German Car Analysis

- Niharika Singh

DATA DESCRIPTION

- Source dataset contains data from one of **Europe's largest car market AutoScaut24 from 2011-21**
- Original dataset consists of **46405 rows** and **9 columns**
- Fields contain information about **make, model, price, offer type, mileage, fuel, gear type, horsepower** and **production year.**
- Data was downloaded from Kaggle.com Collection method – scraping.

  Link - https://www.kaggle.com/datasets/ander289386/cars-germany

OBJECTIVES

Data cleaning	revealing gaps and inaccuracies in the data that can interfere with the analysis and then eliminating them.
Data Analysis	Use Pandas, NumPy, and other Python libraries for investigating our data and searching for interesting information and insights
Statistical Analysis	Using Pandas functions to obtain statistical information about data and identify relationships between variables.
Discovering insights	Discovery of interesting points that are worth paying attention to and on which it is possible to conduct further research.
Graphical representation	Using the Python Matplotlib library to represent findings using various types of visualization

CLEANING

1. *Checking data for problems, inconsistencies*
2. *Checking and Changing data formats where necessary*
3. *Replacing, dropping null values in columns*
4. *Replacing incorrect values*
5. *Verification of results*

DATASET STATS

Price has a strong **positive** correlation with HP. Mileage was **inversely** related to year. Gear type is also correlated with the price.

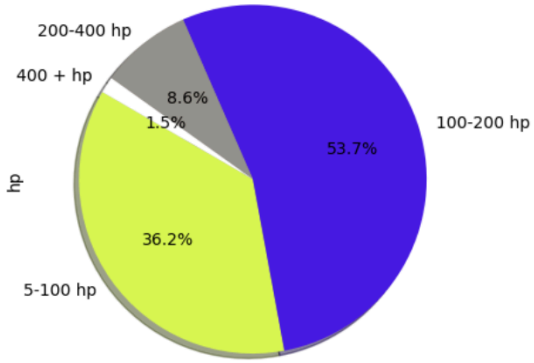
	mileage	price	hp	year
mileage	1.00	-0.30	-0.01	-0.68
price	-0.30	1.00	0.75	0.41
hp	-0.01	0.75	1.00	0.17
year	-0.68	0.41	0.17	1.00

Correlation coefficients

Descriptive statistics

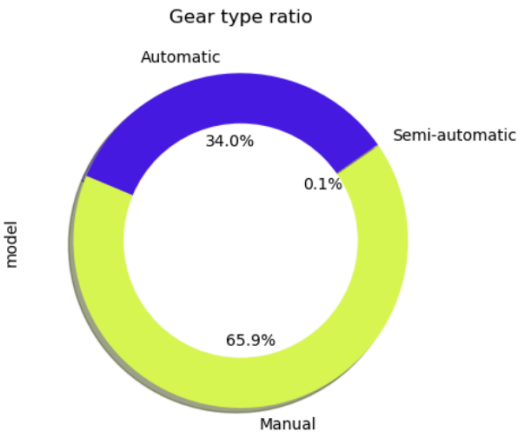
	mileage	price	hp
count	46262.00	46262.00	46262.00
mean	71259.64	16544.75	132.89
std	62635.20	19253.64	75.01
min	0.00	1100.00	5.00
25%	19900.00	7490.00	86.00
50%	60000.00	10999.00	116.00
75%	105000.00	19483.75	150.00
max	1111111.00	1199900.00	850.00

HORSEPOWER



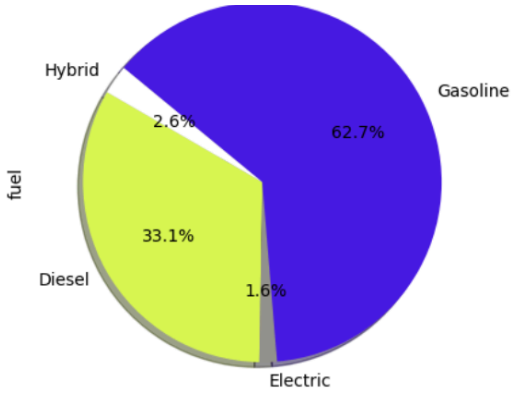
- The lowest value of horsepower – **5 hp.** Microcar(Due, M.Go, M8), Renault Twizy, Ligier JS 50, Estrima Biro
- Mercedes-benz G63 AMG has the **highest 850** horsepower value
- Almost **90 %** of all offers have **less** than **200** horsepower

TRANSMISSION



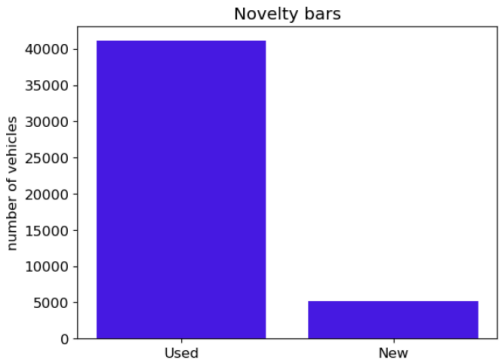
- Most of the offers have Manual gear type – almost **30.50**0 cars
- **15.714** cars on Autoscaut24 have Automatic transmission
- Only **56** offers have Semi-automatic gearbox

FUEL TYPES



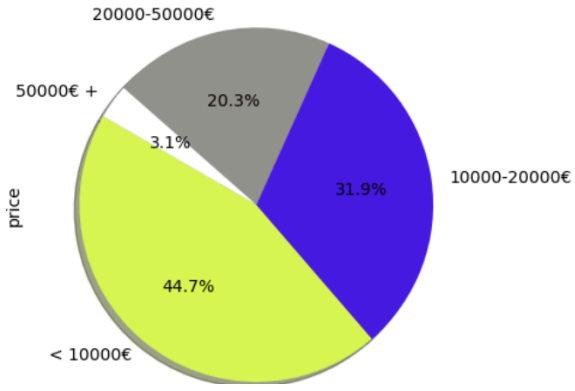
- The most of cars – 28.869 use Gasoline fuel type
- **Diesel – 15.220, Hybrid – 1.203, Electric – 725 cars.**
- Gas vehicles are the least represented - 0.5% of the total number or 245 vehicles

NEW OR USED?



- The dataset includes five different types of offers: pre-registered, used, demonstration, employee's car, and new.
- **Most of the vehicles (41.116) represented on the website are used. 5.146 of the cars are new.**

PRICE CATEGORIES

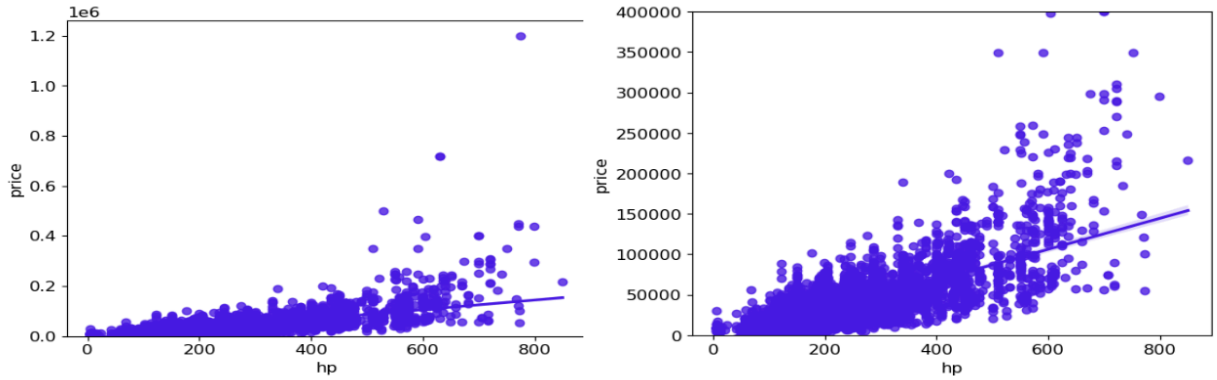


- TThe most of cars – 20.664 are cheaper than 10.000€.
- 76.6%, or 35.434 offers, have price less than 20.000€, and 96.9% or 44805 are priced under 50.000€.
- 1.457 vehicles cost more than €50.000, with the most expensive costing 1.199.900€.
- The cheapest car priced at 1.100€.

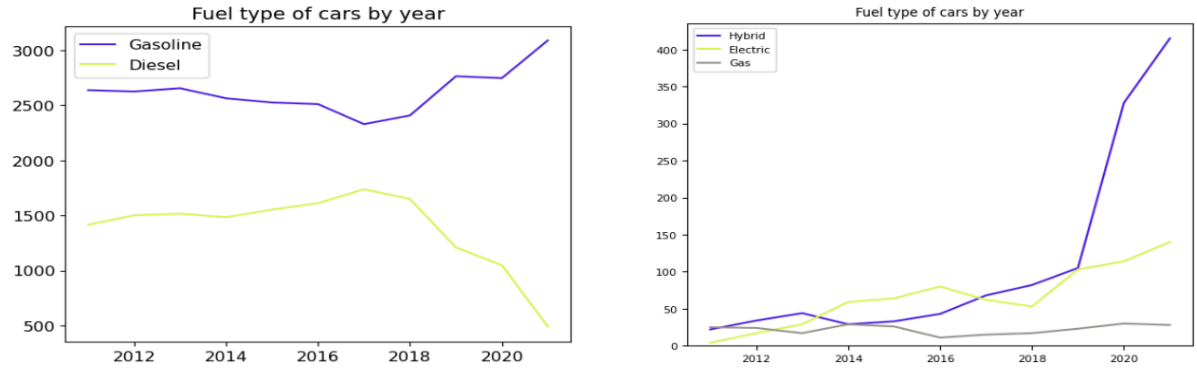
PRICE vs. HORSEPOWER

- Price of the vehicle has strong positive correlation with horsepower

- When the horsepower value of the car is higher than 400, the probability of huge price outliers increases.
- Correlation between variables is stronger when HP value is lower than 450.



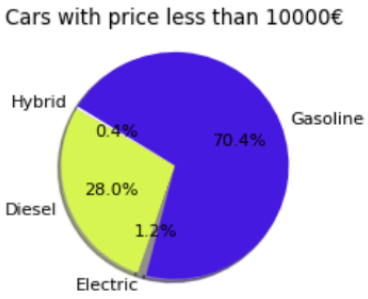
FUEL TYPES BY YEAR

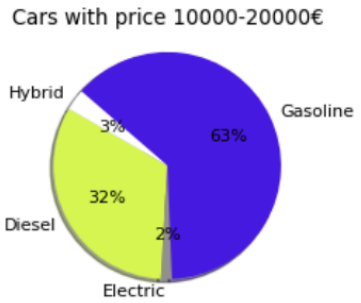


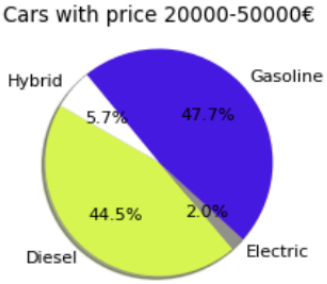
- The number of electric and hybrid vehicles represented increases with the production year.
- An anomaly was detected in the quantity of cars, produced in 2017: gasoline had the lowest value and diesel had its peak at the same time. The trend begins to reverse in 2018.
- Gas cars by year have no pronounced trend. The chart is in flat

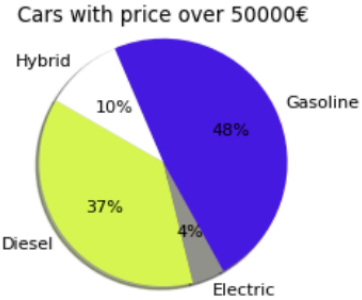
FUEL TYPES IN PRICE CATEGORIES RATIO

- Gasoline cars are widely represented among cars with prices lower than €10.000. This ratio falls (7.4%) in the next group and falls below half in the next two groups.
- Hybrid and electric vehicles' share of the total tends to grow with price in each price group.



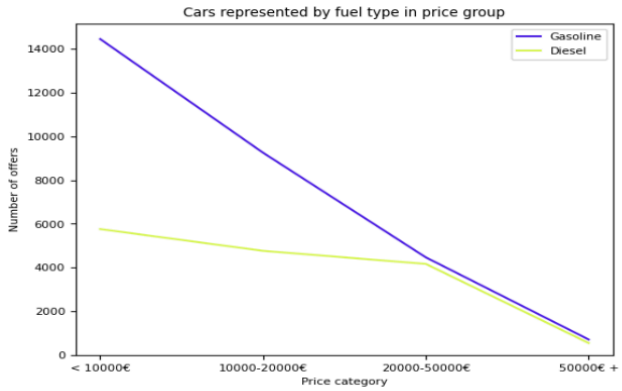


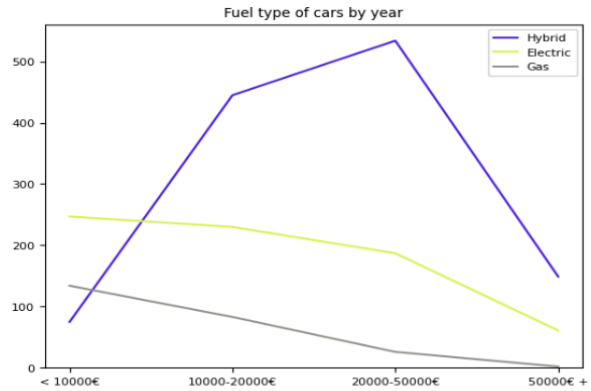




- The proportion of diesel vehicles increases with price in the first three groups while decreasing in the fourth.

REAL PRICE CATEGORIES BY FUEL TYPES





- The total quantity of gasoline, diesel, electric, and gas cars represented tends to decrease when prices increase.
- The number of hybrid vehicles increases in the first three price categories while decreasing in the fourth

PRICE CATEGORIES TRENDS

- Cars with a price lower than 10.000€ tend to be less represented when production year increase
- Vehicles with a price higher than 10.000€ tend to be more represented when production year increases

TOP 10 POPULAR BRANDS

Brand	Average price, €	Cars represented	Average mileage
Volkswagen	16060.09	6921.00	80881.04
Opel	10428.03	4801.00	78416.94
Ford	13793.37	4441.00	69201.89
Skoda	13715.19	2877.00	70644.23
Renault	11287.90	2828.00	57111.05
Audi	30000.23	2681.00	76796.27
BMW	23431.07	2400.00	97489.55
Mercedes-Benz	28379.92	2347.00	96164.90
SEAT	12847.43	1924.00	66158.88
Hyundai	12852.69	1886.00	52792.60

TOP 10 POPULAR MODELS

Brand	Model	Average price, €	Model represented	Average mileage
Volkswagen	Golf	13540.55	1492.00	91738.14
Opel	Corsa	8959.63	1491.00	55012.01
Ford	Fiesta	9864.22	1289.00	52437.59
Opel	Astra	9574.46	1191.00	101118.97
Ford	Focus	12428.89	985.00	83752.35
Volkswagen	up!	7622.25	945.00	55040.93
Skoda	Fabia	10170.06	917.00	55746.15
Volkswagen	Polo	10053.58	893.00	68183.68
smart	forTwo	7878.90	725.00	55304.53
Fiat	500	11022.25	666.00	25588.98

TOP 10 EXPENSIVE CARS

Brand	Model	Price, €	Horsepower	Mileage
Ferrari	F12	1199900	775.00	431
Mercedes-Benz	S 650	717078	630.00	90
Maybach	Pullman	717078	630.00	90
Maybach	Pullman	499800	530.00	3400
Mercedes-Benz	SLS	465000	591.00	350
Lamborghini	Aventador	449900	770.00	3600
Ferrari	812	439900	799.00	6500
Lamborghini	Aventador	439900	770.00	14090
Lamborghini	Aventador	399999	700.00	1200
Porsche	991	399911	700.00	4624

TOP 10 CHEAPEST CARS

Brand	Model	Price,€	HP	Average mileage
Citroen	C1	1100	68.00	204000
Lada	Priora	1190	98.00	100611
Toyota	Aygo	1250	68.00	153000
Peugeot	206	1299	68.00	222000
Opel	Corsa	1300	69.00	133000
Volkswagen	Touran	1396	140.00	233567
Brilliance	BS4	1500	101.00	100524
SEAT	Ibiza	1500	69.00	170000
smart	forTwo	1500	54.00	140000
Fiat	Panda	1500	69.00	157000

TOP 10 POWER

Brand	Model	Price	Horsepower	Mileage
Mercedes-Benz	G 63 AMG	216619	850.00	9234
Audi	R8	295000	799.00	11000
Ferrari	812	439900	799.00	6500
Ferrari	F12	1199900	775.00	431
Tesla	Model S	54450	772.00	71099
Tesla	Model X	99999	772.00	63205
Ford	Mustang	121000	771.00	50
Lamborghini	Aventador	439900	770.00	14090
Lamborghini	Aventador	449900	770.00	3600
Corvette	Z06	148600	767.00	9790

TOP 10 LEAST HORSEPOWER

Brand	Model	AVG Price	HP
Ligier	JS 50	9570.00	5.00
Estrima	Birò	13090.00	5.00
Microcar	M.Go	8262.86	5.86
Microcar	M8	6149.50	6.00
Aixam	City	12935.00	6.50
Microcar	Due	6745.00	7.50
Piaggio	Ape	7849.75	10.25
Renault	Twizy	5413.43	10.74
Tazzari	EV Zero Classic	6500.00	20.00
Zhidou	D2S	12500.00	24.00

CONCLUSION

In this project, an analysis of data on German cars from the Autoscaut24 website was carried out. Relationships, interesting trends between variables were identified.

An anomaly was discovered in 2017. Gasoline cars are the least represented in this year's production, while diesel cars are the most widely represented. Also this year and after, there is a downward trend in the number of cars with a manual transmission and an increase in the number of automatic ones. The analysis was done using Python in Jupiter