

Appendice technique

Fusion congruente de PMB/comScore

Le projet de fusion congruente découle de la volonté de réunir les données sur le lectorat de PMB et les données sur l'utilisation du Web de comScore. La fusion congruente a été réalisée avec l'ensemble des données de l'étude PMB Automne 2011 et la série de données sur l'utilisation mensuelle moyenne de l'Internet de comScore pour les mois de juin, juillet et août 2011. D'un point de vue démographique, la fusion des données se limitait dans les deux cas aux utilisateurs âgés de 18 ans et plus ayant navigué sur le Web au cours des quatre (4) dernières semaines. Le fichier PMB comprenait 21 038 répondants et celui de comScore 38 945 panélistes qui auraient visité 6 800 sites Web.

Les variables suivantes ont été utilisées à titre de variables de contrôle brutes.

Région	Atlantique / Québec / Ontario / Prairies / C.-B.
Genre à l'intérieur des régions	Hommes / Femmes
Âge (répartition par genre)	18–24 / 25–34 / 35–54 / 55+

Il en résulte que chaque source de données a été partagée en 40 cellules définies par ces variables de contrôle. En fin de compte, la fusion se fait au niveau de la cellule, c'est-à-dire, la cellule numéro un de PMB se fusionne avec la cellule numéro un de comScore, et ainsi de suite.

L'annexe 1 détaille et compare les profils des utilisateurs d'Internet de comScore et de PMB et chacun d'eux est extrapolé à partir de l'estimation de son propre univers d'utilisateurs d'Internet.

Annexe 1

	Stats Canada 18+		ComScore Profile (Quarter ending Aug 2011)				PMB 2011 Fall			
	(2011 Fall)		Average Month Internet Users 18+				All 18+		Past month Internet Users 18+	
	V%	(000)	Actual	V%	Weighted	V%	(000)	V%	(000)	V%
TOTAL		26,958	38,945	100.0	20,479,186	100.0	26,958	100.0	20,722	100.0
Region										
Atlantic	7.0	1,900	3,140	8.1	1,406,532	6.9	1,900	7.0	1,464	7.1
Quebec	23.5	6,345	6,638	17.0	4,954,867	24.2	6,345	23.5	4,574	22.1
Ontario	38.6	10,418	17,474	44.9	7,937,356	38.8	10,418	38.6	8,066	38.9
Prairies	17.2	4,646	6,772	17.4	3,563,557	17.4	4,646	17.2	3,638	17.6
BC	13.5	3,649	4,921	12.6	2,616,874	12.8	3,649	13.5	2,979	14.4
Gender										
Female	50.9	13,709	16,622	42.7	10,261,678	50.1	13,709	50.9	10,341	49.9
Male	49.1	13,249	22,323	57.3	10,217,508	49.9	13,249	49.1	10,381	50.1
Age										
18 - 24	12.1	3,275	14,272	36.6	2,269,524	11.1	3,275	12.1	3,145	15.2
25 - 34	17.2	4,631	10,568	27.1	4,560,350	22.3	4,631	17.2	4,315	20.8
35 - 44	17.7	4,772	6,500	16.7	4,621,190	22.6	4,772	17.7	4,194	20.2
45 - 54	19.7	5,316	4,369	11.2	4,251,136	20.8	5,316	19.7	4,232	20.4
55 - 64	15.7	4,244	2,114	5.4	3,066,024	15.0	4,244	15.7	3,084	14.9
65+	17.5	4,720	1,122	2.9	1,710,963	8.4	4,720	17.5	1,752	8.5
Household Income										
Under \$15,000	5.7	1,548	467	1.2	612,691	3.0	1,548	5.7	882	4.3
\$15,000-\$24,999	8.2	2,224	1,414	3.6	1,545,385	7.5	2,224	8.2	1,215	5.9
\$25,000-\$39,999	13.4	3,614	3,808	9.8	2,250,343	11.0	3,614	13.4	2,223	10.7
\$40,000-\$59,999	18.2	4,913	13,065	33.5	4,054,398	19.8	4,913	18.2	3,493	16.9
\$60,000-\$74,999	12.9	3,467	11,613	29.8	2,778,621	13.6	3,467	12.9	2,855	13.8
\$75,000-\$99,999	15.7	4,239	5,751	14.8	4,010,767	19.6	4,239	15.7	3,683	17.8
\$100,000+	25.8	6,952	2,827	7.3	5,226,980	25.5	6,952	25.8	6,372	30.7
Household Size										
1 Person	13.6	3,676	4,190	10.8	3,355,572	16.4	3,676	13.6	2,263	11.1
2	33.3	8,966	13,126	33.7	6,956,957	34.0	8,966	33.3	6,396	31.2
3	19.8	5,338	13,299	34.1	4,441,524	21.7	5,338	19.8	4,357	21.3
4	19.4	5,230	4,633	11.9	3,398,069	16.6	5,230	19.4	4,494	21.9
5 or more	13.9	3,749	3,697	9.5	2,327,065	11.4	3,749	13.9	2,971	14.5
Household with Children										
No	65.2	17,569	20,866	53.6	14,574,476	71.2	17,569	65.2	12,400	59.8
Yes	34.8	9,388	18,079	46.4	5,904,710	28.8	9,388	34.8	8,321	40.2

L'annexe 2 détaille les données d'énumération des 40 cellules et met en évidence les différences entre les estimations des deux sources de données indépendantes en extrapolant des projections d'audience Internet par groupes démographiques.

Annexe 2

			PMB 18+ Gen Pop				Jun-Jul-Aug 2011 ComScore Internet			
			f	g	h	i	L	m	m-i	h-m
			Actual	Weighted	Rounded	P4W Internet	Actual	Weighted	Internet Pop. Delta	PMB Overage
			-----	-----	-----	-----	-----	-----	-----	-----
Atlantic	Male	18 to 24	36	108.4991	108,500	106,163	621	81,207	-24956	27293
		25 to 34	47	135.1990	135,204	117,253	439	120,961	3708	14243
		35 to 54	157	341.0638	341,069	275,610	402	246,135	-29475	94934
		55 +	285	332.8317	332,849	189,909	132	126,237	-63672	206612
	Female	18 to 24	45	105.7992	105,801	105,801	514	95,127	-10674	10674
		25 to 34	59	142.8989	142,898	141,376	403	141,282	-94	1616
		35 to 54	234	337.7423	337,752	314,524	474	319,043	4519	18709
		55 +	400	396.4528	396,485	213,434	155	276,504	63070	119981
Quebec	Male	18 to 24	256	358.7969	358,816	350,137	1456	268,526	-81611	90290
		25 to 34	359	548.9969	549,009	507,912	1095	582,649	74737	-33640
		35 to 54	1175	1184.0439	1,184,113	931,150	1030	1,031,649	100499	152464
		55 +	1110	1025.3419	1,025,379	512,816	352	444,102	-68714	581277
	Female	18 to 24	225	343.2978	343,311	330,094	955	329,752	-342	13559
		25 to 34	452	525.9962	526,009	473,782	699	714,413	240631	-188404
		35 to 54	1248	1108.7985	1,108,859	903,789	831	1,093,085	189296	15774
		55 +	1492	1249.6868	1,249,790	564,953	220	490,704	-74249	759086
Ontario	Male	18 to 24	329	657.9957	658,013	642,086	3715	396,375	-245711	497936
		25 to 34	463	873.5942	873,617	825,406	2667	765,747	-59659	536590
		35 to 54	1312	1956.6631	1,956,721	1,663,463	2861	1,934,573	271110	1430476
		55 +	1302	1594.4133	1,594,449	889,970	841	1,024,976	135006	1277332
	Female	18 to 24	278	629.2951	629,309	602,865	2659	386,850	-216015	242459
		25 to 34	560	897.9938	898,016	847,880	1904	716,308	-131572	181708
		35 to 54	1467	1976.7556	1,976,834	1,684,281	2237	1,726,895	42614	249939
		55 +	1575	1830.7201	1,830,795	910,714	590	986,105	75391	844690
Prairies	Male	18 to 24	131	328.9980	329,005	314,186	1455	223,251	-90935	105754
		25 to 34	224	469.4977	469,507	439,014	1143	454,541	15527	14966
		35 to 54	532	892.9122	892,938	746,866	951	822,447	75581	70491
		55 +	544	646.8784	646,905	381,409	285	353,588	-27821	293317
	Female	18 to 24	135	303.6973	303,709	266,455	1074	189,087	-77368	114622
		25 to 34	258	432.7978	432,804	407,699	839	421,865	14166	10939
		35 to 54	572	838.8014	838,829	712,687	842	703,317	-9370	135512
		55 +	634	731.9892	732,015	369,346	183	395,449	26103	336566
BC	Male	18 to 24	127	228.3985	228,402	223,857	1093	160,077	-63780	68325
		25 to 34	171	300.7976	300,811	283,942	805	337,027	53085	-36216
		35 to 54	501	633.0600	633,089	582,755	686	526,245	-56510	106844
		55 +	656	631.0315	631,060	397,285	294	317,117	-80168	313943
	Female	18 to 24	127	210.4987	210,505	203,792	730	139,278	-64514	71227
		25 to 34	188	303.3980	303,411	270,367	574	305,580	35213	-2169
		35 to 54	651	679.2242	679,251	610,725	555	469,143	-141582	210108
		55 +	721	662.7666	662,795	406,575	184	362,190	-44385	300605
			21038	26957.6237	26,958,634	20,722,328	38945	20,479,407		

Le programme breveté de TNS appelé Congruent Fusion (Fusion congruente) a été utilisé pour fusionner les deux ensembles de données. Plutôt que de conserver les valeurs courantes sur un seul ensemble de données comme le feraient des modèles traditionnels de fusion, la fusion congruente est conçue pour maintenir les valeurs courantes des deux ensembles de données.

Comme les annexes 1 et 2 ont permis de le constater, les prévisions des deux sources de données quant aux estimations du nombre d'utilisateurs Internet sont légèrement différentes. Le *modèle de fusion congruente* intègre la fusion forcée et requiert que les estimations pondérées de l'univers soient identiques à celles au niveau des cellules. De ce fait, avant d'exécuter la fusion, le modèle de fusion congruente intègre la technologie d'imputation de TNS. Fondamentalement, le modèle d'imputation attribue les valeurs propres aux utilisateurs d'Internet d'un fichier à un autre créant ainsi une correspondance parfaite avant la fusion. Dans ce cas-ci, les estimations de PMB ont été attribuées de manière à ce qu'elles correspondent à celles de comScore; les données ont ensuite été fusionnées en préservant le total pondéré de la source initiale, cellule par cellule.

Dans la plupart des cas, les données ont été fusionnées séparément à l'intérieur de chacune des 40 cellules de contrôle. Cependant, pour plusieurs cas, l'extrapolation des estimations de comScore a donné comme résultat une plus grande proportion d'utilisateurs Internet que PMB n'avait estimée. L'imputation ne peut pas corriger de tels écarts, c'est-à-dire qu'elle ne peut pas créer des utilisateurs Internet lorsqu'il y a un manque au niveau des projections démographiques. Pour surmonter ce problème, chaque cellule touchée doit être unifiée à la cellule adjacente en suivant l'échelle d'âge. Conséquemment, les cellules unifiées suivantes ont été créées avant la mise en œuvre du modèle.

Québec — Hommes 18–24 et 25–34 ans
Québec — Femmes 25–34, 35–54 et 55+
C.-B. — Hommes 18–24 et 25–34 ans
C.-B. — Femmes 18–24 et 25–34 ans

En plus des trois variables de contrôle brutes, l'algorithme de la fusion a aussi trouvé des correspondances dans les fichiers sources et cibles sur certaines variables qu'on appelle des variables de contrôle souples dont :

Revenu du foyer	< 15 000 \$
	15 000 \$ – 24 999 \$
	25 000 \$ – 39 999 \$
	40 000 \$ – 59 999 \$
	60 000 \$ – 74 999 \$
	75 000 \$ – 99 999 \$
	> 100 000 \$+
Taille du foyer	1 personne / 2 / 3 / 4 ou plus
Foyer avec enfants	Oui / Non
Âge (cellules unifiées seulement)	18–24 / 25–34 / 35–54 / 55+

Au sein du programme de fusion, les fichiers sources et cibles ont été filtrés sur les variables de contrôle souples, et des points de pénalités sont calculés pour chaque paire de fichiers sources et cibles. La variable du revenu familial s'est vu allouer le plus haut niveau d'importance dans la grille de points. Le couplage parfait de toutes les variables de contrôle souples implique qu'aucun point de pénalité n'a été alloué. Itérativement, le programme de fusion recherche la solution qui permette d'allouer le moins de points de pénalité possible.

La base de données fusionnée comprend les fichiers de 59 897 répondants. Afin de maintenir la taille du fichier à un niveau gérable, les sites Web ayant moins de 50 000 visiteurs uniques ont été supprimés. De ce fait, 6,800 sites Web uniques sont compris dans l'ensemble final de données.

Les dénombrements actuels et pondérés par variable démographique sont détaillés dans le tableau en annexe.

22 décembre 2011

200101071