In [2]:
```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler, LabelEncoder
import warnings
warnings.filterwarnings('ignore')
```

In [3]:
```
df=pd.read_csv('cars.csv')
```

In [4]:
```
df.head()
```

Out[4]:

	First Name	Last Name	Country	Car Brand	Car Model	Car Color	Year of Manufacture	Credit Card Type
0	Yetty	Arghent	Indonesia	Ford	Club Wagon	Teal	1993	mastercard
1	Crystal	Bosworth	China	Cadillac	Escalade ESV	Fuscia	2007	mastercard
2	Monro	Houdhury	Indonesia	Mazda	Miata MX-5	Orange	2009	maestro
3	Bowie	Clair	China	Audi	A4	Orange	2005	instapayment
4	Myrvyn	McAllister	Czech Republic	Nissan	Maxima	Pink	1994	maestro

In [5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30000 entries, 0 to 29999
Data columns (total 8 columns):
 #   Column               Non-Null Count  Dtype 
---  ------               --------------  ----- 
 0   First Name           30000 non-null  object
 1   Last Name            30000 non-null  object
 2   Country              30000 non-null  object
 3   Car Brand            30000 non-null  object
 4   Car Model            30000 non-null  object
 5   Car Color            30000 non-null  object
 6   Year of Manufacture  30000 non-null  int64 
 7   Credit Card Type     30000 non-null  object
dtypes: int64(1), object(7)
memory usage: 1.8+ MB
```

In [6]:
```
df.describe()
```

Out[6]:

	Year of Manufacture
count	30000.000000
mean	2000.247500
std	9.131398
min	1909.000000
25%	1995.000000
50%	2002.000000
75%	2007.000000
max	2013.000000

In [7]:
```
df.isnull()
```

Out[7]:

	First Name	Last Name	Country	Car Brand	Car Model	Car Color	Year of Manufacture	Credit Card Type
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...
29995	False	False	False	False	False	False	False	False
29996	False	False	False	False	False	False	False	False
29997	False	False	False	False	False	False	False	False
29998	False	False	False	False	False	False	False	False
29999	False	False	False	False	False	False	False	False

30000 rows × 8 columns

In [8]:
```
df.isnull().sum()
```

Out[8]:
```
First Name            0
Last Name             0
Country               0
Car Brand             0
Car Model             0
Car Color             0
Year of Manufacture   0
Credit Card Type      0
dtype: int64
```

In [9]:
```
# Data Formatting and Normalization
scaler = MinMaxScaler()
```

In [11]:
```
# Normalize numerical columns
num_columns = ['Year of Manufacture']
df[num_columns]=  scaler.fit_transform(df[num_columns])
```

In [14]:
```
# Convert Categorical Variables to Quantitative Variables
encoder =LabelEncoder()
var_columns=['First Name','Last Name','Country','Car Brand','Car Model','Car Color','Cre
for col in var_columns:
    df[col] = encoder.fit_transform(df[col])
```

In [16]:
```
df.head().T
```

Out[16]:

	0	1	2	3	4
First Name	8087.000000	1845.000000	5684.000000	1041.000000	5766.000000
Last Name	561.000000	2236.000000	9785.000000	3792.000000	13487.000000
Country	92.000000	42.000000	92.000000	42.000000	51.000000
Car Brand	24.000000	10.000000	47.000000	4.000000	54.000000
Car Model	237.000000	347.000000	586.000000	106.000000	575.000000
Car Color	15.000000	3.000000	10.000000	10.000000	11.000000
Year of Manufacture	0.807692	0.942308	0.961538	0.923077	0.817308
Credit Card Type	11.000000	11.000000	10.000000	7.000000	10.000000

In [17]:
```
df.to_csv('new_processed_cars_all_numeric.csv', index=False)
```