

project content

- find the null value if there is any fill with the mean of that column
- check how many type of name in company name and count how many time are they mention
- show all records where origine is this and this
- remove all records where weight is == or any condition may be there
- increase all the value of the this column by the value of threee

```
In [1]: import pandas as pd
df = pd.read_csv(r"C:\Users\PC-chetan\Downloads\car_data.csv")
df
```

```
Out[1]:
```

| | Make | Model | Type | Origin | DriveTrain | MSRP | Invoice | EngineSize | Cylinders | Horsepower |
|-----|-------|-------------------------|-------|--------|------------|----------|----------|------------|-----------|------------|
| 0 | Acura | MDX | SUV | Asia | All | \$36,945 | \$33,337 | 3.5 | 6.0 | |
| 1 | Acura | RSX Type S 2dr | Sedan | Asia | Front | \$23,820 | \$21,761 | 2.0 | 4.0 | |
| 2 | Acura | TSX 4dr | Sedan | Asia | Front | \$26,990 | \$24,647 | 2.4 | 4.0 | |
| 3 | Acura | TL 4dr | Sedan | Asia | Front | \$33,195 | \$30,299 | 3.2 | 6.0 | |
| 4 | Acura | 3.5 RL 4dr | Sedan | Asia | Front | \$43,755 | \$39,014 | 3.5 | 6.0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 423 | Volvo | C70 LPT convertible 2dr | Sedan | Europe | Front | \$40,565 | \$38,203 | 2.4 | 5.0 | |
| 424 | Volvo | C70 HPT convertible 2dr | Sedan | Europe | Front | \$42,565 | \$40,083 | 2.3 | 5.0 | |
| 425 | Volvo | S80 T6 4dr | Sedan | Europe | Front | \$45,210 | \$42,573 | 2.9 | 6.0 | |
| 426 | Volvo | V40 | Wagon | Europe | Front | \$26,135 | \$24,641 | 1.9 | 4.0 | |
| 427 | Volvo | XC70 | Wagon | Europe | All | \$35,145 | \$33,112 | 2.5 | 5.0 | |

428 rows × 15 columns

data cleaning

```
In [2]: df.iloc[:,1:]
```

```
Out[2]:
```

| | Model | Type | Origin | DriveTrain | MSRP | Invoice | EngineSize | Cylinders | Horsepower | M |
|-----|----------------|-------|--------|------------|----------|----------|------------|-----------|------------|-----|
| 0 | MDX | SUV | Asia | All | \$36,945 | \$33,337 | 3.5 | 6.0 | 265 | |
| 1 | RSX Type S 2dr | Sedan | Asia | Front | \$23,820 | \$21,761 | 2.0 | 4.0 | 200 | |
| 2 | TSX 4dr | Sedan | Asia | Front | \$26,990 | \$24,647 | 2.4 | 4.0 | 200 | |
| 3 | TL 4dr | Sedan | Asia | Front | \$33,195 | \$30,299 | 3.2 | 6.0 | 270 | |
| 4 | 3.5 RL 4dr | Sedan | Asia | Front | \$43,755 | \$39,014 | 3.5 | 6.0 | 225 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

| | Model | Type | Origin | DriveTrain | MSRP | Invoice | EngineSize | Cylinders | Horsepower | M |
|-----|-------------------------|-------|--------|------------|----------|----------|------------|-----------|------------|---|
| 423 | C70 LPT convertible 2dr | Sedan | Europe | Front | \$40,565 | \$38,203 | 2.4 | 5.0 | 197 | |
| 424 | C70 HPT convertible 2dr | Sedan | Europe | Front | \$42,565 | \$40,083 | 2.3 | 5.0 | 242 | |
| 425 | S80 T6 4dr | Sedan | Europe | Front | \$45,210 | \$42,573 | 2.9 | 6.0 | 268 | |
| 426 | V40 | Wagon | Europe | Front | \$26,135 | \$24,641 | 1.9 | 4.0 | 170 | |
| 427 | XC70 | Wagon | Europe | All | \$35,145 | \$33,112 | 2.5 | 5.0 | 208 | |

428 rows × 14 columns

In [3]: `df.iloc[:,0]`

Out[3]:

```

0      Acura
1      Acura
2      Acura
3      Acura
4      Acura
...
423    Volvo
424    Volvo
425    Volvo
426    Volvo
427    Volvo
Name: Make, Length: 428, dtype: object

```

In [23]: `df.info()`

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 428 entries, 0 to 427
Data columns (total 15 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Make            428 non-null   object
1   Model           428 non-null   object
2   Type            428 non-null   object
3   Origin          428 non-null   object
4   DriveTrain      428 non-null   object
5   MSRP            428 non-null   object
6   Invoice          428 non-null   object
7   EngineSize      428 non-null   float64
8   Cylinders       426 non-null   float64
9   Horsepower      428 non-null   int64
10  MPG_City        428 non-null   int64
11  MPG_Highway     428 non-null   int64
12  Weight          428 non-null   int64
13  Wheelbase       428 non-null   int64
14  Length          428 non-null   int64
dtypes: float64(2), int64(6), object(7)
memory usage: 50.3+ KB

```

In [27]: `df.isnull().sum()`

Out[27]:

```

Make      0
Model     0
Type      0

```

```
Origin      0
DriveTrain  0
MSRP        0
Invoice     0
EngineSize  0
Cylinders   2
Horsepower  0
MPG_City    0
MPG_Highway 0
Weight      0
Wheelbase   0
Length      0
dtype: int64
```

```
In [32]: df['Cylinders'].mean()
```

```
Out[32]: 5.807511737089202
```

```
In [35]: df.Cylinders.fillna(df['Cylinders'].mean(),inplace=True)
```

```
In [36]: df.isnull().sum()
```

```
Out[36]: Make      0
Model    0
Type      0
Origin    0
DriveTrain 0
MSRP      0
Invoice   0
EngineSize 0
Cylinders 0
Horsepower 0
MPG_City   0
MPG_Highway 0
Weight     0
Wheelbase  0
Length     0
dtype: int64
```

```
In [37]: df.head()
```

```
Out[37]:
```

| | Make | Model | Type | Origin | DriveTrain | MSRP | Invoice | EngineSize | Cylinders | Horsepower | M |
|---|-------|----------------------|-------|--------|------------|----------|----------|------------|-----------|------------|---|
| 0 | Acura | MDX | SUV | Asia | All | \$36,945 | \$33,337 | 3.5 | 6.0 | 265 | |
| 1 | Acura | RSX Type S 2dr | Sedan | Asia | Front | \$23,820 | \$21,761 | 2.0 | 4.0 | 200 | |
| 2 | Acura | TSX 4dr | Sedan | Asia | Front | \$26,990 | \$24,647 | 2.4 | 4.0 | 200 | |
| 3 | Acura | TL 4dr | Sedan | Asia | Front | \$33,195 | \$30,299 | 3.2 | 6.0 | 270 | |
| 4 | Acura | 3.5 RL 4dr | Sedan | Asia | Front | \$43,755 | \$39,014 | 3.5 | 6.0 | 225 | |

```
In [38]: df.Make.value_counts()
```

```
Out[38]: Toyota      28
```

Mercedes-Benz 26
Ford 23
BMW 20
Audi 19
Honda 17
Nissan 17
Volkswagen 15
Chrysler 15
Dodge 13
Mitsubishi 13
Volvo 12
Jaguar 12
Hyundai 12
Subaru 11
Pontiac 11
Mazda 11
Lexus 11
Kia 11
Buick 9
Mercury 9
Lincoln 9
Saturn 8
Cadillac 8
Suzuki 8
Infiniti 8
GMC 8
Acura 7
Porsche 7
Saab 7
Land Rover 3
Oldsmobile 3
Jeep 3
Scion 2
Isuzu 2
MINI 2
Hummer 1
Name: Make, dtype: int64

In [39]: df.head()

Out[39]:

| | Make | Model | Type | Origin | DriveTrain | MSRP | Invoice | EngineSize | Cylinders | Horsepower | M |
|---|-------|----------------|-------|--------|------------|----------|----------|------------|-----------|------------|---|
| 0 | Acura | MDX | SUV | Asia | All | \$36,945 | \$33,337 | 3.5 | 6.0 | 265 | |
| 1 | Acura | RSX Type S 2dr | Sedan | Asia | Front | \$23,820 | \$21,761 | 2.0 | 4.0 | 200 | |
| 2 | Acura | TSX 4dr | Sedan | Asia | Front | \$26,990 | \$24,647 | 2.4 | 4.0 | 200 | |
| 3 | Acura | TL 4dr | Sedan | Asia | Front | \$33,195 | \$30,299 | 3.2 | 6.0 | 270 | |
| 4 | Acura | 3.5 RL 4dr | Sedan | Asia | Front | \$43,755 | \$39,014 | 3.5 | 6.0 | 225 | |

In [42]: df[df['Origin'].isin(['Asia', 'Europe'])]

Out[42]:

| | Make | Model | Type | Origin | DriveTrain | MSRP | Invoice | EngineSize | Cylinders | Horsepower | M |
|---|-------|----------------|-------|--------|------------|----------|----------|------------|-----------|------------|---|
| 0 | Acura | MDX | SUV | Asia | All | \$36,945 | \$33,337 | 3.5 | 6.0 | 265 | |
| 1 | Acura | RSX Type S 2dr | Sedan | Asia | Front | \$23,820 | \$21,761 | 2.0 | 4.0 | 200 | |

| | Make | Model | Type | Origin | DriveTrain | MSRP | Invoice | EngineSize | Cylinders | Horsepower |
|-----|-------|-------------------------|-------|--------|------------|----------|----------|------------|-----------|------------|
| 2 | Acura | TSX 4dr | Sedan | Asia | Front | \$26,990 | \$24,647 | 2.4 | 4.0 | |
| 3 | Acura | TL 4dr | Sedan | Asia | Front | \$33,195 | \$30,299 | 3.2 | 6.0 | |
| 4 | Acura | 3.5 RL 4dr | Sedan | Asia | Front | \$43,755 | \$39,014 | 3.5 | 6.0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 423 | Volvo | C70 LPT convertible 2dr | Sedan | Europe | Front | \$40,565 | \$38,203 | 2.4 | 5.0 | |
| 424 | Volvo | C70 HPT convertible 2dr | Sedan | Europe | Front | \$42,565 | \$40,083 | 2.3 | 5.0 | |
| 425 | Volvo | S80 T6 4dr | Sedan | Europe | Front | \$45,210 | \$42,573 | 2.9 | 6.0 | |
| 426 | Volvo | V40 | Wagon | Europe | Front | \$26,135 | \$24,641 | 1.9 | 4.0 | |
| 427 | Volvo | XC70 | Wagon | Europe | All | \$35,145 | \$33,112 | 2.5 | 5.0 | |

281 rows × 15 columns

In [49]: `df[~(df['Weight'] > 4000)]`

Out[49]:

| | Make | Model | Type | Origin | DriveTrain | MSRP | Invoice | EngineSize | Cylinders | Horsepower |
|-----|-------|-------------------------|-------|--------|------------|----------|----------|------------|-----------|------------|
| 1 | Acura | RSX Type S 2dr | Sedan | Asia | Front | \$23,820 | \$21,761 | 2.0 | 4.0 | |
| 2 | Acura | TSX 4dr | Sedan | Asia | Front | \$26,990 | \$24,647 | 2.4 | 4.0 | |
| 3 | Acura | TL 4dr | Sedan | Asia | Front | \$33,195 | \$30,299 | 3.2 | 6.0 | |
| 4 | Acura | 3.5 RL 4dr | Sedan | Asia | Front | \$43,755 | \$39,014 | 3.5 | 6.0 | |
| 5 | Acura | 3.5 RL w/Navigation 4dr | Sedan | Asia | Front | \$46,100 | \$41,100 | 3.5 | 6.0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 423 | Volvo | C70 LPT convertible 2dr | Sedan | Europe | Front | \$40,565 | \$38,203 | 2.4 | 5.0 | |
| 424 | Volvo | C70 HPT convertible 2dr | Sedan | Europe | Front | \$42,565 | \$40,083 | 2.3 | 5.0 | |
| 425 | Volvo | S80 T6 4dr | Sedan | Europe | Front | \$45,210 | \$42,573 | 2.9 | 6.0 | |
| 426 | Volvo | V40 | Wagon | Europe | Front | \$26,135 | \$24,641 | 1.9 | 4.0 | |
| 427 | Volvo | XC70 | Wagon | Europe | All | \$35,145 | \$33,112 | 2.5 | 5.0 | |

325 rows × 15 columns

In [45]: `df.shape`

Out[45]: (428, 15)

In [4]: `df['MPG_City'] = df['MPG_City'].apply(lambda x:x+3)`

In [53]: df

Out[53]:

| | Make | Model | Type | Origin | DriveTrain | MSRP | Invoice | EngineSize | Cylinders | Horsepower |
|-----|-------|-------------------------|-------|--------|------------|----------|----------|------------|-----------|------------|
| 0 | Acura | MDX | SUV | Asia | All | \$36,945 | \$33,337 | 3.5 | 6.0 | |
| 1 | Acura | RSX Type S 2dr | Sedan | Asia | Front | \$23,820 | \$21,761 | 2.0 | 4.0 | |
| 2 | Acura | TSX 4dr | Sedan | Asia | Front | \$26,990 | \$24,647 | 2.4 | 4.0 | |
| 3 | Acura | TL 4dr | Sedan | Asia | Front | \$33,195 | \$30,299 | 3.2 | 6.0 | |
| 4 | Acura | 3.5 RL 4dr | Sedan | Asia | Front | \$43,755 | \$39,014 | 3.5 | 6.0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 423 | Volvo | C70 LPT convertible 2dr | Sedan | Europe | Front | \$40,565 | \$38,203 | 2.4 | 5.0 | |
| 424 | Volvo | C70 HPT convertible 2dr | Sedan | Europe | Front | \$42,565 | \$40,083 | 2.3 | 5.0 | |
| 425 | Volvo | S80 T6 4dr | Sedan | Europe | Front | \$45,210 | \$42,573 | 2.9 | 6.0 | |
| 426 | Volvo | V40 | Wagon | Europe | Front | \$26,135 | \$24,641 | 1.9 | 4.0 | |
| 427 | Volvo | XC70 | Wagon | Europe | All | \$35,145 | \$33,112 | 2.5 | 5.0 | |

428 rows × 15 columns

In [5]: df

Out[5]:

| | Make | Model | Type | Origin | DriveTrain | MSRP | Invoice | EngineSize | Cylinders | Horsepower |
|-----|-------|-------------------------|-------|--------|------------|----------|----------|------------|-----------|------------|
| 0 | Acura | MDX | SUV | Asia | All | \$36,945 | \$33,337 | 3.5 | 6.0 | |
| 1 | Acura | RSX Type S 2dr | Sedan | Asia | Front | \$23,820 | \$21,761 | 2.0 | 4.0 | |
| 2 | Acura | TSX 4dr | Sedan | Asia | Front | \$26,990 | \$24,647 | 2.4 | 4.0 | |
| 3 | Acura | TL 4dr | Sedan | Asia | Front | \$33,195 | \$30,299 | 3.2 | 6.0 | |
| 4 | Acura | 3.5 RL 4dr | Sedan | Asia | Front | \$43,755 | \$39,014 | 3.5 | 6.0 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 423 | Volvo | C70 LPT convertible 2dr | Sedan | Europe | Front | \$40,565 | \$38,203 | 2.4 | 5.0 | |
| 424 | Volvo | C70 HPT convertible 2dr | Sedan | Europe | Front | \$42,565 | \$40,083 | 2.3 | 5.0 | |
| 425 | Volvo | S80 T6 4dr | Sedan | Europe | Front | \$45,210 | \$42,573 | 2.9 | 6.0 | |
| 426 | Volvo | V40 | Wagon | Europe | Front | \$26,135 | \$24,641 | 1.9 | 4.0 | |
| 427 | Volvo | XC70 | Wagon | Europe | All | \$35,145 | \$33,112 | 2.5 | 5.0 | |

428 rows × 15 columns

In []: