

In []:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans, k_means
```

In []:

```
url = 'C:\\\\Users\\mohit\\Desktop\\LP III\\ML\\1\\sales_data_sample.csv'
df = pd.read_csv(url, encoding = 'latin1')
```

In []:

```
df.shape
```

In []:

```
df.head()
```

In []:

```
df.info()
```

In []:

```
df.describe()
```

In []:

```
df.isnull().sum()
```

In []:

```
df.dtypes
```

In []:

```
df_drop = ['ORDERNUMBER', 'ORDERDATE', 'STATUS', 'CUSTOMERNAME', 'PHONE', 'ADDRESSLINE1', 'AD  
df = df.drop(df_drop, axis = 1)
```

In []:

```
df
```

In []:

```
df.dtypes
```

In []:

```
print(df['DEALSIZE'].unique())  
print(df['PRODUCTLINE'].unique())
```

In []:

```
product_line = pd.get_dummies(df['PRODUCTLINE'])  
deal_size = pd.get_dummies(df['DEALSIZE'])
```

In []:

```
df = pd.concat([df, product_line, deal_size], axis = 1)  
df
```

In []:

```
df_drop = ['PRODUCTLINE', 'DEALSIZE']  
df = df.drop(df_drop, axis = 1)
```

In []:

```
df
```

In []:

```
df['PRODUCTCODE'] = pd.Categorical(df['PRODUCTCODE']).codes
```

In []:

```
x_train = df.values  
x_train.shape
```

In []:

```
model = KMeans(n_clusters = 3)  
model = model.fit(x_train)  
preds = model.predict(x_train)
```

In []:

```
unique, counts = np.unique(preds, return_counts = True)  
unique
```

In []:

```
counts = counts.reshape(1, 3)  
counts
```

In []:

```
counts_df = pd.DataFrame(counts, columns = ['1', '2', '3'])
```

In []:

```
counts_df
```

In []:

```
d = []
r = range(1, 10)

for i in r:
    model = KMeans(n_clusters = i)
    model = model.fit(df)
    d.append(model.inertia_)
```

In []:

```
plt.plot(r, d, 'bo-')
plt.xlabel("n")
plt.ylabel('distortion')
plt.title('elbow method plot')
plt.show()
```