```
#1. Create a dataframe with the {Name-string, Color-categorical, Salary-float64}
#2. Add 10 rows in it


import pandas as pd

df = pd.DataFrame(columns=['Name', 'Color', 'Salary'])

a=['Homit',
   'Rohan',
   'Rahul',
   'Pratham',
   'Dev',
   'Jash',
   'Melvin',
   'Parth',
   'Zubin',
   'Dhruv']
b=['A',
   'B',
   'C',
   'A',
   'B',
   'C',
   'A',
   'B',
   'C',
   'A']
c=['3000',
   '8000',
   '7000',
   '4000',
   '3000',
   '6000',
   '9000',
   '4000',
   '9000',
   '2000']

df['Name']=a
df['Color']=b
df['Salary']=c

df
```

	Name	Color	Salary
0	Homit	A	3000
1	Rohan	B	8000
2	Rahul	C	7000
3	Pratham	A	4000
4	Dev	B	3000
5	Jash	C	6000
6	Melvin	A	9000
7	Parth	B	4000
8	Zubin	C	9000
9	Dhruv	A	2000

```
print(df['Name'].astype(str))
print(df['Color'].astype(str))
print(df['Salary'].astype(float))
```

```
0      Homit
1      Rohan
2      Rahul
3    Pratham
4        Dev
5       Jash
6     Melvin
7      Parth
8      Zubin
9      Dhruv
Name: Name, dtype: object
0    A
1    B
2    C
3    A
4    B
5    C
6    A
7    B
8    C
9    A
Name: Color, dtype: object
0    3000.0
1    8000.0
2    7000.0
3    4000.0
4    3000.0
5    6000.0
6    9000.0
7    4000.0
8    9000.0
9    2000.0
Name: Salary, dtype: float64
```

```
df.dtypes
```

```
Name      object
Color     object
Salary    object
dtype: object
```

```
df['Salary'] = df['Salary'].astype(float)
df.dtypes
```

```
Name       object
Color      object
Salary    float64
dtype: object
```

```
#3. Summarize the numeric and categorical data
df['Color'] = pd.Categorical(df.Color)
df.dtypes
```

```
Name       object
Color    category
Salary    float64
dtype: object
```

```
#Create missing data by adding ‘expenditure’ column
import numpy as np

df['Expenditure']=np.nan
df
```

	Name	Color	Salary	Expenditure
0	Homit	A	3000.0	NaN
1	Rohan	B	8000.0	NaN
2	Rahul	C	7000.0	NaN
3	Pratham	A	4000.0	NaN

```
#Replace the missing values
df['Expenditure']=df['Expenditure'].fillna(500)
#df['Expenditure'] = df['Expenditure'].fillna(df['Salary'].mean())
df
```

	Name	Color	Salary	Expenditure
0	Homit	A	3000.0	500.0
1	Rohan	B	8000.0	500.0
2	Rahul	C	7000.0	500.0
3	Pratham	A	4000.0	500.0
4	Dev	B	3000.0	500.0
5	Jash	C	6000.0	500.0
6	Melvin	A	9000.0	500.0
7	Parth	B	4000.0	500.0
8	Zubin	C	9000.0	500.0
9	Dhruv	A	2000.0	500.0

```
#Convert categorical to numeric data
df2 = pd.get_dummies(df['Color'])

#df = pd.concat([df, df2], axis=1)
df2
```

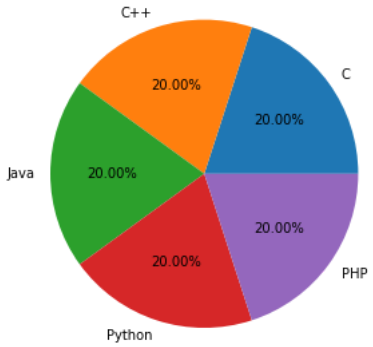
	A	B	C
0	1	0	0
1	0	1	0
2	0	0	1
3	1	0	0
4	0	1	0
5	0	0	1
6	1	0	0
7	0	1	0
8	0	0	1
9	1	0	0

```
df.dtypes
```

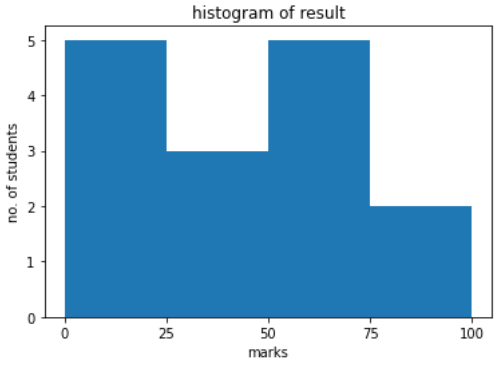
```
Name            object
Color         category
Salary         float64
Expenditure    float64
dtype: object
```

```
#Save as csv
df.to_csv("internalExam.csv")

import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

fig = plt.figure()
```

```
ax = fig.add_axes([0,0,1,1])
ax.axis('equal')

langs = ['C', 'C++', 'Java', 'Python', 'PHP']
students = [20,20,20,20,20]
ax.pie(students, labels = langs,autopct='%1.2f%%')

plt.savefig('temp_rainfall.png')
plt.show()
```




```
from matplotlib import pyplot as plt
import numpy as np
#fig,ax=plt.subplots(1,1)
ax=plt.axes()
a = np.array([22,87,5,43,56,73,55,54,11,20,51,5,79,31,27])
ax.hist(a, bins = [0,25,50,75,100])
ax.set_title("histogram of result")
ax.set_xticks([0,25,50,75,100])
ax.set_xlabel('marks')
ax.set_ylabel('no. of students')
plt.show()
```

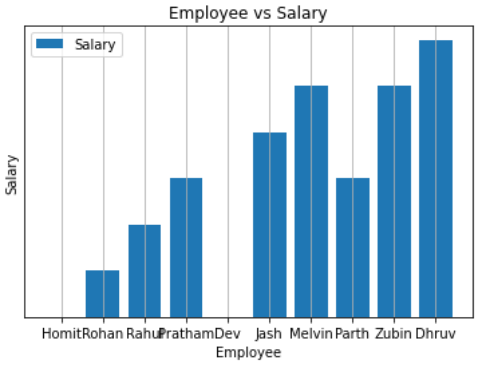



```
import pandas as pd
import numpy as np
dfx = pd.DataFrame((np.random.rand(10, 5)*10), columns=['A', 'B', 'C', 'D', 'E'])
print(dfx)
dfx.plot.box(grid='True')
```

```
          A         B         C         D         E
0  8.070099  7.438033  7.841374  3.881273  8.132982
1  6.857663  8.907998  4.770600  1.031157  3.593411
2  5.477069  0.522547  4.375965  5.469581  5.541921
3  9.872664  1.193163  8.753880  1.380842  4.192267
4  1.350697  5.827984  7.407855  8.746359  1.410679
5  5.544392  2.737899  3.393667  9.852359  9.478049
6  0.661010  6.695827  8.463167  7.989317  8.800858
7  3.991899  5.779331  8.505287  7.033493  5.824527
8  6.753812  2.390489  1.799924  8.309370  6.695958
9  8.875934  7.667562  5.119054  4.057171  8.254622
<matplotlib.axes._subplots.AxesSubplot at 0x7f8efc37e090>
```

Double-click (or enter) to edit

```
#Make a graph
import matplotlib.pyplot as plt
x=df['Name']
y=df['Salary']

ax=plt.axes()
plt.xlabel("Employee")
plt.ylabel("Salary")
plt.title("Employee vs Salary")

ax.set_yticks([2000, 5000])
plt.bar(x,y)
plt.grid(True)
plt.legend(['Salary'])
plt.show()
```




```
st = 'hello'
print(st, 'st')
```

```
hello st
```

```
!pip install nltk
import nltk
nltk.download('wordnet')
nltk.download('punkt')
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: nltk in /usr/local/lib/python3.7/dist-packages (3.7)
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from nltk) (1.2.0)
Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages (from nltk) (7.1.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.7/dist-packages (from nltk) (2022.6.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages (from nltk) (4.64.1)
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
True
```

```
# import these modules

from nltk.stem import WordNetLemmatizer

lemmatizer = WordNetLemmatizer()
```

```
print("rocks :", lemmatizer.lemmatize("rocks"))
print("corpora :", lemmatizer.lemmatize("corpora"))

# a denotes adjective in "pos"
print("better :", lemmatizer.lemmatize("better", pos ="a"))
```

```
---------------------------------------------------------------------------
LookupError                               Traceback (most recent call last)
/usr/local/lib/python3.7/dist-packages/nltk/corpus/util.py in __load(self)
     83                 try:
---> 84                     root = nltk.data.find(f"{self.subdir}/{zip_name}")
     85                 except LookupError:

------------------------------------------------------------ 10 frames ------------------------------------------------------------
LookupError:
**********************************************************************
  Resource omw-1.4 not found.
  Please use the NLTK Downloader to obtain the resource:

  >>> import nltk
  >>> nltk.download('omw-1.4')

  For more information see: https://www.nltk.org/data.html

  Attempted to load corpora/omw-1.4.zip/omw-1.4/

  Searched in:
    - '/root/nltk_data'
    - '/usr/nltk_data'
    - '/usr/share/nltk_data'
    - '/usr/lib/nltk_data'
    - '/usr/share/nltk_data'
    - '/usr/local/share/nltk_data'
    - '/usr/lib/nltk_data'
    - '/usr/local/lib/nltk_data'
**********************************************************************


During handling of the above exception, another exception occurred:

LookupError                               Traceback (most recent call last)
/usr/local/lib/python3.7/dist-packages/nltk/data.py in find(resource_name,
paths)
    581     sep = "*" * 70
    582     resource_not_found = f"\n{sep}\n{msg}\n{sep}\n"
--> 583     raise LookupError(resource_not_found)
    584
    585

LookupError:
**********************************************************************
  Resource omw-1.4 not found.
  Please use the NLTK Downloader to obtain the resource:

  >>> import nltk
  >>> nltk.download('omw-1.4')

  For more information see: https://www.nltk.org/data.html

  Attempted to load corpora/omw-1.4

  Searched in:
    - '/root/nltk_data'
```