**Aide à la décision paris de course de F1**

Avec les données ci-dessous, on peut modéliser la probabilité de gagner la course (Colonne Win) avec une régression simple par exemple et on peut prédire la position exacte du pilote (Colonne Position) avec par exemple un algo de classification non-supervisée où chaque position (1,2,3…20) correspond à un groupe ou autre modèle de type régression logistique multinationale.

Jeu d’entraînement :

| Année | Pilote | Nationalité | Âge | Circuit | Météo | Position | Win |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2021 | A | FR | 23 | Monaco | Pluie | 1 | Y |
| 2021 | B | SP | 27 | Circuit international de Sakhir | Soleil | 2 | N |
| 2021 | C | EN | 23 | J[eddah Corniche Circuit](https://www.google.com/search?client=opera-gx&sxsrf=APq-WBuWEXSqxUA1uJhImvQoXdhb8M94lA:1648063462628&q=Jeddah+Corniche+Circuit&stick=H4sIAAAAAAAAAONgVuLVT9c3NMzJsSwyKTM1m8DC2MDKuIhV3Cs1JSUxQ8E5vygvMzkjVcE5syi5NLMEAOi5D0ExAAAA&sa=X&ved=2ahUKEwjJpP_p-tz2AhUv4YUKHYBIDqQQnLUDegQIARAF) | Soleil | 1 | Y |
| 2021 | D | DE | 28 | Zandvoort | Vent | 3 | N |
| 2020 | A | FR | 23 | Silverstone | Soleil | 5 | N |

Jeu de validation :

[….]

Jeu de test :

| Année | Pilote | Nationalité | Âge | Circuit | Météo | Position | Win |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 2022 | A | FR | 23 | Monaco | Soleil | 1 | 0.99 |
| 2022 | B | SP | 27 | Circuit international de Sakhir | Soleil | 3 | 0.8 |
| 2022 | C | EN | 23 | J[eddah Corniche Circuit](https://www.google.com/search?client=opera-gx&sxsrf=APq-WBuWEXSqxUA1uJhImvQoXdhb8M94lA:1648063462628&q=Jeddah+Corniche+Circuit&stick=H4sIAAAAAAAAAONgVuLVT9c3NMzJsSwyKTM1m8DC2MDKuIhV3Cs1JSUxQ8E5vygvMzkjVcE5syi5NLMEAOi5D0ExAAAA&sa=X&ved=2ahUKEwjJpP_p-tz2AhUv4YUKHYBIDqQQnLUDegQIARAF) | Vent | 1 | 0.83 |
| 2022 | D | DE | 28 | Zandvoort | Vent/Pluie | 6 | 0.6 |
| 2022 | E | LU | 19 | Silverstone | Pluie | 4 | 0.4 |

Une fois le predict réalisé chaque pilote pour chaque course (date, circuit, météo) on stock les résultats dans BQ.

Exemple de report d’aide à la décision :



Afficher ce résultat en fonction de filtre sur Data Studio.

Par exemple, si je veux connaître la prédiction des pilotes sur le prochain circuit je sélectionne ce circuit (filtre du dash) et ensuite le résultat de la prédiction pour ce pilote est affiché.   
Si je veux savoir pour un circuit en particulier, ça me dit quelles sont les positions attendues de chaque pilote. En // je peux avoir des graphs qui pour ce circuit me donne le top 3 des pilotes depuis 5 ans. Peut me donner par exemple si c’est un jour pluvieux les meilleurs pilotes lorsqu’il pleut etc.

Sources intéressantes :

* ML : <https://blog.link-value.fr/prediction-globale-et-estimations-locales-1266ca4c3960>
* BQ : [https://cloud.google.com/bigquery-ml/docs/export-model-tutorial?h](https://cloud.google.com/bigquery-ml/docs/export-model-tutorial?hl)l=fr
* BQ to GCP : https://cloud.google.com/bigquery/docs/visualize-data-studio

Chaque année on a un ensemble de courses de formule 1 qui forment un championnat. Notre but est de prédire pour l’année 2022 le classement des pilotes à une course avec sa probabilité associée lors de la prédiction; on cherchera aussi à prédire qui va gagner le championnat 2022.

De plus, à la fin de chaque course les 10 premiers coureurs gagnent des points. Je crois par contre que les points ne sont pas correctement calculés.

Est ce que les temps aux courses de qualification sont importants pour la prédiction de la course ?

Après lecture des différents jeux de données, idée de jeu de données sur lequel entraîner nos modèles :

|  |  |
| --- | --- |
| Année du championnat |  |
| Circuit |  |
| Date de la course |  |
| Météo lors de la course | On a pas l’info peut être la récupérer ailleurs |
| Pilote (Nom Prénom) |  |
| Nationalité du pilote |  |
| Age du pilote au moment de la course |  |
| Nom de l’écurie |  |
| Nationalité de l’écurie | Est ce que ça vaut vrmt le coup ? je me suis dis que la nationalité différente du pilote peut jouer mais j’en sais rien |
| Numéro de la voiture | Est ce que le pilote change de voiture ? si oui ça peut jouer ? |
| Position au départ de la course |  |
| Numéro du tour de la course |  |
| Temps du tour fait par le pilote |  |
| Numéro de l’arrêt au stand |  |
| Temps de l’arrêt au stand |  |
| Position finale |  |
| Victoire (oui/non) |  |
| Points à la fin de la course |  |

Idée dataviz :

Une page dédiée au pilote : si possible avec des graphs pour que ce soit ludique

Sélection obligatoire sur le pilote

Onglet 1 :

Nom, prénom, date de naissance, nationalité, nom de l’écurie à laquelle il appartient

Des statistiques comme le nombre de fois où il a gagné une course, le nombre de championnat gagné, le nombre de fois où il a été sur le podium, le nombre de points qu’il a gagné au cours de sa période

Onglet 2 :

Sélection sur une course (sur le circuit et sur l’année)

Des statistiques associées, si il a gagné, le nombre de points qu’il a eu, sa position finale, le temps de sa course (par tour? Et en entier?), sa position à chaque tour? , les temps d’arrêt au stand en moyenne avec le nombre d’arrêt qu’il a fait, sa position finale

Etude préliminaire des données :

**Status** → pas à prendre car nécessite que le course soit terminé

**Sprint\_result** → pas à prendre car nécessite que le course soit terminé

**Seasons** → pas nécessaire on a déjà l’info dans Races.csv

**Results** → c’est notre table centrale

**Races** → à merger sur results

**Qualifying** → pas compris l’utilité peut être est-ce les différentes courses de qualification pour déterminer la grid

**Pit\_stops** → pas à prendre car nécessite que la course soit terminé

(ca peut être intéressant pour faire du online prediction pendant que la course est en cours mais pas l’objectif ici)

**Lap\_times** → pas à prendre car nécessite que la course soit terminé

**Drivers** → à merger sur results

**Driver\_standings** → à merger sur results

**Constructors** → à merger sur results

**Constructors\_standings** → à merger sur results

**Constructors\_results** → pareil avec moins de données que **Constructors\_standings** donc pas à prendre.

**Circuits** → à merger sur results

Finalement on garde :

