**Programowanie Laboratorium 10.**

Należy załadować pakiety:

**import** **os import** **pandas** **as** **pd # pd to skrót dla pandas # pandas** - biblioteka do obsługi danych tabelarycznych

Zadanie 1***. Tutorial - pakiet pandas***

## Tworzenie ramek danych i zapisywanie do pliku

* utwórz dwie listy (to są Twoje przykładowe dane)

*imie = ['Bobi','Jess','Jess','Marta','Janek','Melania','Jess', 'Melania']
rok = [1968, 1999, 1955, 2013, 1977, 1978, 2003, 2017]*

* połącz dwie listy ze sobą korzystając z funkcji zip

DataSet1 = list(zip(imie,rok))

* utwórz obiekt **DataFrame** (ramka danych) korzystając z biblioteki pandas

d1 = pd.DataFrame(data = DataSet1, columns=['Imie', 'RokUrodzenia'])

print(d1)

* zapisz dane do pliku DatyUrodzin.csv, DatyUrodzin.txt i DatyUrodzin.xlsx

d1.to\_csv('DatyUrodzin.csv',index=**False**,header=**False**)

d1.to\_csv('DatyUrodzin.txt',index=**False**,header=**False**)

d1.to\_excel('DatyUrodzin.xlsx', index=**False**)

## Odczyt danych operacje na kolumnach i wierszach

* odczytaj na 3 sposoby swoje dane korzystając z funkcji ***read\_csv***, podaj ścieżkę do pliku, zwróć uwagę na nagłówki

sciezkaPlik1 = r'C:\Users\births1880.csv'

sciezkaPlik2 = r'C:\Users\births1880.xlsx'

mojeDane0 = pd.read\_csv(sciezkaPlik1)

mojeDane1 = pd.read\_csv(sciezkaPlik1, header=**None**)

mojeDane2 = pd.read\_csv(sciezkaPlik1, names=['Imie', 'RokUrodzenia']) mojeDane3 = pd.read\_excel(sciezkaPlik2,0,index\_col=**'RokUrodzenia'**)

* usuń plik z danymi

os.remove(sciezkaPlik)

* ponownie zapisz dane do pliku o nowej nazwie, wczytaj plik dane przypisz do zmiennej *mojeDane4* następnie sprawdź typ danych

mojeDane4.dtypes

* posortuj wartości kolumny, wyświetl dwa pierwsze wiersze danych i trzy ostatnie

dSort = mojeDane4.sort\_values(['RokUrodzenia'],ascending=**False**)

dSort.head(2)

dSort.tail(3)

* znajdź minimalną i maksymalną wartość kolumny **RokUrodzenia**, korzystając z funkcji **min** i **max** określ imiona osób które urodziły się najwcześniej i najpóźniej, posortuj dane ze względu na rok urodzenia, wyświetl osoby które urodziły się przed rokiem 2000

MaxValue = dSort['RokUrodzenia'].max()

MaxImie =

dSort['Imie'][dSort['RokUrodzenia'] == MaxValue].values

print(mojeDane4[mojeDane4[**'RokUrodzenia'**] < 2000])

* wskaż imiona które nie powtarzają się, wyświetl podstawowe informacje o danych

mojeDane4['Imie'].unique()

 lub

 **for** i **in** mojeDane4['Imie'].unique():

 print(i)

mojeDane4['Imie'].describe()

* pogrupuj dane względem imion i zsumuj dla pogrupowanych imion daty urodzeń

imieGrup = mojeDane4.groupby('Imie')

imieGrup.sum()

* zmień nazwy kolumn na: 'names', 'Birth', dodaj nową kolumnę 'Nowa' zawierającą liczbę 10, a następnie ją usuń, wstaw indeksy

mojeDane4.columns = ['names', 'Birth']

mojeDane4['Nowa'] = 10

**del** mojeDane4['Nowa']

i = [**'a'**,**'b'**,**'c'**,**'d'**,**'e'**,**'f'**,**'g'**,**'h'**]
mojeDane4.index = i

* zastąp kolumnę 'Birth' liczbami od 1 do 8,

mojeDane4['Nowa'] = [1,2,3,4,5,6,7,8]

* wyświetl wiersz indeksowany jako 'c' i wiersze od 'a' do 'd'

print(mojeDane4.loc[**'c'**])

print(mojeDane4.loc[**'a'**:**'d'**])

* podziel zbiór na podzbiory ze względu na indeksy, wykonaj transpozycję danych, uporządkuj dane: wiersze kolumna Birth kolumny names, zwróć uwagę na inny sposób odwoływania się do zmiennych

mojeDane4.ix[0:3,['names', 'Birth']]

 mojeDane4.T

mojeDane4.crosstab(mojeDane4.Birth,mojeDane4.names)

**Zadanie do samodzielnej realizacji:**

Wczytaj plik filmy.csv, następnie zapoznaj się z danymi i wykonaj polecenia:

* wyświetl wszystkie nazwy kolumn
* wyświetl 5 pierwszych wierszy
* wyświetl wszystkie nazwiska reżyserów filmów
* korzystając z funkcji crosstab wyświetl tytuły filmów w zależności od reżyserów
* wyświetl wszystkie filmy Jamesa Camerona
* wyświetl filmy które są Polskie i są w języku polskim
* ile filmy amerykańskich posiadało budżet większy niż 250 milionów dolarów
* jaki polski film był najdroższy
* podaj 10 filmów które miały największą liczbę lajków na Facebooku
* w jakim kraju nakręcono najwięcej filmów w kolorze
* posortuj dane rosnąco ze względu na rok produkcji
* wymień wszystkie filmy w których grali: Leonardo DiCaprio lub Johnny Depp