In [3]:
```
import pandas as pd
import numpy as np
from sklearn.preprocessing import StandardScaler,LabelEncoder,OneHotEncode
from sklearn.metrics import accuracy_score,classification_report,confusion
from sklearn.ensemble import RandomForestClassifier,GradientBoostingClassi
from sklearn.model_selection import train_test_split
from sklearn.decomposition import IncrementalPCA
from tqdm import tqdm
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score
sns.set_theme(color_codes=True)
```

In [4]:
```
train_data=pd.read_csv(r"C:\Users\Simi\Downloads\credit.csv")
```

In [5]:
```
test_data=pd.read_csv(r"C:\Users\Simi\Downloads\credittest.csv")
```

In [6]:
```
df=pd.concat([train_data,test_data],axis=0)
```

In [7]:
```
print(df)
```

```
        Unnamed: 0 trans_date_trans_time        cc_num  \
0                0      01-01-2019 00:00  2.703190e+15
1                1      01-01-2019 00:00  6.304230e+11
2                2      01-01-2019 00:00  3.885950e+13
3                3      01-01-2019 00:01  3.534090e+15
4                4      01-01-2019 00:03  3.755340e+14
...            ...                   ...           ...
555714      555714      31-12-2020 23:59  3.056060e+13
555715      555715      31-12-2020 23:59  3.556610e+15
555716      555716      31-12-2020 23:59  6.011720e+15
555717      555717      31-12-2020 23:59  4.079770e+12
555718      555718      31-12-2020 23:59  4.170690e+15

                                   merchant        category     amt      f
irst  \
0                fraud_Rippin, Kub and Mann        misc_net    4.97   Jenn
ifer
1           fraud_Heller, Gutmann and Zieme     grocery_pos  107.23  Steph
anie
2                      fraud_Lind-Buckridge   entertainment  220.11     Ed
ward
3        fraud_Kutch, Hermiston and Farrell   gas_transport   45.00     Je
remy
4                       fraud_Keeling-Crist        misc_pos   41.96      T
yler
...                                     ...             ...     ...
...
555714                fraud_Reilly and Sons  health_fitness   43.77    Mic
hael
555715                 fraud_Hoppe-Parisian       kids_pets  111.84
Jose
555716                      fraud_Rau-Robel       kids_pets   86.88
Ann
555717                fraud_Breitenberg LLC          travel    7.99
Eric
555718                    fraud_Dare-Marvin   entertainment   38.13     Sa
muel

           last gender                        street  ...      lat      l
ong  \
0         Banks      F                561 Perry Cove  ...  36.0788  -81.1
781
1          Gill      F  43039 Riley Greens Suite 393  ...  48.8878 -118.2
105
2       Sanchez      M      594 White Dale Suite 530  ...  42.1808 -112.2
620
3         White      M   9443 Cynthia Court Apt. 038  ...  46.2306 -112.1
138
4        Garcia      M              408 Bradley Rest  ...  38.4207  -79.4
629
...         ...    ...                           ...  ...      ...
...
555714    Olson      M           558 Michael Estates  ...  40.4931  -91.8
912
555715  Vasquez      M           572 Davis Mountains  ...  29.0393  -95.4
401
555716   Lawson      F    144 Evans Islands Apt. 683  ...  46.1966 -118.9
017
555717  Preston      M    7020 Doyle Stream Apt. 951  ...  44.6255 -116.4
493
555718     Frey      M     830 Myers Plaza Apt. 384  ...  35.6665  -97.4
```

```
798

        city_pop                                job        dob  \
0           3495        Psychologist, counselling  09-03-1988
1            149  Special educational needs teacher  21-06-1978
2           4154        Nature conservation officer  19-01-1962
3           1939                    Patent attorney  12-01-1967
4             99     Dance movement psychotherapist  28-03-1986
...          ...                                ...        ...
555714       519                       Town planner  13-02-1966
555715     28739                     Futures trader  27-12-1999
555716      3684                           Musician  29-11-1981
555717       129                       Cartographer  15-12-1965
555718    116001                        Media buyer  10-05-1993

                               trans_num   unix_time  merch_lat  merch_lo
ng  \
0       0b242abb623afc578575680df30655b9  1325376018  36.011293  -82.0483
15
1       1f76529f8574734946361c461b024d99  1325376044  49.159047 -118.1864
62
2       a1a22d70485983eac12b5b88dad1cf95  1325376051  43.150704 -112.1544
81
3       6b849c168bdad6f867558c3793159a81  1325376076  47.034331 -112.5610
71
4       a41d7549acf90789359a9aa5346dcb46  1325376186  38.674999  -78.6324
59
...                                  ...         ...        ...
...
555714  9b1f753c79894c9f4b71f04581835ada  1388534347  39.946837  -91.3333
31
555715  2090647dac2c89a1d86c514c427f5b91  1388534349  29.661049  -96.1866
33
555716  6c5b7c8add471975aa0fec023b2e8408  1388534355  46.658340 -119.7150
54
555717  14392d723bb7737606b2700ac791b7aa  1388534364  44.470525 -117.0808
88
555718  1765bb45b3aa3224b4cdcb6e7a96cee3  1388534374  36.210097  -97.0363
72

        is_fraud
0              0
1              0
2              0
3              0
4              0
...          ...
555714         0
555715         0
555716         0
555717         0
555718         0

[1604294 rows x 23 columns]
```

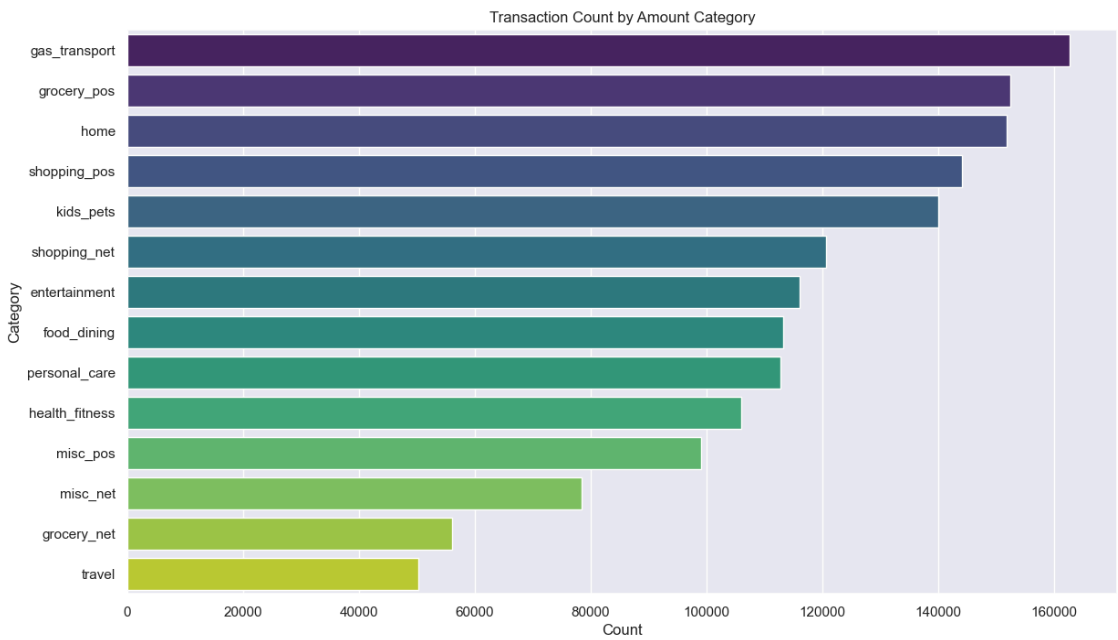
In [8]:
```
df.columns
```

Out[8]:
```
Index(['Unnamed: 0', 'trans_date_trans_time', 'cc_num', 'merchant', 'cate
gory',
       'amt', 'first', 'last', 'gender', 'street', 'city', 'state', 'zi
p',
       'lat', 'long', 'city_pop', 'job', 'dob', 'trans_num', 'unix_time',
       'merch_lat', 'merch_long', 'is_fraud'],
      dtype='object')
```

In [9]:
```
def segment_by_amount(row):
    if row['amt']>100:
        return 'High value'
    elif row['amt']>50:
        return 'Medium Value'
    else:
        return 'Low Value'
    combined_data['AmountSegment']=df.apply(segment_by_amount,axis=1)
    print(df.head())
```

In [10]:
```

plt.figure(figsize=(14, 8))
sns.countplot(y='category', data=df, palette='viridis' ,order=df['category
plt.title('Transaction Count by Amount Category')
plt.xlabel('Count')
plt.ylabel('Category')
plt.show()
```



In [11]:
```
print("Number of columns: {}".format(df.shape[1]))
print("Number of rows: {}".format(df.shape[0]))
```

```
Number of columns: 23
Number of rows: 1604294
```

In [13]:
```
df.isnull().sum()
```

Out[13]:
```
Unnamed: 0               0
trans_date_trans_time    0
cc_num                   0
merchant                 0
category                 0
amt                      0
first                    0
last                     0
gender                   0
street                   0
city                     0
state                    0
zip                      0
lat                      0
long                     0
city_pop                 0
job                      0
dob                      0
trans_num                0
unix_time                0
merch_lat                0
merch_long               0
is_fraud                 0
dtype: int64
```

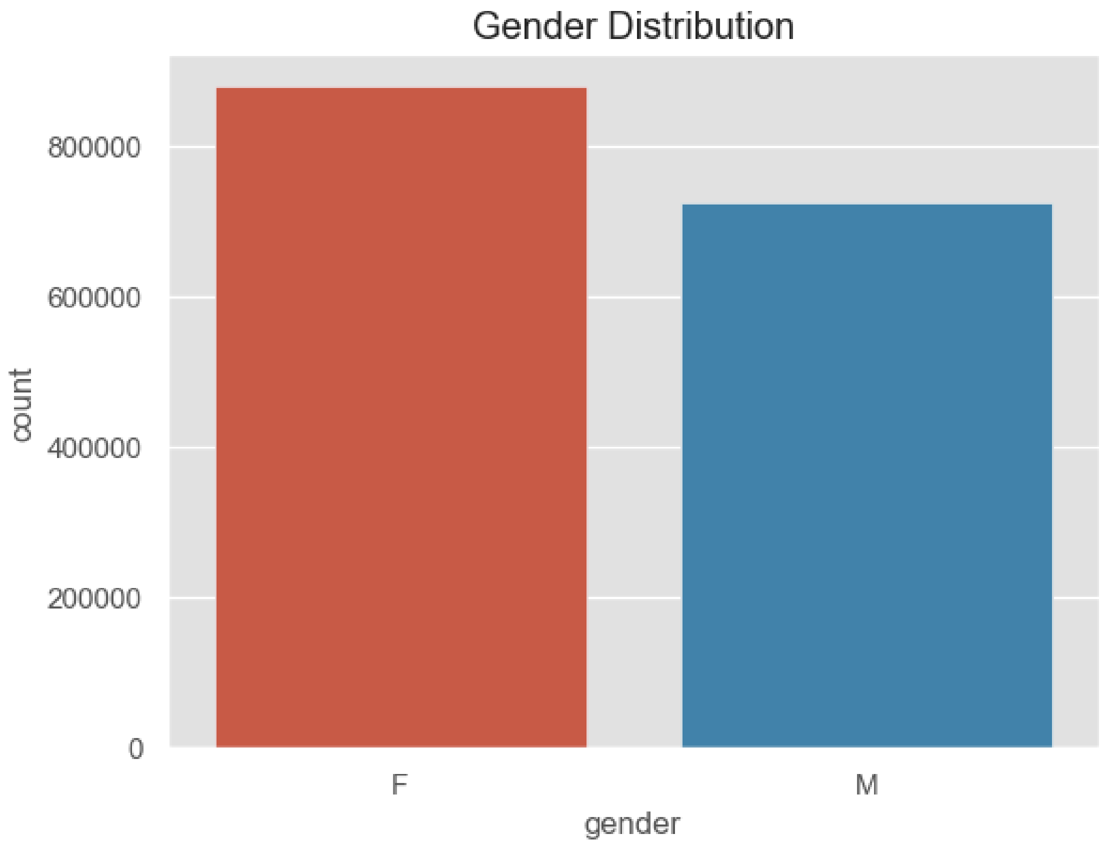
In [14]:
```
sc = StandardScaler()
df['amt'] = sc.fit_transform(pd.DataFrame(df['amt']))
```

In [16]:
```
df.duplicated().any()
```

Out[16]:
```
False
```

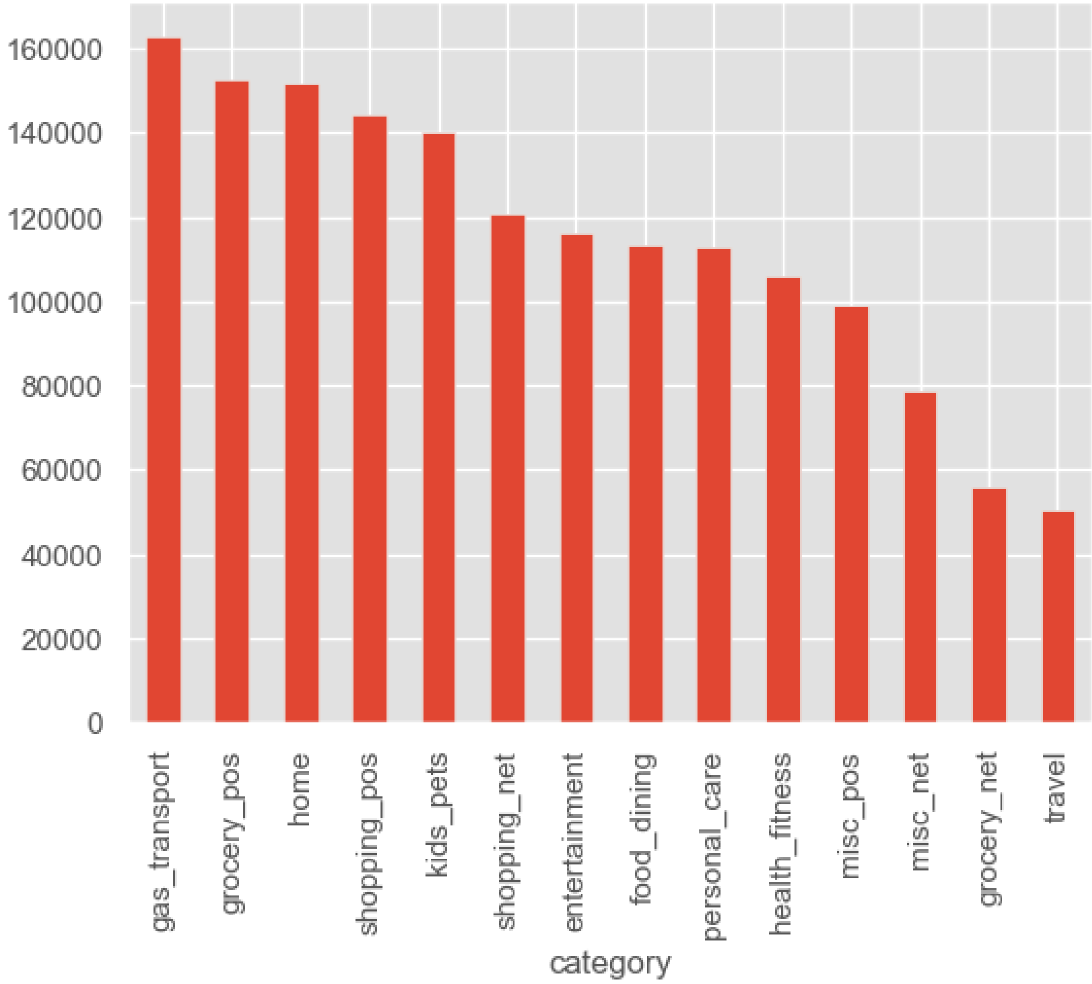
In [17]:
```
df = df.drop_duplicates()
```

In [92]:
```
sns.countplot(x='gender', data=df)
plt.title("Gender Distribution")
plt.show()
```



In [49]:
```
df.category.value_counts().plot.bar()
```

Out[49]: <Axes: xlabel='category'>



In [82]:
```
x.shape
```

Out[82]: (1604294, 7)

In [83]:
```
y.shape
```

Out[83]: (1604294,)

In [84]:
```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2,
```

In [85]:
```
 model=LogisticRegression()
```

In [86]:
```
model.fit(x_train,y_train)
```

Out[86]:
```
▾ LogisticRegression
LogisticRegression()
```

In [89]:
```
y_pred = model.predict(x_test)
```

In [90]:
```
accuracy = accuracy_score(y_test,y_pred)
print("accuracy :", accuracy)
```

```
accuracy : 0.9948388544500856
```

In [ ]: