

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

import warnings
warnings.filterwarnings("ignore", category=FutureWarning)
```

```
df = pd.read_csv('/content/Train.csv')
df.head()
```



	ID	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating
0	1	D	Flight	4	
1	2	F	Flight	4	
2	3	A	Flight	2	
3	4	B	Flight	3	
4	5	C	Flight	2	

```
df.shape
```



```
(10999, 12)
```

```
df.dtypes
```



	0
ID	int64
Warehouse_block	object
Mode_of_Shipment	object
Customer_care_calls	int64
Customer_rating	int64
Cost_of_the_Product	int64
Prior_purchases	int64
Product_importance	object
Gender	object
Discount_offered	int64
Weight_in_gms	int64
Reached.on.Time Y.N	int64

dtype: object

```
df.drop(['ID'], axis=1, inplace=True)
```

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



	0
Warehouse_block	0
Mode_of_Shipment	0
Customer_care_calls	0
Customer_rating	0
Cost_of_the_Product	0
Prior_purchases	0
Product_importance	0
Gender	0
Discount_offered	0
Weight_in_gms	0
Reached.on.Time_Y.N	0

dtype: int64

```
df.duplicated().sum()
```



0

```
df.describe()
```



	Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_pur
count	10999.000000	10999.000000	10999.000000	10999.0
mean	4.054459	2.990545	210.196836	3.5
std	1.141490	1.413603	48.063272	1.5
min	2.000000	1.000000	96.000000	2.0
25%	3.000000	2.000000	169.000000	3.0
50%	4.000000	3.000000	214.000000	3.0
75%	5.000000	4.000000	251.000000	4.0
max	7.000000	5.000000	310.000000	10.0

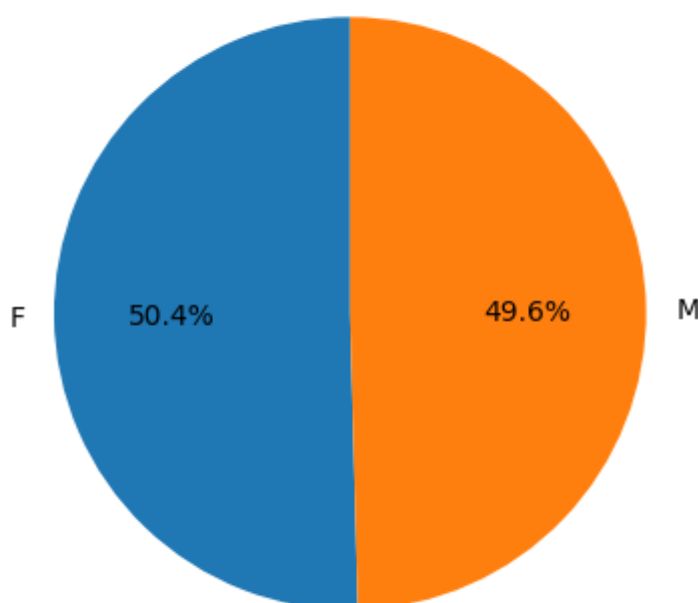
```
df.head()
```

	Warehouse_block	Mode_of_Shipment	Customer_care_calls	Customer_rating
0	D	Flight	4	2
1	F	Flight	4	5
2	A	Flight	2	2
3	B	Flight	3	3
4	C	Flight	2	2

```
plt.pie(df['Gender'].value_counts(), labels = ['F', 'M'], autopct='%1.1f%%', star
plt.title('Gender Distribution')
```

```
Text(0.5, 1.0, 'Gender Distribution')
```

Gender Distribution



```
df.replace([np.inf, -np.inf], np.nan, inplace=True)
```

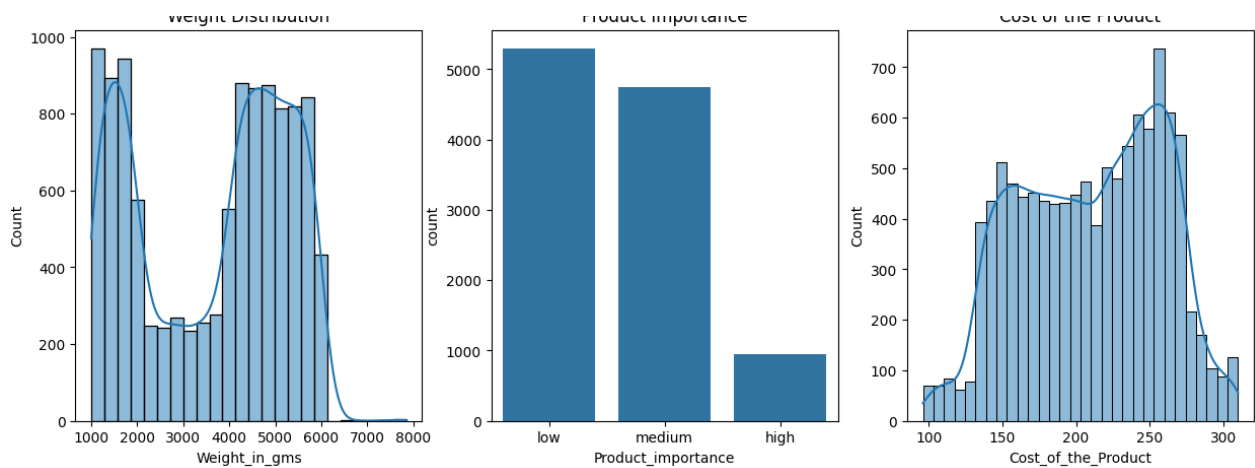
```
fig, ax = plt.subplots(1,3,figsize=(15,5))
sns.histplot(df['Weight_in_gms'], ax=ax[0], kde=True).set_title('Weight Distrik
sns.countplot(x = 'Product_importance', data = df, ax=ax[1]).set_title('Product
sns.histplot(df['Cost_of_the_Product'], ax=ax[2], kde=True).set_title('Cost of
```

```
Text(0.5, 1.0, 'Cost of the Product')
```

Weight Distribution

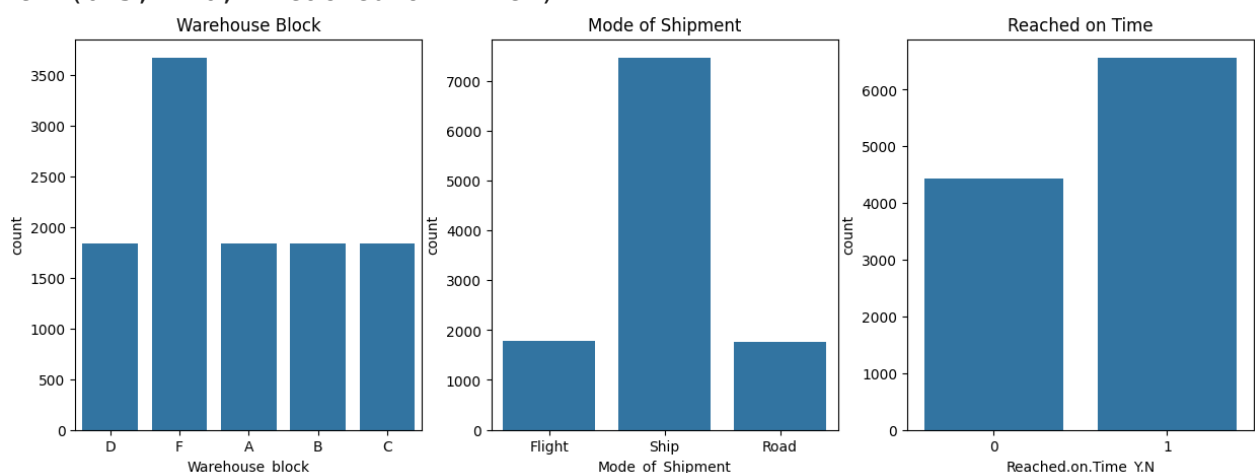
Product Importance

Cost of the Product



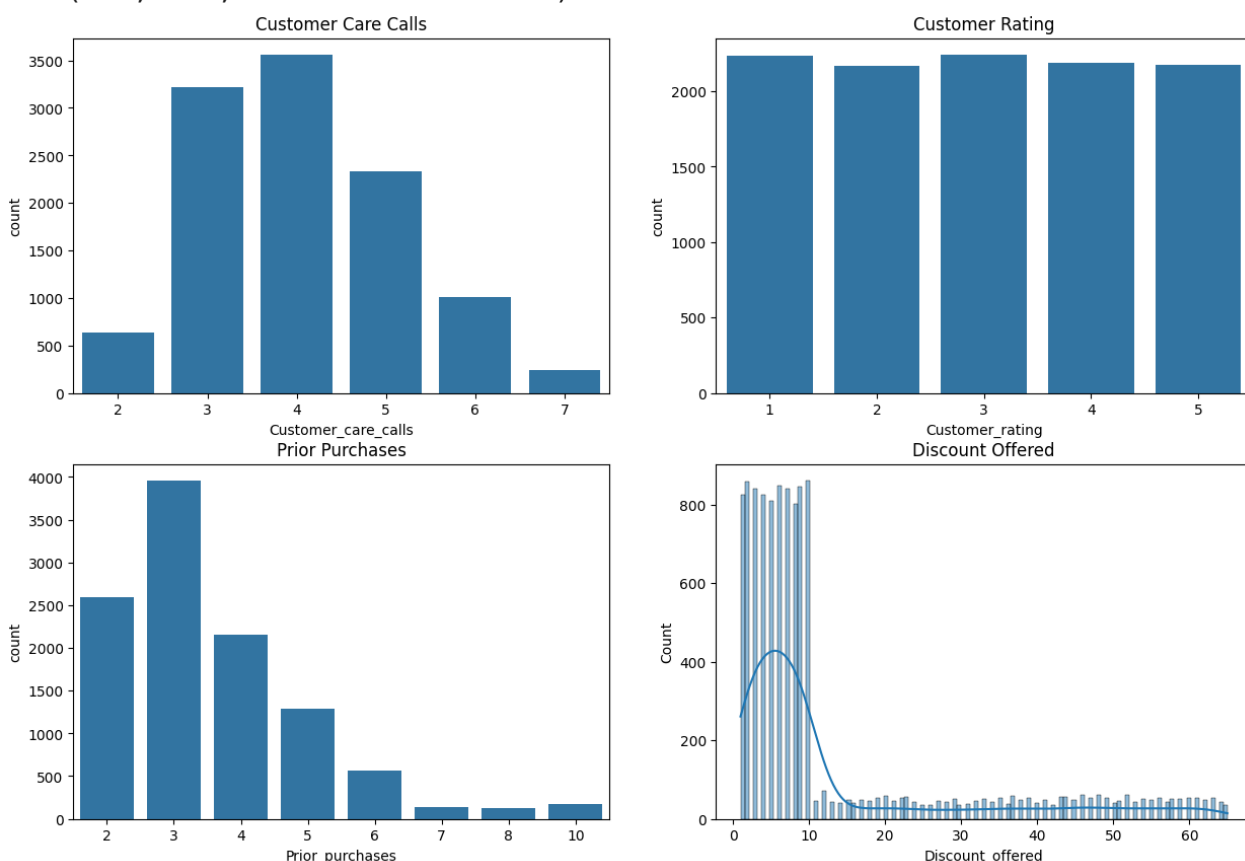
```
fig, ax = plt.subplots(1,3,figsize=(15,5))
sns.countplot(x = 'Warehouse_block', data = df, ax=ax[0]).set_title('Warehouse')
sns.countplot(x = 'Mode_of_Shipment', data = df, ax=ax[1]).set_title('Mode of Shipment')
sns.countplot(x = 'Reached.on.Time_Y.N', data = df, ax=ax[2]).set_title('Reached on Time')
```

Text(0.5, 1.0, 'Reached on Time')

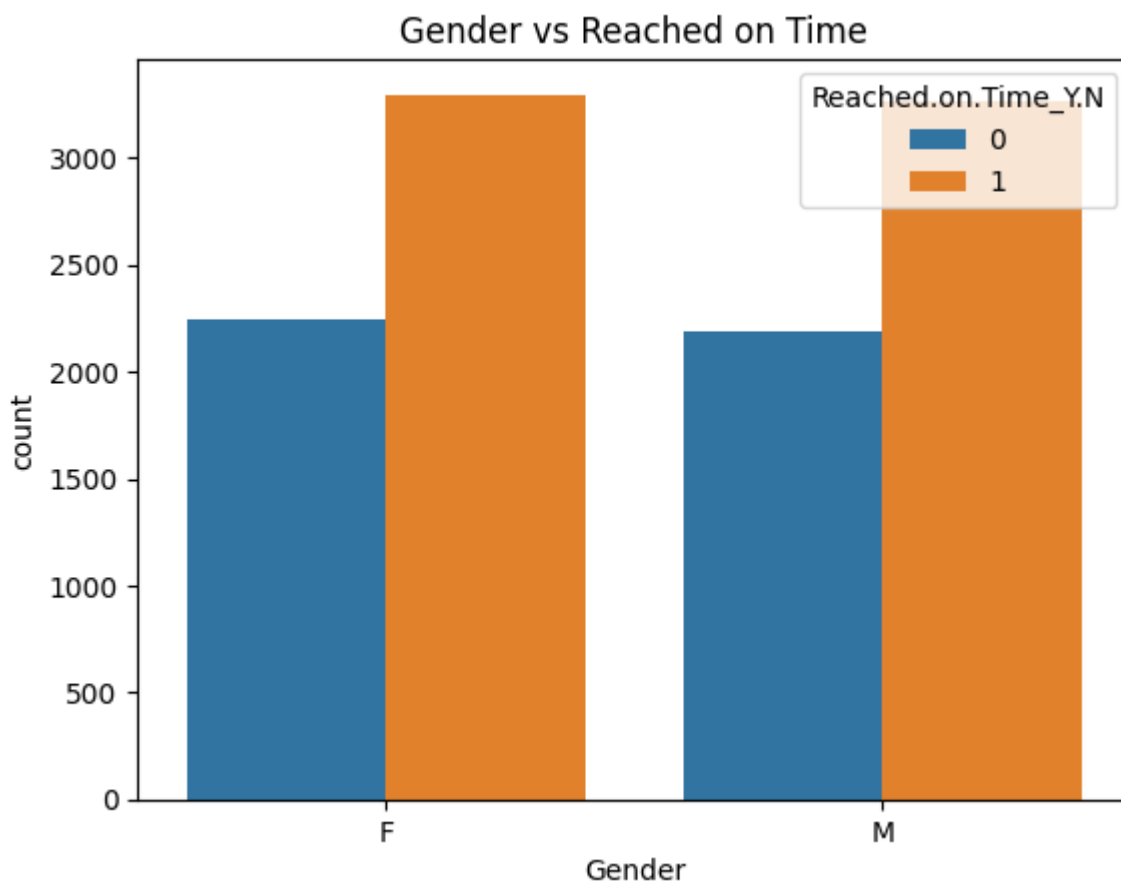


```
fig, ax = plt.subplots(2,2,figsize=(15,10))
sns.countplot(x = 'Customer_care_calls', data = df, ax=ax[0,0]).set_title('Cust
sns.countplot(x = 'Customer_rating', data = df, ax=ax[0,1]).set_title('Customer
sns.countplot(x = 'Prior_purchases', data = df, ax=ax[1,0]).set_title('Prior Pu
sns.histplot(x = 'Discount_offered', data = df, ax=ax[1,1], kde = True).set_tit
```

Text(0.5, 1.0, 'Discount Offered')



```
sns.countplot(x = 'Gender', data = df, hue = 'Reached.on.Time_Y.N').set_title('
    Text(0.5, 1.0, 'Gender vs Reached on Time')
```



```
fig, ax = plt.subplots(1, 3, figsize=(15, 5))
```

```
# Weight Distribution
```

```
sns.violinplot(y=df['Weight_in_gms'], ax=ax[0], inner=None, x=df['Reached.on.Ti
sns.kdeplot(y=df['Weight_in_gms'], ax=ax[0], hue=df['Reached.on.Time_Y.N'], mul
```

```
# Product Importance
```

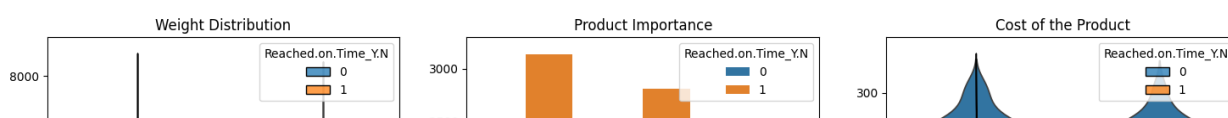
```
sns.countplot(x='Product_importance', data=df, ax=ax[1], hue='Reached.on.Time_Y
```

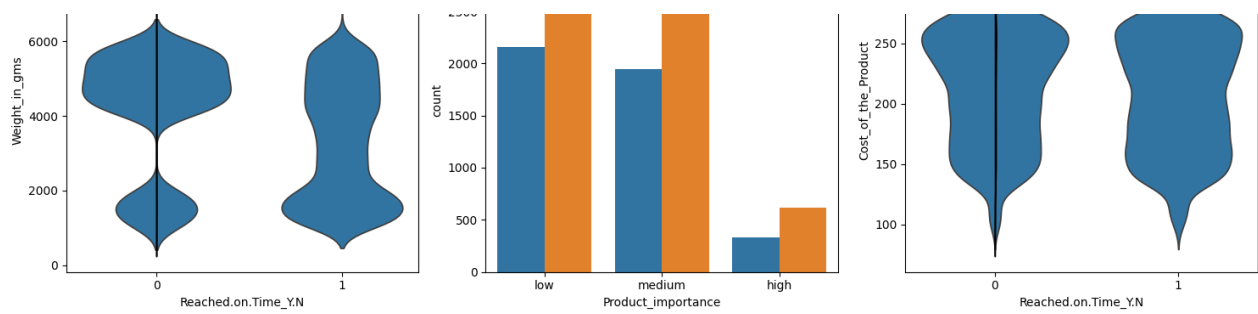
```
# Cost of the Product
```

```
sns.violinplot(y=df['Cost_of_the_Product'], ax=ax[2], inner=None, x=df['Reached
sns.kdeplot(y=df['Cost_of_the_Product'], ax=ax[2], hue=df['Reached.on.Time_Y.N'
```

```
plt.tight_layout()
```

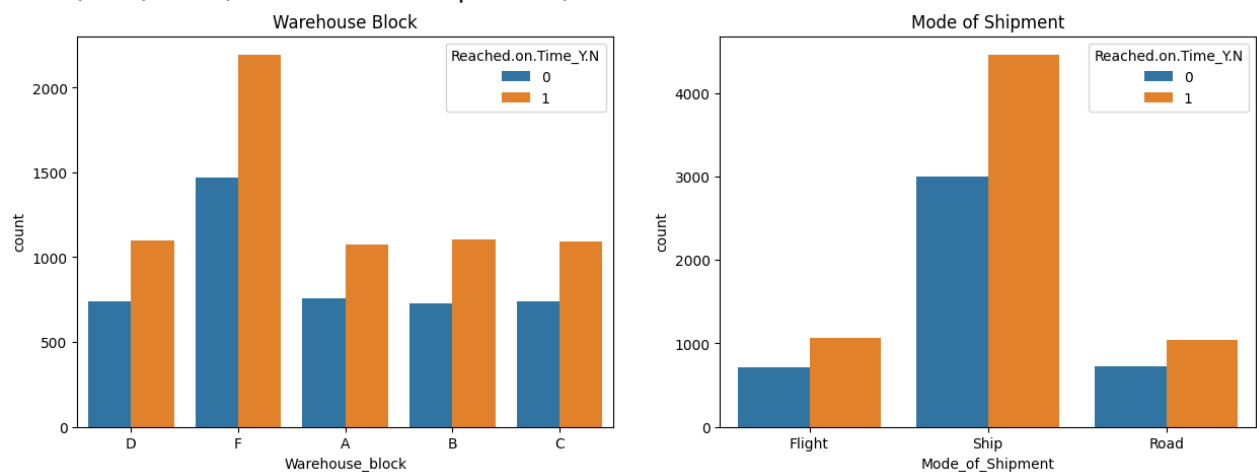
```
plt.show()
```





```
fig, ax = plt.subplots(1,2,figsize=(15,5))
sns.countplot(x = 'Warehouse_block', data = df, ax=ax[0], hue = 'Reached.on.Tin
sns.countplot(x = 'Mode_of_Shipment', data = df, ax=ax[1], hue = 'Reached.on.Ti
```

Text(0.5, 1.0, 'Mode of Shipment')



```
fig, ax = plt.subplots(2,2,figsize=(15,10))
sns.countplot(x = 'Customer_care_calls', data = df, ax=ax[0,0],hue = 'Reached.on.Ti
sns.countplot(x = 'Customer_rating', data = df, ax=ax[0,1],hue = 'Reached.on.Ti
sns.countplot(x = 'Prior_purchases', data = df, ax=ax[1,0],hue = 'Reached.on.Ti
sns.violinplot(x = 'Reached.on.Time_Y.N', y = 'Discount_offered' ,data = df, ax
```

Text(0.5, 1.0, 'Discount Offered')

