```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt


import pandas as pd
```

```
df = pd.read_csv('/content/sales_data_sample.csv',encoding="ISO-8859-1")
```

```
df.head()
```

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	S
0	10107	30	95.70	2	2871.00	2/24/2003 0:00	S
1	10121	34	81.35	5	2765.90	5/7/2003 0:00	S
2	10134	41	94.74	2	3884.34	7/1/2003 0:00	S
			83.26	6	3746.70	8/25/2003 0:00	S
4	10159	49	100.00	14	5205.27	10/10/2003 0:00	S

Saved successfully! ×

5 rows × 25 columns

🪄

```
df.dtypes
```

```
ORDERNUMBER          int64
QUANTITYORDERED      int64
PRICEEACH          float64
ORDERLINENUMBER      int64
SALES              float64
ORDERDATE           object
STATUS              object
QTR_ID               int64
MONTH_ID             int64
YEAR_ID              int64
PRODUCTLINE         object
MSRP                 int64
PRODUCTCODE         object
```

```
CUSTOMERNAME         object
PHONE                object
ADDRESSLINE1         object
ADDRESSLINE2         object
CITY                 object
STATE                object
POSTALCODE           object
COUNTRY              object
TERRITORY            object
CONTACTLASTNAME      object
CONTACTFIRSTNAME     object
DEALSIZE             object
dtype: object
```

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

```
ORDERNUMBER              0
QUANTITYORDERED          0
PRICEEACH                0
ORDERLINENUMBER          0
SALES                    0
ORDERDATE                0
STATUS                   0
QTR_ID                   0
MONTH_ID                 0
YEAR_ID                  0
PRODUCTLINE              0
MSRP                     0
PRODUCTCODE              0
```

Saved successfully!   ✕

```
ADDRESSLINE1             0
ADDRESSLINE2          2521
CITY                     0
STATE                 1486
POSTALCODE              76
COUNTRY                  0
TERRITORY             1074
CONTACTLASTNAME          0
CONTACTFIRSTNAME         0
DEALSIZE                 0
dtype: int64
```

```python
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2823 entries, 0 to 2822
Data columns (total 25 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   ORDERNUMBER      2823 non-null   int64
 1   QUANTITYORDERED  2823 non-null   int64
 2   PRICEEACH        2823 non-null   float64
 3   ORDERLINENUMBER  2823 non-null   int64
 4   SALES            2823 non-null   float64
```

```
 5   ORDERDATE         2823 non-null    object
 6   STATUS            2823 non-null    object
 7   QTR_ID            2823 non-null    int64
 8   MONTH_ID          2823 non-null    int64
 9   YEAR_ID           2823 non-null    int64
 10  PRODUCTLINE       2823 non-null    object
 11  MSRP              2823 non-null    int64
 12  PRODUCTCODE       2823 non-null    object
 13  CUSTOMERNAME      2823 non-null    object
 14  PHONE             2823 non-null    object
 15  ADDRESSLINE1      2823 non-null    object
 16  ADDRESSLINE2      302 non-null     object
 17  CITY              2823 non-null    object
 18  STATE             1337 non-null    object
 19  POSTALCODE        2747 non-null    object
 20  COUNTRY           2823 non-null    object
 21  TERRITORY         1749 non-null    object
 22  CONTACTLASTNAME   2823 non-null    object
 23  CONTACTFIRSTNAME  2823 non-null    object
 24  DEALSIZE          2823 non-null    object
dtypes: float64(2), int64(7), object(16)
memory usage: 551.5+ KB
```
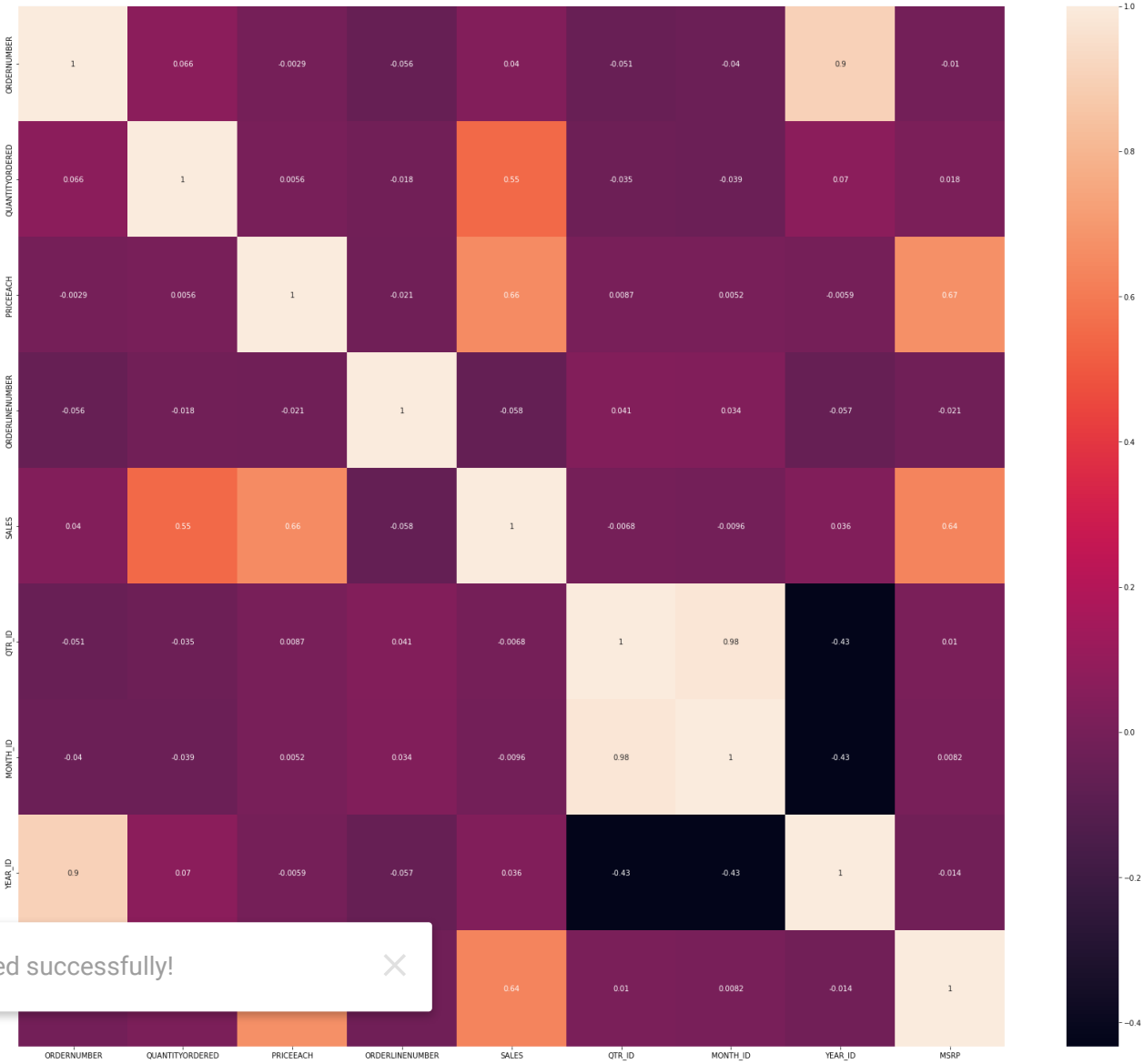
```python
plt.figure(figsize = (30,26))
sns.heatmap(df.corr(),annot = True)
```

Saved successfully!   ✕

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f69cd9aead0>
```


Saved successfully!

```
df.shape
```

```
(2823, 25)
```

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

```
ORDERNUMBER         0
QUANTITYORDERED     0
PRICEEACH           0
ORDERLINENUMBER     0
SALES               0
ORDERDATE           0
STATUS              0
QTR_ID              0
MONTH_ID            0
YEAR_ID             0
PRODUCTLINE         0
MSRP                0
PRODUCTCODE         0
```

```
CUSTOMERNAME          0
PHONE                 0
ADDRESSLINE1          0
ADDRESSLINE2       2521
CITY                  0
STATE              1486
POSTALCODE           76
COUNTRY               0
TERRITORY          1074
CONTACTLASTNAME       0
CONTACTFIRSTNAME      0
DEALSIZE              0
dtype: int64
```

df.dtypes

```
ORDERNUMBER          int64
QUANTITYORDERED      int64
PRICEEACH          float64
ORDERLINENUMBER      int64
SALES              float64
ORDERDATE           object
STATUS              object
QTR_ID               int64
MONTH_ID             int64
YEAR_ID              int64
PRODUCTLINE         object
MSRP                 int64
PRODUCTCODE         object
```

Saved successfully!   ✕

```
ADDRESSLINE1        object
ADDRESSLINE2        object
CITY                object
STATE               object
POSTALCODE          object
COUNTRY             object
TERRITORY           object
CONTACTLASTNAME     object
CONTACTFIRSTNAME    object
DEALSIZE            object
dtype: object
```

```python
country = pd.get_dummies(df['COUNTRY'])
productline = pd.get_dummies(df['PRODUCTLINE'])
Dealsize = pd.get_dummies(df['DEALSIZE'])
```

```python
df = pd.concat([df,country,productline,Dealsize], axis = 1)
```

```python
df.head()
```

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	S
0	10107	30	95.70	2	2871.00	2/24/2003 0:00	S
1	10121	34	81.35	5	2765.90	5/7/2003 0:00	S
2	10134	41	94.74	2	3884.34	7/1/2003 0:00	S
3	10145	45	83.26	6	3746.70	8/25/2003 0:00	S
4	10159	49	100.00	14	5205.27	10/10/2003 0:00	S

5 rows × 54 columns

🪄✨

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

```python
df.dtypes
```

Saved successfully!   ✕

```
ORDERLINENUMBER      int64
SALES                float64
ORDERDATE            object
STATUS               object
QTR_ID               int64
MONTH_ID             int64
YEAR_ID              int64
MSRP                 int64
PRODUCTCODE          object
CUSTOMERNAME         object
PHONE                object
ADDRESSLINE1         object
ADDRESSLINE2         object
CITY                 object
STATE                object
POSTALCODE           object
TERRITORY            object
CONTACTLASTNAME      object
CONTACTFIRSTNAME     object
Australia            uint8
Austria              uint8
Belgium              uint8
Canada               uint8
Denmark              uint8
Finland              uint8
France               uint8
```

```
Germany                uint8
Ireland                uint8
Italy                  uint8
Japan                  uint8
Norway                 uint8
Philippines            uint8
Singapore              uint8
Spain                  uint8
Sweden                 uint8
Switzerland            uint8
UK                     uint8
USA                    uint8
Classic Cars           uint8
Motorcycles            uint8
Planes                 uint8
Ships                  uint8
Trains                 uint8
Trucks and Buses       uint8
Vintage Cars           uint8
Large                  uint8
Medium                 uint8
Small                  uint8
dtype: object
```

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

```python
df.dtypes
```

Saved successfully!   ✕

```
QUANTITYORDERED        int64
PRICEEACH            float64
ORDERLINENUMBER        int64
SALES                float64
ORDERDATE             object
STATUS                object
QTR_ID                 int64
MONTH_ID               int64
YEAR_ID                int64
MSRP                   int64
PRODUCTCODE             int8
CUSTOMERNAME          object
PHONE                 object
ADDRESSLINE1          object
ADDRESSLINE2          object
CITY                  object
STATE                 object
POSTALCODE            object
TERRITORY             object
CONTACTLASTNAME       object
CONTACTFIRSTNAME      object
Australia              uint8
Austria                uint8
Belgium                uint8
Canada                 uint8
Denmark                uint8
Finland                uint8
```

```
France              uint8
Germany             uint8
Ireland             uint8
Italy               uint8
Japan               uint8
Norway              uint8
Philippines         uint8
Singapore           uint8
Spain               uint8
Sweden              uint8
Switzerland         uint8
UK                  uint8
USA                 uint8
Classic Cars        uint8
Motorcycles         uint8
Planes              uint8
Ships               uint8
Trains              uint8
Trucks and Buses    uint8
Vintage Cars        uint8
Large               uint8
Medium              uint8
Small               uint8
dtype: object
```

```python
df.drop('ORDERDATE', axis=1, inplace=True)
```

Saved successfully!   ✕

```
ORDERNUMBER         int64
QUANTITYORDERED     int64
PRICEEACH           float64
ORDERLINENUMBER     int64
SALES               float64
STATUS              object
QTR_ID              int64
MONTH_ID            int64
YEAR_ID             int64
MSRP                int64
PRODUCTCODE         int8
CUSTOMERNAME        object
PHONE               object
ADDRESSLINE1        object
ADDRESSLINE2        object
CITY                object
STATE               object
POSTALCODE          object
TERRITORY           object
CONTACTLASTNAME     object
CONTACTFIRSTNAME    object
Australia           uint8
Austria             uint8
Belgium             uint8
Canada              uint8
Denmark             uint8
```

```
Finland            uint8
France             uint8
Germany            uint8
Ireland            uint8
Italy              uint8
Japan              uint8
Norway             uint8
Philippines        uint8
Singapore          uint8
Spain              uint8
Sweden             uint8
Switzerland        uint8
UK                 uint8
USA                uint8
Classic Cars       uint8
Motorcycles        uint8
Planes             uint8
Ships              uint8
Trains             uint8
Trucks and Buses   uint8
Vintage Cars       uint8
Large              uint8
Medium             uint8
Small              uint8
dtype: object
```

```python
from sklearn.cluster import KMeans
```

Saved successfully!   ✕