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

```
df=pd.read_csv("/content/drive/MyDrive/customer_support_tickets.csv")
```

```
df.head()
```

	Ticket ID	Customer Name	Customer Email	Customer Age	Customer Gender	Product Purchased	Date Purcha
0	1	Marisa Obrien	carrollallison@example.com	32	Other	GoPro Hero	2021-
1	2	Jessica Rios	clarkeashley@example.com	42	Female	LG Smart TV	2021-
2	3	Christopher Robbins	gonzalestracy@example.com	48	Other	Dell XPS	2020-
3	4	Christina Dillon	bradleyolson@example.org	27	Female	Microsoft Office	2020-
4	5	Alexander Carroll	bradleymark@example.com	67	Female	Autodesk AutoCAD	2020-

```
df.shape
```

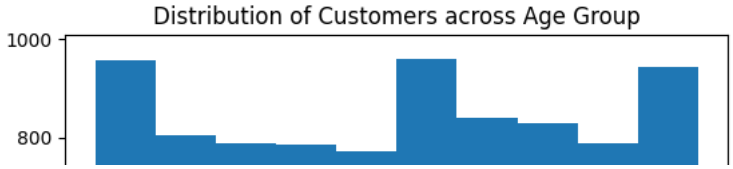
```
(8469, 17)
```

BINNING

```
bin_edges = [0, 18, 35, 50, 100]
bin_labels = ['Teenager', 'Young Adult', 'Adult', 'Senior']
df['Age Category'] = pd.cut(df['Customer Age'], bins=bin_edges, labels=bin_labels)
```

DISTRIBUTION OF CUSTOMERS ACROSS AGE GROUP

```
plt.hist(df['Customer Age'], bins=10)
plt.xlabel('Age Group')
plt.ylabel('Number of Customers')
plt.title('Distribution of Customers across Age Group')
plt.show()
```



CATEGORICAL DATA INTO NUMERICAL

```
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
df['Customer Gender Numerical'] = label_encoder.fit_transform(df['Customer Gender'])
df['Product Purchased Numerical'] = label_encoder.fit_transform(df['Product Purchased'])
```

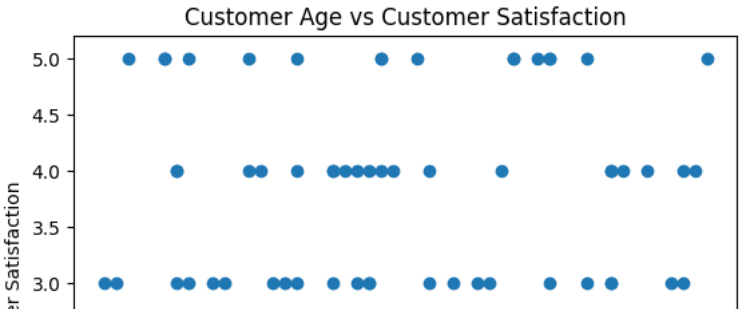
```
df.head()
```

	Ticket ID	Customer Name	Customer Email	Customer Age	Customer Gender	Product Purchased	Da Pur
0	1	Marisa Obrien	carrollallison@example.com	32	Other	GoPro Hero	20
1	2	Jessica Rios	clarkeashley@example.com	42	Female	LG Smart TV	20
2	3	Christopher Robbins	gonzalestracy@example.com	48	Other	Dell XPS	20
3	4	Christina Dillon	bradleyolson@example.org	27	Female	Microsoft Office	20
4	5	Alexander Carroll	bradleymark@example.com	67	Female	Autodesk AutoCAD	20

RELATIONSHIP BETWEEN CUSTOMER AGE AND CUSTOMER SATISFACTION RATING USING THE COVARIANCE MATRIX

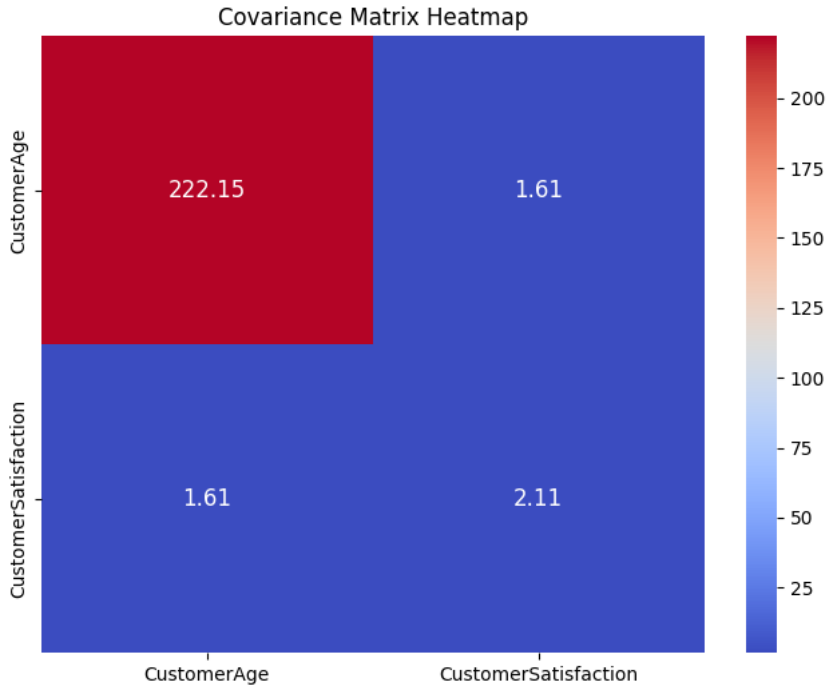
```
np.random.seed(42)
n_samples = 100
customer_age = np.random.randint(18, 70, size=n_samples)
customer_satisfaction = np.random.randint(1, 6, size=n_samples)
data = pd.DataFrame({'CustomerAge': customer_age, 'CustomerSatisfaction': customer_satisfaction})
cov_matrix = np.cov(data['CustomerAge'], data['CustomerSatisfaction'])
print("Covariance Matrix:")
print(cov_matrix)
plt.scatter(data['CustomerAge'], data['CustomerSatisfaction'])
plt.xlabel('Customer Age')
plt.ylabel('Customer Satisfaction')
plt.title('Customer Age vs Customer Satisfaction')
plt.show()
```

```
Covariance Matrix:
[[222.1489899    1.60909091]
 [  1.60909091   2.10868687]]
```



HEATMAP DISTRIBUTION

```
np.random.seed(42)
n_samples = 100
customer_age = np.random.randint(18, 70, size=n_samples)
customer_satisfaction = np.random.randint(1, 6, size=n_samples)
data = pd.DataFrame({'CustomerAge': customer_age, 'CustomerSatisfaction': customer_satisfaction})
cov_matrix = np.cov(data['CustomerAge'], data['CustomerSatisfaction'])
plt.figure(figsize=(8, 6))
sns.heatmap(cov_matrix, annot=True, cmap='coolwarm', fmt=".2f", annot_kws={"size": 12}, xticklabels=data.columns, yticklabels=data.columr
plt.title('Covariance Matrix Heatmap')
plt.show()
```



ASSOCIATION BETWEEN THE CUSTOMER GENDER AND REPEAT PURCHASE USING THE STATISTICAL TEST

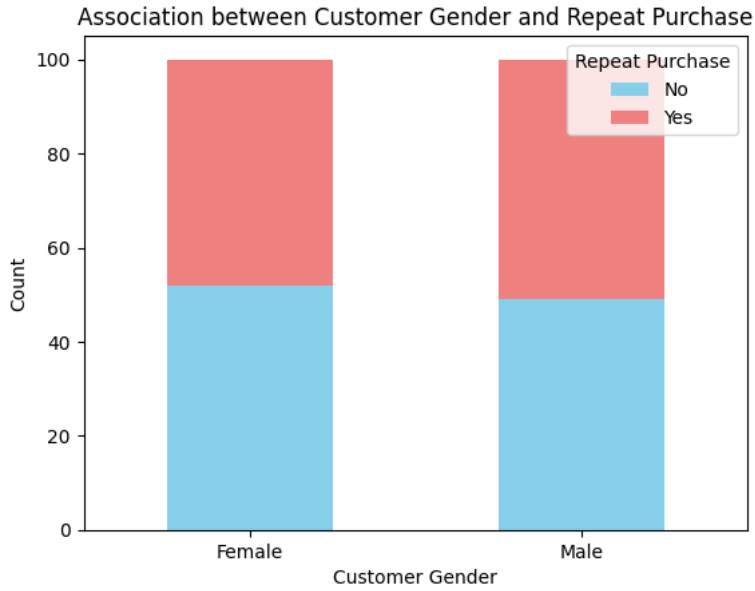
```
from scipy.stats import chi2_contingency
np.random.seed(42)
n_samples = 200
customer_gender = np.random.choice(['Male', 'Female'], size=n_samples)
repeat_purchase = np.random.choice(['Yes', 'No'], size=n_samples)
data = pd.DataFrame({'CustomerGender': customer_gender, 'RepeatPurchase': repeat_purchase})
contingency_table = pd.crosstab(data['CustomerGender'], data['RepeatPurchase'])
chi2, p, dof, expected = chi2_contingency(contingency_table)
print("Chi-Squared:", chi2)
print("p-value:", p)
print("Degrees of Freedom:", dof)
print("Expected Frequencies:")
print(expected)
```

```
Chi-Squared: 0.08000800080008001
p-value: 0.7772865686501679
Degrees of Freedom: 1
Expected Frequencies:
```

```
[[50.5 49.5]
 [50.5 49.5]]
```

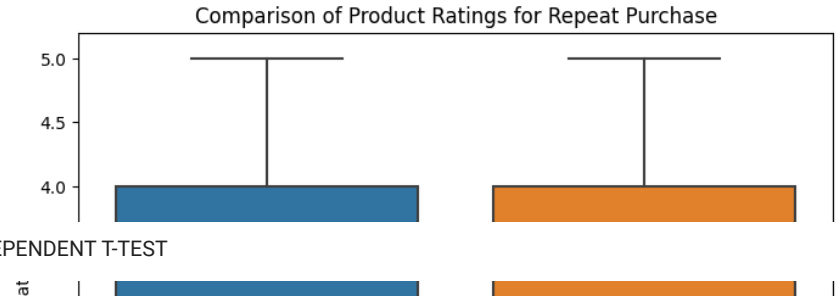
STACKED BAR PLOT

```
np.random.seed(42)
n_samples = 200
customer_gender = np.random.choice(['Male', 'Female'], size=n_samples)
repeat_purchase = np.random.choice(['Yes', 'No'], size=n_samples)
data = pd.DataFrame({'CustomerGender': customer_gender, 'RepeatPurchase': repeat_purchase})
contingency_table = pd.crosstab(data['CustomerGender'], data['RepeatPurchase'])
colors = ['skyblue', 'lightcoral']
contingency_table.plot(kind='bar', stacked=True, color=colors)
plt.title('Association between Customer Gender and Repeat Purchase')
plt.xlabel('Customer Gender')
plt.ylabel('Count')
plt.xticks(rotation=0)
plt.legend(title='Repeat Purchase')
plt.show()
```



COMPARISON OF PRODUCT RATING FOR REPEAT PURCHASE

```
np.random.seed(42)
n_samples = 200
product_rating = np.random.randint(1, 6, size=n_samples)
repeat_purchase = np.random.choice(['Yes', 'No'], size=n_samples)
data = pd.DataFrame({'ProductRating': product_rating, 'RepeatPurchase': repeat_purchase})
plt.figure(figsize=(8, 6))
sns.boxplot(x='RepeatPurchase', y='ProductRating', data=data)
plt.title('Comparison of Product Ratings for Repeat Purchase')
plt.xlabel('Repeat Purchase')
plt.ylabel('Product Rating')
plt.show()
```



INDEPENDENT T-TEST

at

```
from scipy import stats
np.random.seed(42)
n_samples = 200
product_rating = np.random.randint(1, 6, size=n_samples)
repeat_purchase = np.random.choice(['Yes', 'No'], size=n_samples)
data = pd.DataFrame({'ProductRating': product_rating, 'RepeatPurchase': repeat_purchase})
repeat_ratings = data[data['RepeatPurchase'] == 'Yes']['ProductRating']
non_repeat_ratings = data[data['RepeatPurchase'] == 'No']['ProductRating']
t_statistic, p_value = stats.ttest_ind(repeat_ratings, non_repeat_ratings)
print("T-statistic:", t_statistic)
print("p-value:", p_value)
if p_value < 0.05:
    print("There is a significant difference in product ratings between the groups.")
else:
    print("There is no significant difference in product ratings between the groups.")
```

```
T-statistic: 0.0
p-value: 1.0
There is no significant difference in product ratings between the groups.
```