

1. Lire les données

```
import pandas as pd
df = pd.read_csv(source, encoding='latin1')
df.shape
df.columns
```

2. Nettoyer les données

```
df['ville'] = df['ville'].str.title()
```



3. Une variable qualitative

```
t1=df['genre'].value_counts()
t2=df['genre'].value_counts(normalize=True)
t3=round(df['genre'].value_counts(normalize=True)*100)
```

4. Une variable quantitative

```
moyenne=df['age'].mean()
mediane=df['age'].median()
```

5. Deux variables qualitatives

```
t5=pd.crosstab(df["ville"], df["joue_musique"], margins=True, margins_name='Total')
t6=round(pd.crosstab(df["ville"], df["joue_musique"], normalize='columns')*100)
t7=round(pd.crosstab(df["ville"], df["joue_musique"], normalize='index')*100)
```

6. Deux variables une qualitative et l'autre quantitative

```
age_moyen = df.groupby("genre")["age"].mean().reset_index()
```

7. Deux quantitatives

```
correlation = df["age"].corr(df["nb_ordi"])
```

8. Un graphique

```
import matplotlib.pyplot as plt
couleurs=['lightcoral','skyblue']
age_moyen.plot(kind="bar",color=couleurs)
plt.title("Âge moyen selon le genre")
plt.xlabel("Genres")
plt.ylabel("Âges")
plt.grid(axis='y', linestyle='--')
plt.xticks(rotation=0)
plt.show()
```

9. Retenir certaines variables

```
df1=df[['nom', 'genre','ville']]
```

10. Retenir certaines observations

```
df1 = df.query("ville == 'Laval'")
df2 = df.dropna(subset=["connais_python"])
```

12. Sauvegarder les résultats

```
df.to_csv("donnees.csv", index=False, encoding='latin1')
age_moyen.to_csv("age_moyen_tableau.csv", index=False, encoding='latin1')
plt.savefig("mon_graphique.png")
```

11. Regrouper des valeurs

```
groupe={ 'Magog':'Reste du Québec',
          'Carignan':'Reste du Québec',
          'Sherbrooke':'Reste du Québec',
          'Montréal':'Métropole',
          'Laval':'Métropole'}
```

```
df['region'] =
df['ville'].map(groupe).fillna("Autres villes")
```