

WSTĘPNE INFORMACJE O ZBIORZE:

- 1) Typy zmiennych
- 2) Czy zmienne zawierają braki danych? co jest tego powodem?
- 3) Czy dane są tidy?
- 4) Wybór zmiennych , których użyjemy w analizie (minimum 2 zmienne kategoryjne i 3 numeryczne)
- 5) Wstępna eksploracja danych- liczba kategorii, rozkłady, wizualizacja poszczególnych zmiennych i zależności między nimi
- 6) Zadajemy sobie pytanie, które zmienne mogą być ze sobą powiązane i które powiązanie chcemy badać- można postawić wstępną hipotezę
- 7) Dobieramy metody statystyczne oraz tabele i wizualizacje do typu zmiennych oraz przeprowadzonych analiz

ZAWARTOŚĆ RAPORTU:

- 1) Strona tytułowa z autorami i tytułem pracy
- 2) Spis treści
- 3) Opis źródła danych i przedstawienie wszelkich dodatkowych informacji o danych. Z jakiej strony były dane? Z jakiego roku? Czy to dane open-source?
- 4) Skrótowy opis zawartości danych(wyjaśnienie znaczenia zmiennych poddanych analizie, wymienienie zmiennych które pominęliśmy
- 5) Skrótowy opis przeprowadzonego procesu czyszczenia danych- o ile był konieczny(czy występowały braki danych, błędy, bądź wartości odstające?)
- 6) Opis celu i kierunku przeprowadzonych analiz. Jakie problemy będziemy badać oraz na jakie pytania będziemy chcieli odpowiedzieć(albo spróbować odpowiedzieć) poprzez analizę danych? Jakie zależności chcemy ogólnie opisać w pracy?
- 7) Przedstawiając cel analiz można przygotować kilka pytań, na które będziemy starać się odpowiedzieć w raporcie np. jaki kierunek studiów jest najczęściej wybierany przez astronautów? czy płeć różnicuje liczbę podróży kosmicznych? jak zarobki są powiązane z wiekiem? w jaki sposób skorelowane są waga i wzrost? Ogólnie- próbujemy odpowiedzieć na pytanie, jakie są charakterystyki poszczególnych zmiennych oraz powiązania między zmiennymi
- 8) Charakterystyka rozkładu wybranych pojedynczych zmiennych za pomocą tabel bądź wykresów
- 9) Charakterystyka związków między wybranymi zmiennymi za pomocą tabel, wykresów jak również współczynników korelacji czy testów statystycznych
- 10) Wnioski z przeprowadzonych analiz. Jakie wyniki uzyskano? Jak można je interpretować? Jak można odpowiedzieć na zadane na początku pytania?
- 11) W raporcie powinny znaleźć się minimum cztery tabele i cztery typy wykresów. Co najmniej dwa wykresy powinny przedstawić porównanie dwóch(lub więcej) zmiennych.
- 12) W raporcie należy zastosować przynajmniej jedną z metod wymienionych w trzeciej części materiałów (test T, test χ^2 , korelacja Spearmana albo korelacja Pearsona)
- 13) Każda wizualizacja oraz tabelka powinna być w jakiś sposób skomentowana i zinterpretowana w tekście raportu
- 14) W tekście nie powtarzamy dokładnie(liczba po liczbie) danych, które są zawarte w tabeli- tekst służy do podsumowania wyników z tabeli
- 15) Nie przedstawiamy identycznych liczb w tabeli i na wykresie
- 16) Do raportu dołączone powinny być:

- Kod źródłowy PYTHON(służący do wczytania, ewentualnego czyszczenia danych jak również do przedstawionych w raporcie analiz oraz wykresów- nie musi zawierać etapu eksploracji danych)
- plik z danymi albo link do pliku

WYMOGI FORMALNE:

- raport w TeXu lub MS Office
- Rozmiar czcionki 12
- czcionka Times New Roman
- Interlinia 1,15 do 1,5
- Tekst wyjustowany
- Cytowania wg. standardu APA