TP2_Corrige

February 19, 2025

##

Polytechnique Montréal Département Génie Informatique et Génie Logiciel INF8008 – Prétraitement de données . TP2 - Transformation, distribution et statistiques descriptives Hiver 2025 . Janvier 2025

0.1 Introduction

Le TD2 porte sur la transformation, la distribution et les statistiques descriptives. **Nous survolons l'utilisation de fonctions de base de Pandas et de l'analyse de données numériques.** Les données du fichier *Alzheimer_s_Disease_and_Healthy_Aging_Data.csv* sont des données publiques provenant d'enquêtes sur le vieillissement et la santé, faites par le Département de la Santé et des Services sociaux des États-Unis. Contrairement aux données du TP1 qui avaient été traitées au préalable, celles utilisées pour ce TP ne le sont pas. Vous devrez traiter les données brutes pour obtenir une version plus condensée, facilitant l'analyse des tendances et des sous-groupes de population.

Les champs principaux du fichier de données **Alzheimer_s_Disease_and_Healthy_Aging_Data.csv** sont les suivants :

- **YearStart/YearEnd** : années de début et de fin des données
- **LocationAbbr** : abréviation du lieu
- **Class** : catégorie des données (ex. : Santé mentale)
- **Topic** : sujet spécifique (ex. : détresse mentale fréquente)
- **Question** : question étudiée
- **Data_Value_Unit** : unité de mesure des données (ex. : pourcentage)
- **Data_Value** : valeur des données collectées
- **StratificationCategory1 / Stratification1** : catégorie et détail de la première stratification (ex. : âge, genre)
- **StratificationCategory2 / Stratification2** : catégorie et détail de la deuxième stratification (ex. : race, ethnie)

Ces données servent de base pour explorer les tendances, identifier des corrélations, et mieux comprendre les facteurs liés aux maladies neurodégénératives et à la santé mentale des populations vieillissantes. Votre objectif dans ce TP sera de préparer ces données pour qu'elles soient prêtes pour une analyse approfondie.

Voici les librairies python qui sera à utiliser pour ce TP : - pandas - numpy - matplotlib

À noter qu'au niveau de chaque question, il est recommandé de copier le DataFrame obtenu à la question précédente dans un nouveau DataFrame.

Veuillez vous référer à l'énoncé PDF de ce TP pour voir la sortie attendue.

[6]:
```
import pandas as pd
import numpy as np
```

[7]:
```
df = pd.read_csv('/content/drive/MyDrive/INF8008/Hiver 2025/
  ↪Alzheimer_s_Disease_and_Healthy_Aging_Data.csv')
df
```

[7]:

	RowId	YearStart	YearEnd	\
0	BRFSS~2022~2022~42~Q03~TMC01~AGE~RACE	2022	2022	
1	BRFSS~2022~2022~46~Q03~TMC01~AGE~RACE	2022	2022	
2	BRFSS~2022~2022~16~Q03~TMC01~AGE~RACE	2022	2022	
3	BRFSS~2022~2022~24~Q03~TMC01~AGE~RACE	2022	2022	
4	BRFSS~2022~2022~55~Q03~TMC01~AGE~GENDER	2022	2022	
…	…	…	…	
284137	BRFSS~2016~2016~55~Q15~TSC02~AGE~RACE	2016	2016	
284138	BRFSS~2017~2017~56~Q45~TOC13~AGE~RACE	2017	2017	
284139	BRFSS~2015~2015~56~Q42~TCC04~AGE~RACE	2015	2015	
284140	BRFSS~2019~2019~54~Q46~TOC10~AGE~RACE	2019	2019	
284141	BRFSS~2015~2015~56~Q02~TNC02~AGE~RACE	2015	2015	

	LocationAbbr	LocationDesc	Datasource	\
0	PA	Pennsylvania	BRFSS	
1	SD	South Dakota	BRFSS	
2	ID	Idaho	BRFSS	
3	MD	Maryland	BRFSS	
4	WI	Wisconsin	BRFSS	
…	…	…	…	
284137	WI	Wisconsin	BRFSS	
284138	WY	Wyoming	BRFSS	
284139	WY	Wyoming	BRFSS	
284140	WV	West Virginia	BRFSS	
284141	WY	Wyoming	BRFSS	

	Class	\
0	Mental Health	
1	Mental Health	
2	Mental Health	
3	Mental Health	
4	Mental Health	
…	…	
284137	Screenings and Vaccines	
284138	Overall Health	
284139	Cognitive Decline	
284140	Overall Health	
284141	Nutrition/Physical Activity/Obesity	

```
                                                  Topic  \
0                                Frequent mental distress
1                                Frequent mental distress
2                                Frequent mental distress
3                                Frequent mental distress
4                                Frequent mental distress
…                                                      …
284137                        Colorectal cancer screening
284138  Fair or poor health among older adults with ar…
284139  Talked with health care professional about sub…
284140  Disability status, including sensory or mobili…
284141                  Eating 3 or more vegetables daily

                                                 Question Data_Value_Unit  \
0       Percentage of older adults who are experiencin…                 %
1       Percentage of older adults who are experiencin…                 %
2       Percentage of older adults who are experiencin…                 %
3       Percentage of older adults who are experiencin…                 %
4       Percentage of older adults who are experiencin…                 %
…                                                     …              …
284137  Percentage of older adults who had either a ho…                 %
284138  Fair or poor health among older adults with do…                 %
284139  Percentage of older adults with subjective cog…                 %
284140  Percentage of older adults who report having a…                 %
284141  Percentage of older adults who are eating 3 or…                 %

        …           Stratification2                      Geolocation  \
0       …  Native Am/Alaskan Native  POINT (-77.86070029 40.79373015)
1       …    Asian/Pacific Islander  POINT (-100.3735306 44.35313005)
2       …       Black, non-Hispanic    POINT (-114.36373 43.68263001)
3       …       Black, non-Hispanic  POINT (-76.60926011 39.29058096)
4       …                      Male  POINT (-89.81637074 44.39319117)
…     …                         …                                 …
284137  …       Black, non-Hispanic  POINT (-89.81637074 44.39319117)
284138  …                  Hispanic  POINT (-108.1098304 43.23554134)
284139  …    Asian/Pacific Islander  POINT (-108.1098304 43.23554134)
284140  …                  Hispanic   POINT (-80.71264013 38.6655102)
284141  …  Native Am/Alaskan Native  POINT (-108.1098304 43.23554134)

        ClassID  TopicID QuestionID LocationID  StratificationCategoryID1  \
0           C05    TMC01        Q03         42                        AGE
1           C05    TMC01        Q03         46                        AGE
2           C05    TMC01        Q03         16                        AGE
3           C05    TMC01        Q03         24                        AGE
4           C05    TMC01        Q03         55                        AGE
…         …      …        …         …                         …
```

```
284137      C03    TSC02        Q15         55                        AGE
284138      C01    TOC13        Q45         56                        AGE
284139      C06    TCC04        Q42         56                        AGE
284140      C01    TOC10        Q46         54                        AGE
284141      C02    TNC02        Q02         56                        AGE

        StratificationID1 StratificationCategoryID2 StratificationID2
0                    5064                      RACE              NAA
1                  65PLUS                      RACE              ASN
2                  65PLUS                      RACE              BLK
3                  65PLUS                      RACE              BLK
4                  65PLUS                    GENDER             MALE
…                     …                         …                 …
284137        AGE_OVERALL                      RACE              BLK
284138               5064                      RACE              HIS
284139        AGE_OVERALL                      RACE              ASN
284140             65PLUS                      RACE              HIS
284141               5064                      RACE              NAA

[284142 rows x 31 columns]
```

0.1.1 A)

Vous remarquerez que ce jeu de données est assez large, avec 284142 lignes et 31 colonnes.

Avec des ensembles de données de cette taille, on peut souvent trouver des défauts, comme des doublons de lignes. Vérifiez donc s'il existe des valeurs en double dans le DataFrame. (2 points)

[8]:
```
#TODO:
duplicate_rows = df.duplicated()

if True in duplicate_rows.values:
    print("Il existe des valeurs en double.")
else:
    print("Il n'existe pas de valeurs en double.")
```

```
Il n'existe pas de valeurs en double.
```

0.1.2 B)

Il est possible d'extraire la durée du sondage en soustrayant l'année de début de l'année de fin. Utilisez lambda, ainsi que cette soustraction, pour garder les lignes avec une durée de sondage de moins d'1 an. (3 points)

[9]:
```
#TODO:
df['Year_Duration'] = df.apply(lambda row: row['YearEnd'] - row['YearStart'],␣
 ↪axis=1)
df = df[df["Year_Duration"] < 1]
df
```

[9]:

	RowId	YearStart	YearEnd	\
0	BRFSS~2022~2022~42~Q03~TMC01~AGE~RACE	2022	2022	
1	BRFSS~2022~2022~46~Q03~TMC01~AGE~RACE	2022	2022	
2	BRFSS~2022~2022~16~Q03~TMC01~AGE~RACE	2022	2022	
3	BRFSS~2022~2022~24~Q03~TMC01~AGE~RACE	2022	2022	
4	BRFSS~2022~2022~55~Q03~TMC01~AGE~GENDER	2022	2022	
…	…	…	…	
284137	BRFSS~2016~2016~55~Q15~TSC02~AGE~RACE	2016	2016	
284138	BRFSS~2017~2017~56~Q45~TOC13~AGE~RACE	2017	2017	
284139	BRFSS~2015~2015~56~Q42~TCC04~AGE~RACE	2015	2015	
284140	BRFSS~2019~2019~54~Q46~TOC10~AGE~RACE	2019	2019	
284141	BRFSS~2015~2015~56~Q02~TNC02~AGE~RACE	2015	2015	

	LocationAbbr	LocationDesc	Datasource	\
0	PA	Pennsylvania	BRFSS	
1	SD	South Dakota	BRFSS	
2	ID	Idaho	BRFSS	
3	MD	Maryland	BRFSS	
4	WI	Wisconsin	BRFSS	
…	…	…	…	
284137	WI	Wisconsin	BRFSS	
284138	WY	Wyoming	BRFSS	
284139	WY	Wyoming	BRFSS	
284140	WV	West Virginia	BRFSS	
284141	WY	Wyoming	BRFSS	

	Class	\
0	Mental Health	
1	Mental Health	
2	Mental Health	
3	Mental Health	
4	Mental Health	
…	…	
284137	Screenings and Vaccines	
284138	Overall Health	
284139	Cognitive Decline	
284140	Overall Health	
284141	Nutrition/Physical Activity/Obesity	

	Topic	\
0	Frequent mental distress	
1	Frequent mental distress	
2	Frequent mental distress	
3	Frequent mental distress	
4	Frequent mental distress	
…	…	
284137	Colorectal cancer screening	

```
284138  Fair or poor health among older adults with ar…
284139  Talked with health care professional about sub…
284140  Disability status, including sensory or mobili…
284141                  Eating 3 or more vegetables daily

                                                 Question Data_Value_Unit  \
0       Percentage of older adults who are experiencin…                %
1       Percentage of older adults who are experiencin…                %
2       Percentage of older adults who are experiencin…                %
3       Percentage of older adults who are experiencin…                %
4       Percentage of older adults who are experiencin…                %
…                                                    …              …
284137  Percentage of older adults who had either a ho…                %
284138  Fair or poor health among older adults with do…                %
284139  Percentage of older adults with subjective cog…                %
284140  Percentage of older adults who report having a…                %
284141  Percentage of older adults who are eating 3 or…                %

        …                       Geolocation ClassID  TopicID  QuestionID  \
0       …  POINT (-77.86070029 40.79373015)     C05    TMC01         Q03
1       …  POINT (-100.3735306 44.35313005)     C05    TMC01         Q03
2       …    POINT (-114.36373 43.68263001)     C05    TMC01         Q03
3       …  POINT (-76.60926011 39.29058096)     C05    TMC01         Q03
4       …  POINT (-89.81637074 44.39319117)     C05    TMC01         Q03
…     …                                   …     …      …         …
284137  …  POINT (-89.81637074 44.39319117)     C03    TSC02         Q15
284138  …  POINT (-108.1098304 43.23554134)     C01    TOC13         Q45
284139  …  POINT (-108.1098304 43.23554134)     C06    TCC04         Q42
284140  …   POINT (-80.71264013 38.6655102)     C01    TOC10         Q46
284141  …  POINT (-108.1098304 43.23554134)     C02    TNC02         Q02

       LocationID StratificationCategoryID1  StratificationID1  \
0              42                       AGE               5064
1              46                       AGE             65PLUS
2              16                       AGE             65PLUS
3              24                       AGE             65PLUS
4              55                       AGE             65PLUS
…             …                         …                 …
284137         55                       AGE        AGE_OVERALL
284138         56                       AGE               5064
284139         56                       AGE        AGE_OVERALL
284140         54                       AGE             65PLUS
284141         56                       AGE               5064

        StratificationCategoryID2 StratificationID2 Year_Duration
0                            RACE               NAA             0
1                            RACE               ASN             0
```

```
2                          RACE              BLK             0
3                          RACE              BLK             0
4                        GENDER             MALE             0
…                           …                 …              …
284137                     RACE              BLK             0
284138                     RACE              HIS             0
284139                     RACE              ASN             0
284140                     RACE              HIS             0
284141                     RACE              NAA             0

[274881 rows x 32 columns]
```

0.1.3 C)

Maintenant que cette étape est faite, les colonnes YearStart et YearEnd contiennent la même information. Renommez une des deux colonnes à `Year`, et supprimez l'autre. (2 points)

[10]:
```
df = df.rename({"YearStart": "Year"}, axis=1)
df=df.drop(["YearEnd"], axis = 1)
```

0.1.4 D)

Certaines colonnes contiennent des données redondantes ou inutiles pour notre analyse. Éliminez toutes les colonnes inutiles en ne conservant que celles mentionnées dans l'introduction. Combien de colonnes reste-t-il ? (2 points)

[11]:
```
#TODO:

df = df.loc[:, ["Year", "LocationAbbr", "Class", "Topic", "Question",␣
 ↪"Data_Value_Unit", "Data_Value", "Stratification1", "Stratification2",␣
 ↪"StratificationCategory1", "StratificationCategory2"]]
print(" Il existe" , df.columns.shape[0], "colonnes.")
```

```
 Il existe 11 colonnes.
```

0.1.5 E)

Comme vu dans le module 1, le prétraitement des données consiste à gérer les défauts des données collectées, comme les valeurs nulles. La colonne `Data_Value` est importante pour notre analyse.

Vérifiez donc s'il existe des données manquantes dans la colonne `Data_Value`. Quel est le pourcentage de valeurs manquantes ? (3 points)

[12]:
```
cc = df['Data_Value'].count() == len(df)

if not cc:  # cc will be False if there are missing values
    print("Il existe des données manquantes.")
    print("Le pourcentage de données manquantes est de", (1 - df['Data_Value'].
 ↪count() / len(df)) * 100, "%")
```

```
else:
    print("Il n'existe pas de données manquantes.")
```

```
Il existe des données manquantes.
Le pourcentage de données manquantes est de 32.117898290533 %
```

0.1.6 F)

Deux façons de traiter les données manquantes: les remplacer par la valeur médiane ou les éliminer complètement.

Il n'existe pas de solution **unique ou meilleure**. Tout dépend de l'analyse effectuée. Il est essentiel d'examiner les effets de chacun de ces choix sur l'analyse ultérieure. C'est pourquoi, dans ce TP, nous essayerons les deux méthodes.

Vous devez donc:

1. Créez deux copies de l'ensemble de données.
2. Supprimez les valeurs manquantes d'une des copies.
3. Remplacez les valeurs manquantes d'une autre copie par la médiane.

Affichez les nouveaux dataframes. Vous devriez avoir autour de 186595 lignes pour l'un et 274881 lignes pour l'autre. (4 points)

[13]:
```
df1 = df.copy()
df2 = df.copy()
df1["Data_Value"] = df1["Data_Value"].fillna(df["Data_Value"].median())
df2 = df2.dropna(subset=["Data_Value"])
print(df1)
print(df2)
```

```
        Year LocationAbbr                                Class  \
0       2022           PA                       Mental Health
1       2022           SD                       Mental Health
2       2022           ID                       Mental Health
3       2022           MD                       Mental Health
4       2022           WI                       Mental Health
…        …             …                                  …
284137  2016           WI            Screenings and Vaccines
284138  2017           WY                      Overall Health
284139  2015           WY                  Cognitive Decline
284140  2019           WV                      Overall Health
284141  2015           WY  Nutrition/Physical Activity/Obesity

                                          Topic  \
0                      Frequent mental distress
1                      Frequent mental distress
2                      Frequent mental distress
3                      Frequent mental distress
4                      Frequent mental distress
…                                          …
```

```
284137                        Colorectal cancer screening
284138  Fair or poor health among older adults with ar…
284139  Talked with health care professional about sub…
284140  Disability status, including sensory or mobili…
284141                  Eating 3 or more vegetables daily

                                                 Question Data_Value_Unit  \
0       Percentage of older adults who are experiencin…                %
1       Percentage of older adults who are experiencin…                %
2       Percentage of older adults who are experiencin…                %
3       Percentage of older adults who are experiencin…                %
4       Percentage of older adults who are experiencin…                %
…                                                    …              …
284137  Percentage of older adults who had either a ho…                %
284138  Fair or poor health among older adults with do…                %
284139  Percentage of older adults with subjective cog…                %
284140  Percentage of older adults who report having a…                %
284141  Percentage of older adults who are eating 3 or…                %

        Data_Value    Stratification1           Stratification2  \
0             33.0        50-64 years  Native Am/Alaskan Native
1             33.0  65 years or older    Asian/Pacific Islander
2             33.0  65 years or older       Black, non-Hispanic
3              9.0  65 years or older       Black, non-Hispanic
4              5.6  65 years or older                      Male
…              …                  …                         …
284137        70.7            Overall       Black, non-Hispanic
284138        33.0        50-64 years                  Hispanic
284139        33.0            Overall    Asian/Pacific Islander
284140        33.0  65 years or older                  Hispanic
284141        33.0        50-64 years  Native Am/Alaskan Native

       StratificationCategory1 StratificationCategory2
0                    Age Group          Race/Ethnicity
1                    Age Group          Race/Ethnicity
2                    Age Group          Race/Ethnicity
3                    Age Group          Race/Ethnicity
4                    Age Group                 Gender
…                        …                        …
284137               Age Group          Race/Ethnicity
284138               Age Group          Race/Ethnicity
284139               Age Group          Race/Ethnicity
284140               Age Group          Race/Ethnicity
284141               Age Group          Race/Ethnicity

[274881 rows x 11 columns]
        Year LocationAbbr                              Class  \
3       2022           MD                     Mental Health
```

```
4       2022           WI                          Mental Health
6       2022           OK                          Mental Health
7       2022           PA                          Mental Health
8       2022           AZ                         Overall Health
…       …           …                                      …
284131  2015           WI                      Cognitive Decline
284132  2020           US                              Caregiving
284134  2016           WV  Nutrition/Physical Activity/Obesity
284135  2017           WV                         Overall Health
284137  2016           WI               Screenings and Vaccines

                                                    Topic  \
3                                Frequent mental distress
4                                Frequent mental distress
6                                Frequent mental distress
7                                Frequent mental distress
8               Recent activity limitations in past month
…                                                      …
284131  Need assistance with day-to-day activities bec…
284132  Provide care for someone with cognitive impair…
284134                                           Obesity
284135                      Arthritis among older adults
284137                       Colorectal cancer screening

                                                  Question Data_Value_Unit  \
3       Percentage of older adults who are experiencin…                  %
4       Percentage of older adults who are experiencin…                  %
6       Percentage of older adults who are experiencin…                  %
7       Percentage of older adults who are experiencin…                  %
8       Mean number of days with activity limitations …           Number
…                                                      …                …
284131  Percentage of older adults who reported that a…                  %
284132  Percentage of older adults who provided care f…                  %
284134  Percentage of older adults who are currently o…                  %
284135  Percentage of older adults ever told they have…                  %
284137  Percentage of older adults who had either a ho…                  %

        Data_Value    Stratification1           Stratification2  \
3              9.0  65 years or older       Black, non-Hispanic
4              5.6  65 years or older                      Male
6             21.5            Overall  Native Am/Alaskan Native
7             10.0            Overall       White, non-Hispanic
8              6.1  65 years or older       White, non-Hispanic
…              …                  …                        …
284131        29.2        50-64 years                   Female
284132        28.5  65 years or older                 Hispanic
284134        44.5        50-64 years                   Female
284135        57.7            Overall                      NaN
```

```
284137        70.7            Overall       Black, non-Hispanic

       StratificationCategory1 StratificationCategory2
3                    Age Group          Race/Ethnicity
4                    Age Group                  Gender
6                    Age Group          Race/Ethnicity
7                    Age Group          Race/Ethnicity
8                    Age Group          Race/Ethnicity
…                          …                       …
284131               Age Group                  Gender
284132               Age Group          Race/Ethnicity
284134               Age Group                  Gender
284135               Age Group                     NaN
284137               Age Group          Race/Ethnicity

[186595 rows x 11 columns]
```

Note : Pour la suite du travail, chaque étape devra être réalisée sur les deux copies de l'ensemble de données.

0.1.7 G)

Plusieurs classes existent. On va évaluer la santé mentale "Mental Health". Filtrez les données de la colonne "class" pour la valeur "Mental Health", puis déterminez la moyenne de `Data_Value` par `Year` et `Topic`. (2 points)

[14]:
```
df1 = df1[df1["Class"] == "Mental Health"]
df2 = df2[df2["Class"] == "Mental Health"]
grouped_data1 = df1.groupby(['Topic', 'Year'])['Data_Value'].mean().
 ↪reset_index()
grouped_data2 = df2.groupby(['Topic', 'Year'])['Data_Value'].mean().
 ↪reset_index()
```

[14]:
```
```

0.1.8 H)

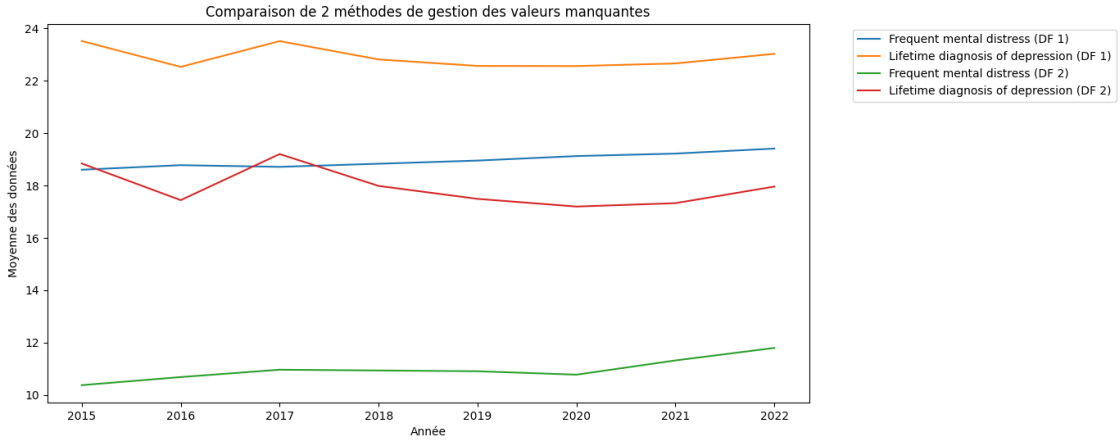
Il est temps de comparer la suppression des données manquantes vs leur remplacement par la médiane. Pour cela, affichez les valeurs moyennes de `Data_Value` par année, pour chaque groupe et chaque topic. (3 points)

[15]:
```
import matplotlib.pyplot as plt
```

[16]:
```
#DF 1: remplacement de NA avec médiane
plt.figure(figsize=(12, 6))
for topic in grouped_data1['Topic'].unique():
    subset = grouped_data1[grouped_data1['Topic'] == topic]
    plt.plot(subset['Year'], subset['Data_Value'], label=f'{topic} (DF 1)')
```

```python
# DF 2: élimination de NA
for topic in grouped_data2['Topic'].unique():
    subset = grouped_data2[grouped_data2['Topic'] == topic]
    plt.plot(subset['Year'], subset['Data_Value'], label=f'{topic} (DF 2)')

plt.title('Comparaison de 2 méthodes de gestion des valeurs manquantes')
plt.xlabel('Année')
plt.ylabel('Moyenne des données')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()
```



0.2 3. LIVRABLES

Vous devez remettre sur Moodle un fichier compressé .zip contenant :

1) Le code : Un Jupyter notebook en Python qui contient le code tel implanté avec les librairies minimales demandées pour ce TP (Python, Pandas, Matplotlib). Le code doit être exécutable sans erreur et accompagné des commentaires appropriés dans le notebook de manière. Tous vos résultats doivent être reproductibles avec le code dans le notebook. *Attention, en aucun cas votre code ne doit avoir été copié de d'ailleurs.*

2) Un fichier pdf représentant votre notebook complètement exécuté sous format pdf (obtenu via latex ou imprimé en pdf avec le navigateur). Assurez-vous que le PDF est entièrement lisible. Tutoriel youtube

ATTENTION: assurez-vous que votre fichier compressé .zip ne dépasse pas la taille limite acceptée sur Moodle.

ÉVALUATION Votre TP sera évalué sur les points suivants :

Critères : 1. Implantation correcte et efficace 2. Qualité du code (noms significatifs, structure, performance, gestion d'exception, etc.) (1 point) 3. Réponses correctes/sensées aux questions de réflexion ou d'analyse

CODE D'HONNEUR - Règle 1: Le plagiat de code est bien évidemment interdit. Toute utilisation de code doit être référencée adéquatement. Vous **ne pouvez pas** soumettre un code, écrit par quelqu'un d'autre. Dans le cas contraire, cela sera considéré comme du plagiat. - **Règle 2**: Vous êtes libres de discuter avec d'autres équipes. Cependant, vous ne pouvez en aucun cas incorporer leur code dans votre TP. - **Règle 3**: Vous ne pouvez pas partager votre code publiquement (par exemple, dans un dépôt GitHub public) tant que le cours n'est pas fini.

0.2.1 Conversion en PDF sur Google Colab

[ ]:
```
%%capture
!sudo apt-get install texlive-xetex texlive-fonts-recommended␣
↪texlive-plain-generic
```

Assurez vous d'avoir téléchargé le TP complété en notebook sur votre ordinateur, puis importé ce fichier dans le répertoire "content" avant de rouler la ligne suivante.

[ ]:
```
!jupyter nbconvert --to pdf /content/TP1.ipynb
```