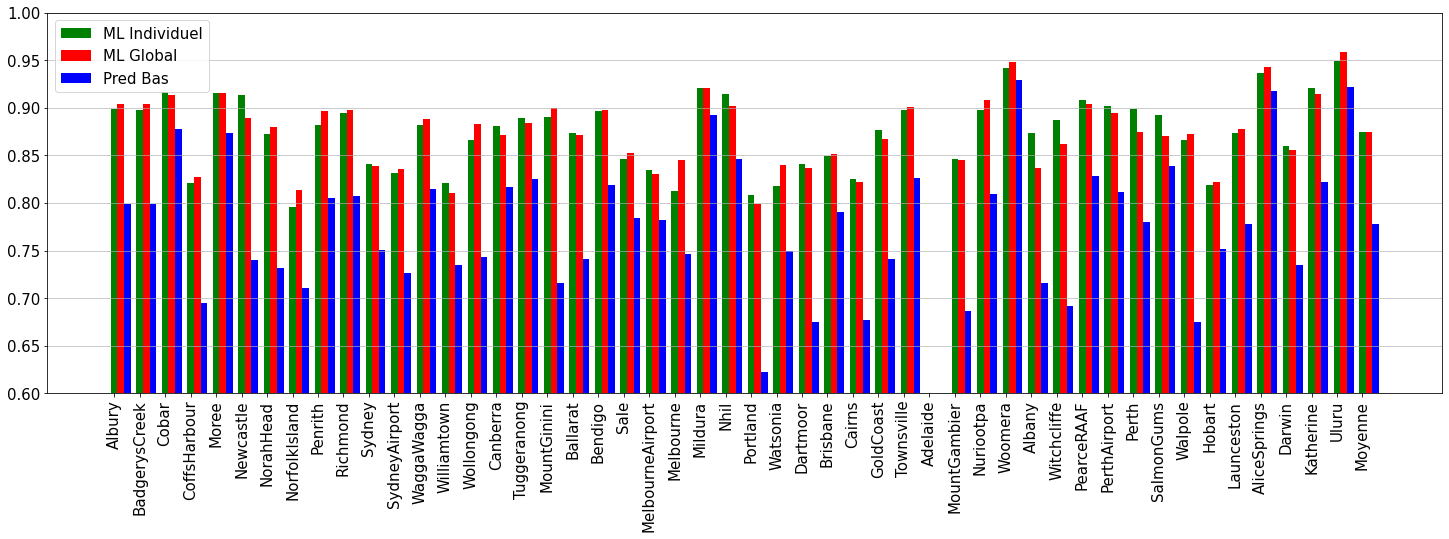
**modélisation : première partie**

Le graphique suivant montre le SCORE de 3 méthodes de prédiction :

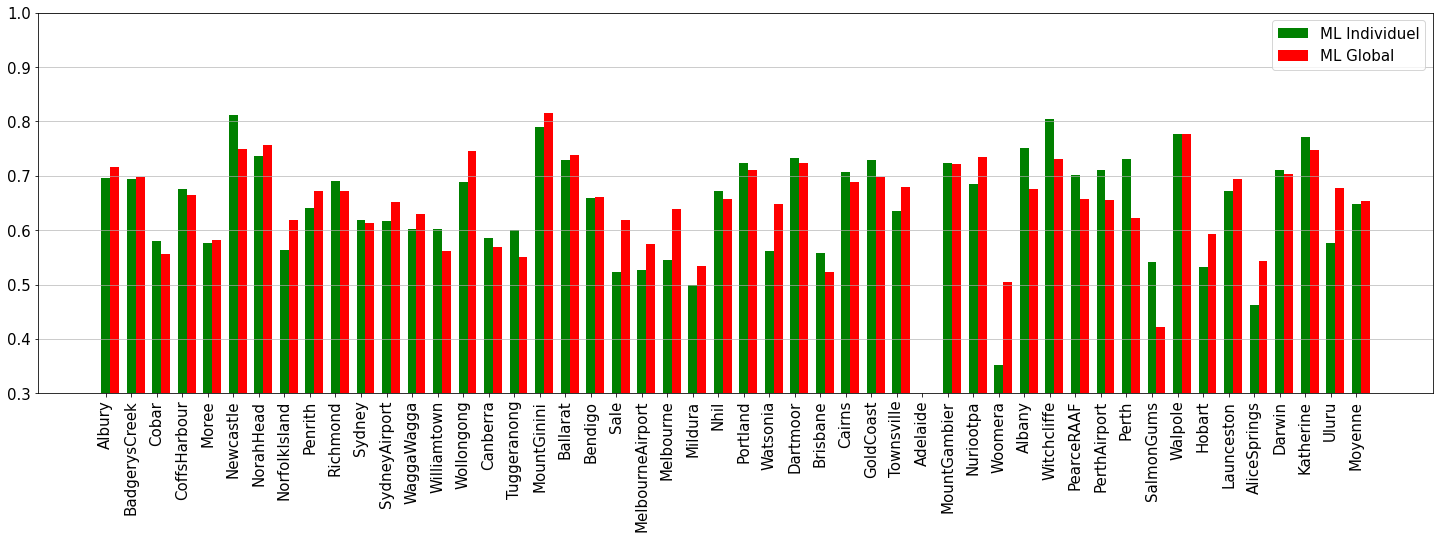
* ML Individuel : un modèle ML pour chaque ville
* ML Global : le même modèle est utilisé pour toutes les villes
* Prediction base : prédire qu'il ne pleuvra jamais.



La moyenne de la ville est légèrement meilleure pour un ML Individuel (0,875 contre 0,8746), et les deux cas montrent une amélioration par rapport à la prédiction de base (0,7782).

Le graphique suivant montre le F1 de 2 méthodes de prédiction :

* ML Individuel : un modèle ML pour chaque ville
* ML Global : le même modèle est utilisé pour toutes les villes

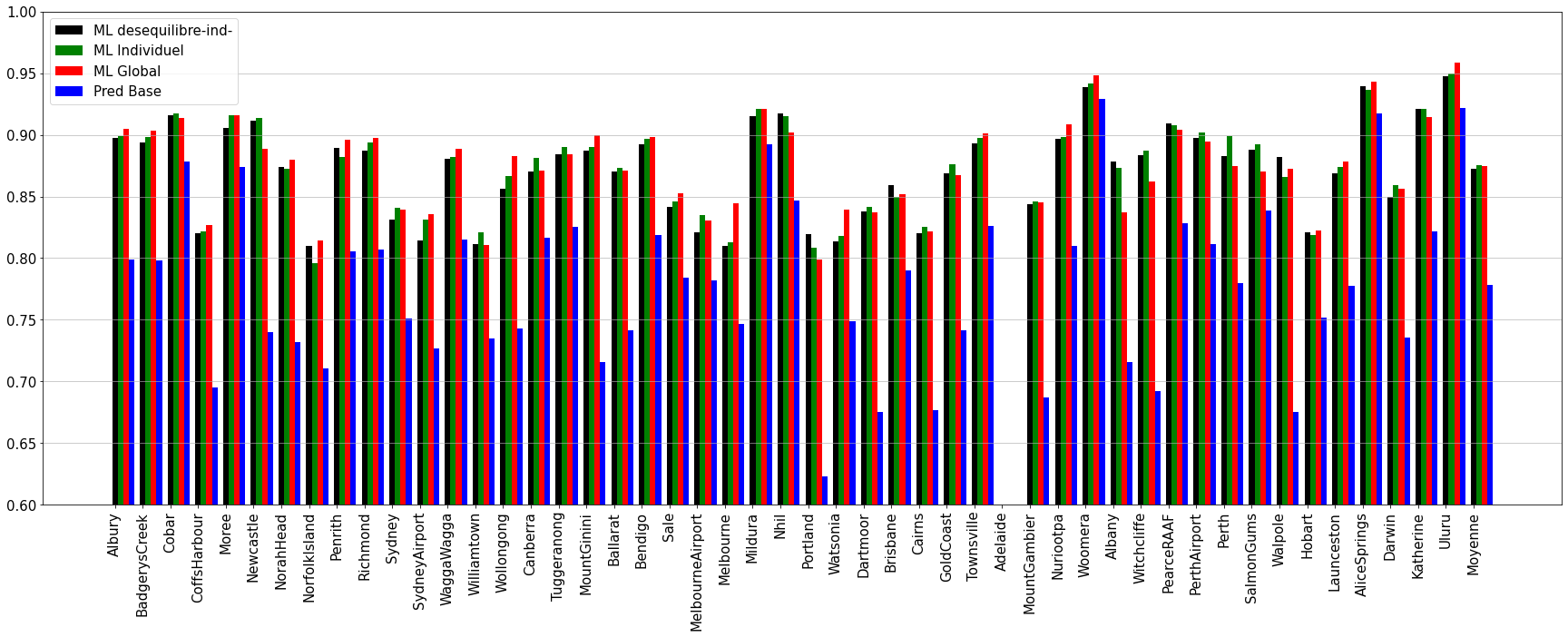


Contrairement au score, dans ce cas le modèle global (0.6537) est meilleur que le modèle individuel (0.6478).

**Ajout d'un modèle pour les bases de données déséquilibrées**

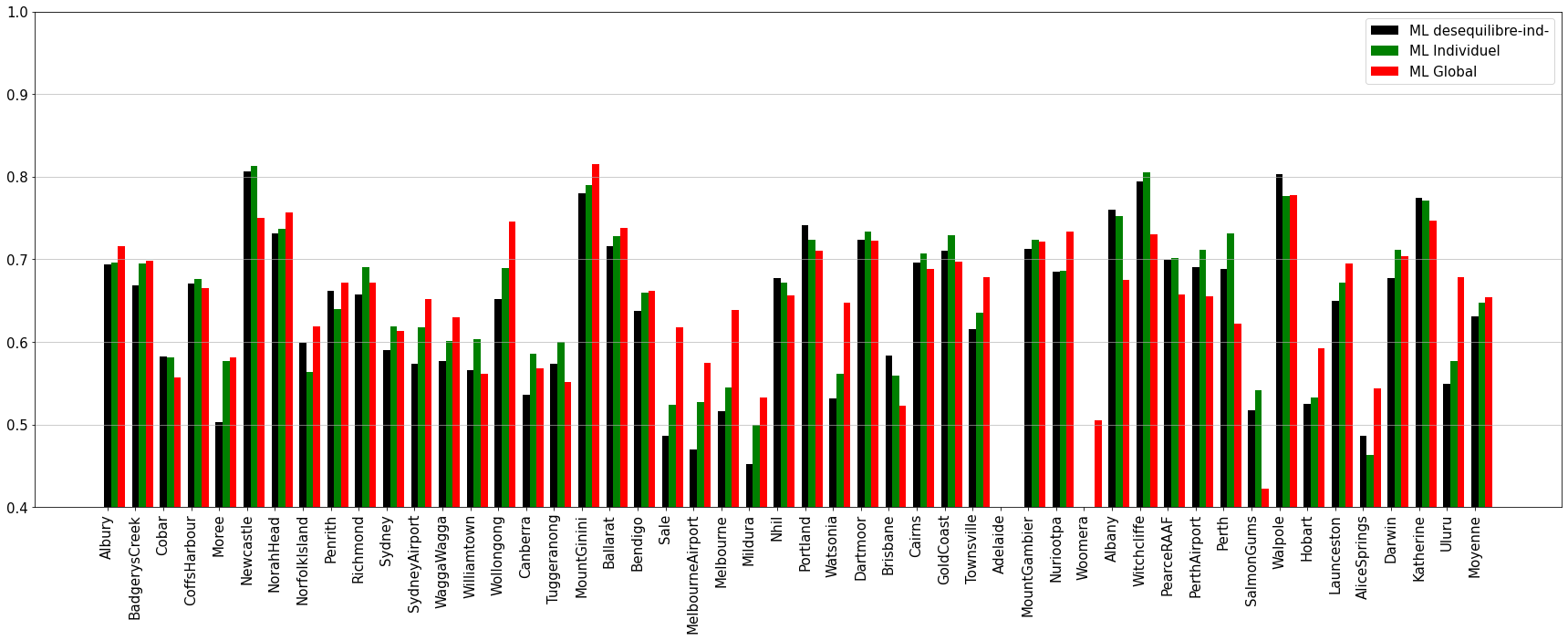
Le graphique suivant montre le SCORE de 3 méthodes de prédiction :

* ML desequilibre-ind- : un modèle de bases de données équilibrées par ville.
* ML Individuel : un modèle ML pour chaque ville
* ML Global : le même modèle est utilisé pour toutes les villes
* Prediction base : prédire qu'il ne pleuvra jamais.

il n'y a pas d'amélioration du résultat moyen.

Le graphique suivant montre le F1 de 2 méthodes de prédiction :

* ML desequilibre-ind- : un modèle de bases de données équilibrées par ville.
* ML Individuel : un modèle ML pour chaque ville
* ML Global : le même modèle est utilisé pour toutes les villes.

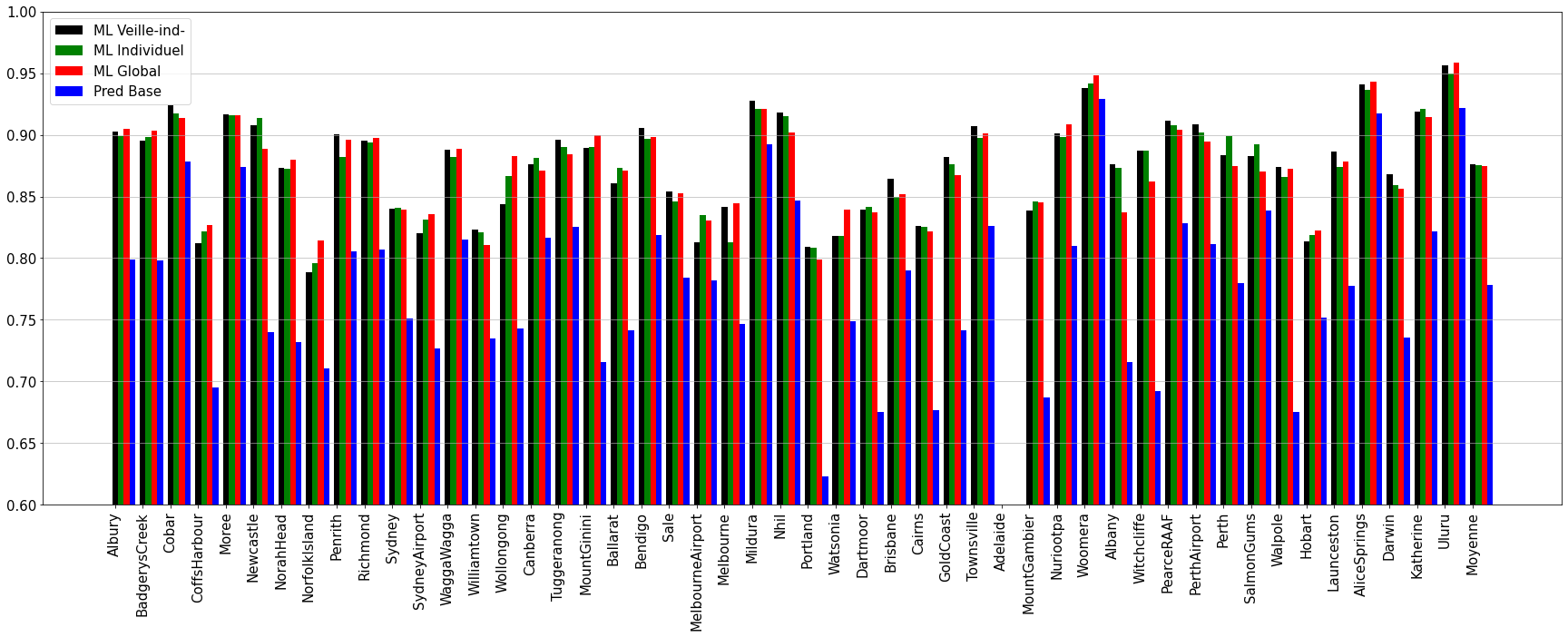


Comme pour le SCORE, aucune amélioration du score moyen n'est observée.

**Ajouter les données de la veille**

Le graphique suivant montre le SCORE de 3 méthodes de prédiction :

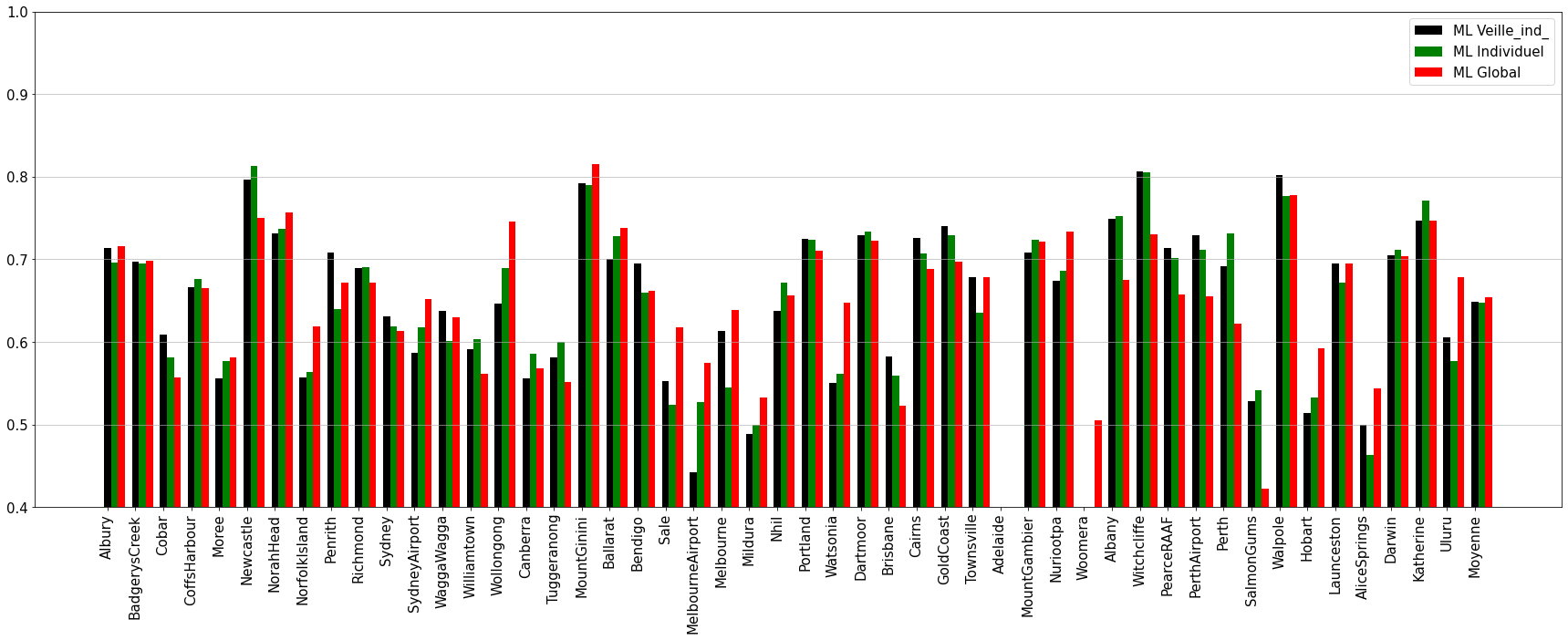
* ML veille-ind- : les données de la veille sont agrégées (duplication des variables) avec une ML par ville.
* ML Individuel : un modèle ML pour chaque ville
* ML Global : le même modèle est utilisé pour toutes les villes
* Prediction base : prédire qu'il ne pleuvra jamais.



On constate qu'avec les données de la veille, le score du modèle s'améliore légèrement par rapport à un modèle individuel (0.8760 contre 0.8750).

Le graphique suivant montre le F1 de 2 méthodes de prédiction :

* ML veille-ind- : les données de la veille sont agrégées (duplication des variables) avec une ML par ville.
* ML Individuel : un modèle ML pour chaque ville
* ML Global : le même modèle est utilisé pour toutes les villes.



Pour F1, on observe également que le modèle est légèrement meilleur qu'un modèle unique (0.6485 contre 0.6477).

**A discuter :**

* Quelle est la meilleure variable pour définir le modèle ? ou une moyenne entre les deux ?
* Chaque ville obtient de meilleurs résultats avec un modèle différent, est-il donc utile de choisir un modèle différent pour chaque ville ? Combien de modèles maximum ?