**Consignes Note d’étape :**

*Cette note d’étape ne doit pas être une version préliminaire du rapport mais plutôt un point sur le travail déjà fait, ce que vous envisagez de faire, les difficultés rencontrées et anticipées, les pistes prévues pour les résoudre, et quelques résultats préliminaires. C’est un document de 3-4 pages maximum, sans annexe, qui permet à votre correspondant ENSAE de voir si vous êtes sur les bons rails.*

Note d’étape

# Présentation du projet

L’objectif à terme de ce projet de statistiques appliquées est de créer une prévision à court terme (un « nowcast ») de la consommation en France à l’aide autant de données officielles qu’alternatives (Sources officielles de vente de détail, sources officielles liées à l’économie, sources de données en temps réel des douanes, autres données alternatives (Google Trends, etc.)

Pour faire cela, il nous faut étudier la littérature sur les modèles de nowcast utilisés par les instituts économiques pour identifier les modèles les mieux adaptés à notre situation et les implémenter. Il nous faut aussi diviser la consommation des ménages en différents items et identifier les composantes qui expliquent le mieux les fluctuations macroéconomiques pour les intégrer à nos modèles. Puis il faut chercher à reproduire les évolutions des variables officielles et comparer les différentes approches pour enfin obtenir un indicateur répliquant en temps réel la consommation des ménages en France.

# Travail effectué

Pour commencer nous avons dû faire un tour d’horizon de la littérature sur le « nowcast » de la consommation afin de trouver un article de recherche intéressant sur la base duquel nous pourrions travailler. Nous avons donc consacré Novembre à la recherche d’un article à répliquer.

Après avoir lu de nombreux articles nous avons décidés que chacun devrait présenter 1 article lors d’une réunion entre nous. A la suite de cette réunion nous avons décidé de présenter à QuantCube un seul article. Cet article de la banque d’Espagne de 2018 nous a semblé particulièrement pertinent puisqu’il cherchait à établir un « nowcast » de la consommation en Espagne et ce à l’aide autant de données officielles qu’alternatives.

Voici le titre de cet article : nowcasting private consumption: traditional indicators, uncertainty measures, credit cards and some internet data.

Nous l’avons donc présenté à QuantCube. Ils l’ont trouvé fort intéressant et nous avons donc décidé de nous baser sur celui-ci pour la suite. Nous avons ensuite déterminé la marche à suivre pour la suite du projet :

Il a été décidé que nous devrions faire l’étude de la variance de la consommation (étudier comment elle se décompose entre les différentes composantes de la consommation) et étudier le calcul de la consommation pour la France. Puis que nous devrions constituer un dataset avec les variables officielles et faire une étude statistique entre les variables officielles et la consommation. Ceci pour pouvoir ensuite faire un premier modèle de prédiction de la consommation à l’aide de la consommation passée. Ensuite il nous faudra faire un second modèle incorporant en plus les données officielles et enfin un troisième en ajoutant les données google trend. Finalement il nous faudra comparer les résultats des 3 modèles.

Nous nous sommes donc mis à la tâche. Nous avons tout d’abord cherché des données fiables. Nous avons commencé par trouver un ensemble de données contenant la consommation des ménages, 20 fonctions de consommation et 577 catégories de biens et services et ce depuis 1959 à un rythme annuel. Nous avons donc analysé cet ensemble de données en décomposant la variance et en faisant une analyse en composantes principales et ce à la fois sur les fonctions de consommations et les différents produits.

Puis nous avons cherché d’autres bases de données afin d’obtenir une meilleure granularité.

## Difficultés rencontrées

Nous avons été confrontés à des problèmes de compréhension lors de la recherche des articles car ils utilisaient des concepts ou des techniques que nous ne connaissions pas. Notamment pour bien les comprendre il eût fallu maîtriser les différentes techniques que nous voyons en cours de séries temporelles. De même nous avons eu quelques difficultés à comprendre exactement ce que recouvrait la consommation et l’évolution des différentes techniques utilisées pour la calculer, ainsi que les différentes manières de l’exprimer (en valeur ou en volume, en prix chaînés ou en prix courant). Néanmoins en approfondissant la question et demandant au besoin à QuantCube de nous éclairer nous avons été en mesure d’avancer.

## Quelques résultats préliminaires

# Ce qu’il reste à faire

Il nous reste à avancer dans l’analyse de la variance de la décomposition de la consommation et des corrélations entre les variables officielles et la consommation des ménages. Et ce notamment en faisant varier les périodes considérées afin de s’assurer qu’il n’y a pas eu de changements à même de biaiser notre prédiction.

## Difficultés anticipées