

eval_projet2_solutions

November 10, 2024

Évaluation du projet

Utilisez ce fichier pour écrire le code Python afin de répondre aux questions se trouvant sur Moodle.

0.0.1 À propos des données :

- **Country:** Le nom du pays.
- **ISO:** Le code du pays pour la norme ISO 3166-1 alpha-3.
- **Temperature (1960) to Temperature (2022):** Colonnes représentant les températures annuelles de surface (en degrés Celsius) pour chaque année de 1960 à 2022.

```
[1252]: # Importation des bibliothèques
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

```
[1253]: # Extraction des données dans un dataframe
df = pd.read_csv("temperature annuelle de surface.csv")
df
```

```
[1253]:
```

	Country	Year	Temperature
0	China	2004	36.31
1	Italy	2009	35.11
2	Spain	2017	7.87
3	Canada	2003	-7.45
4	Japan	1963	19.44
..
495	Australia	2014	22.67
496	Spain	1980	32.28
497	Australia	2018	21.06
498	Canada	1976	12.47
499	Australia	1973	29.69

[500 rows x 3 columns]

Écrire le code pour afficher tous les pays. (2pts) Répondre sur Moodle à la Question #1: Quel est le pays à l'indice 3 ?

```
[1254]: # Question #1 (2 pts)
df['Country']
```

```
[1254]: 0      China
      1      Italy
      2      Spain
      3      Canada
      4      Japan
      ...
     495  Australia
     496      Spain
     497  Australia
     498      Canada
     499  Australia
      Name: Country, Length: 500, dtype: object
```

```
[1255]: # Changement des noms des colonnes et affichage de la première colonne
actuels = ['Country', 'Year', 'Temperature']
nouveaux = ['Pays', 'Année', 'Température (C)']

for i in range(len(actuels)):
    df = df.rename(columns={actuels[i]: nouveaux[i]})

df.head(1)
```

```
[1255]:    Pays  Année  Température (C)
0  China   2004           36.31
```

Écrire le code pour afficher les statistiques descriptives du dataframe. (1pt) Répondre sur Moodle à la Question #2: Quelle est la moyenne des températures pour tous le pays ?

```
[1256]: # Question #2 (1pt):
df.describe()
```

```
[1256]:
```

	Année	Température (C)
count	500.000000	500.000000
mean	1991.374000	14.562280
std	18.044871	14.336185
min	1960.000000	-9.930000
25%	1976.000000	2.850000
50%	1991.500000	14.640000
75%	2007.000000	26.705000
max	2022.000000	39.970000

1. Écrire le code pour obtenir la température moyenne de chaque pays. (20pts) Attention, vous devez obligatoirement utiliser des dataframes, listes et boucles for. Voici l'algorithme du code: L'affichage doit respecter le format suivant:

- United Kingdom: La température moyenne est de valeur°C

- Canada: La température moyenne est de valeur°C
- Germany: La température moyenne est de valeur°C
- Spain: La température moyenne est de valeur°C
- Australia: La température moyenne est de valeur°C
- France: La température moyenne est de valeur°C
- Japan: La température moyenne est de valeur°C
- Italy: La température moyenne est de valeur°C
- United States: La température moyenne est de valeur°C
- China: La température moyenne est de valeur°C

Répondre sur Moodle à la Question #3: Quelle est la température moyenne de chaque pays ?

```
[1257]: # Question #3 (20pts)

# Créer une liste des pays uniques à partir de la sélection de la colonne
↳ 'Pays' du dataframe (3pts)
liste_pays = set(df['Pays'].values)

# Dataframe contenant les pays et leurs températures (4pts)
df_pays_temperatures = df[['Pays', 'Température (C)']].values.tolist()

# Parcourir la liste des pays uniques un pays à la fois
# Pour chacun des pays, obtenir la somme des températures et calculer la
↳ moyenne (2pts)
for pays in liste_pays: #2pts
    # Initialiser la somme des températures d'un pays et le nombre de
↳ températures (2pts)
    somme = 0
    nb_temps = 0
    # Pour chacun des pays du dataframe pays-températures (2pts)
    for ligne in df_pays_temperatures:
        # Si les pays sont les mêmes, on calcul la somme et la moyenne des
↳ températures du pays (2pts)
        if ligne[0] == pays:
            somme += ligne[1]
            nb_temps += 1
            # Calculer la température moyenne pour le pays courant (2pts)
            if nb_temps > 0:
                temp_moyenne = somme / nb_temps
            else:
                temp_moyenne = 0

    # Afficher la température moyenne du pays (3pts)
    print(f"{pays}: La température moyenne est de {round(temp_moyenne,2)}°C")
```

United Kingdom: La température moyenne est de 15.59°C

Canada: La température moyenne est de 16.13°C

Germany: La température moyenne est de 16.0°C

Spain: La température moyenne est de 16.1°C
 Australia: La température moyenne est de 13.67°C
 France: La température moyenne est de 12.45°C
 Japan: La température moyenne est de 12.33°C
 Italy: La température moyenne est de 14.16°C
 United States: La température moyenne est de 13.22°C
 China: La température moyenne est de 15.39°C

Écrire le code permettant d'obtenir le nombre de données manquantes pour la colonne 'Température'. (2pts) Répondre sur Moodle à la Question #4: Combien de températures maquent-il ?

[1258]: *# Question #4 (2pts)*

```
df.isna().sum()
```

[1258]: Pays 0
 Année 0
 Température (C) 0
 dtype: int64

Écrire le code permettant de répondre à la question #5 ci-dessous (6pts). Pour ce faire: - Sélectionner les données pour le Canada. - Enregistrez les données dans un dataframe nommé df_canada. - À l'aide d'une boucle FOR, parcourir les données afin d'identifier la plus basse température. - Filtrer la ligne qui contient la température la plus basse Répondre sur Moodle à la Question #5: En quelle année se trouve la température la plus basse au Canada ?

[1259]: *# Question 5 (6pts)*

```
df_canada = df[df['Pays'] == 'Canada'] #1pt
min = 0
for t in df_canada['Température (C)']: #2pts
    if t < min:
        annee = df_canada[df_canada['Température (C)'] == t] #2pts

print(f"L'année où il a fait le plus froid au Canada est {annee['Année']}")  ␣
↪ #1pt
```

L'année où il a fait le plus froid au Canada est 397 1996
 Name: Année, dtype: int64

Écrire le code pour calculer la température moyenne au Canada de 2010 à 2020 ? (4pts) Algorithme:
 1. Filtrer les données pour le Canada entre 2010 et 2020 2. Calculez la température moyenne pour cette période. Répondre sur Moodle à la Question #6: Quelle est la température moyenne au Canada de 2010 à 2020

[1260]: *# Question #6 (4pts)*

```
# Filtrer les données pour le Canada entre 2012 et 2022
df_10_ans = df_canada[(df_canada['Année'] >= 2010) & (df_canada['Année'] <= ␣
↪ 2020)]
```

```

somme_temp = 0
cpt = 0

for temp in df_10_ans["Température (C)"]:
    somme_temp += temp
    cpt += 1

temp_moyenne = somme_temp / cpt

# Afficher la température moyenne
print(f"La température moyenne au Canada de 2010 à 2020 est de {round(temp_moyenne,2)}C")

```

La température moyenne au Canada de 2010 à 2020 est de 20.59C

Ecrire le code pour créer un graphique en secteur permettant de visualiser la distribution des températures au Canada de 2010 à 2020. Sauvegardez le graphique en format PNG

Éléments du graphique

1. **Titre du graphique** : Proportion des températures annuelles au Canada (2010-2020)
2. **Légende** : Années
3. **Sauvegarde** : Le graphique est sauvegardé en format PNG sous le nom `graphique_secteur_températures_annuelles_canada_2010_2020.png`.

Rémettre l'image du graphique sur Moodle à la Question #7 (5pts)

```

[1261]: # Question #7a (5pts)
df_10_ans = df_canada[(df_canada['Année'] >= 2010) & (df_canada['Année'] <=
↳2020)]

# Vérifier et corriger les valeurs négatives ou nulles
df_5_ans = df_5_ans[df_5_ans['Température (C)'] > 0]

# Créer le graphique en secteur
plt.figure(figsize=(10, 7))

secteurs, etiquettes, autotexts = plt.pie(df_5_ans['Température (C)'],
↳autopct='%1.1f%%', startangle=140)

# Ajouter un titre
plt.title('Proportion des températures annuelles au Canada (2010-2020)')

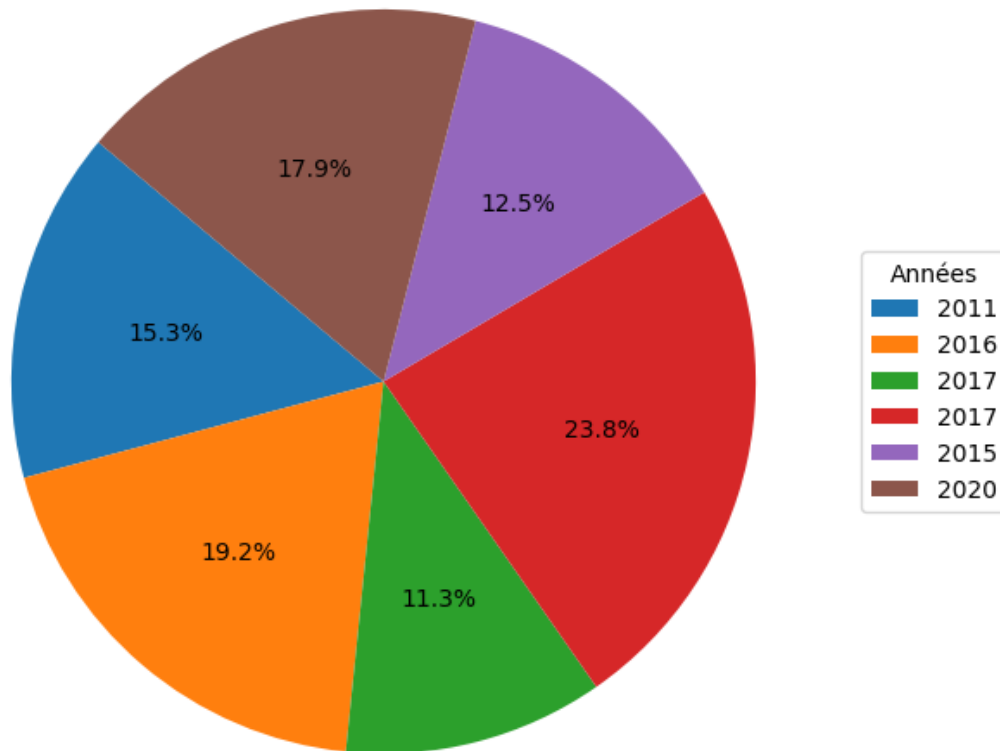
# Ajouter une légende
plt.legend(secteurs, df_5_ans['Année'], title="Années", loc="center left",
↳bbox_to_anchor=(1, 0, 0.5, 1))

# Sauvegarder le graphique en format PNG
plt.savefig('graphique_secteur_températures_annuelles_canada_2010_2020.png')

```

```
# Afficher le graphique  
plt.show()
```

Proportion des températures annuelles au Canada (2010-2020)



Ecrire le code pour créer un histogramme permettant de visualiser la distribution des températures annuelles au Canada entre 2010 et 2020 Sauvegardez le graphique en format PNG

Éléments du graphique

1. **Titre du graphique** : Distribution des températures annuelles au Canada (2010-2020)
2. **Axe des X (horizontal)** : Année de 2010 à 2020
3. **Axe des Y (vertical)** : Température (°C)
 - Couleur : Bleu
4. **Sauvegarde** : Le graphique est sauvegardé en format PNG sous le nom `histogramme_températures_annuelles_canada_2010_2020.png`.

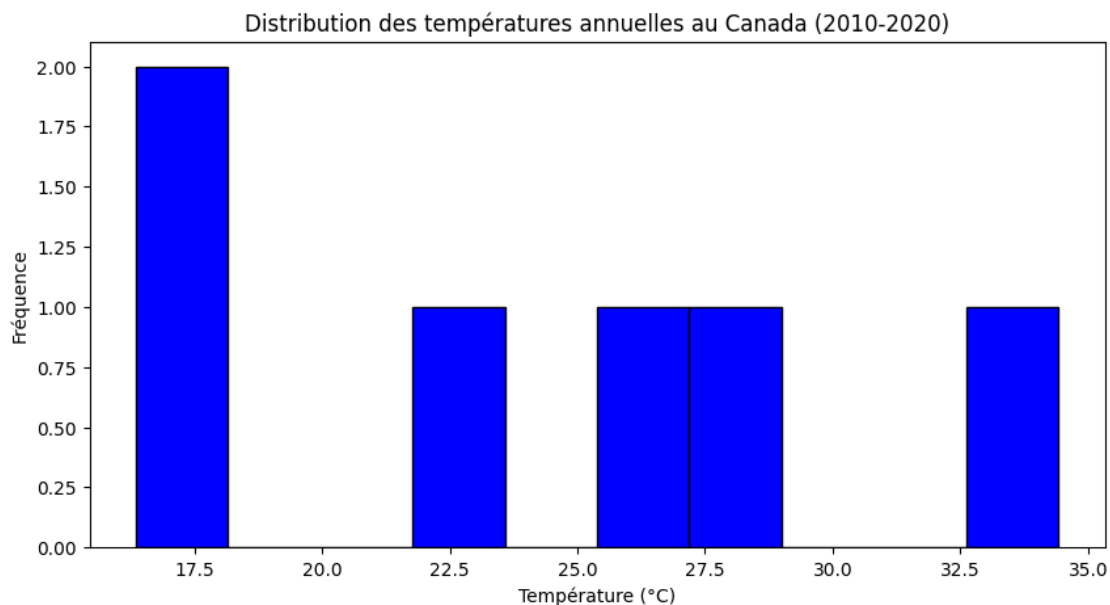
Rémettre l'image du graphique sur Moodle à la Question #7a (5pts)

```
[1262]: # Question #7a (5pts)
# Créer l'histogramme
plt.figure(figsize=(10, 5))
plt.hist(df_5_ans['Température (C)'], bins=10, color='b', edgecolor='black')

# Ajouter des titres et des labels
plt.title('Distribution des températures annuelles au Canada (2010-2020)')
plt.xlabel('Température (°C)')
plt.ylabel('Fréquence')

# Sauvegarder le graphique en format PNG
plt.savefig('histogramme_températures_annuelles_canada_2010_2020.png')

# Afficher le graphique
plt.show()
```



À l'aide de NumPy: Q8. Calculez la température moyenne au Canada Q9. Calculez la médiane des températures Q10. Calculez l'écart-type des températures Q11. Calculez la température minimum Q12. Calculez la température maximum

Répondre sur Moodle aux Questions #8 à #12 (5x1pt):

```
[1263]: # Questions #8 à #12(5pts)
# Convertir la colonne des températures en array NumPy
temperatures = np.array(df_canada['Température (C)'])

# Calcul de la moyenne
moyenne = np.mean(temperatures)
```

```

print(f"Température moyenne: {round(moyenne,2)}°C")

# Calcul de la médiane
mediane = np.median(temperatures)
print(f"Température médiane: {round(mediane,2)}°C")

# Calcul de l'écart-type
ecart_type = np.std(temperatures)
print(f"Écart-type des températures: {round(ecart_type,2)}°C")

# Calcul des valeurs minimales et maximales
min_temp = np.min(temperatures)
max_temp = np.max(temperatures)
print(f"Température minimale: {round(min_temp,2)}°C")
print(f"Température maximale: {round(max_temp,2)}°C")

```

Température moyenne: 16.13°C
 Température médiane: 17.89°C
 Écart-type des températures: 13.69°C
 Température minimale: -9.76°C
 Température maximale: 39.83°C

À l'aide de NumPy: - Calculer les coefficients de la régression linéaire entre les années et les températures - Créer le graphique nuage de points des températures en fonction des années et sauvegardez-le sous le nom de 'regression.png' - Ajouter la droite de régression linéaire entre les années et les températures Répondre sur Moodle aux Questions #13 à #15 (1pt, 2pts et 2pts): 13. Quels sont les valeurs des coefficients de la régression ? 14. Déposez le graphique sur Moodle pour la question 14. 15. En observant un graphique avec la droite de régression linéaire, quelles déductions faites vous au niveau de: 1. La tendance générale 2. La force de la relation 3. Prédictions possibles

```

[1264]: # Questions #13 à #15 regression 5pts
# Calculer les coefficients de la régression linéaire
pente, b = np.polyfit(annees, temperatures, 1)
print(pente,b)

# Tracer les données et la droite de régression
plt.figure(figsize=(10, 5))
plt.scatter(annees, temperatures, color='blue', label='Données')
plt.plot(annees, pente * anneess + b, color='red', label='Régression linéaire')

# Ajouter des titres et des labels
plt.title('Régression linéaire des températures annuelles au Canada')
plt.xlabel('Année')
plt.ylabel('Température (°C)')
plt.legend()

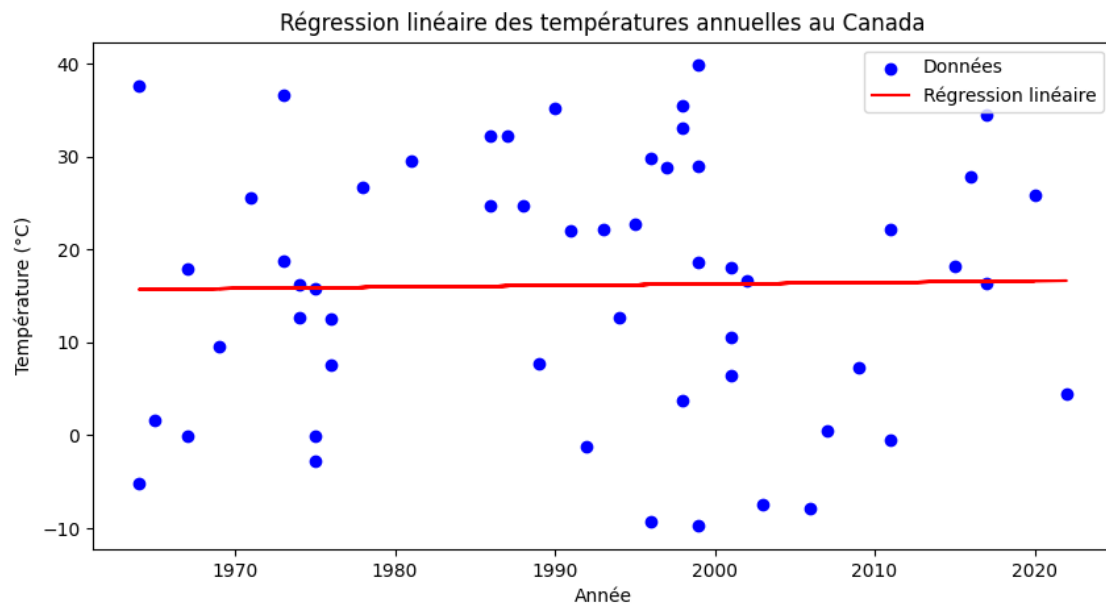
# Sauvegarder le graphique en format PNG
plt.savefig('regression.png')

```



```
# Afficher le graphique  
plt.show()
```

0.01620103539997822 -16.130447331236613



Q13. Quels sont les valeurs des coefficients de la régression ? pente = 0.01620103539997822
ordonnée à l'origine = -16.130447331236613

La pente de la droite de régression indique la tendance générale des données. Si la pente est positive, cela signifie qu'il y a une tendance à la hausse des températures au fil des années. Si la pente est négative, cela indique une tendance à la baisse.

Q14. Force de la relation

La proximité des points de données par rapport à la droite de régression montre la force de la relation linéaire. Si les points sont proches de la droite, cela indique une forte corrélation linéaire. Si les points sont dispersés, la corrélation est plus faible.

Q15. Prédictions

La droite de régression peut être utilisée pour faire des prédictions sur les températures futures en extrapolant les données. Cependant, il est important de noter que les prédictions basées sur une régression linéaire simple peuvent ne pas être précises si les conditions changent ou si la relation n'est pas strictement linéaire.