



المدرسة الوطنية للعلوم التطبيقية بتطوان  
ⵜⴰⴳⴷⴰⵏⵜ ⵜⴰⵎⴰⵏⵏⵜ ⵜⴰⵎⴰⵏⵏⵜ ⵜⴰⵎⴰⵏⵏⵜ  
ÉCOLE NATIONALE DES SCIENCES APPLIQUÉES  
DE TÉTOUAN



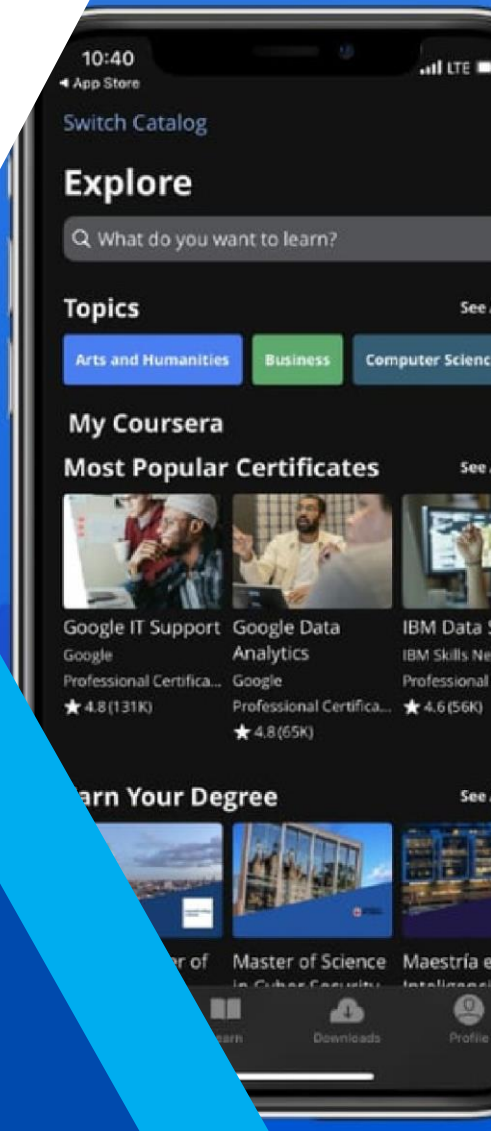
# RAPPORT DU PROJET VISUALISATION :

Analytiques des cours  
Coursera

Filière d'Ingénieur CI1 :  
Big Data et Intelligence Artificielle

Préparé par :

Ben Ayad Douaae  
Amekran Ouadiaa  
Akhrif Salma



# Table DE MATIÈRE

Introduction

Structure du Dashboard

Données et indicateurs clés

Présentation générale du Dashboard

Dashboard

Démo interactive

Problèmes rencontrés et solutions

Conclusion

# Introduction

L'apprentissage en ligne a connu une expansion significative ces dernières années, offrant aux apprenants du monde entier une multitude de cours accessibles à distance. Coursera, en tant que plateforme d'apprentissage en ligne renommée, propose une diversité de cours dans divers domaines.

Le jeu de données associé à Coursera semble constituer une compilation d'informations riches sur ces cours. Chaque ligne de ce jeu de données représente vraisemblablement un cours distinct, fournissant une vue détaillée de ses caractéristiques. Ces attributs incluent des éléments tels que le titre du cours, la note moyenne attribuée par les apprenants, le niveau de difficulté, l'horaire des cours, les objectifs d'apprentissage, les détails de l'instructeur, l'institution proposant le cours, les mots-clés associés, l'URL du cours, la durée estimée pour sa complétion, ainsi que le nombre d'avis laissés par les participants. Cette diversité d'informations permet d'explorer et d'analyser les tendances, la pertinence et la popularité des cours disponibles sur Coursera.

# Structure du Dashboard

## 1. Sections et composant principaux :

**Titre Principal** : « Analyse des Cours Coursera ».

Notre tableau de bord se compose de deux pages principales : la première page, « Datasse », qui contient des informations générales et des statistiques de nos données, et la deuxième page, « Tableau de bord », qui contient les graphiques utilisés dans nos données.

- **Dataset** :

Cette page comprend trois éléments : « À propos des données », qui résume nos données ; « Données », qui contient nos données ; et « Structure », qui contient les indicateurs, leur nombre, leur type, ainsi qu'une vision générale des valeurs de chaque indicateur.

- **Tableau de bord** :

Cette page comprend trois boîtes de valeurs et huit graphiques :

- **Boîtes de valeurs** :

**Total des cours** : Affiche le nombre total de cours disponibles sur la plateforme Coursera.

**Durée moyenne pour terminer les cours** : Présente la durée moyenne nécessaire pour compléter les cours, ce qui peut aider les apprenants à planifier leur emploi du temps.

**Moyenne des évaluations** : Indique la note moyenne attribuée aux cours, ce qui peut être utile pour évaluer la qualité générale des cours disponibles.

- **Graphiques** :

**Durée moyenne passée dans les cours par niveau** : Permet de comparer la durée moyenne passée dans les cours selon le niveau de difficulté.

**Fréquence des mots-clés** : Affiche la fréquence d'utilisation des mots-clés dans les cours, ce qui peut révéler les tendances et les sujets populaires.

**Compétences les plus enseignées sur Coursera** : Présente les compétences les plus couramment enseignées sur la plateforme, offrant un aperçu des domaines d'étude les plus demandés.

**Comparaison de la Popularité** : Compare la popularité des différents cours en termes de nombre d'inscriptions ou d'achèvements.

**Durée vs Note** : Analyse la relation entre la durée des cours et les notes attribuées.

**Note moyenne par catégorie et niveau** : Affiche la note moyenne des cours selon leur catégorie et leur niveau de difficulté.

**Diagramme de dispersion** : Nombre d'avis vs Durée pour compléter : Illustration graphique du nombre d'avis par rapport à la durée nécessaire pour compléter les cours.

**Le nombre total des cours** : Présente le nombre total de cours disponibles sur la plateforme.

Ces sections et composants principaux offrent une vue complète et organisée des données et des analyses liées aux cours Coursera, permettant aux utilisateurs d'explorer, de comprendre et de tirer des insights des données présentées.

## 2. Organisation de l'information :

Le tableau de bord organise les informations relatives à l'analyse des cours sur Coursera, le jeu de données nommé " CourseraDataset-Clean" est extrait d'un fichier Excel intitulé "dataset after cleaning". Ce fichier regroupe une diversité d'informations incluant les catégories de cours, les niveaux de difficulté, les durées moyennes de complétion, ainsi que les évaluations moyennes, parmi d'autres données pertinentes. Ce jeu de données constitue une source précieuse permettant de concevoir un tableau de bord interactif et bien structuré.

# Données et indicateurs clés

## 1. Sources de données :

Pour trouver le jeu de données "Coursera Dataset", nous avons effectué des recherches approfondies sur divers sites web. Parmi les plateformes que nous avons consultées, citons notamment Kaggle, Data Search et data.gov. Finalement, nous avons localisé ce jeu de données sur Kaggle, dont le propriétaire est "Elvin Rustamov", étant donné que nous avons identifié que le sujet de Coursera serait bénéfique pour nous en tant qu'ingénieurs spécialisés dans le big data et l'IA, et étant particulièrement intéressés par l'apprentissage en ligne, l'exploration visuelle de ce jeu de données nous permettra, ainsi qu'à nos collègues, d'obtenir une vision d'ensemble de cette plateforme.

Voici le lien où on a trouvé ce dataset :

<https://www.kaggle.com/datasets/elvinrustam/coursera-dataset>

## 2. Principaux indicateurs :

Les principaux indicateurs que nous avons utilisés pour élaborer notre tableau de bord, nous avons inclus :

**Level** (la catégorie) : Cette colonne indique le niveau de difficulté ou de complexité du cours. Il pourrait catégoriser les cours comme débutant, intermédiaire ou avancé, par exemple.

Cet indicateur nous sera précieux pour déterminer les niveaux des cours proposés par la plateforme Coursera. Il permettra d'obtenir une compréhension claire et détaillée de la diversité des niveaux de difficulté présents dans les cours disponibles sur cette plateforme d'apprentissage en ligne. En intégrant cette information à notre tableau de bord, nous pourrons mieux appréhender la variété des offres éducatives et adapter nos stratégies en conséquence.

**Rating** : La colonne de notation contient probablement la note moyenne du cours, telle que fournie par les utilisateurs ayant suivi la cour. Les notes sont souvent données sur une échelle, par exemple de 1 à 5 étoiles.

Cela va nous aider de prendre une idée sur la qualité des cours fournis par cette plateforme.

**What you will learn** : Cette colonne décrit probablement les objectifs d'apprentissage ou les sujets abordés dans le cours. Elle donne un résumé des connaissances ou compétences que les participants peuvent espérer acquérir.

**Keyword** : Cette colonne peut contenir des mots-clés ou des tags associés au cours, qui peuvent aider les utilisateurs à rechercher des cours pertinents en fonction de thèmes ou de sujets spécifiques.

**Duration to complete (Approx.)** : Cette colonne spécifie le temps approximatif nécessaire pour terminer le cours. Il est donné en termes d'heures.

**Number of Reviews** : Cette colonne contient le nombre d'avis ou de notations soumis par les utilisateurs ayant suivi le cours. Elle donne une indication de la popularité du cours et du niveau de satisfaction des utilisateurs.

**Skill gain** : Cette colonne peut détailler les compétences spécifiques que les participants acquerront à la fin du cours.

# Données et indicateurs clés

Le tableau de bord Coursera Analytics vise à fournir une analyse approfondie des données des cours disponibles sur la plateforme Coursera. Il offre aux utilisateurs la possibilité d'explorer et de comprendre divers aspects des cours, tels que la durée moyenne, les évaluations, les compétences enseignées, et bien plus encore. Ce tableau de bord interactif permet aux utilisateurs de visualiser les données de manière intuitive et de tirer des insights significatifs pour informer les décisions stratégiques dans le domaine de l'éducation en ligne.

## 1. Description du Dashboard :

Le tableau de bord "Coursera course Analytics" comprend plusieurs onglets, chacun offrant une fonctionnalité spécifique pour explorer les données :

L'onglet "Dataset" fournit un aperçu général des données de cours de Coursera, y compris les caractéristiques de chaque cours et un aperçu de la structure des données.

L'onglet "Tableau de bord" est le cœur du système, offrant une gamme de visualisations interactives telles que des graphiques sur la durée moyenne des cours par niveau, la fréquence des mots-clés, les compétences les plus enseignées, et plus encore. Les utilisateurs peuvent également filtrer les données par niveau, catégorie, etc.

En outre, le tableau de bord est conçu avec une interface conviviale et intuitive, permettant aux utilisateurs de naviguer facilement entre les différentes fonctionnalités et visualisations.

## 2. Fonctionnalités principales:

Le tableau de bord propose plusieurs fonctionnalités pour permettre aux utilisateurs d'explorer et d'analyser les données sur les cours Coursera de manière approfondie :

**Affichage des données brutes :** Dans l'onglet "Dataset", les utilisateurs peuvent accéder aux données brutes sous forme de tableau. Une section descriptive fournit également un aperçu des principales caractéristiques du jeu de données, telles que les variables disponibles et leur signification.

**Visualisations interactives :** L'onglet "Tableau de bord" offre une variété de visualisations interactives, notamment des graphiques montrant la durée moyenne passée dans les cours par niveau, la fréquence des mots-clés les plus courants, les compétences les plus enseignées sur Coursera, etc. Ces visualisations permettent aux utilisateurs de détecter rapidement les tendances et les modèles dans les données.

**Filtrage des données :** Les utilisateurs ont la possibilité de filtrer les données par niveau, catégorie et note. Par exemple, ils peuvent sélectionner un niveau ou une catégorie pour affiner leur analyse.



**Boîtes de valeur :** Le tableau de bord présente également des boîtes de valeur qui affichent des statistiques clés telles que le nombre total de cours, la durée moyenne pour terminer les cours et la moyenne des évaluations. Ces boîtes de valeur offrent un aperçu rapide des principales métriques du jeu de données.

En combinant ces fonctionnalités, le tableau de bord offre aux utilisateurs un moyen puissant d'explorer et d'analyser les données sur les cours Coursera, les aidant à tirer des insights précieux pour la prise de décision et la planification stratégique.

### **1. Utilisateurs cibles:**

Le tableau de bord cible principalement les professionnels de l'éducation, les chercheurs en apprentissage en ligne et les responsables de programme qui souhaitent comprendre les tendances et les caractéristiques des cours Coursera. Il est également utile pour les apprenants potentiels qui cherchent à explorer les offres de cours disponibles sur la plateforme et à prendre des décisions éclairées.

### **2. Contexte et problématique:**

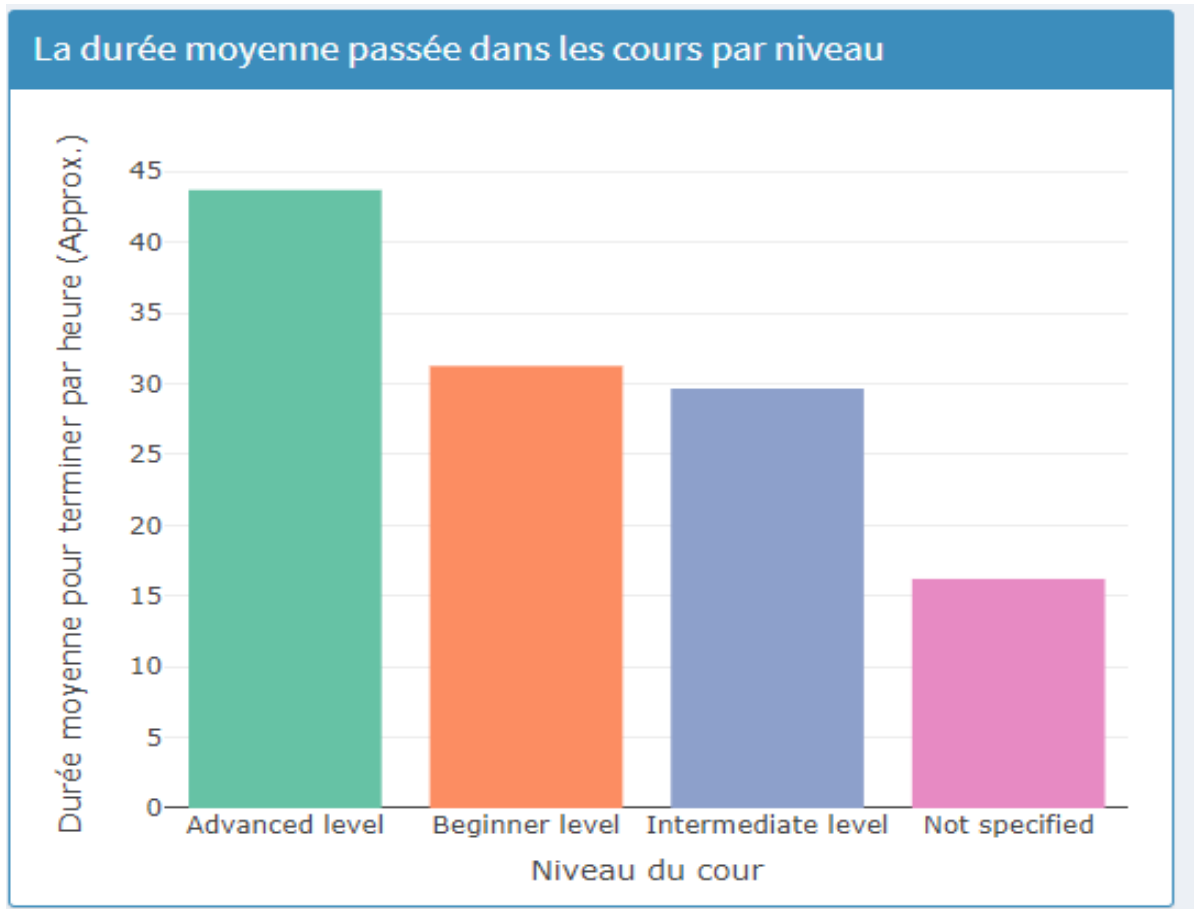
Le développement de ce tableau de bord fait suite à un besoin croissant de comprendre les dynamiques des cours en ligne, en particulier sur la plateforme Coursera. Avec une multitude de cours disponibles, il devient essentiel de pouvoir analyser les données de manière efficace pour identifier les tendances, les préférences des apprenants et les domaines d'intérêt. Le tableau de bord vise à répondre à cette problématique en fournissant des outils interactifs pour explorer et analyser les données de manière approfondie.

### **3. les sources de données:**

Les informations intégrées dans le tableau de bord sont extraites d'un jeu de données exhaustif, comprenant une gamme étendue d'informations telles que les catégories de cours, les niveaux de difficulté, les durées moyennes de complétion, les évaluations, et d'autres métriques associées aux cours disponibles sur Coursera. Ces données servent de pilier pour alimenter les graphiques interactifs et les indicateurs clés, offrant ainsi une représentation fidèle et dynamique des tendances et des performances des cours proposés sur la plateforme Coursera. Ce processus garantit une analyse approfondie et une compréhension enrichie des dynamiques de l'éducation en ligne, permettant aux utilisateurs de prendre des décisions éclairées et de maximiser leur expérience d'apprentissage.

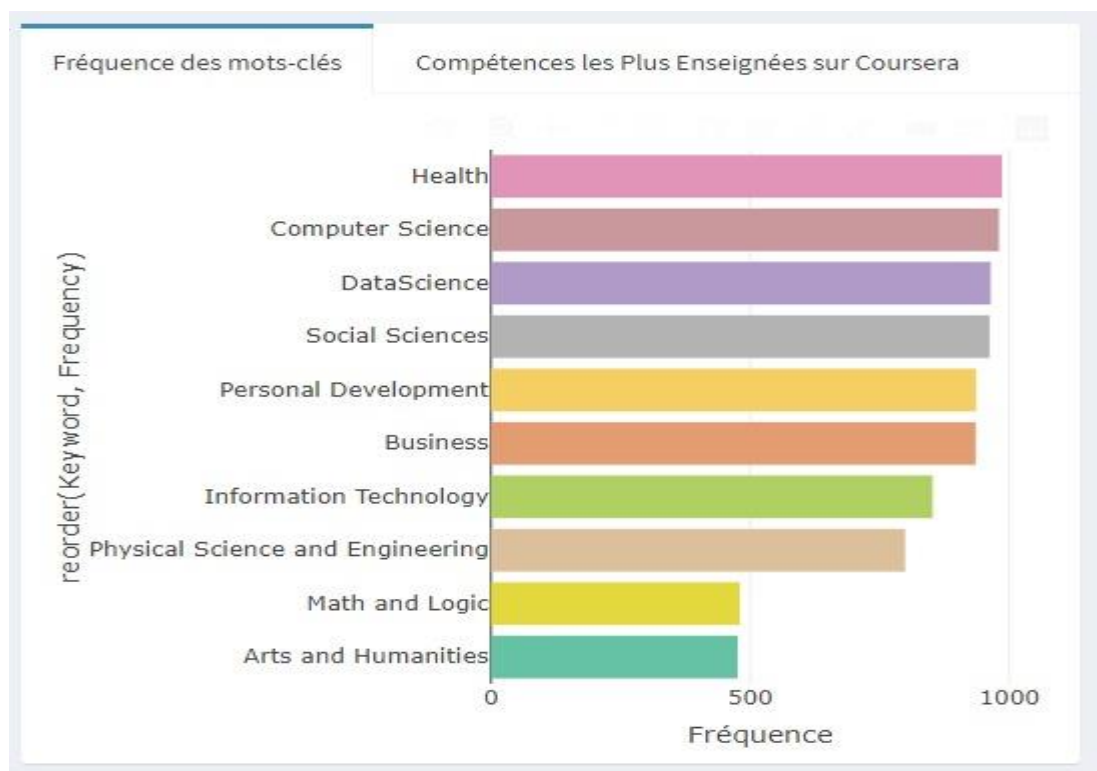
# Démo interactive

## 1. Figure1 : La durée moyenne passée dans les cours par niveau :



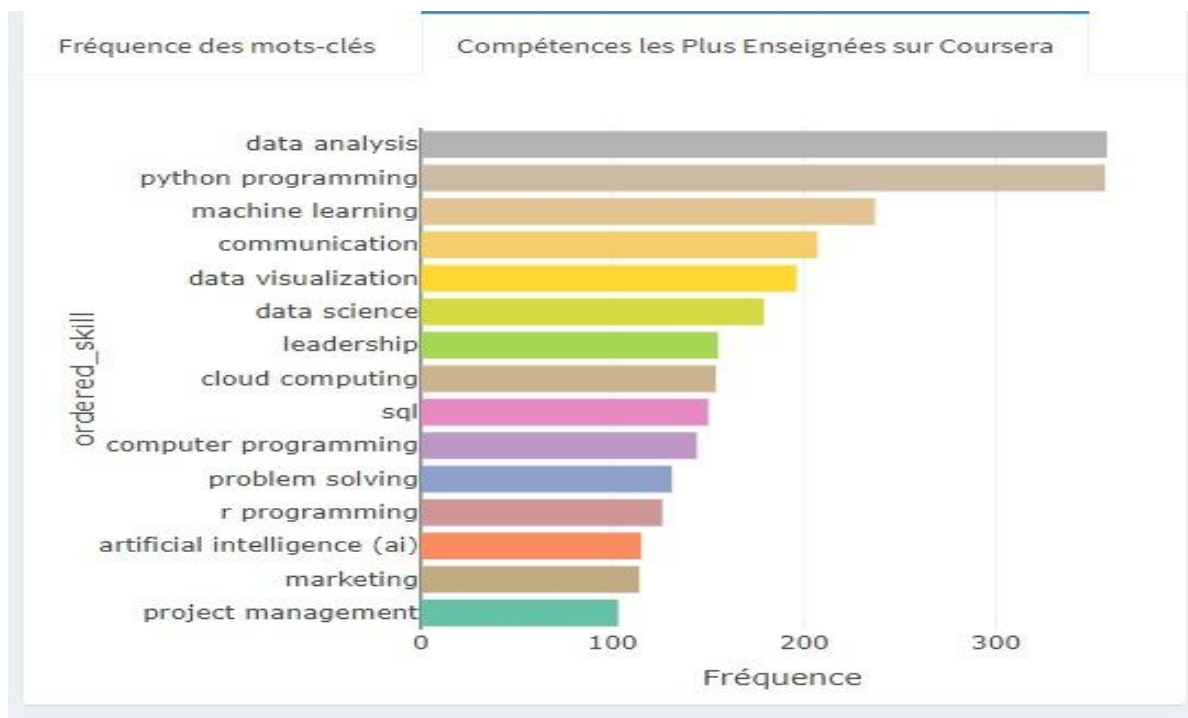
Dans cet histogramme, il est clair que les cours de niveau "Advanced" nécessitent considérablement plus de temps pour être complétés, avec une moyenne de 43,69 heures. Les cours de niveau "Beginner" viennent ensuite avec une moyenne de 31,25 heures, suivis par les cours de niveau "Intermediate" avec une moyenne de 29,65 heures. En revanche, les cours de niveau "Not Specified" affichent une moyenne notablement plus basse, avec seulement 16,20 heures en moyenne. Ces résultats soulignent que les apprenants consacrent beaucoup plus de temps aux cours de niveau "Advanced" sur Coursera par rapport aux autres niveaux.

## 2. Figure 2 : Fréquence des mots-clés :



L'analyse de ce graphe révèle que la santé, le commerce et l'informatique sont les plus populaires, attirant ainsi un grand nombre d'étudiants. Cette préférence peut s'expliquer par la forte demande d'emplois dans ces secteurs, les opportunités de carrière lucratives et la perception générale de leur pertinence dans le monde moderne. D'un autre côté, les mathématiques et la logique, les arts et lettres, ainsi que les sciences physiques et l'ingénierie sont moins prisés, ce qui pourrait refléter un manque d'intérêt perçu ou une idée fausse selon laquelle ils sont moins pertinents sur le marché du travail actuel. Il est essentiel de souligner que la valeur et l'importance de ces domaines moins populaires ne doivent pas être sous-estimées, car ils servent souvent de fondements à de nombreuses industries et disciplines. En résumé, bien que la préférence des étudiants penche vers certains domaines, il est crucial de reconnaître la diversité des opportunités et la contribution significative de chaque domaine d'étude à différents secteurs.

### 3. Figure 3 : Compétences les Plus Enseignées sur Coursera :



Ce graphe représente la fréquence des compétences les plus enseignée dans notre ensemble de données de cours Coursera. Chaque barre horizontale représente un mot-clé spécifique, et la longueur de chaque barre indique le nombre de fois que ce mot-clé apparaît dans notre ensemble de données.

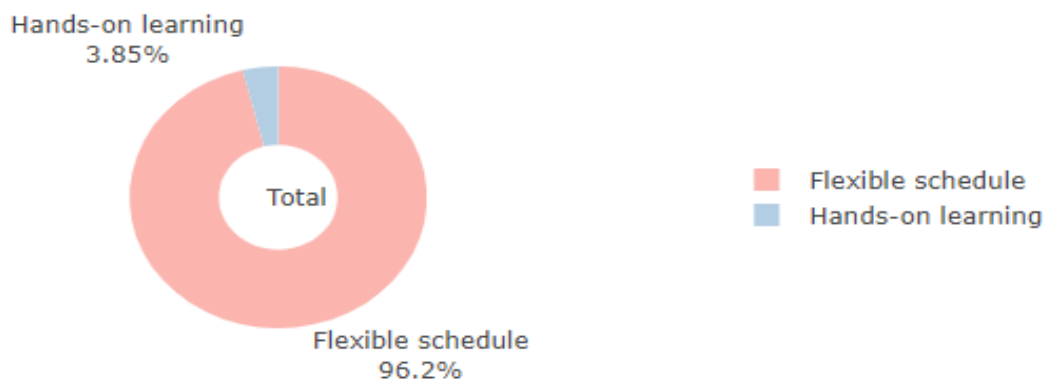
Les mots-clés sont affichés sur l'axe vertical (y), tandis que la fréquence est représentée sur l'axe horizontal (x). Plus la barre est longue, plus ce mot-clé est fréquent dans nos données.

De ce graphe il est notable que les compétences les plus enseignées sur Coursera sont centrées autour de "data analysis", "python programming" et "machine learning". Ces domaines, tous liés au domaine des sciences des données, soulignent une tendance marquée vers des compétences hautement recherchées dans le secteur de la technologie et de l'analyse de données. Cette concentration met en évidence l'importance croissante de ces compétences dans le paysage professionnel actuel, où la maîtrise de l'analyse de données et de la programmation en Python, ainsi que la compréhension des principes du machine learning, sont devenues des atouts cruciaux.

conclusion:Le graphe représente la fréquence des compétences les plus enseignées dans l'ensemble de données de cours Coursera. Il montre une concentration significative autour de mots-clés tels que "data analysis", "python programming" et "machine learning", soulignant ainsi une forte tendance vers des compétences liées aux sciences des données. Ces compétences sont devenues cruciales dans le paysage professionnel actuel, reflétant la demande croissante pour l'analyse de données, la programmation en Python et la compréhension du machine learning.

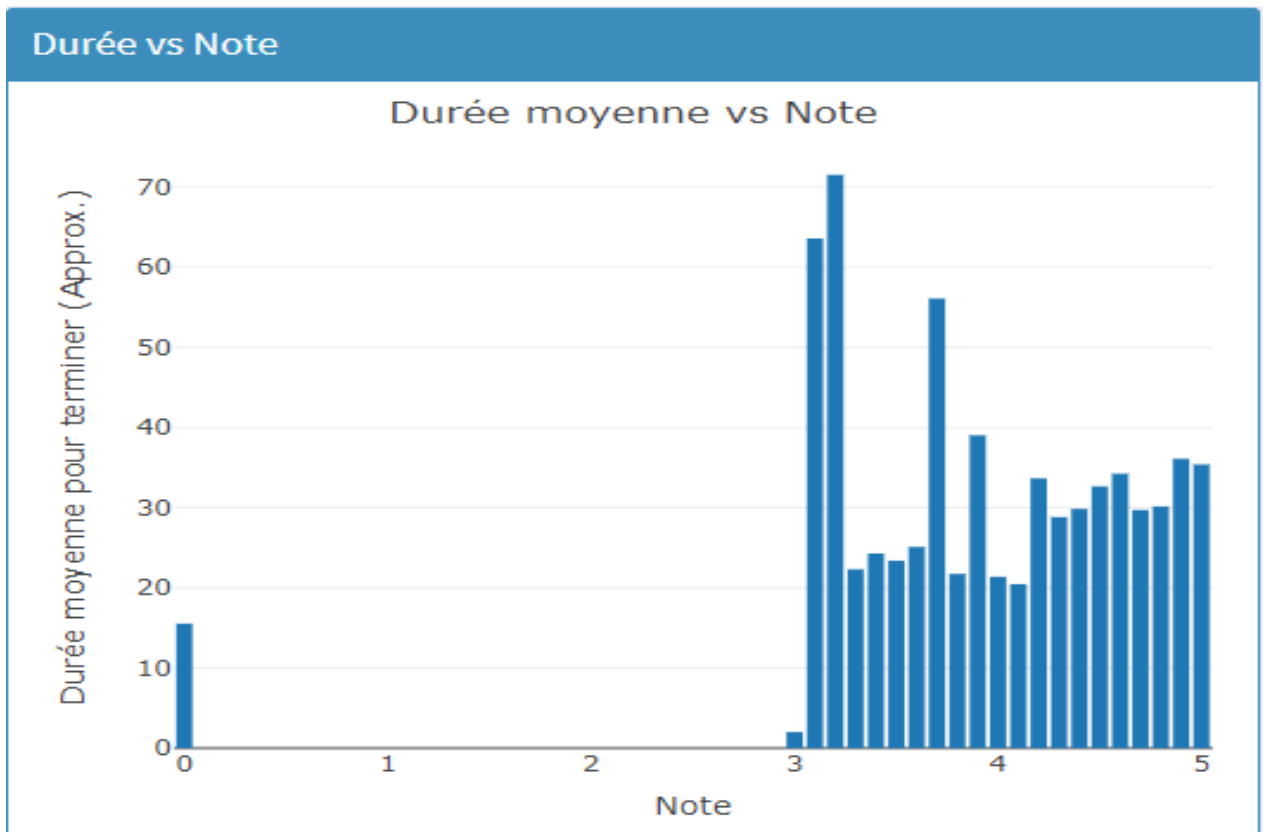
#### 4. Figure 4 : Comparaison de la Popularité :

## Comparaison de la Popularité



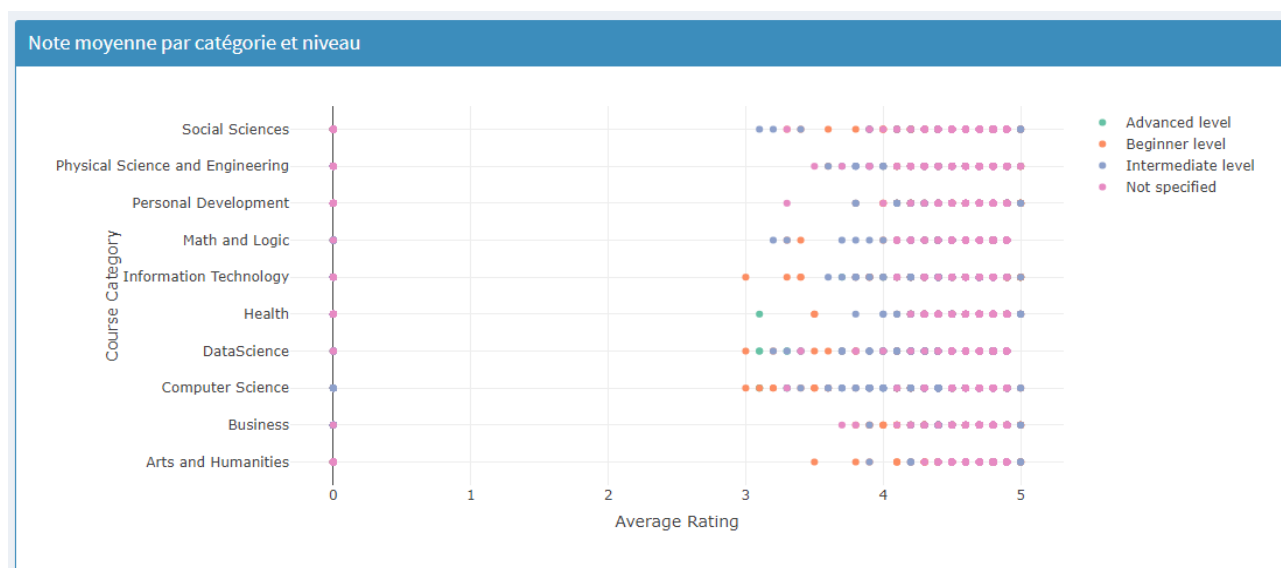
Le graphique circulaire met en lumière une nette préférence des étudiants pour un horaire flexible (96,15 %) par rapport à l'apprentissage pratique (3,85 %). Cette préférence peut découler de la nécessité d'harmoniser les études avec les engagements personnels et professionnels. Elle témoigne également d'une transition vers des méthodes d'apprentissage plus autonomes et auto-dirigées. L'accessibilité et la diversité des ressources en ligne favorisent cette flexibilité. Comprendre ces préférences est crucial pour élaborer des programmes éducatifs adaptés aux besoins et aux réalités diversifiées des étudiants.

### 5. Figure 5 : Durée vs Note :



Cet histogramme illustre la relation entre la note et la durée nécessaire pour terminer un cours. On observe que lorsque la durée moyenne pour terminer le cours est modérée, la note attribuée à ce cours est généralement meilleure. Cette observation souligne l'importance de gérer efficacement son temps d'étude et de maintenir une progression régulière dans l'achèvement des tâches académiques pour obtenir de meilleurs résultats d'apprentissage. Cependant, il est également essentiel de reconnaître que d'autres facteurs peuvent également influencer la performance académique, et que cette corrélation pourrait varier en fonction de divers contextes et situations d'apprentissage.

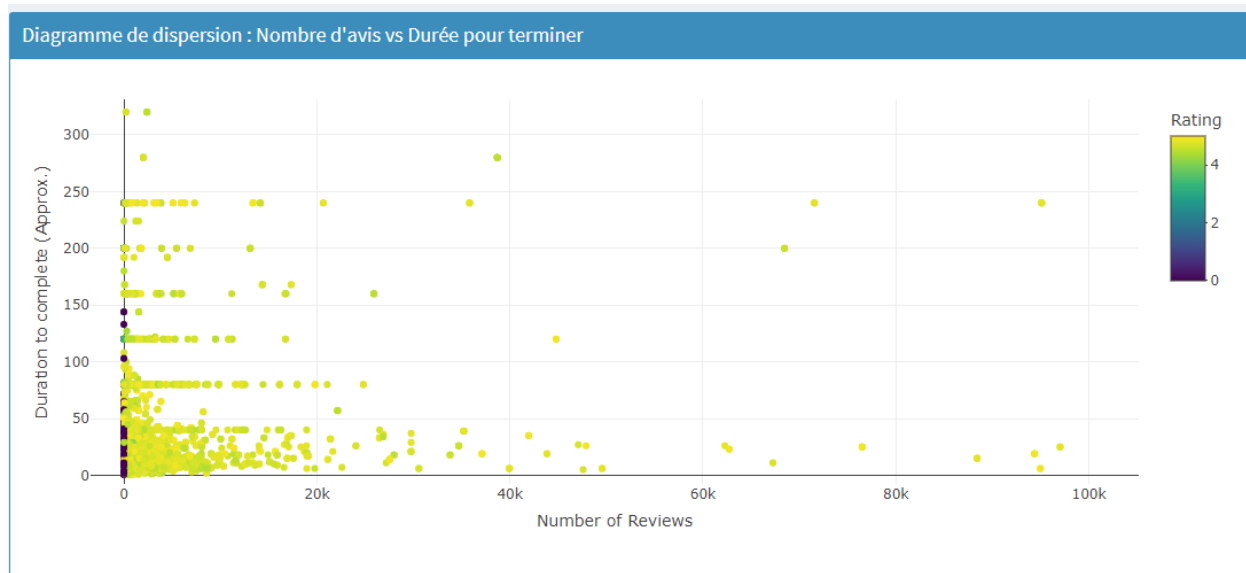
**6. Figure 6 : Note moyenne par catégorie et niveau :**



Ce graphe est un nuage de points interactif qui utilise la bibliothèque Plotly. Il représente visuellement les relations entre les données de différentes variables. L'axe des x du graphique représente les évaluations moyennes des cours (Rating), tandis que l'axe des y représente les catégories de cours (Keyword) et La couleur des points est déterminée par une troisième variable, le niveau des cours (Level).

L'analyse de ce graphique révèle que la majorité des cours disponibles sur la plateforme Coursera obtiennent des évaluations de 3 ou plus, indiquant ainsi une qualité générale bien perçue par les apprenants. Il est intéressant de noter que le niveau "not specified" a obtenu le rating maximum, suggérant que la qualité des cours ne semble pas être spécifiquement liée à un niveau particulier. Cette constatation suggère que Coursera offre des opportunités d'apprentissage de haute qualité adaptées à divers niveaux de compétence, ce qui pourrait être particulièrement bénéfique pour une audience variée en quête de formation en ligne.

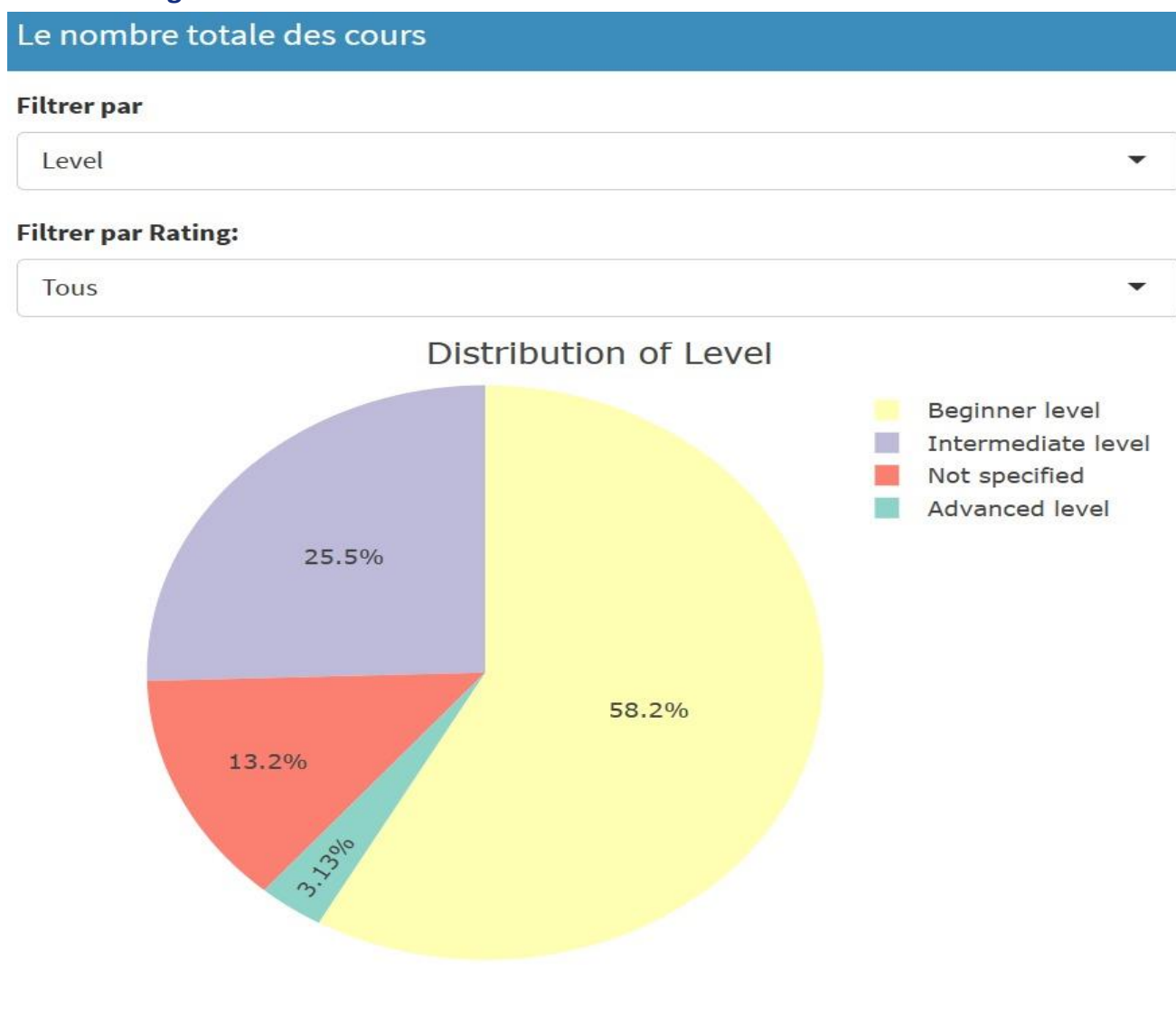
## 7. Figure 7 : Diagramme de dispersion : Nombre d'avis vs Durée pour terminer :



Nous observons sur ce graphique que la majorité des avis concernant les cours de Coursera sont positifs. De plus, une tendance intéressante se dégage : lorsque la durée des cours diminue, le nombre d'avis semble augmenter. Cette corrélation suggère que les cours plus courts pourraient susciter un plus grand intérêt et générer davantage de retours de la part des apprenants. Ces constats soulignent l'importance de la qualité des cours offerts par Coursera, ainsi que la pertinence de la durée des cours dans l'engagement et la satisfaction des apprenants.

En résumé, l'analyse du graphique indique une satisfaction générale des apprenants à l'égard des cours de Coursera, avec une tendance positive des avis. De manière intéressante, une corrélation suggère que des cours plus courts pourraient susciter davantage d'intérêt, ce qui souligne l'importance de la qualité et de la durée des cours dans l'engagement des apprenants. Ces conclusions pourraient influencer des ajustements stratégiques pour maintenir la satisfaction des apprenants et optimiser l'expérience d'apprentissage sur la plateforme.

## 8. Figure 8 : Le nombre totale des cours :





Le diagramme circulaire montre la répartition du nombre de cours par niveau. Voici une analyse de la répartition des cours par niveau :

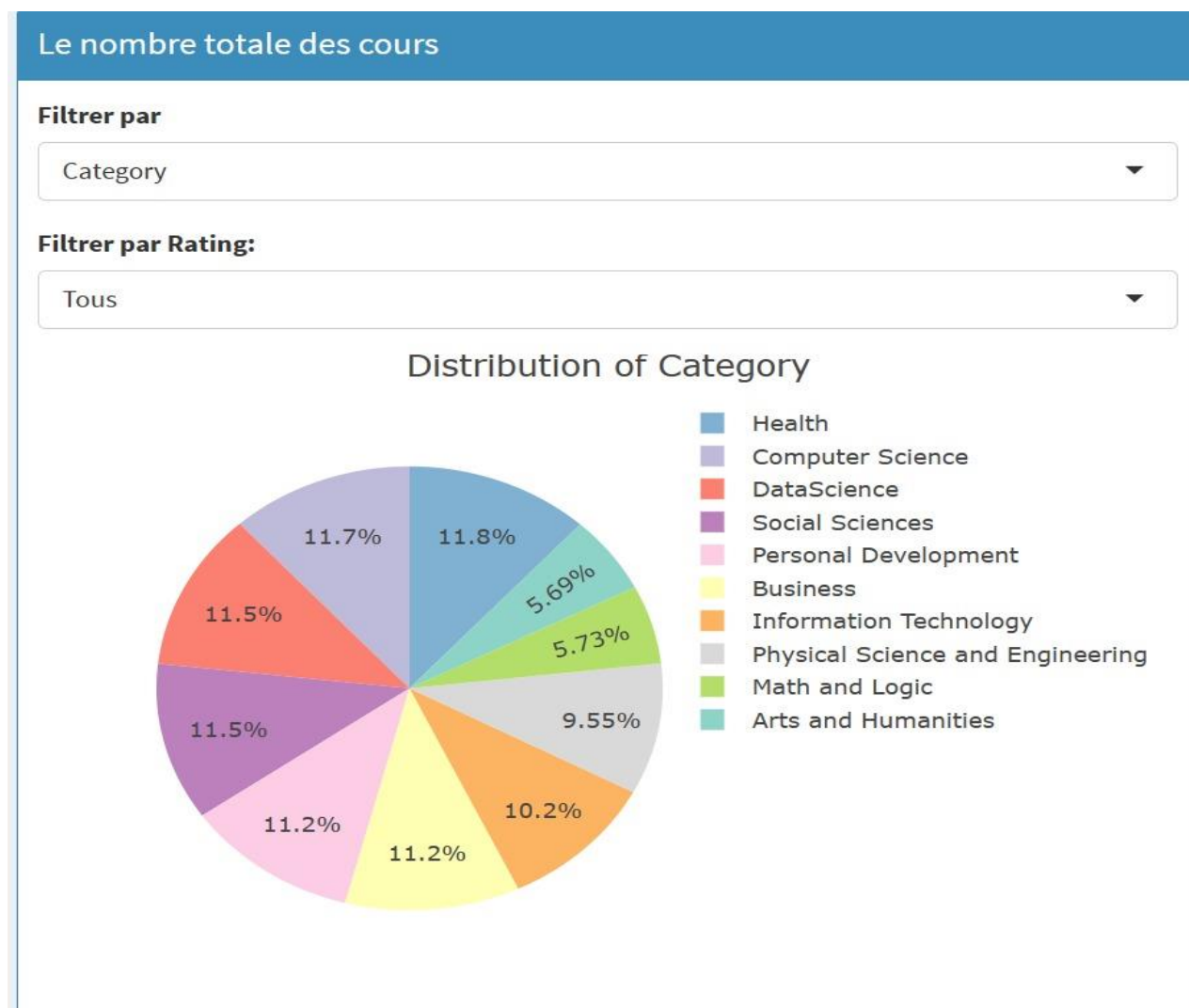
Niveau débutant : 25,5 %

Niveau intermédiaire : 58,2 %

Non spécifié : 13,2 %

Niveau avancé : 3,13 %

On constate qu'il y a plus de cours de niveau intermédiaire que de tout autre niveau. Cela pourrait indiquer une demande pour des cours qui fournissent des connaissances de base avant de passer à des sujets plus complexes.



Ce graph montre la répartition des cours offerts dans un programme éducatif particulier, par catégorie. Voici ce qu'il indique :

La catégorie la plus importante, avec 11,8 %, est l'informatique.

La santé la suit de près avec 11,7 %.

Les sciences sociales, le développement personnel et les affaires représentent chacun environ 11,5 % des cours proposés.

La technologie de l'information arrive à 10,2 %.

Les mathématiques et la logique, les sciences physiques et l'ingénierie, ainsi que les arts et les lettres représentent entre 9,55 % et 5,69 % des cours proposés.

L'informatique apparaît comme la catégorie la plus importante, suivie de près par la santé. Cela peut refléter la demande croissante pour les compétences informatiques et les connaissances en santé dans le monde moderne.

Les sciences sociales, le développement personnel et les affaires sont également bien représentés, ce qui suggère un intérêt généralisé pour ces domaines. La technologie de l'information occupe une place significative, ce qui n'est pas surprenant compte tenu de son importance dans de nombreux secteurs.

Les mathématiques et la logique, les sciences physiques et l'ingénierie, ainsi que les arts et les lettres, bien que représentant une part légèrement moins importante, restent néanmoins des domaines d'étude significatifs.

En conclusion, cette répartition des cours reflète la diversité des intérêts et des besoins éducatifs, mettant en lumière l'importance de différentes disciplines dans l'éducation contemporaine.

# Problèmes rencontrés et solutions

Problèmes	Solutions
L'absence de formation dans le langage R.	Nous avons suivi des tutoriels sur YouTube afin de mieux appréhender le langage R. Ces sessions de formation en ligne se sont avérées cruciales pour approfondir nos compétences et notre compréhension de l'utilisation pratique du langage R. En explorant diverses ressources, nous avons renforcé notre maîtrise du langage, améliorant ainsi notre aptitude à l'appliquer de manière effective dans des contextes réels.
Recherche de données.	Après une exploration minutieuse de diverses plateformes en ligne telles que Kaggle, Data Search et data.gov, nous avons finalement localisé le jeu de données recherché sur Kaggle.
Choisir les graphes utiles pour cette visualisation	On a fait des recherches sur les types des graphes disponibles et leur utilité pour visualiser notre donnée.

# Conclusion

En conclusion de notre projet de visualisation de données basé sur la dataset de Coursera, nous avons réussi à présenter de manière efficace diverses informations sur les cours disponibles. À travers des graphiques interactifs, nous avons pu mettre en évidence les tendances significatives, offrant une vision approfondie de la distribution des évaluations, des durées de cours et des niveaux d'enseignement. Ces différentes visualisations ont grandement amélioré notre compréhension de la plateforme Coursera, soulignant des points clés tels que la prédominance des évaluations positives et la diversité des niveaux de compétence. En parallèle, cette expérience a renforcé nos compétences en visualisation de données tout en élargissant nos perspectives sur l'apprentissage en ligne. Grâce à cette immersion dans les données de Coursera, nous avons acquis des connaissances précieuses qui enrichissent notre compréhension de la plateforme et de ses dynamiques.