

Python - Data Science for SL Lab Cheat Sheet

1. Importing at the beginning of the program?

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
```

2. How to define the Pandas Data Frame?

```
# If the csv file is file.csv, then:
df = pd.read_csv('file.csv')
```

3. How to map the values in the current column to a separate set of values?

```
# If the column name is 'A' and it can have two values: 0 or 1
df['A'] = df['A'].map({
    0: 'Zero',
    1: 'One'
});
```

4. How to print the data headers and first 5 rows?

```
print(df.head(5))
```

5. How to drop one or set of columns - inplace (the preferred way of doing it)?

```
columnsToDrop = ['A', 'B', 'C']
df.drop(columnsToDrop, axis=1, inplace=True)
# Print the data headers
print(df.head(5))
```

6. How to drop one or set of columns - reassign?

```
columnsToDrop = ['A', 'B', 'C']
df = df.drop(columnsToDrop, axis=1)
# Print the data headers
print(df.head(5))
```

7. How to add default values for a column?

```
# If the column name is 'A' and the default value is 'N', then:
df['A'].fillna('N')
```

8. How to visualize a a plot?

```
# There are 2 columns involved here. One is the 'X' column and the other is the 'Y' column.
ax = sns.countplot(x='X', hue='Y', palette='Set1', data=df)
ax.set(title='X vs Y Plot', xlabel='Ex', ylabel='Why')
plt.show()
```

9. How to print crosstab between 'X' and 'Y' where 'X' is an independent column and 'Y' is dependent?

```
print(pd.crosstab(df['X'], df.Y))
```

10. How to cut a variable into intervals and then plot it?

```
# If there are n values in interval
interval = (0, 1, 2, 3)
# There should be n-1 categories
categories = ['A', 'B', 'C']
# If we want to cut the column 'X', then:
df['X_Cuts'] = pd.cut(df.X, interval, labels=categories)
# To plot this, use x='X_Cuts' and hue = as asked in question
```