

for data analysis

```
import pandas as pd
import numpy as np
```

for data visualization

```
In [13]: import seaborn as sns
import matplotlib.pyplot as plt

In [12]: df = pd.read_csv("Users/Hp/AppData/Roaming/Microsoft/Windows/Start Menu/Programs/Anaconda3 (64-bit)/Data_Visualization_with_Python_s2-main/vw.csv")
print((number of rows",df.shape[0])
display(df.shape)
print(number of columns",df.shape[1])
display(df)
print(df)

(15157, 8)
number of rows= 15157
number of columns= 8

model year price transmission mileage fuelType mpg engineSize
0 T-Roc 2019 25000 Automatic 13904 Diesel 49.6 2.0
1 T-Roc 2019 26883 Automatic 4562 Diesel 49.6 2.0
2 T-Roc 2019 20000 Manual 7414 Diesel 50.4 2.0
3 T-Roc 2019 33482 Automatic 4825 Petrol 32.5 2.0
4 T-Roc 2019 22900 Semi-Auto 6500 Petrol 39.8 1.5
... ..
15152 Eos 2012 5990 Manual 7400 Diesel 58.9 2.0
15153 Fox 2008 1799 Manual 28102 Petrol 46.3 1.2
15154 Fox 2008 1590 Manual 7000 Petrol 42.0 1.4
15155 Fox 2006 1250 Manual 82704 Petrol 45.3 1.2
15156 Fox 2007 2295 Manual 7400 Petrol 46.3 1.2

15157 rows x 8 columns

In [13]: display(df.head())

model year price transmission mileage fuelType mpg engineSize
0 T-Roc 2019 25000 Automatic 13904 Diesel 49.6 2.0
1 T-Roc 2019 26883 Automatic 4562 Diesel 49.6 2.0
2 T-Roc 2019 20000 Manual 7414 Diesel 50.4 2.0
3 T-Roc 2019 33482 Automatic 4825 Petrol 32.5 2.0
4 T-Roc 2019 22900 Semi-Auto 6500 Petrol 39.8 1.5

value counts()

In [14]: display(df["transmission"].value_counts())

Manual 9417
Semi-Auto 3780
Automatic 1960
Name: transmission, dtype: int64
```

Pie-Chart

```
In [15]: import matplotlib.pyplot as plt
import seaborn as sns
df["transmission"].value_counts().plot(kind="pie", autopct='%1.2f%%', startangle=98)
plt.title("Percentage of Transmission of Cars")
plt.xlabel("")
plt.ylabel("")

Percentage of Transmission of Cars

Manual 62.17%
Automatic 12.93%
Semi-Auto 24.94%
```

Creating DataFrame

```
In [16]: df = df[df["transmission"]].value_counts()
display(df)

transmission
Manual 9417
Semi-Auto 3780
Automatic 1960

In [17]: df = df.reset_index()
df = df.rename(columns={"index": "transmission type", "transmission": "number of cars"})
display(df)

transmission type number of cars
0 Manual 9417
1 Semi-Auto 3780
2 Automatic 1960

In [18]: df["percentage of cars"] = (df["number of cars"]/df.shape[0])*100
display(df)

transmission type number of cars percentage of cars
0 Manual 9417 62.13
1 Semi-Auto 3780 24.94
2 Automatic 1960 12.93
```

Bar-plot

```
In [19]: sns.barplot(x="transmission type", y="percentage of cars", data=df, alpha=0.50, color="green")
plt.title("Percentage of Transmission of Cars")
plt.show()

Percentage of Transmission of Cars

percentage of cars
60
50
40
30
20
10
0
Manual Semi-Auto Automatic
transmission type
```

Comparison

The pie-chart gives a clear indication that manual type holds greater percentage alone than the other two types combined. If needed, percentages can also be shown as annotations in the bar-chart but the pie-chart shows this comparison even without annotations. On the other hand, x and y-labels in the bar-graph presents a more detailed picture than the pie-chart as labels are not used in pie-charts.

So, for a more detailed picture, bars can be helpful while pie-charts are applicable to make visualizations in an easier/faster way

```
In [ ]:

In [20]: def linear_eqn(x, n, c):
xmp = arange(1,11,1)
c1 =
y = m*x+c

df = pd.DataFrame()

df["x"] = x
df["y"] = m*x + c

display(df)

In [21]: df = pd.read_csv("Users/Hp/AppData/Roaming/Microsoft/Windows/Start Menu/Programs/Anaconda3 (64-bit)/Data_Visualization_with_Python_s2-main/vw.csv")
display(df.head(10))

model year price transmission mileage fuelType mpg engineSize
0 T-Roc 2019 25000 Automatic 13904 Diesel 49.6 2.0
1 T-Roc 2019 26883 Automatic 4562 Diesel 49.6 2.0
2 T-Roc 2019 20000 Manual 7414 Diesel 50.4 2.0
3 T-Roc 2019 33482 Automatic 4825 Petrol 32.5 2.0
4 T-Roc 2019 22900 Semi-Auto 6500 Petrol 39.8 1.5
5 T-Roc 2020 31895 Manual 10 Petrol 42.2 1.5
6 T-Roc 2020 27895 Manual 10 Petrol 42.2 1.5
7 T-Roc 2020 38495 Semi-Auto 10 Petrol 32.5 2.0
8 T-Roc 2019 21995 Manual 10 Petrol 44.1 1.0
9 T-Roc 2019 23285 Manual 10 Petrol 42.2 1.5
```

```
In [27]: plt.figure(figsize=(18,14))

plt.subplot(2, 1, 1)
sns.scatterplot(x="year", y="price", data=df)

plt.subplot(2, 1, 2)
sns.scatterplot(x="engineSize", y="price", data=df)
plt.tight_layout()
plt.show()

price
70000
60000
50000
40000
30000
20000
10000
0
2000.0 2002.5 2005.0 2007.5 2010.0 2012.5 2015.0 2017.5 2020.0
year

price
70000
60000
50000
40000
30000
20000
10000
0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
engineSize
```

In the price Vs. year scatter plot, it is following a certain trend as the price rises with the passing of years. But the plot doesn't show any distinct relationship between engineSize and price. Price sometimes rises, sometimes stays constant, even sometimes falls as the increase in engine size.

```
In [39]: plt.figure(figsize=(18,10))

plt.subplot(1, 2, 1)
sns.regplot(x="mileage", y="price", data=df, scatter_kws={"color": "orange", "edgecolor": "white"})

plt.subplot(1, 2, 2)
sns.regplot(x="year", y="price", data=df, line_kws={"color": "red", "scatter_kws": {"edgecolor": "white"}})

plt.tight_layout()
plt.show()

price
60000
40000
20000
0
-20000
0 50000 100000 150000 200000
mileage

price
60000
40000
20000
0
-20000
2000.0 2002.5 2005.0 2007.5 2010.0 2012.5 2015.0 2017.5 2020.0
year
```

The regression plot also helps us to understand the underlying relation between dependent and independent variables, here, between mileage and price and the year and price. As the graph indicates, price falls with the increase in mileage and rises with the passing of year. The trendline helps the audience to understand the trend.

```
In [42]: penguins = sns.load_dataset("penguins")
display(penguins.head())
print(penguins.shape)

species island bill_length_mm bill_depth_mm flipper_length_mm body_mass_g sex
0 Adelle Torgersen 39.1 18.7 181.0 3750.0 Male
1 Adelle Torgersen 41.3 18.0 186.0 3200.0 Female
2 Adelle Torgersen NaN NaN NaN NaN NaN
3 Adelle Torgersen NaN NaN NaN NaN NaN
4 Adelle Torgersen 36.7 19.3 193.0 3450.0 Female
(344, 7)

In [37]: sns.pairplot(penguins)
plt.show()

bill_length_mm
50
40
30
20
10
0
0 40 80 120 160 200 240 280 320 360 400 440 480 520
bill_depth_mm

flipper_length_mm
250
200
150
100
50
0
0 40 80 120 160 200 240 280 320 360 400 440 480 520
bill_depth_mm

body_mass_g
6000
5000
4000
3000
2000
1000
0
0 40 80 120 160 200 240 280 320 360 400 440 480 520
bill_depth_mm
```

```
In [45]: sns.pairplot(penguins, hue="species")
plt.show()

bill_length_mm
60
50
40
30
20
10
0
0 40 80 120 160 200 240 280 320 360 400 440 480 520
bill_depth_mm

flipper_length_mm
250
200
150
100
50
0
0 40 80 120 160 200 240 280 320 360 400 440 480 520
bill_depth_mm

body_mass_g
6000
5000
4000
3000
2000
1000
0
0 40 80 120 160 200 240 280 320 360 400 440 480 520
bill_depth_mm

species
Adelle
Chinstrap
Gentoo
```

Pair-plot helps to visualize multiple inputs in a single plot. This also helps to compare those inputs among themselves. For example, here, it is shown that in some cases, Chinstrap and Adelle show same characteristics as blue and orange dots are overlapping while the green dots are showing distant characters. In most cases, relation between x and y axis is positive. Adding a categorical column as hue was important as different color points indicate different inputs.

```
In [60]: plt.figure(figsize=(8,5))
sns.boxplot(x="price", data=df, showfliers=False)
plt.show()

price
35000
30000
25000
20000
15000
10000
5000
0
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420 440 460 480 500 520 540 560 580 600 620 640 660 680 700 720 740 760 780 800 820 840 860 880 900 920 940 960 980 1000 1020 1040 1060 1080 1100 1120 1140 1160 1180 1200 1220 1240 1260 1280 1300 1320 1340 1360 1380 1400 1420 1440 1460 1480 1500 1520 1540 1560 1580 1600 1620 1640 1660 1680 1700 1720 1740 1760 1780 1800 1820 1840 1860 1880 1900 1920 1940 1960 1980 2000 2020 2040 2060 2080 2100 2120 2140 2160 2180 2200 2220 2240 2260 2280 2300 2320 2340 2360 2380 2400 2420 2440 2460 2480 2500 2520 2540 2560 2580 2600 2620 2640 2660 2680 2700 2720 2740 2760 2780 2800 2820 2840 2860 2880 2900 2920 2940 2960 2980 3000 3020 3040 3060 3080 3100 3120 3140 3160 3180 3200 3220 3240 3260 3280 3300 3320 3340 3360 3380 3400 3420 3440 3460 3480 3500 3520 3540 3560 3580 3600 3620 3640 3660 3680 3700 3720 3740 3760 3780 3800 3820 3840 3860 3880 3900 3920 3940 3960 3980 4000 4020 4040 4060 4080 4100 4120 4140 4160 4180 4200 4220 4240 4260 4280 4300 4320 4340 4360 4380 4400 4420 4440 4460 4480 4500 4520 4540 4560 4580 4600 4620 4640 4660 4680 4700 4720 4740 4760 4780 4800 4820 4840 4860 4880 4900 4920 4940 4960 4980 5000 5020 5040 5060 5080 5100 5120 5140 5160 5180 5200 5220 5240 5260 5280 5300 5320 5340 5360 5380 5400 5420 5440 5460 5480 5500 5520 5540 5560 5580 5600 5620 5640 5660 5680 5700 5720 5740 5760 5780 5800 5820 5840 5860 5880 5900 5920 5940 5960 5980 6000 6020 6040 6060 6080 6100 6120 6140 6160 6180 6200 6220 6240 6260 6280 6300 6320 6340 6360 6380 6400 6420 6440 6460 6480 6500 6520 6540 6560 6580 6600 6620 6640 6660 6680 6700 6720 6740 6760 6780 6800 6820 6840 6860 6880 6900 6920 6940 6960 6980 7000 7020 7040 7060 7080 7100 7120 7140 7160 7180 7200 7220 7240 7260 7280 7300 7320 7340 7360 7380 7400 7420 7440 7460 7480 7500 7520 7540 7560 7580 7600 7620 7640 7660 7680 7700 7720 7740 7760 7780 7800 7820 7840 7860 7880 7900 7920 7940 7960 7980 8000 8020 8040 8060 8080 8100 8120 8140 8160 8180 8200 8220 8240 8260 8280 8300 8320 8340 8360 8380 8400 8420 8440 8460 8480 8500 8520 8540 8560 8580 8600 8620 8640 8660 8680 8700 8720 8740 8760 8780 8800 8820 8840 8860 8880 8900 8920 8940 8960 8980 9000 9020 9040 9060 9080 9100 9120 9140 9160 9180 9200 9220 9240 9260 9280 9300 9320 9340 9360 9380 9400 9420 9440 9460 9480 9500 9520 9540 9560 9580 9600 9620 9640 9660 9680 9700 9720 9740 9760 9780 9800 9820 9840 9860 9880 9900 9920 9940 9960 9980 10000 10020 10040 10060 10080 10100 10120 10140 10160 10180 10200 10220 10240 10260 10280 10300 10320 10340 10360 10380 10400 10420 10440 10460 10480 10500 10520 10540 10560 10580 10600 10620 10640 10660 10680 10700 10720 10740 10760 10780 10800 10820 10840 10860 10880 10900 10920 10940 10960 10980 11000 11020 11040 11060 11080 11100 11120 11140 11160 11180 11200 11220 11240 11260 11280 11300 11320 11340 11360 11380 11400 11420 11440 11460 11480 11500 11520 11540 11560 11580 11600 11620 11640 11660 11680 11700 11720 11740 11760 11780 11800 11820 11840 11860 11880 11900 11920 11940 11960 11980 12000 12020 12040 12060 12080 12100 12120 12140 12160 12180 12200 12220 12240 12260 12280 12300 12320 12340 12360 12380 12400 12420 12440 12460 12480 12500 12520 12540 12560 12580 12600 12620 12640 12660 12680 12700 12720 12740 12760 12780 12800 12820 12840 12860 12880 12900 12920 12940 12960 12980 13000 13020 13040 13060 13080 13100 13120 13140 13160 13180 13200 13220 13240 13260 13280 13300 13320 13340 13360 13380 13400 13420 13440 13460 13480 13500 13520 13540 13560 13580 13600 13620 13640 13660 13680 13700 13720 13740 13760 13780 13800 13820 13840 13860 13880 13900 13920 13940 13960 13980 14000 14020 14040 14060 14080 14100 14120 14140 14160 14180 14200 14220 14240 14260 14280 14300 14320 14340 14360 14380 14400 14420 14440 14460 14480 14500 14520 14540 14560 14580 14600 14620 14640 14660 14680 14700 14720 14740 14760 14780 14800 14820 14840 14860 14880 14900 14920 14940 14960 14980 15000 15020 15040 15060 15080 15100 15120 15140 15160 15180 15200 15220 15240 15260 15280 15300 15320 15340 15360 15380 15400 15420 15440 15460 15480 15500 15520 15540 15560 15580 15600 15620 15640 15660 15680 15700 15720 15740 15760 15780 15800 15820 15840 15860 15880 15900 15920 15940 15960 15980 16000 16020 16040 16060 16080 16100 16120 16140 16160 16180 16200 16220 16240 16260 16280 16300 16320 16340 16360 16380 16400 16420 16440 16460 16480 16500 16520 16540 16560 16580 16600 16620 16640 16660 16680 16700 16720 16740 16760 16780 16800 16820 16840 16860 16880 16900 16920 16940 16960 16980 17000 17020 17040 17060 17080 17100 17120 17140 17160 17180 17200 17220 17240 17260 17280 17300 17320 17340 17360 17380 17400 17420 17440 17460 17480 17500 17520 17540 17560 17580 17600 17620 17640 17660 17680 17700 17720 17740 17760 17780 17800 17820 17840 17860 17880 17900 17920 17940 17960 17980 18000 18020 18040 18060 18080 18100 18120 18140 18160 18180 18200 18220 18240 18260 18280 18300 18320 18340 18360 18380 18400 18420 18440 18460 18480 18500 18520 18540 18560 18580 18600 18620 18640 18660 18680 18700 18720 18740 18760 18780 18800 18820 18840 18860 18880 18900 18920 18940 18960 18980 19000 19020 19040 19060 19080 19100 19120 19140 19160 19180 19200 19220 19240 19260 19280 19300 19320 19340 19360 19380 19400 19420 19440 19460 19480 19500 19520 19540 19560 19580 19600 19620 19640 19660 19680 19700 19720 19740 19760 19780 19800 19820 19840 19860 19880 19900 19920 19940 19960 19980 20000 20020 20040 20060 20080 20100 20120 20140 20160 20180 20200 20220 20240 20260 20280 20300 20320 20340 20360 20380 20400 20420 20440 20460 20480 20500 20520 20540 20560 20580 20600 20620 20640 20660 20680 20700 20720 20740 20760 20780 20800 20820 20840 20860 20880 20900 20920 20940 20960 20980 21000 21020 21040 21060 21080 21100 21120 21140 21160 21180 21200 21220 21240 21260 21280 21300 21320 21340 21360 21380 21400 21420 21440 21460 21480 21500 21520 21540 21560 21580 21600 21620 21640 21660 21680 21700 21720 21740 21760 21780 21800 21820 21840 21860 21880 21900 21920 21940 21960 21980 22000 22020 22040 22060 22080 22100 22120 22140 22160 22180 22200 22220 22240 22260 22280 22300 22320 22340 22360 22380 22400 22420 22440 22460 22480 22500 22520 22540 22560 22580 22600 22620 22640 22660 22680 22700 22720 22740 22760 22780 22800 22820 22840 22860 22880 22900 22920 22940 22960 22980 23000 23020 23040 23060 23080 23100 23120 23140 23160 23180 23200 23220 23240 23260 23280 23300 23320 23340 23360 23380 23400 23420 23440 23460 23480 23500 23520 23540 23560 23580 23600 23620 23640 23660 23680 23700 23720 23740 23760 23780 23800 23820 23840 23860 23880 23900 23920 23940 23960 23980 24000 24020 24040 24060 24080 24100 24120 24140 24160 24180 24200 24220 24240 24260 24280 24300 24320 24340 24360 24380 24400 24420 24440 24460 24480 24500 24520 24540 24560 24580 24600 24620 24640 24660 24680 24700 24720 24740 24760 24780 24800 24820 24840 24860 24880 24900 24920 24940 24960 24980 25000 25020 25040 25060 25080 25100 25120 25140 25160 25180 25200 25220 25240 25260 25280 25300 25320 25340 25360 25380 25400 25420 25440 25460 25480 25500 25520 25540 25560 25580 25600 25620 25640 25660 25680 25700 25720 25740 25760 25780 25800 25820 25840 25860 25880 25900 25920 25940 25960 25980 26000 26020 26040 26060 26080 26100 26120 26140 26160 26180 26200 26220 26240 26260 26280 26300 26320 26340 26360 26380 26400 26420 26440 26460 26480 26500 26520 26540 26560 26580 26600 26620 26640 26660 26680 26700 26720 26740 26760 26780 26800 26820 26840 26860 26880 26900 26920 26940 26960 26980 27000 27020 27040 27060 27080 27100 27120 27140 27160 27180 27200 27220 27240 27260 27280 27300 27320 27340 27360 27380 27400 27420 27440 27460 27480 27500 27520 27540 27560 27580 27600 27620 27640 27660 27680 27700 27720 27740 27760 27780 27800 27820 27840 27860 27880 27900 27920 27940 27960 27980 28000 28020 28040 28060 28080 28100 28120 28140 28160 28180 28200 28220 28240 28260 28280 28300 28320 28340 28360 28380 28400 28420 28440 28460 28480 28500 28520 28540 28560 28580 28600 28620 28640 28660 28680 28700 28720 28740 28760 28780 28800 28820 28840 28860 28880 28900 28920 28940 28960 28980 29000 29020 29040 29060 29080 29100 29120 29140 29160 29180 29200 29220 29240 29260 29280 29300 29320 29340 29360 29380 29400 29420 29440 29460 29480 29500 29520 29540 29560 29580 29600 29620 29640 29660 29680 29700 29720 29740 29760 29780 29800 29820 29840 29860 29880 29900 29920 29940 
```