Fundamentals of data science (mini project)

analysis of customer

import packege

In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

LOAD DATA

In [19]:
```
data_frame=pd.read_excel("mpdata.xlsx")
data_frame.head()
```

Out[19]:

	customer id	gender	age	income	score spend
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

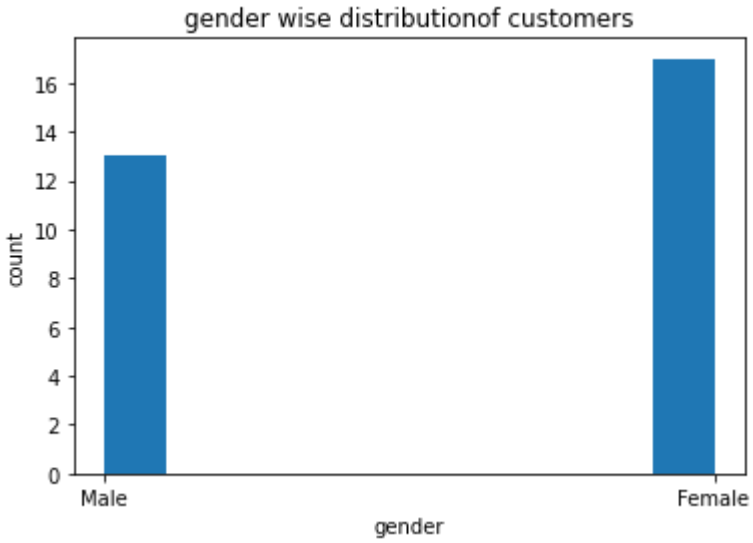
In [20]:
```
data_frame=data_frame[["gender","age","income","score spend"]]
data_frame.head()
```

Out[20]:

	gender	age	income	score spend
0	Male	19	15	39
1	Male	21	15	81
2	Female	20	16	6
3	Female	23	16	77
4	Female	31	17	40

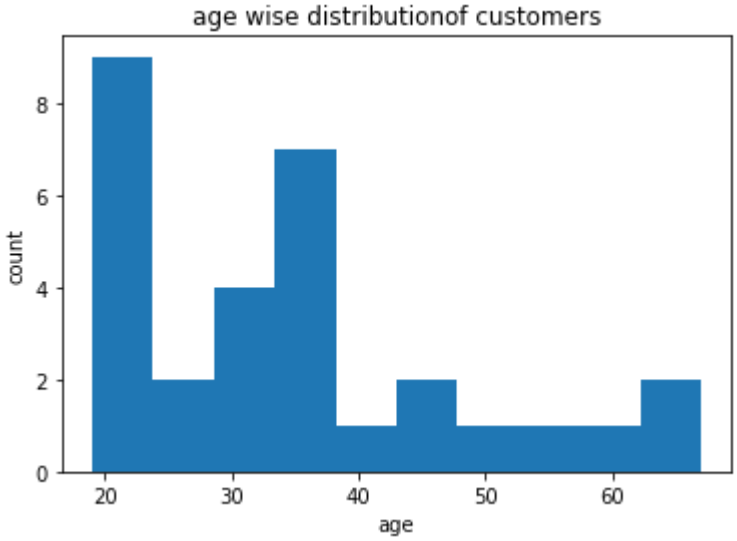
In [22]:
```
ax=plt.subplot()
ax.hist(data_frame["gender"])
ax.set_title("gender wise distributionof customers")
ax.set_xlabel("gender")
ax.set_ylabel("count")
```

Out[22]:
```
Text(0, 0.5, 'count')
```



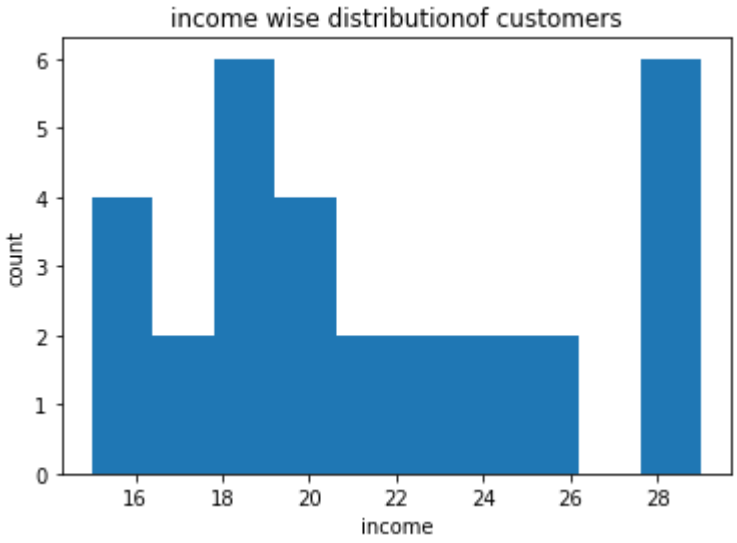
In [23]:
```
ax=plt.subplot()
ax.hist(data_frame["age"])
ax.set_title("age wise distributionof customers")
ax.set_xlabel("age")
ax.set_ylabel("count")
```

Out[23]:
```
Text(0, 0.5, 'count')
```



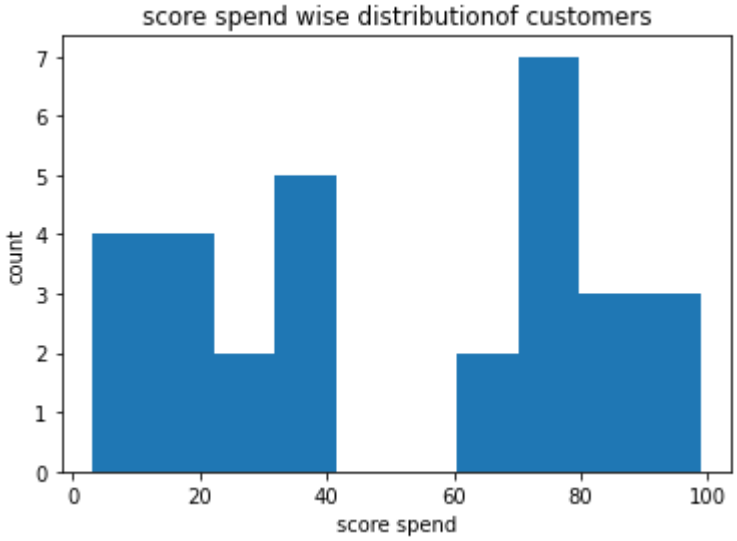
In [24]:
```
ax=plt.subplot()
ax.hist(data_frame["income"])
ax.set_title("income wise distributionof customers")
ax.set_xlabel("income")
ax.set_ylabel("count")
```

Out[24]:
```
Text(0, 0.5, 'count')
```



In [38]:
```
ax=plt.subplot()
ax.hist(data_frame["score spend"])
ax.set_title("score spend wise distributionof customers")
ax.set_xlabel("score spend")
ax.set_ylabel("count")
```

Out[38]:
```
Text(0, 0.5, 'count')
```



In [40]:
```
P_A = (44 + 44) / (44 + 44 + 59 + 53)
P_A_n_B = (44) / (44 + 44 + 59 + 53)
P_B = (53 + 44) / (44 + 44 + 59 + 53)

print("Probability of customer being a male is", P_A)
print("Probability of customer having a spending score greater than 50 is", P_B)
print("Probability of customer being a male and having a spending score greater than 50

# Required probability is: P(A | B) = P(A n B) / P(B)
P_A_given_B = P_A_n_B / P_B
print("Probability of customer being male given that the spending score is greater than
```

```
Probability of customer being a male is 0.44
Probability of customer having a spending score greater than 50 is 0.485
```

```
Probability of customer being a male and having a spending score greater than 50 is 0.22
Probability of customer being male given that the spending score is greater than 50 is
0.4536082474226804
```

In [54]:

```
# Using marginal probability, we gets
P_A = (42 + 33 + 17 + 20) / (42 + 33 + 24 + 27 + 17 + 20 + 20 + 17)
P_B = (17 + 20 + 20 + 17) / (42 + 33 + 24 + 27 + 17 + 20 + 20 + 17)
P_C = (33 + 27 + 20 + 17) / (42 + 33 + 24 + 27 + 17 + 20 + 20 + 17)
P_A_n_B = (17 + 20) / (42 + 33 + 24 + 27 + 17 + 20 + 20 + 17)
P_A_n_B_n_C = 20 / (42 + 33 + 24 + 27 + 17 + 20 + 20 + 17)

print("Probability of a female customer visiting the mall is", P_A)
print("Probability of a customer having an annual income greater than 70 k$ is", P_B)
print("Probability of a customer having a spending score greater than 50 is", P_C)
print("Probability of a female customer having an annual income greater than 70 k$ is",
print("Probability of a female customer having an annual income greater than 70 k$ and
      P_A_n_B_n_C)

# Required probability is P(A n B / C) = P(A n B n C) / P(C)
P_A_n_B_given_C = P_A_n_B_n_C / P_C
print("Probability of a female customer with an annual income greater than 70 k$ given
```

```
Probability of a female customer visiting the mall is 0.56
Probability of a customer having an annual income greater than 70 k$ is 0.37
Probability of a customer having a spending score greater than 50 is 0.485
Probability of a female customer having an annual income greater than 70 k$ is 0.185
Probability of a female customer having an annual income greater than 70 k$ and a spendi
ng score greater than 50 is 0.1
Probability of a female customer with an annual income greater than 70 k$ given that the
spending score is greater than 50 is 0.2061855670103093
```

In [ ]: