**Ensemble d’objectifs du projet nommé « Canadian Fuel Consumption Ratings », dont les tâches des phases de ces objectifs ont été modélisées et implémentées en utilisant le langage de programmation R, et dont les résultats fournis par ces objectifs modélisés seront montrés dans le présent document**

**Objectif 1 : Analyse des tendances sur plusieurs années**

1. **Collecte et préparation des données :**
   * **Méthode** : Techniques de nettoyage de données avec dplyr, readr, et writexl pour éliminer les valeurs manquantes et les erreurs.
   * Utilisation de la technique d’échantillonnage double pour les neuf jeux de données.
   * **Outils** : le logiciel R (les librairies dplyr, readr et writexl).

**Analyse descriptive :**

* + **Technique** : Analyse statistique descriptive (moyenne, médiane, écart-type, quartiles, médiane, mode, étendue, coefficient de variation) pour calculer les statistiques sur la consommation de carburant (en ville, sur autoroute et combiné) et sur les émissions de CO2.
  + Utilisation de fonctions R permettant de calculer les statistiques descriptives (moyenne, médiane, écart-type, quartiles, médiane, mode, étendue, coefficient de variation) pour calculer les statistiques sur la consommation de carburant (en ville, sur autoroute et combiné) et sur les émissions de CO2.
  + **Outils** : le logiciel R (les librairies dplyr, stats et summarytools).

1. **Visualisation des tendances :**

* **Méthode** : Création de graphiques de lignes simples, d’histogrammes sous forme de barres, de graphiques en aires empilées, de boxplots, de graphiques de type nuages de points, et de graphiques à barres groupées, pour visualiser les tendances de consommation de carburant (en ville, sur autoroute et combiné) et d’émissions de CO2 au fil des années (années 2015 à 2023).
* **Outils** : le logiciel R (la librairie ggplot2).

1. **Analyse des tendances annuelles :**
   * **Technique** : Analyse des séries temporelles pour observer les variations annuelles. Utilisation de tests d'hypothèses pour vérifier la significativité des changements annuels, par exemple avec le test t pour comparer les moyennes des consommations de carburant (en ville, sur autoroute et combiné) et des émissions de CO2 par année (2015 à 2023) et le test de Kruskal-Wallis pour comparer les médianes des données de tous les échantillons, de toutes les années (2015 à 2023).
   * **Outils** : le logiciel R (les librairies tseries, car et stats).
2. **Rédaction d'un rapport sur les tendances :**
   * **Approche** : Compilation des résultats obtenus et interprétation des tendances observées.
   * **Outils** : le logiciel Word pour la rédaction du rapport final.
3. **Corroboration des résultats avec un logiciel statistique :**
   * **Méthode** : Utilisation d'un outil logiciel complémentaire (comme StatGraphics) pour valider les résultats obtenus de l’analyse des tendances sur plusieurs années.

**Objectif 2 : Impact des caractéristiques des véhicules**

1. **Analyse des caractéristiques des véhicules :**
   * **Technique** : Analyse statistique descriptive (moyenne, médiane, écart-type, quartiles, médiane, mode, étendue, coefficient de variation) pour explorer les relations entre les caractéristiques des véhicules (taille du moteur, nombre de cylindres, type de transmission) et la consommation de carburant (en ville, sur autoroute et combiné) et les émissions de CO2.
   * **Outils** : le logiciel R (librairies dplyr et stats).
2. **Modélisation statistique :**
   * **Méthode** : Régression linéaire simple pour quantifier l'impact de plusieurs caractéristiques des véhicules (taille du moteur, nombre de cylindres, type de transmission) sur la consommation de carburant (en ville, sur autoroute et combiné) et les émissions de CO2.
   * **Outils** : le logiciel R (librairies lm et car).
3. **Comparaison entre différents types de véhicules :**
   * **Technique** : Comparaison des moyennes de consommation de carburant (en ville, sur autoroute et combiné) et des émissions de CO2 selon différentes catégories de véhicules nommées par le champ Vehicle Class (test de Kruskal-Wallis pour plusieurs groupes et test t pour deux groupes).
   * **Outils** : le logiciel R (les librairies car et stats).
4. **Analyse des interactions entre caractéristiques :**
   * **Méthode** : Utiliser des interactions dans des modèles de régression linéaire simple pour examiner comment certaines caractéristiques (taille du moteur, nombre de cylindres, type de transmission) interagissent et influencent la consommation de carburant (ville, autoroute et combiné) et les émissions de CO2.
   * **Outils** : le logiciel R (les librairies lm et car).
5. **Rédaction d'un rapport d'analyse :**
   * **Approche** : Compilation des résultats pour documenter l'impact des caractéristiques des véhicules.
   * **Outils** : le logiciel Word pour la rédaction du rapport final.
6. **Corroboration des résultats avec un logiciel statistique :**
   * **Méthode** : Comparer les résultats des analyses effectués tout en utilisant le langage de programmation R avec ceux obtenus en utilisant le logiciel StatGraphics pour valider les conclusions.

**Objectif 3 : Comparaison des types de carburants**

1. **Classification des véhicules par type de carburant** :
   * **Technique** : Regrouper les données en fonction du type de carburant (données provenant du champ Fuel Type) pour chaque type de véhicule (données provenant du champ Vehicle Class).
   * **Outils** : le logiciel R (la librairie dplyr).
2. **Analyse comparative** :
   * **Méthode** : Comparer les médianes de consommation de carburant (champs City, Hwy et Comb) et des émissions de CO2 pour différents types de carburants (champ Fuel Type) en utilisant des tests d'hypothèses comme le test t et le test de Kruskal-Wallis.
   * **Outils** : le logiciel R (les librairies car et stats).
3. **Visualisation des résultats** :
   * **Technique** : Créer des diagrammes en histogrammes sous format de barres et en boîtes à moustaches pour comparer les performances des différents carburants (Champ Fuel Type), lors des consommations des véhicules (en ville, sur autoroute et combiné).
   * **Outils** : le logiciel R (la librairie ggplot2).
4. **Analyse des coûts associés** :
   * **Méthode** : Utilisation de l'analyse statistique descriptive (moyenne, médiane, écart-type, mode, étendue, coefficient de variation) pour calculer les coûts moyens associés à chaque type de carburant sur une période donnée.
   * **Outils** : le logiciel R (les librairies dplyr et summarytools).
5. **Rédaction d'un rapport de comparaison** :
   * **Approche** : Documenter les résultats et observations clés des différences entre les différents types de carburant (champ fuel Type).
6. **Proposition de recommandations** :
   * **Méthode** : Basée sur tous les résultats obtenus, formuler des recommandations optimales pour le choix du type de carburant à utiliser normalement, en tenant compte de la consommation de carburant et des coûts associés.

**Objectif 4 : Émissions de CO2 et impact environnemental**

1. **Analyse de la corrélation** :

* **Technique** : Calcul de la corrélation entre la consommation de carburant (en ville, sur autoroute et combiné) et les émissions de CO2 en utilisant des mesures comme le coefficient de corrélation de Pearson.
* **Approche** :
* La corrélation de Pearson mesure la force et la direction de la relation linéaire entre deux variables quantitatives.
* **Analyse des résultats** : Déterminer si une relation significative existe entre la consommation de carburant et les émissions de CO2.
* **Outils** : le logiciel R (librairies cor et stats).

1. **Modélisation de l'impact environnemental** :

* **Technique** : Régression linéaire simple pour modéliser l'impact environnemental basé sur la consommation de carburant (en ville, sur autoroute et combiné) et d'autres facteurs comme la taille du moteur (champ Engine Size (L)) et le type de carburant (champ Fuel Type).
* **Approche** :
* Créer un modèle prédictif pour estimer les émissions de CO2 en fonction de la consommation de carburant et des caractéristiques des véhicules.
* Analyser les médianes (test de Kruskal-Wallis) dans les modèles de régression linéaire simple pour évaluer la significativité des prédicteurs.
* **Outils** : le logiciel R (librairies lm et car).

1. **Étude de cas** :

* **Méthode** :
* Sélectionner quelques véhicules représentatifs, par exemple, différents types de voitures (champ Vehicle Class) et marques (champ Make) (sélectionner 50% de la taille des 18 échantillons créés à la phase 1 de l’objectif 1), pour illustrer la relation entre la consommation de carburant (en ville, sur autoroute et combiné) et les émissions de CO2 (champ CO2 Émissions (g/km)). Utiliser le coefficient de corrélation de Pearson, celui de Spearman, celui de Kendall et le coefficient de corrélation de distance.
* Comparer les performances des véhicules en matière de consommations de carburant (en ville, sur autoroute et combiné) et d'émissions de CO2 (comparaison entre ces deux variables).
* **Outils** : le logiciel R (librairies dplyr, readr, cor, energy et summarytools).

1. **Analyse des tendances des émissions** :

* **Technique** : Analyse des séries temporelles pour évaluer l'évolution des émissions de CO2 au fil des ans.
* **Approche** :
* Appliquer des tests statistiques (par exemple, test t et le test de Mann-Kendall pour les tendances) afin de vérifier la significativité des variations des consommations de carburant (en ville, sur autoroute et combiné) et les émissions de CO2 associées dans le temps (années 2015 à 2023).
* Utiliser des graphiques de séries temporelles sous forme de lignes pour visualiser les émissions de CO2 et les consommations de carburant (en ville, sur autoroute et combiné) au cours de la période étudiée (2015-2023).
* Ajouter des histogrammes sous forme de barres, des graphiques en aires empilées, des boxplots, des nuages de points, et des graphiques à barres groupées pour compléter l'analyse visuelle des tendances des émissions de CO2 selon le type de carburant consommé (en ville, sur autoroute et combiné), et celle des tendances de consommation de carburant et des émissions de CO2 selon l’année en question.
* **Outils** : le logiciel R (les librairies ggplot2, tseries et stats).

1. **Rédaction d'un rapport sur l'impact environnemental** :

* **Approche** :
* Compiler les résultats de l'analyse de la corrélation, de la modélisation et des études de cas dans un rapport détaillé.
* Interpréter les résultats pour offrir une vue d'ensemble claire de l'impact environnemental des véhicules étudiés.
* **Outils** : le logiciel R pour la génération de visualisations et le logiciel Word pour la rédaction du rapport final.

1. **Proposition de recommandations** :

* **Méthode** :
* Formuler des recommandations pour les décideurs basés sur les analyses effectuées.
* Exemples : recommandations sur les politiques incitant à l'adoption de véhicules moins polluants, les carburants alternatifs, ou des améliorations de l'efficacité énergétique des véhicules.
* **Approche** :
* Les recommandations doivent être appuyées par les résultats des analyses statistiques et modélisations, en mettant en avant les impacts positifs des stratégies d'optimisation environnementale.
* **Outils** : le logiciel R pour la synthèse des résultats et logiciels de traitement de texte comme Word pour rédiger les recommandations.

**Objectif 5 : Cote de smog et de CO2 et santé publique**

1. **Récupération, calcul et analyse des cotes de smog et de CO2 et analyse sur les tendances de variation des cotes de smog et de CO2**

* **Technique :** Remplissage des colonnes manquantes concernant les cotes de smog et de CO2 dans les jeux de données entre 2015 et 2023, issus du projet intitulé « Canadian Fuel Consumption Ratings ». Pour chaque véhicule, les cotes de smog et de CO2 ont été calculées à l’aide d’intervalles prédéfinis basés sur les émissions de CO2 et les caractéristiques des véhicules, notamment :
* **Fuel Consumption (City) (L/100 km)**
* **Fuel Consumption (Hwy) (L/100 km)**
* **Fuel Consumption (Comb) (L/100 km)**
* **CO2 Émissions (g/km)**

Ces caractéristiques ont permis de déterminer les valeurs manquantes pour les cotes de smog et de CO2 en fonction des critères suivants :

* **CO2 Rating (Cote de CO2):**
* 10 : Moins de 100 g/km
* 9 : Entre 100 et 125 g/km
* 8 : Entre 126 et 150 g/km
* 7 : Entre 151 et 175 g/km
* 6 : Entre 176 et 200 g/km
* 5 : Entre 201 et 225 g/km
* 4 : Entre 226 et 250 g/km
* 3 : Entre 251 et 275 g/km
* 2 : Entre 276 et 300 g/km
* 1 : Plus de 300 g/km
* **Smog Rating (Cote de Smog):**
* 10 : Véhicules électriques ou hybrides de haute performance (faible pollution)
* 8-9 : Véhicules hybrides et certains modèles à faible consommation d'essence
* 6-7 : Véhicules économiques à essence ou diesel propre
* 4-5 : Véhicules standards à essence avec moteur moyen
* 2-3 : Véhicules à forte consommation de carburant (SUV, grosses cylindrées)
* 1 : Véhicules très polluants, grosses voitures ou camions avec de fortes émissions
* **Approche :**
* Utilisation des données gouvernementales associées au projet « Canadian Fuel Consumption Ratings » pour récupérer les informations manquantes sur les côtes de CO2 et de smog des véhicules.
* Le remplissage des valeurs manquantes dans les colonnes des cotes de CO2 et de smog a été effectué en fonction des niveaux d'émissions et des caractéristiques des véhicules mentionnées ci-dessus.
* Pour évaluer les tendances de variation des cotes de CO2 et de smog sur la période 2015-2023, des analyses statistiques seront réalisées à l’aide du test de **Kruskal-Wallis** et du test de **Dunn** sur les 18 échantillons provenant des 9 jeux de données utilisés dans ce projet. Ces tests permettront d’évaluer la significativité des variations des cotes de CO2 et de smog en fonction des années, des types de véhicules, de la taille du moteur et du nombre de cylindres.
* **Outils :** le logiciel R (la librairie dplyr pour la gestion des données, la librairie readr pour l'importation des fichiers, la librairie writexl pour l'exportation des résultats, et la librairie stats pour les tests statistiques).

1. **Analyse des tendances de la cote de smog et de la côte de CO2**

* **Technique** : Analyse des séries temporelles pour suivre l'évolution de la cote de smog sur la période 2015-2023.
* **Approche** :
* Appliquer des méthodes comme le lissage exponentiel pour repérer des tendances de variation des côtes de smog et de CO2 à long terme (années 2015 à 2023) et selon le type de carburant (champ fuel type).
* Utiliser le test de Mann-Kendall pour évaluer la significativité des tendances observées dans la cote de smog et de CO2, au fil des ans (années 2015 à 2023), et selon le type de carburant consommé (champ Fuel Type).
* Visualiser, à l’aide de graphiques à lignes, d’histogrammes sous forme de barres, de graphiques de type boîte à moustaches et de graphiques de type nuage de points les changements dans la cote de smog des véhicules sur plusieurs années (années 2015 à 2023), selon le type de carburant (Fuel Type).
* **Outils** : le logiciel R (la librairie ggplot2 pour les visualisations, la librairie stats pour les tests statistiques, et la librairie forecast pour les méthodes de séries temporelles).

1. **Impact sur la santé publique**
   * **Technique** : Analyse contextuelle des données sur la santé publique liées à la pollution atmosphérique.
   * **Approche** :
   * Croiser les données de la cote de smog avec les indicateurs de santé publique, comme les taux de maladies respiratoires dans les régions les plus touchées par la pollution.
   * Utiliser des analyses de corrélation pour voir s’il existe une relation entre la hausse de la cote de smog et l’augmentation des problèmes de santé liés à la pollution (par exemple, les hospitalisations pour des maladies respiratoires).
   * **Outils** : le logiciel R (la librairie dplyr pour le traitement des données, et la librairie stats pour les analyses de corrélation).
2. **Consultation d’informations provenant d'experts sur Internet**

* **Méthode** :
* Rechercher et consulter des pages web d’experts en santé publique (par exemple, l’Organisation mondiale de la santé, ou des experts nationaux) pour obtenir des informations pertinentes sur l’impact du smog sur la santé.
* **Approche** :
* Compléter l’analyse quantitative avec des éléments qualitatifs basés sur des études et publications récentes concernant les effets du smog sur la santé publique.
* **Outils** : le logiciel R pour organiser les résultats de la recherche et les incorporer dans l’analyse.

1. **Rédaction d'un rapport sur la cote de smog**

* **Approche** :
* Compiler les résultats de l'analyse des tendances de la cote de smog, en expliquant les implications sur la santé publique et l'environnement.
* Inclure des sections dédiées aux tendances observées, à leur impact sur la santé publique et aux mesures à prendre pour améliorer la qualité de l’air.
* **Outils** : Utilisation du logiciel Word pour la rédaction du rapport, avec des visualisations produites à l’aide du logiciel R (la librairie ggplot2).

1. **Proposition de recommandations**

* **Méthode** :
* Sur la base des résultats de l'analyse, formuler des recommandations pour les décideurs publics et les fabricants de véhicules.
* Exemples : recommandations sur l’adoption de véhicules moins polluants ou des politiques incitatives pour réduire les émissions de smog.
* **Approche** :
* Les recommandations seront appuyées par les résultats des analyses de tendance et par l'impact de la cote de smog sur la santé publique.
* **Outils** : le logiciel R pour générer des résumés clairs des résultats, suivi de la rédaction des recommandations dans un rapport (en utilisant le logiciel Word).

**Objectif 6 : Analyse des modèles de véhicules**

1. **Identification des modèles de véhicules**

* **Technique** : Classification des données pour regrouper les véhicules par marque et modèle.
* **Approche** :
* Utiliser les fonctions de regroupement en R (comme group\_by de dplyr) pour créer des catégories basées sur les marques (champ Make) et modèles (champ Model) des véhicules.
* Identifier les différentes marques et modèles présents dans les 18 échantillons associés aux 9 jeux de données, créées avec la technique d’échantillonnage double lors de la réalisation de la phase 1 de l’objectif 1.
* **Outils** : le logiciel R (la librairie dplyr pour le traitement et la classification des données).

1. **Évaluation des performances par marque et modèle**

* **Technique** : Analyse descriptive et inférentielle pour évaluer la consommation de carburant (en ville, sur autoroute et combiné) et les émissions de CO2 pour chaque marque et modèle de véhicule.
* **Approche** :
* Calculer les statistiques descriptives (moyennes, médianes, écart-types, médianes, modes, étendues, coefficients de variation) pour chaque marque et modèle des véhicules, en termes de consommation de carburant (en ville, sur autoroute et combiné) et d'émissions de CO2.
* **Outils** : R (la librairie dplyr pour les statistiques descriptives, la librairie car et stats pour les tests T et de Kruskal-Wallis).

1. **Comparaison entre marques et modèles de véhicules**

* **Technique** : Analyse comparative pour identifier les marques et les modèles de véhicules offrant des performances supérieures, en termes de consommation de carburant (en ville, sur autoroute et combiné) et d’émissions de CO2.
* **Approche** :
* Calculer les statistiques descriptives (moyenne, médiane, écart-type, mode, étendue, coefficient de variation) des indices de performance (par exemple, consommation de carburant, en ville, sur autoroute et combiné par100 km et les émissions de CO2 produits) et identifier les marques et les modèles de véhicules qui se distinguent par leur meilleur indice de performance, en termes de consommation de carburant (en ville, sur autoroute et combiné) et d’émissions de CO2 minimes.
* Utiliser des diagrammes en boîte à moustaches, de graphiques à lignes simples, des graphiques à nuage de points et d’histogrammes sous forme de barres pour visualiser la distribution de la consommation de carburant (en ville, sur autoroute et combiné) et des émissions de CO2, par marque et par modèle de véhicules.
* **Outils** : R (la librairie ggplot2 pour les visualisations, la librairie car pour les tests statistiques).

1. **Analyse des modèles émergents**

* **Technique** : Analyse des tendances pour évaluer l'impact des nouveaux modèles sur la consommation de carburant.
* **Approche** :
* Identifier les nouveaux modèles introduits sur le marché entre 2015 et 2023.
* Comparer la consommation de carburant et les émissions de ces nouveaux modèles par rapport aux modèles plus anciens, en utilisant des méthodes de régression linéaire simple pour quantifier l'impact.
* **Outils** : le logiciel R (la librairie dplyr pour le traitement des données, la librairie lm pour la régression linéaire simple).

1. **Rédaction d'un rapport sur les modèles de véhicules**

* **Approche** :
* Compiler les résultats des analyses effectuées sur les performances de chaque modèle, les comparaisons entre marques et l'analyse des modèles émergents dans un rapport.
* Inclure des visualisations pour illustrer les résultats et les tendances observées.
* **Outils** : Utilisation du logiciel Word pour la rédaction du rapport, intégrant les visualisations réalisées avec le logiciel R (la librairie ggplot2).

1. **Corroborer les résultats obtenus en utilisant un logiciel de calcul et d'analyse statistique et probabiliste comme StatGraphics**

* **Technique** : Validation croisée des résultats obtenus.
* **Approche** :
* Utiliser StatGraphics pour réaliser des analyses complémentaires, comme des graphiques et des tests statistiques, afin de vérifier les résultats obtenus avec R.
* Comparer les résultats des deux outils pour assurer la robustesse des conclusions.
* **Outils** : StatGraphics pour les analyses, R pour les analyses principales.

**Objectif 7 : Importance de l’année du véhicule**

1. **Analyse chronologique des données**

* **Technique** : Analyse des séries temporelles pour segmenter les données par année.
* **Approche** :
* Regrouper les données par année et calculer des statistiques descriptives (moyenne, médiane, écart-type, mode, étendue, coefficient de variation et coefficient de corrélation de Pearson) pour la consommation de carburant et les émissions de CO2 selon l’année des véhicules (2015 à 2023).
* Visualiser les performances des véhicules analysées (consommation de carburant en ville, sur autoroute et combiné et émissions de CO2) au fil des ans (années 2015 à 2023) à l'aide de graphiques linéaires simples, des histogrammes sous forme de barres, des graphiques type nuage de points, des graphiques de type boîtes à moustaches et des graphiques en barres groupées (par types de consommation de carburant et par émissions de CO2) pour observer les tendances de consommation de carburant (en ville, sur autoroute et combiné) et d’émissions de CO2, selon l’année des véhicules (années 2015 à 2023).
* **Outils** : le logiciel R (la librairie dplyr pour le traitement des données, la librairie ggplot2 pour les visualisations).

1. **Évaluation des effets des réglementations**

* **Technique** : Analyse des impacts pour étudier les effets des changements réglementaires sur la performance énergétique.
* **Approche** :
* Identifier les principales réglementations adoptées durant la période d'étude et les relier aux données de consommation et d'émissions.
* Utiliser des modèles de régression pour évaluer l'impact des changements réglementaires sur la performance énergétique des véhicules, en tenant compte d'autres variables pertinentes.
* **Outils** : le logiciel R (la librairie lm pour la régression).

1. **Comparaison des années récentes aux précédentes**

* **Technique** : Tests statistiques pour évaluer les améliorations significatives des véhicules récents.
* **Approche** :
* Comparer les performances des véhicules récents (2020-2023) avec celles des années précédentes (2015 à 2019) en utilisant des tests d'hypothèses (par exemple, le tests t et le test de Kruskal-Wallis).
* Analyser les résultats pour déterminer si les différences observées sont significatives.
* **Outils** : le logiciel R (la librairie stats pour les tests statistiques).

1. **Identification des tendances par année**

* **Technique** : Analyse de régression pour identifier les tendances spécifiques par année de modèle.
* **Approche** :
* Utiliser une régression linéaire simple pour modéliser l'évolution de la consommation de carburant (en ville, sur autoroute et combiné) et des émissions de CO2 en fonction de l'année (champ Year), de la marque (champ Make) et du modèle (champ Model) des véhicules.
* Interpréter les coefficients de régression pour comprendre l'impact de l'année sur les performances.
* **Outils** : le logiciel R (la librairie lm pour la régression, la librairie ggplot2 pour les graphiques).

1. **Rédaction d'un rapport sur l'importance de l’année**

* **Approche** :
* Documenter les résultats des analyses concernant l'évolution des performances des véhicules, les effets des réglementations, et les tendances par année dans un rapport structuré.
* Inclure des visualisations et des conclusions claires pour soutenir les recommandations.
* **Outils** : Utilisation du logiciel Word pour la rédaction du rapport.

1. **Sensibilisation des consommateurs**

* **Approche** :
* Rédiger un court document pour sensibiliser les consommateurs à l'importance de choisir des véhicules récents, en soulignant les avantages en termes de consommation de carburant et d'émissions.
* Inclure des graphiques et des points clés basés sur les résultats de l’analyse pour renforcer le message.
* **Outils** : Word ou PowerPoint pour la création du document, intégrant des éléments visuels issus des analyses.

**Objectif 8 : Évolution selon le type de transmission et le nombre de cylindres**

1. **Classification des types de transmission**

* **Technique** : Classification des données pour regrouper les véhicules par type de transmission (provenant du champ Transmission) et par nombre de cylindres (provenant du champ Cylinders).
* **Approche** :
* Identification des valeurs du champ Transmission et du champ Cylinders: Extraire et utiliser les valeurs exactes telles qu'elles apparaissent dans les 18 échantillons associés aux 9 jeux de données.
* Classification par types de transmission et par nombre de cylindres : Classifier les véhicules en fonction des valeurs brutes des champs Transmission et Cylinders.
* Création d'une nouvelle variable : La nouvelle variable correspondra à la valeur exacte des données classifiées des champs Transmission et Cylinders.
* **Outils** : le logiciel R (la librairie dplyr pour la manipulation des données).

1. **Analyse de l’efficacité énergétique par transmission et par nombre de cylindres**

* **Technique** : Analyse comparative pour évaluer l'influence de chaque type de transmission (champ Transmission) et de chaque nombre de cylindres (champ Cylinders) sur la consommation de carburant (en ville, sur autoroute et combiné) et sur les émissions de CO2.
* **Approche** :
* Calculer des statistiques descriptives (moyenne, médiane, écart-type, mode, étendue, coefficient de variation) pour la consommation de carburant (en ville, sur autoroute et combiné) et pour les émissions de CO2 des véhicules analysées, en fonction des types de transmission (champ Transmission) et des nombres de cylindres (champ Cylinders).
* Utiliser des tests d'hypothèses (comme le test de Kruskal-Wallis et test de Dunn) pour déterminer si les différences de consommation de carburant et les différences d’émissions de CO2 des véhicules analysées, entre les types de transmission et les nombres de cylindres sont significatives.
* **Outils** : le logiciel R (la librairie car pour le test de Kruskal-Wallis et pour le test de Dunn).

1. **Comparaison des performances au fil du temps**

* **Technique** : Analyse des séries temporelles pour examiner les tendances d’efficacité énergétique (consommations de carburant en ville, autoroute et combiné et émissions de CO2) que donnent les différents types de transmission et nombres de cylindres des véhicules analysées, des années 2015 à 2023.
* **Approche** :
* Segmenter les données par type de transmission, nombre de cylindres et année, puis calculer des statistiques descriptives d’efficacité énergétique (moyenne, médiane, écart-type, mode, étendue et coefficient de variation).
* Visualiser les tendances à l'aide de graphiques linéaires simples pour montrer l'évolution de l'efficacité énergétique (consommations de carburant en ville, autoroute et combiné et émissions de CO2) que donne chaque type de transmission et de chaque nombre de cylindres des véhicules analysées au fil des ans (années 2015 à 2023).
* **Outils** : le logiciel R (la librairie dplyr pour le traitement des données, la librairie ggplot2 pour les graphiques).

1. **Étude des préférences des consommateurs**

* **Technique** : Analyse descriptive et enquête pour étudier les tendances d'achat liées aux types de transmission et aux nombres de cylindres.
* **Approche** :
* Analyser les données de vente pour identifier les préférences des consommateurs concernant les types de transmission et les nombres de cylindres.
* Si disponible, intégrer des données d’enquête pour évaluer les raisons des préférences des consommateurs pour un type de transmission et pour un nombre de cylindres particuliers.
* **Outils** : le logiciel R (la librairie dplyr pour la manipulation des données, la librairie ggplot2 pour visualiser les préférences).

1. **Rédaction d'un rapport sur les transmissions**

* **Approche** :
* Compiler les résultats des analyses sur l'influence de la transmission et le nombre de cylindres sur la consommation de carburant et l'efficacité énergétique dans un rapport structuré.
* Inclure des graphiques et des statistiques pour soutenir les conclusions et recommandations.
* **Outils** : Utilisation du logiciel Word pour la rédaction du rapport.

1. **Sensibilisation du public**

* **Approche** :
* Rédiger un court document pour sensibiliser le public sur l'impact des choix de transmission et de nombre de cylindres sur la consommation de carburant et l'efficacité énergétique.
* Inclure des graphiques et des points clés basés sur les résultats de l’analyse pour renforcer le message.
* **Outils** : Word ou PowerPoint pour la création du document, intégrant des éléments visuels issus des analyses.

**Objectif 9 : Différences entre les classes de véhicules**

1. **Classification des véhicules par classe de véhicule**

* **Technique** : Classification des données pour segmenter les véhicules en différentes classes (provenant du champ Vehicle Class).
* **Approche :**
* Identification des valeurs du champ Vehicle Class : Extraire et utiliser les valeurs exactes telles qu'elles apparaissent dans les 18 échantillons associés aux 9 jeux de données.
* Classification des classes de véhicules : Classifier les véhicules en fonction des valeurs brutes du champ Vehicle Class.
* Création d'une nouvelle variable : La nouvelle variable correspondra à la valeur exacte de la classe des véhicules (champ Vehicle Class) dont les données sont dans les 18 échantillons associés aux 9 jeux de données.
* **Outils** : le logiciel R (la librairie dplyr pour la manipulation des données).

1. **Analyse comparative des classes**

* **Technique** : Analyse statistique pour évaluer la consommation de carburant (en ville, sur autoroute et combiné) et les émissions de CO2 pour chaque classe de véhicules (années 2015 à 2023).
* **Approche** :
* Calculer des statistiques descriptives (moyenne, médiane, écart-type, mode, étendue et coefficient de variation) pour la consommation de carburant (en ville, sur autoroute et combiné) et les émissions de CO2 par classe de véhicule analysée.
* **Outils** : le logiciel R (la librairie dplyr pour la manipulation des données et les calculs de statistiques descriptives).

1. **Identification des tendances par classe**

* **Technique** : Analyse des séries temporelles pour examiner l'évolution des performances de consommation de carburant (en ville, sur autoroute et combiné) et d’émissions de CO2 des classes de véhicules au fil des ans (années 2015 à 2023).
* **Approche** :
* Segmenter les données par classe de véhicule (champ Vehicle Class) et par année (champ Year), puis calculer les statistiques descriptives (moyenne, médiane, écart-type, mode, étendue et coefficient de variation) sur la consommation de carburant (en ville, sur autoroute et combiné) et sur les émissions de CO2, pour chaque classe et année de véhicule.
* Visualiser les tendances à l'aide de graphiques linéaires simples, d’histogrammes sous forme de barres, de graphiques en aires empilées, de boxplots, de graphique à nuages de points, et de graphiques à barres groupées pour montrer l'évolution des performances de consommation de carburant (en ville, sur autoroute et combiné) et d’émissions de CO2 de chaque classe de véhicule et de chaque année de véhicule.
* **Outils** : le logiciel R (la librairie dplyr pour le traitement des données, la librairie ggplot2 pour les graphiques).

1. **Évaluation de l'impact des caractéristiques de classes**

* **Technique** : Analyse de régression pour étudier l'impact de la taille du moteur (Engine Size (L)), du nombre de cylindres (Cylinders), du type de transmission (Transmission) et du type de carburant (Fuel Type) de chaque classe de véhicule sur la consommation de carburant (en ville, sur autoroute et combiné) et sur les émissions de CO2.
* **Approche** :
* Utiliser des modèles de régression pour examiner la relation entre les caractéristiques des classes de véhicules (par exemple, classe de véhicule, taille du moteur, nombre de cylindres, type de transmission et type de carburant) et la consommation de carburant et les émissions de CO2.
* Analyser les coefficients du modèle pour identifier l'impact significatif des caractéristiques des véhicules (classe de véhicule, taille du moteur, nombre de cylindres, type de transmission et type de carburant) sur l'efficacité énergétique (consommation de carburant en ville, sur autoroute et combiné et émissions de CO2).
* **Outils** : le logiciel R (la librairie lm pour la régression).

1. **Rédaction d'un rapport sur les classes de véhicules**

* **Approche** :
* Compiler les résultats des analyses sur les différences entre les classes de véhicules dans un rapport structuré.
* Inclure des graphiques et des statistiques pour soutenir les conclusions et recommandations.
* **Outils** : Utilisation du logiciel Word pour la rédaction du rapport.

1. **Sensibilisation des consommateurs**

* **Approche** :
* Rédiger un court document pour sensibiliser le public sur le choix de la classe de véhicules en tenant compte de la consommation de carburant et des émissions de CO2.  
  Intégrer des graphiques et des points clés basés sur les résultats de l’analyse pour renforcer le message.
* **Outils** : Word ou PowerPoint pour la création du document, intégrant des éléments visuels issus des analyses.

**Objectif 10 : Analyse des performances des moteurs au fil du temps**

1. **Segmentation des données par taille du moteur et nombre de cylindres**

* **Technique** : Classification des données pour classer les véhicules selon la taille du moteur (champ Engine Size) et le nombre de cylindres (champ Cylinders).
* **Approche** :
* Identification des valeurs des champs Engine Size et Cylindres : Extraire et utiliser les valeurs exactes telles qu'elles apparaissent dans les 18 échantillons associés aux 9 jeux de données.
* Segmentation des données par taille du moteur (champ Engine Size) et par nombre de cylindres (champ Cylinders).
* Classifier les données des véhicules en fonction des valeurs brutes des champs Engine Size et Cylindres.
* Création de nouvelles variables : Les nouvelles variables correspondront à la valeur exacte de la taille du moteur (champ Engine Size) et du nombre de cylindres (champ Cylinders).
* **Outils** : le logiciel R (la librairie dplyr pour la manipulation des données).

1. **Évaluation des performances au fil du temps**

* **Technique** : Analyse statistique pour analyser l'évolution de la consommation de carburant (en ville, sur autoroute, et combinée) et des émissions de CO2 en fonction de la taille du moteur (champ Engine Size) et du nombre de cylindres (champ Cylinders).
* **Approche** :
* Calculer des statistiques descriptives (moyenne, médiane, écart-type, mode, étendue et coefficient de variation) pour la consommation de carburant et les émissions de CO2 par taille de moteur (champ Engine Size) et par nombre de cylindres (champ Cylinders).
* Utiliser des tests d'hypothèses (comme le test de Kruskal-Wallis et le test de Dunn) pour déterminer si les différences de consommation de carburant (en ville, sur autoroute, et combinée) et des émissions de CO2, en fonction de la taille du moteur (champ Engine Size) et du nombre de cylindres (champ Cylinders), sont significatives entre les données provenant de chacun des 18 échantillons associés aux 9 jeux de données de consommation de carburant et d’émissions de CO2, entre les années 2015 à 2023.
* Visualiser l’évolution de la consommation de carburant (en ville, sur autoroute, et combinée) et des émissions de CO2 en fonction de la taille du moteur (champ Engine Size) et du nombre de cylindres (champ Cylinders) au fil du temps (années 2015 à 2023) à l'aide de graphiques linéaires simples, d’histogrammes sous forme de barres, de graphiques en aires empilées, de boxplots, de graphique à nuages de points, et de graphiques à barres groupées pour montrer l'évolution des performances de consommation de carburant (en ville, sur autoroute et combiné) et d’émissions de CO2 de chaque classe de véhicule et de chaque année de véhicule.
* **Outils** : le logiciel R (la librairie car pour les tests de Kruskal-Wallis et les tests de Dunn, la librairie ggplot2 pour les graphiques).

1. **Étude des technologies émergentes**

* **Technique** : Analyse qualitative et quantitative pour identifier les nouvelles technologies qui ont amélioré l'efficacité des moteurs.
* **Approche** :
* Rechercher des données et des études de cas sur les innovations technologiques (comme l'injection directe, les turbocompresseurs, etc.).
* Évaluer l'impact de ces technologies sur la consommation de carburant et les performances des moteurs à travers des études de corrélation.
* **Outils** : Revue de littérature et recherche documentaire, avec des outils d'analyse de données comme R pour l'analyse des tendances.

1. **Analyse des performances par taille de moteur et par nombre de cylindres**

* **Technique** : Analyse comparative pour comparer les performances (de consommation de carburant en ville, sur autoroute et combiné et émissions de CO2) des tailles de moteurs (champ Engine Size) et des nombres de cylindres (champ Cylinders) de toutes les voitures dont les données sont dans les 18 échantillons associés aux 9 jeux de données.
* **Approche** :
* Calculer les statistiques descriptives (moyenne, médiane, écart-type, mode, étendue et coefficient de variation) pour la consommation de carburant (en ville, sur autoroute et combinée) et les émissions de CO2 par type de moteur (champ Engine Size) et par nombre de cylindres (champ Cylinders).
* Utiliser des tests d'hypothèses (comme le test de Kruskal-Wallis et le test de Dunn) pour évaluer les différences significatives entre les tailles de moteurs (champ Engine Size) et entre les nombres de cylindres (champ Cylinders) et les consommations de carburant (en ville, sur autoroute et combinée) et les émissions de CO2 qu’ils produisent.
* Visualiser les performances (de consommation de carburant en ville, sur autoroute et combiné et émissions de CO2) des tailles de moteurs (champ Engine Size) et des nombres de cylindres (champ Cylinders) de toutes les voitures analysées au fil des ans (années 2015 à 2023), à l'aide de graphiques linéaires simples, d’histogrammes sous forme de barres, de graphiques en aires empilées, de boxplots, de graphique à nuages de points, et de graphiques à barres groupées pour montrer l'évolution des performances de consommation de carburant (en ville, sur autoroute et combiné) et d’émissions de CO2 de des différentes tailles de moteur (champ Engine Size) et des différents nombres de cylindres (champ Cylinders) de chaque classe de véhicule et de chaque année de véhicule (années 2015 à 2023).
* **Outils** : le logiciel R (la librairie car pour les tests de Kruskal-Wallis et les tests de Dunn, la librairie ggplot2 pour les graphiques).

1. **Rédaction d'un rapport sur les performances des moteurs**

* **Approche** :
* Compiler les résultats des analyses sur les performances des moteurs au fil du temps dans un rapport structuré.
* Inclure des graphiques et des statistiques pour soutenir les conclusions et recommandations.
* **Outils** : Utilisation du logiciel Word pour la rédaction du rapport.

1. **Sensibilisation des consommateurs**

* **Approche** :
* Rédiger un court document pour sensibiliser le public sur l'importance du choix de moteur, en tenant compte de la consommation de carburant et des émissions de CO2.
* Intégrer des graphiques et des points clés basés sur les résultats de l’analyse pour renforcer le message.
* **Outils** : Word ou PowerPoint pour la création du document, intégrant des éléments visuels issus des analyses.

Il est à noter que, lors de la rédaction du rapport Word du travail noté 2 de SCI 1402, Gonzalo Alfredo Romero Francia a expliqué, avec trop de détails et résultats de tests statistiques de type test d’hypothèses ANOVA à deux critères différents, le fait que les données provenant des 18 échantillons créés à partir des 9 jeux de données associés au projet nommé « Canadian Fuel Consumption Ratings », par le tout premier code R implémentant la tâche 1 de la phase 1 de l’objectif 1 de ce projet en science de données nommé ci-dessus, ne suivent jamais une distribution normale, alors il a bien expliqué le fait que, au lieu d’utiliser l’analyse statistique de type test d’hypothèses ANOVA à deux critères différents, il allait utiliser les statistiques de type test d’hypothèses non paramétriques nommés Test t, Test de Kruskal-Wallis, Test de Dunn et Test de Mann-Kendall. Cela est déjà expliqué en détail à la section nommé « Ensemble d’informations importantes pour la réalisation des tâches des phases de chacun des 10 objectifs du projet nommé « Canadian Fuel Consumption Ratings », dont ces tâches des phases de ces objectifs seront modélisées et implémentées en utilisant le langage de programmation R » du document Word du travail noté 2 de SCI 1402.