```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv('/content/archive.zip')
df
```

	Unnamed: 0	Gender	customer_type	age	type_of_travel	customer_class	flight_distance	inflight_wifi_service	departure_arri
0	0	Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	
1	1	Male	disloyal Customer	25	Business travel	Business	235	3	
2	2	Female	Loyal Customer	26	Business travel	Business	1142	2	
3	3	Female	Loyal Customer	25	Business travel	Business	562	2	
4	4	Male	Loyal Customer	61	Business travel	Business	214	3	
...	...	...	...	...	...	...	...	...	
129875	129875	Male	disloyal Customer	34	Business travel	Business	526	3	
129876	129876	Male	Loyal Customer	23	Business travel	Business	646	4	
129877	129877	Female	Loyal Customer	17	Personal Travel	Eco	828	2	
129878	129878	Male	Loyal Customer	14	Business travel	Business	1127	3	
129879	129879	Female	Loyal Customer	42	Personal Travel	Eco	264	2	

129880 rows × 24 columns

```
df.isna().sum()
```

```
Unnamed: 0                          0
Gender                              0
customer_type                       0
age                                 0
type_of_travel                      0
customer_class                      0
flight_distance                     0
inflight_wifi_service               0
departure_arrival_time_convenient   0
ease_of_online_booking              0
gate_location                       0
food_and_drink                      0
online_boarding                     0
seat_comfort                        0
inflight_entertainment              0
onboard_service                     0
leg_room_service                    0
baggage_handling                    0
checkin_service                     0
inflight_service                    0
cleanliness                         0
departure_delay_in_minutes          0
arrival_delay_in_minutes          393
satisfaction                        0
dtype: int64
```

```
df.dtypes
```

```
Unnamed: 0                          int64
Gender                             object
customer_type                      object
age                                 int64
type_of_travel                     object
customer_class                     object
flight_distance                     int64
inflight_wifi_service               int64
departure_arrival_time_convenient   int64
ease_of_online_booking              int64
gate_location                       int64
food_and_drink                      int64
online_boarding                     int64
seat_comfort                        int64
inflight_entertainment              int64
onboard_service                     int64
```

```
leg_room_service                    int64
baggage_handling                    int64
checkin_service                     int64
inflight_service                    int64
cleanliness                         int64
departure_delay_in_minutes          int64
arrival_delay_in_minutes          float64
satisfaction                       object
dtype: object
```

```
df.head()
```

	Unnamed: 0	Gender	customer_type	age	type_of_travel	customer_class	flight_dista
0	0	Male	Loyal Customer	13	Personal Travel	Eco Plus	
1	1	Male	disloyal Customer	25	Business travel	Business	
2	2	Female	Loyal Customer	26	Business travel	Business	1
3	3	Female	Loyal Customer	25	Business travel	Business	
4	4	Male	Loyal Customer	61	Business travel	Business	

5 rows × 24 columns

```
df.tail()
```

	Unnamed: 0	Gender	customer_type	age	type_of_travel	customer_class	flight_
129875	129875	Male	disloyal Customer	34	Business travel	Business	
129876	129876	Male	Loyal Customer	23	Business travel	Business	
129877	129877	Female	Loyal Customer	17	Personal Travel	Eco	
129878	129878	Male	Loyal Customer	14	Business travel	Business	
129879	129879	Female	Loyal Customer	42	Personal Travel	Eco	

5 rows × 24 columns

```
df['arrival_delay_in_minutes'].fillna(df['arrival_delay_in_minutes'].mode()[0],inplace=True)
df.isna().sum()
```

```
Unnamed: 0                          0
Gender                              0
customer_type                       0
age                                 0
type_of_travel                      0
customer_class                      0
flight_distance                     0
inflight_wifi_service               0
departure_arrival_time_convenient   0
ease_of_online_booking              0
gate_location                       0
food_and_drink                      0
online_boarding                     0
seat_comfort                        0
inflight_entertainment              0
onboard_service                     0
leg_room_service                    0
baggage_handling                    0
checkin_service                     0
inflight_service                    0
cleanliness                         0
departure_delay_in_minutes          0
arrival_delay_in_minutes            0
satisfaction                        0
dtype: int64
```

```
from sklearn.preprocessing import LabelEncoder
lb = LabelEncoder()
for colomn in df.select_dtypes(include='object'):
  df[colomn]=lb.fit_transform(df[colomn])
```

```
df.dtypes
```

```
Unnamed: 0                          int64
Gender                              int64
customer_type                       int64
age                                 int64
type_of_travel                      int64
customer_class                      int64
flight_distance                     int64
inflight_wifi_service               int64
departure_arrival_time_convenient   int64
ease_of_online_booking              int64
gate_location                       int64
food_and_drink                      int64
online_boarding                     int64
seat_comfort                        int64
inflight_entertainment              int64
onboard_service                     int64
leg_room_service                    int64
baggage_handling                    int64
checkin_service                     int64
inflight_service                    int64
cleanliness                         int64
departure_delay_in_minutes          int64
arrival_delay_in_minutes          float64
satisfaction                        int64
dtype: object
```

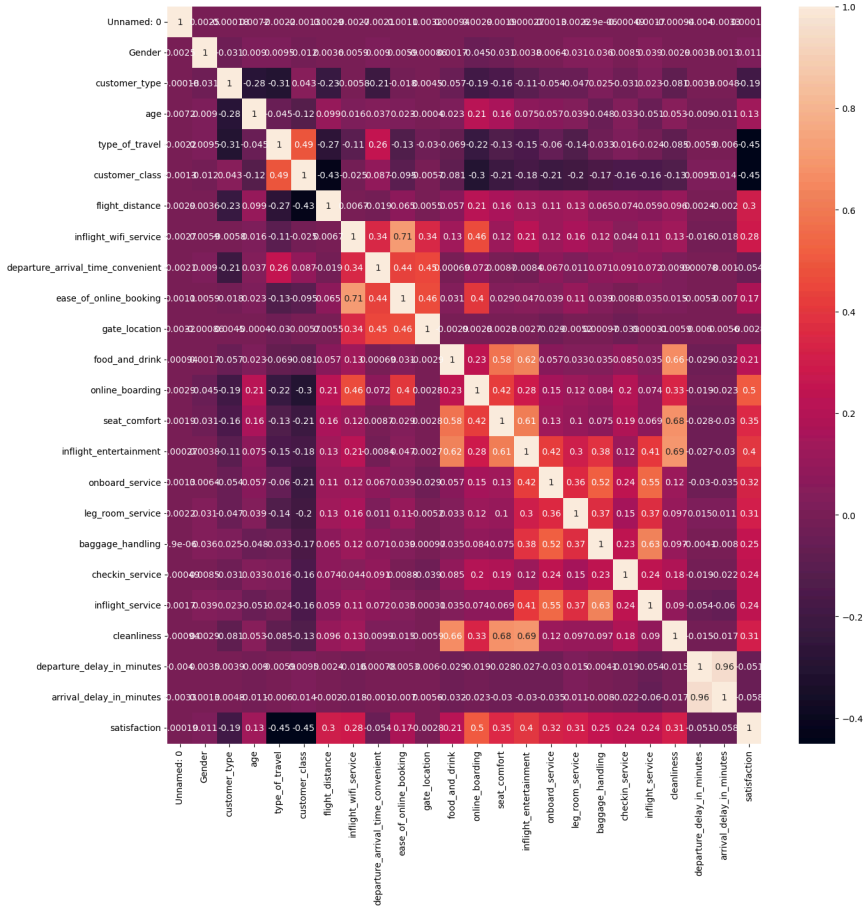
```
corr_matrix = df.corr()
corr_matrix
```

	Unnamed: 0	Gender	customer_type	age	type_of
Unnamed: 0	1.000000	0.002538	-0.000180	0.007166	-(
Gender	0.002538	1.000000	-0.030958	0.008996	(
customer_type	-0.000180	-0.030958	1.000000	-0.284172	-(
age	0.007166	0.008996	-0.284172	1.000000	-(
type_of_travel	-0.002249	0.009503	-0.308236	-0.044808	'
customer_class	-0.001307	-0.011574	0.042994	-0.116921	(
flight_distance	0.002895	0.003616	-0.226021	0.099459	-(
inflight_wifi_service	-0.002745	0.005901	-0.005757	0.016116	-(
departure_arrival_time_convenient	-0.002134	0.008995	-0.206873	0.036960	(
ease_of_online_booking	0.001116	0.005893	-0.018059	0.022565	-(
gate_location	0.003236	-0.000863	0.004465	-0.000398	-(
food_and_drink	0.000937	0.001730	-0.056997	0.023194	-(
online_boarding	0.002947	-0.045022	-0.189083	0.207572	-(
seat_comfort	0.001868	-0.030756	-0.156239	0.159136	-(
inflight_entertainment	0.000269	0.003843	-0.106001	0.074947	-(
onboard_service	0.001279	0.006447	-0.054172	0.057078	-(
leg_room_service	0.002237	0.031047	-0.046841	0.039119	-(
baggage_handling	0.000007	0.036356	0.024874	-0.047991	-(
checkin_service	-0.000489	0.008462	-0.031243	0.033475	(
inflight_service	0.001669	0.038504	0.023292	-0.051347	-(
cleanliness	-0.000944	0.002867	-0.081302	0.052565	-(
departure_delay_in_minutes	-0.003972	0.003491	0.003859	-0.009041	-(
arrival_delay_in_minutes	-0.003327	0.001286	0.004769	-0.011206	-(
satisfaction	0.000192	0.011236	-0.186017	0.134091	-(

24 rows × 24 columns

```
plt.figure(figsize=(15,15))
sns.heatmap(df.corr(),annot=True)
```

<Axes: >




```
df.drop(['Unnamed: 0'],axis=1,inplace=True)
df
```

	Gender	customer_type	age	type_of_travel	customer_class	flight_distance
0	1	0	13	1	2	460
1	1	1	25	0	0	235
2	0	0	26	0	0	1142

```
x = df.iloc[:,:-1].values
y = df.iloc[:,-1].values
x
```

```
array([[ 1.,  0., 13., ...,  5., 25., 18.],
       [ 1.,  1., 25., ...,  1.,  1.,  6.],
       [ 0.,  0., 26., ...,  5.,  0.,  0.],
       ...,
       [ 0.,  0., 17., ...,  2.,  0.,  0.],
       [ 1.,  0., 14., ...,  4.,  0.,  0.],
       [ 0.,  0., 42., ...,  1.,  0.,  0.]])
```

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30,random_state=42)
```

```
from sklearn.preprocessing import StandardScaler
scaler=StandardScaler()
scaler.fit(x_train)
x_train=scaler.transform(x_train)
x_test=scaler.transform(x_test)
```

```
from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier(n_neighbors=5)
knn.fit(x_train,y_train)
y_pred=knn.predict(x_test)
y_pred
```

```
array([0, 1, 0, ..., 0, 0, 0])
```

```
from sklearn.metrics import accuracy_score
score=accuracy_score(y_test,y_pred)
score
```

```
0.9287804126886356
```