

Analiza poziomów wagi wśród obywateli Meksyku, Peru oraz Kolumbii

Szymon Bańczyk

Spis treści

[Spis rysunków 2](#_heading=h.gjdgxs)

[Wstęp 3](#_heading=h.30j0zll)

[Pytania badawcze 3](#_heading=h.1fob9te)

[Cele analizy 3](#_heading=h.3znysh7)

[Hipotezy badawcze 3](#_heading=h.2et92p0)

[Źródło danych 3](#_heading=h.tyjcwt)

[Obróbka danych 5](#_heading=h.1t3h5sf)

[Analiza 6](#_heading=h.17dp8vu)

[Podsumowanie i wnioski 12](#_heading=h.1y810tw)

# Spis rysunków

[Rysunek 1 Typy kolumn 4](#_heading=h.3dy6vkm)

[Rysunek 2 „Surowe” dane przed obróbką 5](#_heading=h.4d34og8)

[Rysunek 3 Dane po obróbce 5](#_heading=h.2s8eyo1)

[Rysunek 4 Dane z podziałem na płeć 6](#_heading=h.3rdcrjn)

[Rysunek 5 Nadwaga w historii rodziny a grupa wagowa 7](#_heading=h.26in1rg)

[Rysunek 6 Wiek i płeć a waga 8](#_heading=h.lnxbz9)

[Rysunek 7 Alkohol a rodzaj wagi 8](#_heading=h.35nkun2)

[Rysunek 8 Rodzaj wagi a monitorowanie kalorii 9](#_heading=h.1ksv4uv)

[Rysunek 9 Kolumna CzyĆwiczy 9](#_heading=h.44sinio)

[Rysunek 10 Aktywność fizyczna a poprawna waga 10](#_heading=h.2jxsxqh)

[Rysunek 11 Środek transportu a średnia waga 11](#_heading=h.z337ya)

[Rysunek 12 Palenie a zaburzenia wagi 11](#_heading=h.3j2qqm3)

# Wstęp

Niniejszy raport jest projektem zaliczeniowym przedmiotu „Raportowanie i wizualizacja” prowadzonym przez dr hab. Krzysztofa Kania na Uniwersytecie Ekonomicznym w Katowicach. Przedstawia on analizę poziomów wagi w niektórych krajach Ameryki Południowej. Problem nieodpowiedniej wagi może prowadzić do różnych problemów zdrowotnych. W analizie skupiono się na poszukiwaniu czynników innych niż te związane z odżywianiem prowadzących do zaburzeń wagi. Analiza wraz z obróbką danych została przeprowadzona w programie Microsoft Power BI, a raport w programie Microsoft Word. Autorem raportu jest Szymon Bańczyk.

# Pytania badawcze

Raport odpowiada na pytania, które zostały zadane po zaznajomieniu się z danymi:

* Jaki jest rozkład rodzaju wagi w zależności płci?
* U ilu osób z nadwagą w historii rodziny pojawiła się nadwaga?
* Jaki jest rozkład rodzaju wagi w zależności od wieku?
* Jak spożywanie alkoholu wpływa na rodzaj wagi?
* Jak często osoby z poprawną wagą uprawiają sport?
* Jaka jest średnia waga osób korzystających z poszczególnych środków transportu?

# Cele analizy

* Wizualna reprezentacja danych zawartych w zbiorze,
* Przedstawienie czynników wpływających na rodzaj wagi,
* Określenie, czy wraz z wiekiem ilość osób z nadwagą wzrasta.

# Hipotezy badawcze

Założone zostało, że:

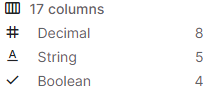
* Płeć nie wpływa na ilość osób w danej kategorii wagi.
* Rodzina z nadwagą w historii wpływa na posiadanie nadwagi.
* Monitorowanie kalorii wpływa na wagę.
* Aktywność fizyczna jest kluczowym czynnikiem, by waga była w normie.
* Palenie wpływa na zaburzenia wagi.

# Źródło danych

Dane użyte w raporcie zostały pobrane z kaggle.com[[1]](#footnote-0). Autorem zbioru jest Fatemeh Mehrparvar. Zawiera on dane przeprowadzone w ankiecie na terenie Meksyku, Peru oraz Kolumbii. Każdy wiersz w zbiorze reprezentuje 1 z 2111 osób, które wzięły udział w badaniu dotyczącym wagi. Kolumny od 1 do 16 reprezentują natomiast odpowiedź na jedno pytanie zadane w ankiecie. Kolumna 17 to rodzaj wagi, obliczony za pomocą BMI. Dane zostały udostępnione na licencji Attribution 4.0 International (CC BY 4.0)[[2]](#footnote-1), co pozwala na ich użytek w raporcie.

Dane zawierają następujące kolumny:

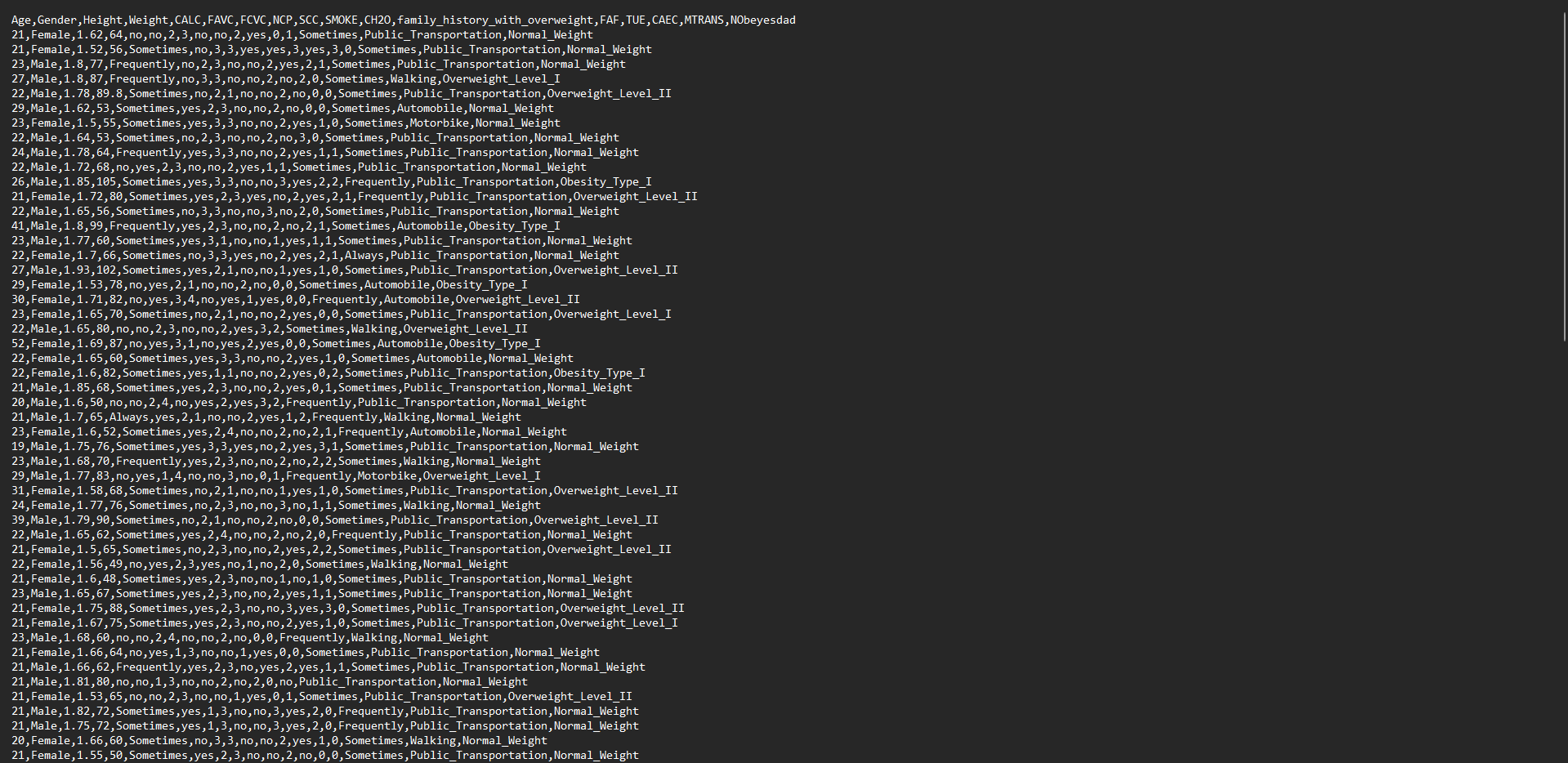
1. Gender: Feature, Categorical, "Gender"
2. Age : Feature, Continuous, "Age"
3. Height: Feature, Continuous
4. Weight: Feature Continuous
5. family\_history\_with\_overweight: Feature, Binary, " Has a family member suffered or suffers from overweight? "
6. FAVC : Feature, Binary, " Do you eat high caloric food frequently? "
7. FCVC : Feature, Integer, " Do you usually eat vegetables in your meals? "
8. NCP : Feature, Continuous, " How many main meals do you have daily? "
9. CAEC : Feature, Categorical, " Do you eat any food between meals? "
10. SMOKE : Feature, Binary, " Do you smoke? "
11. CH2O: Feature, Continuous, " How much water do you drink daily? "
12. SCC: Feature, Binary, " Do you monitor the calories you eat daily? "
13. FAF: Feature, Continuous, " How often do you have physical activity? "
14. TUE : Feature, Integer, " How much time do you use technological devices such as cell phone, videogames, television, computer and others? "
15. CALC : Feature, Categorical, " How often do you drink alcohol? "
16. MTRANS : Feature, Categorical, " Which transportation do you usually use? "
17. NObeyesdad : Target, Categorical, "Obesity level"



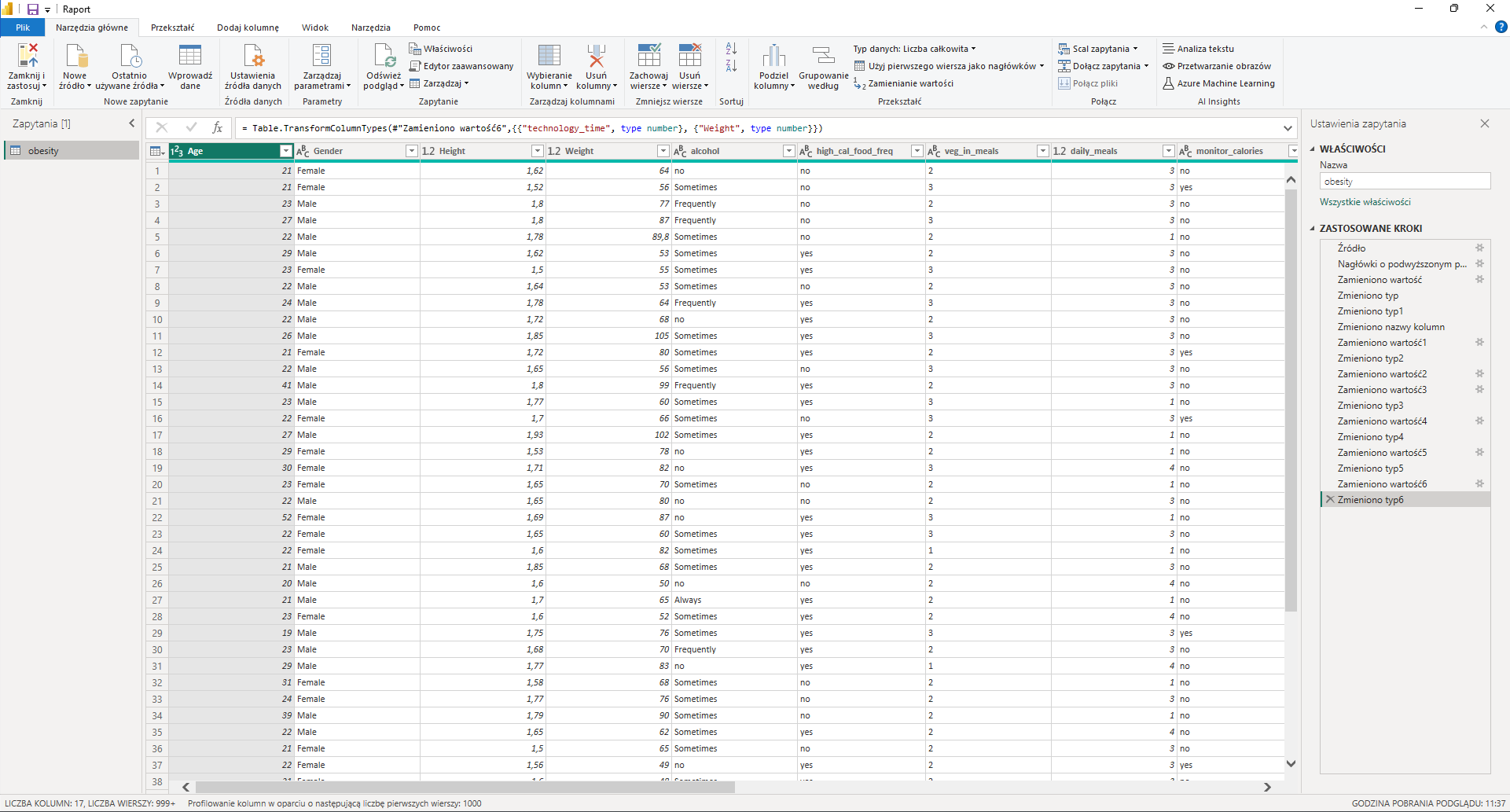
*Rysunek 1 Typy kolumn*

# Obróbka danych

Po wczytaniu danych do Power BI zauważono, że w niektórych kolumnach, w których powinny być tylko liczby całkowite, znajdują się liczby dziesiętne. Dla kolumn, w których liczby powinny być całkowite, postanowiono o „odcięciu” cyfr od znaku kropki w prawo. Z kolei dla liczb dziesiętnych zmieniono format na europejski, tak by można je było przekonwertować na liczby dziesiętne w programie Power BI. Kolejnym krokiem było zmienienie nazw kolumn na takie, które byłyby zrozumiane dla autora, np. „CALC” zmieniono na „alcohol”.



*Rysunek 2 „Surowe” dane przed obróbką*

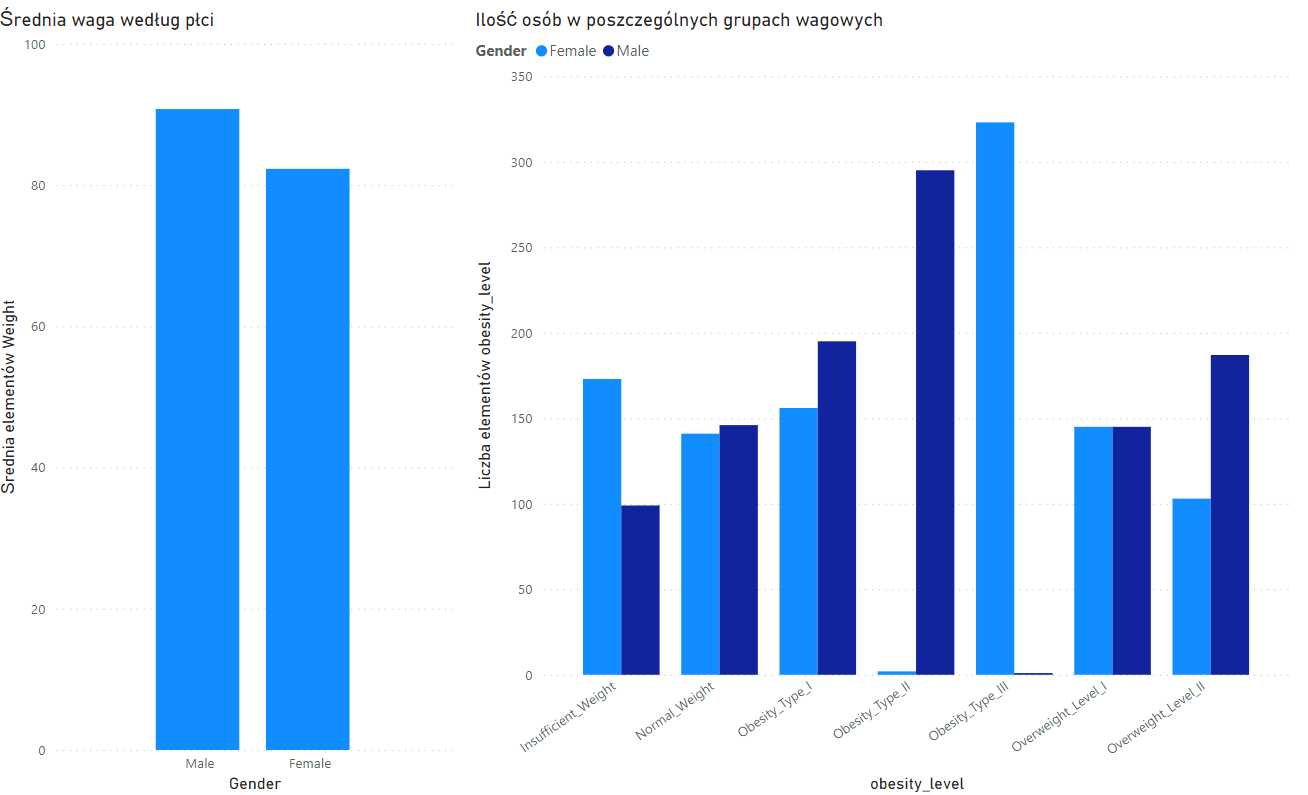


*Rysunek 3 Dane po obróbce*

# Analiza

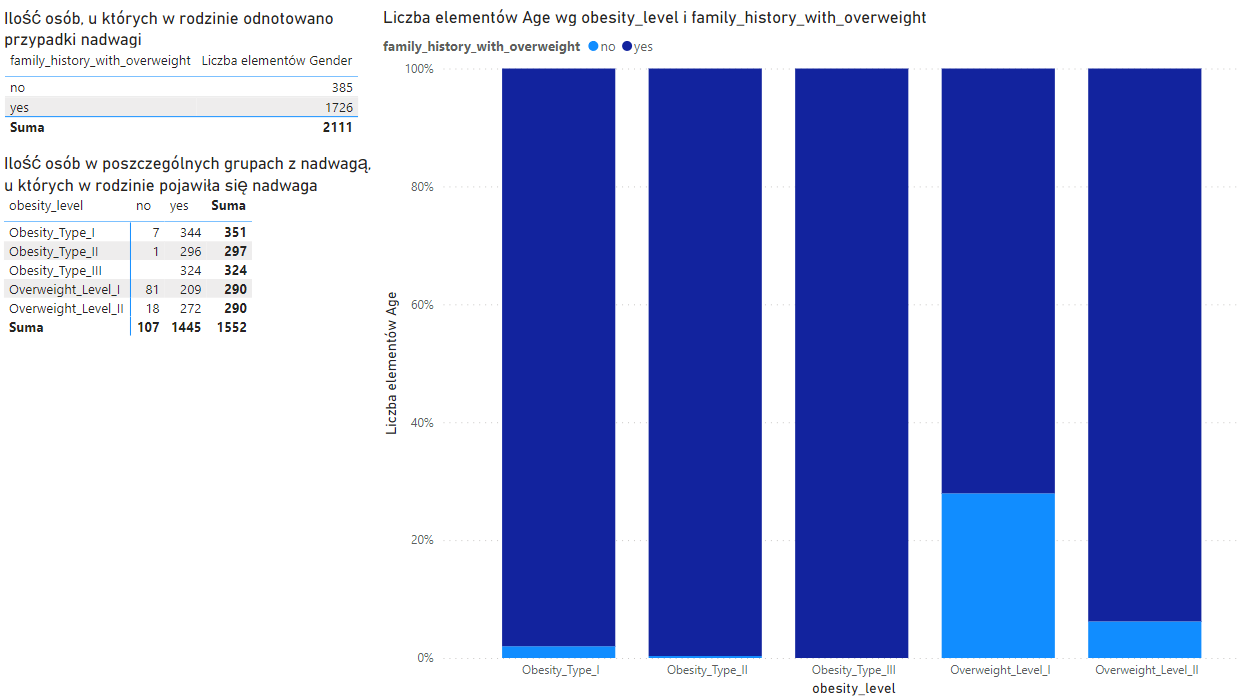
W celu analizy danych powstało łącznie 8 dashboardów.

Pierwszy z nich reprezentuje rozkład średniej wagi z podziałem na płeć oraz ilość osób w poszczególnych grupach wagowych.



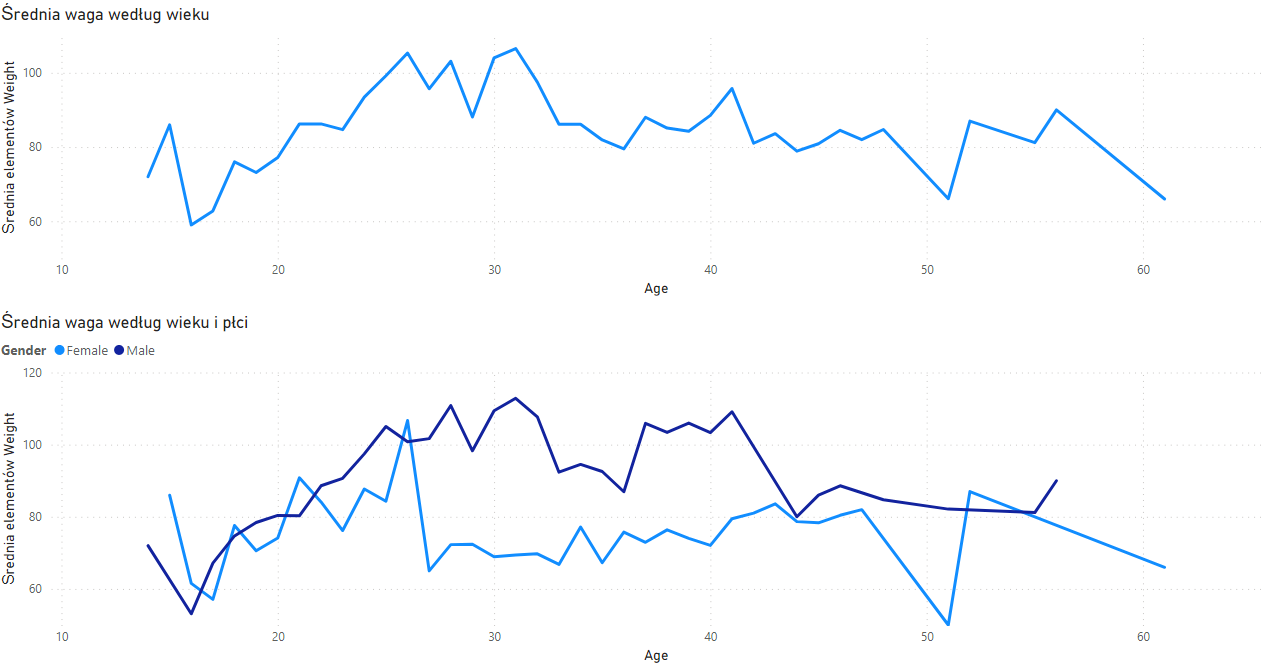
*Rysunek 4 Dane z podziałem na płeć*

Na wykresie po lewej stronie można zauważyć, że średnia dla obu płci jest prawie że równa, co może świadczyć o tym, że płeć nie ma znaczenia w przypadku wagi. Warto jednakże zauważyć niektóre dysproporcje na wizualizacji po prawej stronie. Zdecydowana większość osób z otyłością stopnia drugiego jest płci męskiej, natomiast zdecydowana większość osób z otyłością stopnia trzeciego jest płci żeńskiej. Poza tym widać, że prawie dwukrotnie więcej osób z prawidłową wagą to kobiety. Reszta kategorii jest niemalże równa, a odchylenia mogą wynikać z nierównej ilości badanych osób danej płci.



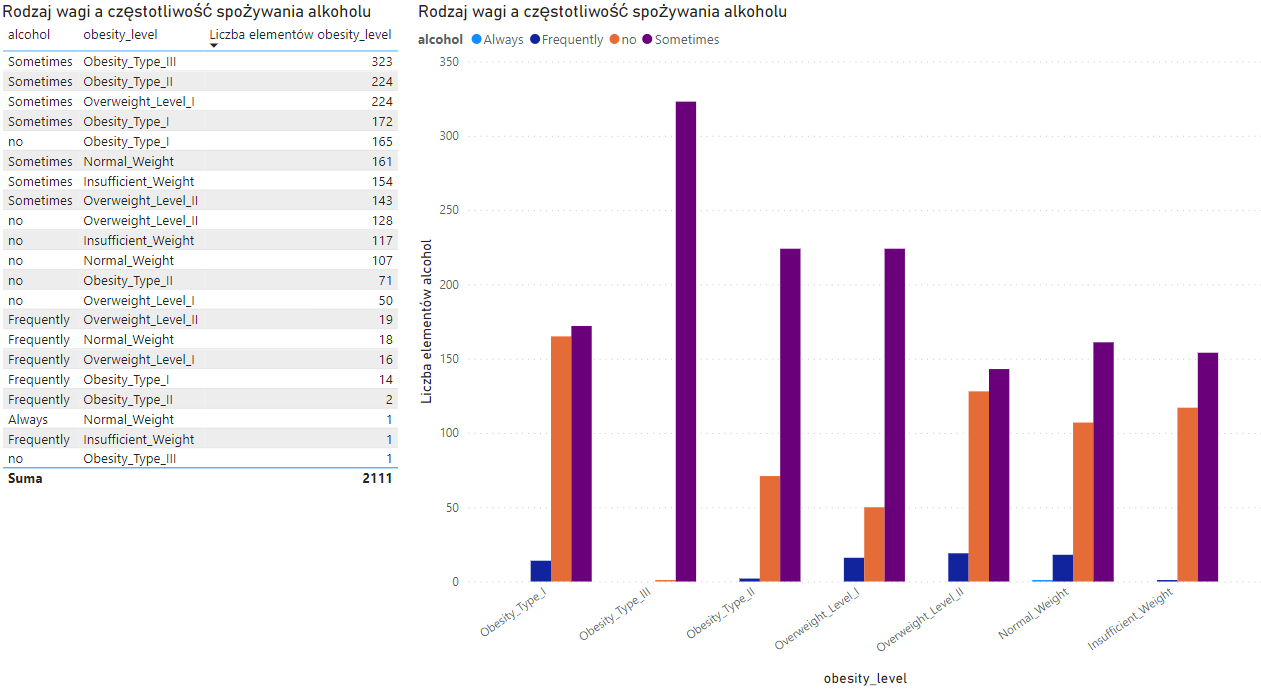
*Rysunek 5 Nadwaga w historii rodziny a grupa wagowa*

Drugi panel przedstawia rozkład osób, u których w rodzinie odnotowano nadwagę w przeszłości. Jak widać zdecydowana większość osób z nadwagą lub otyłością sama posiada nadwagę lub otyłość. Dla otyłości poziomu drugiego i trzeciego, to właściwie pewnik, że gdy w historii rodziny odnotowano nadwagę, te osoby również są zagrożone tą chorobą. Można bezpiecznie stwierdzić, że gdy w rodzinie odnotowano nadwagę, to niższe pokolenia również dotyka ten problem.



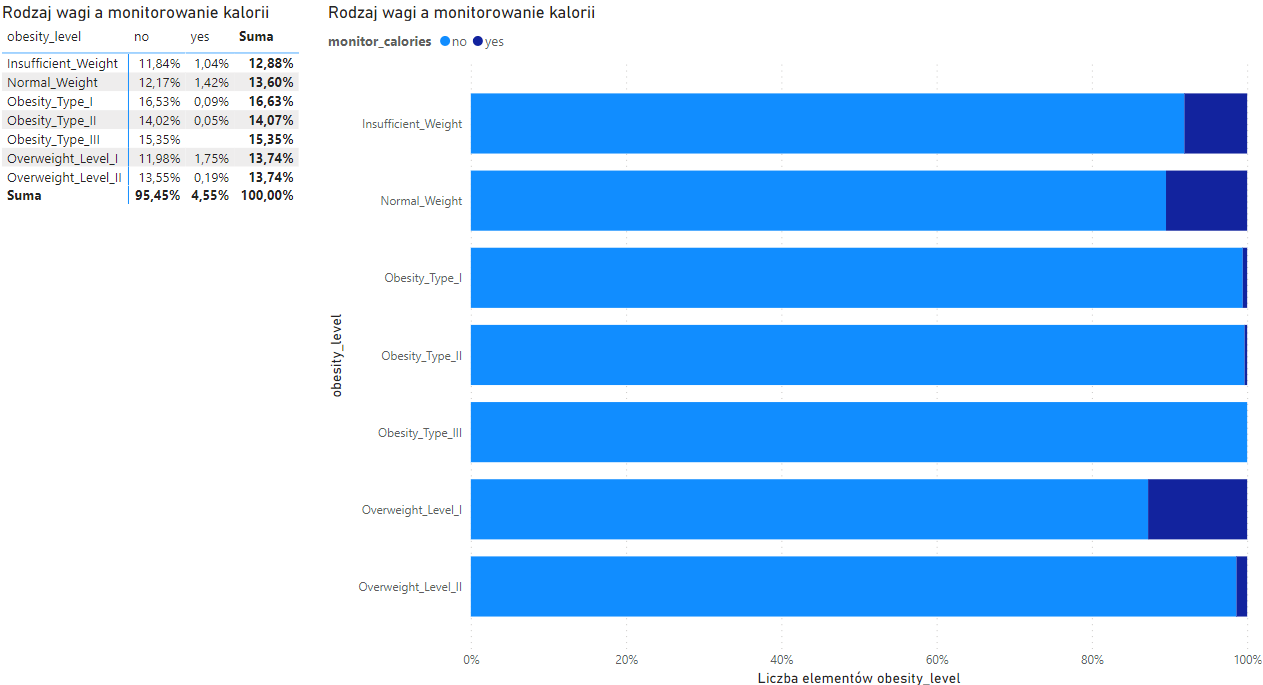
*Rysunek 6 Wiek i płeć a waga*

Na powyższej wizualizacji można zobaczyć jak kształtuje się średnia waga. Niezależnie od płci, najwięcej kilogramów mają osoby w wieku 26-31 lat. Wykres poniżej wizualizuje średnią wagę w zależności od wieku oraz płci. Przez całą długość życia kobiety mają średnio niższą wagę niż mężczyźni. Najprawdopodobniej jest to spowodowane czynnikami genetycznymi i biologicznymi.



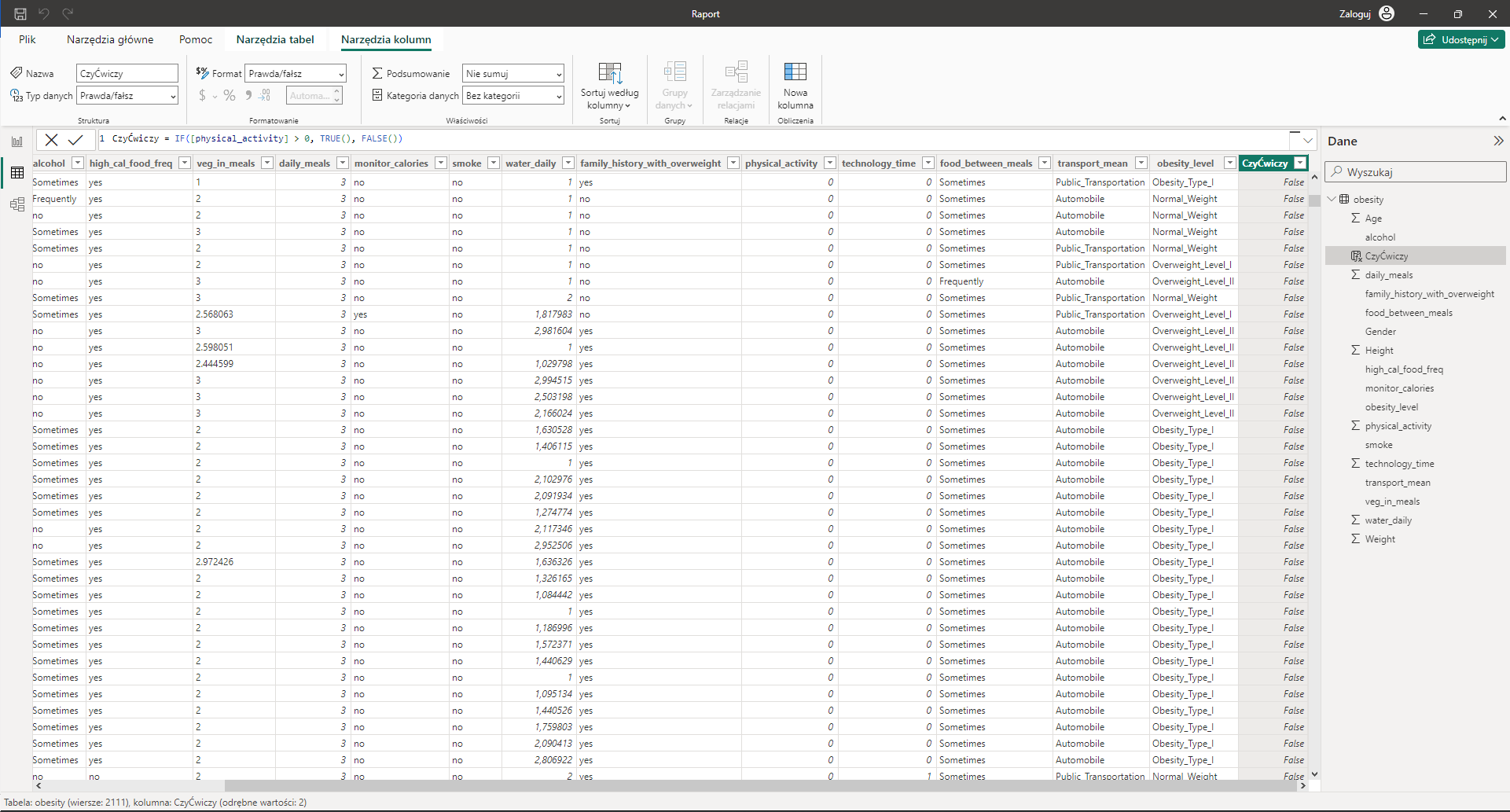
*Rysunek 7 Alkohol a rodzaj wagi*

Czwarty dashboard obrazuje jak częstotliwość spożywania alkoholu wpływa na rodzaj wagi. Można zauważyć, że w osoby z każdego rodzaju wagi w większości odpowiadały, że spożywają alkohol tylko czasami („Sometimes”). W żadnej grupie odpowiedź „Często” („Frequently”) nie przewyższa, a nawet nie dorównuje odpowiedzi „Czasami” („Sometimes”). Warto również zauważyć, że osoby z otyłością poziomu pierwszego, nadwagą poziomu drugiego oraz osoby z wagą prawidłową lub niedowagą, na pytanie odnośnie alkoholu odpowiadały równie często „nie” jak i „czasami”. Oznacza to, że częstotliwość spożywania alkoholu nie ma wpływu na rodzaj wagi. Nie oznacza to jednak, że spożywanie alkoholu jest zdrowe.



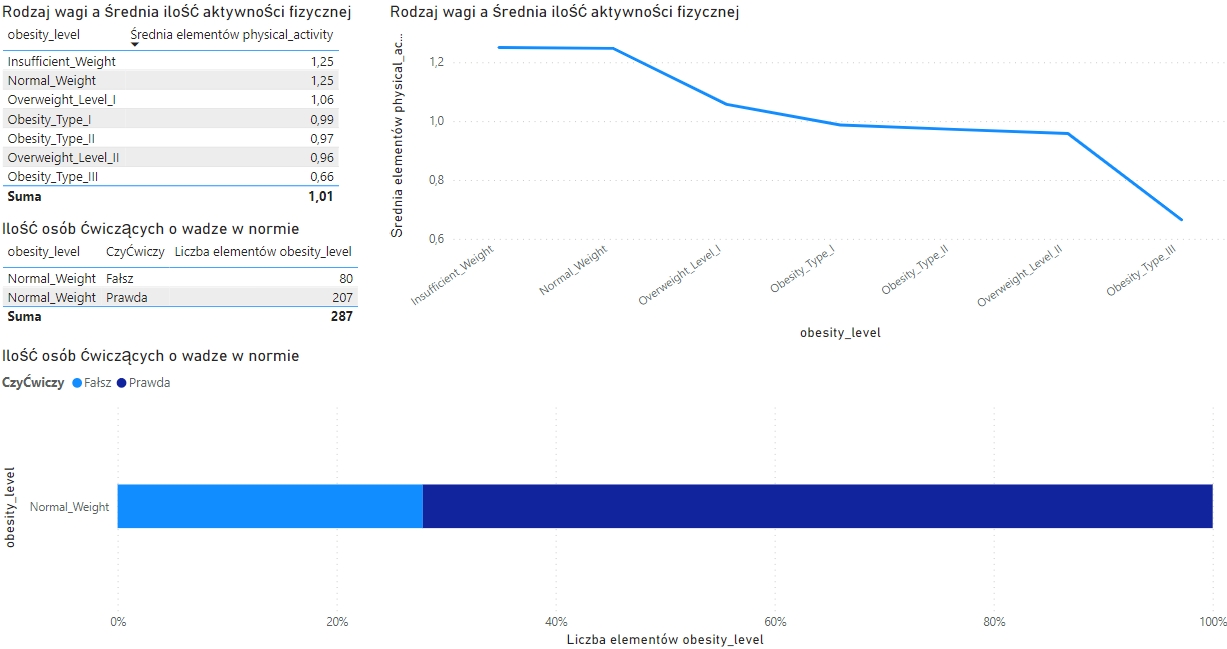
*Rysunek 8 Rodzaj wagi a monitorowanie kalorii*

Piąty pulpit ukazuje dość zaskakujące dane z dwóch powodów. Po pierwsze, jak bardzo mały procent badanych monitoruje swoje kalorie – łącznie tylko 5%. Drugim zaskoczeniem jest to, że osoby z poprawną wagą są co prawda drugą najliczniejszą grupą monitorującą swoje kalorie, lecz wciąż jest to tylko 10,45%. Pozostałe 89,55% nie liczy swoich dziennych kalorii. Z danych tych wynika, że nie jest to istotny czynnik decydujący o nadwadze.



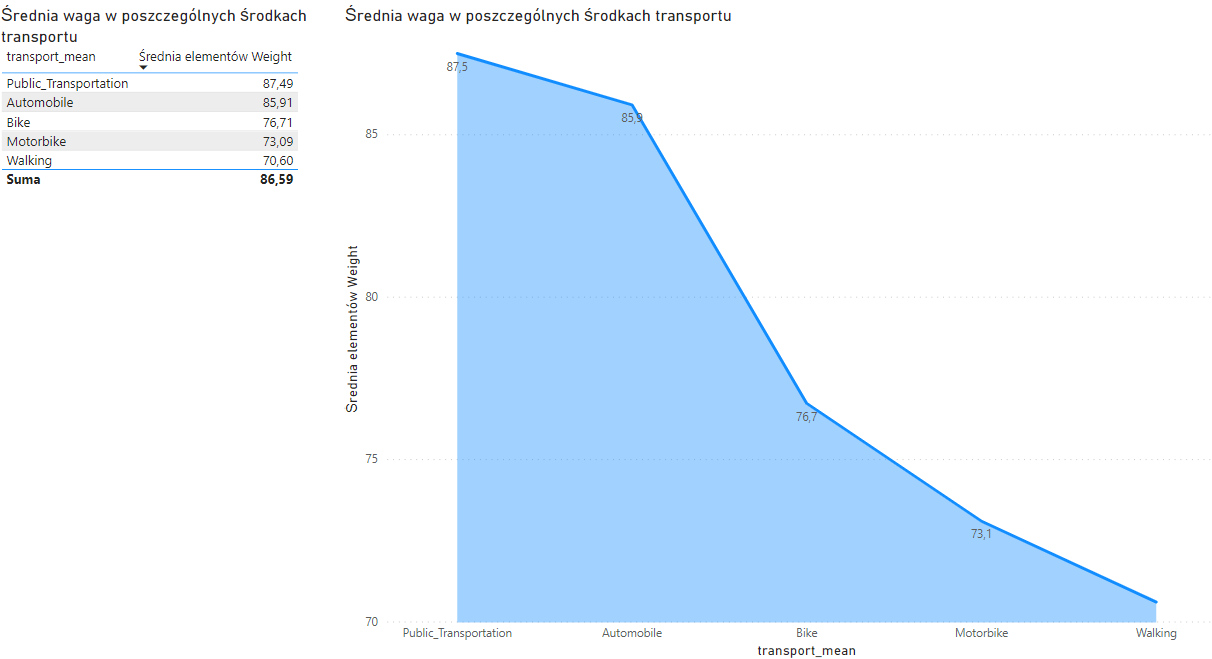
*Rysunek 9 Kolumna CzyĆwiczy*

Na potrzeby szóstej wizualizacji potrzebna była kolumna wyliczeniowa, sprawdzająca czy dana osoba jest aktywna fizycznie.



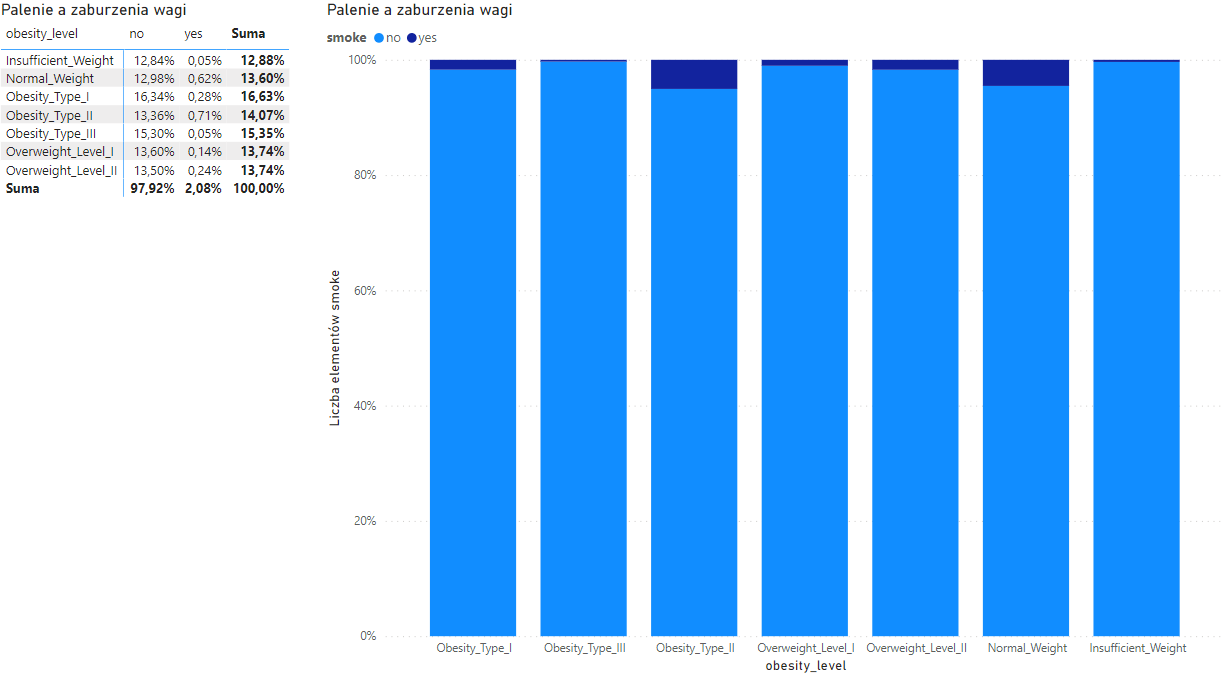
*Rysunek 10 Aktywność fizyczna a poprawna waga*

Tym razem pod lupę zostanie wzięta aktywność fizyczna. Pierwsza tabela oraz wykres liniowy ukazują, że im osoba posiada większą nadwagę, tym rzadziej uprawia sport. Odpowiadając na założoną hipotezę, należy spojrzeć na drugą tabelę oraz wykres słupkowy. Dane na nim wskazują, że większość, bo 72,13% badanych posiadających wagę w normie, uprawia regularnie jakąś aktywność fizyczną. W związku z powyższym, hipotezę mówiącą o zależności uprawiania sportu a poprawnej wadze, można przyjąć za poprawną.



*Rysunek 11 Środek transportu a średnia waga*

Przedostatni panel skupia się na tym, jaka jest przeciętna waga osób podróżujących poszczególnymi środkami transportu. Jak widać na załączonym rysunku, im sposób poruszania się wymaga większego wysiłku, tym niższa jest średnia waga. Można więc założyć, że rodzaj wybieranego transportu wpływa na średnią wagę jak i na rodzaj wagi.



*Rysunek 12 Palenie a zaburzenia wagi*

Ostatnia wizualizacja jest równie zaskakująca co piąta. Zaskakująca jest liczba osób, która odpowiedziała, że pali – tylko 2,08%. Palenie jest nawykiem wiążącym się z bardzo poważnymi skutkami, lecz dziwi głównie jak mało osób odpowiedziało „tak” w stosunku do liczby osób palących zaobserwowanych w codziennym życiu. Wracając do treści dashboardu, możemy stwierdzić, że palenie nie ma wpływu na nadwagę, niedowagę lub wagę prawidłową, tym samym obalając ostatnią hipotezę.

# Podsumowanie i wnioski

Dane dostarczyły odpowiedzi na wiele pytań dotyczących czynników wpływających na niepoprawną wagę w krajach Ameryki Południowej. Jak wcześniej zakładano, płeć nie ma znaczenia na znaczenia na większość rodzajów wag, jednak warto odnotować, że otyłość stopnia drugiego jest zdominowana przez mężczyzn, natomiast otyłość stopnia trzeciego – przez kobiety. Okazuje się jednak, że najistotniejszym czynnikiem decydującym o tym, czy człowiek jest narażony na nadwagę lub otyłość jest historia rodzinna z przypadkami nadwagi. Równie ważnym czynnikiem jest aktywność fizyczna. Ważnym jest, by regularnie ćwiczyć lub uprawiać jakikolwiek sport, aby utrzymywać wagę w normie. Jak się również okazało, średnia waga zależy od rodzaju wybranego środka transportu.

Analiza dostarczyła również zaskakujących wniosków obalających hipotezy. Najbardziej nieprzewidywalną informacją jest to, że osoby o prawidłowej wadze niekoniecznie monitorują swoje kalorie. Z raportu wynika również, że palenie nie decyduje o zaburzeniach wagi.

1. <https://www.kaggle.com/datasets/fatemehmehrparvar/obesity-levels> [↑](#footnote-ref-0)
2. <https://creativecommons.org/licenses/by/4.0/> [↑](#footnote-ref-1)