



# Projet de Big Data

Adrien Huet and Charlie Brosse



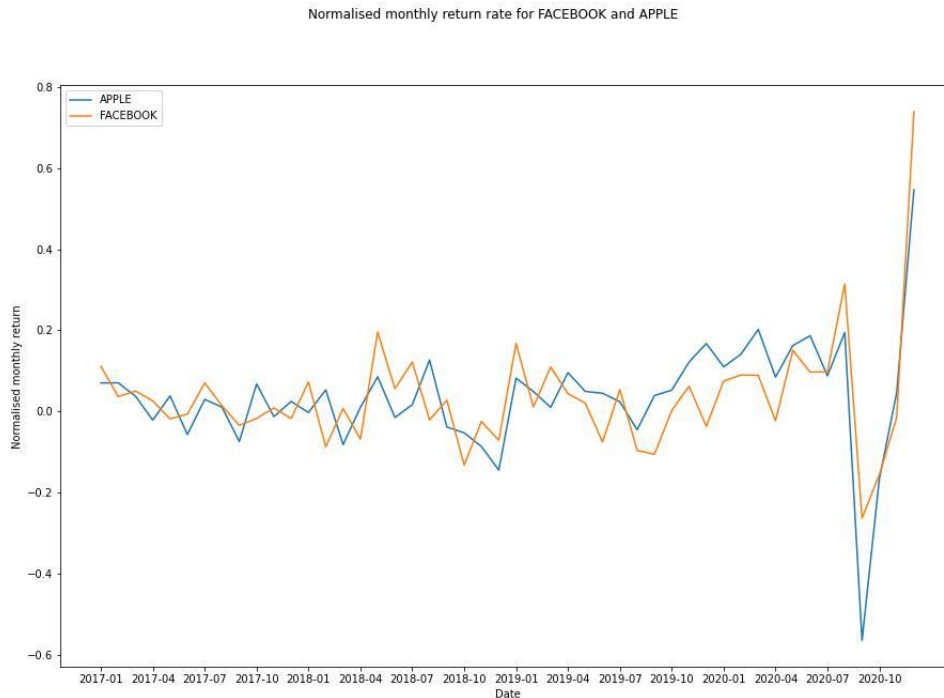
# Exploration des données et analyse

- Dataset : valeurs des actions de 7 entreprises échantillonnées par jour
  - Pour chaque journée : valeur maximale, minimale, d'ouverture et de fermeture de bourse
  - Nombre d'actions échangées
- 
- Très forte corrélation entre les colonnes
  - Pas de valeurs manquantes
  - Croissance globale de la valeur des actions (visualisation de la *daily\_evolution* et *daily return*)
  - Ajout de *Adj\_close*

# Exemple d'analyse: évolution par entreprise



# Exemple d'analyse: corrélation entre 2 entreprises



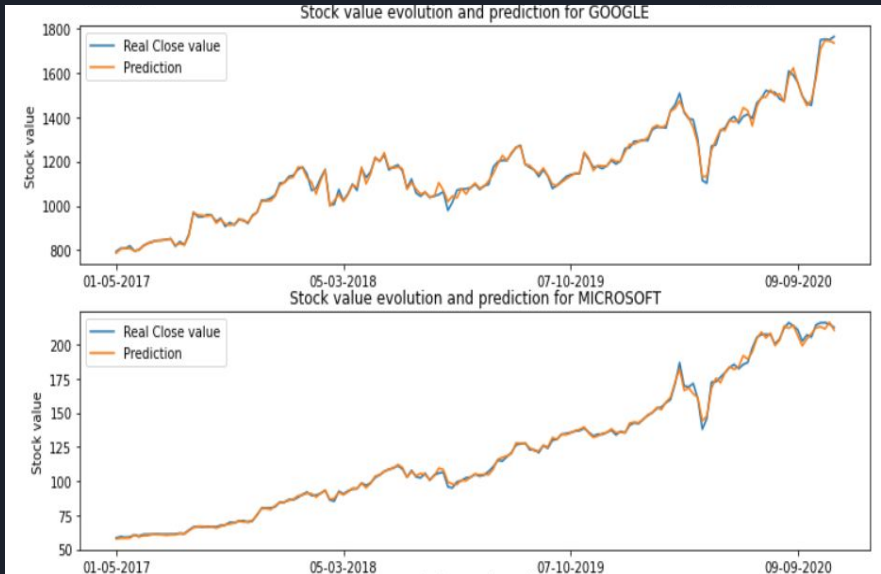
**Corrélation: 0.9**



# Choix pour la régression

- Etude de la prédiction des valeurs de fermeture *Adj Close*
- Colonnes features : *Date* et *Open*
- Conversion en *double* des valeurs, en *unixtime* pour les dates
- Fonction d'erreur : RMSE
- Etude du maximum d'erreur
- Split 80 / 20 random

# Résultats de la régression



RMSE

14.8208

GOOGLE

1.7897

MICROSOFT