

R à Québec 2017 - Atelier d'introduction à R

Deuxième Partie - Étude de cas - Jeudi 25 mai 13h à 16h

Table des matières

- [Description Sommaire](#)
- [Mandat](#)
- [Énoncé](#)
 - [Question 1 - Extraction, traitement, visualisation et analyse des données](#)
 - [Question 2 - Création de fonctions utilitaires](#)
 - [Question 3 - Communication de vos résultats](#)
 - [Question 4 - Analyse de la compétition](#)
 - [Question 5 - Ajustement des lois de distribution sur données empiriques](#)
 - [Question 6 - Simulation et analyse de rentabilité](#)
- [Auteurs](#)
- [Liens utiles](#)

Description Sommaire

Dans le cadre de cette étude de cas, nous nous placerons dans la peau d'un analyste du département de tarification oeuvrant au sein d'une compagnie se spécialisant dans le transport de colis par voies aériennes en mettant à profit le jeu de données d'[OpenFlights](#).

Mandat

Notre mandat consistera dans un premier temps à analyser les bases de données que nous avons à notre disposition afin de créer des fonctions qui permettront de facilement intégrer les informations qu'elles contiennent lors de la tarification d'une livraison spécifique. Une fois cette tarification complétée, nous devons fournir des chartes pour facilement estimer les prix d'une livraison qui s'avèreront être des outils indispensables au département de marketing et au reste de la direction. Après avoir transmis les documents en question, votre gestionnaire voulant s'assurer que la nouvelle tarification sera efficiente et profitable vous demandera d'analyser les prix de la concurrence pour en extrapoler leur tarification. Finalement, vous serez appelé à comparer ces deux tarifications et la compétitivité de votre nouvelle tarification comparativement au reste du marché en procédant à une analyse stochastique.

Énoncé

Question 1 - Extraction, traitement, visualisation et analyse des données

1. Extraire les bases de données [airports.dat](#), [routes.dat](#) et [province](#).
2. Nettoyer le jeu de données en ne conservant que les données relatives au Canada.
3. Extraire des informations générales sur la distribution des variables présente dans le jeu de données et vous informer sur la signification de ces dernières. Ainsi que sur les différentes modalités qu'elles peuvent prendre. Identifier les données manquantes.
4. Corriger les modalités des variables absentes et faire une sélection des variables qui vous semble utile pour le reste du traitement.
5. Créer une carte du Canada affichant les différents aéroports à l'aide du paquetage ggmap.
6. Créer une seconde carte montrant toutes les routes possibles entre ces différents aéroports.
7. Calculer un indice d'achalandage des aéroports en fonction de la quantité de routes en destination.
8. Créer une carte permettant de visualiser l'indice grâce à un graphique à bulles.

Note:

Les données disponibles avec le code source ont été simplifiées, les données d'origine sont disponibles en référence.

Question 2 - Création des fonctions utilitaires

1. Écrivez le code source de la fonction **airportsDist(sourceIATA, destIATA)** qui permettra de calculer la distance par vol d'oiseau entre deux aéroports.
2. Écrivez le code source de la fonction **arrivalTime(sourceIATA, destIATA)** qui permettra de calculer l'heure d'arrivée dans la ville de destination d'un colis qui serait posté immédiatement.
3. Écrivez le code source de la fonction **shippingCost(sourceIATA, destIATA, weight, *percentCredit, *dollarCredit)** qui permettra de tarifier la livraison d'un colis donné en fonction de son poids, sa provenance et sa destination, et en tenant compte des frais fixes, des taxes (variant par province), de la marge de profits, des rabais applicables (% et \$) ainsi que des normes suivantes:
 - Distance minimale de 100 km
 - Poids maximale de 30 Kg
 - Aucune surcharge pour colis de poids inférieur à 2 Kg
 - Un rabais pour un envoi ayant un coût de base de plus de 100\$
 - Un rabais selon la province de destination
 - 15 % pour le Québec
 - 5 % pour la Colombie-Britannique
 - 10 % pour l'Ontario
 - 2.5 % pour l'Alberta
 - Un rabais pour les longues distances
 - Un rabais de 15 % pour une distance entre 2000 et 2500 Km
 - Un rabais de 12.25 % pour une distance de entre 2500 et 3000 Km
 - Un rabais de 10 % pour une distance de plus de 3000 Km

- Un prix ne devrait jamais être négatif ou en deçà des FraisFixes * (1 + profitMargin)

Notes:

- N'oubliez pas de tenir compte des différents fuseaux horaires.
- Faites en sorte que votre fonction accomplisse la vérification des conditions et renvoie un message d'aide pour remédier à la situation.
- Votre fonction devrait être complètement paramétrable sans nécessairement exiger la définition de tous les paramètres afin de permettre son utilisation. Pensez donc à définir des valeurs par défaut aux arguments que vous jugerez optionnels.

Question 3 - Communication de vos résultats

1. À partir de la fonction **shippingCost**, créer des chartes graphiques permettant d'analyser la tarification d'un trajet en fonction de sa **distance** sachant que le poids moyen des colis est de **Y**.
2. Pendant que vous étiez en train de créer les chartes de distance, vous vous souvenez que vous avez reçu une demande spéciale du département de marketing pour les comptoirs de dépôt de la ville de Montréal. Celle-ci aimerait avoir différentes chartes variant selon le **poids** pour les trajets suivants:
 - Montréal (YUL) - Québec (YQB)
 - Montréal (YUL) - Vancouver (YVR)
 - Montréal (YUL) - Toronto (YYZ)
 - Montréal (YUL) - Calgary (YYC)

Question 4 - Analyse de la compétition

Grâce à la base de données [benchmark.csv](#), vous serez en mesure d'extrapoler grâce à des techniques statistiques la tarification de la compétition.

1. Visualiser la distribution des prix en fonction du poids.
2. Visualiser ensuite la distribution des prix en fonction de la distance.
3. Visualiser maintenant la distribution des prix en fonction de poids et de la distance.
4. Tester l'indépendances des variables à l'aide du test de la fonction *chisq.test* qui effectue le test de Chi carré.
5. Vous constatez qu'un modèle linéaire serait suffisant pour faire l'approximation.
6. En utilisant la fonction *lm*, déterminer la droite de régression afin de déterminer les paramètres de la loi sous-jacente.

Note:

Vous savez que vos concurrents utilisent principalement le poids et la distance pour déterminer le prix des livraisons.

Question 5 - Ajustement des lois de distribution sur les données empiriques

En reprenant les données de la compétition, vous êtes aussi en mesure d'extraire la distribution suivie par le poids des colis et des distances.

1. Utiliser le paquetage *actuar* et la fonction *optim* ou *fitdistr* pour ajuster les distributions suivantes à la distribution empirique en fonction du poids.
 - Loi Normale

- Loi Gamma
- Loi Log-normale
- Loi Weibull

2. Présenter les différentes distributions obtenues.
3. Faire un choix de distribution à partir de la déviance.

Note:

Les paramètres initiaux de vos optimisations peuvent impacter le résultat de la fonction.

Question 6 - Simulation et analyse de rentabilité

Encore une fois, pour le bureau de Montréal, qui est inquiet que la nouvelle tarification nous fasse perdre des parts de marché qui sont actuellement de 15%. On vous demande de faire une analyse de rentabilité. Pour ce faire, vous décidez donc d'utiliser des techniques stochastiques afin de simuler un bassin de clients au cours de la prochaine année.

1. Simuler le nombre de colis envoyés au cours de chacun des prochains mois à l'aide d'une distribution Poisson. En raison de la saisonnalité, le nombre de colis moyens envoyés par mois au cours des trois dernières années était:

Mois	Nb Colis (Avg. 3yrs)
Janvier	2000
Février	1700
Mars	1500
Avril	1350
Mai	1600
Juin	1650
Juillet	1750
Août	2000
Septembre	2300
Octobre	2425
Novembre	2500
Décembre	3500

2. Pour chacun des colis, vous devez ensuite générer un poids à l'aide de la distribution retenue précédemment ainsi que la destination à l'aide de la distribution empirique des indices d'aéroports.
3. Calculer ensuite les prix chargés par votre compagnie ainsi que la compétition pour chacun des transports.
4. Déterminer les revenus totaux de votre compagnie ainsi que de la compétition au cours de la prochaine année.
5. Déterminer comment la nouvelle tarification impactera la part de marché.

Notes:

- Pensez à définir un point de départ à votre générateur de nombres aléatoires afin que vos résultats soient reproductibles.
- Distribution Poisson *moyenne* = λ .
- On considère toutes les distributions comme équivalentes et ayant la même probabilité d'être la destination choisie par le client.
- Nous considérons que le marché est efficient et que les clients feront leur choix de compagnie qu'en fonction du prix chargé, en allant évidemment avec la compagnie chargeant le moins cher.

Auteurs

- Superviseur:
 - [Vincent Goulet](#) - Vincent.Goulet@act.ulaval.ca
- Concepteurs:
 - [David Beauchemin](#) - david.beauchemin.5@ulaval.ca
 - [Samuel Cabral Cruz](#) - samuelpcabralcruz@gmail.com

Liens utiles

- [Colloque R](#)
- [Taxes de vente Canada](#)
- [Actuar](#)
- [Optim](#)
- [sqldf](#)
- [Loi Normale](#)
- [Loi Gamma](#)
- [Loi Log - normale](#)
- [Loi Weibull](#)
- [Loi Pareto](#)
- [Loi Poisson](#)

