Projet data

1. Analyse descriptive : (df.describe())

#### runtime\_minutes **(durée des films)**

* **Moyenne** : 110.5 min | **Écart-type** : ≈ 20 min
* Les films ont des **durées assez homogènes**.
* L’écart-type modéré montre une **faible variabilité**, confirmée par les quartiles resserrés entre 96 et 120 min.

➡️ **Conclusion** : Variable **homogène**, la majorité des films ont une durée standard proche de la moyenne.

#### ✅ movie\_averageRating **(note moyenne)**

* **Moyenne** : 6.40 | **Écart-type** : ≈ 1.02
* La **dispersion faible** montre que les films sont notés de manière **assez cohérente**.
* Peu de films avec des notes très éloignées de la moyenne.

➡️ **Conclusion** : Variable **très homogène**, les notes sont globalement similaires entre les films.

#### ✅ movie\_numberOfVotes **(nombre de votes)**

* **Moyenne** : 134k | **Écart-type** : 213k
* Forte variabilité : certains films ont très peu de votes, d'autres plusieurs millions.
* L’écart-type **très élevé** par rapport à la moyenne est un indicateur clair d’**hétérogénéité extrême**.

➡️ **Conclusion** : Variable **fortement hétérogène**, fortement influencée par quelques films très populaires.

#### ✅ approval\_Index

* **Moyenne** : 5.03 | **Écart-type** : 1.36
* Variabilité **modérée**, les films se concentrent autour d’un indice moyen.
* Pas d’extrêmes marqués, ce qui traduit une **bonne cohérence globale**.

➡️ **Conclusion** : Variable **relativement homogène**, avec une légère variabilité acceptable.

#### ✅ Production budget $

* **Moyenne** : 37.7 M$ | **Écart-type** : 44.5 M$
* Très **grande dispersion** des budgets : certains films sont produits avec des millions, d’autres avec plusieurs centaines de millions.
* L’écart-type supérieur à la moyenne indique une **hétérogénéité forte**.

➡️ **Conclusion** : Variable **très hétérogène**, avec des écarts de budget considérables entre films.

#### ✅ Domestic gross $

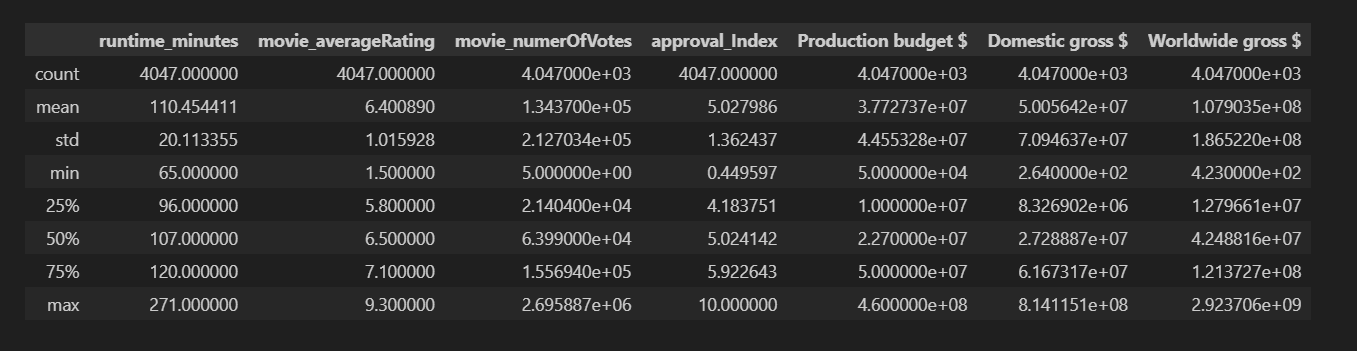
* **Moyenne** : 50 M$ | **Écart-type** : 70.9 M$
* Variabilité **très élevée** : les recettes internes varient fortement d’un film à l’autre.
* Présence de films à succès qui tirent la moyenne vers le haut.

➡️ **Conclusion** : Variable **hautement hétérogène**, les performances domestiques sont très disparates.

#### ✅ Worldwide gross $

* **Moyenne** : 108 M$ | **Écart-type** : 186 M$
* C’est la variable la plus dispersée : certaines recettes mondiales explosent les compteurs.
* Les différences de revenus mondiaux sont **extrêmement importantes**.

➡️ **Conclusion** : Variable **extrêmement hétérogène**, dominée par une minorité de blockbusters mondiaux.



2-distributions :

**1. Runtime (minutes)**

* **Distribution** : La durée des films varie de 65 à 271 minutes, avec une médiane de 107 minutes et une moyenne de 110.45 minutes. Cette proximité entre la médiane et la moyenne suggère une **distribution relativement symétrique** autour de la médiane. Cependant, quelques films plus longs tirent légèrement la moyenne vers des valeurs plus élevées.
* **Asymétrie** : L’asymétrie positive (Skewness = 1.50) montre que la distribution est **légèrement décalée vers la droite**, ce qui signifie que la majorité des films ont une durée proche de 107 minutes, mais quelques films plus longs augmentent la moyenne.
* **Outliers** : 124 films sont considérés comme des valeurs aberrantes (au-dessus de 156 minutes), ce qui suggère qu'il existe quelques films particulièrement longs qui se distinguent de la majorité des films plus courts.
* **Conclusion** : En somme, la **tendance générale** des films montre une durée modérée autour de 107 minutes, mais la présence de films **extrêmement longs** influence légèrement la moyenne et génère une petite **asymétrie vers des durées plus longues**.

**2. Production budget ($)**

* **Distribution** : Le budget de production des films varie de 50 000 $ à 460 000 000 $. La médiane de 22 700 000 $ est bien inférieure à la moyenne de 37 727 367 $, ce qui indique une **minorité de films à budget élevé** tirant la moyenne vers le haut. La plupart des films ont un budget bien plus modeste.
* **Asymétrie** : L’asymétrie positive élevée (Skewness = 2.48) montre que la **distribution est fortement biaisée** vers les films à budget plus élevé. Ce biais est alimenté par quelques films à budgets exceptionnellement élevés.
* **Outliers** : 270 films sont considérés comme des valeurs aberrantes (au-dessus de 110 000 000 $), ce qui signifie que les films ayant des budgets très élevés sont relativement rares, mais ont une forte influence sur la distribution.
* **Conclusion** : La **tendance générale** montre une large concentration de films à budgets modérés, mais une petite proportion de films à **très gros budgets** tire la moyenne vers des valeurs plus élevées et crée une **asymétrie marquée**.

**3. Domestic gross ($)**

* **Distribution** : Les recettes domestiques des films varient de 264 $ à 814 115 070 $, avec une médiane de 27 288 872 $ et une moyenne de 50 056 416 $. Cela indique que la majorité des films ont des revenus domestiques plus faibles, mais que quelques films à gros succès influencent fortement la moyenne.
* **Asymétrie** : L’asymétrie positive (Skewness = 3.75) et l'aplatissement élevé (Kurtosis = 22.11) confirment que la distribution est fortement biaisée vers les films à succès, avec quelques films générant des recettes domestiques exceptionnellement élevées.
* **Outliers** : 315 films sont considérés comme des valeurs aberrantes, principalement au-dessus de 141 692 580 $, illustrant l'impact des blockbusters et autres films à gros succès au box-office.
* **Conclusion** : La **tendance générale** montre que la plupart des films génèrent des revenus domestiques relativement modestes, mais quelques films **à très gros succès** génèrent des recettes bien supérieures à la moyenne, ce qui introduit une **forte asymétrie**.

**4. Worldwide gross ($)**

* **Distribution** : Les recettes mondiales des films varient de 423 $ à 2 923 706 026 $, avec une médiane de 42 488 161 $ et une moyenne de 107 903 473 $. L’écart entre la médiane et la moyenne suggère l’influence de films avec des recettes mondiales exceptionnelles, particulièrement au sommet de la distribution.
* **Asymétrie** : L’asymétrie très positive (Skewness = 4.64) et l’aplatissement extrême (Kurtosis = 35.21) indiquent que la **majorité des films génèrent des recettes mondiales modérées**, mais quelques films **extrêmement rentables** créent un biais important vers des valeurs élevées.
* **Outliers** : 392 films sont considérés comme des valeurs aberrantes, principalement au-dessus de 284 236 827 $, montrant l’impact des films ayant des recettes mondiales très élevées.
* **Conclusion** : La **tendance générale** montre des films générant des recettes mondiales modérées, mais une **minorité de films à très gros succès** génère des recettes exceptionnelles, ce qui crée une distribution très biaisée et une **asymétrie marquée**.

**5. Movie average rating**

* **Distribution** : Les évaluations moyennes des films vont de 1.5 à 9.3, avec une médiane de 6.5 et une moyenne de 6.40. Cela suggère que la majorité des films ont des évaluations proches de la moyenne, avec une légère concentration autour de la médiane.
* **Asymétrie** : L’asymétrie négative (Skewness = -0.63) montre que la majorité des films ont des évaluations **supérieures à la moyenne**, bien que quelques films mal notés tirent la distribution légèrement vers la gauche.
* **Outliers** : Seules 73 évaluations sont considérées comme des valeurs aberrantes, ce qui signifie qu'il y a relativement peu de films avec des évaluations exceptionnellement faibles ou élevées.
* **Conclusion** : La **tendance générale** des films montre des évaluations plutôt modérées autour de 6.5, mais avec quelques films ayant des **évaluations exceptionnellement basses ou élevées**, qui introduisent une légère asymétrie vers la gauche.

**6. Movie number of votes**

* **Distribution** : Le nombre de votes des films varie de 5 à 2 695 887, avec une médiane de 63 990 et une moyenne de 134 370. Cela indique que la majorité des films ont un nombre de votes relativement faible, mais quelques films très populaires ont un nombre de votes exceptionnellement élevé.
* **Asymétrie** : L’asymétrie positive (Skewness = 4.47) et l'aplatissement élevé (Kurtosis = 30.72) montrent que la **distribution est fortement influencée** par des films populaires qui ont reçu une grande quantité de votes.
* **Outliers** : 353 films sont considérés comme des valeurs aberrantes, principalement au-dessus de 357 129 votes, ce qui souligne l'impact des films les plus populaires, qui attirent une quantité exceptionnelle de votes.
* **Conclusion** : La **tendance générale** montre une majorité de films avec un nombre de votes modéré, mais quelques films à **très grand succès** génèrent une **forte concentration de votes**, créant ainsi une distribution fortement asymétrique.

**7. Approval Index**

* **Distribution** : L'index d'approbation des films varie de 0.45 à 10.0, avec une médiane de 5.02 et une moyenne de 5.03. Cette petite différence entre la moyenne et la médiane montre que la distribution de l'index d'approbation est presque symétrique.
* **Asymétrie** : L’asymétrie est très faible (Skewness = -0.04), ce qui indique une **répartition plutôt symétrique** autour de la moyenne, avec très peu de films ayant un index d'approbation extrême.
* **Outliers** : Il y a 69 films considérés comme des valeurs aberrantes, mais leur nombre reste relativement faible par rapport à d’autres variables.
* **Conclusion** : La **tendance générale** montre une **répartition symétrique** des index d’approbation, avec une majorité de films ayant des valeurs proches de la médiane (autour de 5), sans valeurs extrêmes.

**Etude comparative par rapport a la variable genre :**

**1. Durée du film (runtime\_minutes) :**

* **Action** (113.17 minutes) a une durée moyenne plus élevée que d'autres genres comme **Comedy** (102.34 minutes) ou **Family** (102.25 minutes).
* Les genres comme **Animation** (87.36 minutes) et **Documentary** (103.79 minutes) tendent à avoir des films plus courts.
* Il semble qu'il y ait une tendance vers des films plus longs pour des genres comme **Biography** (123.46 minutes) et **Drama** (116.18 minutes), ce qui est logique, car ces genres peuvent offrir des récits plus profonds et complexes.

**2. Note moyenne (movie\_averageRating) :**

* Le genre avec la **note moyenne la plus élevée** est **Film-Noir** (7.25), ce qui pourrait refléter la reconnaissance artistique et la qualité des films dans ce genre.
* **Musical** (4.45) et **Family** (6.15) ont les **notes les plus basses**, ce qui pourrait indiquer que ce sont des genres plus polarisants ou que les films populaires dans ces catégories ne sont pas perçus aussi positivement par les spectateurs.
* **Romance** (7.00) et **Biography** (7.07) se distinguent également par des notes relativement élevées, indiquant que ces genres rencontrent une appréciation plus marquée.

**3. Nombre moyen de votes (movie\_numerOfVotes) :**

* Les genres **Romance** (260441) et **Mystery** (242774) génèrent un grand nombre de votes, ce qui peut indiquer que ce sont des genres populaires ou que les films dans ces catégories attirent une large audience.
* **Musical** (1042) a le nombre de votes le plus bas, ce qui pourrait être lié à un public cible plus restreint ou une faible production dans ce genre.

**4. Indice d'approbation (approval\_Index) :**

* **Romance** (6.29) et **Drama** (5.05) ont des indices d'approbation relativement élevés, ce qui suggère une satisfaction générale pour ces genres.
* Les genres comme **Musical** (2.16) et **Family** (3.90) montrent des indices plus bas, peut-être en raison de la réception partagée du public.

**5. Budget de production (Production budget $) :**

* Les genres comme **Action** (66427469.81 $) et **Sci-Fi** (40500000.00 $) ont des budgets de production élevés, ce qui est attendu pour des genres nécessitant des effets spéciaux coûteux et une production de grande envergure.
* **Family** (15000000.00 $) et **Documentary** (7991911.76 $) ont des budgets de production plus faibles, car ces genres peuvent nécessiter moins de ressources ou cibler des segments plus petits du marché.

**6. Recettes au box-office (Domestic gross $ et Worldwide gross $) :**

* **Action** et **Adventure** dominent les recettes avec des moyennes élevées pour les recettes mondiales (**Worldwide gross $**), ce qui est logique étant donné que ce sont des genres populaires et bien financés.
* Les genres comme **Documentary** (Domestic gross $: 13758136.74) et **Family** (Domestic gross $: 4142165.00) semblent avoir des recettes plus faibles, ce qui peut être dû à un public plus restreint ou à des sorties moins larges.

**Conclusion :**

* Les genres comme **Action**, **Adventure**, et **Sci-Fi** tendent à avoir de longues durées, des budgets de production élevés, et génèrent de fortes recettes au box-office.
* Des genres comme **Romance** et **Biography** ont une note moyenne élevée, mais peuvent avoir des budgets et des recettes plus faibles.
* **Documentary** et **Family** semblent moins populaires au niveau des votes et des recettes, mais avec des budgets plus modestes.

**Etude basée sur le feature director\_profession :**

**1. Durée du Film (en minutes) :**

* La durée moyenne des films varie selon les professions. Les **artistes musicaux** ont la durée moyenne la plus longue (123,75 minutes), tandis que les **gestionnaires de lieux** ont la durée moyenne la plus courte (95 minutes). Cela pourrait suggérer que des rôles comme celui des artistes musicaux sont associés à des productions de films plus longues, comme les films musicaux, tandis que les gestionnaires de lieux pourraient être impliqués dans des productions de plus petite envergure avec des durées plus courtes.

**2. Note Moyenne du Film :**

* Les **concepteurs de costumes** ont la note moyenne la plus élevée (7,50), ce qui indique que leur travail est souvent associé à des films bien notés. En revanche, les **départements de casting** et les **départements de transport** ont des notes plus basses (environ 5,7-5,8), ce qui pourrait refléter des films qui ne sont pas aussi acclamés par la critique en termes d’interprétation ou de production, mais qui restent néanmoins importants pour leur rôle dans la production.

**3. Nombre de Votes :**

* Les **producteurs** et les **scénaristes** reçoivent le plus grand nombre de votes, ce qui montre que leurs films sont plus populaires et ont un large public, les **producteurs** étant aussi fréquemment impliqués dans des films à gros budget. D'un autre côté, les **départements de casting** et **de transport** ont moins de votes, ce qui pourrait indiquer que leurs rôles ne sont pas directement liés à la reconnaissance publique des films ou des acteurs impliqués.

**4. Indice d'Approbation :**

* L'**indice d'approbation** (une mesure d'approbation ou de satisfaction) varie, avec les **concepteurs de costumes** ayant l'indice d'approbation le plus élevé (6,55), ce qui suggère que leurs contributions sont très appréciées dans l'industrie. En revanche, les **départements de casting** et **de transport** ont les indices d'approbation les plus bas (environ 3,5), ce qui pourrait signifier que ces départements ne reçoivent pas toujours le même niveau de reconnaissance ou d'appréciation, malgré leur rôle essentiel dans le processus de production.

**5. Budget de Production :**

* Le **budget de production** est en corrélation avec l'envergure des films associés à ces rôles. Les **départements musicaux** et les **effets spéciaux** ont des budgets relativement élevés (plus de 56 millions de dollars), ce qui est cohérent avec des productions à gros budget nécessitant des éléments visuels ou musicaux importants. En revanche, des départements comme le **casting** et le **transport** ont des budgets beaucoup plus bas, ce qui suggère que ces rôles sont souvent associés à des projets plus petits ou moins coûteux.

**6. Recettes Domestiques et Mondiales :**

* Les **recettes mondiales** et **domestiques** montrent que des professions comme **réalisateurs**, **producteurs**, et **scénaristes** sont associées à des films ayant de grosses recettes, notamment les **réalisateurs**, qui ont une recette moyenne mondiale de plus de 109 millions de dollars. En revanche, les **directeurs de casting** et les **départements de casting** ont des recettes beaucoup plus faibles, ce qui reflète leur rôle plus en retrait dans le processus global de production cinématographique.

**Points Clés :**

* **Les rôles associés à des tâches créatives de haute visibilité**, comme les **concepteurs de costumes**, **réalisateurs** et **producteurs**, travaillent sur des films à gros budget, ce qui se traduit par des recettes mondiales et domestiques élevées.
* **Les rôles de soutien**, comme ceux des **départements de casting** et **de transport**, contribuent à la production des films, mais sont associés à des budgets et des recettes plus faibles, probablement en raison de la nature moins visible de leurs contributions.
* L'**indice d'approbation** et les **notes des films** montrent que, généralement, les rôles les plus visibles dans la production (comme les **réalisateurs**, **concepteurs de costumes**) bénéficient d'une plus grande reconnaissance, ce qui se reflète dans des évaluations plus positives.