



Évaluation 2

2-06-2023

Titre du cours

Introduction à la valorisation de données

Numéro du cours

420-J04-RO

Programme

LEA.DA Big Data en Finance

Prénom et nom de l'enseignant

Abderrazak Sahraoui

Courriel

asahraoui@crosemont.qc.ca

Consignes Générales

L'évaluation comporte quatre parties.

L'évaluation est à faire en équipe de deux ou trois.

Chargez vos fichiers dans la remise de cette évaluation sur Teams.

L'échéance de la remise est lundi 12 à 17H00.

PARTIE 1

Créer un classeur Excel et le nommer **TransformationsEval3**. Dans ce classeur on voudrait importer le fichier **trucks.csv** et procéder à des transformations par **Power Query** :

	A _C driverid	A _C truckid	A _C model	1 ₂ jun13_miles	1 ₂ jun13_gas	1 ₂ may13_miles	1 ₂ may13_gas
1	A1	A1	Freightliner	9217	1914	8769	
2	A2	A2	Ford	12058	2335	14314	
3	A3	A3	Ford	13652	2899	12075	
4	A4	A4	Kenworth	12687	2439	10680	
5	A5	A5	Hino	10233	1825	14634	
6	A6	A6	Caterpillar	14488	2883	13317	
7	A7	A7	Ford	10938	2231	12080	
8	A8	A8	Navistar	11392	2280	8922	
9	A9	A9	Volvo	12601	2515	9707	
10	A10	A10	Peterbilt	13699	2583	11583	

1. Transformer les colonnes **xxx_miles** et **xxx_gas** en lignes de valeurs

	A _C driverid	A _C truckid	A _C model	A _C Attribut	1 ₂ Valeur
1	A1	A1	Freightliner	jun13_miles	9217
2	A1	A1	Freightliner	jun13_gas	1914
3	A1	A1	Freightliner	may13_miles	8769
4	A1	A1	Freightliner	may13_gas	1892
5	A1	A1	Freightliner	apr13_miles	14234
6	A1	A1	Freightliner	apr13_gas	3008
7	A1	A1	Freightliner	mar13_miles	11519
8	A1	A1	Freightliner	mar13_gas	2262
9	A1	A1	Freightliner	feb13_miles	8676
10	A1	A1	Freightliner	feb13_gas	1596
11	A1	A1	Freightliner	jan13_miles	10025
12	A1	A1	Freightliner	jan13_gas	1878

2. Fractionner la colonne **Attribut** en deux colonnes

	A _C driverid	A _C truckid	A _C model	A _C Attribut.1	A _C Attribut.2	1 ₂ Valeur
1	A1	A1	Freightliner	jun13	miles	9217
2	A1	A1	Freightliner	jun13	gas	1914
3	A1	A1	Freightliner	may13	miles	8769
4	A1	A1	Freightliner	may13	gas	1892
5	A1	A1	Freightliner	apr13	miles	14234
6	A1	A1	Freightliner	apr13	gas	3008
7	A1	A1	Freightliner	mar13	miles	11519
8	A1	A1	Freightliner	mar13	gas	2262
9	A1	A1	Freightliner	feb13	miles	8676

3. Transformer les colonnes **Attribut.2** et **Valeur** en deux colonnes **miles** et **gas**

9	A1	A1	Freightliner	aug12	9754	1967
---	----	----	--------------	-------	------	------

4. Renommer la colonne **Attribut.1** en **Mois**

9	A1	A1	Freightliner	aug12	9754	1967
10	A1	A1	Freightliner	dec09	8795	2642

5. Transformer le tableau pour afficher **totalmiles** et **totalgas** par **driverid**

8	A15	A15			651036	130413
---	-----	-----	--	--	--------	--------

6. Ajouter une colonne pour afficher le quotient de **totalmiles** par **totalgas**

9	A16	A16	638086	129902	4,912056781
---	-----	-----	--------	--------	-------------

7. Trier le tableau par ordre croissant sur **milespergas**

10	A84	A84	651620	146354	4,452355248
----	-----	-----	--------	--------	-------------

• PARTIE 2

Dans le classeur Excel TransformationsEval3, on voudrait importer le fichier geolocation.csv et procéder à des transformations par Power Query :

	truckid	driverid	event	latitude	longitude	city
1	A54	A54	normal	38.440467	-122.714431	Santa Rosa
2	A20	A20	normal	36.977173	-121.899402	Aptos
3	A40	A40	overspeed	37.957702	-121.29078	Stockton
4	A31	A31	normal	39.409608	-123.355566	Willits
5	A71	A71	normal	33.683947	-117.794694	Irvine
6	A50	A50	normal	38.40765	-122.947713	Occidental
7	A51	A51	normal	37.639097	-120.996878	Modesto
8	A19	A19	normal	37.962146	-122.345526	San Pablo

1. Filtrer le tableau pour ne garder que les lignes où event est différent de normal

	truckid	driverid	event	latitude	longitude	city
1	A40	A40	overspeed	37.957702	-121.29078	Stockton
2	A5	A5	unsafe following distance	41.526509	-124.038407	Klamath
3	A48	A48	overspeed	38.752124	-121.288006	Roseville
4	A94	A94	unsafe following distance	37.941325	-122.484704	San Quentin
5	A49	A49	lane departure	39.150171	-123.207783	Ukiah
6	A97	A97	overspeed	38.440467	-122.714431	Santa Rosa
7	A75	A75	lane departure	34.106676	-117.806726	San Dimas
8	A9	A9	unsafe tail distance	32.715329	-117.157255	San Diego
9	A80	A80	unsafe tail distance	37.77493	-122.419416	San Francisco
10	A91	A91	unsafe following distance	38.440467	-122.714431	Santa Rosa

2. Transformer le tableau pour afficher le nombre d'événements dangereux par driverid

	truckid	driverid	NombreEvents
1	A40	A40	2
2	A5	A5	9
3	A48	A48	5
4	A94	A94	6
5	A49	A49	6
6	A97	A97	20
7	A75	A75	6
8	A9	A9	6
9	A80	A80	5
10	A91	A91	2

3. Fusionner le tableau avec le tableau résultant de la partie 1 comme suit :

	A6C truckid	A6C driverid	123 NombreEvents	1.2 trucks.totalmiles	ABC 123 trucks.milespergas
1	A40	A40	2	643126	5,656490497
2	A1	A1	3	628507	4,757740231
3	A5	A5	9	676574	5,197658429
4	A10	A10	6	675377	5,365585674
5	A48	A48	5	631133	5,534458114
6	A100	A100	4	634338	4,931761816
7	A94	A94	6	676750	4,96854053
8	A11	A11	5	652452	5,420747412
9	A49	A49	6	627335	5,522169309
10	A12	A12	5	668241	4,650865459

4. Ajouter une colonne pour afficher le facteur de risque calculé par la formule **nombreEvents * 1000000/totalmiles**.

	A6C truckid	A6C driverid	123 NombreEvents	1.2 trucks.totalmiles	ABC 123 trucks.milespergas	ABC 123 riskfactor
1	A40	A40	2	643126	5,656490497	3,109810519
2	A1	A1	3	628507	4,757740231	4,773216527
3	A5	A5	9	676574	5,197658429	13,30231431
4	A10	A10	6	675377	5,365585674	8,883927051
5	A48	A48	5	631133	5,534458114	7,922260443
6	A100	A100	4	634338	4,931761816	6,305786505
7	A94	A94	6	676750	4,96854053	8,865903214
8	A11	A11	5	652452	5,420747412	7,663398993
9	A49	A49	6	627335	5,522169309	9,5642679
10	A12	A12	5	668241	4,650865459	7,482330477

5. Trier le tableau par ordre décroissant sur **riskfactor**.

	A6C truckid	A6C driverid	123 NombreEvents	1.2 trucks.totalmiles	ABC 123 trucks.milespergas	ABC 123 riskfactor
1	A97	A97	20	631199	4,136351722	31,68572827
2	A73	A73	10	642222	4,845751624	15,57093964
3	A35	A35	9	637069	5,679141	14,12719815
4	A50	A50	9	640069	4,471348036	14,06098405
5	A5	A5	9	676574	5,197658429	13,30231431
6	A95	A95	8	642920	4,29690424	12,44322777
7	A71	A71	8	672026	4,588461013	11,90430132
8	A92	A92	7	618049	5,651973919	11,32596283
9	A86	A86	7	629558	5,140046211	11,118912
10	A44	A44	7	652599	5,001601803	10,7263419

• PARTIE 3

Dans le classeur Excel TransformationsEval3, on voudrait importer le fichier geolocation.csv et procéder à des transformations par Power Query pour focaliser sur le conducteur A97:

1. Filtrer le tableau pour ne garder que les lignes du driverid **A97**

	A6C truckid	A6C driverid	A6C event	A6C latitude	A6C longitude	A6C city
1	A97	A97	overspeed	38.440467	-122.714431	Santa Rosa
2	A97	A97	normal	34.579434	-118.116461	Palmdale
3	A97	A97	normal	39.409608	-123.355566	Willits
4	A97	A97	normal	34.106676	-117.806726	San Dimas
5	A97	A97	normal	38.693723	-119.779683	Markleeville
6	A97	A97	lane departure	34.106676	-117.806726	San Dimas
7	A97	A97	normal	38.440467	-122.714431	Santa Rosa
8	A97	A97	lane departure	37.639097	-120.996878	Modesto
9	A97	A97	normal	37.639097	-120.996878	Modesto
10	A97	A97	normal	36.977173	-121.899402	Aptos
11	A97	A97	unsafe tail distance	34.106676	-117.806726	San Dimas

2. Transformer le tableau pour ne garder que les colonnes ci-dessous :

	AB_C city	AB_C state	1 ² ₃ event_ind	1 ² ₃ idling_ind
1	Santa Rosa	California	1	0
2	Palmdale	California	0	0
3	Willits	California	0	0
4	San Dimas	California	0	1
5	Markleeville	California	0	0
6	San Dimas	California	0	1
7	Santa Rosa	California	0	0
8	Modesto	California	1	0
9	Modesto	California	0	0
10	Aptos	California	0	0

3. Transformer le tableau pour afficher le total des événements dangereux et le total **Idling** par **city**

	AB_C state	AB_C city	1.2 TotalEvents	1.2 TotalIdlings
1	California	Santa Rosa	2	4
2	California	Palmdale	0	0
3	California	Willits	0	0
4	California	San Dimas	2	3
5	California	Markleeville	1	1
6	California	Modesto	2	3
7	California	Aptos	1	2
8	California	Lodi	3	1
9	California	Gilroy	2	0
10	California	Antelope	0	0
11	California	Hollister	0	1
12	California	Arbuckle	0	2
13	California	Oceano	0	2
14	California	San Diego	1	0

• PARTIE 4

On voudrait utiliser le tableau produit en partie 3 et produire la carte (map) correspondante en utilisant l'outil Power BI.

