

## Ułożenie listy indywidualnych profili demograficznych na potrzeby drugiego losowania

### I. Wprowadzenie

1. Naszym celem jest ułożenie listy 75 indywidualnych profili demograficznych panelistów i panelistek w taki sposób, aby była ona maksymalnie zbliżona do składu idealnego lub w pełni z nim zgodna. Profile demograficzne obejmują kryteria: płeć i wiek (traktowane tutaj łącznie), wykształcenie, transport oraz osiedle.
2. Na potrzeby ustalenia listy indywidualnych profili demograficznych korzystamy z dwóch zbiorów danych:
  - 1) "Skład idealny" (A) - zawierający dane o liczbie cech wyodrębnionych w ramach każdego z kryteriów demograficznych, które powinni posiadać paneliści/cki, aby struktura panelu odzwierciedlała strukturę demograficzną miasta;
  - 2) "Profile w bazie danych" (B) - zawierający dane o cechach osób, które zarejestrowały się do udziału w panelu - w czterech wymienionych wyżej kryteriach demograficznych.
3. Dla każdej cechy sprawdzamy liczbę osób w zbiorach A i B. Dla przedstawienia metody, na próbie zarejestrowanych 547 osób (przed upływem terminu rejestracji), przedstawia się to następująco:

Płeć i wykształcenie (PW)	W bazie (B)	Skład idealny (A)	Transport (T)	W bazie (B)	Skład idealny (A)	Wykształcenie (W)	W bazie (B)	Skład idealny (A)
K 18-24	26	2	auto	191	31	średnie	183	42
K 25-39	90	11	zbiorkom	276	21	wyższe	352	25
K 40-64	91	16	pieszo	28	18	podstawowe	12	8
K 65+	58	12	rower	52	5			
M 18-24	24	2						
M 25-39	106	10						
M 40-64	102	14						
M 65+	50	8						

Osiedle	W bazie	Skład idealny	Osiedle	W bazie	Skład idealny
Bieńkowice	0	1	Nowy Dwór	10	2
Biskupin - Sępolno - Dąbie - Bartoszewice	23	2	Ołbin	15	3
Borek	17	1	Oftaszyn	9	1
Brochów	7	1	Oporów	14	1
Gaj	28	2	Osobowice - Rędzin	2	1
Gajowice	14	3	Pawłowice	3	1
Gądów - Popowice Płd.	31	3	Pilczyce - Kozanów - Popowice Płn.	35	3
Grabiszyn - Grabiszynek	6	2	Plac Grunwaldzki	3	1
Huby	16	2	Polanowice - Poświętne - Ligota	5	1
Jagodno	5	1	Powstańców Śląskich	25	3
Jerzmanowo - Jarnołtów - Strachowice - Osiniec	0	1	Pracze Odrzańskie	2	1
Karłowice - Różanka	34	3	Przedmieście Oławskie	28	2
Klecina	4	1	Przedmieście Świdnickie	14	2
Kleczków	6	1	Psie Pole - Zawidawie	22	3
Kowale	3	1	Sołtysowice	7	1
Krzyki - Partynice	21	2	Stare Miasto	7	1
Księżę	4	1	Strachocin - Swojczyce - Wojnów	6	1
Kuźniki	3	1	Szczepin	30	2
Leśnica	22	3	Świniary	0	1
Lipa Piotrowska	4	1	Tarnogaj	7	1
Małlice	9	1	Widawa	3	1
Muchobór Mały	11	1	Wojszyce	7	1
Muchobór Wielki	10	1	Zacisze - Zalesie - Szczytniki	4	1
Nadodrze	9	3	Żerniki	2	1

## II. Opis metody tworzenia profili

1. Przyjęta przez nas metoda wyboru profili opiera się na tym, że w ramach każdego z kryteriów demograficznych losujemy po kolei profile posiadające poszczególne wyodrębnione w tych kryteriach cechy (np. kryterium - wykształcenie, cecha - podstawowe).
2. Jeżeli na liście profili znajduje się już n profili posiadających cechę X, a profili posiadających inną cechę z tego kryterium (Z) jest n-1 albo mniej, kolejny profil losujemy spośród profili posiadających cechę Z.

*Przykład: w składzie rzeczywistym mamy po 3 profile z wykształceniem podstawowym i średnim oraz 2 profile z wykształceniem wyższym. W kolejnym kroku profile losujemy więc spośród profili z wykształceniem wyższym.*

Po wypełnieniu w ten sposób wszystkich miejsc przypadających na daną cechę w składzie idealnym, opisaną zasadę doboru cech stosujemy do pozostałych cech z danego kryterium.

*Przykład: jeżeli w składzie rzeczywistym mamy już 12 profili z wykształceniem podstawowym, kolejne losowania przeprowadzamy już tylko spośród profili posiadających wykształcenie wyższe albo średnie.*

- Sortowanie profili demograficznych i oznaczanie cech na potrzeby wyboru profili wykonujemy przy pomocy programu Excel. Losowanie numerów przyporządkowanych do poszczególnych wykonujemy przy użyciu strony [www.random.org](http://www.random.org).
- Ze wszystkich kryteriów najmniej cech ma kryterium W (cechy: wyższe, średnie i podstawowe). Jest więc prawdopodobne, że w zbiorze B wśród profili posiadających te cechy znajdą się profile posiadające wszystkie cechy z kryteriów PW, T i O. Jedynie profile z wykształceniem podstawowym jest niewiele, jednak po wypełnieniu miejsc przypadających na tę cechę, wybór osób będzie przeprowadzany z obszernych cech "średnie" i "wyższe". Dlatego w kolejnych krokach losujemy profile spośród profili posiadających kolejne cechy w kryterium W. Jednocześnie przyjęliśmy następującą kolejność uzupełniania cech w kryterium W: podstawowe, średnie i wyższe - tak, aby najpierw uzupełnić cechy, które posiada najmniejsza liczba profili w zbiorze B.
- Aby zachować zasadę określoną w pkt II.1, wylosowawszy profil posiadający daną cechę w innym kryterium niż W, oznaczamy tę cechę jako czasowo wykorzystaną - w kolejnym kroku losujemy zaś profil spośród profili nieposiadających cech oznaczonych jako czasowo wykorzystane.

*Przykład: w pierwszym kroku losowaliśmy profil spośród tych posiadających wykształcenie podstawowe i wylosowaliśmy profil z cechą: kobieta, 18-24 lata. W drugim kroku losujemy więc profil spośród tych z wykształceniem średnim, z wyłączeniem kobiet w wieku 18-24 lata.*

- Kiedy w ramach danego kryterium wylosujemy już tyle samo profili posiadających każdą z cech, z powrotem usuwamy oznaczenie tych cech jako czasowo wykorzystane. W kolejnym kroku losujemy więc spośród profili posiadających którąkolwiek z cech z danego kryterium (z wyłączeniem cech trwale wykorzystanych - patrz pkt. II.6).

*Przykład: w krokach od pierwszego do trzeciego wylosowaliśmy profile z cechami: auto, zbiorkom, rower i wszystkie te cechy oznaczyliśmy jako wykorzystane. W czwartym kroku losujemy tylko spośród profili z cechą "pieszo". Po wylosowaniu czwartego profilu, w składzie rzeczywistym mamy po jednym profilu posiadającym każdą z cech w kryterium T. Oznaczamy więc cechy: auto, zbiorkom i rower jako niewykorzystane, a w kroku piątym losujemy spośród profili posiadających którąkolwiek z tych cech.*

- Kiedy podczas układania listy profili dla danej cechy wypełnione zostaną wszystkie miejsca przypadające na nią w składzie idealnym, oznaczamy tę cechę jako trwale wykorzystaną. Losując kolejne profile do listy, pomijamy więc profile posiadające tę cechę.

*Przykład: w krokach od pierwszego do dwudziestego uzupełniliśmy po 5 miejsc dla każdej z cech z kryterium T. Na cechę "rower" w składzie idealnym przypada 5 miejsc. Dlatego w kroku dwudziestym pierwszym i kolejnych nie losujemy już profili posiadających cechę "rower".*

### **III. Losowanie spośród profili posiadających cechy oznaczone jako czasowo wykorzystane**

Może okazać się, że wśród profili z określoną cechę w kryterium W nie ma profili nieposiadających cech oznaczonych jako wykorzystane. Wtedy kolejny profil losujemy spośród profili posiadających takie cechy (robimy więc wyjątek od zasady przyjętej w pkt II.1). Pomijając wykorzystanie cech stosujemy łącznie wymienione poniżej zasady.

1. Losujemy spośród profili posiadających mniej oznaczonych cech.

*Przykład: losujemy spośród profili z wykształceniem podstawowym. W 10-tym kroku w grupie tej znajduje się 7 profili posiadających trzy oznaczone cechy oraz 5 profili z dwoma oznaczonymi cechami. Losujemy więc spośród 5 profili posiadających 2 oznaczone cechy;*

2. Jeżeli wśród profili z tą samą liczbą oznaczonych cech znajdują się profile z cechami oznaczonymi w różnych kryteriach, stosujemy następującą kolejność pomijania wykorzystania cech w kryteriach: PW, T, O.

*Przykład: przeprowadzamy opisane wyżej losowanie. Wśród 5 profili posiadających 2 oznaczone cechy są następujące kombinacje cech: 3 profile z oznaczonymi cechami z kryteriów PW i O oraz 2 oznaczonymi cechami z kryteriów T i O. Losujemy więc spośród profili posiadających oznaczone cechy z kryteriów PW i O, ponieważ PW ma pierwszeństwo przed T.*

Przyjęta kolejność jest uzasadniona tym, że:

- na poszczególne cechy w kryterium O przypada najmniejsza liczba miejsc w składzie idealnym (1-3 miejsca). Przydzielanie poszczególnym cechom miejsc w składzie rzeczywistym poza kolejnością zwiększa więc ryzyko, że nie znajdą się w nim w ogóle profile reprezentujące najmniej popularne cechy. Dlatego fakt wykorzystania cechy z kryterium O pomijamy w ostatniej kolejności;
  - liczba cech w kryterium T (4) jest z kolei znacznie mniejsza niż cech w kryterium PW (8). Dlatego, jeżeli pominiemy w pierwszej kolejności wykorzystanie cechy z kryterium PW, posiadamy większą liczbę cech, którym możemy przydzielać miejsca poza kolejnością.
3. Jeżeli jednak po wytypowaniu cechy zgodnie z kolejnością określoną w pkt III.2 okaże się, że liczba miejsc w składzie rzeczywistym przypadających na nią poza kolejnością byłaby większa niż liczba takich miejsc przypadających na cechę oznaczoną jako wykorzystaną w kolejnym kryterium, pomijamy wykorzystanie cechy z tego drugiego kryterium.

*Przykład: przeprowadzamy opisane wyżej losowanie. Po wytypowaniu zgodnie z pkt III.2 cechy z kryterium PW okazuje się, że cecha ta już w poprzednim kroku uzyskała dodatkowe miejsce. Łącznie otrzymałaby więc już drugie miejsce poza kolejnością. Tymczasem cecha z kryterium T otrzymałaby dopiero pierwsze dodatkowe miejsce. Losujemy więc spośród profili posiadających oznaczone cechy z kryteriów T i O.*

4. Jeżeli trzeba dokonać wyboru spośród profili posiadających oznaczone cechy w tych samych dwóch kryteriach, a liczba miejsc w składzie rzeczywistym przypadających poza kolejnością jest taka sama dla ich cech z różnych kryteriów, losujemy spośród tych profili, gdzie większa

liczba dodatkowych miejsc przypadaby na cechę z kryterium mającego pierwszeństwo zgodnie z pkt III.2.

*Przykład: przeprowadzamy opisane wyżej losowanie. Wśród 5 profili posiadających dwie oznaczone cechy wszystkie mają oznaczone cechy z kryteriów PW i O. W dwóch z nich oznaczone cechy z kryterium PW już w poprzednich krokach uzyskały dodatkowe miejsce, a cechy z kryterium O otrzymałyby dopiero pierwsze dodatkowe miejsce. W trzech pozostałych oznaczone cechy z kryterium PW otrzymałyby zaś pierwsze dodatkowe miejsce, a cechy z kryterium O drugie dodatkowe miejsce. Losuję więc spośród pierwszej grupy profili, ponieważ kryterium PW zgodnie z pkt III.2 ma pierwszeństwo względem kryterium O.*

5. Jeżeli profile posiadają oznaczone różne cechy w ramach jednego kryterium, losujemy spośród tych profili, w które posiadają cechy, które otrzymały mniejszą liczbę miejsc w składzie rzeczywistym poza kolejnością.

*Przykład: przeprowadzamy opisane wyżej losowanie. Wśród 5 profili posiadających dwie oznaczone cechy wszystkie mają oznaczone cechy z kryteriów PW oraz O. We wszystkich cecha z kryterium O otrzymałyby pierwsze dodatkowe miejsce. W dwóch z nich oznaczona jest jednak cecha "kobieta, 18-24", której zostało przydzielone już dodatkowe miejsce, a w trzech cecha "kobieta, 25-39", która otrzymałyby pierwsze dodatkowe miejsce. Losujemy więc spośród drugiej grupy profili.*

#### **IV. Uzupełnienie brakujących miejsc na liście profili**

Stosując opisane wyżej zasady, na próbie 547 profili ze zbioru B udało się nam zapełnić 66 miejsc na liście profili indywidualnych (88% listy profili). W kroku 67 okazuje się, że nie ma już profili, które nie posiadałyby cech oznaczonych jako trwale wykorzystane. Dlatego w celu uzupełnienia 9 ostatnich miejsc wykonujemy opisane poniżej kroki.

1. Liczymy różnicę pomiędzy miejscami przypadającymi na dane cechy w składzie idealnym, a miejscami, które otrzymały już w składzie rzeczywistym.
2. Sprawdzamy, dla których cech z kryterium PW różnica z pkt IV.1 jest największa, a następnie losujemy brakujące profile spośród profili ze zbioru B posiadających tę cechę (a w kolejnych krokach spośród profili posiadających kolejne cechy z kryterium PW. Ze względu na to, że każdy z profili posiada już inną trwale oznaczoną cechę, musimy pomijać ich wykorzystanie. Stosujemy wtedy następujące zasady:
  - losujemy spośród profili posiadających mniej oznaczonych cech - *przykład: losujemy spośród profili z cechą "kobieta, 25-29". W grupie tej znajduje się 5 profili posiadających trzy oznaczone cechy oraz 3 profile z dwoma oznaczonymi cechami. Losujemy więc spośród 3 profili posiadających 2 oznaczone cechy;*
  - jeżeli wśród profili z tą samą liczbą oznaczonych cech znajdują się profile z cechami oznaczonymi w różnych kryteriach, stosujemy następującą kolejność pomijania wykorzystania cech w kryteriach: O, W i T - *przykład: losujemy spośród profili z cechą "kobieta, 40-64". W grupie tej znajdują się 4 profile z oznaczoną cechą z kryterium O, 3 z kryterium W i 2 z kryterium T. Losujemy więc spośród 2 profili z oznaczoną cechą z kryterium O;*
  - jednocześnie sprawdzamy, które z profili posiadających daną cechę z kryterium PW posiadają cechy w innych kryteriach, dla których możliwość zapełnienia miejsc skończyła się wcześniej

oraz losujemy spośród tych profili - *przykład: przeprowadzamy opisane wyżej losowanie. Wśród 4 profili z oznaczoną cechą z kryterium O dwa mają cechę "auto", a dwa cechę "pieszo", dla której możliwość zapewnienia miejsc skończyła się wcześniej. Losujemy więc spośród drugiej grupy profili.*

Przyjęta kolejność uzupełniania brakujących profili uzasadniona jest tym, jakie kryteria uważamy za priorytetowe - za priorytetowe przyjęliśmy kolejno kryteria: PW, T, W i O.

- 3. Stosując opisane wyżej zasady, na próbie 547 profili udało się zapłacić miejsca w składzie rzeczywistym: w kryterium PW - zgodnie ze składem idealnym, w kryterium T - różnica 1 profilu (17 cech "rower" i 32 cech "auto" zamiast odpowiednio 18 i 31), w kryterium W różnica 1 profilu 26 cech "wyższe" i 41 cech "średnie" zamiast odpowiednio 25 i 24), w kryterium O - różnica 5 profili.**