Practical 3: Descriptive Statistics - Cheat Sheet

Theory Concepts

```
Measures of Central Tendency:
- Mean: Average of values.
- Median: Middle value when data is sorted.
- Mode: Most frequent value.

Measures of Variability (Dispersion):
- Standard Deviation (std): Spread around the mean.
- Minimum & Maximum: Range of values.
- Percentiles: Data value below which a certain % of data falls.

Grouped Statistics:
-  Grouping allows analysis within categories (e.g., class-wise stats, species-wise stats).

Useful Functions:
- mean(), median(), std(), min(), max(), describe()
- groupby('column').describe() for grouped stats
```

Part A: Summary Stats Grouped by Categorical Variable

```
import pandas as pd
import numpy as np

df = pd.read_csv('student.csv')
df

df.mean()
df.median()
df.std()
df.min()
df.max()

np.std(df['marks'])

group_by_class = df.groupby('class')
group_by_class.mean()
group_by_class.describe()

group_by_class.get_group('TE').describe()

group_by_age = df.groupby('age')
group_by_age.groups
group_by_age.get_group(20).describe()
```

Practical 3: Descriptive Statistics - Cheat Sheet

Part B: Summary Stats for Each Species in Iris Dataset

```
df = pd.read_csv('iris.csv')
df

df.isnull().sum()
df.describe()
df['sepal_width'].describe()
df.groupby('species').describe()

species_list = ['Iris-setosa', 'Iris-versicolor', 'Iris-virginica']

for species in species_list:
    print(f"\nSummary Statistics for {species}:\n")
    print(df[df['species'] == species].describe())
    print("\n" + "-"*50 + "\n")
```