**Plan pour la présentation**

**Support :**

* PowerPoint ou notebook RMarkdown

**Contenu (doit tenir en une vingtaine de minutes) :**

* Présentation de Kobe Bryant (contexte et quelques statistiques sur sa carrière)
* **I. Présentation du jeu de données et de l'objectif**
  + **Origine :** Kaggle
  + **Objectif :** Prédire le résultat d'un shot de KB (mesure via la LogLoss et comparaison avec le leaderboard)
  + **Variables dans le jeu de données initial :**
    - Information sur le shot : Type de shot, localisation, moment du match
    - Informations sur le match : date, adversaire, playoffs ou saison régulière, domicile ou extérieur
* **II. Présentation de la phase d'ajout de données :**
  + **Ajout de données extérieures** (nombreuses statistiques détaillées disponibles sur Internet par rapport à la NBA) :
    - statistiques globales sur le match de KB (temps total joués, nombre de points total, passes, rebonds,...)
    - résultats du match (victoire ou défaite, écarts de points, série en cours)
    - mais aussi le résultat des shots manquants (permet de calculer la LogLoss sans soumettre à Kaggle)
  + **Calcul de données supplémentaires :**
    - Réussite au shot de KB (depuis le début du match, sur le match précédent)
    - Dynamique de réussite sur les derniers / prochains shots tentés
    - Écart de temps entre les shots
* **III. Statistiques descriptives en utilisant le Shiny :**
  + Axé sur l'influence des différentes variables sur le pourcentage de réussite de KB au shot
  + **Exemple :** Croisement de la réusiite de KB au shot de KB avec la nature du shot, la localisation du shot, le moment du match, l'adversaire, le fait d'être en playoffs, la dynamique de shot qui précéde,..
* **IV. Trois grands types de variables explicative possibles pour nos modèles :**
  + ***Light* :** on utilise uniquement les données d'origine du Kaggle
  + ***Prévision* :** on utilise les données *Light* et on rajoute toutes les données qui ont lieu **avant** le shot d'un point de vue temporelle (pourcentage de réussite depuis le début du match, statistiques sur le match précédent, dynamique sur les derniers shots tentés,...)
  + ***Full*** : on utilise les données *Prévision* et on rajoute des informations qui ont lieu **après** le shot d'un point de vue temporelle (résultat du match, statistiques de KB sur ce match, dynamique pour les shots qui auront lieu après)
  + **En résumé :** *Light* < *Prévision* < *Full*
  + **Note :** Dans tous les cas on s'interdit d'utiliser des statistiques relatives aux nombre de points marquées ou à la réussite au shot de KB sur le match courant (exemple : pourcentage de réussite à 2 points de KB sur le match).
* **V. Présentation des différents modèles testés :**
  + **Type de modèle :** Régression logitique, forêt aléatoires, XGBoost, autres... ?
  + **Résultats des modèles** via les mesures suivantes : LogLoss, rang Kaggle associé (on garde pour la conclusion pour le suspens ?), taux de shots bien classés et AUC.
  + En plus de l'effet choix du type de modèle, analyse de l'impacts des effets suivants :
    - **Effet ajout de variables :** passage de *light* à *prévision* puis à *full*
    - **Effet paramétrage du modèle :** passage du paramétrage de base à un paramétrage plus avancé (important notamment pour XGBoost)
  + **Note :** En pratique je pense qu'il faut présenter les type de modèle les uns après les autres puis faire une conclusion qui compare les différents modèles (tableaux et graphes récapitulatifs pour comparaison de performances.
  + **Analyse rapide des variables les plus importantes** dans notre meilleure modèle avec notamment un zoom sur l'effet de la variable "Temps avant le prochain shot de KB" (uniquement dans les modèles *full*)
* **VI. Conclusion :**
  + Les rangs Kaggle de nos différents modèles !