

```

# -*- coding: utf-8 -*-
"""Car_price_prediction.ipynb

Automatically generated by Colab.

Original file is located at
https://colab.research.google.com/drive/1NhOLrHJ9kCj7KVX2VIbM9qZZC\_gVB92E
"""

# Import Libraries:
import pandas as pd

"""1.Calculate aggregate statistics such as mean, median, and standard
deviation of car prices within each group
2.What is the average price of cars for each make and model?
3.Is there a relationship between the year of the car and its price?
4.How does the condition of the car affect its price?
5.What is the correlation between price and numerical features (mileage,
year)?
6.What is the average price of cars for each make and model?
7.Which brands and models have the highest and lowest average prices?
8.How have car prices changed over the years?
9.Is there a noticeable trend in pricing based on the manufacturing year?
10.Is there a correlation between the price of the car and its mileage?
11.Do cars with higher mileage tend to have lower prices?
12.How does the condition of the car (Excellent, Good, Fair) affect its
price?
13.Are there significant differences in prices based on the condition of
the car?
14.Can you compare the average prices of cars from different
manufacturers?
15.How do prices vary between different car models within the same
manufacturer?
16.Calculate summary statistics like mean, median, and standard deviation
of car prices to describe the central tendency and variability of the
data.
"""

# Import Data:
df = pd.read_csv("/content/CarPricesPrediction.csv")

# see the dataset:
df

# see the top 5 heads of data:
df.head(5)

# see the shape of data :
df.shape

# see the columns:
df.columns

# see the size of data:
df.size

# get the information of the data:

```

```

df.info()

# check the null values of the data:
df.isnull().sum() # there is no null values in the dataset

# see the statistical summary of the data:
df.describe

# drop the unnecessary column of the data:
df.drop(columns=['Unnamed: 0'], inplace=True)
#df.drop(columns=['Unnamed: 0'], axis=1)

df.head(2)

# see the datatypes :
df.dtypes

# check the duplicates:
df_no_duplicates = df.drop_duplicates() # there is no duplicates in this
dataset
df.shape

# frequency (count)of each category (make) of car:
df['Make'].value_counts()

# check the count of each category of condition column of data:
df['Condition'].value_counts()

# set both of them together:
df[['Make','Condition']].value_counts().reset_index()

"""**1. How many manufacturers for cars making and their names?**"""

unique_makes = df['Make'].unique()
unique_makes

"""There are 5 unique manufacturer in this data above

**2.how many unique models for cars in data and their names?**
"""

unique_model = df['Model'].unique()
unique_model

"""there are 5 unique models in this data above"""

df['Make'].nunique()

df['Model'].nunique()

"""**3.Which is the maximum and minimum value of manufacturing(Make)of
cars?**"""

df['Make'].max()

"""maximum value of make column or manufacturers is "Toyota"
"""

```

```

df['Make'].min()

"""minimum value of make column or manufacturers is "Chevrolet"

maximum value of make column or manufacturers is "Toyota"

**4.Which is the maximum and minimum value of model of cars or which
model is highest and lowest in frequency ?**
"""

df['Model'].max()

"""Maximum number of cars of model 'Silerado'."""

df['Model'].min()

"""Minimum number of cars of model 'Altima'

**4.What is the average price,median price and standard deviation or
variance of cars for each make and model?**
"""

agg= df.groupby(['Make','Model'])['Price'].agg(['mean','median','std'])
agg

"""Average, median price and sd for checking the variation of cars for
each 'make' and 'model'

**5.What is the average price and other stats of cars for the
'conditions' in the data?**
"""

agg= df.groupby(['Condition'])['Price'].agg(['mean','median','std'])
agg

"""from above, we do groupby 'condition' of cars with price to derived
their mean, median and sd.

**6.What is the average mileage and median & sd of cars for each model?**
"""

agg= df.groupby(['Model'])['Mileage'].agg(['mean','median','std'])
agg

"""from above, we do groupby of model with mileage to find their mean,
median and sd.

**7.What is the average price of cars for each make and model?**
"""

avg_price_cars= df.groupby(['Make','Model'])['Price'].agg(['mean'])
avg_price_cars.sort_values(by= 'mean')

"""average price of cars for each 'make' and 'model', highest mean of
make is "Honda" with model "Camry" and lowest mean of 'make' is 'Honda'
and 'model' 'silverado'.

**8.What is the average mileage of cars for each make and model?**

```

```
"""
```

```
avg_mileage_cars= df.groupby(['Make','Model'])['Mileage'].agg(['mean'])
avg_mileage_cars.sort_values(by='mean')
```

```
"""average mileage of cars for each 'make' and 'model', highest mean of
make is "Toyota" with model "Camry" and lowest mean of 'make' is 'Ford'
and 'model' 'Civic'.
```

```
**9.Which brands and models have the highest and lowest average prices?
***
```

```
"""
```

```
avg_price_cars_mod= df.groupby('Model')['Price'].mean().sort_values()
avg_price_cars_mod
```

```
avg_price_cars_mod.idxmax()
```

```
avg_price_cars_mod.idxmin()
```

```
"""The Highest average price of model is F-150.
The Lowest average price of model is Altima.
```

```
"""
```

```
avg_price_cars_mak= df.groupby('Make')['Price'].mean().sort_values()
avg_price_cars_mak
```

```
avg_price_cars_mak.idxmin()
```

```
avg_price_cars_mak.idxmax()
```

```
"""The Highest average price of make is "Nissan".
The Lowest average price of make is "Toyota".
```

```
**10.How many cars manufactures(Make) in a year?**
"""
```

```
year = 2014
filtered_df = df[df['Year']==year]
car_makes = filtered_df['Make'].unique()
car_makes
```

```
"""from above, in the year 2014, manufacturers makes number of cars in
the year 2014 is five (make) which are, 'Toyota', 'Ford', 'Chevrolet',
'Honda', 'Nissan'."""
```

```
year==2016
manufacturers_in_year = df[df['Year'] == year]['Make'].nunique()
manufacturers_in_year
```

```
"""from above, in the year 2016, manufacturers makes number of cars in
the year 2014 is five (make) 5.
```

```
**11.How many types of model in a year 2016 & 2019?**
"""
```

```
filtered_2016 = df[df['Year']==2016]
```

```

filtered_2019 = df[df['Year']==2019]
car_model_2016 = filtered_2016['Model'].unique()
car_model_2019= filtered_2019['Model'].unique()
car_model_2016
car_model_2019

"""from above , types of model in a year 2016 & 2019 are Civic',
'Altima', 'Silverado', 'Camry', 'F-150'.

**12. What are categories of car conditions, mention with their names?**
"""

car_cond =df['Condition'].unique()
car_cond

"""The categories of car conditions are 'Excellent', 'Good', 'Fair'.

**13.What types of cars from which manufacturer where mileage is greater
than 13000 and price is less than 27800?**
"""

df[(df['Mileage'] >13000) & (df['Price'] < 27800)]

"""these are the cars from the manufacturers where mileage is greater
than 13000 and price is less than 27800.

1.Is there a relationship between the year of the car and its price?
2.How does the condition of the car affect its price?
3.What is the correlation between price and numerical features (mileage,
year)?
4.What is the average price of cars for each make and model?
5.Which brands and models have the highest and lowest average prices?
6.How have car prices changed over the years?
7.Is there a noticeable trend in pricing based on the manufacturing year?
8.Is there a correlation between the price of the car and its mileage?
9.Do cars with higher mileage tend to have lower prices?
10.How does the condition of the car (Excellent, Good, Fair) affect its
price?
11.Are there significant differences in prices based on the condition of
the car?
12.Can you compare the average prices of cars from different
manufacturers?
13.How do prices vary between different car models within the same
manufacturer?
14.Calculate summary statistics like mean, median, and standard deviation
of car prices to describe the central tendency and variability of the
data.

**14.How have car prices changed over the years?**
"""

car_p_changed_yrs=df.groupby('Year')['Price'].mean()
car_p_changed_yrs

"""Car prices changes over the years from year 2010 is 27908.86 , then in
year 2011 is 27043.65 , then increasing next 2 years then decreasing for
further some years and so on..

```

```

**15.Is there a noticeable trend in pricing based on the manufacturing
year?**
"""

price_per_change= car_p_changed_yrs.pct_change()*100
price_per_change.reset_index()

price_per_change= car_p_changed_yrs.pct_change()*100
price_per_change

df1= df[['Price','Year']]
df2=price_per_change

merged_df = pd.merge(df1, df2, on='Year')
merged_df

"""there is percentage change in years is fluctuating from year 2011 to
2022.

**16.How does the condition of the car (Excellent, Good, Fair) affect its
price?**
"""

avg_p_by_condition=df.groupby('Condition')['Price'].mean().sort_values(as
cending=False)
avg_p_by_condition

"""from above , the condition affects the price of the cars , fair
condition getting high price however, good condition getting moderate and
excellent condition lower price.

***Are there significant differences in prices based on the condition of
the car? ***ASK

**17.Can you compare the average prices of cars from different
manufacturers? ***
"""

avg_p_make=
df.groupby('Make')['Price'].mean().sort_values(ascending=True)
avg_p_make

"""average price of the cras from different manufacturers , the higher
mean price for make is "Nissan" 22587.50 and the lower mean price for
make is "Toyota" "21881.91"

**18.How do prices vary between different car models with the same
manufacturer? ***
"""

avg_p_vary=
df.groupby(['Make','Model'])['Price'].mean().sort_values(ascending=True)
avg_p_vary

"""from above prices vary between different car models with the same
manufacturer.. make "Honda" model "silverado" has lower prices 21291.34
however,
"Honda" "Camry" has higher price 24099.20

```

```

**19.Calculate summary statistics like mean, median, and standard
deviation of car prices to describe the central tendency and variability
of the data.**
"""

mean_price= df['Price'].mean()
mean_price

"""mean price of price which is 22195.20"""

median_price= df['Price'].median()
median_price

"""median price of price is 22247.875."""

sd_price = df['Price'].std()
sd_price

"""standard deviation or variation of price is 4245.1915

**20.Calculate summary statistics like mean, median, and standard
deviation of car mileage to describe the central tendency and variability
of the data.**
"""

mean_mile=df["Mileage"].mean()
mean_mile

"""mean mileage of mileage is 78796.927"""

median_mile=df["Mileage"].median()
median_mile

"""median mileage of mileage is 78056.5"""

sd_mile=df["Mileage"].std()
sd_mile

"""variation in milaege or sd of milaege is 39842.2599"""

df.columns

corr= df['Mileage'].corr(df['Price'])
corr

corr= df['Year'].corr(df['Price'])
corr

df.head(2)

#from sklearn.preprocessing import LabelEncoder

#label_encoder = LabelEncoder()

#encoded_Model= label_encoder.fit_transform(Model)

```

```
#mean_car_price_by_fuel_type =  
df.groupby('Fuel_Type')['Car_Price'].mean()  
#df['Fuel_Type_Encoded'] =  
df['Fuel_Type'].map(mean_car_price_by_fuel_type)
```