

# Wstęp do Eksploracji Danych - Projekt II

Zuzanna Ostas, Katarzyna Grzesiuk, Paweł Malinowski

## Opis oraz cel projektu

Celem naszego projektu jest analiza wyników ankiet przeprowadzonych w ramach diagnozy społecznej w roku 2015. Użyte zostały dane pochodzące ze strony internetowej: <http://www.diagnoza.com/>

Badania przez nas przeprowadzone skupiły się głównie na aspektach związanych z zarobkami, wykształceniem, wykonywaną pracą. Zainteresowaniem naszym było również przeanalizowanie, jak dane aspekty mają wpływ na poziom zadowolenia z życia polskich obywateli. Wyniki naszych testów w celu zapewnienia większej czytelności podzieliliśmy na cztery rozdziały:

1. Analiza ogólna społeczeństwa
2. Analiza zarobków
3. Analiza zdrowia fizycznego
4. Analiza zdrowia psychicznego

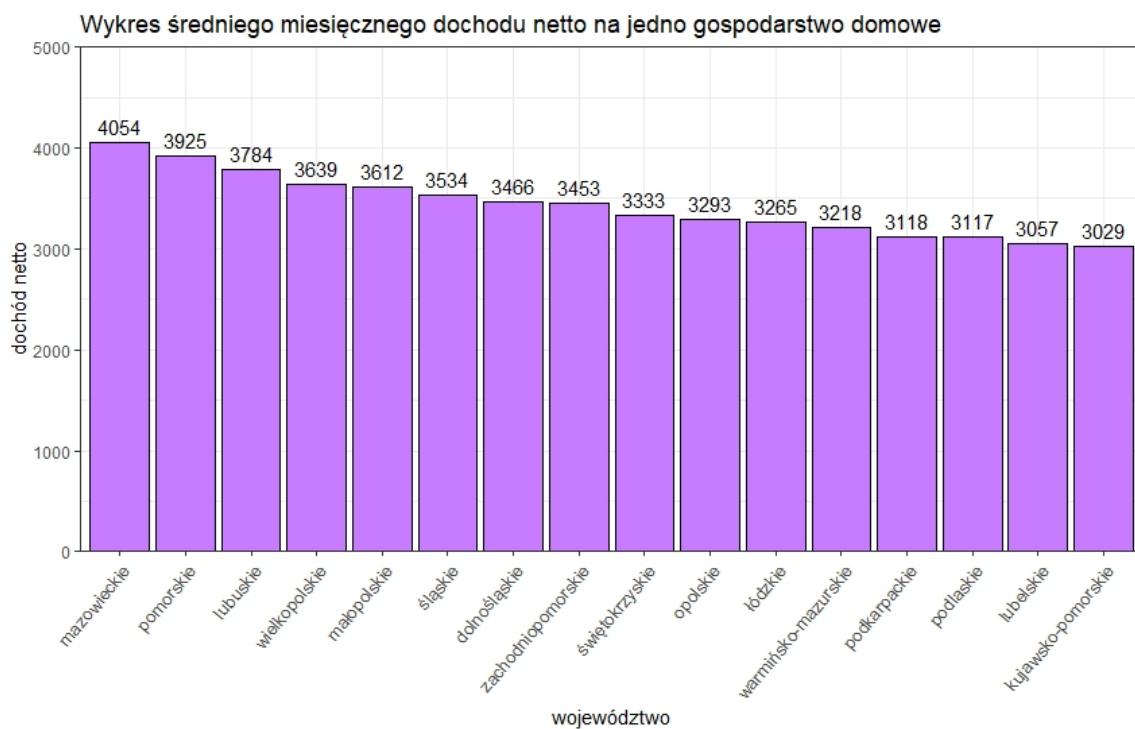
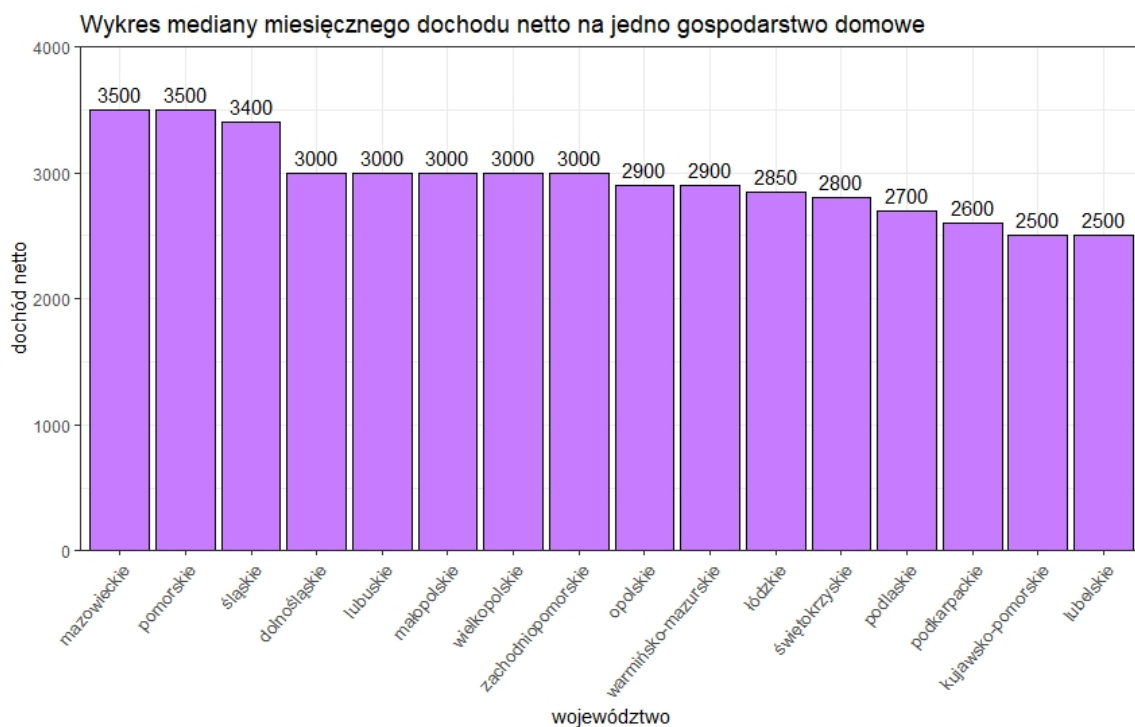
Rozdziały są w naturalny sposób ze sobą połączone, każdy z nich obejmuje dziedziny życia, które znajdowały się w naszym kręgu zainteresowań i zostały wymienione na początku tego opisu.

## 1 Ogólna analiza

W początkowej analizie dostępnych danych skupiliśmy się na zobrazowaniu podstawowych różnic względem miejsca zamieszkania oraz płci ankietowanych.

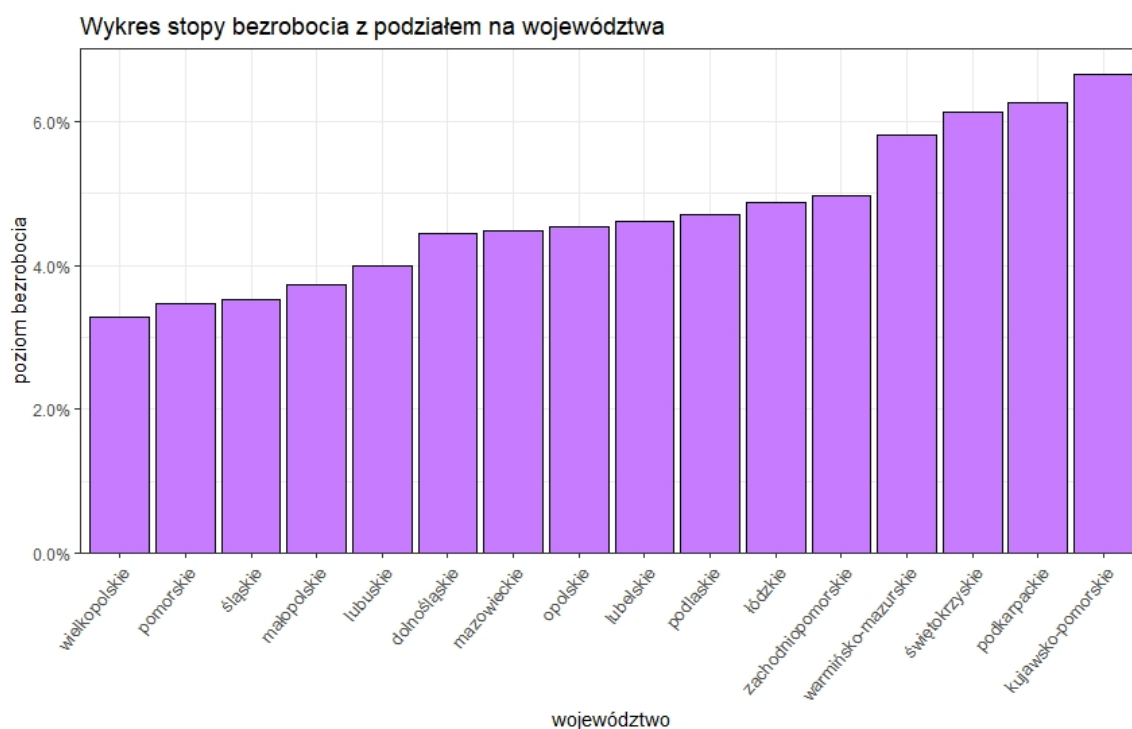
### Podział na województwa

Na poniższych wykresach możemy zobaczyć średnią oraz medianę miesięcznego dochodu netto przypadające na jedno gospodarstwo domowe.



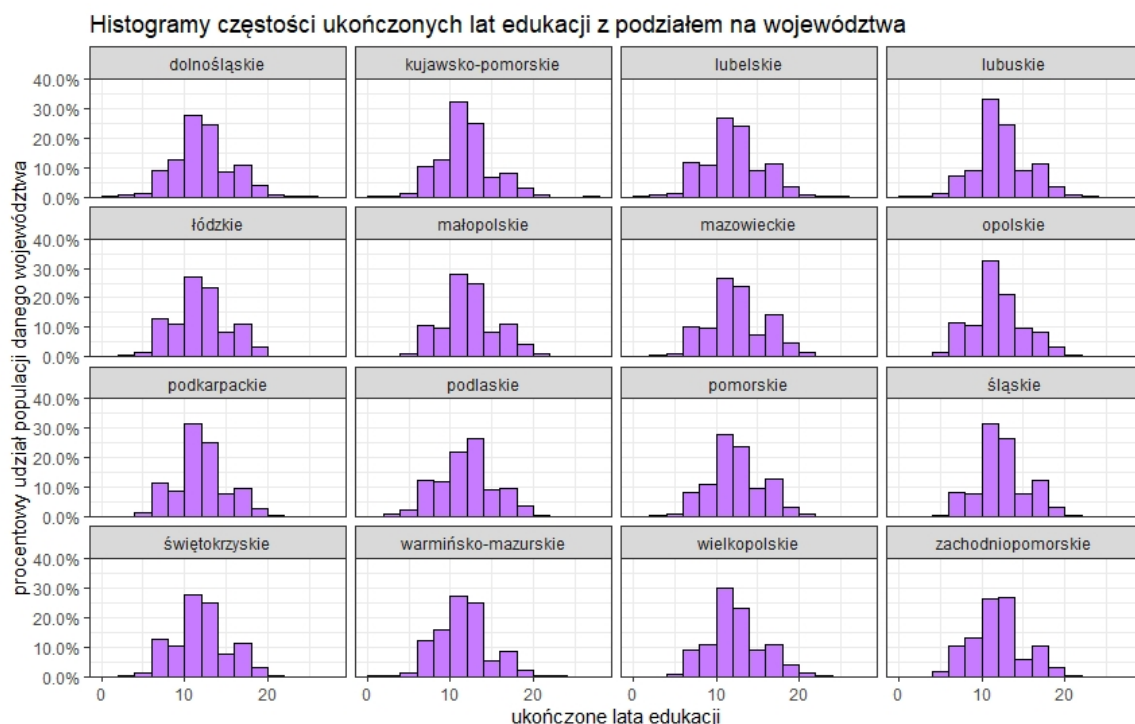
Z wykresów tych możemy wywnioskować, że gospodarstwa znajdujące się w makroregionie północno-zachodnim i południowo-zachodnim wraz z województwem mazowieckim wykazują dochody większe niż gospodarstwa w regionie wschodnim i centralnym. Szczególnie widoczna jest różnica między województwem kujawsko-pomorskim i mazowieckim. Sięga ona około 1000 złotych dla mediany, jak i średniej.

W celu głębszego przeanalizowania tych różnic zbadaliśmy również procentową stopę bezrobocia. Wyniki naszej analizy przedstawiamy na poniższym wykresie.



Z tego wykresu również możemy odczytać wspomniany wyżej podział. Znajdujące się w Polsce centralnej województwo kujawsko-pomorskie wykazuje zgodnie z danymi z roku 2015 dwukrotnie większy wskaźnik bezrobocia niż sąsiadujące z nim województwo wielkopolskie, które znajduje się w Polsce zachodniej.

W naszych badaniach chcieliśmy również dokonać obserwacji poziomu edukacji w poszczególnych województwach. W tym celu posłużyliśmy się zadeklarowaną przez ankietowanych liczbą ukończonych lat edukacji i na podstawie tych danych wykonaliśmy histogramy odpowiadające poszczególnym województwom.

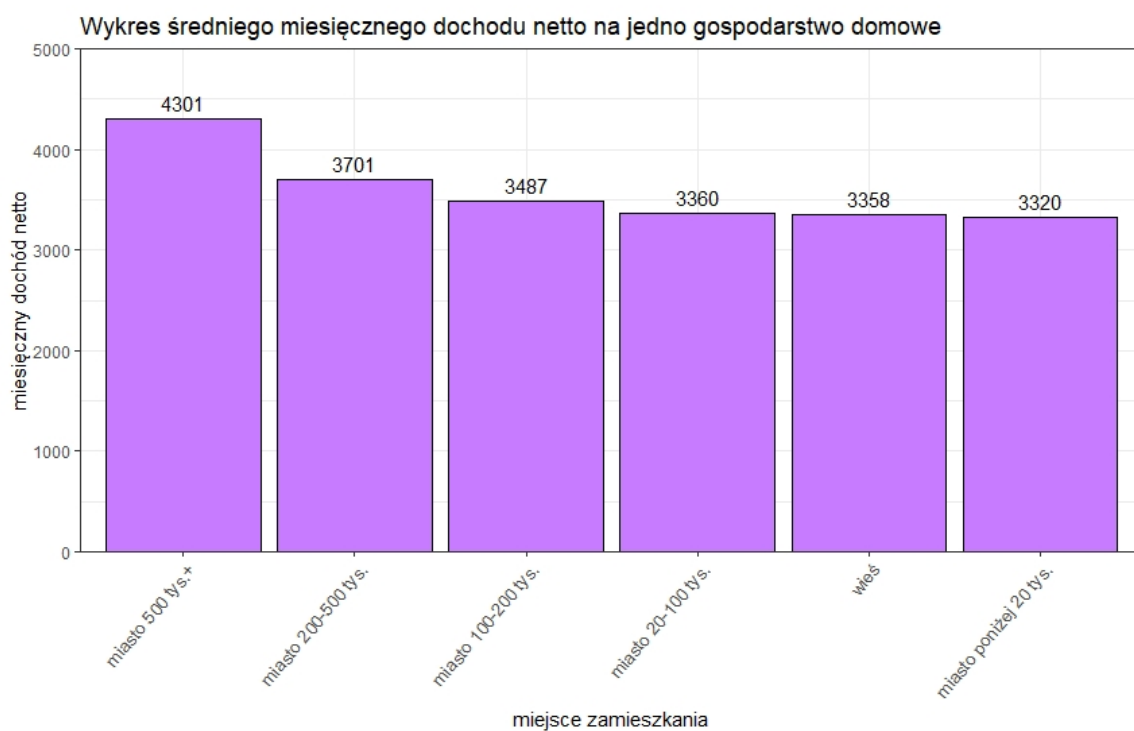
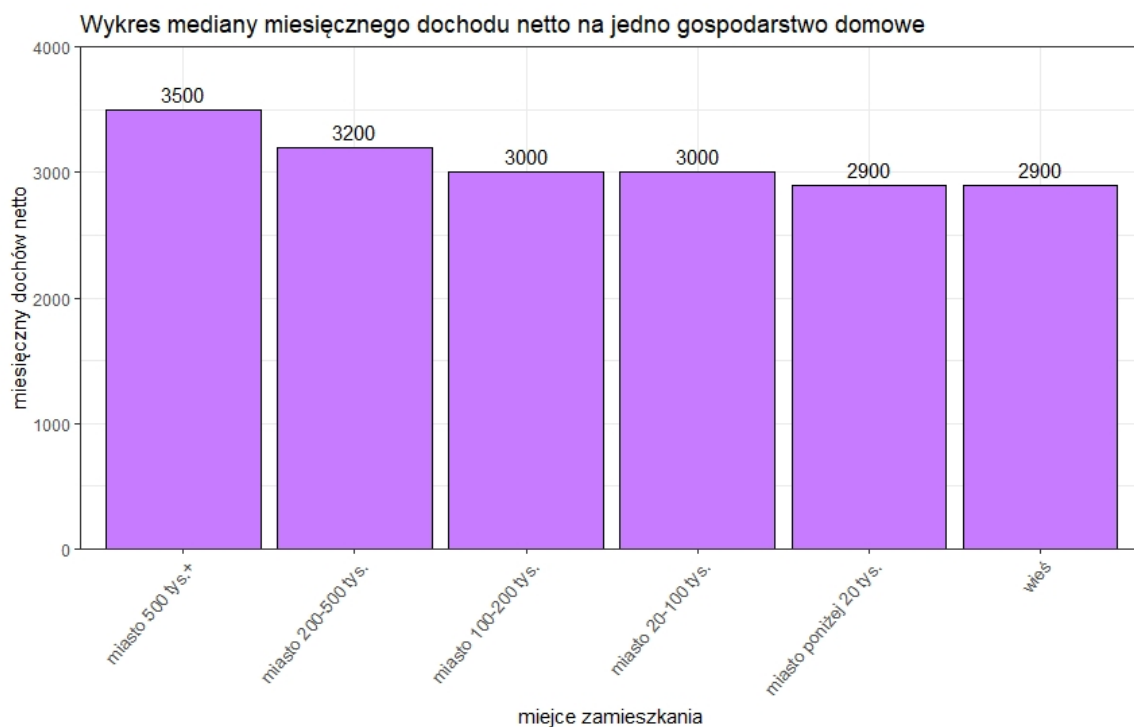


Jak możemy zauważyć, różnice pomiędzy województwami nie są znaczące. W każdym województwie największy odsetek populacji ukończył 10-12 lub 12-14 lat edukacji. Wynik ten jest związany 12-13 letnim okresem potrzebnym na uzyskanie wykształcenia średniego.

## Analiza na podstawie klasy miejscowości, w której znajduje się miejsce zamieszkania badanego

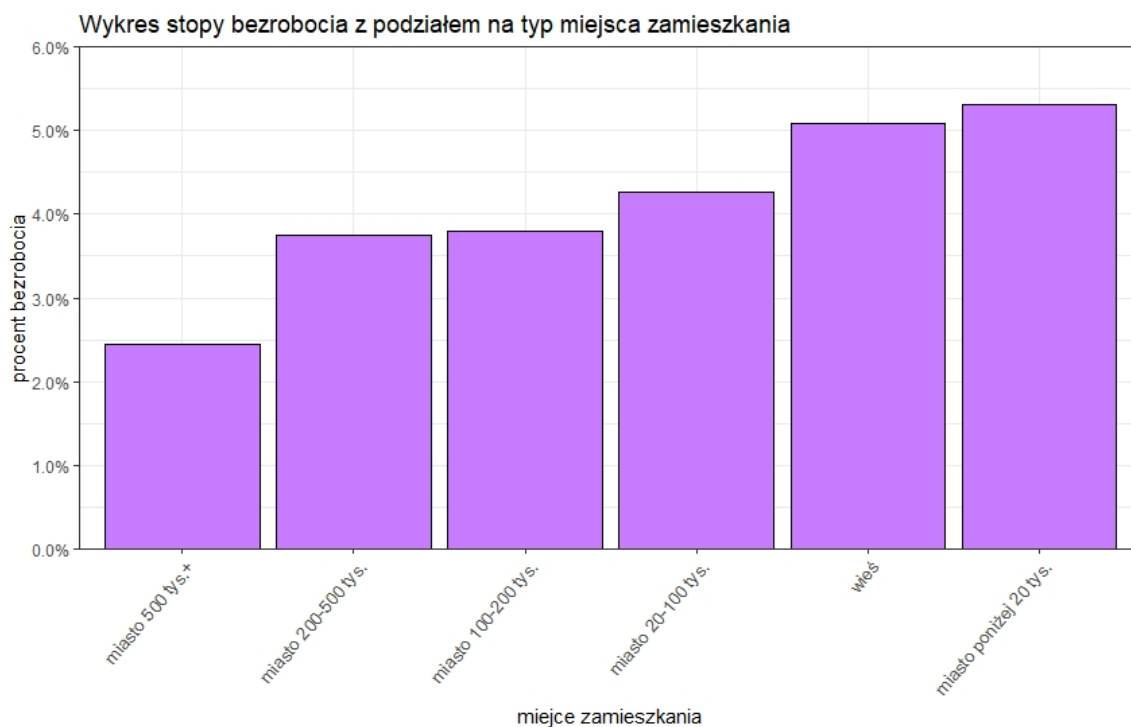
Naturalnym pytaniem nasuwającym się po analizie województw jest zbadanie wpływu wielkości miejsca zamieszkania na przebadane wcześniej aspekty życia.

Analizując medianę oraz średnią miesięcznego dochodu netto na jedno gospodarstwo domowe otrzymaliśmy poniższe wykresy.



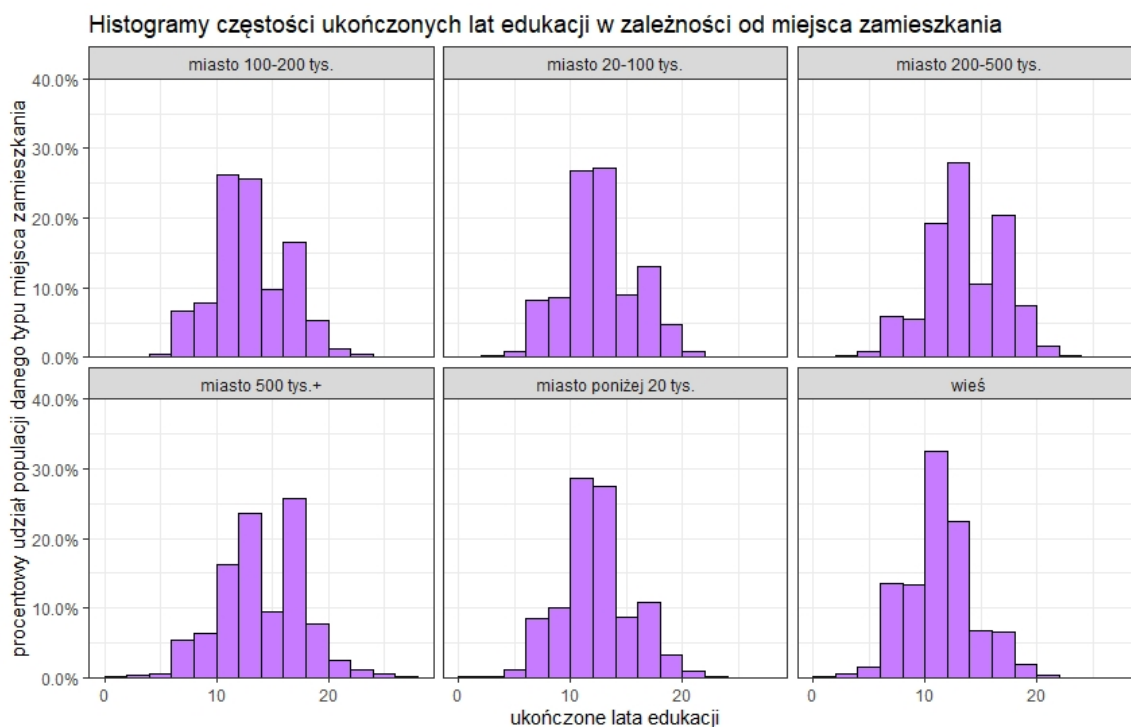
Z wykresów tych możemy wywnioskować, że dochód wykazywany przez gospodarstwo domowe jest w znacznym stopniu związany z tym, gdzie dane gospodarstwo się znajduje i jest tym wyższy, im bardziej zurbanizowane jest miejsce zamieszkania.

Zbadaliśmy również wskaźnik stopy bezrobocia.



Łatwo możemy zauważyć, że wsie oraz miasta o populacji poniżej 20 tysięcy mieszkańców wykazują większą stopę bezrobocia niż bardziej zurbanizowane tereny. Możemy zauważyć dwukrotnie większą stopę bezrobocia pomiędzy ankietowanymi z małych miast i wsi a respondentami z miast z ludnością powyżej 500 tysięcy.

Pozostało nam zbadanie poziomu edukacji, który tak samo, jak w przypadku województw również przedstawimy za pomocą histogramów.

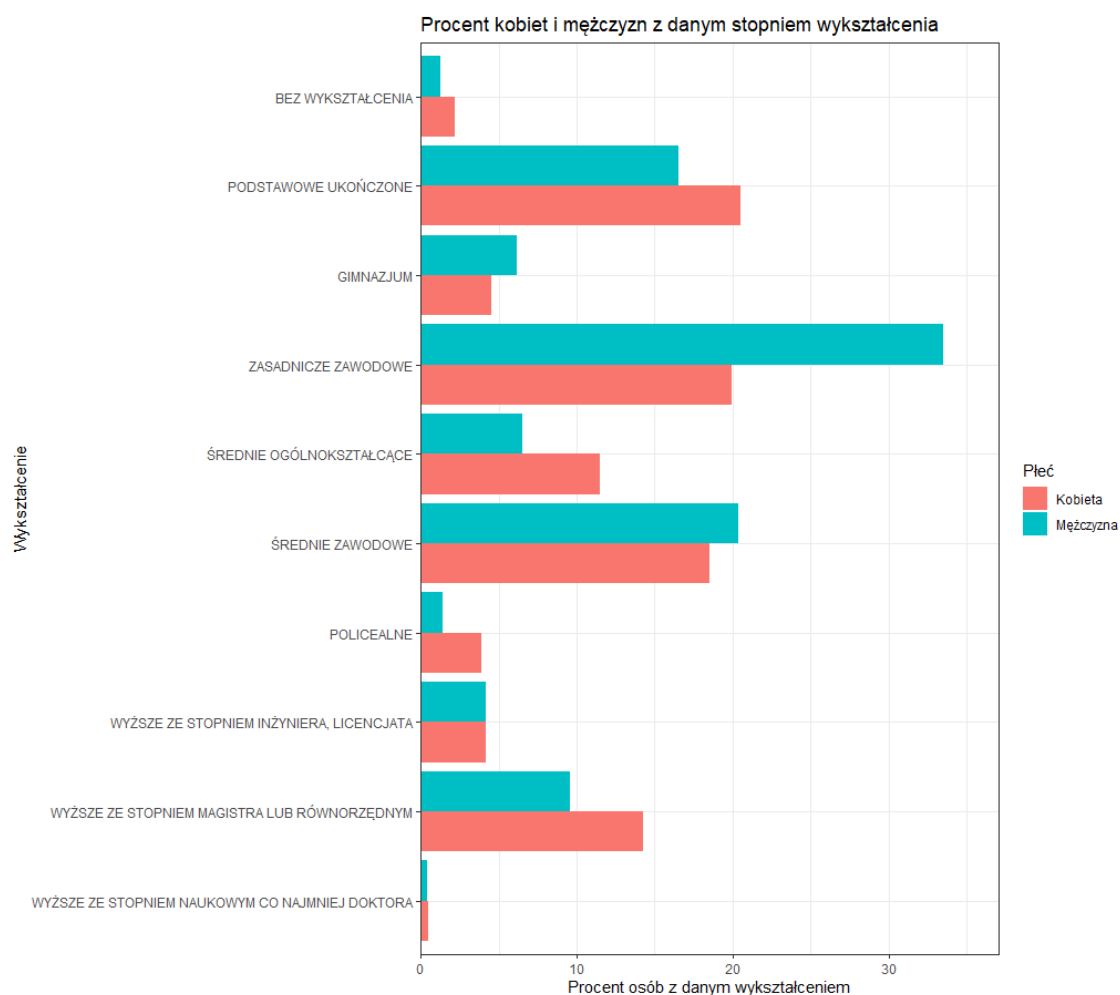


Z wykresów tych możemy odczytać różnice w liczbie ukończonych lat edukacji. Anketowani mieszkający w małych ośrodkach miejskich oraz na wsi odbywali głównie od 10 do 12 lub od 12 do 14 lat edukacji. Możemy wnioskować zatem, że uzyskali oni głównie wykształcenie średnie lub zasadnicze zawodowe. Największy odsetek respondentów, którzy odbyli przynajmniej 14 lat edukacji, znajduje się w wśród mieszkańców miast największych. Przy czym w miastach o ludności powyżej 500 tysięcy, osoby posiadające od 16 do 18 lat edukacji stanowią grupę najliczniejszą i stanowią oni około 25% populacji.

## Analiza pod względem podziału na płci

Postanowiliśmy również sprawdzić, czy poziom uzyskanej edukacji jest uzależniony od płci ankietowanego oraz czy płeć powoduje u ankietowanych wybór różnych stanowisk pracy.

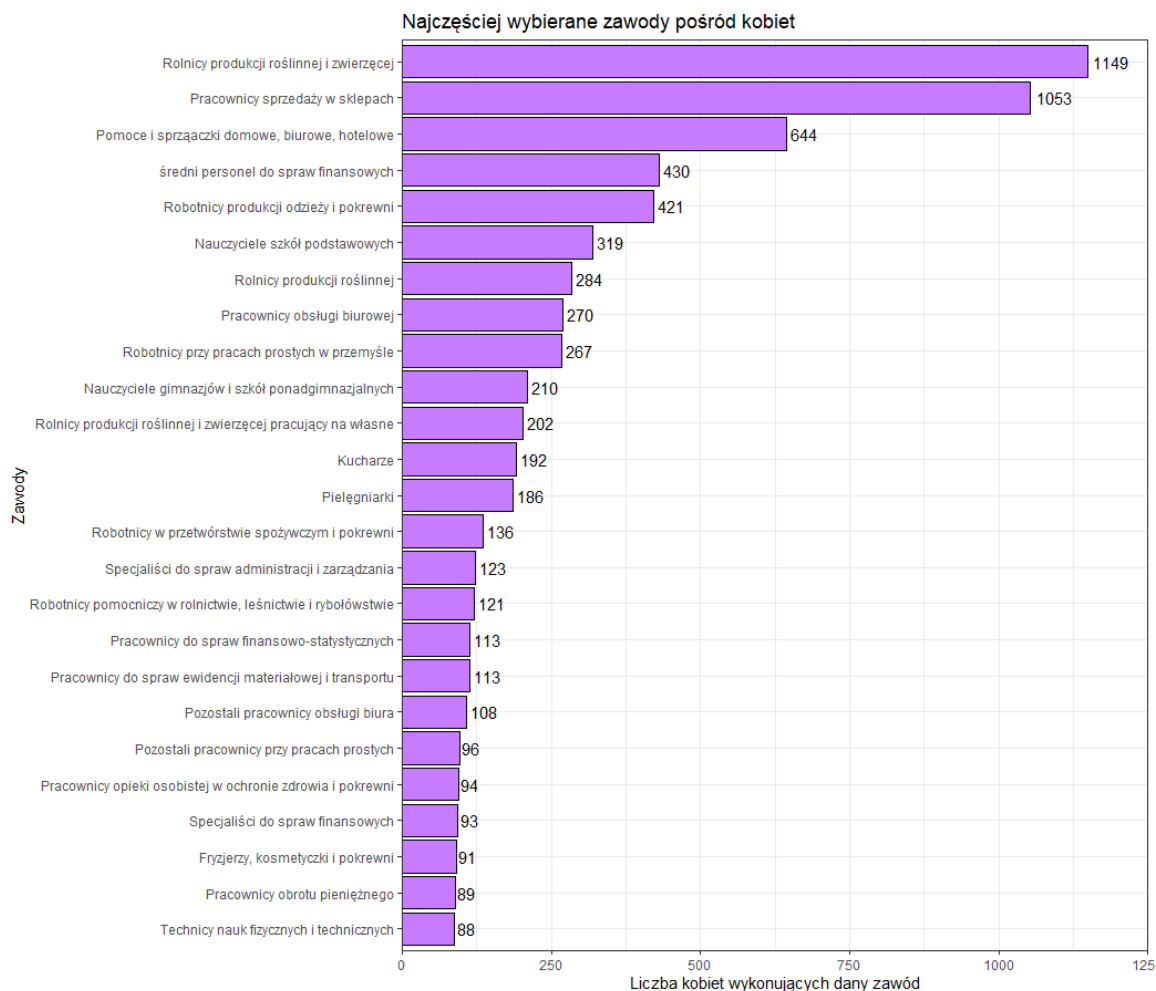
Aby zbadać poziom wykształcenia względem płci, wykonaliśmy wykres słupkowy, który przypisuje procentowy udział populacji podzielonej na płeć do odpowiedniego stopnia wykształcenia.



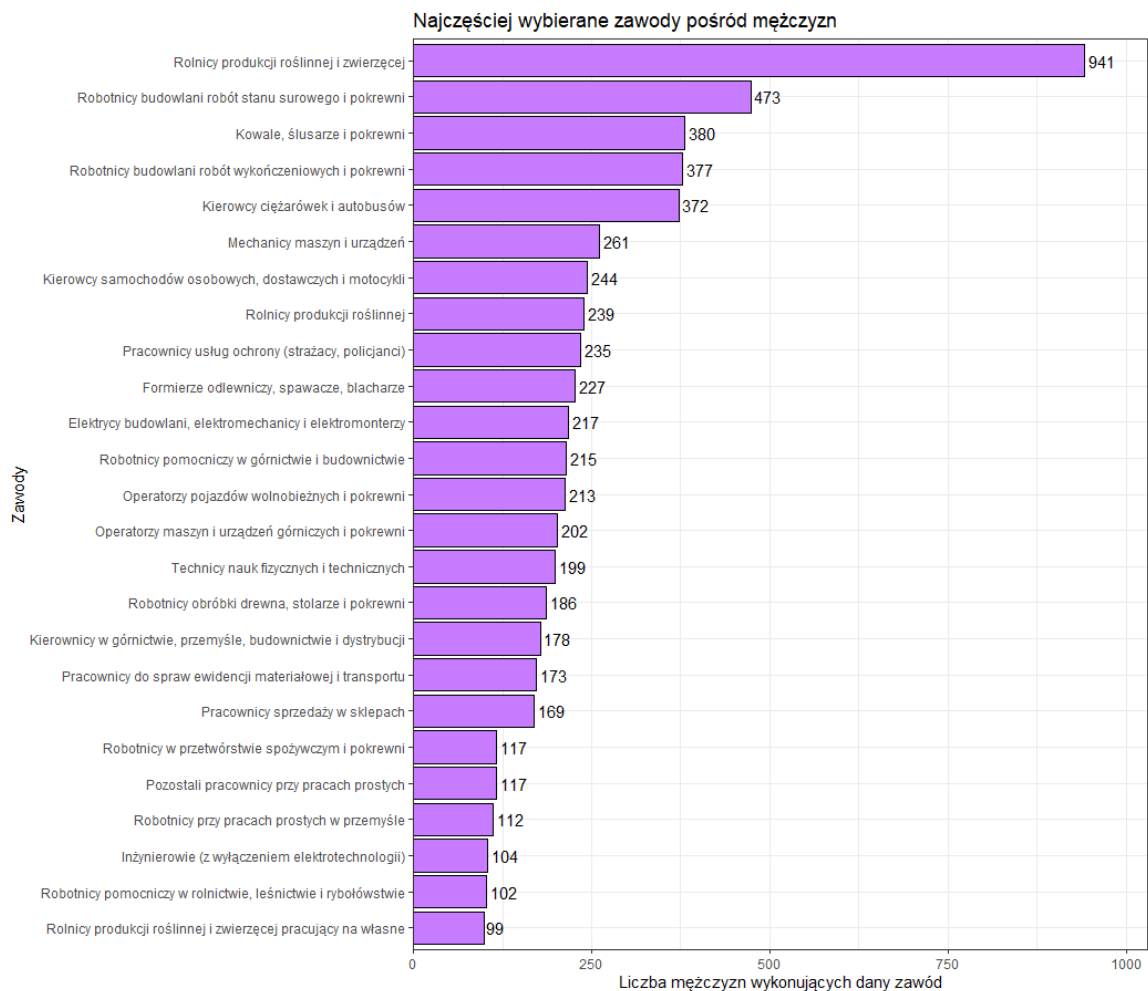
Analiza powyższego wykresu pozwala nam stwierdzić, że anketowani mężczyźni częściej niż kobiety kończą wykształcenie zawodowe zasadnicze jak i średnie. Kobiety natomiast częściej wybierają wykształcenie ogólnokształcące oraz kontynuują naukę

do uzyskania stopnia magistra lub stopnia mu równoważnego.

Ankietowani wśród pytań związanych z zawodem odpowiadali również na pytanie, w jakiej branży pracują. Wyniki uzyskane z tych pytań pozwoliły nam stworzyć wykres przedstawiający najczęściej wybierane branże. Ponieważ naszym celem było zbada-  
nie różnic zawodowych pomiędzy kobietami a mężczyznami mieszkającymi w Polsce, wykonaliśmy dwa wykresy osobne dla każdej z płci.







Z wykresów tych możemy wywnioskować, że prace wykonywane przez kobiety są związane głównie z relacjami międzyludzkimi, finansami i przemysłem lekkim. Kobiety, często pracują na stanowiskach biurowych oraz w placówkach edukacyjnych. Mężczyźni natomiast najczęściej wybierają pracę w przemyśle ciężkim oraz budownictwie. Wnioski te możemy połączyć z poprzednim wykresem dotyczącym edukacji. Ukończenia wykształcenia zawodowego u mężczyzn przyczynia się do dalszego wyboru kariery w sektorach technicznych. Natomiast uzyskanie wykształcenia wyższego przez kobiety wiąże się z częstym wyborem pracy w edukacji i finansach. Trudno nie zauważyć dość znacznego wkładu osób pracujących w sektorze rolnym. Dotyczy to obydwu płci i po przeprowadzonej przez nas analizie jest związane z dużą ilością danych pobranych z ośrodków wiejskich. Spostrzeżenie to przedstawiliśmy na poniższym wykresie.



Z analizy tej wynika, że 50,6% badanych to mieszkańcy wsi. Jest to wartość wyższa względem wyników badań przeprowadzonych przez GUS w latach 1950-2020, z których wynika, że w roku 2015 procentowy udział ludności wiejskiej na tle całej populacji wynosił 39,7%.

Dane te uzyskaliśmy z raportu:

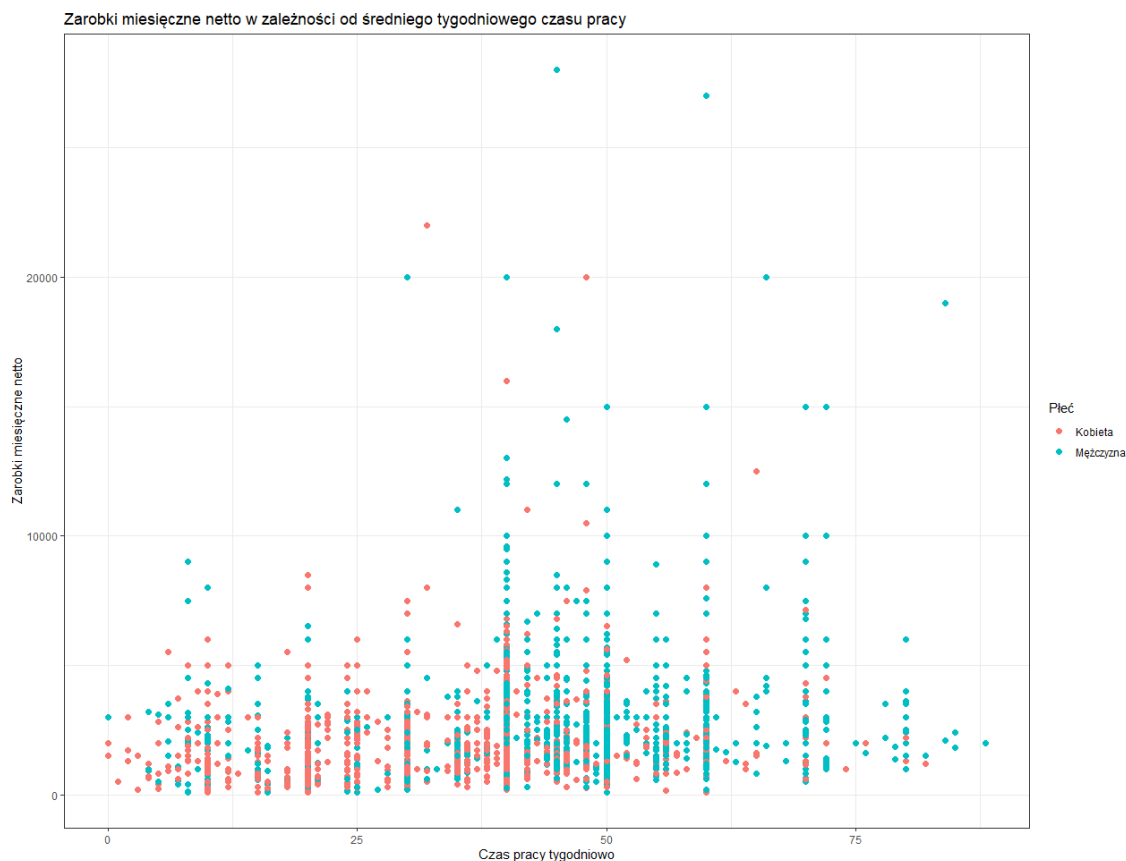
<https://stat.gov.pl/obszary-tematyczne/ludnosc/ludnosc/ludnosc-stand-istruktura-ludnosci-oraz-ruch-naturalny-w-przekroju-terytorialnym-stand-w-dniu-31-12-2020,6,29.html> plik Tabela.I.XLS

## 2 Zarobki

W tej części raportu omówione zostaną zagadnienia związane z zarobkami Polaków. Zbadane zostaną zarobki w zależności od średniego czasu pracy, stopnia wykształcenia oraz zadowolenia z wykonywanej pracy.

### Zależność pomiędzy zarobkami a czasem pracy

Zacniemy od zbadania jaka jest zależność pomiędzy czasem pracy a zarobkami. W tym celu powstał wykres kropkowy widoczny poniżej. Jedna oś przedstawia wartości liczbowe dla średniego tygodniowego czasu pracy, a druga informuje o miesięcznych zarobkach netto. Na wykresie kolorami zostały zaznaczone płci ankietowanych.

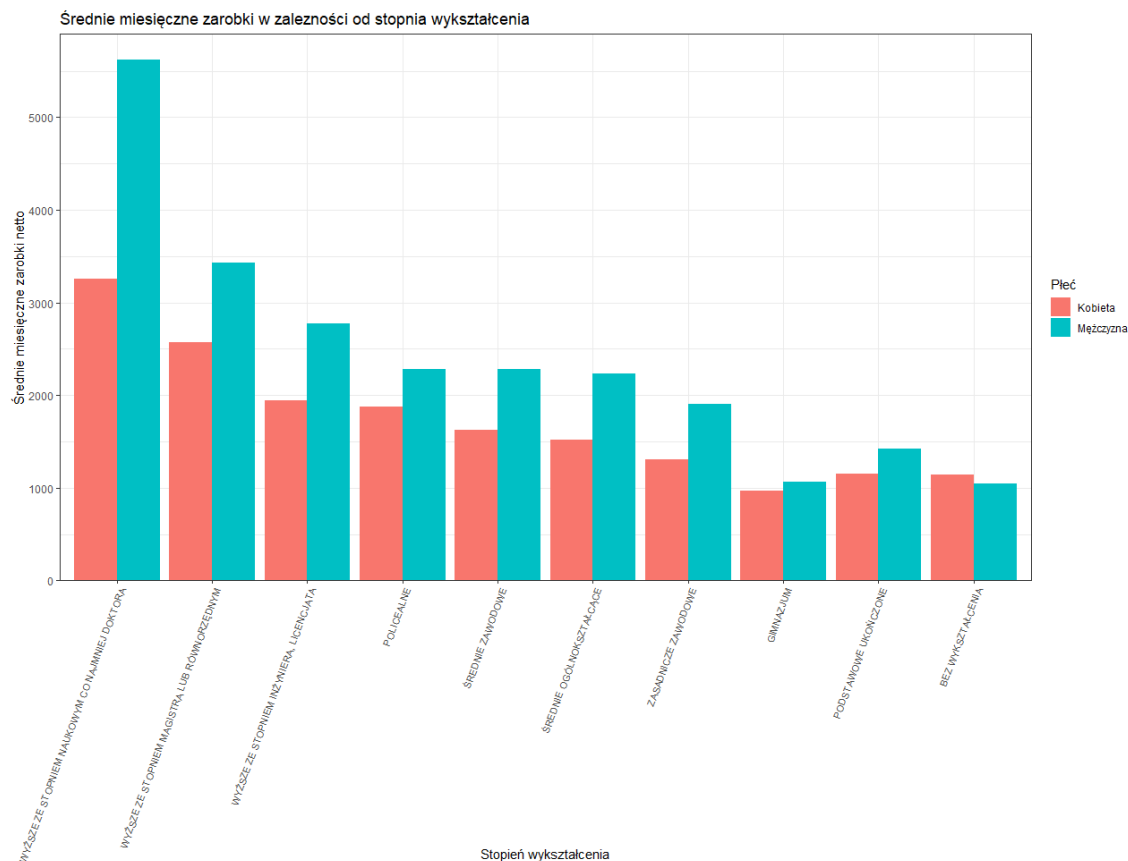


Pierwszym, mało zaskakującym wynikiem jest fakt, że tylko pojedyncze jednostki zarabiają miesięcznie powyżej 10000 zł netto. W grupie tej dominują mężczyźni. Przy tak wysokich zarobkach czas pracy tygodniowo wynosi co najmniej 30 godzin. W grupie osób zarabiających poniżej 10000 zł netto możemy zauważyć, że mężczyźni zazwyczaj pracują więcej w tygodniu niż kobiety. Zdecydowanie dominują oni w grupie osób pracujących więcej niż 50 godzin tygodniowo, natomiast w grupie ankietowanych, których czas pracy wynosił mniej niż 30 godzin tygodniowo jest zdecydowanie więcej kobiet. Zazwyczaj również wartości zarobków dla mężczyzn są wyższe w stosunku do zarobków kobiet pracujących tyle samo czasu tygodniowo.

Możemy zatem wnioskować, że zazwyczaj mężczyźni poświęcają więcej czasu na pracę tygodniowo niż kobiety, jednak nie zawsze idzie to w parze z wysokimi zarobkami. Z wykresu nie wynika jednoznaczna zależność pomiędzy czasem pracy a otrzymywanym wynagrodzeniem.

## Zależność pomiędzy zarobkami a stopniem wykształcenia

W celu zbadania zależności wysokości zarobków od stopnia wykształcenia powstał wykres słupkowy pokazujący średnią miesięczną pensję liczoną w grupach o różnym stopniu wykształcenia. Ponownie zastosowany został podział na płci.

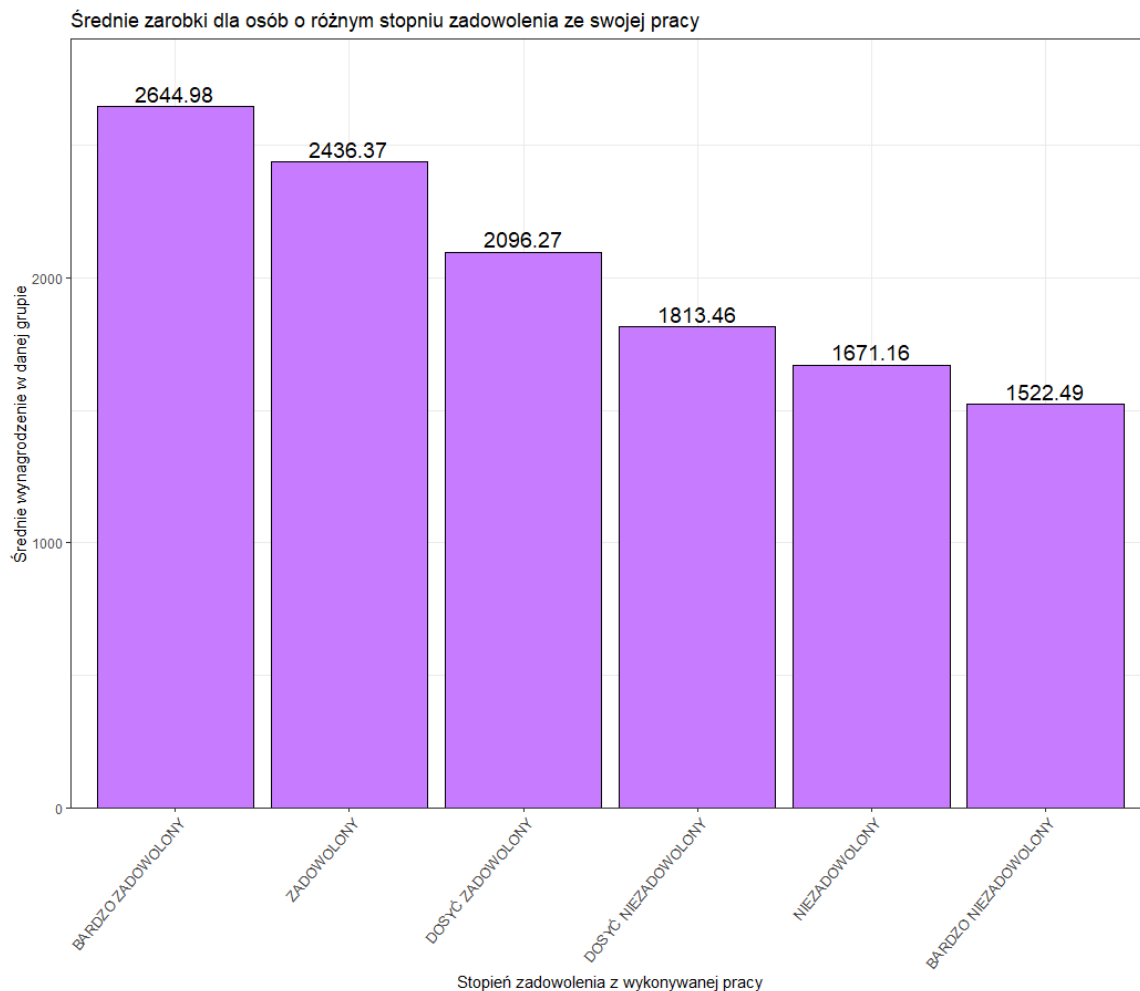


Z wykresu od razu wylania się dosyć naturalna prawidłowość pokazująca, że generalnie im wyższy stopień wykształcenia tym wyższe zarobki. Zależność ta jest prawdziwa dla obu płci dla stopni wykształcenia wyższych niż wykształcenie gimnazjalne. Możemy zauważyć, że dla wykształcenia policealnego, średniego ogólnokształcącego oraz średniego i zasadniczego zawodowego zarobki są podobne, zwłaszcza wśród mężczyzn, co również wydaje się naturalne gdyż w istocie są to podobne poziomy wykształcenia. W oczy rzuca się również, że mężczyźni zarabiają średnio zdecydowanie więcej od kobiet. Najwyraźniej widać to na przykładzie wykształcenia wyższego ze stopniem naukowym co najmniej doktora.

Głównym wnioskiem z powyższego wykresu jest zatem to, że w istocie patrząc na średnie zarobki, wyższy stopień wykształcenia wiąże się z większym wynagrodzeniem. Dodatkowo wyraźnie widać, nierówności w średnich wynagrodzeniach kobiet i mężczyzn o tym samym.

## Zależność pomiędzy zadowoleniem z pracy a zarobkami

Do zbadania zależności pomiędzy zarobkami a zadowoleniem z pracy policzone zostało średnie miesięczne wynagrodzenie dla grup ankietowanych zaznaczających różny stopień zadowolenia z wykonywanej pracy. Rezultat został przedstawiony na wykresie słupkowym. Na jednej osi znajdują się różne możliwe kategorie odnoszące się do zadowolenia z pracy, a na drugiej wartości liczbowe reprezentujące średnie miesięczne wynagrodzenie netto.



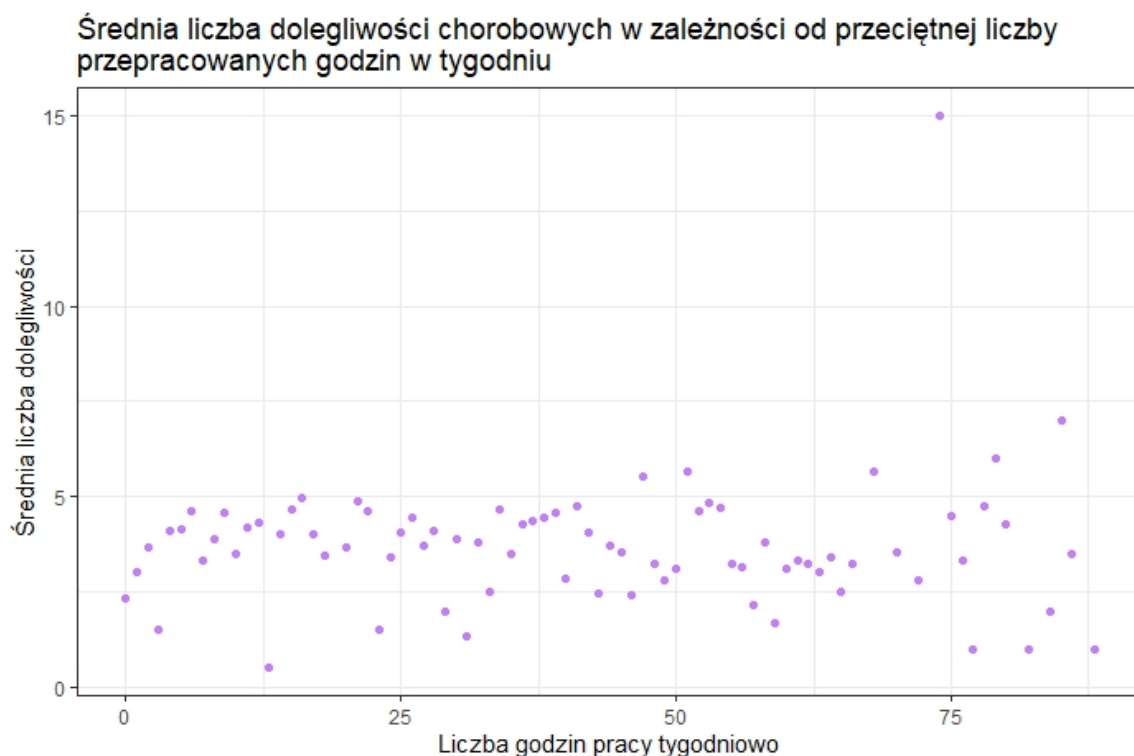
Analizując powyższy wykres od razu można zauważyć, że im wyższy stopień zadowolenia z pracy tym wyższe średnie wynagrodzenie. Różnica pomiędzy najniższym a najwyższym średnim wynagrodzeniem to ponad 1000 zł netto. Ponadto pomiędzy każdym słupkiem jest znacząca różnica. Możemy zatem wysnuć prosty wniosek, ponownie dosyć naturalny, że im lepiej płatna praca tym większe zadowolenie pracownika z jej wykonywania. Oczywiście pieniądze nie będą jedynym czynnikiem, który na to wpływa, lecz jak pokazuje wykres są one na pewno istotne.

### 3 Zdrowie fizyczne

Kolejnym tematem, który postanowiliśmy zbadać, jest zależność między pracą a zdrowiem fizycznym. Podejrzewając, że długie godziny pracy wpływają niekorzystnie na zdrowie, wpływ pracy na badanych sprawdzaliśmy na podstawie przeciętnej liczby godzin przepracowanych w ciągu tygodnia.

#### Zależność między średnim tygodniowym czasem pracy a liczbą dolegliwości

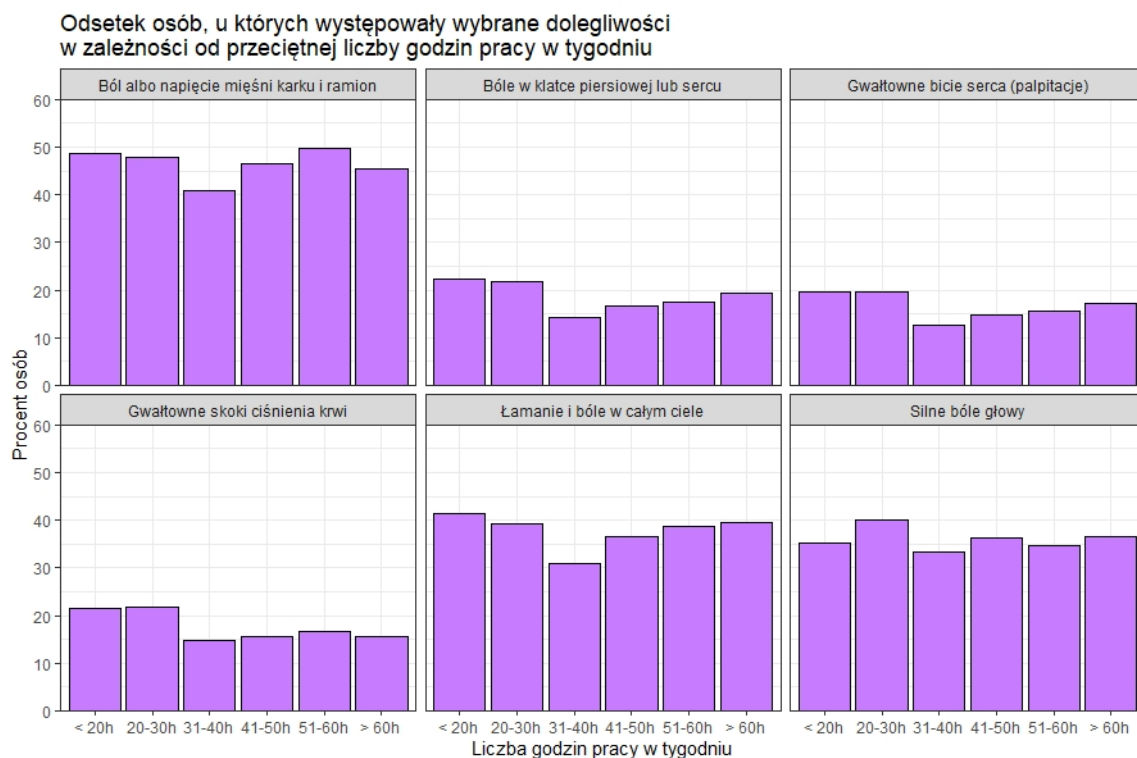
Zaczelśmy od sprawdzenia jak dużo dolegliwości występuje u osób pracujących. Dla odpowiednich grup została policzona średnia liczba występujących przypadłości spośród 15 wymienionych w ankiecie.



Na podstawie powyższego wykresu jednak nie można zauważyć rzeczywiście silnej zależności między liczbą dolegliwości a czasem pracy, w każdej grupie średnia ta skupia się na wartościach pomiędzy 3 a 5. Przy większej liczbie przepracowanych godzin widzimy większy rozrzut, jednak wynika to prawdopodobnie z tego, że grupy te mają mało reprezentantów wśród ankietowanych, więc wyniki zależą tak naprawdę od pojedynczych osób.

### **Zależność między średnim tygodniowym czasem pracy a występowaniem poszczególnych dolegliwości**

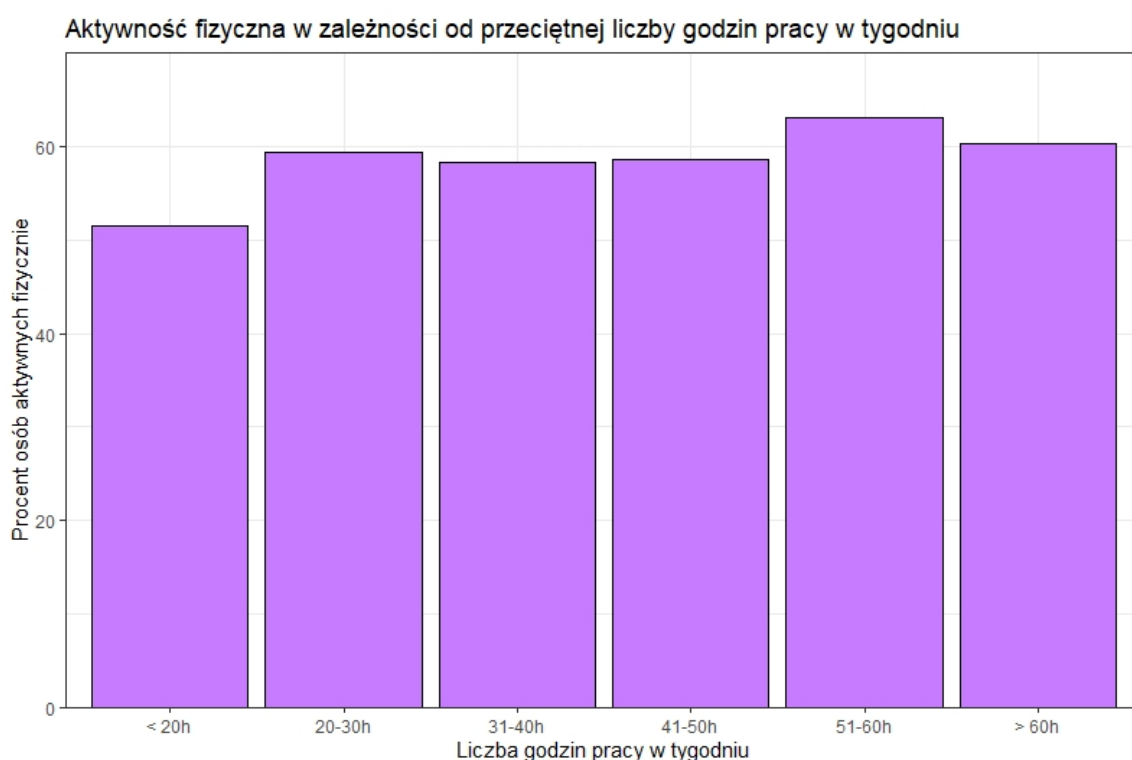
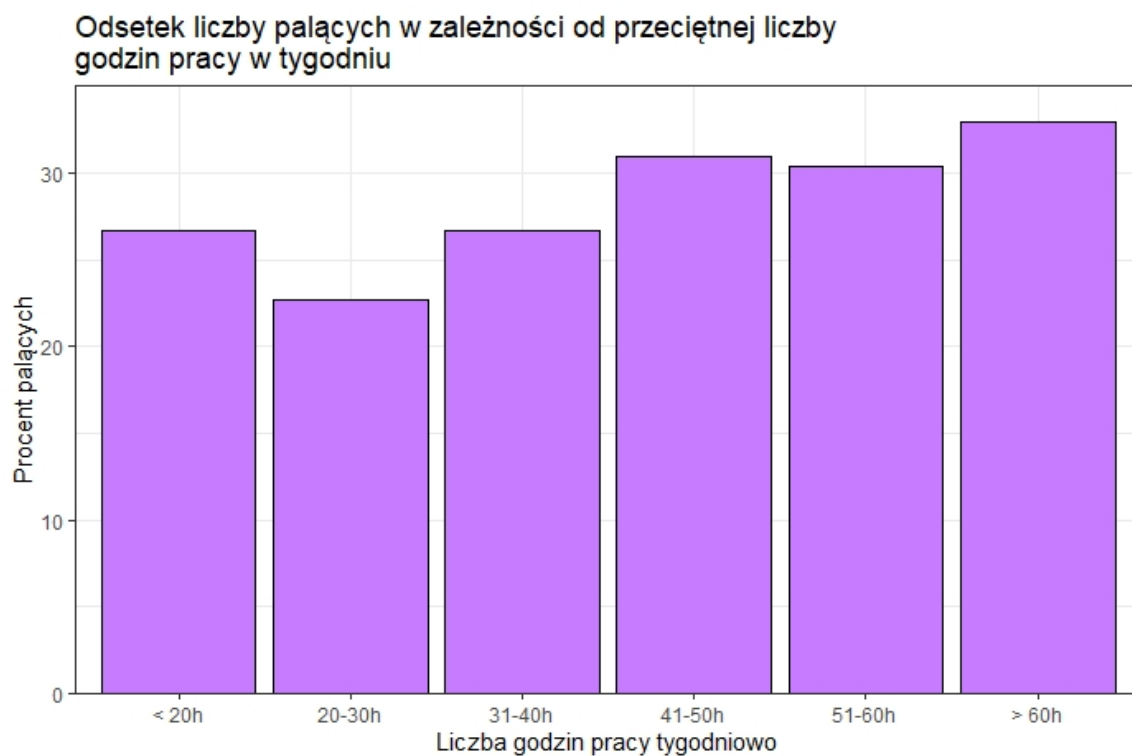
W dalszych badaniach sprawdziliśmy występowanie już tylko kilku wybranych dolegliwości, jednak najpierw podzieliliśmy godziny pracy na przedziały: poniżej 20h, 21-30h, 31-40h, 41-50h, 51-60h i powyżej 60h tygodniowo, a następnie badaliśmy różne zależności dla każdej z tych grup. Dla każdej choroby utworzyliśmy wykres pokazujący odsetek osób, u których ona występuje.



Wyniki przedstawione na wykresach okazują się raczej sprzeczne z oczekiwaniami. Myśląc nad wpływem pracy na zdrowie, spodziewaliśmy się raczej, że branie nadgodzin czy praca na więcej niż jeden etat nie sprzyja zdrowiu i że osoby w przedziałach powyżej 50 godzin tygodniowo powinny wypadać gorzej na tle innych. Jednak z wykresów widzimy, że grupy te właściwie się nie wyróżniają. Wyniki są wręcz przeciwne od wstępnych przewidywań, gdyż na większości wykresów najwyższymi słupkami są pierwsze dwa, czyli te reprezentujące osoby pracujące do 30 godzin tygodniowo. Widzimy też, że grupą o najmniejszym odsetku dolegliwości są osoby w przedziale 31-40h. Nasuwa się więc wniosek, że być może 7-8 godzin pracy dziennie jest rzeczywiście optymalnym rozwiązaniem. Z wykresu widzimy również, że niezależnie od tego ile się pracuje, jest się narażonym na łamanie i bóle w całym ciele, silne bóle głowy i przede wszystkim na ból i napięcie mięśni karku i ramion, który dokucza aż 40-50 procentom osób pracujących.

## **Zależność między średnim tygodniowym czasem pracy a dbaniem o zdrowie**

Poza występowaniem dolegliwości, zbadaliśmy również jak osoby pracujące dbają o swoje zdrowie i czy to też jest uzależnione od liczby godzin spędzanych w pracy. W tym celu sprawdziliśmy dwie rzeczy: czy ankietowane osoby palą papierosy oraz czy uprawiają jakąś formę aktywności fizycznej. W tym celu ponownie podzieliliśmy wszystkich na grupy w zależności od godzin pracy w tygodniu i dla każdej z tych grup obliczyliśmy odsetek osób palących i odsetek osób uprawiających jakikolwiek sport. Wyniki przedstawione są na poniższych wykresach.



Tym razem widzimy, że odsetek liczby palących raczej rośnie wraz ze zwiększaniem liczby godzin pracy. Najwięcej osób palących jest w grupie pracujących ponad 60 godzin tygodniowo, gdzie odsetek tych osób sięga około 33%. Najlepiej wypada w tej kategorii grupa 20-30h, w której wynik to około 23% palących.



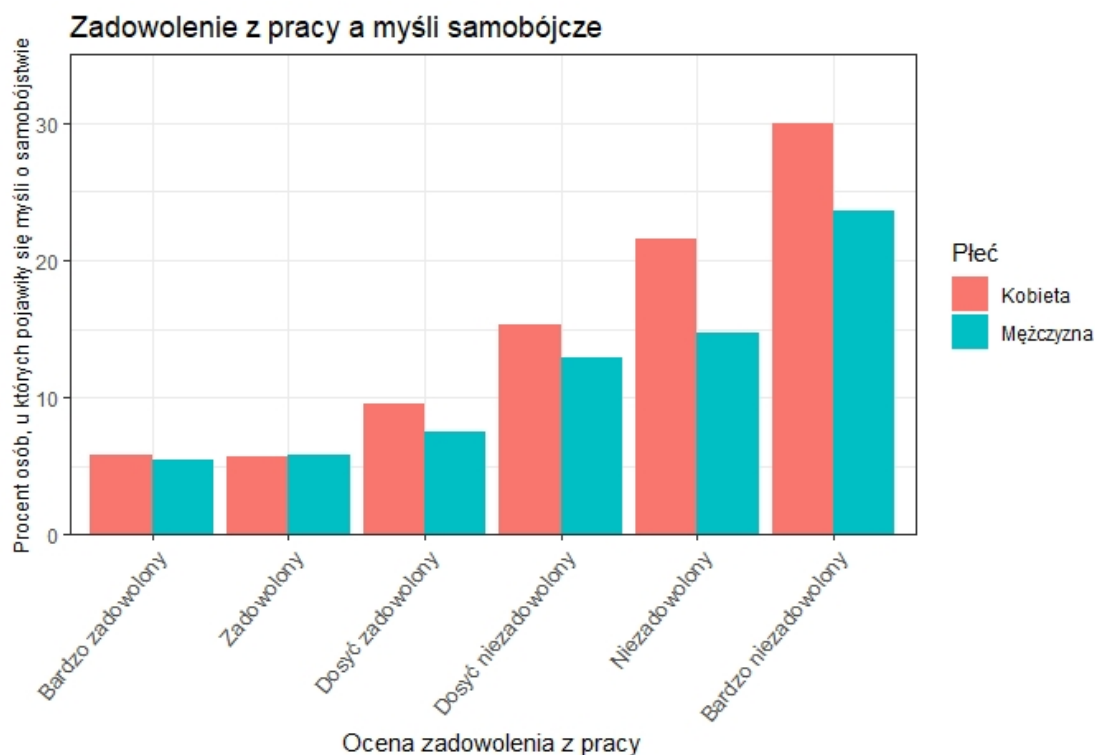
W kwestii aktywności fizycznej, wyniki są bardzo zbliżone dla każdej grupy, jednak ponownie zaskoczeniem jest najniższy wynik osób pracujących poniżej 20 godzin tygodniowo. Mogłoby się wydawać, że kiedy mniej czasu poświęcamy na pracę, to mamy go więcej na zajęcia takie jak sport, jednak ankieta nie potwierdziła takich przypuszczeń.

## 4 Zdrowie psychiczne

Kontynuując wątek zdrowia, postanowiliśmy dalej sprawdzić zależność między zdrowiem psychicznym ankietowanych a pracą. W tym celu podzieliliśmy badane osoby ze względu na ich ocenę zadowolenia ze swojej pracy, na 6 kategorii zgodnie z sześcioma stopniami zadowolenia bądź niezadowolenia które można było zaznaczyć w ankiecie. Zdrowie psychiczne natomiast postanowiliśmy sprawdzić na dwa sposoby – zbadać zarówno szczęście jak i nieszczęście ankietowanych w zależności od zadowolenia z pracy.

### Zależność między zadowoleniem z pracy a występowaniem myśli samobójczych

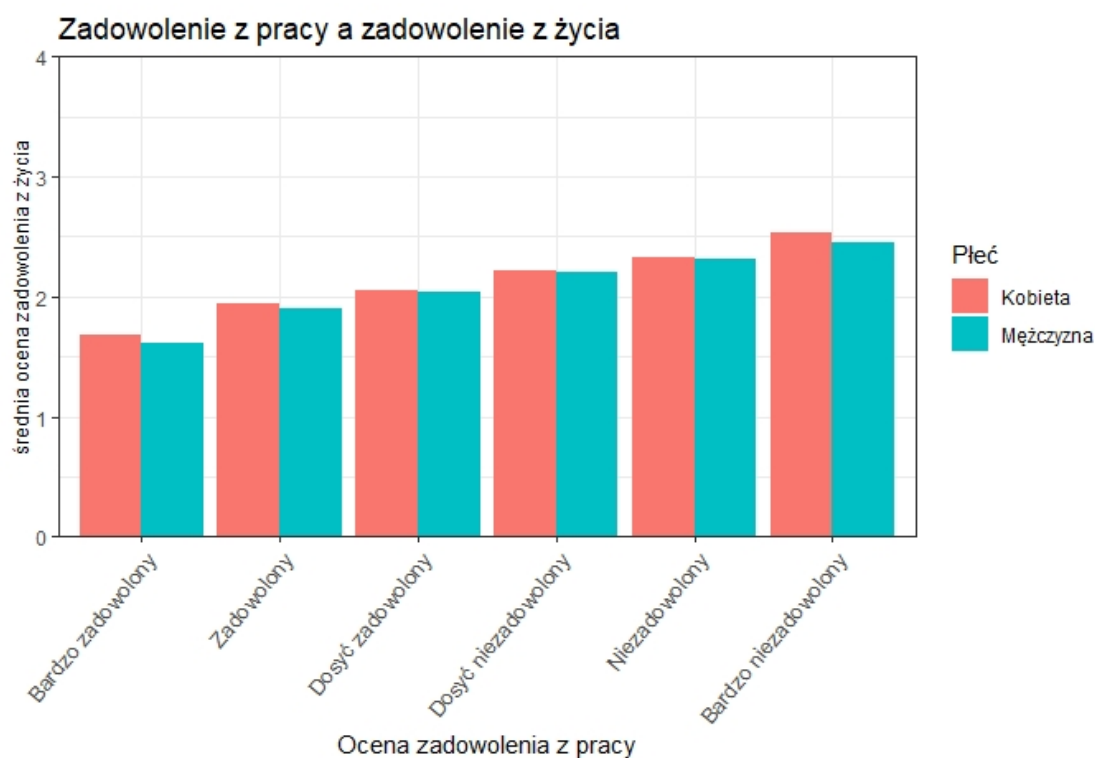
Zaczynając od nieszczęścia, analizowanym aspektem było pojawianie się myśli samobójczych u badanej osoby. W odpowiedzi na pytanie *Jak często w minionych miesiącach zdarzało się Panu być tak załamany, że myślał Pan o samobójstwie?* można było wybrać jedną z czterech kategorii: *nigdy, rzadko, dosyć często i bardzo często*. Uprościliśmy te kategorie do dwóch grup – jedna u której występowały myśli samobójcze i druga w której nie występowały. Następnie obliczono jaki był odsetek osób z pierwszej grupy w zależności od zadowolenia z pracy i z podziałem na płeć.



Od razu rzucają się w oczy dwie rzeczy. Po pierwsze, widać, że im mniejsze jest zadowolenie z pracy, tym większy jest odsetek osób, które rozważały popełnienie samobójstwa. Dla osób bardzo niezadowolonych z pracy, wynik sięgał około 24% u mężczyzn i aż 30% wśród kobiet. Drugą rzeczą jest różnica między płciami, gdyż w prawie wszystkich kategoriach widać znaczną przewagę odsetka kobiet, które myślały o samobójstwie.

## Zależność między zadowoleniem z pracy a zadowoleniem z życia

Przechodząc dalej do badania szczęścia ankietowanych, sprawdziliśmy jak ludzie oceniają swoje zadowolenie z życia w zależności od zadowolenia z pracy. Tutaj ankieta dawała 4 możliwości odpowiedzi: 1 – *bardzo szczęśliwy*, 2 – *dosyć szczęśliwy*, 3 – *niezbyt szczęśliwy*, 4 – *nieszczęśliwy*. Korzystając z tej skali obliczyliśmy jak badani z poszczególnych grup z podziałem na płeć oceniają swoje szczęście. Na wykresie przedstawiliśmy średnią ocenę dla każdej grupy, przy czym należy pamiętać, że w tym przypadku wyższy słupek oznacza mniejsze zadowolenie z życia.



Widzimy wyraźny, choć niezbyt silny, wzrost zadowolenia z życia wraz ze wzrostem zadowolenia z pracy. Okazuje się także, że generalnie mężczyźni są bardziej zadowoleni z życia, chociaż słupki są bardzo zbliżonej wysokości. Wyniki nie są może tak wyraziste jak na poprzednim wykresie, jednak zdecydowanie można dostrzec zależność między zadowoleniem z życia a pracy.

## Wnioski

Udostępniona nam baza danych dotycząca diagnozy społecznej dostarczyła nam wielu przydatnych informacji, które wykorzystaliśmy w przeprowadzonych analizach. Pierwszym z nasuwających się wniosków jest ciągle widoczne zróżnicowanie polskiego społeczeństwa pod względem podziału regionalnego, stopnia urbanizacji miejsca zamieszkania oraz płci. Cały czas możemy zaobserwować przewagę Polski zachodniej, metropolii i płci męskiej w kwestiach związanych z zarobkami czy wykształceniem. W kontekście pracy rozważyliśmy również zdrowie psychiczne i fizyczne. Wbrew naszym przypuszczeniom, czas przeznaczony na pracę nie wykazywał większego związku ze stanem zdrowia fizycznego badanej próbki społeczeństwa. Udało nam się jednak wykazać wpływ zadowolenia z pracy na ogólne zadowolenie z życia. Przeprowadzone analizy pokazują jak ważną rolę w życiu człowieka odgrywa praca.