**Résumé de ce que je vais écrire sur le PowerPoint :**

***Projet exploration de données :***

***FIFA Football***

***PROBLÉMATIQUE : recrutement des joueurs***

Nous sommes une nouvelle équipe de football et souhaitons recruter des joueurs pour un championnat (FIFA).

Pour ce fait, nous devons mettre en place quelques critères d’éligibilité des joueurs en fonction des postes pour ne pouvoir sélectionner que les meilleurs.

**Points essentiels :**

1. **Collecter les données :** Kaggle
2. **Nettoyer :** Python
3. **Structurer :** Lien, sens
4. **Analyser:** Graphes, présentation

**1.Collecte des données :**

Lien Kaggle : \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Non de la table: FIFA\_DATA

(+ une petite description de la table et les colonnes)

**2.Nettoyage des données :**

\_ suppression des colonnes vides ou inutiles

\_ suppression des lignes vides

\_ remplissage des données vides (façon aléatoire)

\_ normaliser les monnaies

\_ conversion des unités

- Masse

- Taille

\_ gérer les types de chaque colonne

\_ substitution des données (les postes)

\_ création d’une colonne groupe de postes (défenseurs, ailiers, etc…)

\_ modification des types des données dans ces colonnes :

-Wage : float

-Height : float

-Weight :float

-Crossing : int

-International reputation : int

-body type : corriger les valeurs erronées

\_ Table nettoyée : fifa\_data\_clean (#en excel pour simplifier l’étude et en csv lors de la vraie utilisation.

**2.Structurer :**

#En général, la structure veut dire la création de la BDD dans le cas où il y a une BDD, dans notre cas c’est bcp plus les bons liens entre les colonnes de notre table, c a d que chaque donnée dans la table doit être utilisable, ajoute un sens à notre étude,... .

**2.Analyser + Résoudre notre problématique :**

2.1.Notre analyse portera sur les différents grands **postes** :

- Attaque (ST, CF)

- Milieu de terrain (CM, RM, LM)

- Defense (GK, CB, RB, LB)

- Ailiers (LW, RW)

2.2. Notre analyse va se baser sur ces **sous-problématiques** :

\_ rapport **âge** et **IMC**

\_ **critères physiques** **d’eligibilité** à chaque **poste:**

sur quels aspects majeures sont évalués les joueurs par postes / critères techniques d’éligibilité à chaque poste ( on pourrait faire une moyenne des compétences et voir à l’aide d’une distributions quels sont ceux qui sont vitaux pour un joueur en fonction de son poste)

· On le fait d’abord pour les groupes de postes ( gardiens, défenseurs , ailiers, attaquants, milieux de terrain)

· Ensuite les familles de poste

- Le poste en lui-même

- Le pieds de préférence par poste

\_ par poste on évaluera la **compétence** des joueurs par préférence de **pieds** et par **utilisation de leurs pieds faibles** et on fera un lien avec la **réputation internationale = #Graphe à trois variable(bien)**

\_ on voit les pays qui fournissent le plus de joueurs par groupe de postes pour aller les chercher dans ces pays **#une bonne idée sur laquelle on doit parler au call.**

\_ voir par poste le moyenne des salaires

\_ Voir le budget de chaque équipe pour les différents grands postes

\_ **#Je propose aussi** :Voir le **marking** par rapport à la **réputation** (**#En vrai on peut faire un graphe développé pour la réputation vu qu’elle combine beaucoup de facteurs (les buts, money, pieds, … ).**

**\_ #Je propose :** Voir les pays qui retournent souvent dans un club, c'est -à -dire le pays dans lequel un club choisit toujours ou souvent les joueurs, celà nous permettra de voir quel pays donne beaucoup de joueurs.

#Important : **On peut créer des graphes par nous même, c’est-à-dire on va créer les graphes avec les colonnes de notre table mais aussi on peut créer des graphes avec des informations hors la table comme le dernier graphe que je propose.**