In [1]:
```
#IMPORTING THE NECESSARY LIBRARIES
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
sns.set_style('darkgrid')
```

In [2]:
```
train = pd.read_csv(r'C:\\Users\\MY PC\\Downloads\\Trains.csv')
train.head()
```

Out[2]:

	VehicleID	Location	Maker	Model	Year	Colour	Amount (Million Naira)	Type	Distance
0	VHL12546	Abuja	Honda	Accord Coupe EX V-6	2,011	Silver	2.2	Nigerian Used	NaN
1	VHL18827	Ibadan	Hyundai	Sonata	2,012	Silver	3.5	Nigerian Used	125,000
2	VHL19499	Lagos	Lexus	RX 350	2,010	Red	9.2	Foreign Used	110,852
3	VHL17991	Abuja	Mercedes-Benz	GLE-Class	2,017	Blue	22.8	Foreign Used	30,000
4	VHL12170	Ibadan	Toyota	Highlander	2,002	Red	2.6	Nigerian Used	125,206

In [3]:
```
test = pd.read_csv(r'C:\\Users\\MY PC\\Downloads\\Tests.csv')
test.head()
```

Out[3]:

	VehicleID	Location	Maker	Model	Year	Colour	Type	Distance
0	VHL18518	Abuja	BMW	323i	2,008	White	Foreign Used	30524.0
1	VHL17149	Lagos	Toyota	Camry	2,013	White	Foreign Used	NaN
2	VHL10927	Lagos	Toyota	Highlander Limited V6	2,005	Gold	Foreign Used	NaN
3	VHL12909	Lagos	Toyota	Camry	2,011	Gray	Foreign Used	166839.0
4	VHL12348	Lagos	Lexus	ES 350 FWD	2,013	Red	Foreign Used	88862.0

In [4]:
```
1 train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7205 entries, 0 to 7204
Data columns (total 9 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   VehicleID               7205 non-null   object 
 1   Location                7205 non-null   object 
 2   Maker                   7205 non-null   object 
 3   Model                   7205 non-null   object 
 4   Year                    7184 non-null   object 
 5   Colour                  7205 non-null   object 
 6   Amount (Million Naira)  7188 non-null   float64
 7   Type                    7008 non-null   object 
 8   Distance                4845 non-null   object 
dtypes: float64(1), object(8)
memory usage: 506.7+ KB
```

In [5]:
```
1 test.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2061 entries, 0 to 2060
Data columns (total 8 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   VehicleID  2061 non-null   object 
 1   Location   2061 non-null   object 
 2   Maker      2061 non-null   object 
 3   Model      2061 non-null   object 
 4   Year       2059 non-null   object 
 5   Colour     2061 non-null   object 
 6   Type       2007 non-null   object 
 7   Distance   1385 non-null   float64
dtypes: float64(1), object(7)
memory usage: 128.9+ KB
```

In [6]:
```
1 train.isnull().sum()
```

Out[6]:
```
VehicleID                  0
Location                   0
Maker                      0
Model                      0
Year                      21
Colour                     0
Amount (Million Naira)    17
Type                     197
Distance                2360
dtype: int64
```

In [7]:
```
1 train['Year']=train['Year'].str.replace(',','')
2 train.head()
3
```

Out[7]:

	VehicleID	Location	Maker	Model	Year	Colour	Amount (Million Naira)	Type	Distance
0	VHL12546	Abuja	Honda	Accord Coupe EX V-6	2011	Silver	2.2	Nigerian Used	NaN
1	VHL18827	Ibadan	Hyundai	Sonata	2012	Silver	3.5	Nigerian Used	125,000
2	VHL19499	Lagos	Lexus	RX 350	2010	Red	9.2	Foreign Used	110,852
3	VHL17991	Abuja	Mercedes-Benz	GLE-Class	2017	Blue	22.8	Foreign Used	30,000
4	VHL12170	Ibadan	Toyota	Highlander	2002	Red	2.6	Nigerian Used	125,206

```
1
```

In [8]:
```
test['Year'] = test['Year'].str.replace(',','')
test.head()
```

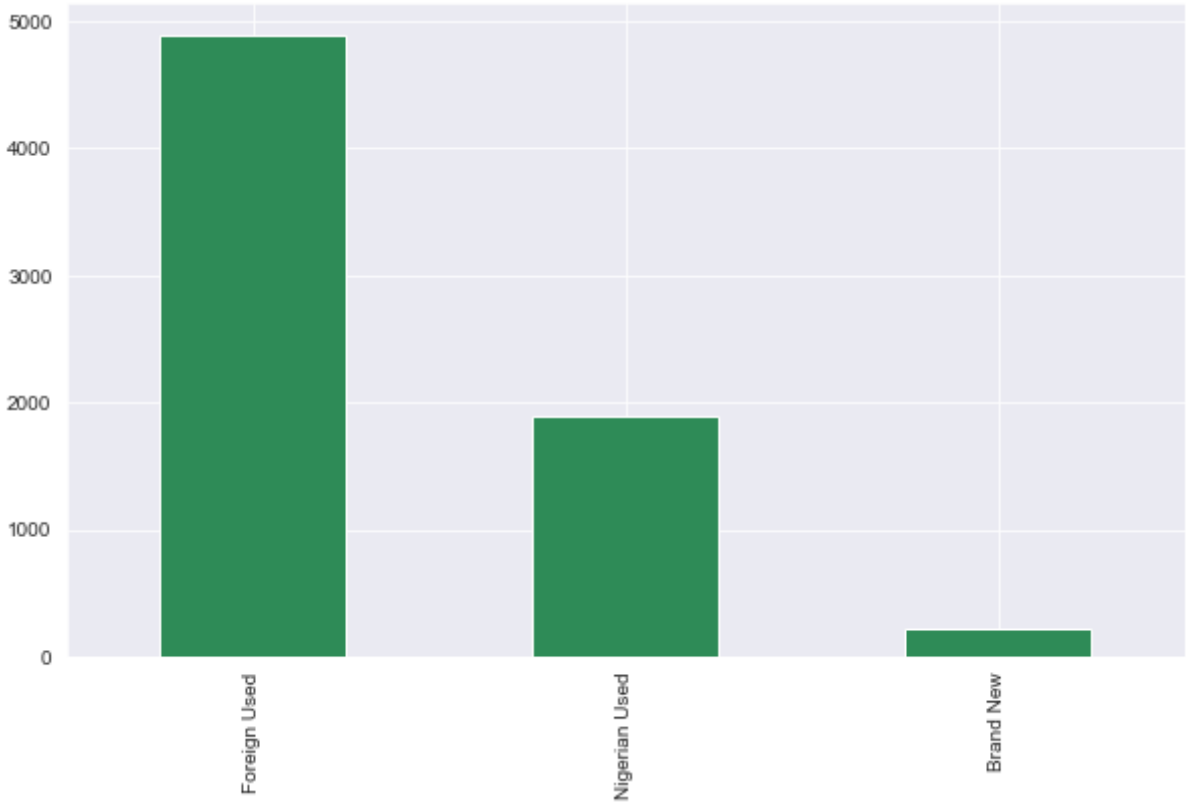
Out[8]:

	VehicleID	Location	Maker	Model	Year	Colour	Type	Distance
0	VHL18518	Abuja	BMW	323i	2008	White	Foreign Used	30524.0
1	VHL17149	Lagos	Toyota	Camry	2013	White	Foreign Used	NaN
2	VHL10927	Lagos	Toyota	Highlander Limited V6	2005	Gold	Foreign Used	NaN
3	VHL12909	Lagos	Toyota	Camry	2011	Gray	Foreign Used	166839.0
4	VHL12348	Lagos	Lexus	ES 350 FWD	2013	Red	Foreign Used	88862.0

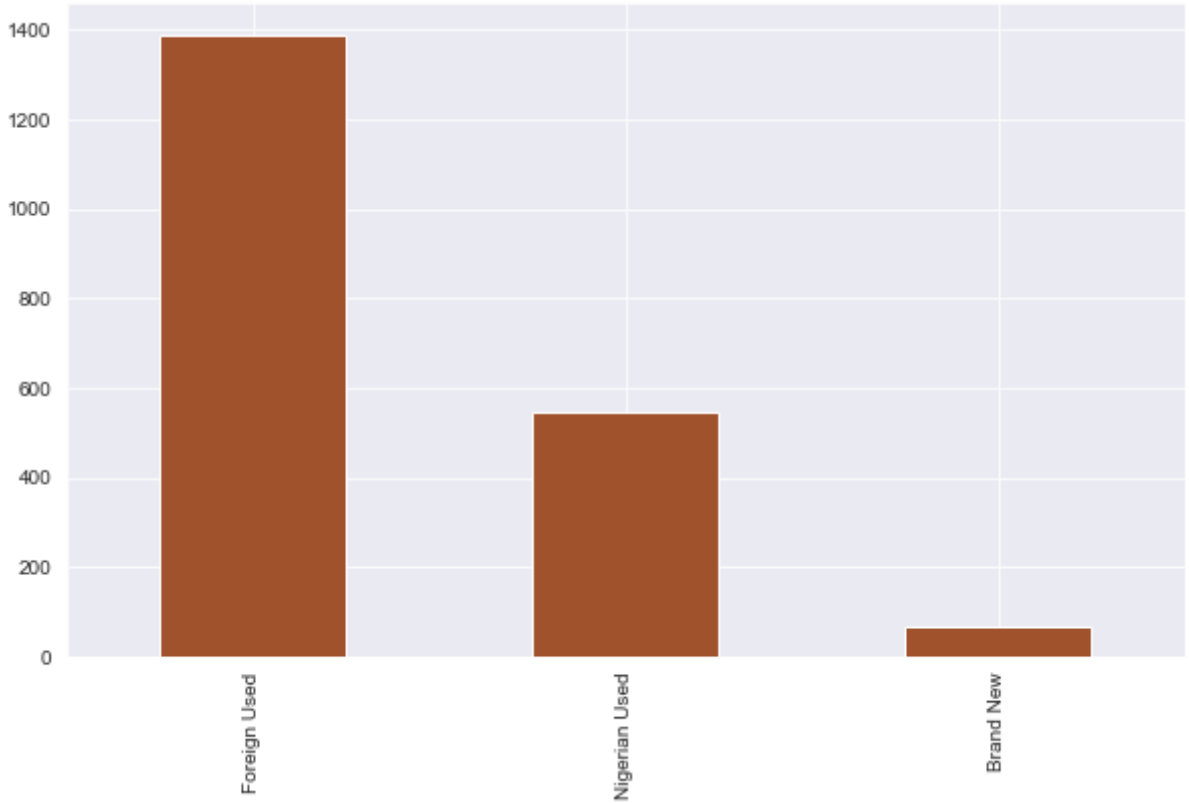
In [9]:
```
test.isnull().sum()
```

Out[9]:
```
VehicleID      0
Location       0
Maker          0
Model          0
Year           2
Colour         0
Type          54
Distance     676
dtype: int64
```

In [10]:
```
plt.figure(figsize=(10,6))
train['Type'].value_counts().plot(kind='bar',color='seagreen');
```



In [11]:
```
1 plt.figure(figsize=(10,6))
2 test['Type'].value_counts().plot(kind='bar',color='sienna');
```



In [12]:
```
1 train['Maker'].value_counts(normalize=True)*100
```

Out[12]:
```
Toyota           38.056905
Lexus            22.192922
Mercedes-Benz    16.835531
Honda             4.968772
Hyundai           2.248439
Acura             2.137405
Land Rover        1.721027
Ford              1.665510
BMW               1.540597
Nissan            1.401804
Peugeot           1.040944
Kia               0.916031
Volkswagen        0.749480
Pontiac           0.416378
Mazda             0.388619
Dodge             0.333102
Audi              0.319223
Mitsubishi        0.319223
Chevrolet         0.291464
Infiniti          0.249827
Jeep              0.249827
Jaguar            0.166551
Rolls-Royce       0.138793
Mini              0.124913
GMC               0.111034
Suzuki            0.111034
Cadillac          0.097155
Scion             0.097155
Porsche           0.097155
Volvo             0.097155
Bentley           0.083276
Maserati          0.083276
Lincoln           0.083276
Buick             0.069396
Chrysler          0.055517
Lamborghini       0.055517
Opel              0.055517
Rover             0.041638
GAC               0.041638
```

```
Renault           0.041638
Fiat              0.041638
Citroen           0.027759
Subaru            0.027759
Saturn            0.027759
JAC               0.027759
Hummer            0.027759
Skoda             0.013879
Saab              0.013879
IVM               0.013879
King              0.013879
MG                0.013879
Tata              0.013879
BAW               0.013879
Ferrari           0.013879
Brabus            0.013879
Name: Maker, dtype: float64
```

test['Maker']

In [13]:
```
1 test['Maker'].value_counts(normalize=True)
```

Out[13]:
```
Toyota           0.395924
Lexus            0.212033
Mercedes-Benz    0.156720
Honda            0.056283
Hyundai          0.021834
BMW              0.019893
Land Rover       0.018923
Acura            0.017952
Ford             0.015526
Nissan           0.011645
Volkswagen       0.010674
Peugeot          0.009704
Mazda            0.007763
Kia              0.006308
Pontiac          0.005822
Audi             0.004852
Chevrolet        0.003396
Mitsubishi       0.002911
Dodge            0.002911
Lincoln          0.002426
Infiniti         0.002426
Jaguar           0.001941
Mini             0.001456
Opel             0.001456
Volvo            0.000970
Chrysler         0.000970
Scion            0.000970
Lamborghini      0.000485
Jeep             0.000485
Rolls-Royce      0.000485
Buick            0.000485
Cadillac         0.000485
Subaru           0.000485
Fiat             0.000485
Maserati         0.000485
Renault          0.000485
GMC              0.000485
Porsche          0.000485
Seat             0.000485
```

```
Brabus           0.000485
Name: Maker, dtype: float64
```

In [14]:
```
train['Distance']= train['Distance'].str.replace(',','')
train.head()
```

Out[14]:

	VehicleID	Location	Maker	Model	Year	Colour	Amount (Million Naira)	Type	Distance
0	VHL12546	Abuja	Honda	Accord Coupe EX V-6	2011	Silver	2.2	Nigerian Used	NaN
1	VHL18827	Ibadan	Hyundai	Sonata	2012	Silver	3.5	Nigerian Used	125000
2	VHL19499	Lagos	Lexus	RX 350	2010	Red	9.2	Foreign Used	110852
3	VHL17991	Abuja	Mercedes-Benz	GLE-Class	2017	Blue	22.8	Foreign Used	30000
4	VHL12170	Ibadan	Toyota	Highlander	2002	Red	2.6	Nigerian Used	125206

In [15]:
```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7205 entries, 0 to 7204
Data columns (total 9 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   VehicleID               7205 non-null   object 
 1   Location                7205 non-null   object 
 2   Maker                   7205 non-null   object 
 3   Model                   7205 non-null   object 
 4   Year                    7184 non-null   object 
 5   Colour                  7205 non-null   object 
 6   Amount (Million Naira)  7188 non-null   float64
 7   Type                    7008 non-null   object 
 8   Distance                4845 non-null   object 
dtypes: float64(1), object(8)
memory usage: 506.7+ KB
```

In [16]:
```
train['Distance']= train['Distance'].astype(float)
```

In [17]:
```
train['Distance'].mean()
```

Out[17]:
```
103198.90361197111
```

In [18]:
```
train['Distance'].median()
```

Out[18]:
```
80830.0
```

In [19]:
```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7205 entries, 0 to 7204
Data columns (total 9 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   VehicleID               7205 non-null   object 
 1   Location                7205 non-null   object 
 2   Maker                   7205 non-null   object 
 3   Model                   7205 non-null   object 
 4   Year                    7184 non-null   object 
 5   Colour                  7205 non-null   object 
 6   Amount (Million Naira)  7188 non-null   float64
 7   Type                    7008 non-null   object 
 8   Distance                4845 non-null   float64
dtypes: float64(2), object(7)
memory usage: 506.7+ KB
```

In [20]:
```
train['Distance'].fillna(train['Distance'].median(),inplace=True)
```

In [21]:
```
test['Distance'].fillna(test['Distance'].median(),inplace=True)
```

In [22]:
```
train['Year'] = train['Year'].str.replace(',','')
train.Year.head()
```

Out[22]:
```
0    2011
1    2012
2    2010
3    2017
4    2002
Name: Year, dtype: object
```

In [23]:
```
train['Year']= train['Year'].astype(float).head()
```

In [24]:
```
test['Year'] = test['Year'].astype(float)
```

In [25]:
```
Avg_val= train['Year'].median()
```

In [26]:
```
train['Year'].fillna(Avg_val,inplace=True)
```

In [27]:
```
test['Year'].fillna(test['Year'].median(),inplace=True)
```

In [28]:
```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7205 entries, 0 to 7204
Data columns (total 9 columns):
 #   Column                 Non-Null Count  Dtype  
---  ------                 --------------  -----  
 0   VehicleID              7205 non-null   object 
 1   Location               7205 non-null   object 
 2   Maker                  7205 non-null   object 
 3   Model                  7205 non-null   object 
 4   Year                   7205 non-null   float64
 5   Colour                 7205 non-null   object 
 6   Amount (Million Naira)  7188 non-null   float64
 7   Type                   7008 non-null   object 
 8   Distance               7205 non-null   float64
dtypes: float64(3), object(6)
memory usage: 506.7+ KB
```

In [29]:
```
1 train['Amount (Million Naira)']
```

Out[29]:
```
0        2.20
1        3.50
2        9.20
3       22.80
4        2.60
        ...
7200     5.70
7201     4.00
7202     2.85
7203     8.65
7204     3.38
Name: Amount (Million Naira), Length: 7205, dtype: float64
```

In [30]:
```
1 train['Amount (Million Naira)'].fillna(train['Amount (Million Naira)'].median(),inplace=True)
2 train['Amount (Million Naira)'].isnull().sum()
```

Out[30]:
```
0
```

In [31]:
```
1 train['Type'].fillna(method='bfill',inplace=True)
```

In [32]:
```
1 test['Type'].fillna(method= 'bfill',inplace=True)
```

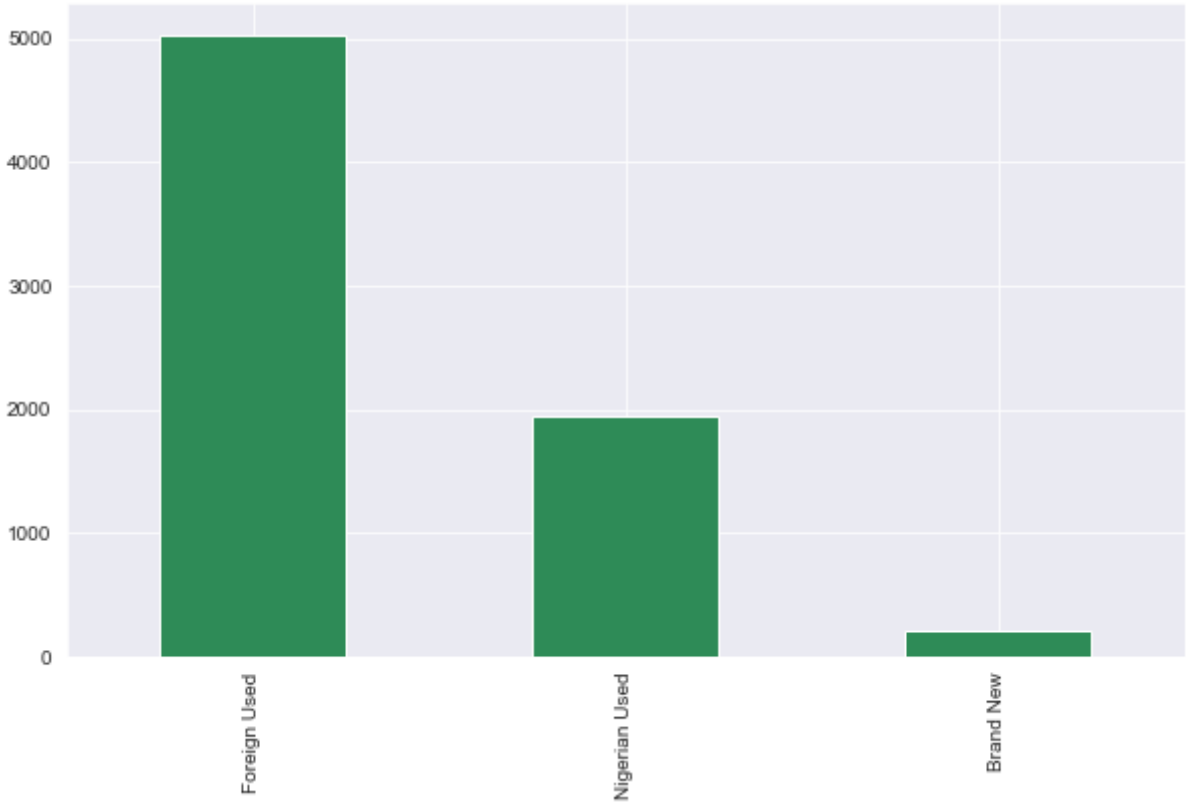
In [33]:
```
1 train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7205 entries, 0 to 7204
Data columns (total 9 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   VehicleID               7205 non-null   object 
 1   Location                7205 non-null   object 
 2   Maker                   7205 non-null   object 
 3   Model                   7205 non-null   object 
 4   Year                    7205 non-null   float64
 5   Colour                  7205 non-null   object 
 6   Amount (Million Naira)  7205 non-null   float64
 7   Type                    7205 non-null   object 
 8   Distance                7205 non-null   float64
dtypes: float64(3), object(6)
memory usage: 506.7+ KB
```

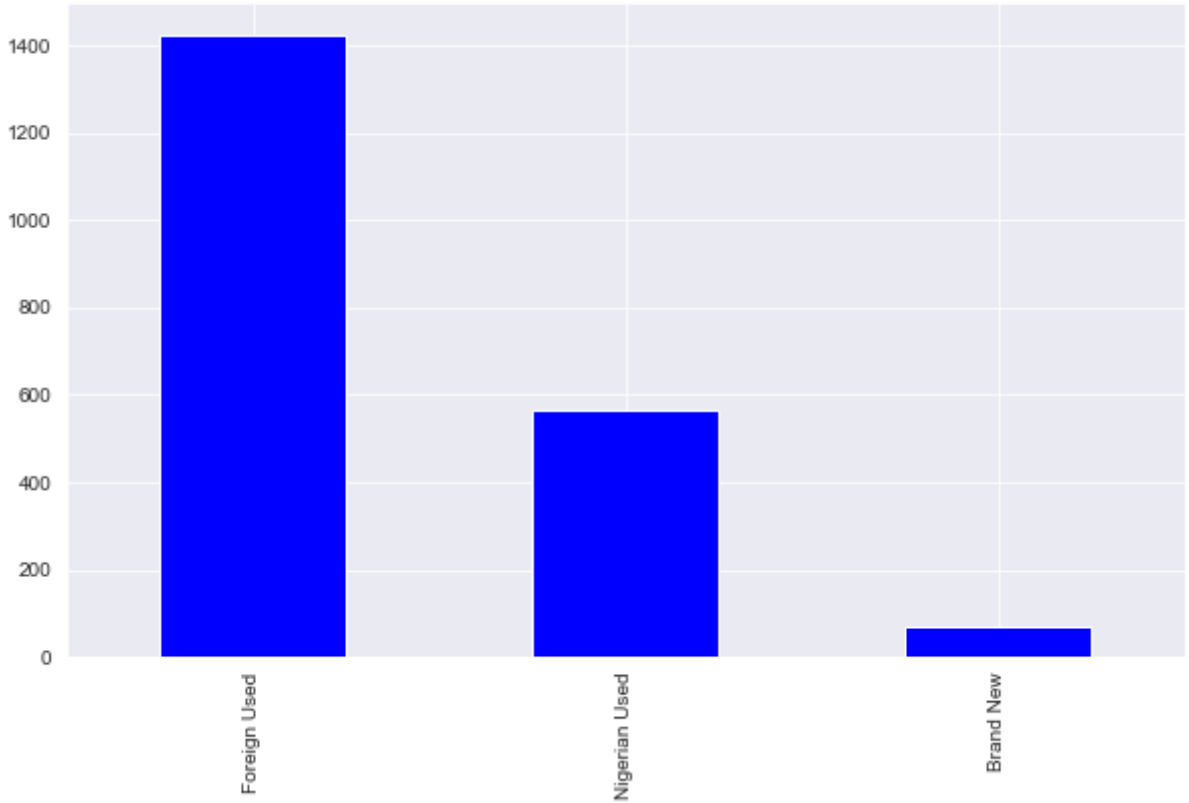
In [34]:
```
1 test.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2061 entries, 0 to 2060
Data columns (total 8 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   VehicleID  2061 non-null   object 
 1   Location   2061 non-null   object 
 2   Maker      2061 non-null   object 
 3   Model      2061 non-null   object 
 4   Year       2061 non-null   float64
 5   Colour     2061 non-null   object 
 6   Type       2061 non-null   object 
 7   Distance   2061 non-null   float64
dtypes: float64(2), object(6)
memory usage: 128.9+ KB
```

In [35]:
```
plt.figure(figsize=(10,6))
train['Type'].value_counts().plot(kind='bar',color='seagreen');
```



In [36]:
```
plt.figure(figsize=(10,6))
test['Type'].value_counts().plot(kind='bar',color='blue');
```



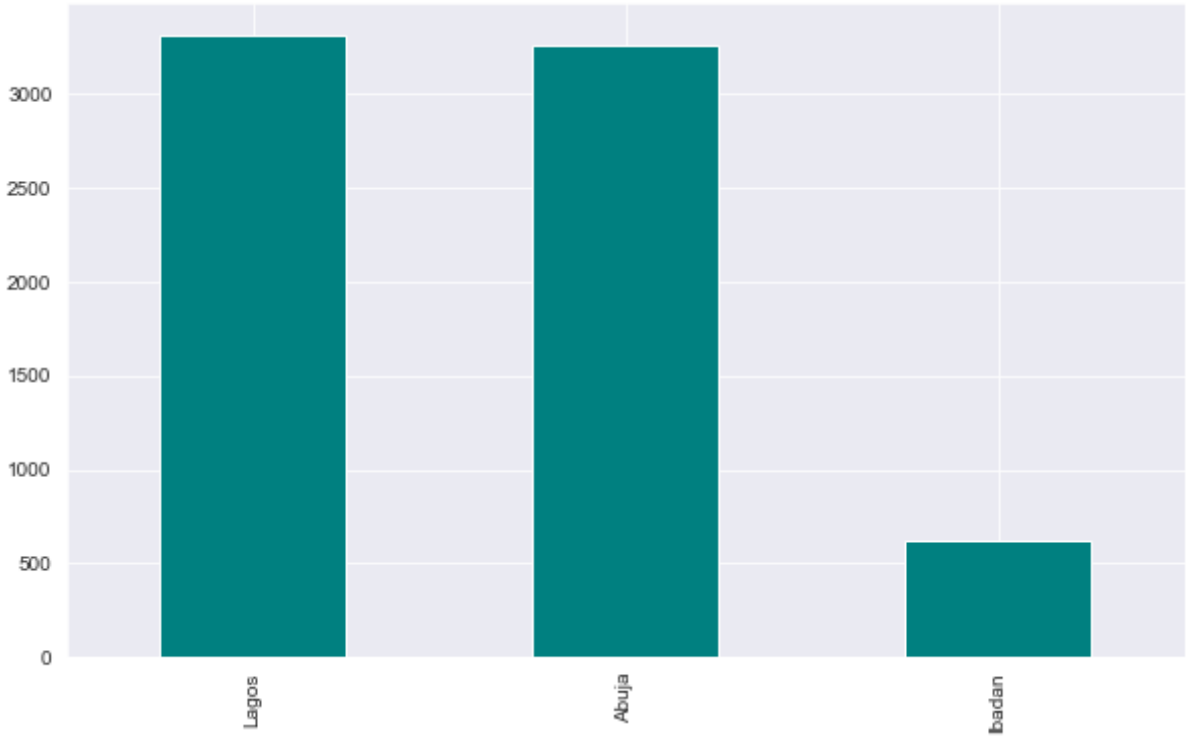
In [37]:
```
train['Location'].nunique()
```

Out[37]: 3

In [38]:
```
test['Location'].nunique()
```

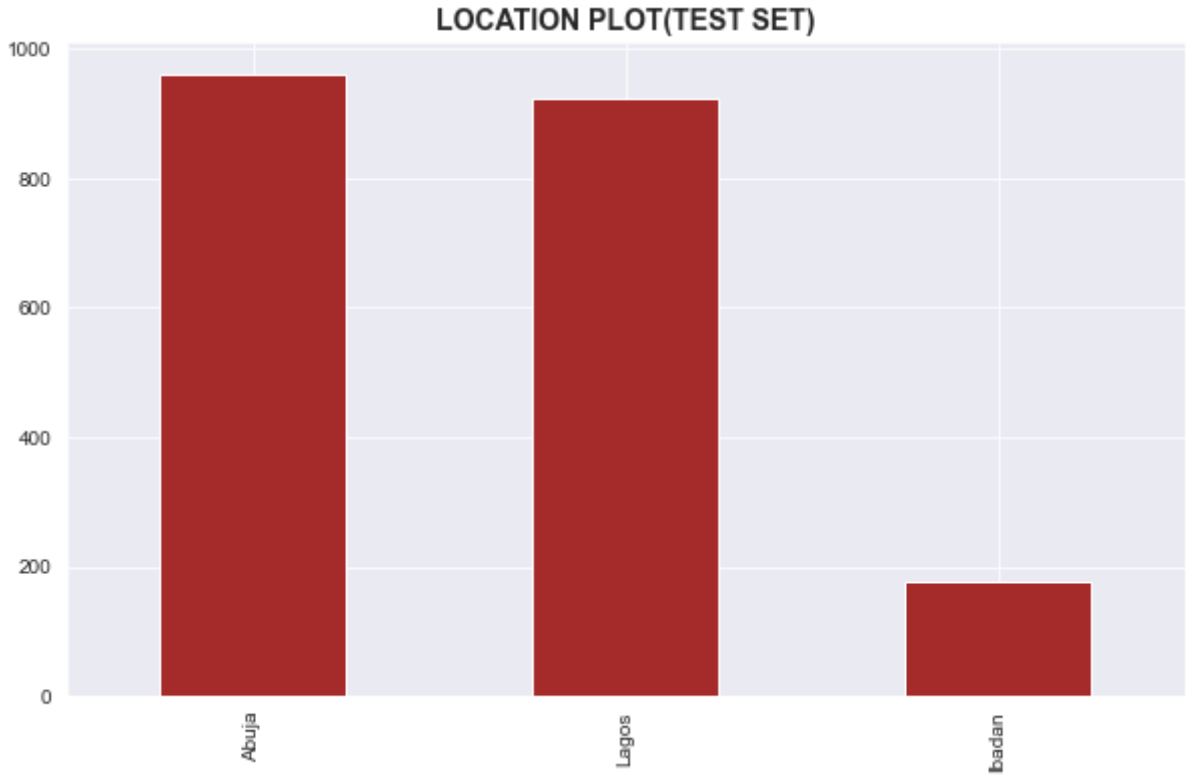
Out[38]: 3

In [39]:
```
plt.figure(figsize=(10,6))
train['Location'].value_counts().plot(kind='bar',color='teal');
```

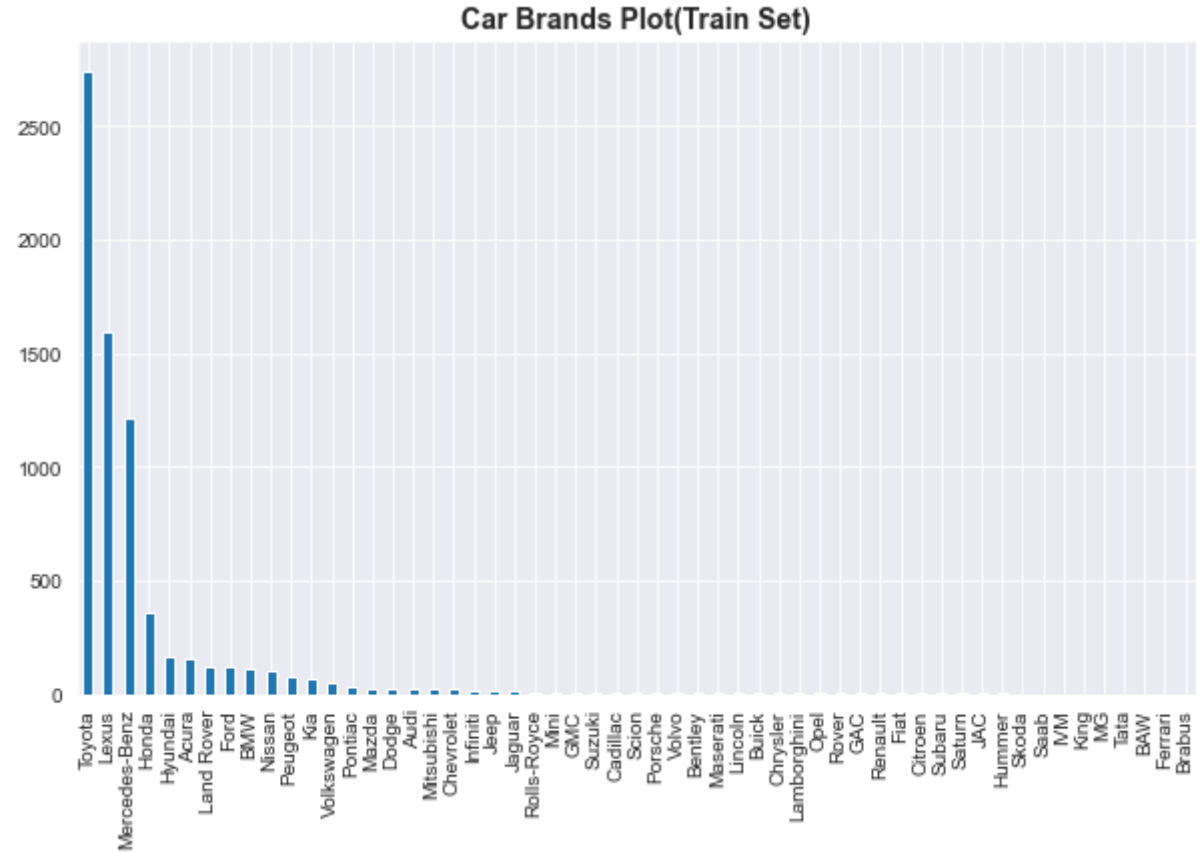


In [40]:
```
plt.figure(figsize=(10,6))
test['Location'].value_counts().plot(kind='bar',color='brown');
plt.title('LOCATION PLOT(TEST SET)',fontweight='bold',fontsize=14)
```

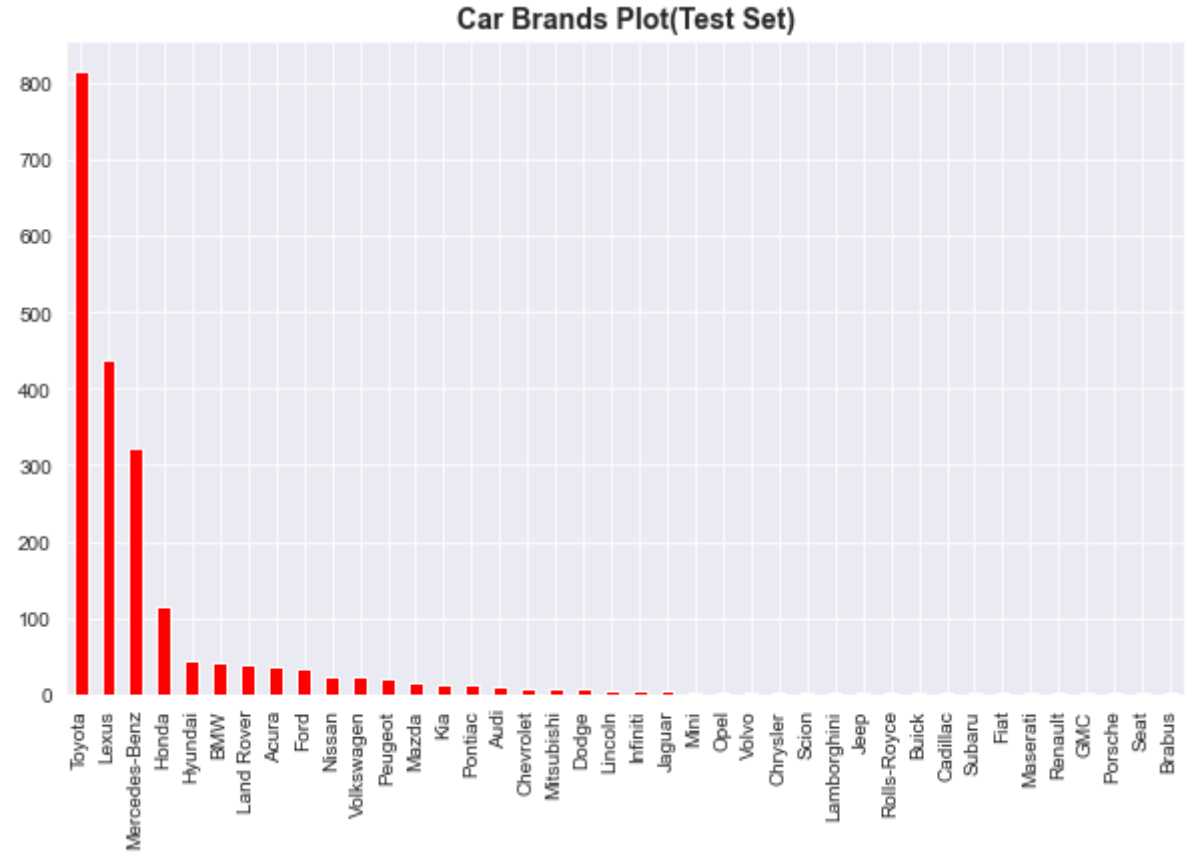
Out[40]: Text(0.5, 1.0, 'LOCATION PLOT(TEST SET)')



In [41]:
```
plt.figure(figsize=(10,6))
train['Maker'].value_counts().plot(kind='bar');
plt.title('Car Brands Plot(Train Set)',fontweight='bold',fontsize=14);
```



In [42]:
```
plt.figure(figsize=(10,6))
test['Maker'].value_counts().plot(kind='bar',color='red');
plt.title('Car Brands Plot(Test Set)',fontweight='bold',fontsize=14);
```



In [43]:
```
1 train['Model'].value_counts(normalize=True)*100
```

Out[43]:
```
Camry                    9.035392
ES 350                   4.163775
Corolla                  3.913949
C300                     2.761971
RX                       2.192922
                           ...
Accent 1.6               0.013879
Land Cruiser Prado EXR   0.013879
Lx                       0.013879
CLK                      0.013879
320i SV Premium          0.013879
Name: Model, Length: 1223, dtype: float64
```

In [44]:
```
1 test['Model'].value_counts(normalize=True)*100
```

Out[44]:
```
Camry                      8.879185
ES 350                     4.415332
Corolla                    3.590490
C300                       2.668607
RX                         2.280446
                             ...
XF Premium                 0.048520
Land Cruiser Prado GX      0.048520
Accord EX V6 Automatic     0.048520
Tacoma Access Cab I4 AWD   0.048520
Torrent                    0.048520
Name: Model, Length: 587, dtype: float64
```

In [45]:
```
1 train.describe()
```

Out[45]:

	Year	Amount (Million Naira)	Distance
count	7205.000000	7205.000000	7.205000e+03
mean	2010.999584	11.833375	9.587196e+04
std	0.128524	25.290819	9.756484e+04
min	2002.000000	0.450000	1.000000e+00
25%	2011.000000	3.500000	6.700000e+04
50%	2011.000000	5.650000	8.083000e+04
75%	2011.000000	11.500000	9.557200e+04
max	2017.000000	456.000000	1.985400e+06

In [46]:
```
1 test.describe()
```

Out[46]:

	Year	Distance
count	2061.000000	2061.000000
mean	2011.077147	96650.133916
std	4.964864	87473.957491
min	1982.000000	1.000000
25%	2008.000000	67415.000000
50%	2011.000000	82000.000000
75%	2014.000000	97000.000000
max	2022.000000	985216.000000

In [47]:
```
1 car_grp = train.groupby(['Maker'])
2 car_grp
```

Out[47]:
```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000026D237EA1C0>
```

In [48]:
```
1 car_grp.get_group('Lexus')
```

Out[48]:

	VehicleID	Location	Maker	Model	Year	Colour	Amount (Million Naira)	Type	Distance
2	VHL19499	Lagos	Lexus	RX 350	2010.0	Red	9.20	Foreign Used	110852.0
6	VHL16314	Lagos	Lexus	LX 570 AWD	2011.0	Black	79.00	Foreign Used	80830.0
15	VHL14203	Lagos	Lexus	ES 350	2011.0	Blue	3.55	Nigerian Used	80830.0
18	VHL18666	Lagos	Lexus	RX 350 AWD	2011.0	Black	14.00	Foreign Used	80830.0
27	VHL10241	Lagos	Lexus	ES 350	2011.0	Gray	4.20	Foreign Used	85474.0
...	...	...	...	...	...	...	...	...	...
7176	VHL15677	Lagos	Lexus	RX 330 AWD	2011.0	Black	4.90	Foreign Used	80830.0
7177	VHL16611	Abuja	Lexus	RX 350	2011.0	Gray	9.00	Foreign Used	52525.0
7184	VHL12389	Lagos	Lexus	LX 570 (5 Seats) AWD	2011.0	Black	56.00	Foreign Used	33217.0
7193	VHL13713	Lagos	Lexus	GX	2011.0	Blue	6.55	Foreign Used	89635.0
7197	VHL18453	Lagos	Lexus	RX	2011.0	Black	5.00	Foreign Used	30000.0

1599 rows × 9 columns

In [49]:
```
1 print(train.loc[train['Amount (Million Naira)']==456.0])
2
```

```
     VehicleID Location        Maker                Model    Year Colour  \
1577  VHL15837    Abuja  Rolls-Royce  Rolls-Royce Phantom  2011.0   Gray

      Amount (Million Naira)       Type  Distance
1577                   456.0  Brand New   80830.0
```

In [50]:
```
1 print(train.loc[train['Amount (Million Naira)']==0.450])
```

```
     VehicleID Location       Maker Model    Year  Colour  \
1428  VHL10434    Lagos  Volkswagen  Golf  2011.0  Silver

      Amount (Million Naira)          Type  Distance
1428                    0.45  Nigerian Used  123143.0
```

In [221]:
```
1 car_grp.get_group('Ferrari')
```

Out[221]:

	Location	Maker	Model	Year	Colour	Amount (Million Naira)	Type	Distance
1071	0	12	460	2011	1	95	1	80830

In [51]:
```
1 car_grp.get_group('Rolls-Royce')
```

Out[51]:

	VehicleID	Location	Maker	Model	Year	Colour	Amount (Million Naira)	Type	Distance
1234	VHL19383	Abuja	Rolls-Royce	Phantom Coupe	2011.0	Blue	120.0	Foreign Used	31000.0
1577	VHL15837	Abuja	Rolls-Royce	Rolls-Royce Phantom	2011.0	Gray	456.0	Brand New	80830.0
1713	VHL17982	Abuja	Rolls-Royce	Ghost Base EWB	2011.0	Gray	240.0	Foreign Used	26400.0
3474	VHL10352	Abuja	Rolls-Royce	Rolls-Royce Cullinan Base	2011.0	Black	444.0	Brand New	80830.0
4046	VHL11297	Abuja	Rolls-Royce	Rolls-Royce Phantom	2011.0	White	450.0	Brand New	80830.0
4914	VHL19719	Abuja	Rolls-Royce	Phantom Base EWB	2011.0	Red	380.0	Foreign Used	10500.0
5145	VHL19098	Abuja	Rolls-Royce	Rolls-Royce Phantom	2011.0	Brown	430.0	Brand New	4000.0
5247	VHL17959	Abuja	Rolls-Royce	Rolls-Royce Ghost	2011.0	Green	454.0	Brand New	80830.0
5527	VHL15407	Abuja	Rolls-Royce	Rolls-Royce Ghost	2011.0	White	450.0	Brand New	80830.0
5908	VHL15076	Abuja	Rolls-Royce	Wraith 6.6 RWD	2011.0	Black	185.0	Foreign Used	80830.0

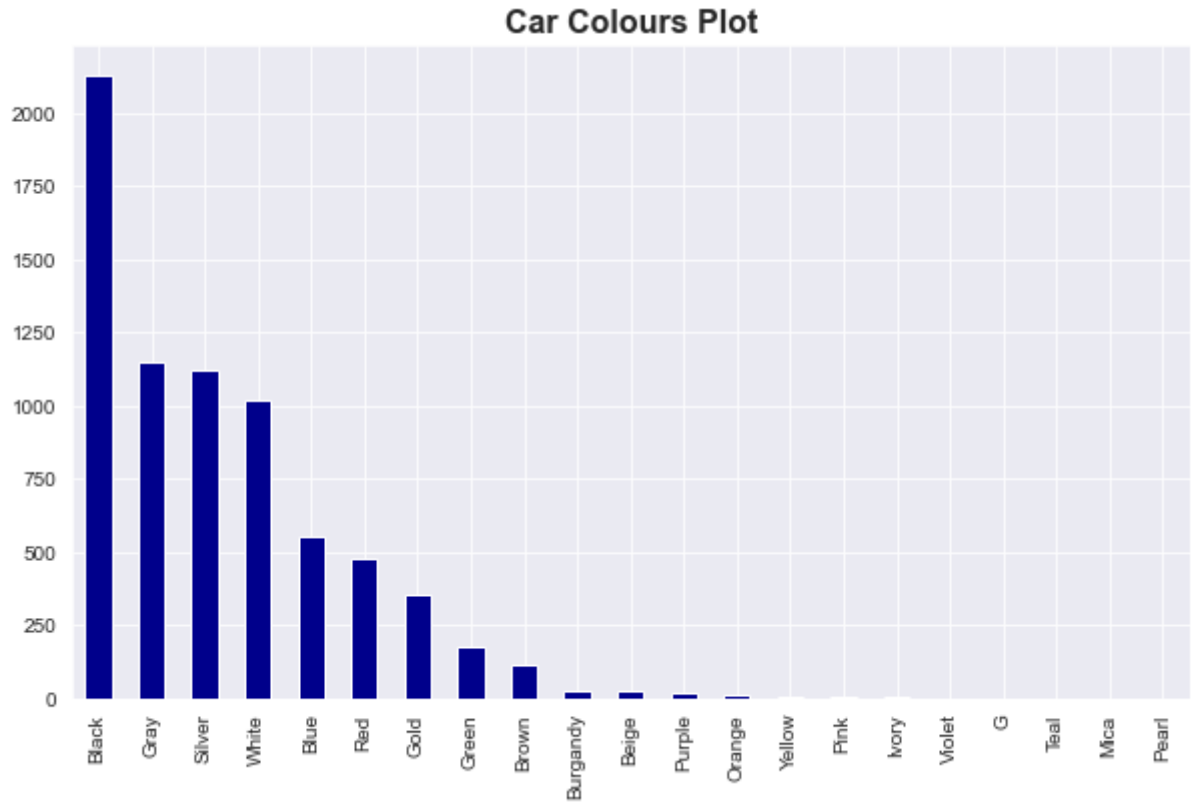
In [52]:
```
1 train.isnull().sum()
```

Out[52]:
```
VehicleID                0
Location                 0
Maker                    0
Model                    0
Year                     0
Colour                   0
Amount (Million Naira)   0
Type                     0
Distance                 0
dtype: int64
```

In [53]:
```
1 test.isnull().sum()
```

Out[53]:
```
VehicleID    0
Location     0
Maker        0
Model        0
Year         0
Colour       0
Type         0
Distance     0
dtype: int64
```

In [54]:
```
1 plt.figure(figsize=(10,6))
2 train['Colour'].value_counts().plot(kind='bar',color='darkblue');
3 plt.title('Car Colours Plot',fontweight='bold',fontsize=16);
```



In [55]:
```
train.drop(columns=['VehicleID'],inplace=True)
```

In [56]:
```
test.drop(columns=['VehicleID'],inplace=True)
```

In [57]:
```
train['Location'] = train['Location'].astype('category')
train['Maker'] = train['Maker'].astype('category')
train['Model']  = train['Model'].astype('category')
train['Colour'] = train['Colour'].astype('category')
train['Type'] = train['Type'].astype('category')
```

In [58]:
```
train['Location'] = train['Location'].cat.codes
train['Maker'] = train['Maker'].cat.codes
train['Model'] = train['Model'].cat.codes
train['Colour'] = train['Colour'].cat.codes
train['Type'] = train['Type'].cat.codes
```

In [59]:
```
train['Year'] = train['Year'].astype(int)
train['Amount (Million Naira)'] = train['Amount (Million Naira)'].astype(int)
train['Distance'] = train['Distance'].astype(int)

train.dtypes
```

Out[59]:
```
Location                 int8
Maker                    int8
Model                   int16
Year                    int32
Colour                   int8
Amount (Million Naira)  int32
Type                     int8
Distance                int32
dtype: object
```

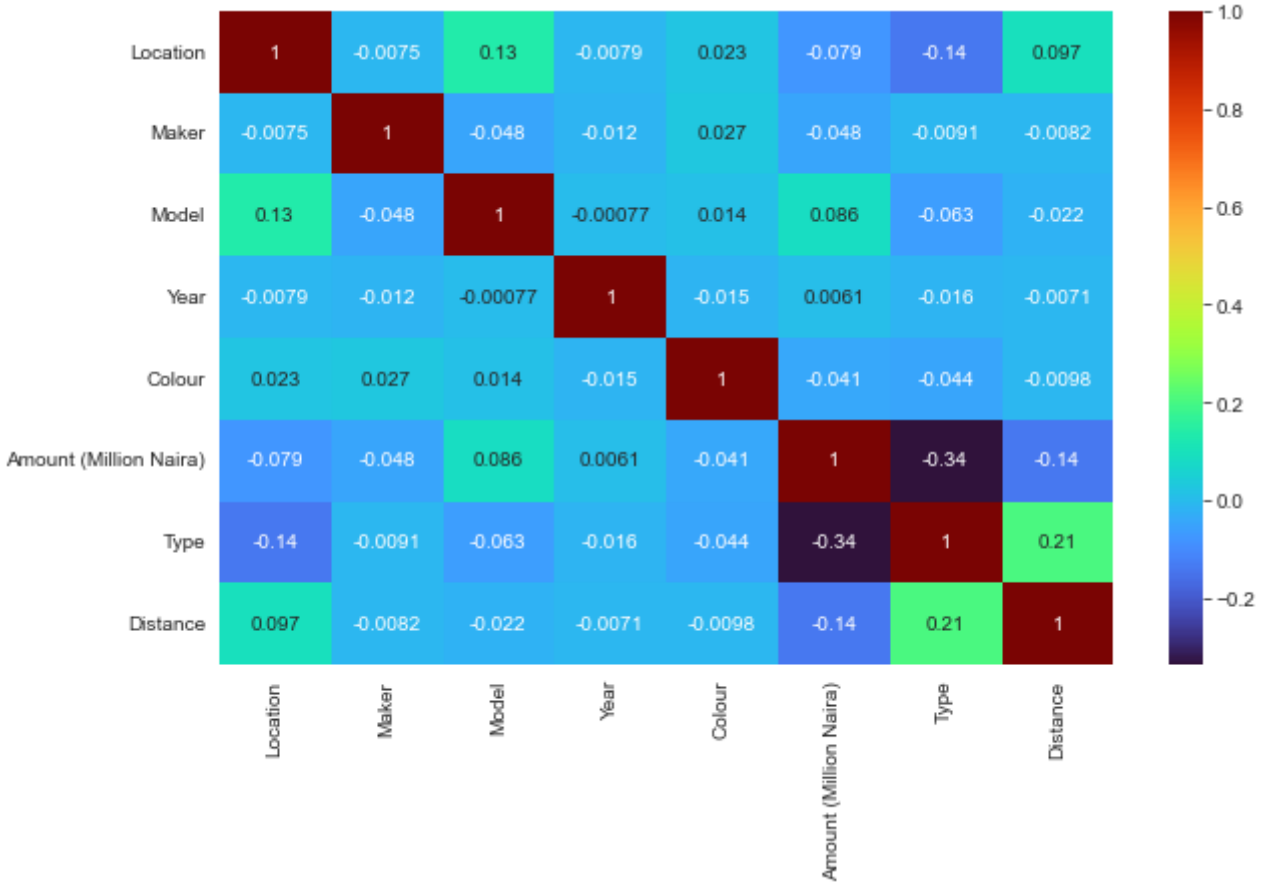
In [60]:
```
test['Location'] = test['Location'].astype('category')
test['Maker'] = test['Maker'].astype('category')
test['Model']  = test['Model'].astype('category')
test['Colour'] = test['Colour'].astype('category')
test['Type'] = test['Type'].astype('category')
```

In [61]:
```
test['Location'] = test['Location'].cat.codes
test['Maker'] = test['Maker'].cat.codes
test['Model'] = test['Model'].cat.codes
test['Colour'] = test['Colour'].cat.codes
test['Type'] = test['Type'].cat.codes
```

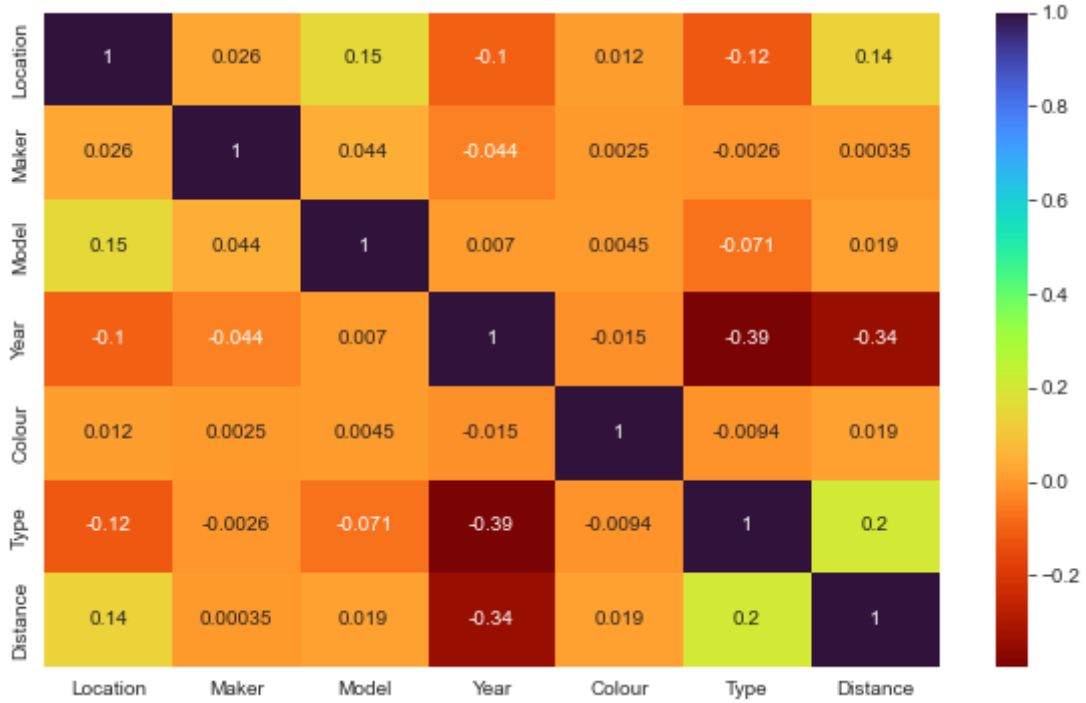
In [62]:
```
test['Distance']= test['Distance'].astype(int)
test['Year']= test['Year'].astype(int)

test.dtypes
```

Out[62]:
```
Location     int8
Maker        int8
Model       int16
Year        int32
Colour       int8
Type         int8
Distance    int32
dtype: object
```

In [63]:
```
plt.figure(figsize=(10,6))
train.corr()
sns.heatmap(train.corr(),cmap='turbo',annot=True);
```



In [64]:
```
plt.figure(figsize=(10,6))
test.corr()
sns.heatmap(test.corr(),cmap='turbo_r',annot=True);
```



In [65]:
```
y = train['Amount (Million Naira)'].values
X = train.drop(columns=['Amount (Million Naira)']).values
```

In [66]:
```
import sklearn
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
```

In [87]:
```
regressor = LinearRegression(fit_intercept=True)
regressor
```

Out[87]:
```
LinearRegression()
```

In [68]:
```
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.3,random_state=0)
```

In [69]:
```
numeric =['Location','Maker','Model','Year','Colour','Type','Distance']
```

In [70]:
```
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test  =  sc.fit_transform(X_test)
```

In [71]:
```
regressor.fit(X_train,y_train)
```

Out[71]:
```
LinearRegression()
```

In [72]:
```
y_pred= regressor.predict(X_test)
y_pred
```

Out[72]:
```
array([12.87795459, -4.82391667, 13.81753653, ...,  8.01401199,
       -2.3032464 , 12.10203203])
```

In [140]:
```
mse= mean_squared_error(y_test,y_pred)
rmse = mean_squared_error(y_test,y_pred,squared=False)

print('Mean sqaured error is :',mse)
print('Root mean squared error is:',rmse)
```

```
Mean sqaured error is : 644.0560626103866
Root mean squared error is: 25.378259645026617
```

In [141]:
```
r2_score(y_test,y_pred)
```

Out[141]:
```
0.15089349257789464
```

In [142]:
```
from sklearn.tree import DecisionTreeRegressor
```

In [173]:
```
dt = DecisionTreeRegressor(criterion='mse',splitter='best')
dt
```

Out[173]:
```
DecisionTreeRegressor()
```

In [174]:
```
dt.fit(X_train,y_train)
```

Out[174]:
```
DecisionTreeRegressor()
```

In [175]:
```
yhat = dt.predict(X_test)
yhat
```

Out[175]:
```
array([16.,  1.,  4., ...,  2.,  5.,  8.])
```

In [176]:
```
r2_score(y_test,yhat)
```

Out[176]:
```
0.3898293564546699
```

In [177]:
```
mse= mean_squared_error(y_test,yhat)
rmse = mean_squared_error(y_test,yhat,squared=False)

print('Mean sqaured error is :',mse)
print('Root mean squared error is:',rmse)
```

```
Mean sqaured error is : 462.820740115812
Root mean squared error is: 21.513268931424903
```

In [159]:
```
from sklearn.ensemble import RandomForestRegressor
```

In [263]:
```
random_ = RandomForestRegressor(random_state=0,n_estimators=1000,criterion='mse',max_features='sqrt',oob_score=1,n_j
random_
```

Out[263]:
```
RandomForestRegressor(max_features='sqrt', n_estimators=1000, n_jobs=-1,
                      oob_score=1, random_state=0)
```

In [264]:
```
random_.fit(X_train,y_train)
```

Out[264]:
```
RandomForestRegressor(max_features='sqrt', n_estimators=1000, n_jobs=-1,
                      oob_score=1, random_state=0)
```

In [265]:
```
yrand = random_.predict(X_test)
yrand
```

Out[265]:
```
array([15.868     ,  1.18166667, 11.32845   , ...,  2.915     ,
        4.84      ,  7.45526667])
```

In [266]:
```
mse= mean_squared_error(y_test,yrand)
rmse = mean_squared_error(y_test,yrand,squared=False)

print('Mean sqaured error is :',mse)
print('Root mean squared error is:',rmse)
```

```
Mean sqaured error is : 248.36999485349915
Root mean squared error is: 15.759758718124436
```

In [277]:
```
print('The regression score is:',r2_score(y_test,yrand)*100)
```

```
The regression score is: 67.25555566952598
```

In [268]:
```
test_pred = random_.predict(test)
```

In [269]:
```
test_pred
```

Out[269]:
```
array([4.314, 5.947, 5.947, ..., 5.947, 4.314, 4.314])
```

In [270]:
```
sub = pd.read_csv(r'C:\\Users\\MY PC\\Downloads\\SampleSubmission.csv')
```

In [271]:
```
1 sub.head()
```

Out[271]:

	VehicleID	Amount (Million Naira)
0	VHL18518	1.0
1	VHL17149	1.0
2	VHL10927	1.0
3	VHL12909	1.0
4	VHL12348	1.0

In [272]:
```
1 sub['Amount (Million Naira)'] = test_pred
2 sub.head(10)
```

Out[272]:

	VehicleID	Amount (Million Naira)
0	VHL18518	4.314
1	VHL17149	5.947
2	VHL10927	5.947
3	VHL12909	5.947
4	VHL12348	5.947
5	VHL10798	5.947
6	VHL11022	4.314
7	VHL12206	5.947
8	VHL11697	4.314
9	VHL12313	5.947

In [275]:
```
1 sub['Amount (Million Naira)'].max()
2
```

Out[275]: 7.149016666666667

In [274]:
```
1 sub.to_csv(r'C:\\Users\\MY PC\\Downloads\\Deelox.csv',index=False)
```

In [ ]:
```
1
```