

## **Projet 1:**

Bienvenue chez YNOV Data! Votre première mission consiste à construire un modèle des prix immobiliers en Californie en vous basant sur les données de recensement. Ces données comportent un certain nombre de variables quantitatives telles que la population, le revenu médian et le prix médian des habitations pour chacun des “**Blocks**” recensés en Californie. Ces “**Blocks**” constituent la plus petite subdivision pour laquelle les services américains du recensement publient des données(en général, ils regroupent de 600 à 3000 personnes). Nous les appellerons “districts” dans ce qui suit.

Votre modèle devra apprendre de ces [données](#) et être capable de prédire le prix médian des habitations dans chaque district, en fonction des autres variables connues.

Pour l’instant nous allons juste nous focaliser sur l’analyse exploratoire et le nettoyage des données du dataset, le reste du travail sera fait par la suite lorsque nous aurons assimiler des notions plus approfondies sur le machine learning.

## **Projet 2:**

Des relevés minutieux ont été effectués par des data analystes de la ville de Seattle en 2016. Voici [les données](#) et [leur source](#). Cependant, ces relevés sont coûteux à obtenir, et à partir de ceux déjà réalisés, **vous voulez tenter de prédire les émissions de CO2 et la consommation totale d’énergie** de bâtiments **non destinés à l’habitation** pour lesquels elles n’ont pas encore été mesurées.