```
#problem solving survey of collected data of startup using dataset
```

```
#step 1:
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
from google.colab import files
uploaded = files.upload()
```

```
Choose files No file chosen    Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving 50_Startups.csv to 50_Startups.csv
```

```
df = pd.read_csv('50_Startups.csv')
```

```
---------------------------------------------------------------------------
FileNotFoundError                         Traceback (most recent call last)
/tmp/ipython-input-3169187586.py in <cell line: 0>()
----> 1 df = pd.read_csv('50_Startups.csv')

 4 frames
/usr/local/lib/python3.12/dist-packages/pandas/io/common.py in get_handle(path_or_buf, mode, encoding, compression, memory_map, is_text, errors, storage_options)
    871         if ioargs.encoding and "b" not in ioargs.mode:
    872             # Encoding
--> 873             handle = open(
    874                 handle,
    875                 ioargs.mode,

FileNotFoundError: [Errno 2] No such file or directory: '50_Startups.csv'
```

Next steps: Explain error

```
import pandas as pd
df = pd.read_csv('50_Startups.csv')
```

```
---------------------------------------------------------------------------
FileNotFoundError                         Traceback (most recent call last)
/tmp/ipython-input-3909762965.py in <cell line: 0>()
      1 import pandas as pd
----> 2 df = pd.read_csv('50_Startups.csv')

 4 frames
/usr/local/lib/python3.12/dist-packages/pandas/io/common.py in get_handle(path_or_buf, mode, encoding, compression, memory_map, is_text, errors, storage_options)
    871         if ioargs.encoding and "b" not in ioargs.mode:
    872             # Encoding
--> 873             handle = open(
    874                 handle,
    875                 ioargs.mode,

FileNotFoundError: [Errno 2] No such file or directory: '50_Startups.csv'
```

Next steps: Explain error

```
df.head()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-1344414169.py in <cell line: 0>()
----> 1 df.head()
      2

NameError: name 'df' is not defined
```

```
df.info()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-3771845804.py in <cell line: 0>()
----> 1 df.info()

NameError: name 'df' is not defined
```

```
df.shape
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-857714779.py in <cell line: 0>()
----> 1 df.shape

NameError: name 'df' is not defined
```

```
df  = df.drop('State',axis=True)
```

```
df.head()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
```

```
corr = df.corr()
sns.heatmap(corr,annot = True)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-2889783282.py in <cell line: 0>()
----> 1 corr = df.corr()
      2 sns.heatmap(corr,annot = True)

NameError: name 'df' is not defined
```

```
x = df.drop('Profit', axis=True)
y = df['Profit']
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-1960602326.py in <cell line: 0>()
----> 1 x = df.drop('Profit', axis=True)
      2 y = df['Profit']

NameError: name 'df' is not defined
```

```
x.head() #before standardize
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-3190858357.py in <cell line: 0>()
----> 1 x.head() #before standardize

NameError: name 'x' is not defined
```

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
x = sc.fit_transform(x)
```

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=0)
```

```
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-4183739024.py in <cell line: 0>()
----> 1 x_train.shape,x_test.shape,y_train.shape,y_test.shape

NameError: name 'x_train' is not defined
```

```
from sklearn.linear_model import LinearRegression
regression = LinearRegression()
regression.fit(x_train, y_train)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-2128641244.py in <cell line: 0>()
      1 from sklearn.linear_model import LinearRegression
      2 regression = LinearRegression()
----> 3 regression.fit(x_train, y_train)

NameError: name 'x_train' is not defined
```

```
x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

```
from sklearn.linear_model import LinearRegression
regression = LinearRegression()
regression.fit(x_train, y_train)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-2128641244.py in <cell line: 0>()
      1 from sklearn.linear_model import LinearRegression
      2 regression = LinearRegression()
----> 3 regression.fit(x_train, y_train)

NameError: name 'x_train' is not defined
```

```
y_pred = regression.predict(x_test).round(1)
```

```
calculation = pd.DataFrame(np.c_[y_test, y_pred], columns=["Original Salary", "predict Salary"])
calculation.head(5)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-2835968914.py in <cell line: 0>()
----> 1 calculation = pd.DataFrame(np.c_[y_test, y_pred], columns=["Original Salary", "predict Salary"])
      2 calculation.head(5)

NameError: name 'pd' is not defined
```

```
print("Training Accuracy", regression.score(x_train,y_train))
print("test Accuracy", regression.score(x_test,y_test))
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-4156230847.py in <cell line: 0>()
----> 1 print("Training Accuracy", regression.score(x_train,y_train))
      2 print("test Accuracy", regression.score(x_test,y_test))

NameError: name 'x_train' is not defined
```

```
regression.intercept_
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
/tmp/ipython-input-2787009467.py in <cell line: 0>()
----> 1 regression.intercept_

AttributeError: 'LinearRegression' object has no attribute 'intercept_'
```

```
regression.coef_
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
/tmp/ipython-input-1325692839.py in <cell line: 0>()
----> 1 regression.coef_

AttributeError: 'LinearRegression' object has no attribute 'coef_'
```

```
feature=[1653490.0,136897]
```

```
feature = [165349.20,136897.80,471784.10]
scale_feature = sc.transform([feature])
print(scale_feature)
```

```
y_pred_test = regressor.predict(scale_feature)
y_pred_test #by using library
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-2515437553.py in <cell line: 0>()
----> 1 y_pred_test = regressor.predict(scale_feature)
      2 y_pred_test #by using library

NameError: name 'regressor' is not defined
```

```
Start coding or generate with AI.
```

```
y_pred_test = regression.predict(scale_feature)
y_pred_test #by using library
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/tmp/ipython-input-666096526.py in <cell line: 0>()
----> 1 y_pred_test = regression.predict(scale_feature)
      2 y_pred_test #by using library

NameError: name 'regression' is not defined
```