

ANALYSE DES DONNEES



Formation préparée et présentée par
Najah DAOUES inspecteur
pédagogique d'informatique de
Tataouine et Médenine

ANALYSE DES DONNEES

Compétences à développer

- Acquérir des notions de base pour l'automatisation des traitements des données en vue de les analyser.
- Acquérir des notions de base pour l'analyse de données.
- Développer le raisonnement et la pensée logique pour résoudre des problèmes de gestion de données.
- Utiliser à bon escient les services d'internet pour créer et partager des documents en ligne utiles dans le processus de gestion de données.

Objectifs spécifiques :

- Assimiler l'intérêt de l'analyse des données.
- Découvrir les étapes d'analyse des données.
- Reconnaître différents outils et méthodes de collecte des données.
- Créer un questionnaire en ligne.
- Utiliser un environnement de programmation pour résoudre des problèmes relatifs à la gestion de données.



C'est quoi l'analyse des données ?

□ **Activité N° 1 : Découvrir l'analyse des données**

En groupe, présentez une phrase dans laquelle vous exprimez vos idées portant sur l'analyse de données.

En plénière, chaque groupe présente ses réponses.

L'intérêt de l'analyse des données

□ Activité N° 2 : Comprendre l'intérêt de l'analyse des données

Parmi les propositions ci-dessous, cochez les propositions qui présentent l'intérêt de l'analyse de données.

L'analyse des données permet d'examiner et d'interpréter des données afin de (/ d') :

<input type="checkbox"/>	Chercher des réponses à des questions.	<input type="checkbox"/>	Analyser des situations de la vie courante, d'une entreprise, etc.
<input type="checkbox"/>	Proposer des solutions à des problèmes.	<input type="checkbox"/>	Obtenir des renseignements sur les lacunes de dysfonctionnement d'un système.
<input type="checkbox"/>	Evaluer des politiques et des décisions.	<input type="checkbox"/>	Aider à la prise de décision.
<input type="checkbox"/>	Automatiser les données.	<input type="checkbox"/>	Stocker les données de façon structurée.
<input type="checkbox"/>	Augmenter la productivité des entreprises.	<input type="checkbox"/>	

Les étapes de l'analyse des données

□ **Activité N° 3 : Les étapes de l'analyse des données**

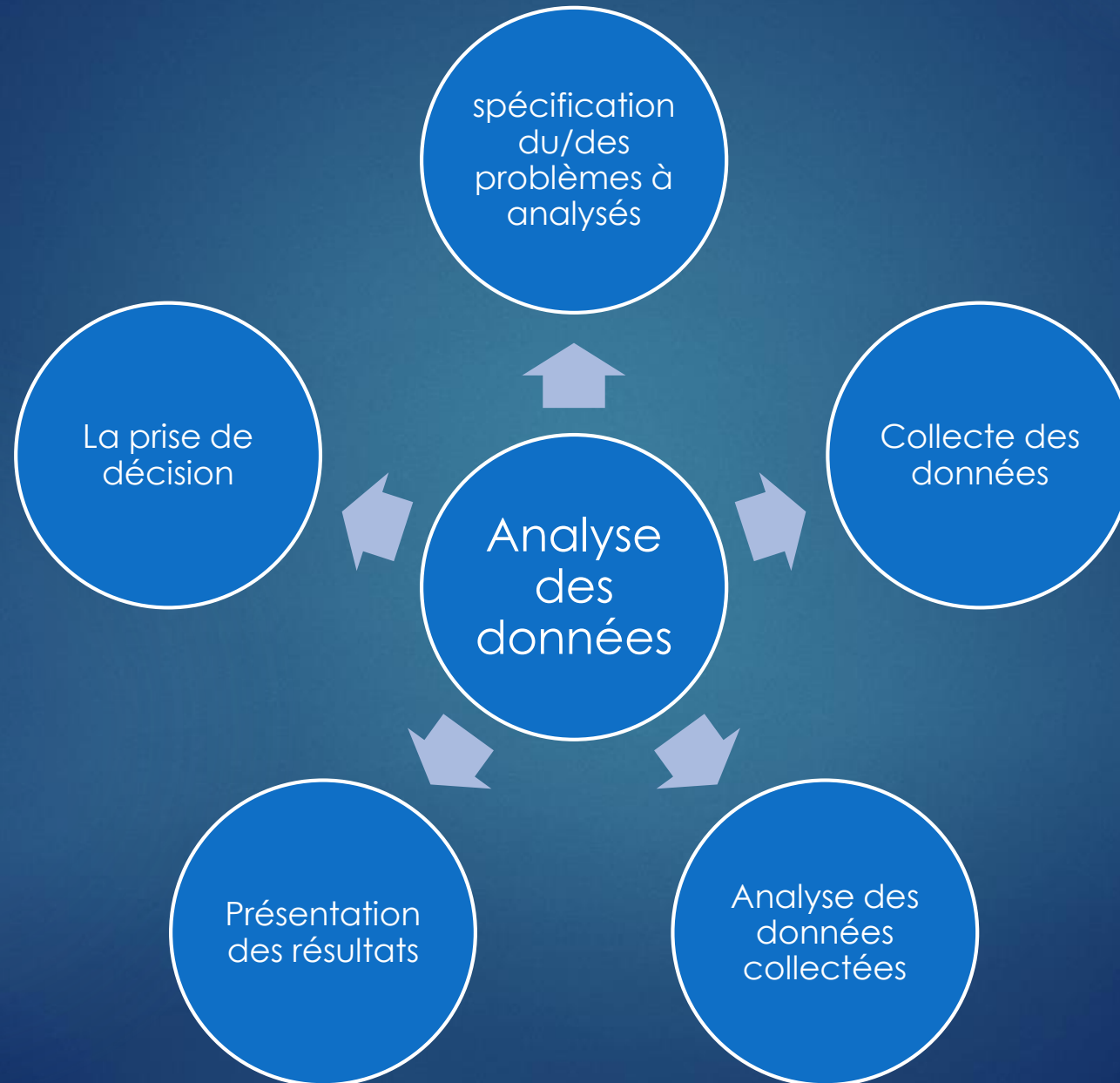
En groupe, visionnez la vidéo intitulée « Vidéo1.mp4 » située dans le dossier « Formation analyse des données » du lecteur C, puis dégagez les principales étapes de l'analyse des données.

En plénière, chaque groupe présente ses réponses.

Les étapes de l'analyse des données

unicef 

Les étapes de l'analyse des données



SPECIFICATION DU PROBLEME



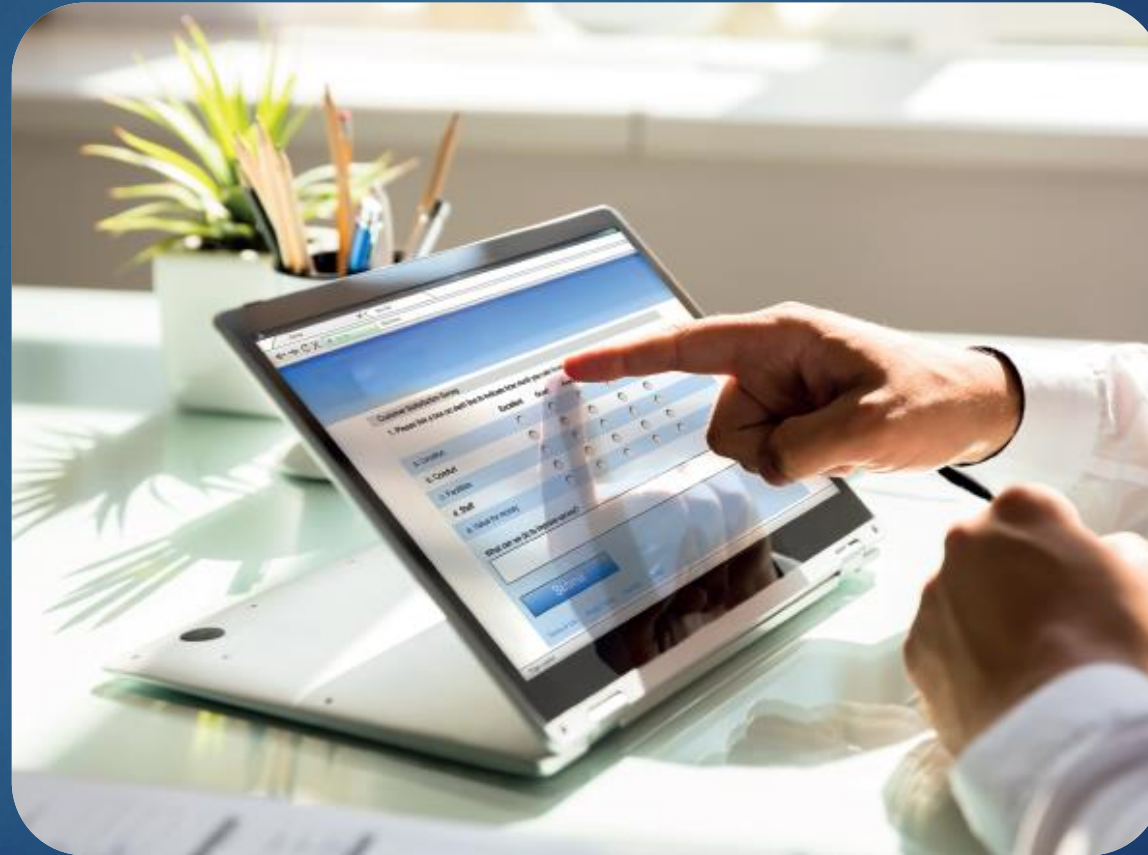
Problème à analysé

Texte :

Au lycée et pendant la récréation, vous avez remarqué que la majorité des élèves sont attachés à leurs smartphones et leurs téléphones. Vous avez remarqué que ce phénomène s'accroît de jour en jour et certains de vos camarades vous avouent qu'ils sont incapables de se séparer de leurs téléphones. Un bon nombre de vos camarades se lèvent même plusieurs fois la nuit pour consulter leurs comptes Facebook ou pour jouer leur jeu préféré. Ils déclarent aussi qu'ils dépensent la majorité de leur argent de poche sur les frais de connexion, de communication téléphonique et l'achat des accessoires et des nouveaux appareils.

Pendant la séance d'informatique, vous avez exposé ce phénomène à votre professeur, il vous a demandé d'analyser ce phénomène dans le milieu scolaire (Ecole primaire, collège et lycée) en utilisant une méthode de collecte des données, un outil d'analyse de données, de présenter les résultats et de proposer des recommandations.

COLLECTE DES DONNES



Méthodes de collecte des données

□ Activité N° 4 : Découverte de la méthode quantitative et la méthode qualitative

Pour comprendre la différence entre la méthode quantitative et la méthode qualitative, consultez l'adresse URL suivante « <https://fr.surveymonkey.com/mp/quantitative-vs-qualitative-research/> » puis classez chacun des mots ci-dessous dans la colonne de la méthode appropriée.

Méthodes de collecte des données

/ Entretien / Questionnaire / Études de cas / Formulaire / Groupes de discussion / Conférences / Vision globale / Étude de comportement / Aide à la prise de décision / Réponses à des questions ouvertes /
/ Réaliser sur un nombre restreint de personnes /
/ Analyse du contenu / N'aide pas à la prise de décision / Avis d'experts /
/ Généraliser les résultats de l'étude sur le reste de la population /

Méthode quantitative	Méthode qualitative
<ul style="list-style-type: none">• Questionnaire• Formulaire• Vision globale• Aide à la prise de décision• Généraliser les résultats de l'étude sur un groupe de personnes très important	<ul style="list-style-type: none">• Entretien / Étude de cas• Groupes de discussion• Études de comportement• Réponses à des questions ouvertes réalisées sur un nombre restreint de personnes• Analyse du contenu• N'aide pas à la prise de décision• Avis d'experts / Conférences

Les étapes de collecte des données

Création du questionnaire sur papier.

Création du formulaire en ligne.

Partage du formulaire.

Exportation des données collectées en formats xls ou csv.

La nature des questions dans un questionnaire

□ Activité N° 5 : Découverte de la nature des questions dans un questionnaire

Sur le Web et en binôme, consultez l'adresse URL suivante :

« <https://forms.gle/TtgMcAv8zsgvKw8v5> »

puis classez les numéros des questions selon les deux natures suivantes :

N° des questions ouvertes	N° des questions fermées
/ 1 / 8 / 11 (Autre) / 12 (Autre) / / 14 / 15 /	/ 2 / 3 / 4 / 5 / 6 / 7 / 9 / / 10 / 11 / 12 / 13 /

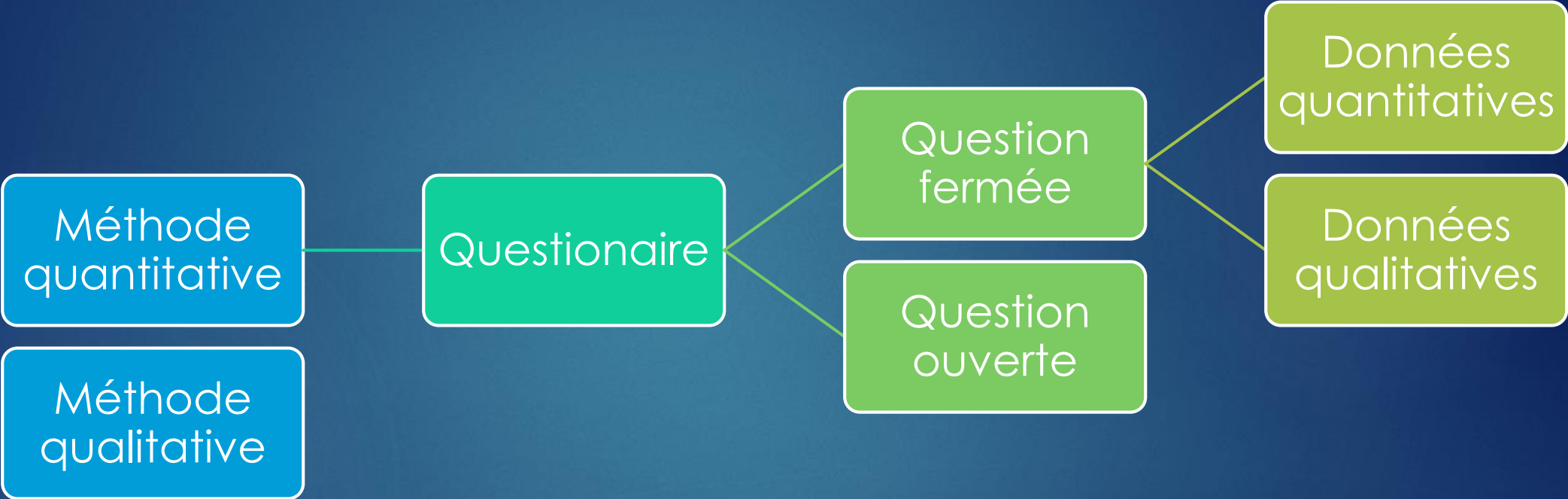
La nature des questions dans un questionnaire

□ Activité N° 8 : La nature des données (Données quantitatives / données qualitatives)

En groupe, et en vous basant sur les réponses correctes de l'activité précédente, classez les numéros des questions fermées selon leurs réponses possibles en des données quantitatives (Numérique) ou en des données qualitatives.

N° des questions fermées avec des données quantitatives	N° des questions fermées avec des données qualitatives
/ 2 / 9 / 10 /	/ 3 / 4 / 5 / 6 / 7 / 11 / 12 / 13 /

Conclusion



Création des questionnaires

□ **Activité N° 5 : Création d'un questionnaire**

Pour le sujet suivant, créez un questionnaire sur papier pour la collecte des données.

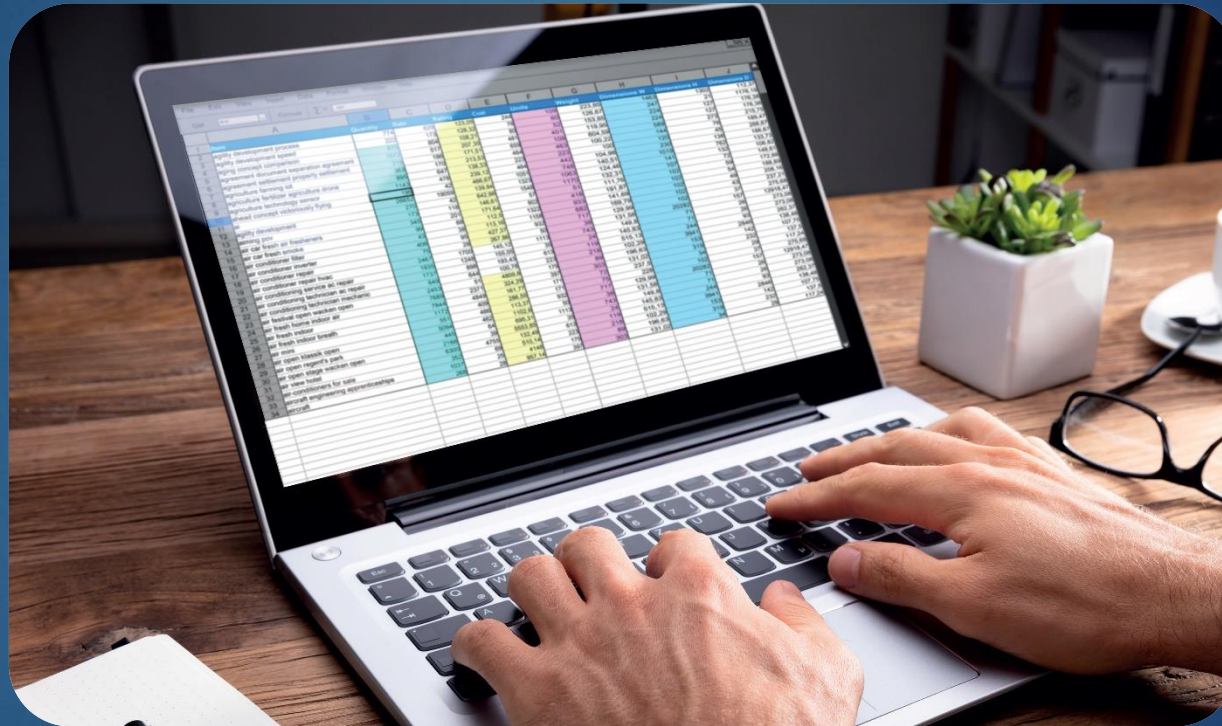
Sujet du questionnaire : L'impact de l'utilisation des smartphones sur la vie des élèves (**voir fiche projet**).

Création et partage du formulaire en ligne



Analyse de quelques questionnaires

L'ANALYSE DES DONNEES COLLECTEES



La qualité des données

- ✓ Validité.
- ✓ Fiabilité.
- ✓ Précision.
- ✓ Intégrité.
- ✓ Actualité.

Comment analyser les données sous python

Installation de
python

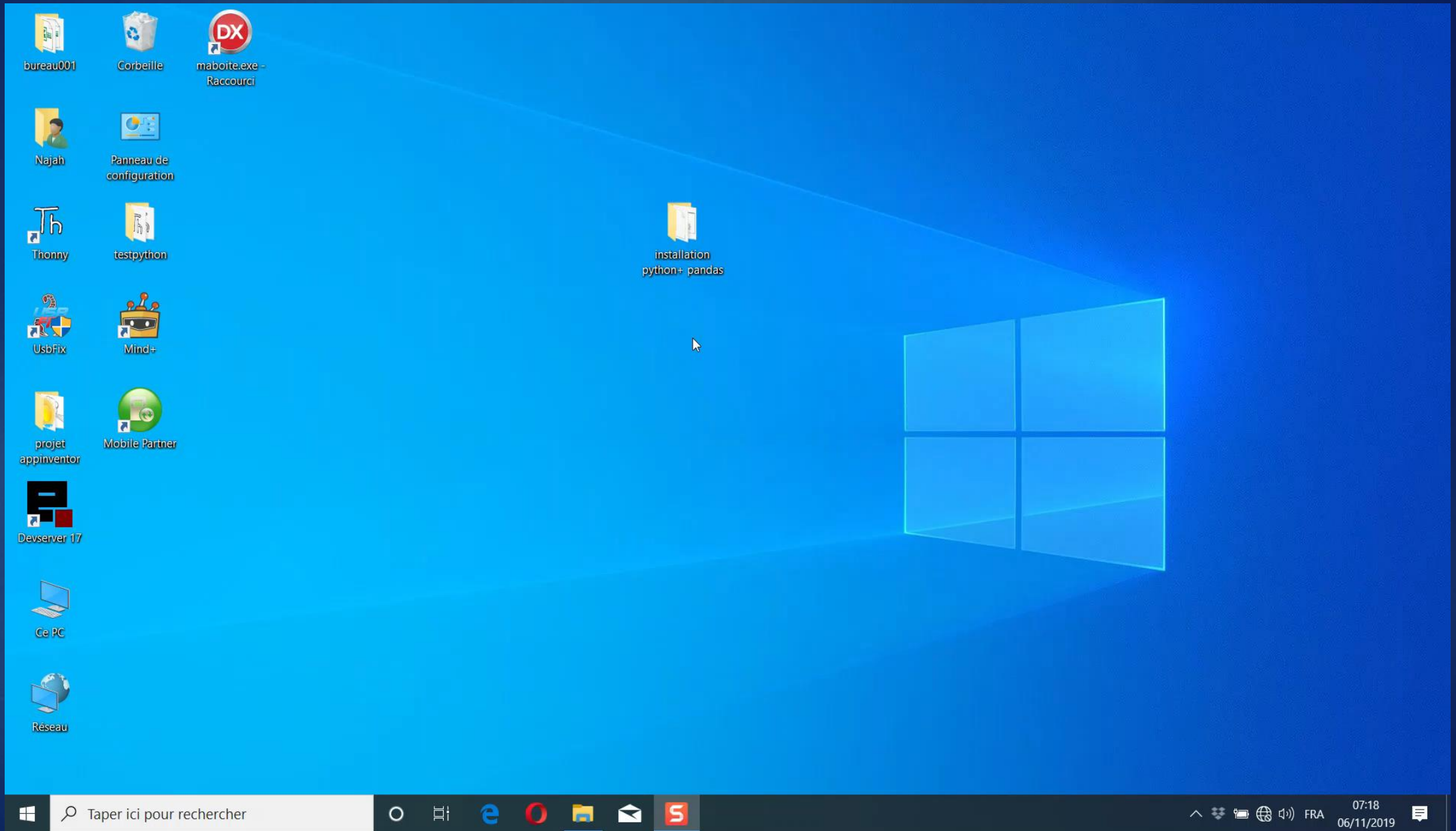


Installation de la
bibliothèque de
pandas



Lancement du
python et analyse
des données

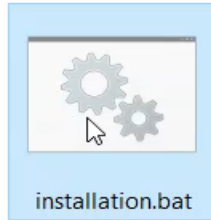
Installation de python



Installation de pandas



installation
python.mp4



installation.bat



numpy-1.17.2-cp
37-cp37m-win32.
whl



pandas-0.25.1-cp
37-cp37m-win32.
whl



python_dateutil-
2.8.0-py2.py3-no
ne-any.whl



python-3.7.4.exe



pytz-2019.2-py2.
py3-none-any.wh
l



six-1.12.0-py2.py
3-none-any.whl

Les étapes d'analyse des données avec pandas

Importation de la
bibliothèque “
pandas”

➤ `import pandas as pd`

Lecture des
données à partir
d'un fichier texte
ou CSV : création
d'un dataframe.

➤ `df = pd.read_csv('données.csv',
sep = ';', encoding='ANSI')`

Utilisation des
méthodes
prédéfinies de
python et
pandas

Analyse des données du projet

□ Question:

calculer le nombre total d'utilisation des appareils électroniques par utilisateur

Analyse des données du projet

□ Question:

calculer le nombre total d'utilisation des appareils électroniques par utilisateur

⇒ Démarche à suivre :

- Importation de la bibliothèque “ pandas” (Voir tutoriel).
- **import pandas as pd**
- Lecture des données à partir d'un fichier texte ou CSV : création d'un dataframe (*Voir tutoriel*).
- **df = pd.read_csv('données.csv', sep = ';' , encoding='ANSI')**
- Affichage de la liste des données (instruction facultative).
- **print (df)**
- Affichage les noms des colonnes (instruction facultative).
- **print (df.dtypes)**
- Taper la formule qui permet de calculer la somme des 4 colonnes (Education + Jeux + Réseaux Sociaux+ Communication).

```
print (df.Education+df.Jeux+df.Réseaux+df.Communication)
```

Analyse des données du projet

□ Question :

Calculer le nombre total des heures d'utilisation journalière par activité.

Analyse des données du projet

❑ Question :

Calculer le nombre total des heures d'utilisation journalière par activité.

➔ Démarche à suivre :

```
import pandas as pd
```

```
df = pd.read_csv('données.csv', sep = ';', encoding='ANSI')
```

```
print ("-----Le nombre total des heures d'utilisation journalière par activité ----- \n")
```

```
print ("Education :", df.Education.sum ())
```

```
print ("Jeux :", df.Jeux.sum ())
```

```
print (" Réseaux Sociaux :", df.Réseaux.sum ())
```

```
print ("Communication :", df.communication.sum ())
```

Analyse des données du projet

□ Question :

Calculer le nombre total des heures d'utilisation mensuelle par activité.

Analyse des données du projet

❑ Question :

Calculer le nombre total des heures d'utilisation mensuelle par activité.

➔ Démarche à suivre :

```
import pandas as pd
```

```
df = pd.read_csv('données.csv', sep = ';' , encoding='ANSI')
```

```
print ("-----Le nombre total des heures d'utilisation mensuelle par activité ----- \n")
```

```
print ("Education :", df.Education.sum ()*30)
```

```
print ("Jeux :", df.Jeux.sum ()*30)
```

```
print (" Réseaux Sociaux :", df.Réseaux.sum ()*30)
```

```
print ("Communication :", df.communication.sum ()*30)
```


Analyse des données du projet

□ Question :

Calculer la moyenne des dépenses mensuelles par les utilisateurs.

Analyse des données du projet

□ Question :

Calculer la moyenne des dépenses mensuelles par les utilisateurs.

⇒ Démarche à suivre :

```
import pandas as pd
```

```
df = pd.read_csv('données.csv', sep = ';' , encoding='ANSI')
```

```
print ("La moyenne des dépenses %.2f " %df.Frais.mean ())
```

```
##%.xf c'est à dire afficher un nombre sur x chiffres après le point décimal
```

Analyse des données du projet

□ **Question:**

Calculer la valeur maximale des dépenses.

Analyse des données du projet

□ Question:

Calculer la valeur maximale des dépenses.

⇒ Démarche à suivre :

```
import pandas as pd
```

```
df = pd.read_csv('données.csv', sep = ';' , encoding='ANSI')
```

```
print ("La valeur maximale des dépenses %.2f " %df.Frais.max ())
```

```
##%.xf c'est à dire afficher un nombre sur x chiffres après le point décimal
```

Analyse des données du projet

□ **Question N° :**

La liste des élèves qui sont dépendants de leurs appareils électroniques.

Analyse des données du projet

□ Question N° :

La liste des élèves qui sont dépendants de leurs appareils électroniques.

➔ Démarche à suivre :

```
import pandas as pd
```

```
df = pd.read_csv('données.csv', sep = ';' , encoding='ANSI')
```

```
print ("La liste des élèves qui sont dépendants : ", df[df['Dépendance']=='Oui'])
```


Analyse des données du projet

□ Question :

Extraire le nom et l'âge des élèves qui ont la valeur maximale de dépense.

Analyse des données du projet

□ Question :

Extraire le nom et l'âge des élèves qui ont la valeur maximale de dépense.

⇒ Démarche à suivre :

```
import pandas as pd
```

```
df = pd.read_csv('données.csv', sep = ';', encoding='ANSI')
```

```
maxi=df.Frais.max ()
```

```
dfres=df[df['Frais']==maxi]
```

```
print("Le nom et l'âge des élèves qui ont la valeur maximale de dépense \n ", dfres  
[['Prénom', 'Age']])
```

Analyse des données du projet

□ Question :

Afficher le nom des élèves qui ont une utilisation journalière de smartphone ≥ 5 heures.

Analyse des données du projet

□ Question :

Afficher le nom des élèves qui ont une utilisation journalière de smartphone ≥ 5 heures.

⇒ Démarche à suivre :

```
import pandas as pd
```

```
df = pd.read_csv('données.csv', sep = ';', encoding='ANSI')
```

```
df["Total"] = df.Education+df.Jeux+df.Réseaux+df.communication
```

```
dres=df[df["Total"]>=5]
```

```
print(dres["Prénom"])
```

Analyse des données du projet

□ Question :

Afficher le nom des élèves qui ont une utilisation journalière de smartphone $\geq X$ heures avec X est un nombre saisi par l'utilisateur.

Analyse des données du projet

□ Question :

Afficher le nom des élèves qui ont une utilisation journalière de smartphone $\geq X$ heures avec X est un nombre saisi par l'utilisateur.

⇒ Démarche à suivre :

```
import pandas as pd
```

```
df = pd.read_csv('données.csv', sep = ';' , encoding='ANSI')
```

```
df["Total"] = df.Education+df.Jeux+df.Réseaux+df.communication
```

```
X=eval(input("Saisir le nombre d'heures"))
```

```
dres=df[df["Total"]>=X]
```

```
print(dres["Prénom"])
```


Analyse des données du projet

□ Question :

Afficher le nom de l'activité (Education / Jeux / Réseaux sociaux/ Communication) la plus utilisée par les élèves .

Analyse des données du projet

□ Question :

Afficher le nom de l'activité (Education / Jeux / Réseaux sociaux/ Communication) la plus utilisée par les élèves .

⇒ Démarche à suivre :

```
import pandas as pd
```

```
df = pd.read_csv('données.csv', sep = ';', encoding='ANSI')
```

```
somedu = df.Education.sum ()
```

```
somjeu = df.Jeux.sum ()
```

```
somres = df.Réseaux.sum ()
```

```
somcom = df.communication.sum ()
```

```
if ((somedu>somjeu) & (somedu>somres) & (somedu >somcom )):
```

```
    print ("Education")
```

```
else :
```

```
    if ((somjeu>somedu) & (somjeu>somres) & (somjeu>somcom )):
```

```
        print ("Jeux")
```

```
    else :
```

```
        if ((somres>somjeu) & (somres>somedu) & (somres>somcom )):
```

```
            print ("Réseaux Sociaux")
```

```
        else:
```

```
            if ((somcom>somres) & (somcom>somjeu) (somcom>somedu )):
```

```
                print ("Communication")
```

Analyse des données du projet

❑ Question :

Pour chaque élève afficher d'une décision sur son indépendance des smartphones. L'indépendance et calculer de la façon suivantes:

Observation	Total utilisation
Neutre	0 Heure
Dépendant	>=5 Heures
Indépendant	Entre [1 et 4] heures

➔ Démarche à suivre :

```
import pandas as pd
```

```
df = pd.read_csv('données.csv', sep = ';' , encoding='ANSI')
```

```
df["Total"] = df.Education+df.Jeux+df.Réseaux+df.communication
```

Analyse des données du projet

```
for x in df["Total"] :  
    if (x== 0) :  
        print( " Neutre")  
    else :  
        if(1<=x<=4) :  
            print( " Indépendant")  
        else :  
            if(x>=5) :  
                print( " dépendant")
```