

# Data Analysis summary

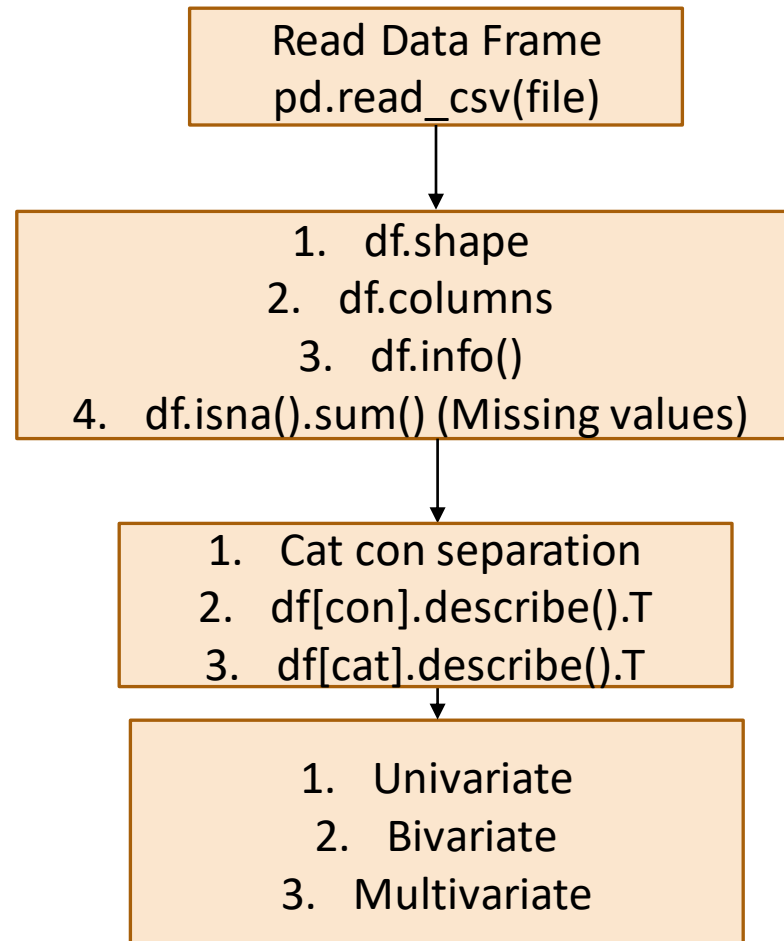
---

UTKARSH GAIKWAD

CLASS STARTING SHARP AT 8:10 PM

# Steps to follow in data analysis

---

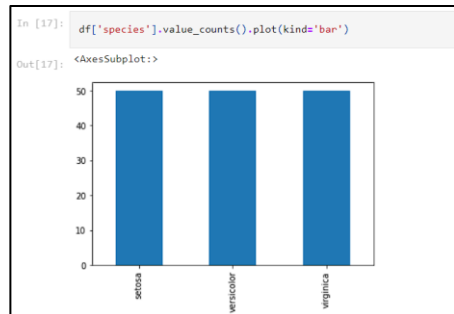


# Univariate analysis

---

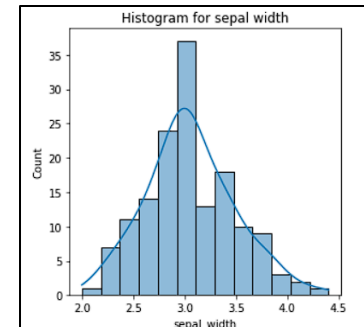
Categorical Features  
(Containing Text)

Countplot  
`df.value_counts()`  
`df.value_counts().plot(kind='bar')`



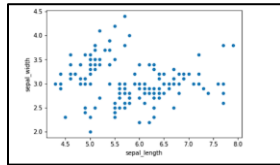
Continuous Features  
(Numerical Features)

Histogram  
`import seaborn as sns`  
`sns.histplot(data=df,x='column_name',kde=True)`



# Bivariate Analysis

Continuous vs Continuous



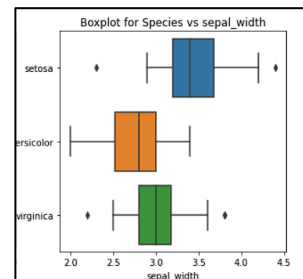
Scatterplot

```
Import matplotlib.pyplot as plt  
plt.scatter(df['c1'], df['c2'])
```

Categorical vs Continuous

Boxplot

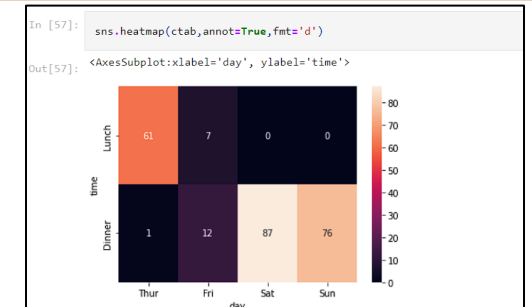
```
Import seaborn as sns  
sns.boxplot(data=df, x='c1', y='c2')
```



Categorical vs Categorical

Crosstab

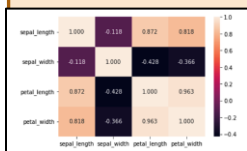
```
ctab = pd.crosstab(df['cat1'], df['cat2'])  
sns.heatmap(ctab, annot=True, fmt='d')
```



Correlation heatmap

Import seaborn as sns

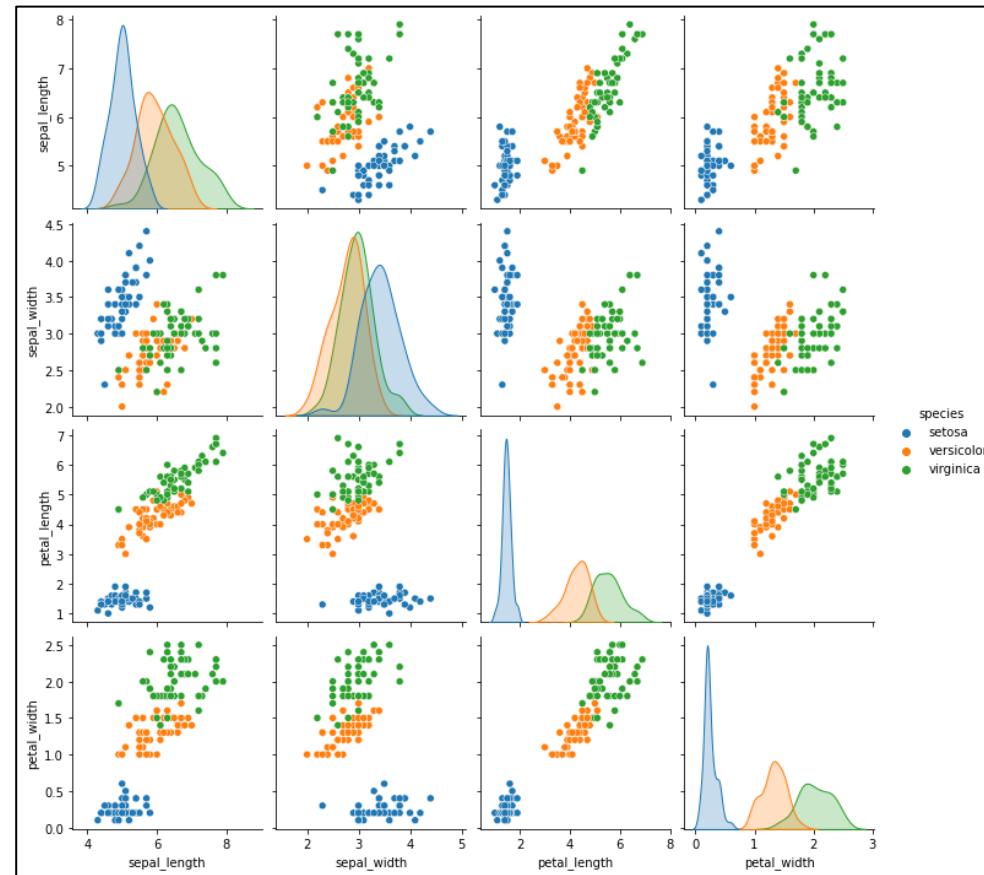
```
sns.heatmap(df.corr(),annot=True,fmt='.3f')
```



# Multivariate analysis

All continuous variables

Import seaborn as sns  
`sns.pairplot(df, hue='c1')`



# Thank you

---

FOR ANY QUERIES PING ME ON SKYPE GROUP