour capturer cette réticence des français à voyager.

L’impact du COVID sur ces industries pouvant avoir des répercussions durables, nous voulons donc estimer la demande en temps réelle pour les vols d’avion et les réservations de chambre d’hôtel des français à l’étranger.

Data acquisition et Méthodologie

- Data acquisition

Ayant décidé de nous focaliser sur les secteur du tourisme et voyage, nous avons défini une liste de catégories et mots clés à récupérer sur Google Trends.

Pour définir les catégories, nous avons utilisé un listing créé par Patrick Trasborg sur Github (<https://github.com/pat310/google-trends-api/wiki/Google-Trends-Categories>). Parmi toutes les catégories calculées par Google, nous avons retenu : "Travel", "Air Travel", "Travel Agencies & Services", "Hotels & Accommodations", "Car Rental & Taxi Services".

Pour définir les keywords, nous avons inclus les recherches types les plus appropriées à nos domaines d'activités. Par exemple une requête type pour effectuer des réservations de billets d'avion serait "acheter billet avion Londres". Nous avons défini nos keywords en incluant une partie "achat" pour mesurer un comportement de consommation ("acheter" dans notre exemple), et une partie variable comme la destination ("Londres" ici).

Nous avons ensuite récupéré les données journalières de toutes 33 catégories, depuis janvier 2019 à janvier 2021. Les features variant entre 0 et 100, où 100 correspond au nombre maximal de requêtes effectuées en France, du 01/2019 au 01/2021.

	hotel Londres	hotel Barcelone	hotel Miami	hotel New York	hotel Lisbonne	hotel Rome	booking	airbnb	air france	skyscanner	...	voyage Barcelone	visiter Barcelone	voyage Miami	visiter Miami	voyage New York	visiter New York	voyage Lisbonne	visiter Lisbonne	voyage Rome	visiter Rome
date																					
2019-01-07	43.0	64.0	100.0	93.0	70.0	74.0	53.0	48.0	91.0	68.0	...	18.0	0.0	32.0	28.0	88.0	50.0	16.0	80.0	74.0	100.0
2019-01-08	43.0	64.0	100.0	93.0	70.0	74.0	53.0	48.0	91.0	68.0	...	18.0	0.0	32.0	28.0	88.0	50.0	16.0	80.0	74.0	100.0
2019-01-09	43.0	64.0	100.0	93.0	70.0	74.0	53.0	48.0	91.0	68.0	...	18.0	0.0	32.0	28.0	88.0	50.0	16.0	80.0	74.0	100.0
2019-01-10	43.0	64.0	100.0	93.0	70.0	74.0	53.0	48.0	91.0	68.0	...	18.0	0.0	32.0	28.0	88.0	50.0	16.0	80.0	74.0	100.0
2019-01-11	43.0	64.0	100.0	93.0	70.0	74.0	53.0	48.0	91.0	68.0	...	18.0	0.0	32.0	28.0	88.0	50.0	16.0	80.0	74.0	100.0

Nous avons également collecté des données des marchés financiers, pour observer l'impact du choc de demande sur les cours de bourse des entreprises. Nous avons récupéré l'historique des cotations journalières d'entreprises du secteur touristique : Airbus, Air france, Mariotte, Hitton, Booking.com.

	AF.PA Close	AF.PA Volume	AIR.PA Close	AIR.PA Volume	MAR Close	MAR Volume	BKNG Close	BKNG Volume	HLT Close	HLT Volume	All Close	Avg All Close	Avg All Volume
date													
2019-01-01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2019-01-02	10.48866	2479675.0	95.112098	1483033.0	107.459999	1802100.0	1721.699951	312600.0	70.940002	3629500.0	101.286049	498.803013	1806808.25
2019-01-03	10.27848	3061087.0	91.767299	1964509.0	101.739998	3344300.0	1663.119995	445200.0	67.300003	5629200.0	96.753648	480.981824	2845802.25
2019-01-04	10.47284	2822536.0	96.185603	2006859.0	107.809998	3045300.0	1717.550049	541000.0	69.940002	4110700.0	101.997800	497.871413	2425964.75

Enfin, nous avons récupéré les données officielles du nombre de cas par jour, mis en ligne par Hannah Ritchie sur la plateforme *ourworldindata.org*.

- Data preprocessing

Avec les séries Google Trends téléchargées, nous avons construit 5 sous-indices pour mesurer les tendances de consommation. Ces indices sont une moyenne des keywords qui les composent. Les catégories sont définies de la façon suivante :

```
'hotel' : ['hotel Londres', 'hotel Barcelone', 'hotel Miami', 'hotel New York', 'hotel Lisbonne', 'hotel Rome'],  
'voyage' : ['voyage Angleterre', 'voyage Barcelone', 'voyage Espagne', 'voyage Italie', 'voyage Rome', 'voyage USA',  
            'voyage Lisbonne', 'voyage Londres', 'voyage Miami', 'voyage New York', 'voyage Portugal'],  
'visiter' : ['visiter Angleterre', 'visiter Barcelone', 'visiter Espagne', 'visiter Italie', 'visiter Rome', 'visiter USA',  
             'visiter Lisbonne', 'visiter Londres', 'visiter Miami', 'visiter New York', 'visiter Portugal'],  
'other' : ['booking', 'airbnb', 'air france', 'skyscanner', 'billet avion']
```

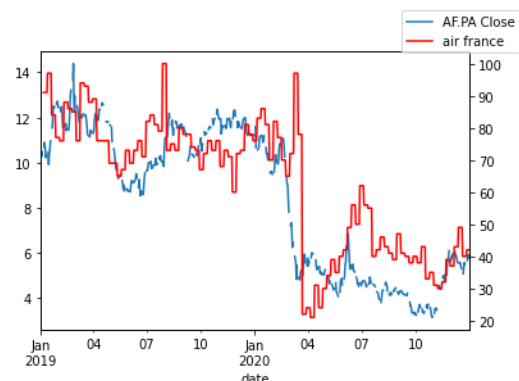
Pour les cours de bourse des entreprises, nous avons simplement construit un indice boursier en prenant la moyenne des cours de bourse normalisés (moyenne de 0, écart-type de 1) des entreprises. Nous formalisons les cours car le prix des actions ne dépend pas uniquement de la capitalisation, mais aussi du nombre d'actions.

Résultats

- Impact de la pandémie sur le comportement des consommateurs

Le prix de l'action de l'Air France a chuté lors du début du confinement en Mars, ainsi que le volume de recherche Google.

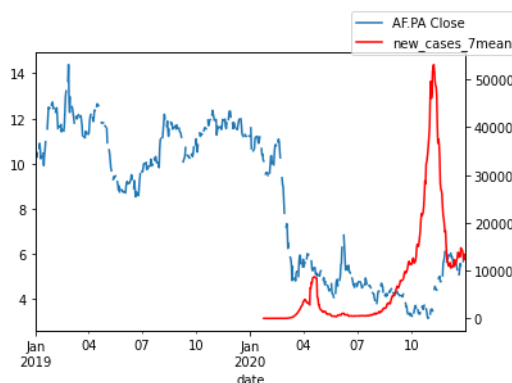
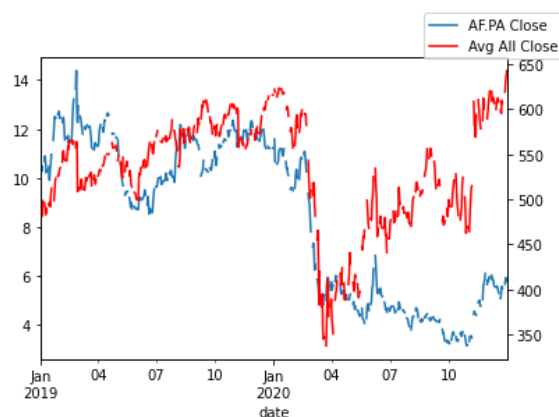
L'intention d'achats de billets d'avions est très basse en mars 2020, en comparaison de 2019. De plus, les deux courbes évoluent dans le même sens, mais le cours de bourse d'Air France a été plus touché par le COVID que les recherches sur Google.



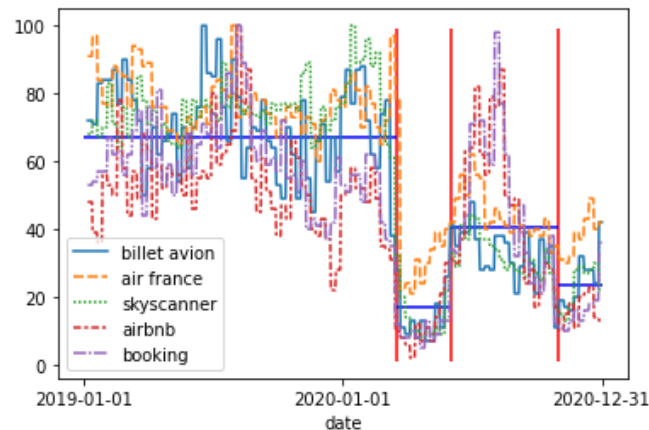
Grâce à l'API Google Trends, nous avons obtenu les données d'entreprises du secteur touristique. La figure ci-contre montre le prix de l'action Air France face à ses concurrents (Airbus, Mariotte, Booking.com, Hilton).

Le cours de bourse des concurrents est resté bas jusqu'à Décembre 2020. Le nombre de touristes étrangers a diminué énormément à cause des règles sanitaires, mais français ont pu continuer de voyager en France.

Sur le graphique ci-dessous, la courbe de rouge représente le nombre de cas COVID en France. Au mois de Novembre 2020 en France, le nombre de cas était à son maximum, et l'action Air France était alors au plus bas.



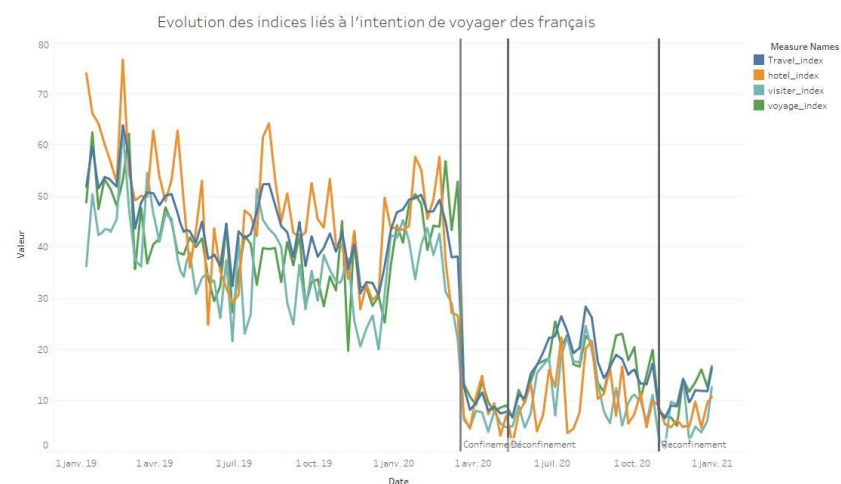
Nous avons ensuite étudié les requêtes pour les keywords : “billet avion”, “air france”, “skyscanner”, “airbnb”, “booking”. La 1^e ligne rouge verticale correspond au début du 1^e confinement en France (17/03/2020), la 2^e est la fin de ce confinement (02/06/2020), et la dernière est la date du début du re-confinement (30/10/2020). La fréquence moyenne des tous les keywords entre la période sont les lignes bleues horizontales.



On voit que les recherches liées aux keywords correspondant aux entreprises du secteur touristique ont brusquement chuté après chaque confinement de la population.

- Indicateur entreprise

Nous avons construit dans un premier temps les différents indices liés à l'intention de voyager des français. Ce sont en effet des indices tirés de Google Trends qui traduisent l'envie de voyager (à travers d'un ensemble de mots clés commençant par “travel”, “voyager” ou “visiter”) et l'intention de réserver un hébergement (mots clés commençant par “hôtel”).

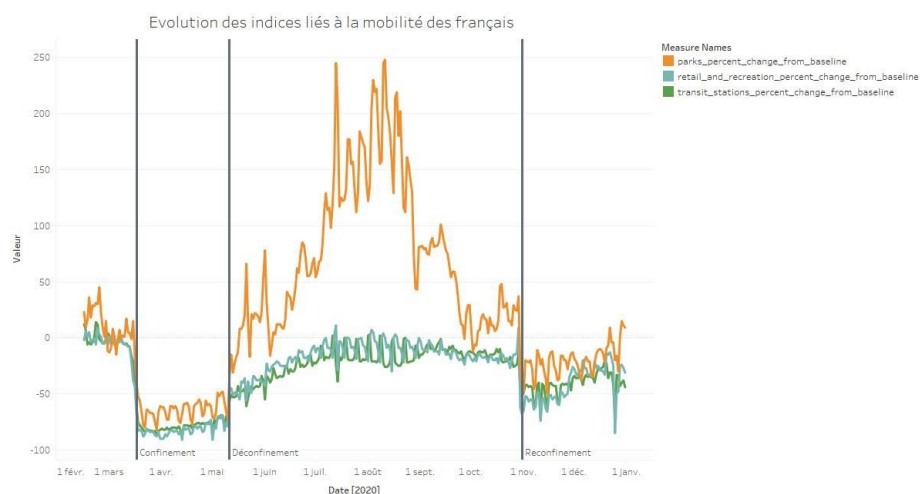


En mettant en relation ces indices aux dates liées aux mesures gouvernementales contre le COVID, nous constatons que l'intention de voyager des français a fortement baissé suite au confinement qui a commencé le 17 mars 2020. Le déconfinement marque le début d'une petite période de rebondissement où les français se sont exprimés de leur envie de voyager.

De la même manière, on voit que ces indices sont en chute dès l'annonce du reconfinement fin octobre 2020.

Nous pouvons donc conclure que les données collectées de Google Trends nous donnent de très bons indicateurs en temps réel sur l'intention des français de voyager et ainsi de consommer les services en ligne tels que réservation du billet d'avion ou de chambre d'hôtel.

Dans notre deuxième graphique, les indices liés à la mobilité des français sont présentés. Parmi ceux collectés par Google Mobility, nous avons fait le choix de présenter les trois les plus pertinents. En effet, ce sont les indices les plus liés aux activités extérieures des français. "Parks" traduit les activités en plein air (parcs nationaux, plages, jardins publics...). "Retail_and_recreation" correspond aux lieux de loisirs comme cafés, restaurants, librairies, musées, théâtres... Enfin, "transit_stations" représente la fréquentation dans les gares et les stations de transport public.



Nous constatons une très forte hausse de fréquentation des activités en plein air après le déconfinement. L'intérêt du public pour les voyages en plein air est beaucoup plus élevé (hausse de 200%) que pour les autres activités (hausse de 30%), et si les autres loisirs peinent à retrouver leur niveau d'avant la crise, le plein air connaît une vraie explosion.

Le confinement ainsi que les mesures de distanciation sociale ont certainement joué dans la préférence des français pour les activités en plein air.

L'indicateur de mobilité montre que le désir de voyager est toujours bien présent, mais qu'il est exprimé sous une autre forme plus adaptée au contexte de restriction. La pandémie a fortement changé le comportement des consommateurs du secteur du tourisme. Ils ne voyagent plus de la même façon qu'auparavant. Les acteurs de l'industrie devraient être à l'écoute du public pour mieux comprendre leurs envies, leurs besoins et pour proposer des services qui correspondent mieux à leurs attentes.

Conclusion

Grâce à l'utilisation des données de Google Trends, nous avons pu voir un lien avec les performances financières des entreprises du secteur touristique.

Ce secteur a été particulièrement affecté par la crise du COVID, et a permis d'observer en direct un changement dans les habitudes de consommation des français sur internet. Ce choc de demande a largement affecté les performances boursières d'entreprises, et nous avons vu qu'il était possible de suivre en temps réel cette information à l'aide de données alternatives à haute fréquence : les recherches en ligne sur Google.

Les implications pour les entreprises sont donc de réaliser l'intérêt de suivre en temps réel l'évolution de certains types de queries, de larges chocs ou fortes variations donnant des informations très précieuses pour l'avenir.

Annexe

- Data sources

Chiffre officiels du COVID : <https://ourworldindata.org/coronavirus-source-data>

Data Google Mobility : <https://ourworldindata.org/covid-mobility-trends>

- Recherches

Bortoli, and Combes (2015). Apports de Google Trends pour prévoir la conjoncture française: des pistes limitées. Note de conjoncture, INSEE, mars.

Robin F., Utilisation des données Google Trends dans l'enquête de conjoncture commerce de détail de la Banque De France.

Gawlik, Kabaria and Kaur (2011), Predicting tourism trends with Google Insights

McKinsey (2020), French consumer sentiment during the coronavirus crisis