In [ ]:

```
import warnings
warnings.filterwarnings('ignore')
import operator
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

import requests
from bs4 import BeautifulSoup

import re
import time
from matplotlib import ticker
import joblib


from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = 'all'

from selenium import webdriver
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.chrome.options import Options
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By
from selenium.webdriver.support.wait import WebDriverWait
```

In [ ]:

```
from selenium import webdriver
from selenium.webdriver.chrome.service import Service

service_path = Service("D:\Softwares and setups\chrome driver\chromedriver.exe")
driver= webdriver.Chrome(service= service_path)
driver.get("https://www.cars24.com/buy-used-car?registrationCity=Andhra%20Pradesh&registrat
time.sleep(3)
last_height = driver.execute_script("return document.body.scrollHeight")
while True:
    driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
    time.sleep(3)
    new_height = driver.execute_script("return document.body.scrollHeight")
    if new_height == last_height:
       break
    last_height = new_height
contents = driver.find_elements(By.CLASS_NAME, "_1l4fi")
con_list = []
for content in contents:
    car_transmission = content.find_element(By.CLASS_NAME, "cvakB").text.split()[-1]
    car_info = content.find_element(By.CLASS_NAME, "_3FpCg").text
    Kilometers_travelled = content.find_element(By.CLASS_NAME, "bVR0c").text.split()[0]
    owned = content.find_element(By.CLASS_NAME, "bVR0c").text.split()[2]
    fuel_varient = content.find_element(By.CLASS_NAME, "bVR0c").text.split()[-1]
    price = content.find_element(By.CLASS_NAME, "_7udZZ").text
    emi_available = content.find_element(By.CLASS_NAME, "_2HFRN").text.split()[0]
    cont_info =  {'car_info' : car_info,
                  'car_transmission' : car_transmission,
                  'Kilometers_travelled' : Kilometers_travelled,
                  'owned' : owned,
                  'fuel_varient' : fuel_varient,
                  'price' : price,
                  'emi_available' : emi_available
    }
    con_list.append(cont_info)
resale_df = pd.DataFrame(con_list)
n = resale_df[['car_transmission']]
resale_df = resale_df[n.replace(n.apply(pd.Series.value_counts)).gt(10).all(1)]
driver. quit()
```

In [6]:

```
resale_df
```

Out[6]:

	car_info	car_transmission	Kilometers_travelled	owned	fuel_varient	price	emi_av
1	2012 Hyundai i10	Automatic	14,075	1st	Petrol	₹4,00,099	₹7,822
2	2014 Hyundai i10	Manual	18,897	1st	Petrol	₹3,51,399	₹6,870
3	2022 Hyundai AURA	MT	2,903	1st	Petrol	₹8,21,999	₹15,436
4	2020 Maruti Ertiga	Manual	17,115	1st	Petrol	₹11,75,599	₹22,077
5	2014 Maruti Alto 800	Manual	15,260	1st	Petrol	₹2,82,999	₹5,533
...	...	...	...	...	...	...	
1897	2016 Maruti Alto 800	Manual	24,174	1st	Petrol	₹3,08,799	₹6,037
1898	2017 Hyundai Elite i20	Manual	37,241	1st	Petrol	₹6,14,199	₹12,008
1899	2016 Honda BR-V	Manual	62,742	2nd	Petrol	₹7,90,599	₹15,456
1900	2020 Mahindra XUV500	Automatic	7,725	1st	Diesel	₹17,07,999	₹32,074
1901	2017 Hyundai Elite i20	Manual	16,146	1st	Petrol	₹6,05,999	₹11,847

1858 rows × 7 columns

In [7]:

```
resale_df['car_transmission'].value_counts()
```

Out[7]:

```
Manual       1495
Automatic     345
MT             18
Name: car_transmission, dtype: int64
```

In [10]:

```
resale_df.loc[resale_df['car_transmission'] == 'MT']
```

Out[10]:

	car_info	car_transmission	Kilometers_travelled	owned	fuel_varient	price	emi_
3	2022 Hyundai AURA	MT	2,903	1st	Petrol	₹8,21,999	₹15,4
8	2022 Nissan MAGNITE	MT	3,412	1st	Petrol	₹6,93,099	₹13,5
29	2020 KIA SONET	MT	12,947	1st	Diesel	₹11,97,299	₹22,4
62	2021 Renault Kiger	MT	4,647	1st	Petrol	₹9,63,799	₹18,0
63	2021 Toyota URBAN CRUISER	MT	4,287	2nd	Petrol	₹10,79,799	₹20,2
65	2022 Renault Kiger	MT	2,510	1st	Petrol	₹8,35,999	₹15,6
128	2021 Toyota URBAN CRUISER	MT	9,537	1st	Petrol	₹10,81,899	₹20,3
163	2021 Renault Kiger	MT	8,200	1st	Petrol	₹9,83,899	₹18,4
166	2022 Volkswagen TAIGUN	MT	7,754	1st	Petrol	₹15,26,899	₹28,6
240	2021 Renault Kiger	MT	11,118	1st	Petrol	₹8,75,299	₹16,4
307	2021 Mahindra XUV 3OO	MT	9,194	1st	Diesel	₹13,19,499	₹24,7
317	2021 Toyota URBAN CRUISER	MT	22,622	1st	Petrol	₹10,56,499	₹19,8
426	2022 Volkswagen TAIGUN	MT	7,252	1st	Petrol	₹14,42,999	₹27,0
1029	2020 Volkswagen Polo	MT	23,135	1st	Petrol	₹9,24,699	₹17,3
1056	2021 MG HECTOR	MT	18,388	1st	Petrol	₹16,53,399	₹31,0
1094	2021 Hyundai NEW I20	MT	4,206	1st	Petrol	₹10,12,899	₹19,0

	car_info	car_transmission	Kilometers_travelled	owned	fuel_varient	price	emi_
1475	2022 Hyundai NEW I20	MT	5,009	1st	Petrol	₹8,89,799	₹16,7
1838	2021 Hyundai NEW I20	MT	33,848	2nd	Petrol	₹7,51,199	₹14,6

In [11]:

```
resale_df['car_transmission'] = resale_df['car_transmission'].replace({'MT':'Manual'})
```

In [12]:

```
resale_df['car_transmission'].value_counts()
```

Out[12]:

```
Manual       1513
Automatic     345
Name: car_transmission, dtype: int64
```

In [13]:

```
resale_df['fuel_varient'].value_counts()
```

Out[13]:

```
Petrol    1542
Diesel     311
CNG          5
Name: fuel_varient, dtype: int64
```

In [21]:

```
resale_df['car_info_year'] = resale_df['car_info'].apply(lambda x: x.split()[0])
resale_df
resale_df['car_info_year'].value_counts()
```

Out[21]:

	car_info	car_transmission	Kilometers_travelled	owned	fuel_varient	price	emi_av
1	2012 Hyundai i10	Automatic	14,075	1st	Petrol	₹4,00,099	₹7,822
2	2014 Hyundai i10	Manual	18,897	1st	Petrol	₹3,51,399	₹6,870
3	2022 Hyundai AURA	Manual	2,903	1st	Petrol	₹8,21,999	₹15,436
4	2020 Maruti Ertiga	Manual	17,115	1st	Petrol	₹11,75,599	₹22,077
5	2014 Maruti Alto 800	Manual	15,260	1st	Petrol	₹2,82,999	₹5,533
...	...	...	...	...	...	...	
1897	2016 Maruti Alto 800	Manual	24,174	1st	Petrol	₹3,08,799	₹6,037
1898	2017 Hyundai Elite i20	Manual	37,241	1st	Petrol	₹6,14,199	₹12,008
1899	2016 Honda BR-V	Manual	62,742	2nd	Petrol	₹7,90,599	₹15,456
1900	2020 Mahindra XUV500	Automatic	7,725	1st	Diesel	₹17,07,999	₹32,074
1901	2017 Hyundai Elite i20	Manual	16,146	1st	Petrol	₹6,05,999	₹11,847

1858 rows × 8 columns

Out[21]:

```
2018    328
2017    303
2019    282
2016    198
2020    174
2014    132
2021    118
2015    107
2013     69
2012     59
2011     35
2010     27
```

```
2022     18
2009      6
2008      2
Name: car info year, dtype: int64
```

In [33]:

```
resale_df = resale_df[resale_df.car_info_year != '2008']
resale_df = resale_df[resale_df.car_info_year != '2009']
resale_df['car_info_year'].value_counts()
```

Out[33]:

```
2018    328
2017    303
2019    282
2016    198
2020    174
2014    132
2021    118
2015    107
2013     69
2012     59
2011     35
2010     27
2022     18
Name: car_info_year, dtype: int64
```

In [35]:

```
resale_df['car_brand_elements_beta'] = resale_df['car_info'].apply(lambda x: x.split()[1:])
resale_df['car_brand_total'] = resale_df['car_brand_elements_beta'].apply(lambda x: '    '.
resale_df
```

Out[35]:

	car_info	car_transmission	Kilometers_travelled	owned	fuel_varient	price	emi_av
1	2012 Hyundai i10	Automatic	14,075	1st	Petrol	₹4,00,099	₹7,822
2	2014 Hyundai i10	Manual	18,897	1st	Petrol	₹3,51,399	₹6,870
3	2022 Hyundai AURA	Manual	2,903	1st	Petrol	₹8,21,999	₹15,436
4	2020 Maruti Ertiga	Manual	17,115	1st	Petrol	₹11,75,599	₹22,077
5	2014 Maruti Alto 800	Manual	15,260	1st	Petrol	₹2,82,999	₹5,533
...	...	...	...	...	...	...	
1897	2016 Maruti Alto 800	Manual	24,174	1st	Petrol	₹3,08,799	₹6,037
1898	2017 Hyundai Elite i20	Manual	37,241	1st	Petrol	₹6,14,199	₹12,008
1899	2016 Honda BR-V	Manual	62,742	2nd	Petrol	₹7,90,599	₹15,456
1900	2020 Mahindra XUV500	Automatic	7,725	1st	Diesel	₹17,07,999	₹32,074
1901	2017 Hyundai Elite i20	Manual	16,146	1st	Petrol	₹6,05,999	₹11,847

1850 rows × 10 columns

In [36]:

```
resale_df['varient_beta'] = resale_df['car_brand_elements_beta'].apply(lambda x: x[1:])
resale_df
```

Out[36]:

	car_info	car_transmission	Kilometers_travelled	owned	fuel_varient	price	emi_av
1	2012 Hyundai i10	Automatic	14,075	1st	Petrol	₹4,00,099	₹7,822
2	2014 Hyundai i10	Manual	18,897	1st	Petrol	₹3,51,399	₹6,870
3	2022 Hyundai AURA	Manual	2,903	1st	Petrol	₹8,21,999	₹15,436
4	2020 Maruti Ertiga	Manual	17,115	1st	Petrol	₹11,75,599	₹22,077
5	2014 Maruti Alto 800	Manual	15,260	1st	Petrol	₹2,82,999	₹5,533
...	...	...	...	...	...	...	
1897	2016 Maruti Alto 800	Manual	24,174	1st	Petrol	₹3,08,799	₹6,037
1898	2017 Hyundai Elite i20	Manual	37,241	1st	Petrol	₹6,14,199	₹12,008
1899	2016 Honda BR-V	Manual	62,742	2nd	Petrol	₹7,90,599	₹15,456
1900	2020 Mahindra XUV500	Automatic	7,725	1st	Diesel	₹17,07,999	₹32,074
1901	2017 Hyundai Elite i20	Manual	16,146	1st	Petrol	₹6,05,999	₹11,847

1850 rows × 11 columns

In [37]:

```
resale_df['Car Brand'] = resale_df['car_brand_elements_beta'].apply(lambda x: x[0])
resale_df['Varient'] = resale_df['varient_beta'].apply(lambda x: '  '.join(x))
resale_df
```

Out[37]:

	car_info	car_transmission	Kilometers_travelled	owned	fuel_varient	price	emi_av
1	2012 Hyundai i10	Automatic	14,075	1st	Petrol	₹4,00,099	₹7,822
2	2014 Hyundai i10	Manual	18,897	1st	Petrol	₹3,51,399	₹6,870
3	2022 Hyundai AURA	Manual	2,903	1st	Petrol	₹8,21,999	₹15,436
4	2020 Maruti Ertiga	Manual	17,115	1st	Petrol	₹11,75,599	₹22,077
5	2014 Maruti Alto 800	Manual	15,260	1st	Petrol	₹2,82,999	₹5,533
...	...	...	...	...	...	...	
1897	2016 Maruti Alto 800	Manual	24,174	1st	Petrol	₹3,08,799	₹6,037
1898	2017 Hyundai Elite i20	Manual	37,241	1st	Petrol	₹6,14,199	₹12,008
1899	2016 Honda BR-V	Manual	62,742	2nd	Petrol	₹7,90,599	₹15,456
1900	2020 Mahindra XUV500	Automatic	7,725	1st	Diesel	₹17,07,999	₹32,074
1901	2017 Hyundai Elite i20	Manual	16,146	1st	Petrol	₹6,05,999	₹11,847

1850 rows × 13 columns

In [44]:

```
resale_df['Car Brand'].value_counts()
resale_df['Varient'].value_counts()
```

Out[44]:

```
Maruti suzuki    727
Hyundai          496
Honda            137
Renault          107
Tata              91
Ford              67
Mahindra          44
Volkswagen        39
Toyota            28
KIA               27
Datsun            21
Jeep              21
MG                16
Nissan            14
Skoda             13
Mercedes           1
BMW                1
Name: Car Brand, dtype: int64
```

Out[44]:

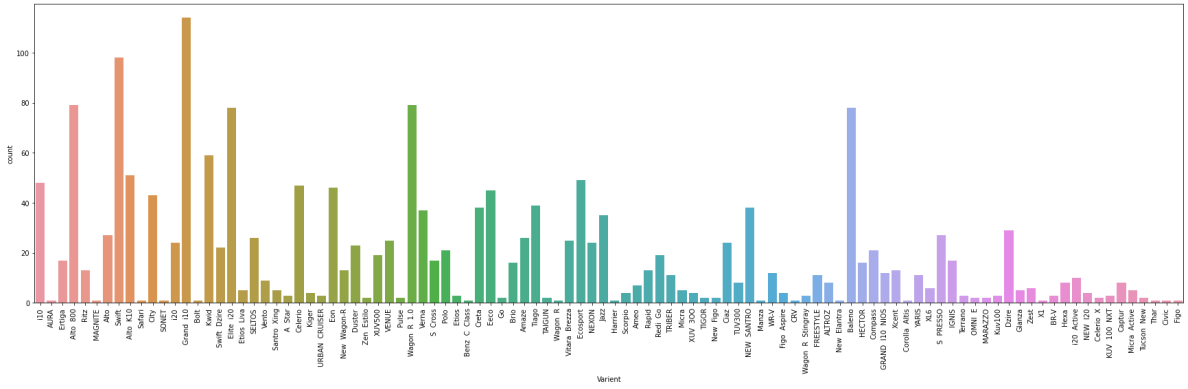
```
Grand  i10       114
Swift             98
Alto  800         79
Wagon  R  1.0     79
Baleno            78
                 ...
Bolt               1
SONET              1
Safari             1
MAGNITE            1
Figo               1
Name: Varient, Length: 102, dtype: int64
```

In [42]:

```
resale_df['Car Brand'] = resale_df['Car Brand'].replace({'Maruti':'Maruti suzuki'})
```

In [43]:

```
plt.figure(figsize=(30,8));
sns.countplot(resale_df['Varient']);
plt.xticks(rotation = 90);
```



In [45]:

```
resale_df = resale_df[resale_df['Car Brand'] != 'BMW']
resale_df = resale_df[resale_df['Car Brand'] != 'Mercedes']
resale_df['Car Brand'].value_counts()
```

Out[45]:

```
Maruti suzuki    727
Hyundai          496
Honda            137
Renault          107
Tata              91
Ford              67
Mahindra          44
Volkswagen        39
Toyota            28
KIA               27
Datsun            21
Jeep              21
MG                16
Nissan            14
Skoda             13
Name: Car Brand, dtype: int64
```

In [48]:

```
resale_df.drop(['varient_beta'], axis=1,inplace=True)
resale_df.drop(['car_brand_total'], axis=1,inplace=True)
resale_df.drop(['car_brand_elements_beta'], axis=1,inplace=True)
resale_df.drop(['car_info'], axis=1,inplace=True)
resale_df
```

Out[48]:

	car_transmission	Kilometers_travelled	owned	fuel_varient	price	emi_available	ca
1	Automatic	14,075	1st	Petrol	₹4,00,099	₹7,822/month	
2	Manual	18,897	1st	Petrol	₹3,51,399	₹6,870/month	
3	Manual	2,903	1st	Petrol	₹8,21,999	₹15,436/month	
4	Manual	17,115	1st	Petrol	₹11,75,599	₹22,077/month	
5	Manual	15,260	1st	Petrol	₹2,82,999	₹5,533/month	
...	...	...	...	...	...	...	
1897	Manual	24,174	1st	Petrol	₹3,08,799	₹6,037/month	
1898	Manual	37,241	1st	Petrol	₹6,14,199	₹12,008/month	
1899	Manual	62,742	2nd	Petrol	₹7,90,599	₹15,456/month	
1900	Automatic	7,725	1st	Diesel	₹17,07,999	₹32,074/month	
1901	Manual	16,146	1st	Petrol	₹6,05,999	₹11,847/month	

1848 rows × 9 columns

In [49]:

```
third_column = resale_df.pop('car_info_year')
resale_df.insert(0, 'Year released', third_column)
```

In [50]:

```
second_column = resale_df.pop('Varient')
resale_df.insert(0, 'Varient', second_column)
first_column = resale_df.pop('Car Brand')
resale_df.insert(0, 'Car Brand', first_column)
resale_df
```

Out[50]:

	Car Brand	Varient	Year released	car_transmission	Kilometers_travelled	owned	fuel_varient
1	Hyundai	i10	2012	Automatic	14,075	1st	Petrol
2	Hyundai	i10	2014	Manual	18,897	1st	Petrol
3	Hyundai	AURA	2022	Manual	2,903	1st	Petrol
4	Maruti suzuki	Ertiga	2020	Manual	17,115	1st	Petrol
5	Maruti suzuki	Alto 800	2014	Manual	15,260	1st	Petrol
...	...	...	...	...	...	...	...
1897	Maruti suzuki	Alto 800	2016	Manual	24,174	1st	Petrol
1898	Hyundai	Elite i20	2017	Manual	37,241	1st	Petrol
1899	Honda	BR-V	2016	Manual	62,742	2nd	Petrol
1900	Mahindra	XUV500	2020	Automatic	7,725	1st	Diesel
1901	Hyundai	Elite i20	2017	Manual	16,146	1st	Petrol

1848 rows × 9 columns

In [51]:

```
resale_df.describe()
```

Out[51]:

	Car Brand	Varient	Year released	car_transmission	Kilometers_travelled	owned	fuel_varient
count	1848	1848	1848	1848	1848	1848	1848
unique	15	100	13	2	1828	4	3
top	Maruti suzuki	Grand i10	2018	Manual	18,954	1st	Petrol
freq	727	114	328	1506	2	1310	1533

₹

In [53]:

```
resale_df = resale_df[resale_df['owned'] != '4th']
```

In [54]:

```
resale_df['owned'].value_counts()
```

Out[54]:

```
1st    1310
2nd     464
3rd      73
Name: owned, dtype: int64
```

In [55]:

```
resale_df.isnull().sum()
```

Out[55]:

```
Car Brand               0
Varient                 0
Year released           0
car_transmission        0
Kilometers_travelled    0
owned                   0
fuel_varient            0
price                   0
emi_available           0
dtype: int64
```

In [56]:

```
resale_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1847 entries, 1 to 1901
Data columns (total 9 columns):
 #   Column                Non-Null Count  Dtype 
---  ------                --------------  ----- 
 0   Car Brand             1847 non-null   object
 1   Varient               1847 non-null   object
 2   Year released         1847 non-null   object
 3   car_transmission      1847 non-null   object
 4   Kilometers_travelled  1847 non-null   object
 5   owned                 1847 non-null   object
 6   fuel_varient          1847 non-null   object
 7   price                 1847 non-null   object
 8   emi_available         1847 non-null   object
dtypes: object(9)
memory usage: 144.3+ KB
```

In [57]:

```
resale_df['Price INR(₹)'] = resale_df['price'].apply(lambda x: x.strip('₹'))
resale_df['Price INR(₹)'] = resale_df['Price INR(₹)'].apply(lambda x:re.sub(r',','',x))
resale_df
```

Out[57]:

	Car Brand	Varient	Year released	car_transmission	Kilometers_travelled	owned	fuel_varient
1	Hyundai	i10	2012	Automatic	14,075	1st	Petrol
2	Hyundai	i10	2014	Manual	18,897	1st	Petrol
3	Hyundai	AURA	2022	Manual	2,903	1st	Petrol
4	Maruti suzuki	Ertiga	2020	Manual	17,115	1st	Petrol
5	Maruti suzuki	Alto 800	2014	Manual	15,260	1st	Petrol
...	...	...	...	...	...	...	...
1897	Maruti suzuki	Alto 800	2016	Manual	24,174	1st	Petrol
1898	Hyundai	Elite i20	2017	Manual	37,241	1st	Petrol
1899	Honda	BR-V	2016	Manual	62,742	2nd	Petrol
1900	Mahindra	XUV500	2020	Automatic	7,725	1st	Diesel
1901	Hyundai	Elite i20	2017	Manual	16,146	1st	Petrol

1847 rows × 10 columns

In [58]:

```
resale_df.drop(['price'], axis=1,inplace=True)
```

In [59]:

```
resale_df['Kilometers_travelled'] = resale_df['Kilometers_travelled'].apply(lambda x:re.sub
resale_df['Kilometers_travelled']=resale_df['Kilometers_travelled'].astype('int64')
resale_df['emi_available'] = resale_df['emi_available'].apply(lambda x:x[1:])
resale_df['emi_available'] = resale_df['emi_available'].apply(lambda x:re.sub(r',','',x))
resale_df['EMI available (₹)'] = resale_df['emi_available'].apply(lambda x:re.sub(r'/month'
resale_df['Price INR(₹)']=resale_df['Price INR(₹)'].astype('int64')
resale_df.head()
```

Out[59]:

	Car Brand	Varient	Year released	car_transmission	Kilometers_travelled	owned	fuel_varient
1	Hyundai	i10	2012	Automatic	14075	1st	Petrol
2	Hyundai	i10	2014	Manual	18897	1st	Petrol
3	Hyundai	AURA	2022	Manual	2903	1st	Petrol
4	Maruti suzuki	Ertiga	2020	Manual	17115	1st	Petrol
5	Maruti suzuki	Alto 800	2014	Manual	15260	1st	Petrol
...	...	...	...	...	...	...	...
1897	Maruti suzuki	Alto 800	2016	Manual	24174	1st	Petrol
1898	Hyundai	Elite i20	2017	Manual	37241	1st	Petrol
1899	Honda	BR-V	2016	Manual	62742	2nd	Petrol
1900	Mahindra	XUV500	2020	Automatic	7725	1st	Diesel
1901	Hyundai	Elite i20	2017	Manual	16146	1st	Petrol

1847 rows × 10 columns

In [2]:

```
resale_df.head()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Input In [2], in <cell line: 1>()
----> 1 resale_df.head()

NameError: name 'resale_df' is not defined
```

In [60]:

```
third_column = resale_df.pop('EMI available (₹)')
resale_df.insert(7, 'EMI (₹/month)', third_column)
resale_df
```

Out[60]:

	Car Brand	Varient	Year released	car_transmission	Kilometers_travelled	owned	fuel_varient
1	Hyundai	i10	2012	Automatic	14075	1st	Petrol
2	Hyundai	i10	2014	Manual	18897	1st	Petrol
3	Hyundai	AURA	2022	Manual	2903	1st	Petrol
4	Maruti suzuki	Ertiga	2020	Manual	17115	1st	Petrol
5	Maruti suzuki	Alto 800	2014	Manual	15260	1st	Petrol
...	...	...	...	...	...	...	...
1897	Maruti suzuki	Alto 800	2016	Manual	24174	1st	Petrol
1898	Hyundai	Elite i20	2017	Manual	37241	1st	Petrol
1899	Honda	BR-V	2016	Manual	62742	2nd	Petrol
1900	Mahindra	XUV500	2020	Automatic	7725	1st	Diesel
1901	Hyundai	Elite i20	2017	Manual	16146	1st	Petrol

1847 rows × 10 columns

In [62]:

```
resale_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1847 entries, 1 to 1901
Data columns (total 10 columns):
 #   Column                Non-Null Count  Dtype 
---  ------                --------------  ----- 
 0   Car Brand             1847 non-null   object
 1   Varient               1847 non-null   object
 2   Year released         1847 non-null   object
 3   car_transmission      1847 non-null   object
 4   Kilometers_travelled  1847 non-null   int64 
 5   owned                 1847 non-null   object
 6   fuel_varient          1847 non-null   object
 7   EMI (₹/month)         1847 non-null   int64 
 8   emi_available         1847 non-null   object
 9   Price INR(₹)          1847 non-null   int64 
dtypes: int64(3), object(7)
memory usage: 158.7+ KB
```

In [63]:

```
resale_df.drop(['emi_available'], axis=1,inplace=True)
third_column = resale_df.pop('fuel_varient')
resale_df.insert(3, 'fuel_varient', third_column)
second_column = resale_df.pop('owned')
resale_df.insert(5, 'owned', second_column)
resale_df
```

Out[63]:

	Car Brand	Varient	Year released	fuel_varient	car_transmission	owned	Kilometers_travelled
1	Hyundai	i10	2012	Petrol	Automatic	1st	14075
2	Hyundai	i10	2014	Petrol	Manual	1st	18897
3	Hyundai	AURA	2022	Petrol	Manual	1st	2903
4	Maruti suzuki	Ertiga	2020	Petrol	Manual	1st	17115
5	Maruti suzuki	Alto 800	2014	Petrol	Manual	1st	15260
...	...	...	...	...	...	...	...
1897	Maruti suzuki	Alto 800	2016	Petrol	Manual	1st	24174
1898	Hyundai	Elite i20	2017	Petrol	Manual	1st	37241
1899	Honda	BR-V	2016	Petrol	Manual	2nd	62742
1900	Mahindra	XUV500	2020	Diesel	Automatic	1st	7725
1901	Hyundai	Elite i20	2017	Petrol	Manual	1st	16146

1847 rows × 9 columns

In [64]:

```
resale_df.to_csv('cars_24_south_india')
joblib.dump(resale_df,"cars_24_south_india.pkl")
```

Out[64]:

```
['cars_24_south_india.pkl']
```

In [72]:

```
resale_df.loc[resale_df['Year released'] == 2022]
```

Out[72]:

	Car Brand	Varient	Year released	fuel_varient	car_transmission	owned	Kilometers_travell
3	Hyundai	AURA	2022	Petrol	Manual	1st	29
8	Nissan	MAGNITE	2022	Petrol	Manual	1st	34
65	Renault	Kiger	2022	Petrol	Manual	1st	25
152	Maruti suzuki	Swift	2022	Petrol	Manual	1st	33
166	Volkswagen	TAIGUN	2022	Petrol	Manual	1st	77
189	Maruti suzuki	Celerio	2022	Petrol	Manual	1st	40
426	Volkswagen	TAIGUN	2022	Petrol	Manual	1st	72
433	Hyundai	Creta	2022	Petrol	Automatic	1st	154
568	Maruti suzuki	Alto	2022	Petrol	Manual	1st	6
681	Maruti suzuki	S PRESSO	2022	Petrol	Manual	1st	141
867	Maruti suzuki	Ciaz	2022	Petrol	Manual	1st	44
1130	Hyundai	NEW SANTRO	2022	Petrol	Manual	1st	56
1314	Maruti suzuki	Alto	2022	Petrol	Manual	1st	21
1428	Tata	ALTROZ	2022	Petrol	Manual	1st	230
1441	Maruti suzuki	Alto	2022	Petrol	Manual	1st	48
1475	Hyundai	NEW I20	2022	Petrol	Manual	1st	50
1872	Maruti suzuki	Alto	2022	Petrol	Manual	1st	50
1879	Maruti suzuki	XL6	2022	Petrol	Automatic	1st	3

In [76]:

```
n = resale_df.pop('Days in market')
resale_df.insert(3,'Days in market',n)
resale_df
```

Out[76]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_tr
1	Hyundai	i10	2012	3880	Petrol	Automatic	1st	
2	Hyundai	i10	2014	3128	Petrol	Manual	1st	
3	Hyundai	AURA	2022	188	Petrol	Manual	1st	
4	Maruti suzuki	Ertiga	2020	933	Petrol	Manual	1st	
5	Maruti suzuki	Alto 800	2014	3144	Petrol	Manual	1st	
...	...	...	...	...	...	...	...	
1897	Maruti suzuki	Alto 800	2016	2389	Petrol	Manual	1st	
1898	Hyundai	Elite i20	2017	2062	Petrol	Manual	1st	
1899	Honda	BR-V	2016	2420	Petrol	Manual	2nd	
1900	Mahindra	XUV500	2020	944	Diesel	Automatic	1st	
1901	Hyundai	Elite i20	2017	2028	Petrol	Manual	1st	

1847 rows × 10 columns

In [ ]:

```
resale_df['EMI (₹/month)']=resale_df['EMI (₹/month)'].astype('int64')
resale_df['Year released'] = resale_df['Year released'].astype('int64')
resale_df['Days in market'] = resale_df['Year released'].apply(lambda x: ((2022-x)*365)+ ra
resale_df['Year released'] = resale_df['Year released'].astype('object')
```

In [78]:

```
resale_df['Year released'].value_counts()
```

Out[78]:

```
2018    328
2017    303
2019    282
2016    198
2020    174
2014    132
2021    118
2015    106
2013     68
2012     58
2011     35
2010     27
2022     18
Name: Year released, dtype: int64
```

Type *Markdown* and LaTeX: α^2

Type *Markdown* and LaTeX: α^2

Type *Markdown* and LaTeX: α^2

Data visualisation

In [ ]:

```

```

In [151]:

```
bins = [i*200 for i in range(24)]
labels = [str(i*200) for i in range(24)][1:]
resale_df['Days in market range'] = pd.cut(x = resale_df['Days in market'], bins = bins, la
```

In [ ]:

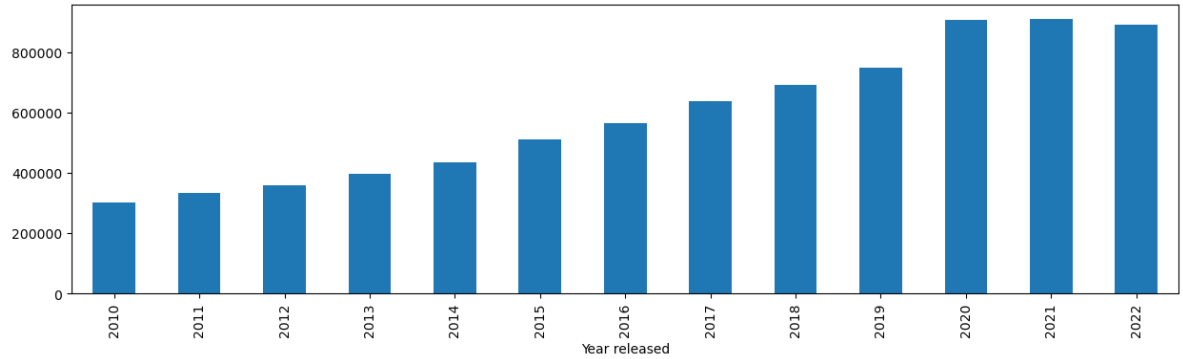
```

```

Let us see the how price of the car varies. The first one is based on car model

In [169]:

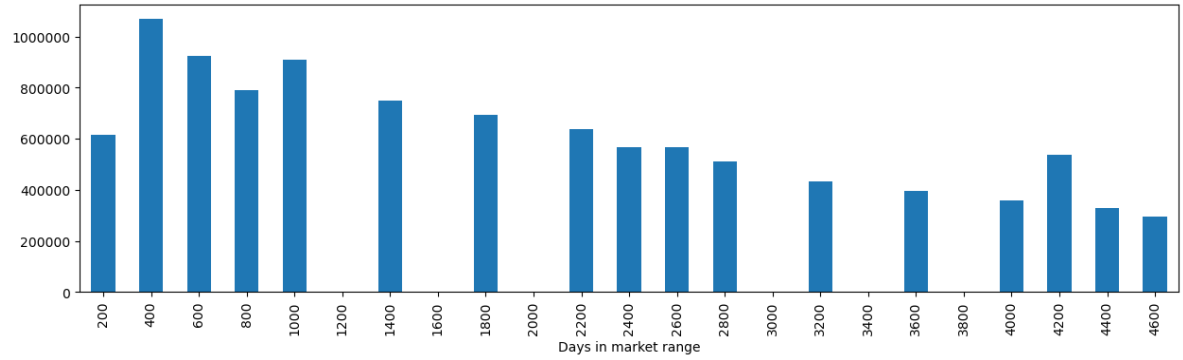
```
plt.figure(dpi = 100, figsize=(15,4));
m= resale_df.groupby(['Year released'])['Price INR(₹)'].mean().plot(kind = 'bar');
```



as expected the model with the latest varient has the higher price. But when you see it in tems of days and not in years there is some discrepancies.

In [168]:

```
plt.figure(dpi = 100, figsize=(15,4));
# plt.xticks(color='w');
m= resale_df.groupby(['Days in market range'])['Price INR(₹)'].mean().plot(kind = 'bar');
# plt.xlabel("Kilometers travelled (ascending order)");
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
m.yaxis.set_major_formatter(formatter);
```



Why the car that is released just months ago has low value? Why the cars that has been over 4000 days is priced higher? Well if we look at the bars with discrepancies, some logic is observed

In [171]:

```
resale_df.loc[resale_df['Days in market range'] == '200']
```

Out[171]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_t
3	Hyundai	AURA	2022	188	Petrol	Manual	1st	
8	Nissan	MAGNITE	2022	184	Petrol	Manual	1st	
152	Maruti suzuki	Swift	2022	184	Petrol	Manual	1st	
1130	Hyundai	NEW SANTRO	2022	190	Petrol	Manual	1st	
1314	Maruti suzuki	Alto	2022	183	Petrol	Manual	1st	
1441	Maruti suzuki	Alto	2022	197	Petrol	Manual	1st	
1872	Maruti suzuki	Alto	2022	183	Petrol	Manual	1st	

The cars that has been less than 200 days were primarily low cost cars to begin with.

In [172]:

```
resale_df.loc[resale_df['Days in market range'] == '1000']
```

Out[172]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_t
4	Maruti suzuki	Ertiga	2020	933	Petrol	Manual	1st	
29	KIA	SONET	2020	948	Diesel	Manual	1st	
37	Maruti suzuki	Swift	2020	951	Petrol	Manual	2nd	
49	Hyundai	Elite i20	2020	965	Petrol	Manual	1st	
54	KIA	SELTOS	2020	965	Diesel	Automatic	1st	
...	...	...	...	...	...	...	...	
1875	Hyundai	VENUE	2020	923	Petrol	Automatic	1st	
1889	Mahindra	XUV500	2020	914	Diesel	Automatic	1st	
1891	Tata	NEXON	2020	938	Petrol	Automatic	1st	
1895	Maruti suzuki	S PRESSO	2020	946	Petrol	Manual	1st	
1900	Mahindra	XUV500	2020	944	Diesel	Automatic	1st	

174 rows × 12 columns

In [173]:

```
resale_df.loc[resale_df['Days in market range'] == '1000']['Varient'].value_counts()
```

Out[173]:

```
S  PRESSO           18
VENUE               15
SELTOS              14
Swift                9
GRAND  I10  NIOS     9
Elite  i20           7
ALTROZ               6
HECTOR               6
Eeco                 6
NEXON                6
Baleno               5
NEW  SANTRO          5
XL6                  4
Kwid                 4
City                 4
TRIBER               4
Dzire                3
S  Cross             3
Alto                 3
Vitara  Brezza       3
Ertiga               3
New  Wagon-R         3
Polo                 3
Ecosport             3
Grand  i10           2
YARIS                2
IGNIS                2
Creta                2
XUV500               2
Rapid                2
Ciaz                 2
Tiago                2
Scorpio              1
Vento                1
SONET                1
XUV  3OO             1
Redi  Go             1
Celerio  X           1
KUV  100  NXT        1
Duster               1
Captur               1
Glanza               1
Celerio              1
Jazz                 1
Name: Varient, dtype: int64
```

In [ ]:

```

```

The majority of the cars that has around 1000 days above were dominated by high end cars that cars baring few .

In [ ]:

```
```

In [174]:

```
resale_df.loc[resale_df['Days in market range'] == '4200']
```

Out[174]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_trave
1821	Honda	Civic	2011	4200	Petrol	Manual	1st	25

And the final one is honda civic which started around 14 lakhs even in 2011.

Type *Markdown* and LaTeX: α^2

now let us see if we can find any patterns when it comes to kilometers travelled

In [159]:

```
resale_df['Kilometers_travelled'].max()
```

Out[159]:

```
235466
```

In [160]:

```
resale_df['Kilometers_travelled'].min()
```

Out[160]:

```
350
```

In [161]:

```
resale_df['Kilometers_travelled'].mean()
```

Out[161]:

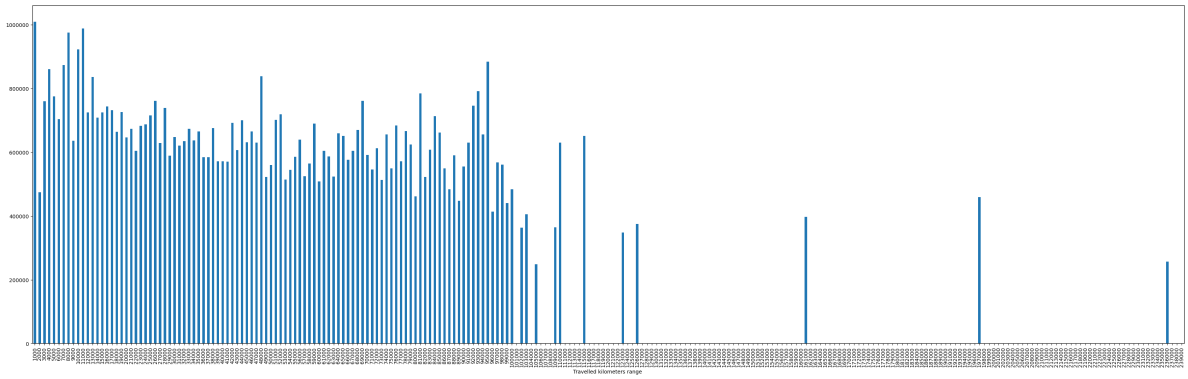
```
45932.03410936654
```

In [165]:

```
bins = [i*1000 for i in range(240)]
labels = [str(i*1000) for i in range(240)][1:]
resale_df['Travelled kilometers range '] = pd.cut(x = resale_df['Kilometers_travelled'], bi
```

In [203]:

```
plt.figure(dpi = 100, figsize=(40,12));
# plt.xticks(color='w');
m= resale_df.groupby(['Travelled kilometers range '])['Price INR(₹)'].mean().plot(kind = 'b
# plt.xlabel("Kilometers travelled (ascending order)");
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
m.yaxis.set_major_formatter(formatter);
```



with some difficulty we can observe that car that travelled just under 2000 has a less value. Let us explore

In [190]:

```
resale_df.loc[resale_df['Travelled kilometers range '] == '2000']
```

Out[190]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_travel
217	Maruti suzuki	Alto	2021	555	Petrol	Manual	1st	1
624	Datsun	Redi Go	2021	568	Petrol	Manual	1st	1

As we can see these cars are of low end value

In [240]:

```
resale_df.loc[resale_df['Car Brand'] == 'Hyundai']
```

Out[240]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_trav
1	Hyundai	i10	2012	3880	Petrol	Automatic	1st	
2	Hyundai	i10	2014	3128	Petrol	Manual	1st	
3	Hyundai	AURA	2022	188	Petrol	Manual	1st	
9	Hyundai	i10	2016	2380	Petrol	Manual	1st	
12	Hyundai	i10	2011	4249	Petrol	Manual	2nd	
...	...	...	...	...	...	...	...	
1881	Hyundai	Verna	2017	2008	Petrol	Manual	1st	
1888	Hyundai	Elite i20	2017	2019	Petrol	Manual	1st	
1894	Hyundai	NEW I20	2021	596	Petrol	Manual	1st	
1898	Hyundai	Elite i20	2017	2062	Petrol	Manual	1st	
1901	Hyundai	Elite i20	2017	2028	Petrol	Manual	1st	

496 rows × 12 columns

In [260]:

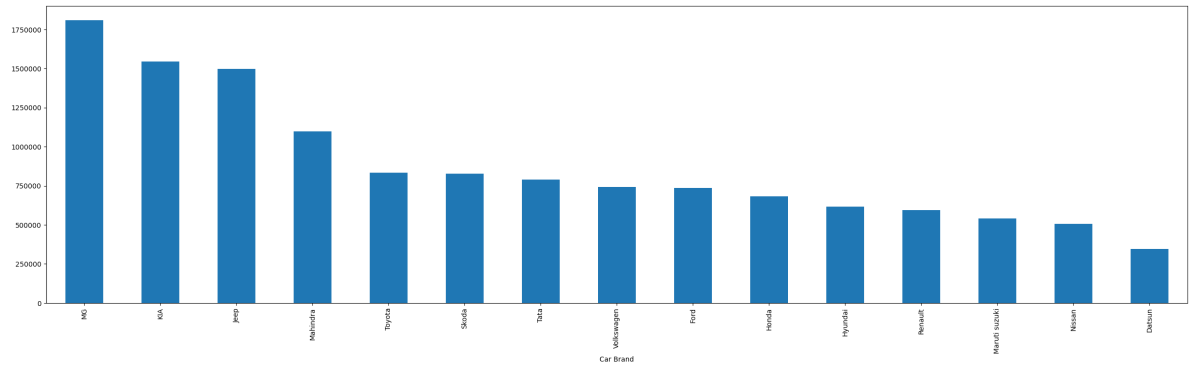
```
resale_df['Varient'].value_counts()
```

Out[260]:

```
Grand  i10       114
Swift             98
Wagon  R  1.0     79
Baleno            78
Alto  800         78
                ...
Bolt               1
SONET              1
Safari             1
MAGNITE            1
Figo               1
Name: Varient, Length: 100, dtype: int64
```

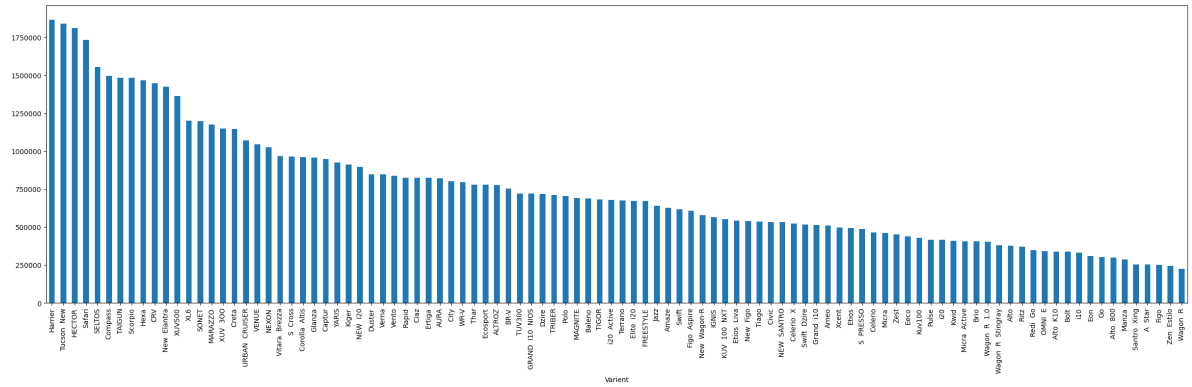
In [87]:

```
plt.figure(dpi = 100, figsize=(30,8));
m= resale_df.groupby(['Car Brand'])['Price INR(₹)'].mean().sort_values(ascending = False).p
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



In [88]:

```
plt.figure(dpi = 100, figsize=(30,8));
m= resale_df.groupby(['Varient'])['Price INR(₹)'].mean().sort_values(ascending = False).plo
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



In [279]:

```
k = resale_df['Varient'].value_counts().to_markdown()
print(k)
```

Zen Estilo	2
Pulse	2
Go	2
Celerio X	2
TAIGUN	2
Thar	1
Civic	1
Harrier	1
Corolla Altis	1
New Elantra	1
CRV	1
Manza	1
AURA	1
Wagon R	1
Bolt	1
SONET	1
Safari	1
MAGNITE	1
Figo	1

In [248]:

```
resale_df.loc[resale_df['Varient'].str.contains("i10"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("AURA"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Ertiga"),'Body type'] = 'MUV'

resale_df.loc[resale_df['Varient'].str.contains("Alto 800"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Elite i20"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("BR-V"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Grand  i10"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Swift"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Wagon  R  1.0"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Baleno"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Kwid"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Alto  K10"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Ecosport"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Celerio"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Eon"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Eeco"),'Body type'] = 'Minivan'

resale_df.loc[resale_df['Varient'].str.contains("City"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Tiago"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Creta"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("NEW  SANTRO"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Verna"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Jazz"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Dzire"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("S  PRESSO"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("SELTOS"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Amaze"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Vitara  Brezza"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("VENUE"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("NEXON"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("i20"),'Body type'] = 'Hatchback'
```

```
resale_df.loc[resale_df['Varient'].str.contains("Ciaz"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Duster"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Swift  Dzire"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Polo"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Compass"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Redi  Go"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("XUV500"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("S  Cross"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("IGNIS"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Brio"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("HECTOR"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("New  Wagon-R"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Rapid"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Ritz"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Xcent"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("GRAND  I10  NIOS"),'Body type'] = 'Hatchba

resale_df.loc[resale_df['Varient'].str.contains("WR-V"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("TRIBER"),'Body type'] = 'MUV'

resale_df.loc[resale_df['Varient'].str.contains("YARIS"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("FREESTYLE"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("i20  Active"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Vento"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Captur"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("ALTROZ"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Hexa"),'Body type'] = 'MUV'

resale_df.loc[resale_df['Varient'].str.contains("TUV300"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Ameo"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("XL6"),'Body type'] = 'MUV'

resale_df.loc[resale_df['Varient'].str.contains("Zest"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Micra  Active"),'Body type'] = 'Hatchback'
resale_df.loc[resale_df['Varient'].str.contains("Micra"),'Body type'] = 'Hatchback'
```

```
resale_df.loc[resale_df['Varient'].str.contains("Glanza"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Etios  Liva"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Santro  Xing"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("XUV  3OO"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Scorpio"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Figo  Aspire"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("NEW  I20"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Kiger"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Terrano"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("A  Star"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("URBAN  CRUISER"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Etios"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Kuv100"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Wagon  R  Stingray"),'Body type'] = 'Hatch

resale_df.loc[resale_df['Varient'].str.contains("MARAZZO"),'Body type'] = 'Minivan'

resale_df.loc[resale_df['Varient'].str.contains("OMNI  E"),'Body type'] = 'Minivan'

resale_df.loc[resale_df['Varient'].str.contains("Tucson  New"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("TIGOR"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("New  Figo"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Zen  Estilo"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Pulse"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Go"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Celerio  X"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("TAIGUN"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Thar"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Civic"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Harrier"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Corolla  Altis"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("New  Elantra"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("CRV"),'Body type'] = 'SUV'
```

```
resale_df.loc[resale_df['Varient'].str.contains("Manza"),'Body type'] = 'Sedan'

resale_df.loc[resale_df['Varient'].str.contains("Wagon  R"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Bolt"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("Safari"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("MAGNITE"),'Body type'] = 'SUV'

resale_df.loc[resale_df['Varient'].str.contains("Figo"),'Body type'] = 'Hatchback'

resale_df.loc[resale_df['Varient'].str.contains("SONET"),'Body type'] = 'SUV'
```

In [387]:

```
resale_df.to_csv('cars_24_south_india')
joblib.dump(resale_df,"cars_24_south_india.pkl")
```

In [46]:

```
resale_df = joblib.load("cars_24_south_india.pkl")
```

In [47]:

```
resale_df
```

Out[47]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_tr
1	Hyundai	i10	2012	3880	Petrol	Automatic	1st	
2	Hyundai	i10	2014	3128	Petrol	Manual	1st	
3	Hyundai	AURA	2022	188	Petrol	Manual	1st	
4	Maruti suzuki	Ertiga	2020	933	Petrol	Manual	1st	
5	Maruti suzuki	Alto 800	2014	3144	Petrol	Manual	1st	
...	...	...	...	...	...	...	...	
1897	Maruti suzuki	Alto 800	2016	2389	Petrol	Manual	1st	
1898	Hyundai	Elite i20	2017	2062	Petrol	Manual	1st	
1899	Honda	BR-V	2016	2420	Petrol	Manual	2nd	
1900	Mahindra	XUV500	2020	944	Diesel	Automatic	1st	
1901	Hyundai	Elite i20	2017	2028	Petrol	Manual	1st	

1847 rows × 13 columns

In [48]:

```
resale_df['Body type'].value_counts()
```

Out[48]:

```
Hatchback    1159
SUV           342
Sedan         255
Minivan        49
MUV            42
Name: Body type, dtype: int64
```

In [24]:

```
hatchback_df = resale_df.loc[resale_df['Body type'] == 'Hatchback'].drop(['Body type'], axi

hatchback_df

SUV_df = resale_df.loc[resale_df['Body type'] == 'SUV'].drop(['Body type'], axis=1)

SUV_df

Sedan_df = resale_df.loc[resale_df['Body type'] == 'Sedan'].drop(['Body type'], axis=1)

Sedan_df

Minivan_df = resale_df.loc[resale_df['Body type'] == 'Minivan'].drop(['Body type'], axis=1)

Minivan_df

MUV_df = resale_df.loc[resale_df['Body type'] == 'MUV'].drop(['Body type'], axis=1)

MUV_df
```

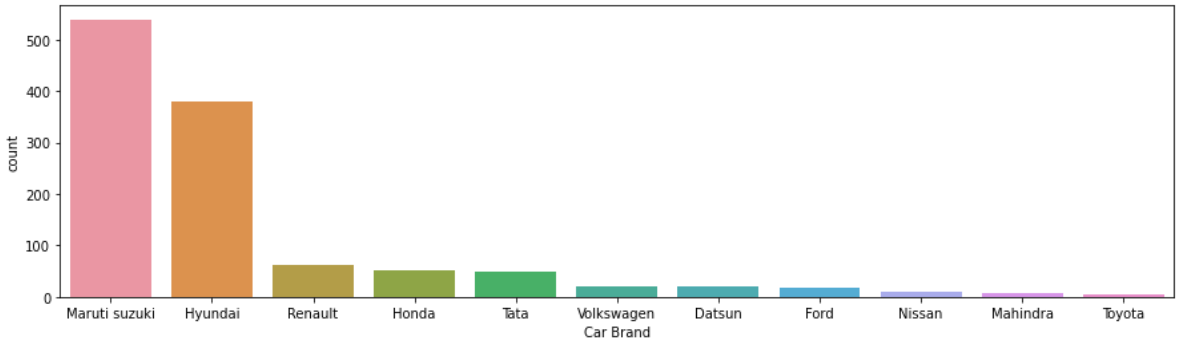
255 rows × 12 columns

Out[24]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_travelled	(₹/
147	Maruti suzuki	Eeco	2018	1682	Petrol	Manual	1st	22471	
259	Maruti suzuki	Eeco	2020	946	Petrol	Manual	1st	14836	
265	Maruti suzuki	Eeco	2020	957	Petrol	Manual	1st	13130	

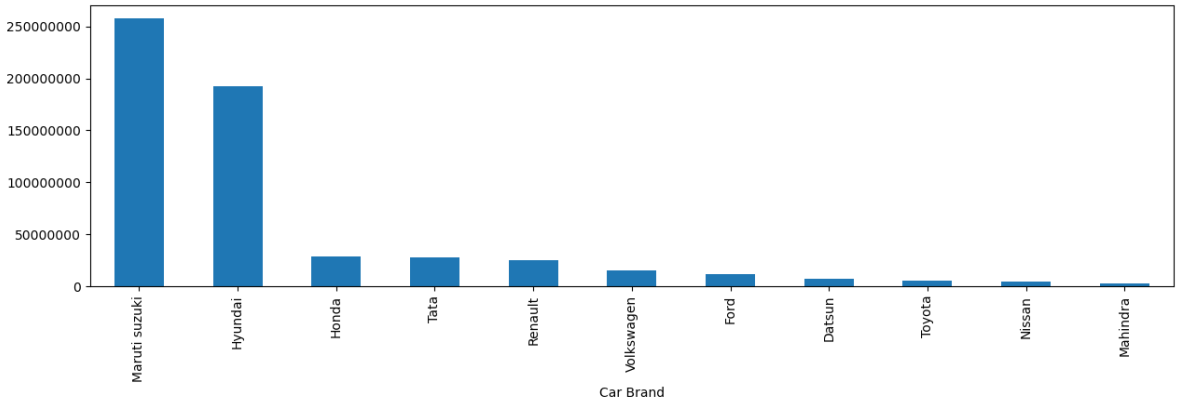
In [102]:

```
plt.figure(figsize=(15,4));
sns.countplot(x = 'Car Brand',
              data = hatchback_df,
              order = hatchback_df['Car Brand'].value_counts().index);
```



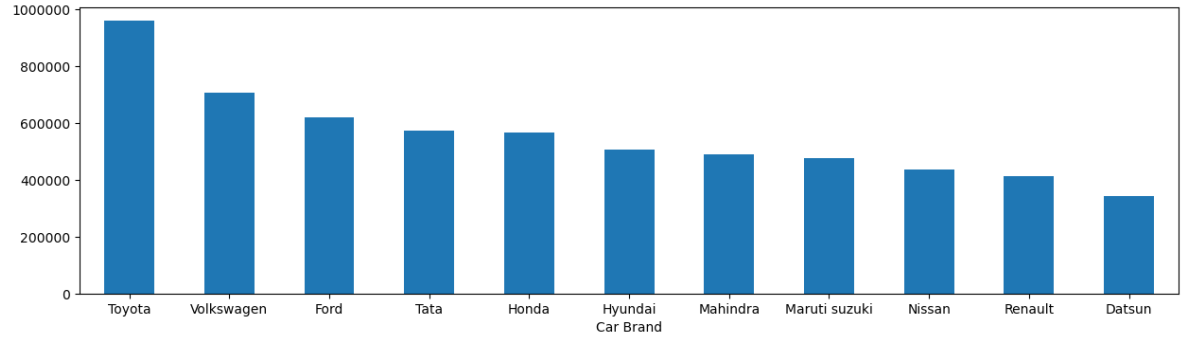
In [331]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= hatchback_df.groupby(['Car Brand'])['Price INR(₹)'].sum().sort_values(ascending = False)
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



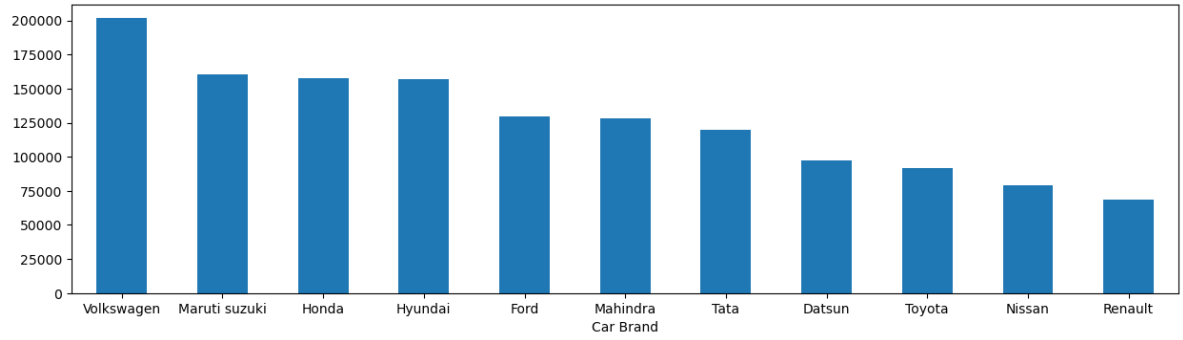
In [56]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= hatchback_df.groupby(['Car Brand'])['Price INR(₹)'].mean().sort_values(ascending = False
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 0);
m.yaxis.set_major_formatter(formatter);
```



In [339]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= hatchback_df.groupby(['Car Brand'])['Price INR(₹)'].std().sort_values(ascending = False)
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 0);
m.yaxis.set_major_formatter(formatter);
```



In [ ]:

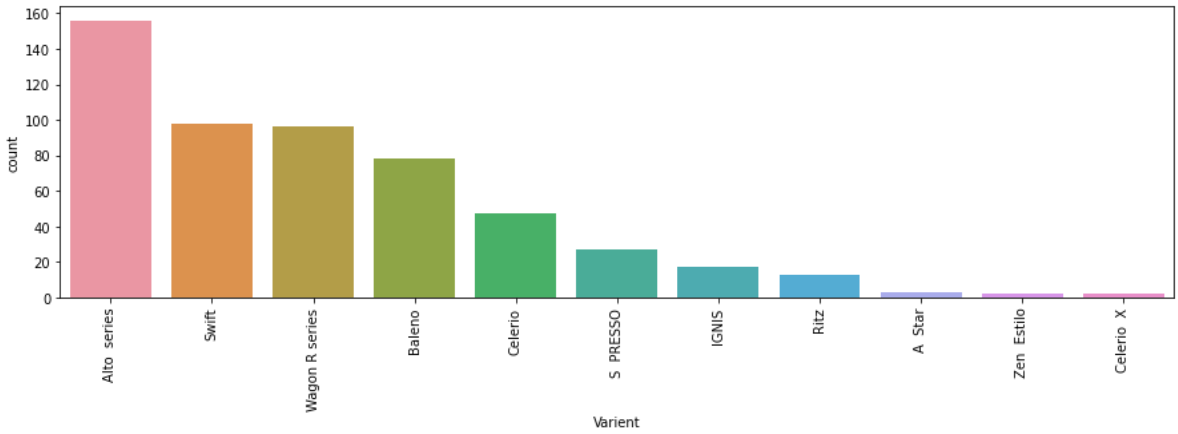
```
hatchback_df
```

If you we calculate the market of each cars by multiplying the number of cars with average price we will get a better idea of these car value

If you we calculate the market of each cars by multiplying the number of cars with average price we will get a better idea of these car value

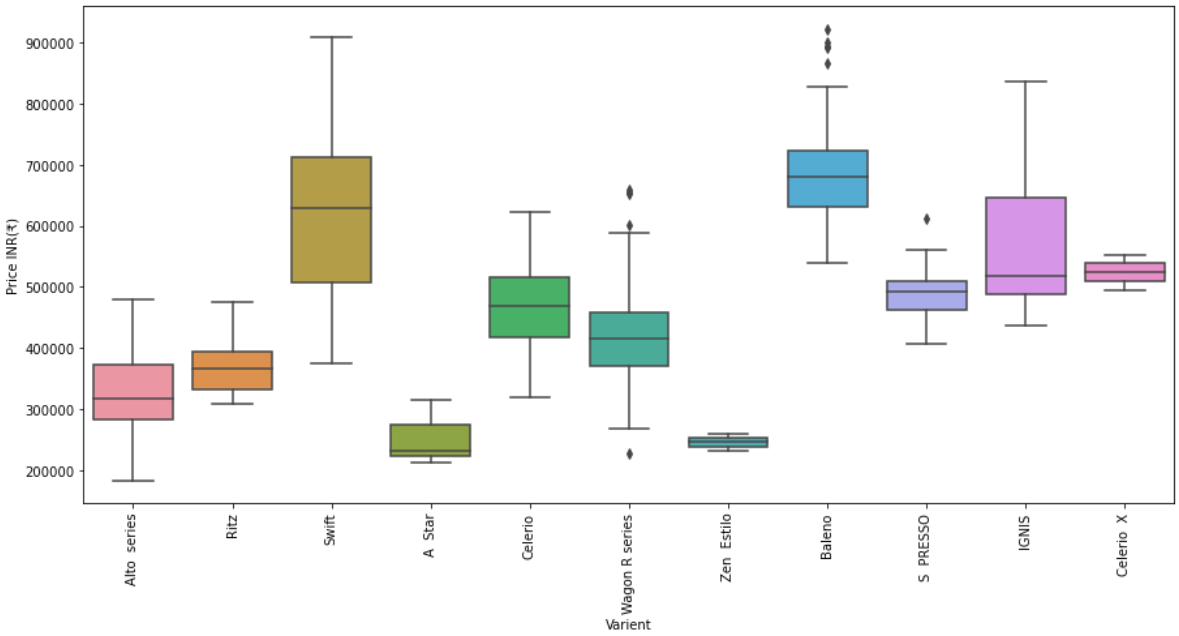
In [97]:

```
plt.figure(figsize=(15,4));
m = sns.countplot(x = 'Varient',
                  data = maurti_varient_df,
                  order = maurti_varient_df['Varient'].value_counts().index);
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



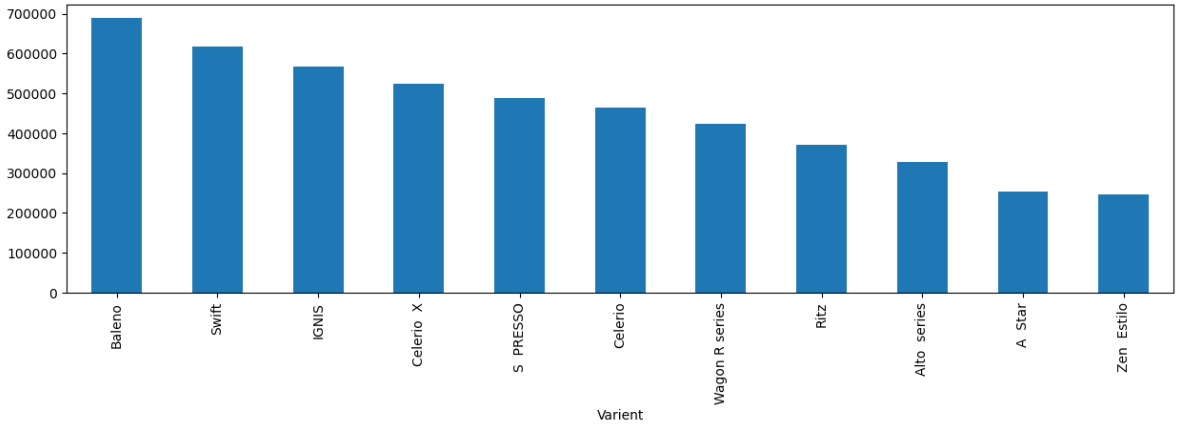
In [81]:

```
plt.figure(figsize=(15,7));
sns.boxplot(x = maurti_varient_df['Varient'], y = maurti_varient_df['Price INR(₹)']);
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



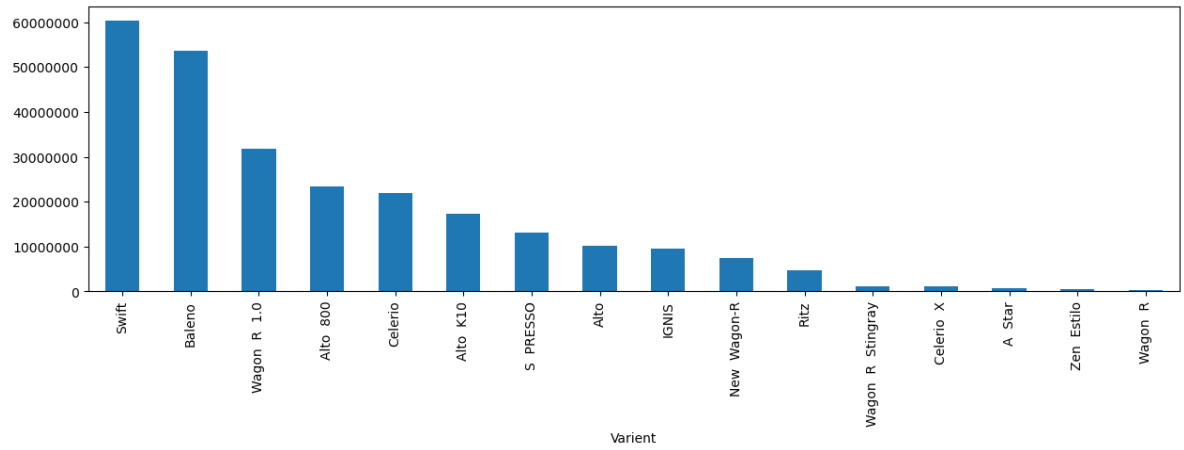
In [325]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= maurti_varient_df.groupby(['Varient'])['Price INR(₹)'].mean().sort_values(ascending = Fa
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



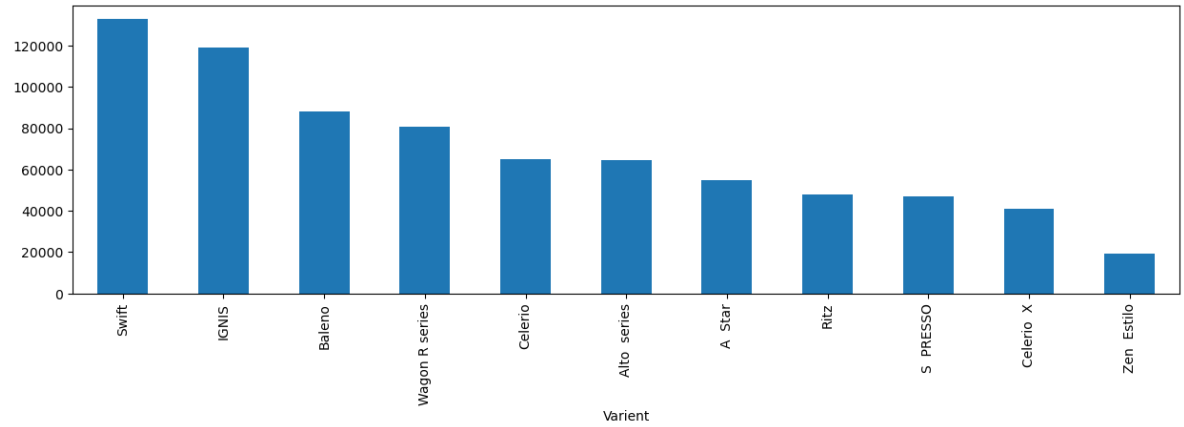
In [30]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= maurti_varient_df.groupby(['Varient'])['Price INR(₹)'].sum().sort_values(ascending = Fal
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



In [336]:

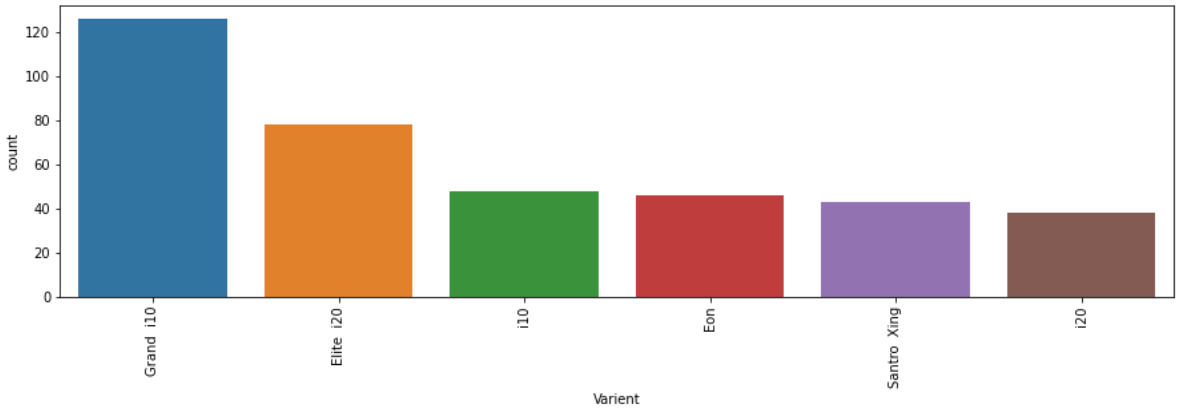
```
plt.figure(dpi = 100, figsize=(15,4));
m= maurti_varient_df.groupby(['Varient'])['Price INR(₹)'].std().sort_values(ascending = Fal
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



In [ ]:

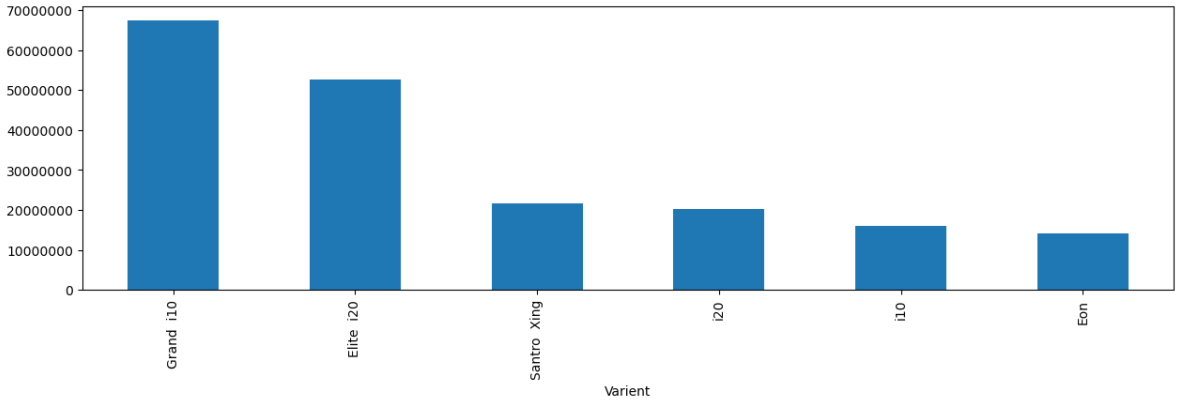
In [305]:

```
plt.figure(figsize=(15,4));
m = sns.countplot(x = 'Varient',
              data = Hyundai_hatchback_varient_df,
              order = Hyundai_hatchback_varient_df['Varient'].value_counts().index);
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



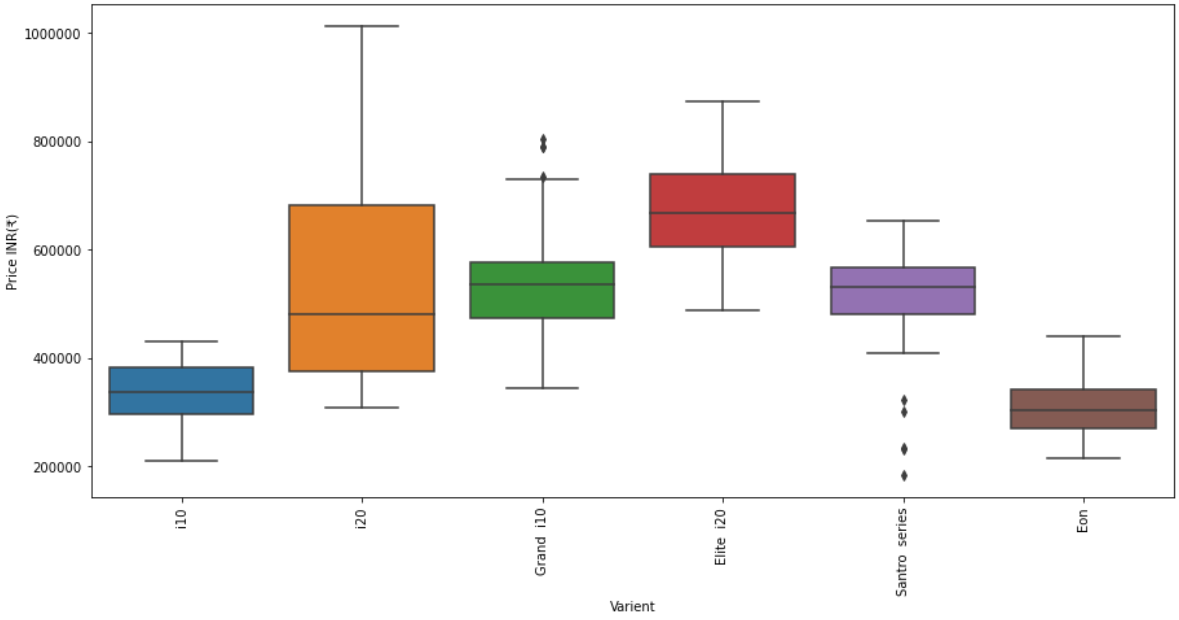
In [328]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= Hyundai_hatchback_varient_df.groupby(['Varient'])['Price INR(₹)'].sum().sort_values(asce
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



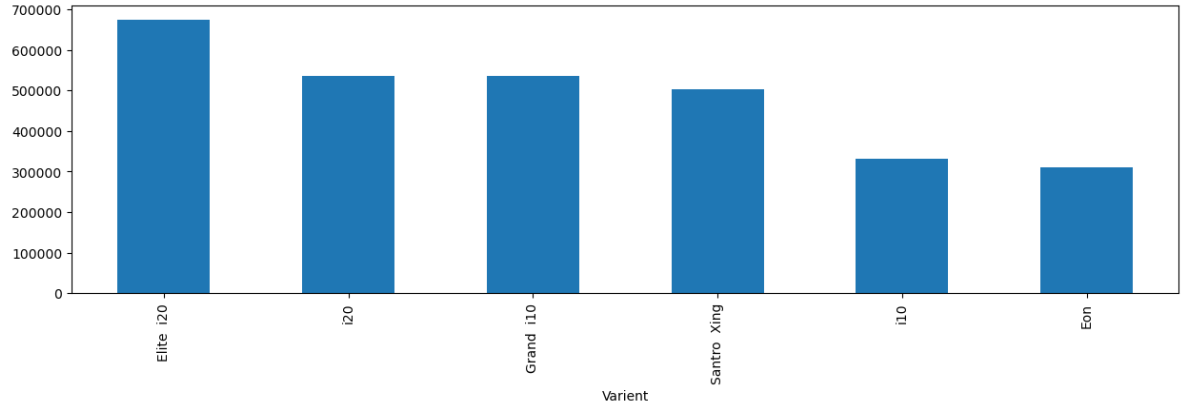
In [108]:

```
plt.figure(figsize=(15,7));
m = sns.boxplot(x = Hyundai_varient_df['Varient'], y = Hyundai_varient_df['Price INR(₹)']);
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



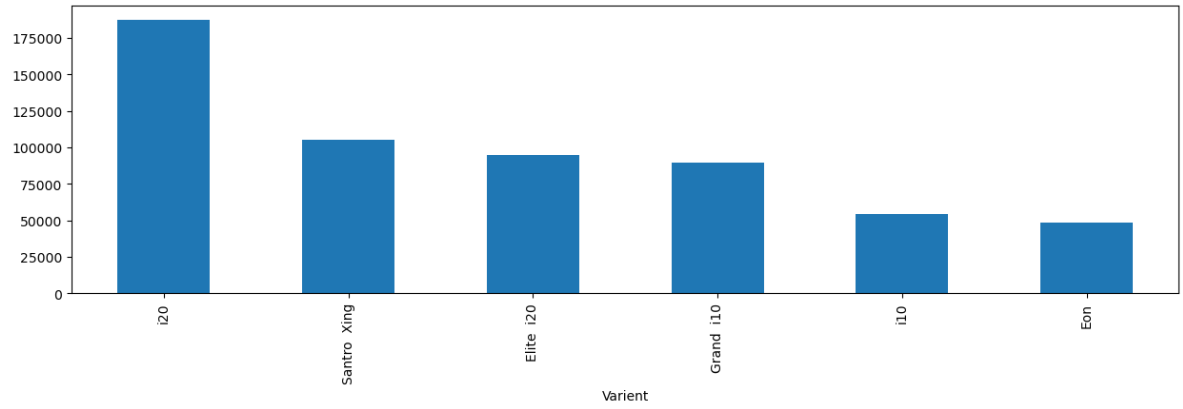
In [327]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= Hyundai_hatchback_varient_df.groupby(['Varient'])['Price INR(₹)'].mean().sort_values(asc
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



In [338]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= Hyundai_hatchback_varient_df.groupby(['Varient'])['Price INR(₹)'].std().sort_values(asce
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



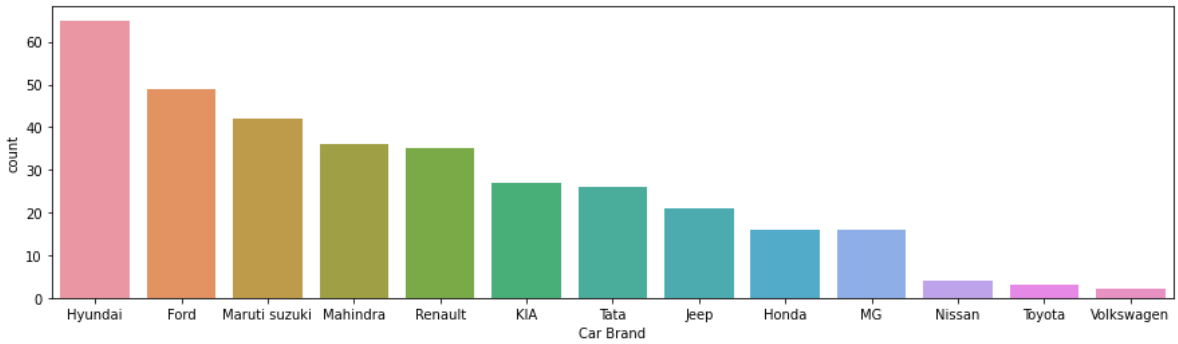
In [ ]:

SUV

In [ ]:

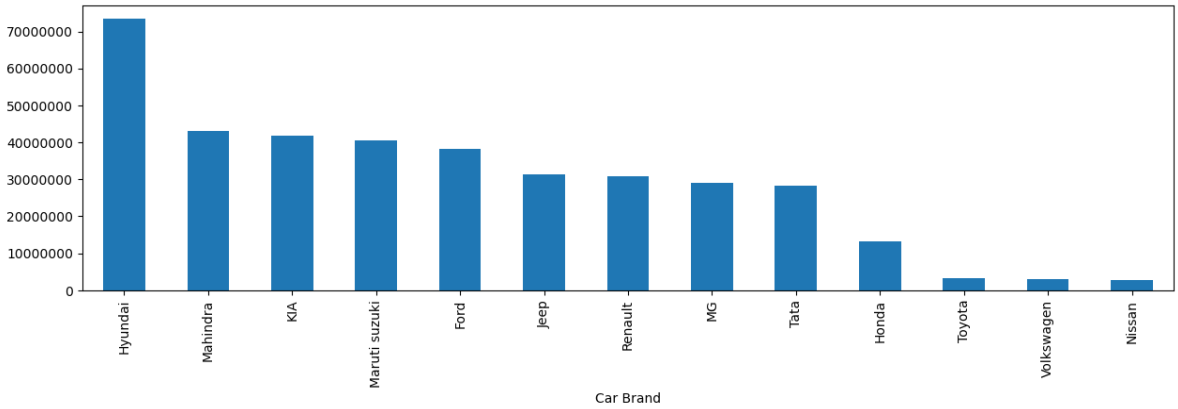
In [329]:

```
plt.figure(figsize=(15,4));
sns.countplot(x = 'Car Brand',
              data = SUV_df,
              order = SUV_df['Car Brand'].value_counts().index);
```



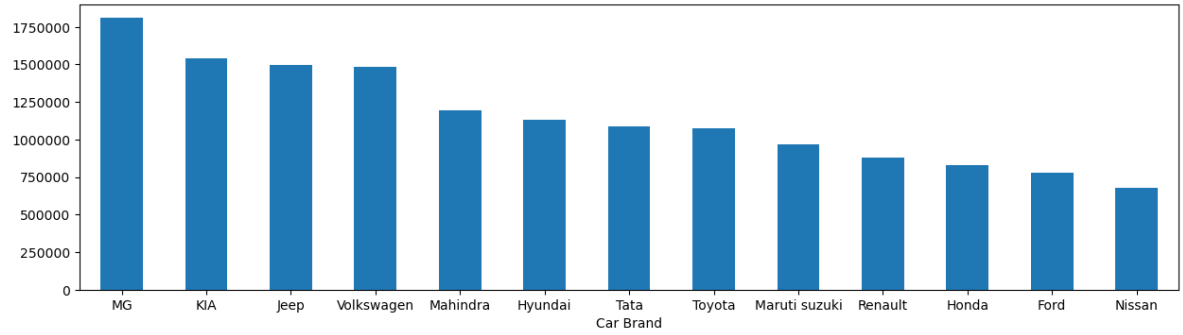
In [340]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= SUV_df.groupby(['Car Brand'])['Price INR(₹)'].sum().sort_values(ascending = False).plot(
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



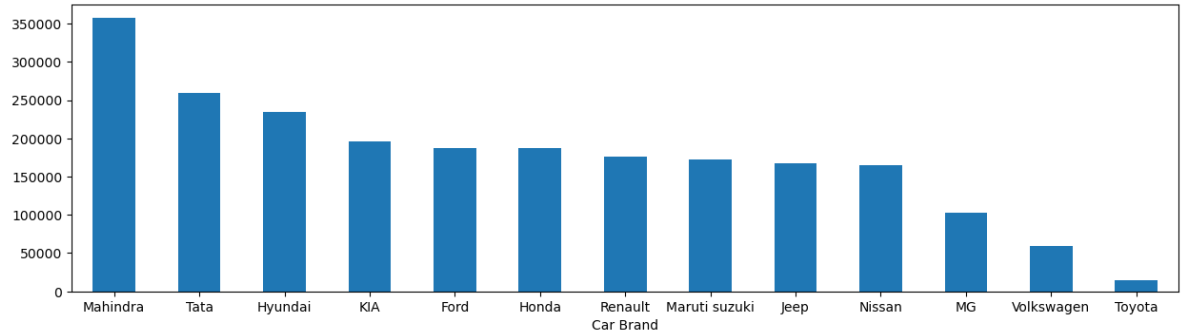
In [335]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= SUV_df.groupby(['Car Brand'])['Price INR(₹)'].mean().sort_values(ascending = False).plot
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 0);
m.yaxis.set_major_formatter(formatter);
```



In [334]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= SUV_df.groupby(['Car Brand'])['Price INR(₹)'].std().sort_values(ascending = False).plot(
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 0);
m.yaxis.set_major_formatter(formatter);
```



In [341]:

```
Mahindra_varient_df = SUV_df.loc[SUV_df['Car Brand'] == 'Mahindra']
```

In [342]:

```
Mahindra_varient_df
```

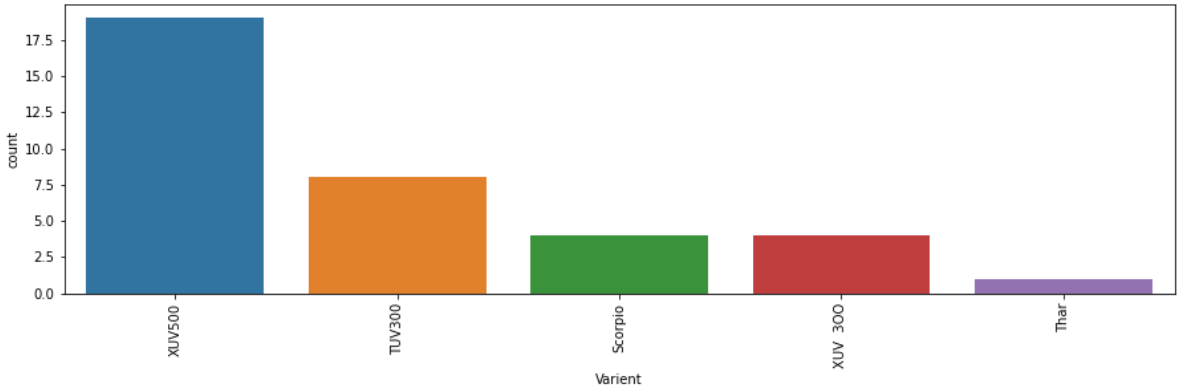
Out[342]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometer
94	Mahindra	XUV500	2019	1288	Diesel	Manual	1st	
149	Mahindra	XUV500	2019	1312	Diesel	Manual	1st	
172	Mahindra	XUV500	2017	2023	Diesel	Manual	2nd	
210	Mahindra	XUV500	2016	2387	Diesel	Automatic	2nd	
211	Mahindra	XUV500	2017	2024	Diesel	Manual	1st	
221	Mahindra	XUV500	2018	1699	Diesel	Manual	1st	
232	Mahindra	Scorpio	2014	3114	Diesel	Manual	2nd	
235	Mahindra	XUV500	2017	2057	Diesel	Automatic	2nd	
238	Mahindra	XUV500	2015	2773	Diesel	Manual	1st	
239	Mahindra	XUV500	2018	1659	Diesel	Manual	1st	
260	Mahindra	XUV500	2018	1689	Diesel	Automatic	2nd	
264	Mahindra	Scorpio	2021	572	Diesel	Manual	1st	
294	Mahindra	XUV 3OO	2020	965	Diesel	Automatic	1st	
307	Mahindra	XUV 3OO	2021	584	Diesel	Manual	1st	
325	Mahindra	XUV500	2015	2774	Diesel	Manual	1st	
330	Mahindra	TUV300	2016	2407	Diesel	Manual	1st	
351	Mahindra	TUV300	2016	2408	Diesel	Manual	1st	
358	Mahindra	XUV500	2017	2052	Diesel	Manual	1st	
361	Mahindra	TUV300	2016	2397	Diesel	Manual	1st	
389	Mahindra	XUV500	2017	2051	Diesel	Manual	2nd	
455	Mahindra	XUV500	2018	1674	Diesel	Manual	2nd	
560	Mahindra	TUV300	2016	2416	Diesel	Manual	1st	
766	Mahindra	XUV500	2018	1690	Diesel	Automatic	1st	
802	Mahindra	Scorpio	2020	919	Diesel	Manual	1st	
1102	Mahindra	TUV300	2016	2406	Diesel	Manual	1st	
1154	Mahindra	Scorpio	2018	1696	Diesel	Manual	1st	
1467	Mahindra	TUV300	2016	2372	Diesel	Manual	1st	
1729	Mahindra	TUV300	2016	2371	Diesel	Manual	1st	
1763	Mahindra	XUV500	2017	2011	Diesel	Manual	2nd	
1773	Mahindra	TUV300	2016	2387	Diesel	Manual	2nd	
1798	Mahindra	XUV 3OO	2019	1276	Diesel	Manual	1st	

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometer
1810	Mahindra	Thar	2017	2028	Diesel	Manual	2nd	
1831	Mahindra	XUV 3OO	2019	1298	Diesel	Manual	1st	
1846	Mahindra	XUV500	2015	2768	Diesel	Manual	1st	
1889	Mahindra	XUV500	2020	914	Diesel	Automatic	1st	
1900	Mahindra	XUV500	2020	944	Diesel	Automatic	1st	

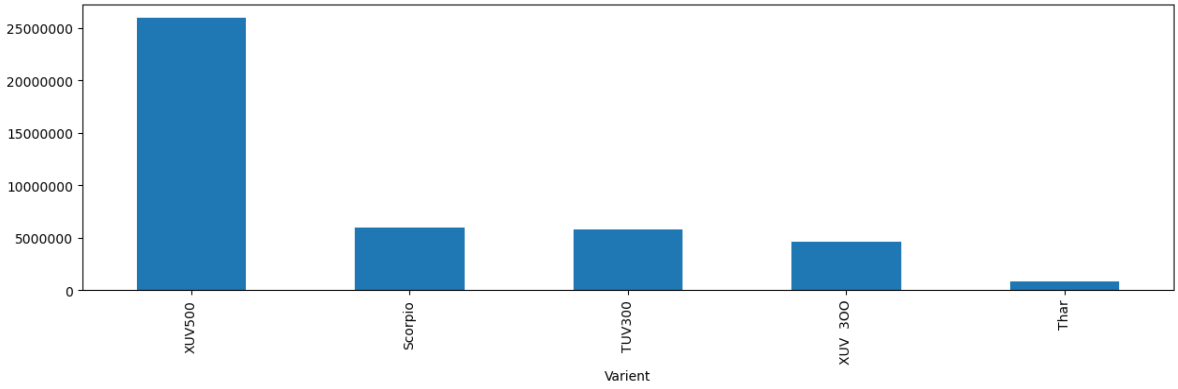
In [343]:

```
plt.figure(figsize=(15,4));
m = sns.countplot(x = 'Varient',
                  data = Mahindra_varient_df,
                  order = Mahindra_varient_df['Varient'].value_counts().index);
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



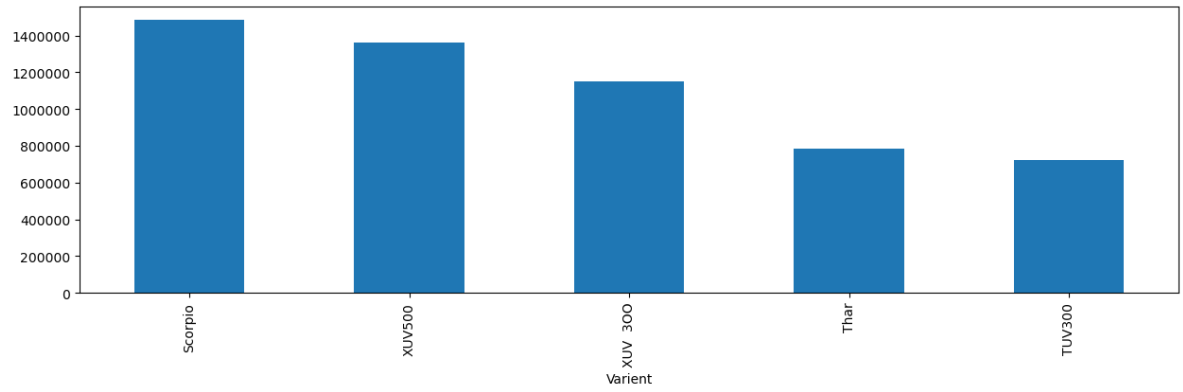
In [345]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= Mahindra_varient_df.groupby(['Varient'])['Price INR(₹)'].sum().sort_values(ascending = F
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



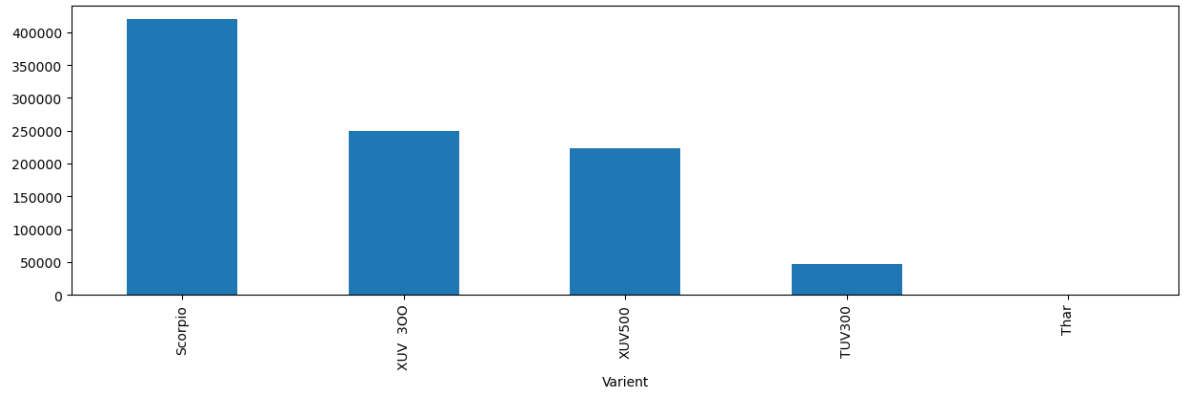
In [346]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= Mahindra_varient_df.groupby(['Varient'])['Price INR(₹)'].mean().sort_values(ascending =
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



In [347]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= Mahindra_varient_df.groupby(['Varient'])['Price INR(₹)'].std().sort_values(ascending = F
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```



In [ ]:

In [ ]:

In [83]:

```
Hyundai_varient_df = hatchback_df.loc[hatchback_df['Car Brand'] == 'Hyundai']
```

In [28]:

```
maurti_varient_df = hatchback_df.loc[hatchback_df['Car Brand'] == 'Maruti suzuki']
```

In [26]:

```
Hyundai_varient_df
```

Out[26]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_tra
1	Hyundai	i10	2012	3880	Petrol	Automatic	1st	
2	Hyundai	i10	2014	3128	Petrol	Manual	1st	
9	Hyundai	i10	2016	2380	Petrol	Manual	1st	
12	Hyundai	i10	2011	4249	Petrol	Manual	2nd	
16	Hyundai	i10	2011	4242	Petrol	Automatic	1st	
...	...	...	...	...	...	...	...	
1866	Hyundai	GRAND I10 NIOS	2020	925	Petrol	Automatic	1st	
1888	Hyundai	Elite i20	2017	2019	Petrol	Manual	1st	
1894	Hyundai	NEW I20	2021	596	Petrol	Manual	1st	
1898	Hyundai	Elite i20	2017	2062	Petrol	Manual	1st	
1901	Hyundai	Elite i20	2017	2028	Petrol	Manual	1st	

379 rows × 12 columns

In [106]:

```
Hyundai_varient_df['Varient'].value_counts()
```

Out[106]:

```
Grand  i10        126
Elite  i20         78
i10                48
Eon                46
Santro  series     43
i20                38
Name: Varient, dtype: int64
```

In [103]:

```
Hyundai_varient_df = Hyundai_varient_df.set_index('Varient')
Hyundai_varient_df = Hyundai_varient_df.rename(index={'Santro  Xing':'Santro  series'})
```

In [105]:

```
Hyundai_varient_df = Hyundai_varient_df.reset_index('Varient')
```

In [109]:

```
Hyundai_varient_df
```

Out[109]:

	Varient	Car Brand	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_trave
0	i10	Hyundai	2012	3880	Petrol	Automatic	1st	1
1	i10	Hyundai	2014	3128	Petrol	Manual	1st	1
2	i10	Hyundai	2016	2380	Petrol	Manual	1st	1
3	i10	Hyundai	2011	4249	Petrol	Manual	2nd	2
4	i10	Hyundai	2011	4242	Petrol	Automatic	1st	6
...	...	...	...	...	...	...	...	
374	Grand i10	Hyundai	2020	925	Petrol	Automatic	1st	3
375	Elite i20	Hyundai	2017	2019	Petrol	Manual	1st	1
376	i20	Hyundai	2021	596	Petrol	Manual	1st	
377	Elite i20	Hyundai	2017	2062	Petrol	Manual	1st	3
378	Elite i20	Hyundai	2017	2028	Petrol	Manual	1st	1

379 rows × 12 columns

In [29]:

```
maurti_varient_df
```

Out[29]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_trav
5	Maruti suzuki	Alto 800	2014	3144	Petrol	Manual	1st	
6	Maruti suzuki	Ritz	2011	4208	Petrol	Manual	1st	
10	Maruti suzuki	Alto	2010	4597	Petrol	Manual	1st	
11	Maruti suzuki	Alto 800	2016	2411	Petrol	Manual	1st	1(
13	Maruti suzuki	Swift	2021	601	Petrol	Manual	1st	
...	...	...	...	...	...	...	...	
1887	Maruti suzuki	Ritz	2010	4574	Petrol	Manual	1st	
1892	Maruti suzuki	Baleno	2016	2428	Petrol	Manual	2nd	
1895	Maruti suzuki	S PRESSO	2020	946	Petrol	Manual	1st	
1896	Maruti suzuki	Wagon R 1.0	2010	4595	Petrol	Manual	2nd	
1897	Maruti suzuki	Alto 800	2016	2389	Petrol	Manual	1st	

539 rows × 12 columns

In [249]:

```
ps = maurti_varient_df['Varient'].value_counts()
```

In [255]:

```
ps[1]
```

Out[255]:

```
98
```

In [318]:

```
mean_list = {name_list[i] : round(maurti_varient_df.loc[maurti_varient_df['Varient'] == f'{
```

In [313]:

```
total_list = {name_list[i] : maurti_varient_df.loc[maurti_varient_df['Varient'] == f'{name_
```

In [314]:

```
total_list
```

Out[314]:

```
{'Alto  series': 50947744,
 'Ritz': 4812587,
 'Swift': 60419602,
 'A  Star': 761597,
 'Celerio': 21867053,
 'Wagon R series': 40757304,
 'Zen  Estilo': 492698,
 'Baleno': 53761122,
 'S  PRESSO': 13164473,
 'IGNIS': 9632083,
 'Celerio  X': 1047998}
```

In [48]:

```
maurti_varient_df = maurti_varient_df.set_index('Varient')
maurti_varient_df = maurti_varient_df.rename(index={'Wagon  R':'Wagon R series'})
```

In [57]:

```
maurti_varient_df = maurti_varient_df.reset_index('Varient')
```

In [56]:

```
maurti_varient_df = maurti_varient_df.set_index('Varient')
maurti_varient_df = maurti_varient_df.rename(index={'Alto  800':'Alto  series'})
```

In [58]:

```
maurti_varient_df['Varient'].value_counts()
```

Out[58]:

```
Alto  series      156
Swift              98
Wagon R series     96
Baleno             78
Celerio            47
S  PRESSO          27
IGNIS              17
Ritz               13
A  Star             3
Zen  Estilo         2
Celerio  X          2
Name: Varient, dtype: int64
```

In [ ]:

```
plt.figure(figsize=(15,4));
sns.countplot(x = 'Car Brand',
              data = hatchback_df,
              order = hatchback_df['Car Brand'].value_counts().index);
```

In [ ]:

```

```

In [ ]:

```

```

In [112]:

```
Hyundai_varient_df.to_csv('Hyundai_varient_df')
joblib.dump(Hyundai_varient_df,"Hyundai_varient_df.pkl")
```

Out[112]:

```
['Hyundai_varient_df.pkl']
```

In [113]:

```
maurti_varient_df.to_csv('maruti_varient_df')
joblib.dump(maurti_varient_df,"maruti_varient_df.pkl")
```

Out[113]:

```
['maruti_varient_df.pkl']
```

In [107]:

```
resale_df.to_csv('cars_24_south_india')
joblib.dump(resale_df,"cars_24_south_india.pkl")
```

Out[107]:

```
['cars_24_south_india.pkl']
```

In [ ]:

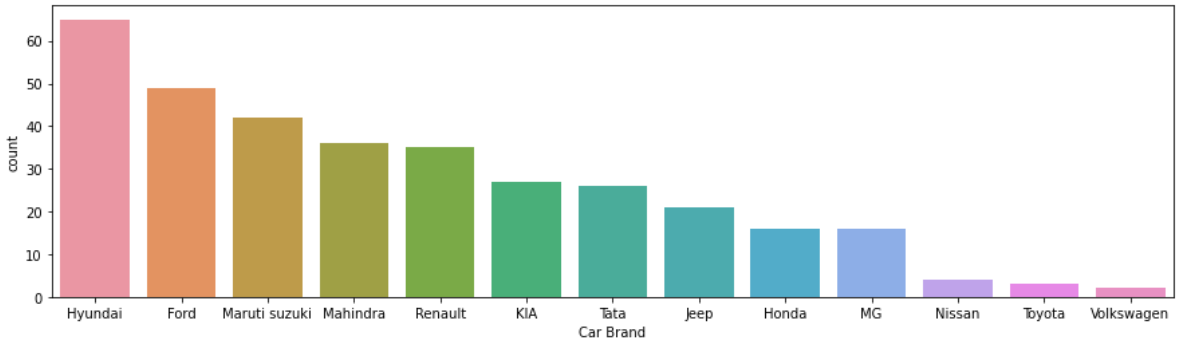
```

```

In [6]:

```
plt.figure(figsize=(15,4));
sns.countplot(x = 'Car Brand',
              data = SUV_df,
              order = SUV_df['Car Brand'].value_counts().index);

resale_df = joblib.load("cars_24_south_india.pkl")
```



In [ ]:

```
plt.figure(dpi = 100, figsize=(15,4));
m= SUV_df.groupby(maurti_varient_df['Varient'])['Price INR(₹)'].mean().sort_values(ascendin
formatter = ticker.ScalarFormatter();
formatter.set_scientific(False);
plt.xticks(rotation = 90);
m.yaxis.set_major_formatter(formatter);
```

In [17]:

```
resale_df = joblib.load("cars_24_south_india.pkl")
```

In [114]:

```
maurti_varient_df = joblib.load("maruti_varient_df.pkl")
Hyundai_varient_df = joblib.load("Hyundai_hatchback_varient_df.pkl")
```

In [111]:

```
maurti_varient_df
```

Out[111]:

	Varient	Car Brand	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_trave
0	Alto series	Maruti suzuki	2014	3144	Petrol	Manual	1st	1
1	Ritz	Maruti suzuki	2011	4208	Petrol	Manual	1st	4
2	Alto series	Maruti suzuki	2010	4597	Petrol	Manual	1st	
3	Alto series	Maruti suzuki	2016	2411	Petrol	Manual	1st	10
4	Swift	Maruti suzuki	2021	601	Petrol	Manual	1st	8
...	...	...	...	...	...	...	...	
534	Ritz	Maruti suzuki	2010	4574	Petrol	Manual	1st	7
535	Baleno	Maruti suzuki	2016	2428	Petrol	Manual	2nd	5
536	S PRESSO	Maruti suzuki	2020	946	Petrol	Manual	1st	2
537	Wagon R series	Maruti suzuki	2010	4595	Petrol	Manual	2nd	2
538	Alto series	Maruti suzuki	2016	2389	Petrol	Manual	1st	2

539 rows × 12 columns

In [22]:

```
maurti_varient_df
```

Out[22]:

	Car Brand	Varient	Year released	Days in market	fuel_varient	car_transmission	owned	Kilometers_tr
1	Hyundai	i10	2012	3880	Petrol	Automatic	1st	
2	Hyundai	i10	2014	3128	Petrol	Manual	1st	
3	Hyundai	AURA	2022	188	Petrol	Manual	1st	
4	Maruti suzuki	Ertiga	2020	933	Petrol	Manual	1st	
5	Maruti suzuki	Alto 800	2014	3144	Petrol	Manual	1st	
...	...	...	...	...	...	...	...	
1897	Maruti suzuki	Alto 800	2016	2389	Petrol	Manual	1st	
1898	Hyundai	Elite i20	2017	2062	Petrol	Manual	1st	
1899	Honda	BR-V	2016	2420	Petrol	Manual	2nd	
1900	Mahindra	XUV500	2020	944	Diesel	Automatic	1st	
1901	Hyundai	Elite i20	2017	2028	Petrol	Manual	1st	

1847 rows × 13 columns

In [ ]:

```

```