SPRAWOZDANIE
ANALIZA
BAZY DANYCH
„WEATHER”

Autorzy:
Kaczmarkiewicz Natalia
Machońko Adam
Szymański Jakub

Analiza i wizualizacja danych
Kognitywistyka rok II

Spis treści:

I. Słowem wstępu

W pracy zostanie przedstawiona analiza wybranych zmiennych ze zbioru danych dążąca do uzyskania właściwych informacji z nich płynących. W dodatku dokonana zostanie próba odpowiedzi na przykładowe pytania, które można skonstruować wykorzystując informacje zawarte w bazie danych, skupiające się na niektórych, opisanych w pracy zmiennych i pomiarach z obszernego zbioru danych ,,Weather", poparta wykresami wykonanymi odpowiednim programem, obliczeniami oraz opisem słownym.

II. Wstępna analiza wybranych zmiennych

Analiza przykładowych zmiennych, które zostaną także użyte w pytaniach, znajdujących się w kolejnym zadaniu.

A) Zmienna 1: średnia temperatura w Chicago przez 2 lata:

Liczba obserwacji:	731
Średnia:	52
Mediana:	54
Odchylenie standardowe:	19,54

Średnia temperatura w Chicago na przestrzeni 2 lat waha się od 32,46°F do 71,54°F.

B) Zmienna 2: średnia temperatura w San Diego przez 2 lata:

Liczba obserwacji:	731
Mediana:	66
Średnia:	66
Odchylenie standardowe:	5,97

Średnia temperatura w San Diego na przestrzeni 2 lat waha się od 60,03°F do 71,97°F.

C) Zmienna 3: Średnia widzialność w Auckland w ciągu 2 lat:

Liczba obserwacji:	731
Odchylenie standardowe:	0,311
Średnia:	5,94
Mediana:	6

Średnia widzialność w Auckland na przestrzeni 2 lat waha się od 5,629 do 6,251.

D) Zmienna 4: Średnia wilgotność w Auckland w ciągu 2 lat:

Liczba obserwacji:	731
Średnia:	80,41
Mediana:	81
Odchylenie standardowe:	7,94

Średnia wilgotność w Auckland na przestrzeni 2 lat waha się od 72,47g/m^3 do 88,53g/m^3.

Nie zaobserwowano braków danych, w żadnej z wymienionych zmiennych.

III. Część główna – pytania i odpowiedzi

1. Uśrednienie danych dla miesięcy:

Tabela 1:Uśrednienie danych dla miesięcy - Auckland

Auckland:												
	2016											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Temperatura (°F)	68.81	70.59	67.64	62.80	60.90	54.73	52.84	52.16	56.13	58.39	60.97	63.55
Punkt rosy (°F)	61.94	63.59	60.70	55.90	55.87	50.83	47.39	46.83	51.27	51.77	53.07	54.90
Wilgotność (g/m^3)	79.12	78.97	79.23	79.20	84.20	86.97	82.03	82.06	84.16	79.00	76.17	73.81
Ciśnienie (Hg)	29.98	29.99	30.13	30.14	29.87	30.08	29.93	29.98	29.97	29.90	29.83	29.97
Widzialność	5.94	6.00	6.00	6.00	5.93	5.83	5.77	5.94	5.97	5.97	6.00	6.00
Wiatr	9.19	8.31	8.48	7.30	9.90	8.70	1248	909	970	1139	1357	11.19
	2017											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Temperatura (°F)	65.77	68.57	66.64	63.03	56.45	53.70	51.97	54.51	56.30	59.06	62.00	68.19
Punkt rosy (°F)	56.22	60.89	60.67	58.16	51.00	49.53	47.90	49.97	50.13	53.03	53.63	58.03
Wilgotność (g/m^3)	72.25	78.32	81.48	84.40	82.68	86.13	86.32	85.64	80.60	81.00	75.13	71.16
Ciśnienie (Hg)	29.89	30.01	30.01	30.00	30.09	30.04	19.85	29.95	29.79	30.00	30.10	29.97
Widzialność	6.00	5.96	6.00	6.00	5.87	5.66	5.90	5.97	6.00	6.00	6.00	6.00
Wiatr	13.29	8.25	9.13	7.63	8.35	7.43	8.19	8.00	11.40	12.29	8.77	9.48

Tabela 2: Uśrednienie danych dla miesięcy - Mumbai

Mumbai												
	2016											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Temperatura (°F)	74.48	75.72	81.80	83.77	87.26	85.03	80.32	81.16	80.23	80.94	79.90	78.13
Punkt rosy (°F)	56.45	62.72	67.29	72.53	77.65	77.87	77.52	76.87	75.83	71.61	58.97	59.19
Wilgotność (g/m^3)	48.80	59.31	57.54	65.50	71.29	77.27	87.39	83.58	82.47	69.71	45.27	47.77
Ciśnienie (Hg)	29.94	29.92	29.87	29.80	29.74	29.68	29.67	29.69	29.77	29.82	29.88	29.90
Widzialność	1.65	1.83	1.94	2.13	2.48	1.83	1.77	1.87	1.77	1.87	1.87	1.71
Wiatr	3.84	4.34	5.19	6.13	7.42	8.77	8.13	10.19	6.57	5.39	4.37	3.77
	2017											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Temperatura (°F)	74.90	77.96	79.94	83.10	85.87	84.30	81.84	81.70	82.90	84.35	80.83	75.81
Punkt rosy (°F)	57.71	59.25	66.39	73.60	77.32	78.13	77.71	77.00	77.17	74.23	61.57	62.23
Wilgotność (g/m^3)	49.74	48.43	57.58	66.93	70.55	78.60	83.19	82.68	80.77	70.87	48.30	58.23
Ciśnienie (Hg)	29.93	29.92	29.84	29.80	29.75	29.65	29.67	29.68	29.74	29.77	29.85	29.90
Widzialność	1.58	1.89	2.00	2.10	2.55	2.13	1.71	1.77	1.97	1.90	1.83	1.65
Wiatr	3.94	4.25	5.42	5.57	7.00	8.63	10.84	8.51	5.27	4.13	4.10	4.09

Tabela 3: Uśrednienie danych dla miesięcy - Beijing

Bejing												
	2016											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Temperatura (°F)	22.48	32.14	45.84	59.90	68.32	76.63	80.87	80.52	70.53	55.09	38.33	30.84
Punkt rosy (°F)	1.48	6.79	17.97	32.27	44.84	59.37	69.39	67.26	57.70	45.06	25.10	14.94
Wilgotność (g/m^3)	41.81	35.07	34.87	38.80	45.03	57.67	69.42	66.26	66.23	72.39	62.13	55.58
Ciśnienie (Hg)	30.42	30.32	30.15	29.88	29.82	29.67	29.67	29.75	29.93	30.16	30.26	30.35
Widzialność	7.48	10.79	7.65	7.93	8.52	6.73	4.87	7.19	7.63	5.52	4.30	4.94
Wiatr	7.55	8.38	6.06	7.17	6.77	5.33	4.94	5.06	4.87	4.97	5.37	4.65
	2017											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Temperatura (°F)	27.74	34.54	45.81	60.10	71.65	76.47	81.16	78.45	70.97	54.13	37.87	29.32
Punkt rosy (°F)	9.13	8.61	20.06	31.37	45.45	55.40	72.09	68.06	56.93	44.13	15.87	6.68
Wilgotność (g/m^3)	49.39	36.54	38.45	36.43	40.90	50.53	74.97	73.32	64.53	72.45	46.03	41.22
Ciśnienie (Hg)	30.39	30.30	30.16	29.88	29.81	29.72	29.65	29.72	29.88	30.27	30.26	30.36
Widzialność	6.52	8.82	8.77	8.53	8.42	8.20	5.35	6.41	6.37	5.68	9.93	9.97
Wiatr	6.55	7.36	6.74	6.67	6.55	5.63	4.90	4.74	4.67	4.45	6.43	6.32

Tabela 4: Uśrednienie danych dla miesięcy - Chicago

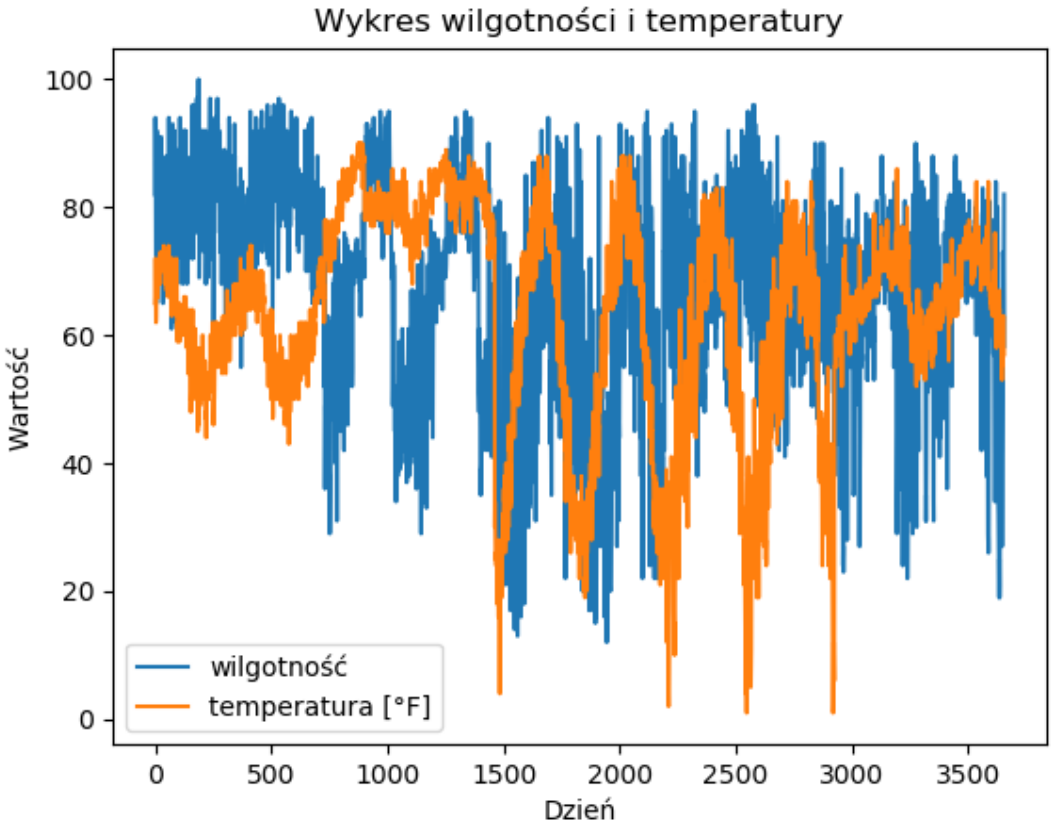
Chicago												
	2016											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Temperatura (°F)	25.03	30.62	43.55	48.00	59.90	71.77	75.71	76.06	69.80	57.13	47.10	25.29
Punkt rosy (°F)	18.45	21.59	34.55	35.63	46.00	57.00	64.00	65.16	58.53	48.58	37.90	18.81
Wilgotność (g/m^3)	73.84	68.89	71.39	65.63	63.87	63.27	70.48	70.68	69.70	73.06	72.03	75.58
Ciśnienie (Hg)	30.02	30.01	29.95	30.01	29.98	29.95	29.97	30.00	30.05	30.08	30.08	30.12
Widzialność	8.68	8.59	8.61	8.67	8.84	9.63	9.32	9.51	9.70	9.39	8.93	7.97
Wiatr	10.68	11.69	10.77	10.80	9.77	8.73	7.97	7.19	9.20	9.19	9.17	11.35
	2017											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Temperatura (°F)	28.97	38.29	39.65	54.10	58.09	72.57	74.90	71.74	69.73	57.84	40.00	27.42
Punkt rosy (°F)	22.93	27.00	26.68	39.53	44.48	54.93	62.03	58.42	54.33	45.74	29.73	16.68
Wilgotność (g/m^3)	76.65	65.21	63.09	62.30	63.81	57.60	67.19	64.84	60.07	66.81	67.33	64.19
Ciśnienie (Hg)	29.99	29.94	30.15	29.94	29.86	29.89	29.99	30.00	30.04	29.98	30.08	30.09
Widzialność	7.52	8.96	8.16	8.93	9.29	9.80	9.55	9.81	9.73	9.13	8.63	9.00
Wiatr	10.00	10.89	12.29	11.47	10.87	10.06	8.13	7.71	8.00	10.77	10.17	10.58

z

Tabela 5: Uśrednienie danych dla miesięcy – San Diego

San Diego												
	2016											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Temperatura (°F)	58.87	64.34	63.00	65.70	65.32	69.23	71.26	73.09	72.53	70.87	66.10	59.32
Punkt rosy (°F)	46.39	41.79	50.09	49.63	53.19	58.00	61.68	62.39	57.77	55.32	46.57	45.06
Wilgotność (g/m^3)	64.77	48.66	63.90	60.73	66.77	70.43	74.52	71.97	63.90	61.26	53.77	62.22
Ciśnienie (Hg)	30.10	30.08	30.03	29.99	29.98	29.95	29.95	29.92	29.95	29.97	30.04	30.07
Widzialność	9.03	9.03	9.16	9.37	10.00	9.33	9.77	9.74	9.60	9.80	9.40	8.90
Wiatr	4.65	4.00	5.87	6.27	6.68	6.20	6.58	6.74	6.50	5.22	4.63	5.09
	2017											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Temperatura (°F)	58.03	60.07	62.94	65.63	65.23	68.20	73.26	73.32	73.20	71.39	65.47	60.52
Punkt rosy (°F)	44.90	49.21	48.06	50.83	54.13	59.37	63.61	63.35	61.47	54.77	52.80	38.06
Wilgotność (g/m^3)	64.23	68.79	61.13	61.37	71.09	76.40	73.87	73.29	69.43	61.74	65.87	49.48
Ciśnienie (Hg)	30.11	30.06	30.07	29.99	29.95	29.91	29.96	29.92	29.89	29.93	30.04	30.10
Widzialność	9.13	8.71	8.77	9.60	9.71	9.17	9.71	9.19	9.47	9.45	9.33	8.97
Wiatr	5.35	6.54	5.29	6.37	7.45	6.70	6.26	6.23	6.70	5.68	4.03	3.58

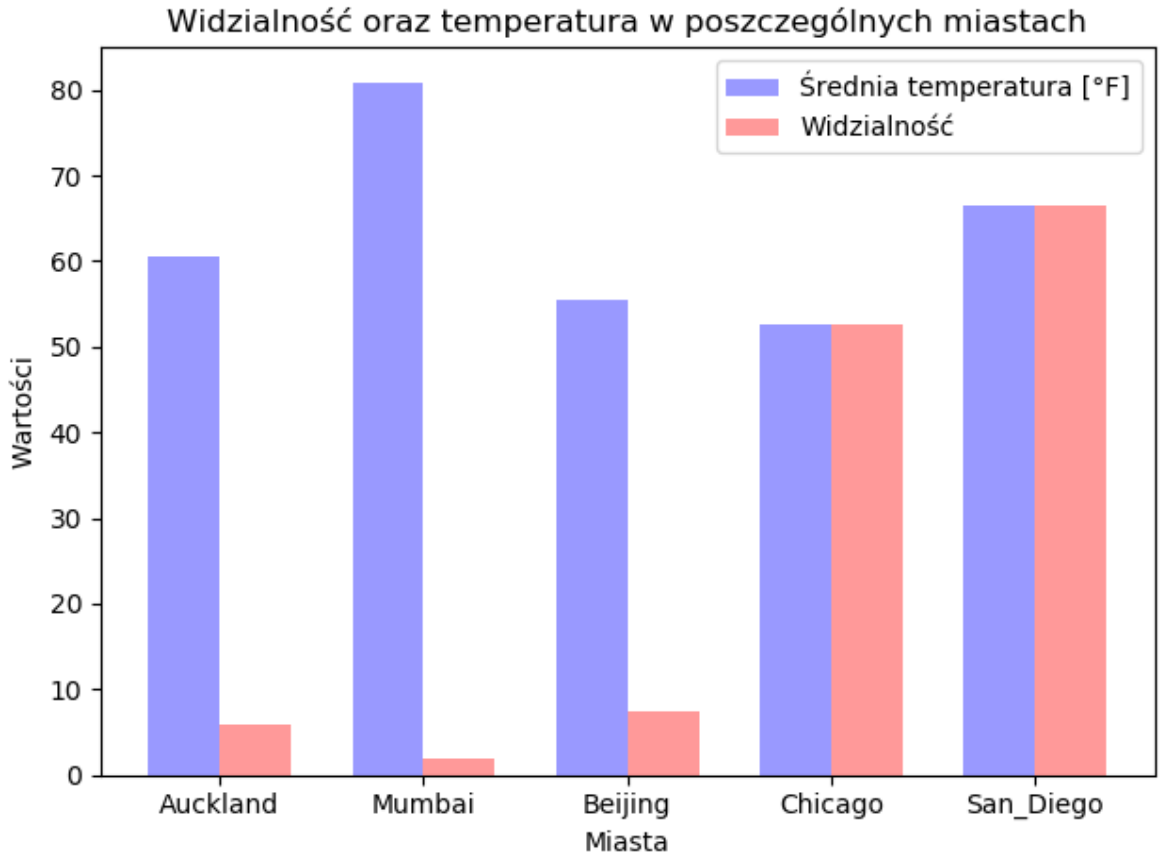
2. Jaki związek ma poziom wilgotności z temperaturą w ciągu dnia? (wykres)



Rysunek 1

Istnieje związek pomiędzy poziomem wilgotności a temperaturą w ciągu dnia, co obrazuje wykres.

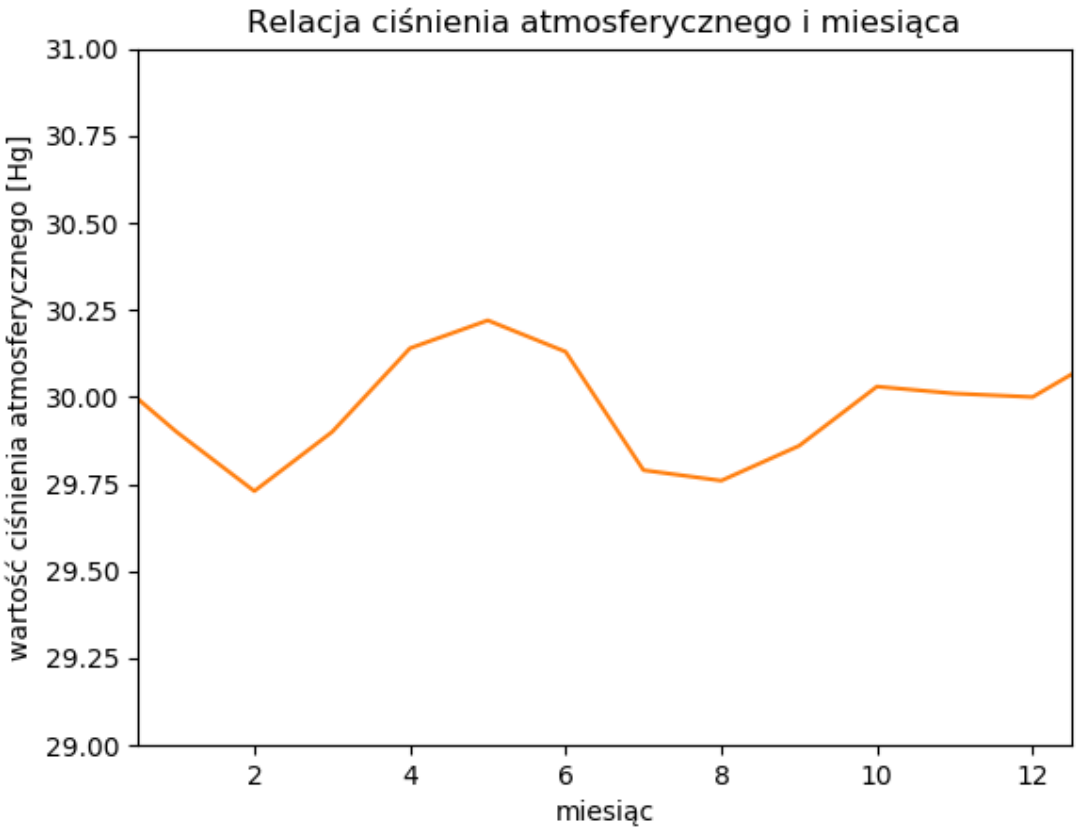
3. Czy w miastach o wysokiej średniej temperaturze widzialność jest większa? (wykres)



Rysunek 2

Widzialność nie ma związku ze średnią temperaturą.
W miastach o wysokiej średniej temperaturze widzialność nie jest większa.

4. Czy ciśnienie atmosferyczne ma związek z miesiącem ? (wykres)



Rysunek 3

Nie ma związku pomiędzy miesiącem a ciśnieniem atmosferycznym.

5. Czy istnieje istotna statystycznie różnica między Auckland a Mumbajem odnośnie wilgotności?

Hipoteza zerowa:

Średnia wilgotność w Auckland jest taka sama jak średnia wilgotność w Mumbaju.

Hipoteza alternatywna:

Miasta będą różnić się pod względem średniej wilgotności,

statistic=21.284667242744483
pvalue=8.934019233865703e-88

$p<0,05$ → Odrzucamy hipotezę zerową.

Wniosek:

Auckland i Mumbaj różnią się pod względem średniej wilgotności.

Istnieje istotna statystycznie różnica pomiędzy Auckland a Mumbajem odnośnie wilgotności. Średnia wilgotność w Auckland jest wyższa od średniej wilgotności w Mumbaju.

6. Czy istnieje różnica między Chicago a San Diego odnośnie temperatury?

Hipoteza zerowa:

Średnie temperatury w Chicago i San Diego nie różnią się.

Hipoteza alternatywna:

Miasta różnią się pod względem średniej temperatury.

statistic=-18.40379894610621
pvalue=3.4572505313339807e-68

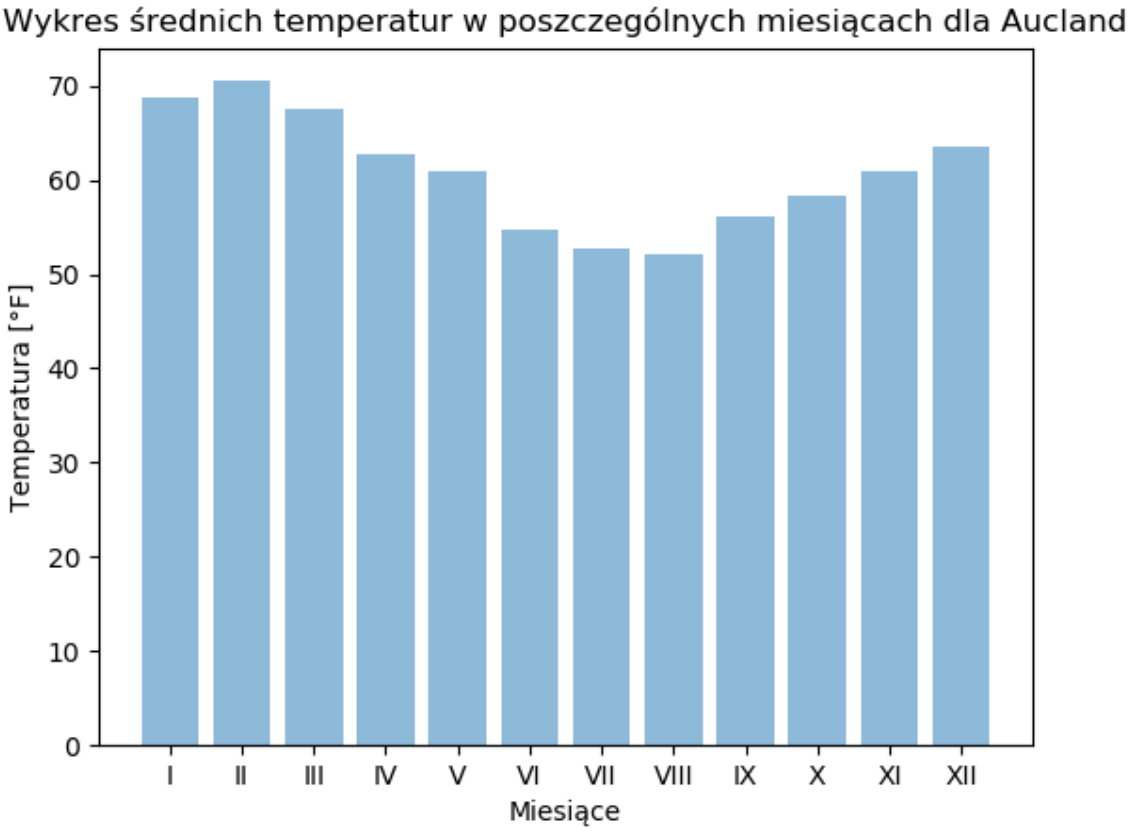
$p<0,05$ → Odrzucamy hipotezę zerową.

Wniosek:

Chicago i San Diego różnią się pod względem średniej temperatury.

Istnieje istotna statystycznie różnica w średnich temperaturach dla Chicago i San Diego. Temperatury osiągają średnio wyższe wartości w San Diego, a średnio niższe wartości w Chicago.

7. Wykres z temperatur miesięcznych w Auckland. (wykres)



Rysunek 4

Najniższa średnia temperatura miała wartość 52,1°F i została odnotowana w Sierpniu.

Najwyższa średnia temperatura miała wartość 70,5°F i została odnotowana w Lutym.

Średnia temperatura roczna wynosiła 60,75°F.

IV. Wnioski

Z bazy danych zostały uzyskane pewne informacje, która ta baza danych zawiera, dotyczące wybranych zmiennych i pytań do tych zmiennych skierowanych.

V. Kod programu

Link: https://github.com/ADmek/Analiza-danych?fbclid=IwAR09nrrgDGzT8wN3ZLr1iTmC9EqhVcuHMZ8yQpIR5HWWkcqnp5Otv79VSQk.

-

```
import matplotlib.pyplot as plt; plt.rcdefaults()
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.ticker import MaxNLocator
from collections import namedtuple
import pandas as pd
import scipy.stats as st
import matplotlib.pyplot as plt
dane=pd.read_csv("weather.csv")
#print(dane.head(0))
#print(dane)
#ZADANIE1
#podział na miesiace danych z Auckland rok 2016
'''columns =
["city","date","year","month","day","high_temp","avg_temp","low_temp","high_dewpt","avg_dewpt","lo
w_dewpt","high_humidity","avg_humidity","low_humidity","high_hg","avg_hg","low_hg","high_vis","avg
_vis","low_vis","high_wind","avg_wind",]
df = pd.DataFrame(dane, columns=columns)
df1=df[df['city'].str.contains('Auckland', na = False)]
rok2016 =df1.loc[df1['year'] == 2016]
aucklandstyczen2016 =rok2016.loc[rok2016['month'] == 1]
aucklandluty2016 =rok2016.loc[rok2016['month'] == 2]
aucklandmarzec2016 =rok2016.loc[rok2016['month'] == 3]
aucklandkwiecień2016 =rok2016.loc[rok2016['month'] == 4]
aucklandmaj2016 =rok2016.loc[rok2016['month'] == 5]
aucklandczerwiec2016 =rok2016.loc[rok2016['month'] == 6]
aucklandlipiec2016 =rok2016.loc[rok2016['month'] == 7]
aucklandsierpien2016 =rok2016.loc[rok2016['month'] == 8]
aucklandwrzesien2016 =rok2016.loc[rok2016['month'] == 9]
aucklandpazdziernik2016 =rok2016.loc[rok2016['month'] == 10]
aucklandlistopad2016 =rok2016.loc[rok2016['month'] == 11]
aucklandgrudzien2016 =rok2016.loc[rok2016['month'] == 12]
#print(aucklandgrudzien2016)
#podział na miesiace danych z Auckland rok 2017
rok2017 =df1.loc[df1['year'] == 2017]
aucklandstyczen2017 =rok2017.loc[rok2017['month'] == 1]
aucklandluty2017 =rok2017.loc[rok2017['month'] == 2]
aucklandmarzec2017 =rok2017.loc[rok2017['month'] == 3]
aucklandkwiecień2017 =rok2017.loc[rok2017['month'] == 4]
aucklandmaj2017 =rok2017.loc[rok2017['month'] == 5]
aucklandczerwiec2017 =rok2017.loc[rok2017['month'] == 6]
aucklandlipiec2017 =rok2017.loc[rok2017['month'] == 7]
aucklandsierpien2017 =rok2017.loc[rok2017['month'] == 8]
aucklandwrzesien2017 =rok2017.loc[rok2017['month'] == 9]
aucklandpazdziernik2017 =rok2017.loc[rok2017['month'] == 10]
aucklandlistopad2017 =rok2017.loc[rok2017['month'] == 11]
aucklandgrudzien2017 =rok2017.loc[rok2017['month'] == 12]
#print(aucklandgrudzien2017)
```

```
#podział na miesiace danych z Mumbaju 2016
df2=df[df['city'].str.contains('Mumbai', na = False)]
rok2016 =df2.loc[df2['year'] == 2016]
mumbaistyczen2016 =rok2016.loc[rok2016['month'] == 1]
mumbailuty2016 =rok2016.loc[rok2016['month'] == 2]
mumbaimarzec2016 =rok2016.loc[rok2016['month'] == 3]
mumbaikwiecień2016 =rok2016.loc[rok2016['month'] == 4]
mumbaimaj2016 =rok2016.loc[rok2016['month'] == 5]
mumbaiczerwiec2016 =rok2016.loc[rok2016['month'] == 6]
mumbailipiec2016 =rok2016.loc[rok2016['month'] == 7]
mumbaisierpien2016 =rok2016.loc[rok2016['month'] == 8]
mumbaiwrzesien2016 =rok2016.loc[rok2016['month'] == 9]
mumbaipazdziernik2016 =rok2016.loc[rok2016['month'] == 10]
mumbailistopad2016 =rok2016.loc[rok2016['month'] == 11]
mumbaigrudzien2016 =rok2016.loc[rok2016['month'] == 12]
#print(mumbaigrudzien2016)

#podział na miesiace danych z Mumbaju 2017
rok2017 =df2.loc[df2['year'] == 2017]
mumbaistyczen2017 =rok2017.loc[rok2017['month'] == 1]
mumbailuty2017 =rok2017.loc[rok2017['month'] == 2]
mumbaimarzec2017 =rok2017.loc[rok2017['month'] == 3]
mumbaikwiecień2017 =rok2017.loc[rok2017['month'] == 4]
mumbaimaj2017 =rok2017.loc[rok2017['month'] == 5]
mumbaiczerwiec2017 =rok2017.loc[rok2017['month'] == 6]
mumbailipiec2017 =rok2017.loc[rok2017['month'] == 7]
mumbaisierpien2017 =rok2017.loc[rok2017['month'] == 8]
mumbaiwrzesien2017 =rok2017.loc[rok2017['month'] == 9]
mumbaipazdziernik2017 =rok2017.loc[rok2017['month'] == 10]
mumbailistopad2017 =rok2017.loc[rok2017['month'] == 11]
mumbaigrudzien2017 =rok2017.loc[rok2017['month'] == 12]
#print(mumbaigrudzien2017)

#podział na miesiące Bejing 2016
df3=df[df['city'].str.contains('Beijing', na = False)]
rok2016 =df3.loc[df3['year'] == 2016]
beijingstyczen2016 =rok2016.loc[rok2016['month'] == 1]
beijingluty2016 =rok2016.loc[rok2016['month'] == 2]
beijingmarzec2016 =rok2016.loc[rok2016['month'] == 3]
beijingkwiecień2016 =rok2016.loc[rok2016['month'] == 4]
beijingmaj2016 =rok2016.loc[rok2016['month'] == 5]
beijingczerwiec2016 =rok2016.loc[rok2016['month'] == 6]
beijinglipiec2016 =rok2016.loc[rok2016['month'] == 7]
beijingsierpien2016 =rok2016.loc[rok2016['month'] == 8]
beijingwrzesien2016 =rok2016.loc[rok2016['month'] == 9]
beijingpazdziernik2016 =rok2016.loc[rok2016['month'] == 10]
beijinglistopad2016 =rok2016.loc[rok2016['month'] == 11]
beijinggrudzien2016 =rok2016.loc[rok2016['month'] == 12]
#print(beijingmaj2016)

#podział na miesiące Bejing 2017
rok2017 =df3.loc[df3['year'] == 2017]
beijingstyczen2017 =rok2017.loc[rok2017['month'] == 1]
beijingluty2017 =rok2017.loc[rok2017['month'] == 2]
beijingmarzec2017 =rok2017.loc[rok2017['month'] == 3]
beijingkwiecień2017 =rok2017.loc[rok2017['month'] == 4]
beijingmaj2017 =rok2017.loc[rok2017['month'] == 5]
beijingczerwiec2017 =rok2017.loc[rok2017['month'] == 6]
beijinglipiec2017 =rok2017.loc[rok2017['month'] == 7]
beijingsierpien2017 =rok2017.loc[rok2017['month'] == 8]
```

```
beijingwrzesien2017 =rok2017.loc[rok2017['month'] == 9]
beijingpazdziernik2017 =rok2017.loc[rok2017['month'] == 10]
beijinglistopad2017 =rok2017.loc[rok2017['month'] == 11]
beijinggrudzien2017 =rok2017.loc[rok2017['month'] == 12]
#print(beijingmaj2017)

#podział na miesiące Chicago 2016
df4=df[df['city'].str.contains('Chicago', na = False)]
rok2016 =df4.loc[df4['year'] == 2016]
chicagostyczen2016 =rok2016.loc[rok2016['month'] == 1]
chicagoluty2016 =rok2016.loc[rok2016['month'] == 2]
chicagomarzec2016 =rok2016.loc[rok2016['month'] == 3]
chicagokwiecień2016 =rok2016.loc[rok2016['month'] == 4]
chicagomaj2016 =rok2016.loc[rok2016['month'] == 5]
chicagoczerwiec2016 =rok2016.loc[rok2016['month'] == 6]
chicagolipiec2016 =rok2016.loc[rok2016['month'] == 7]
chicagosierpien2016 =rok2016.loc[rok2016['month'] == 8]
chicagowrzesien2016 =rok2016.loc[rok2016['month'] == 9]
chicagopazdziernik2016 =rok2016.loc[rok2016['month'] == 10]
chicagolistopad2016 =rok2016.loc[rok2016['month'] == 11]
chicagogrudzien2016 =rok2016.loc[rok2016['month'] == 12]
#print(chicagogrudzien2016)

#podział na miesiące Chicago 2017
rok2017 =df4.loc[df4['year'] == 2017]
chicagostyczen2017 =rok2017.loc[rok2017['month'] == 1]
chicagoluty2017 =rok2017.loc[rok2017['month'] == 2]
chicagomarzec2017 =rok2017.loc[rok2017['month'] == 3]
chicagokwiecień2017 =rok2017.loc[rok2017['month'] == 4]
chicagomaj2017 =rok2017.loc[rok2017['month'] == 5]
chicagoczerwiec2017 =rok2017.loc[rok2017['month'] == 6]
chicagolipiec2017 =rok2017.loc[rok2017['month'] == 7]
chicagosierpien2017 =rok2017.loc[rok2017['month'] == 8]
chicagowrzesien2017 =rok2017.loc[rok2017['month'] == 9]
chicagopazdziernik2017 =rok2017.loc[rok2017['month'] == 10]
chicagolistopad2017 =rok2017.loc[rok2017['month'] == 11]
chicagogrudzien2017 =rok2017.loc[rok2017['month'] == 12]
#print(chicagogrudzien2017)

#podział na miesiące San Diego 2016
df4=df[df['city'].str.contains('San Diego', na = False)]
rok2016 =df4.loc[df4['year'] == 2016]
sandiegostyczen2016 =rok2016.loc[rok2016['month'] == 1]
sandiegoluty2016 =rok2016.loc[rok2016['month'] == 2]
sandiegomarzec2016 =rok2016.loc[rok2016['month'] == 3]
sandiegokwiecień2016 =rok2016.loc[rok2016['month'] == 4]
sandiegomaj2016 =rok2016.loc[rok2016['month'] == 5]
sandiegoczerwiec2016 =rok2016.loc[rok2016['month'] == 6]
sandiegolipiec2016 =rok2016.loc[rok2016['month'] == 7]
sandiegosierpien2016 =rok2016.loc[rok2016['month'] == 8]
sandiegowrzesien2016 =rok2016.loc[rok2016['month'] == 9]
sandiegopazdziernik2016 =rok2016.loc[rok2016['month'] == 10]
sandiegolistopad2016 =rok2016.loc[rok2016['month'] == 11]
sandiegogrudzien2016 =rok2016.loc[rok2016['month'] == 12]
#print(sandiegogrudzien2016)

#podział na miesiące San Diego 2017
rok2017 =df4.loc[df4['year'] == 2017]
sandiegostyczen2017 =rok2017.loc[rok2017['month'] == 1]
sandiegoluty2017 =rok2017.loc[rok2017['month'] == 2]
```

```
sandiegomarzec2017 =rok2017.loc[rok2017['month'] == 3]
sandiegokwiecień2017 =rok2017.loc[rok2017['month'] == 4]
sandiegomaj2017 =rok2017.loc[rok2017['month'] == 5]
sandiegoczerwiec2017 =rok2017.loc[rok2017['month'] == 6]
sandiegolipiec2017 =rok2017.loc[rok2017['month'] == 7]
sandiegosierpien2017 =rok2017.loc[rok2017['month'] == 8]
sandiegowrzesien2017 =rok2017.loc[rok2017['month'] == 9]
sandiegopazdziernik2017 =rok2017.loc[rok2017['month'] == 10]
sandiegolistopad2017 =rok2017.loc[rok2017['month'] == 11]
sandiegogrudzien2017 =rok2017.loc[rok2017['month'] == 12]
#print(sandiegogrudzien2017)'''

#urednienie danych dla miesiecyaucklandstyczen2016 =rok2016.loc[rok2016['month'] == 1]
'''print(aucklandstyczen2016.mean())
print(aucklandluty2016.mean())
print(aucklandmarzec2016.mean())
print(aucklandkwiecień2016.mean())
print(aucklandmaj2016.mean())
print(aucklandczerwiec2016.mean())
print(aucklandlipiec2016.mean())
print(aucklandsierpien2016.mean())
print(aucklandwrzesien2016.mean())
print(aucklandpazdziernik2016.mean())
print(aucklandlistopad2016.mean())
print(aucklandgrudzien2016.mean())'''

'''print(aucklandstyczen2017.mean())
print(aucklandluty2017.mean())
print(aucklandmarzec2017.mean())
print(aucklandkwiecień2017.mean())
print(aucklandmaj2017.mean())
print(aucklandczerwiec2017.mean())
print(aucklandlipiec2017.mean())
print(aucklandsierpien2017.mean())
print(aucklandwrzesien2017.mean())
print(aucklandpazdziernik2017.mean())
print(aucklandlistopad2017.mean())
print(aucklandgrudzien2017.mean())'''

'''print(mumbaistyczen2016.mean())
print(mumbailuty2016.mean())
print(mumbaimarzec2016.mean())
print(mumbaikwiecień2016.mean())
print(mumbaimaj2016.mean())
print(mumbaiczerwiec2016.mean())
print(mumbailipiec2016.mean())
print(mumbaisierpien2016.mean())
print(mumbaiwrzesien2016.mean())
print(mumbaipazdziernik2016.mean())
print(mumbailistopad2016.mean())
print(mumbaigrudzien2016.mean())'''

'''print(mumbaistyczen2017.mean())
print(mumbailuty2017.mean())
print(mumbaimarzec2017.mean())
print(mumbaikwiecień2017.mean())
print(mumbaimaj2017.mean())
print(mumbaiczerwiec2017.mean())
print(mumbailipiec2017.mean())
print(mumbaisierpien2017.mean())
print(mumbaiwrzesien2017.mean())
```

```
print(mumbaipazdziernik2017.mean())
print(mumbailistopad2017.mean())
print(mumbaigrudzien2017.mean())'''
'''
print(beijingstyczen2016.mean())
print(beijingluty2016.mean())
print(beijingmarzec2016.mean())
print(beijingkwiecień2016.mean())
print(beijingmaj2016.mean())
print(beijingczerwiec2016.mean())
print(beijinglipiec2016.mean())
print(beijingsierpien2016.mean())
print(beijingwrzesien2016.mean())
print(beijingpazdziernik2016.mean())
print(beijinglistopad2016.mean())
print(beijinggrudzien2016.mean())

print(beijingstyczen2017.mean())
print(beijingluty2017.mean())
print(beijingmarzec2017.mean())
print(beijingkwiecień2017.mean())
print(beijingmaj2017.mean())
print(beijingczerwiec2017.mean())
print(beijinglipiec2017.mean())
print(beijingsierpien2017.mean())
print(beijingwrzesien2017.mean())
print(beijingpazdziernik2017.mean())
print(beijinglistopad2017.mean())
print(beijinggrudzien2017.mean())

print(chicagostyczen2016.mean())
print(chicagoluty2016.mean())
print(chicagomarzec2016.mean())
print(chicagokwiecień2016.mean())
print(chicagomaj2016.mean())
print(chicagoczerwiec2016.mean())
print(chicagolipiec2016.mean())
print(chicagosierpien2016.mean())
print(chicagowrzesien2016.mean())
print(chicagopazdziernik2016.mean())
print(chicagolistopad2016.mean())
print(chicagogrudzien2016.mean())

print(chicagostyczen2017.mean())
print(chicagoluty2017.mean())
print(chicagomarzec2017.mean())
print(chicagokwiecień2017.mean())
print(chicagomaj2017.mean())
print(chicagoczerwiec2017.mean())
print(chicagolipiec2017.mean())
print(chicagosierpien2017.mean())
print(chicagowrzesien2017.mean())
print(chicagopazdziernik2017.mean())
print(chicagolistopad2017.mean())
print(chicagogrudzien2017.mean())

print(sandiegostyczen2016.mean())
print(sandiegoluty2016.mean())
print(sandiegomarzec2016.mean())
print(sandiegokwiecień2016.mean())
print(sandiegomaj2016.mean())
```

```
print(sandiegoczerwiec2016.mean())
print(sandiegolipiec2016.mean())
print(sandiegosierpien2016.mean())
print(sandiegowrzesien2016.mean())
print(sandiegopazdziernik2016.mean())
print(sandiegolistopad2016.mean())
print(sandiegogrudzien2016.mean())
print(sandiegostyczen2017.mean())
print(sandiegoluty2017.mean())
print(sandiegomarzec2017.mean())
print(sandiegokwiecień2017.mean())
print(sandiegomaj2017.mean())
print(sandiegoczerwiec2017.mean())
print(sandiegolipiec2017.mean())
print(sandiegosierpien2017.mean())
print(sandiegowrzesien2017.mean())
print(sandiegopazdziernik2017.mean())
print(sandiegolistopad2017.mean())
print(sandiegogrudzien2017.mean())'''
#ZADANIE 2
'''wilgotnosc = dane['avg_humidity']
print(wilgotnosc)
temperatura = dane['avg_temp']
print(temperatura)
print(st.pearsonr(wilgotnosc,temperatura))
#plt.plot(wilgotnosc, temperatura)
plt.title('Wykres wilgotnosci i temperatury')
plt.plot(wilgotnosc,label='wilgotnosc')
plt.plot(temperatura,label='temperatura [°F]')
plt.legend()
plt.xlabel('Dzien')
plt.ylabel('Wartosc')
plt.show()'''
#ZADANIE3
#Srednia temperatura
'''columns = ["city","avg_temp","avg_vis"]
df = pd.DataFrame(dane, columns=columns)
Aucklandtemp=df[df['city'].str.contains('Auckland', na = False)]
sredniatempauckland = Aucklandtemp['avg_temp']
print(sredniatempauckland.mean())
columns = ["city","avg_temp","avg_vis"]
df = pd.DataFrame(dane, columns=columns)
Mumbaitemp=df[df['city'].str.contains('Mumbai', na = False)]
sredniatempmumbai = Mumbaitemp['avg_temp']
print(sredniatempmumbai.mean())
columns = ["city","avg_temp","avg_vis"]
df = pd.DataFrame(dane, columns=columns)
Beijingtemp=df[df['city'].str.contains('Beijing', na = False)]
sredniatempbeijing = Beijingtemp['avg_temp']
print(sredniatempbeijing.mean())
columns = ["city","avg_temp","avg_vis"]
df = pd.DataFrame(dane, columns=columns)
Chicagotemp=df[df['city'].str.contains('Chicago', na = False)]
sredniatempchicago = Chicagotemp['avg_temp']
print(sredniatempchicago.mean())
columns = ["city","avg_temp","avg_vis"]
```

```
df = pd.DataFrame(dane, columns=columns)
Sandiegotemp=df[df['city'].str.contains('San Diego', na = False)]
sredniatempsandiego = Sandiegotemp['avg_temp']
print(sredniatempsandiego.mean())
#Srednia widzialnoć
columns = ["city","avg_temp","avg_vis"]
df = pd.DataFrame(dane, columns=columns)
Aucklandvis=df[df['city'].str.contains('Auckland', na = False)]
sredniavisauckland = Aucklandvis['avg_vis']
print(sredniavisauckland.mean())
columns = ["city","avg_temp","avg_vis"]
df = pd.DataFrame(dane, columns=columns)
Mumbaitemp=df[df['city'].str.contains('Mumbai', na = False)]
sredniavismumbai = Mumbaitemp['avg_vis']
print(sredniavismumbai.mean())
columns = ["city","avg_temp","avg_vis"]
df = pd.DataFrame(dane, columns=columns)
Beijingvis=df[df['city'].str.contains('Beijing', na = False)]
sredniavisbeijing = Beijingvis['avg_vis']
print(sredniavisbeijing.mean())
columns = ["city","avg_temp","avg_vis"]
df = pd.DataFrame(dane, columns=columns)
Chicagovis=df[df['city'].str.contains('Chicago', na = False)]
sredniavischicago = Chicagovis['avg_temp']
print(sredniavischicago.mean())
columns = ["city","avg_temp","avg_vis"]
df = pd.DataFrame(dane, columns=columns)
Sandiegovis=df[df['city'].str.contains('San Diego', na = False)]
sredniavissandiego = Sandiegovis['avg_temp']
print(sredniavissandiego.mean())
n_groups = 5
average_temperature = (60.6, 80.9, 55.5, 52.7, 66.6)
#std_men = (5.9, 1.9, 4, 1, 2)
average_visibility = (5.9, 1.9, 7.3, 52.7, 66.6)
#std_women = (3, 5, 2, 3, 3)
fig, ax = plt.subplots()
index = np.arange(n_groups)
bar_width = 0.35
opacity = 0.4
error_config = {'ecolor': '0.3'}
rects1 = ax.bar(index, average_temperature, bar_width,
                alpha=opacity, color='b',
                 error_kw=error_config,
                label='Srednia temperatura [°F]')
rects2 = ax.bar(index + bar_width, average_visibility, bar_width,
                alpha=opacity, color='r',
                  error_kw=error_config,
                label='Widzialnosc')
ax.set_xlabel('Miasta')
ax.set_ylabel('Wartosci')
ax.set_title('Widzialnosc oraz temperatura w poszczegolnych miastach')
ax.set_xticks(index + bar_width / 2)
ax.set_xticklabels(('Auckland', 'Mumbai', 'Beijing', 'Chicago', 'San_Diego'))
ax.legend()
fig.tight_layout()
plt.show()'''
```

```
#ZADANIE4
'''cisnienieatm = dane['avg_hg']
print(cisnienieatm)
miesiac = dane['month']
#print(miesiac)
plt.plot(miesiac, label ='miesiac')
plt.plot(cisnienieatm, label= "cisnienie atmosferyczne")
plt.title('Relacja cisnienia atmosferycznego i miesiąca')
plt.ylabel('cisnienie atmosferyczne')
plt.ylim(29,31)
plt.ylabel("Wartosc cisnienia atmosferycznego [Hg]")
plt.xlabel('miesiac')
plt.xlim(0.5,12.5)
plt.tight_layout()
plt.show()
#nie wiem czy ta korelacja jest tutaj poprawna raczej nie, wykres korelacji jeli jaka jest można
opisać
#z wykresu
print(st.spearmanr(miesiac,cisnienieatm))'''
#ZADANIE5
#srednia wilgotnosc w Auckland
'''columns =
["city","date","year","month","day","high_temp","avg_temp","low_temp","high_dewpt","avg_dewpt","lo
w_dewpt","high_humidity","avg_humidity","low_humidity","high_hg","avg_hg","low_hg","high_vis","avg
_vis","low_vis","high_wind","avg_wind",]
df = pd.DataFrame(dane, columns=columns)
Aucklandwilgotnosc=df[df['city'].str.contains('Auckland', na = False)]
sredniaaucklandwilgotnosc = Aucklandwilgotnosc['avg_humidity']
print(sredniaaucklandwilgotnosc.mean())
#srednia wilgotnosc w Mumbaju
columns =
["city","date","year","month","day","high_temp","avg_temp","low_temp","high_dewpt","avg_dewpt","lo
w_dewpt","high_humidity","avg_humidity","low_humidity","high_hg","avg_hg","low_hg","high_vis","avg
_vis","low_vis","high_wind","avg_wind",]
df = pd.DataFrame(dane, columns=columns)
Mumbaiwilgotnosc=df[df['city'].str.contains('Mumbai', na = False)]
sredniamumbaiwilgotnosc = Mumbaiwilgotnosc['avg_humidity']
print(sredniamumbaiwilgotnosc.mean())
print(st.ttest_ind(sredniaaucklandwilgotnosc,sredniamumbaiwilgotnosc))'''
#ZADANIE6
'''columns =
["city","date","year","month","day","high_temp","avg_temp","low_temp","high_dewpt","avg_dewpt","lo
w_dewpt","high_humidity","avg_humidity","low_humidity","high_hg","avg_hg","low_hg","high_vis","avg
_vis","low_vis","high_wind","avg_wind",]
df = pd.DataFrame(dane, columns=columns)
Chicagotemp=df[df['city'].str.contains('Chicago', na = False)]
sredniatempchicago = Chicagotemp['avg_temp']
print(sredniatempchicago.mean())
columns =
["city","date","year","month","day","high_temp","avg_temp","low_temp","high_dewpt","avg_dewpt","lo
w_dewpt","high_humidity","avg_humidity","low_humidity","high_hg","avg_hg","low_hg","high_vis","avg
_vis","low_vis","high_wind","avg_wind",]
df = pd.DataFrame(dane, columns=columns)
Sandiegotemp=df[df['city'].str.contains('San Diego', na = False)]
sredniatempsandiego = Sandiegotemp['avg_temp']
print(sredniatempsandiego.mean())
```

```
print(st.ttest_ind(sredniatempchicago,sredniatempsandiego))'''

#ZADANIE7
columns =
["city","date","year","month","day","high_temp","avg_temp","low_temp","high_dewpt","avg_dewpt","lo
w_dewpt","high_humidity","avg_humidity","low_humidity","high_hg","avg_hg","low_hg","high_vis","avg
_vis","low_vis","high_wind","avg_wind",]
df = pd.DataFrame(dane, columns=columns)
df1=df[df['city'].str.contains('Auckland', na = False)]
rok2016 =df1.loc[df1['year'] == 2016]
aucklandstyczen2016 =rok2016.loc[rok2016['month'] == 1]
aucklandluty2016 =rok2016.loc[rok2016['month'] == 2]
aucklandmarzec2016 =rok2016.loc[rok2016['month'] == 3]
aucklandkwiecień2016 =rok2016.loc[rok2016['month'] == 4]
aucklandmaj2016 =rok2016.loc[rok2016['month'] == 5]
aucklandczerwiec2016 =rok2016.loc[rok2016['month'] == 6]
aucklandlipiec2016 =rok2016.loc[rok2016['month'] == 7]
aucklandsierpien2016 =rok2016.loc[rok2016['month'] == 8]
aucklandwrzesien2016 =rok2016.loc[rok2016['month'] == 9]
aucklandpazdziernik2016 =rok2016.loc[rok2016['month'] == 10]
aucklandlistopad2016 =rok2016.loc[rok2016['month'] == 11]
aucklandgrudzien2016 =rok2016.loc[rok2016['month'] == 12]

aucklandstyczen2016sredniatemp = aucklandstyczen2016['avg_temp']
aucklandluty2016sredniatemp =aucklandluty2016['avg_temp']
aucklandmarzec2016sredniatemp =aucklandmarzec2016['avg_temp']
aucklandkwiecień2016sredniatemp =aucklandkwiecień2016['avg_temp']
aucklandmaj2016sredniatemp =aucklandmaj2016['avg_temp']
aucklandczerwiec2016sredniatemp =aucklandczerwiec2016['avg_temp']
aucklandlipiec2016sredniatemp =aucklandlipiec2016['avg_temp']
aucklandsierpien2016sredniatemp =aucklandsierpien2016['avg_temp']
aucklandwrzesien2016sredniatemp =aucklandwrzesien2016['avg_temp']
aucklandpazdziernik2016sredniatemp =aucklandpazdziernik2016['avg_temp']
aucklandlistopad2016sredniatemp =aucklandlistopad2016['avg_temp']
aucklandgrudzien2016sredniatemp =aucklandgrudzien2016['avg_temp']

print(aucklandstyczen2016sredniatemp.mean())
print(aucklandluty2016sredniatemp.mean())
print(aucklandmarzec2016sredniatemp.mean())
print(aucklandkwiecień2016sredniatemp.mean())
print(aucklandmaj2016sredniatemp.mean())
print(aucklandczerwiec2016sredniatemp.mean())
print(aucklandlipiec2016sredniatemp.mean())
print(aucklandsierpien2016sredniatemp.mean())
print(aucklandwrzesien2016sredniatemp.mean())
print(aucklandpazdziernik2016sredniatemp.mean())
print(aucklandlistopad2016sredniatemp.mean())
print(aucklandgrudzien2016sredniatemp.mean())

#kod wykresu
#1 to styczen , 12 grudzien
objects = ('1', '2', '3', '4', '5', '6','7' , '8','9', '10','11','12')
y_pos = np.arange(len(objects))
performance = [68.8,70.5,67.6,62.8,60.9,54.7,52.8,52.1,56.1,58.3,60.9,63.5]

plt.bar(y_pos, performance, align='center', alpha=0.5)
plt.xticks(y_pos, objects)
plt.ylabel('Temperatura [°F]')
plt.title('Wykres rednich temperatur w poszczególnych miesiącach')

plt.show()
```