EDA on Used Car Sales

Agenda

- Problem Statement
- Introduction to Use Cars Sales Dataset
- Exploring the Dataset
- Observations and Conclusions
- Actionable Insights

Problem Statement

- The used cars sales company has earned its name because of sincerity in work and quality of services.
- But for recent past for past few months their sales is down.

Introduction

- Dataset contains one of the used cars company sales.
- ~ 10k records available
- Sales history from 1950 to till 2016
- Visual Data Analysis

Id	Feature	Description
01	car	Car brand name
02	model	Available car different Variants
03	year	purchasing Year
04	body	Body type-Hatchback, Sedan, Crossover etc
05	mileage	car Mileage
06	engV	Engine version
07	engType	Car Fuel type - Petrol, Diesel, gas etc
08	drive	Wheel Drive Front, back
09	registration	Check if the vechile is registered
10	price	Price of Car in $

Data Pre Profiling

Overview | Warnings 7 | Reproduction

Dataset statistics

Number of variables	10
Number of observations	9576
Missing cells	945
Missing cells (%)	1.0%
Duplicate rows	113
Duplicate rows (%)	1.2%
Total size in memory	748.2 KiB
Average record size in memory	80.0 B

Variable types

Categorical	5
Numeric	4
Boolean	1

Overview | Warnings 7 | Reproduction

Warnings

Dataset has 113 (1.2%) duplicate rows	Duplicates
`car` has a high cardinality: 87 distinct values	High cardinality
`model` has a high cardinality: 888 distinct values	High cardinality
`engV` has 434 (4.5%) missing values	Missing
`drive` has 511 (5.3%) missing values	Missing
`price` has 267 (2.8%) zeros	Zeros
`mileage` has 348 (3.6%) zeros	Zeros

	car	price	body	mileage	engV	engType	registration	year	model	drive
count	9576	9576.000000	9576	9576.000000	9142.000000	9576	9576	9576.000000	9576	9065
unique	87	NaN	6	NaN	NaN	4	2	NaN	888	3
top	Volkswagen	NaN	sedan	NaN	NaN	Petrol	yes	NaN	E-Class	front
freq	936	NaN	3646	NaN	NaN	4379	9015	NaN	199	5188
mean	NaN	15633.317316	NaN	138.862364	2.646344	NaN	NaN	2006.605994	NaN	NaN
std	NaN	24106.523436	NaN	98.629754	5.927699	NaN	NaN	7.067924	NaN	NaN
min	NaN	0.000000	NaN	0.000000	0.100000	NaN	NaN	1953.000000	NaN	NaN
25%	NaN	4999.000000	NaN	70.000000	1.600000	NaN	NaN	2004.000000	NaN	NaN
50%	NaN	9200.000000	NaN	128.000000	2.000000	NaN	NaN	2008.000000	NaN	NaN
75%	NaN	16700.000000	NaN	194.000000	2.500000	NaN	NaN	2012.000000	NaN	NaN
max	NaN	547800.000000	NaN	999.000000	99.990000	NaN	NaN	2016.000000	NaN	NaN

Action Points:

- Remove **113 duplicate records**.
- **engV** has 434 (4.5%) missing values. Replace the missing values with median **2**. There are 28 records (~0.3%) with value 99.99 and the percentage is very less and leave them as it is.
- **drive** has 511 (5.3%) missing values. Replace the missing values with mode **front**.
- **price** has 267 (2.8%) zeros. Replace the zero values with median **9200**.
- **mileage** has 348 (3.6%) zeros. Replace the zero values with median **128**.

Data Post Profiling

Overview | Warnings 4 | Reproduction

Dataset statistics

Number of variables	11
Number of observations	9463
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	813.4 KiB
Average record size in memory	88.0 B

Variable types

Numeric	5
Categorical	5
Boolean	1

Overview | Warnings 4 | Reproduction

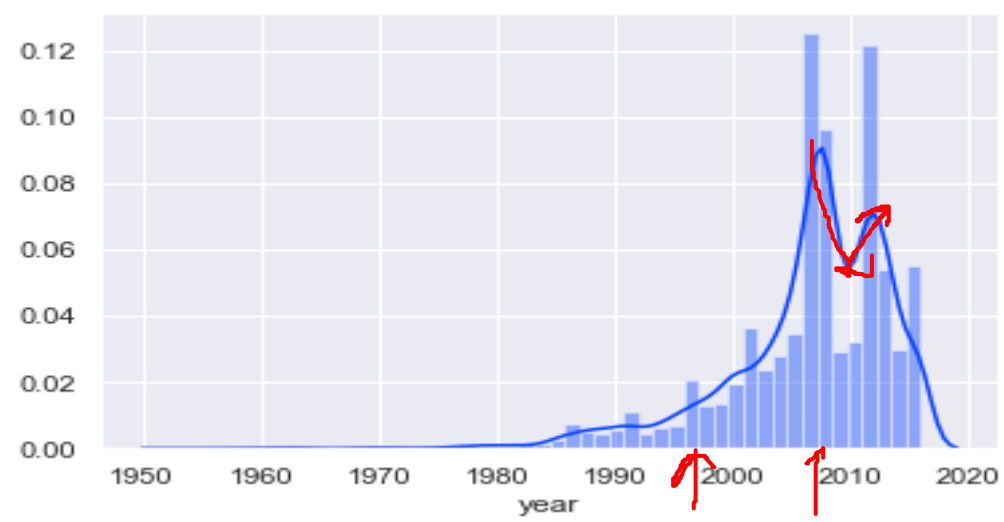
Warnings

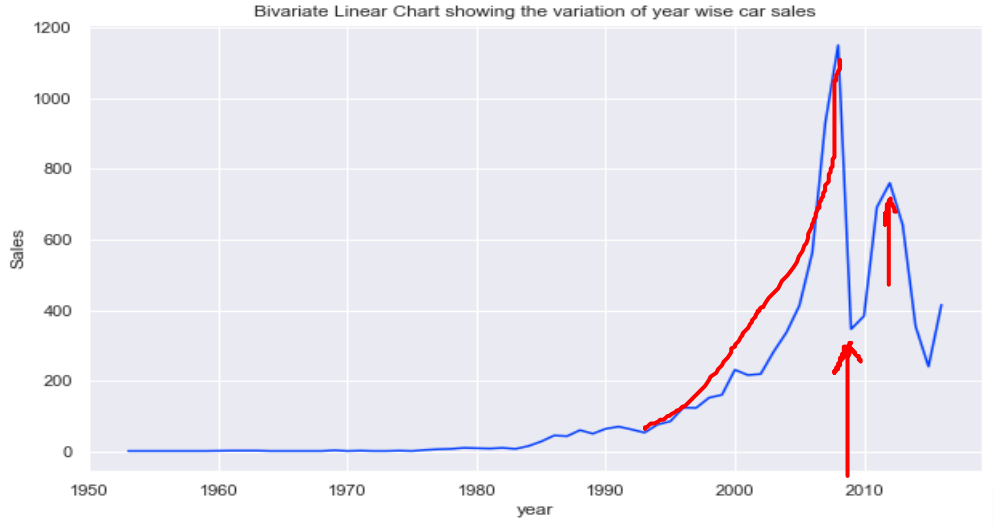
- `car` has a high cardinality: 87 distinct values — High cardinality
- `model` has a high cardinality: 888 distinct values — High cardinality
- `df_index` is uniformly distributed — Uniform
- `df_index` has unique values — Unique

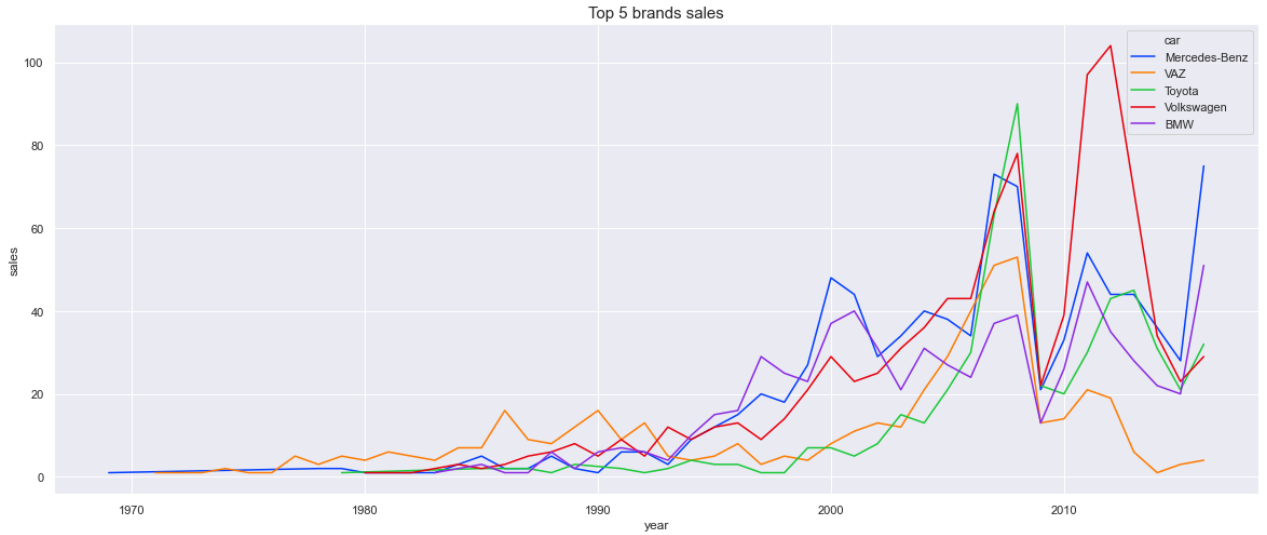
	car	price	body	mileage	engV	engType	registration	year	model	drive	class
count	9463	9463.000000	9463	9463.000000	9463.000000	9463	9463	9463.000000	9463	9463	9463
unique	87	NaN	6	NaN	NaN	4	2	NaN	888	3	3
top	Volkswagen	NaN	sedan	NaN	NaN	Petrol	yes	NaN	E-Class	front	Economy
freq	927	NaN	3622	NaN	NaN	4341	8902	NaN	192	5681	5174
mean	NaN	15537.188707	NaN	144.150586	2.612026	NaN	NaN	2006.525837	NaN	NaN	NaN
std	NaN	22947.612648	NaN	95.079722	5.826202	NaN	NaN	7.062584	NaN	NaN	NaN
min	NaN	259.350000	NaN	1.000000	0.100000	NaN	NaN	1953.000000	NaN	NaN	NaN
25%	NaN	5500.000000	NaN	80.000000	1.600000	NaN	NaN	2004.000000	NaN	NaN	NaN
50%	NaN	9200.000000	NaN	129.000000	2.000000	NaN	NaN	2008.000000	NaN	NaN	NaN
75%	NaN	16500.000000	NaN	195.000000	2.500000	NaN	NaN	2012.000000	NaN	NaN	NaN
max	NaN	547800.000000	NaN	999.000000	99.990000	NaN	NaN	2016.000000	NaN	NaN	NaN

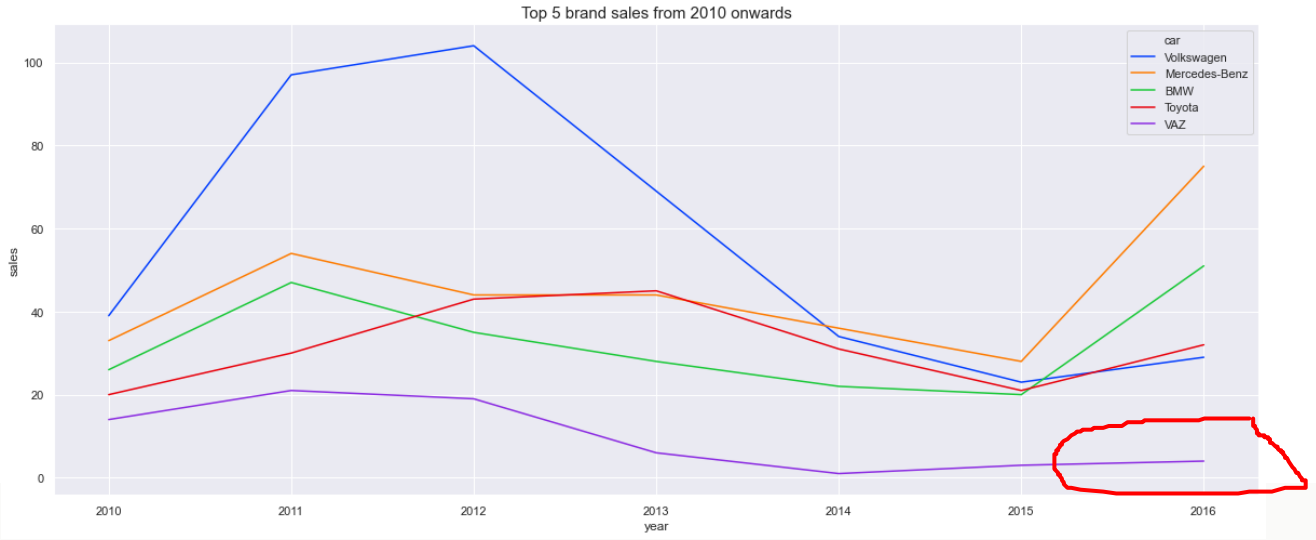
Data Exploration

Year wise car sales





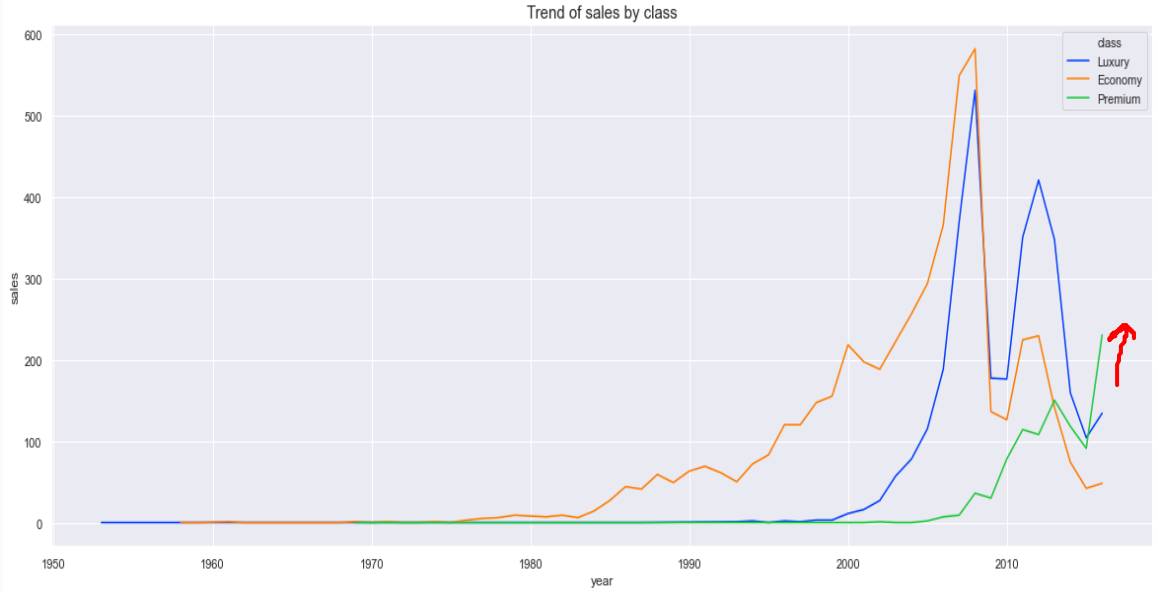


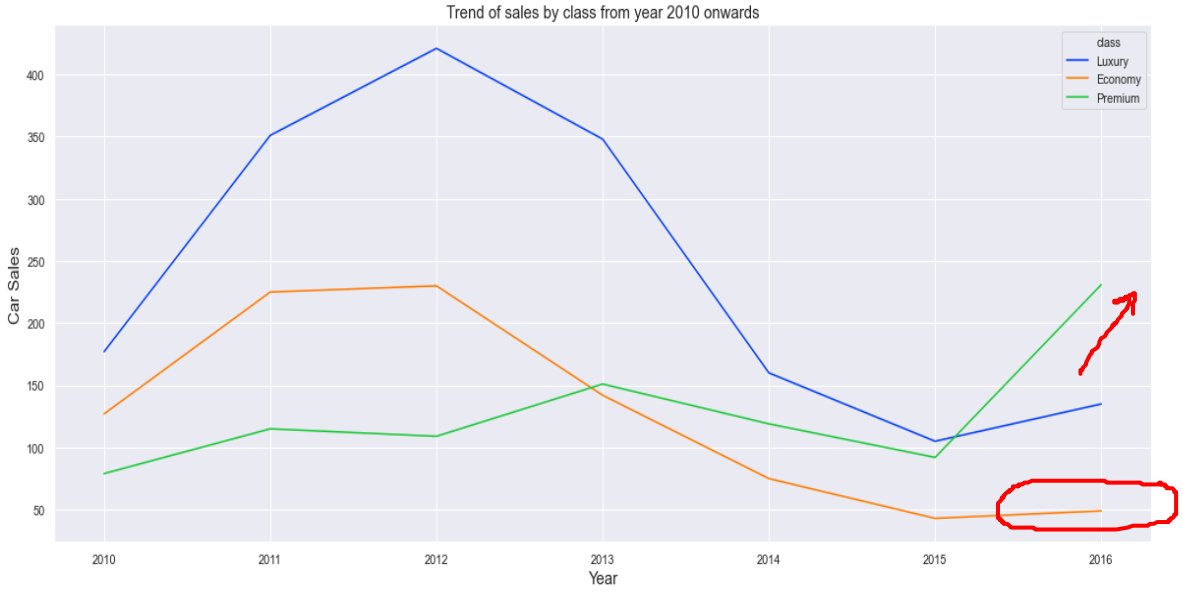


Data Exploration

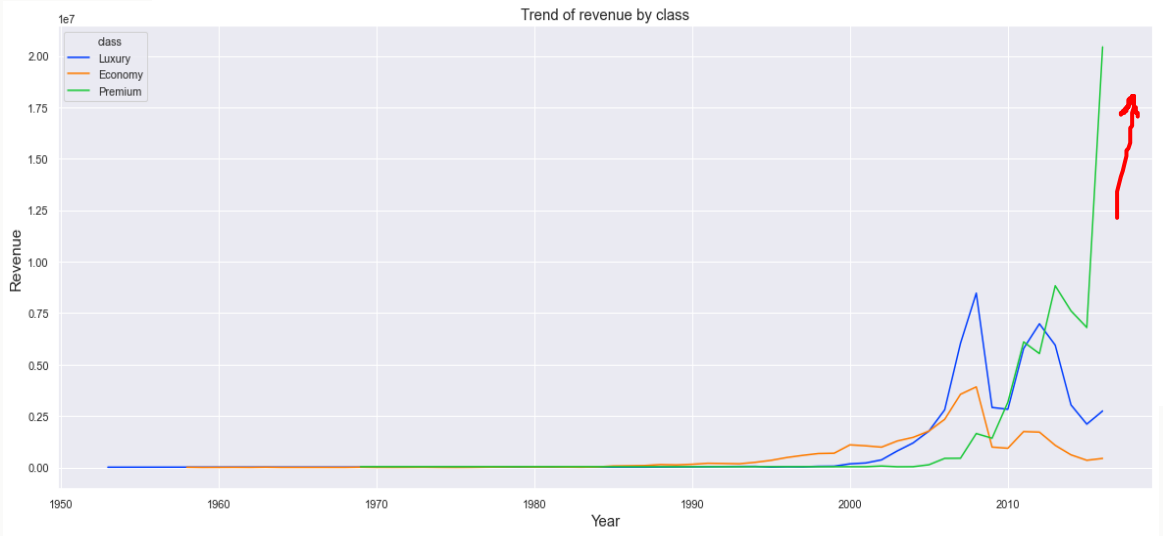
Sales and Revenue based on Class

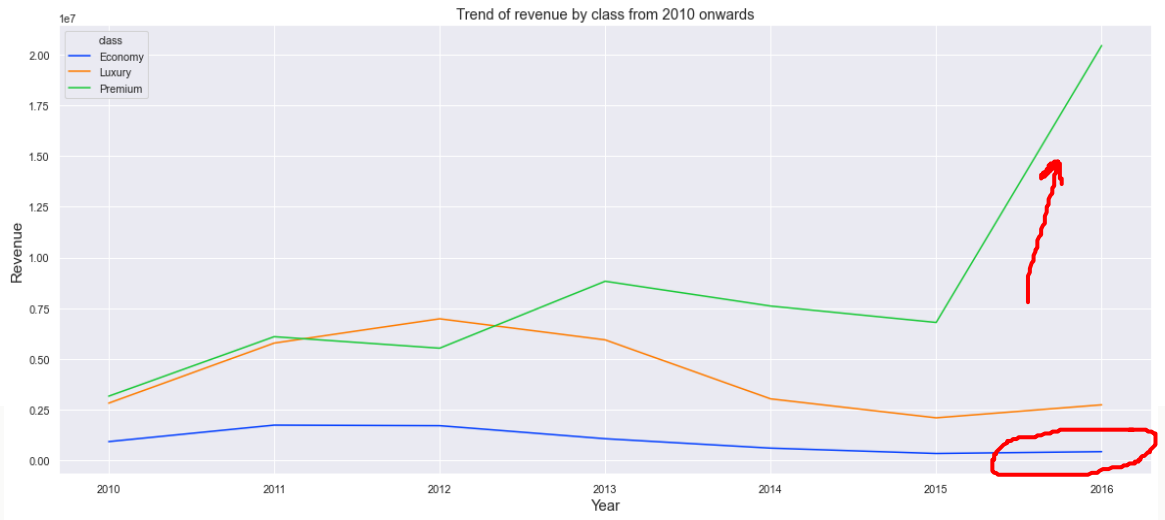
Sales





Revenue



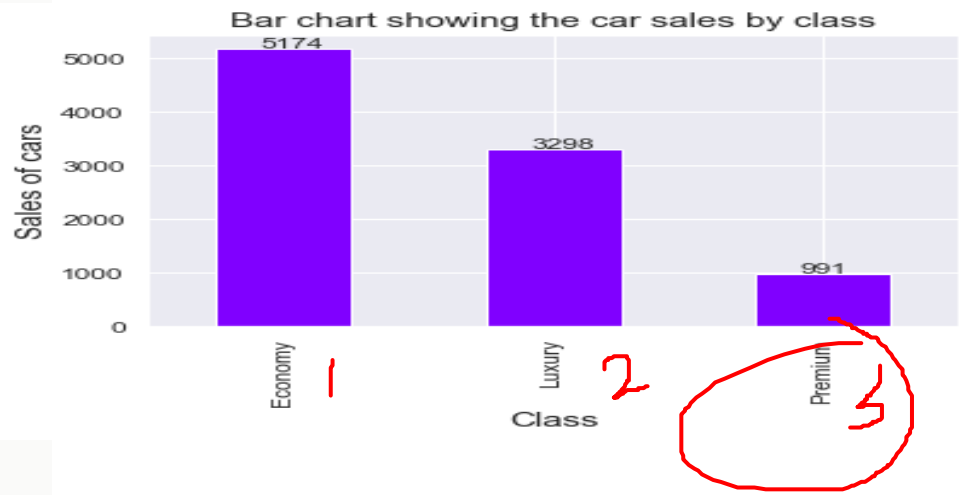


Data Exploration

Sales and Revenue based on Class

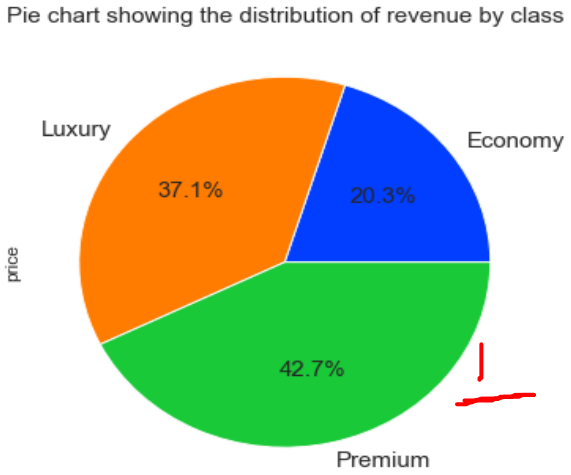
Sales

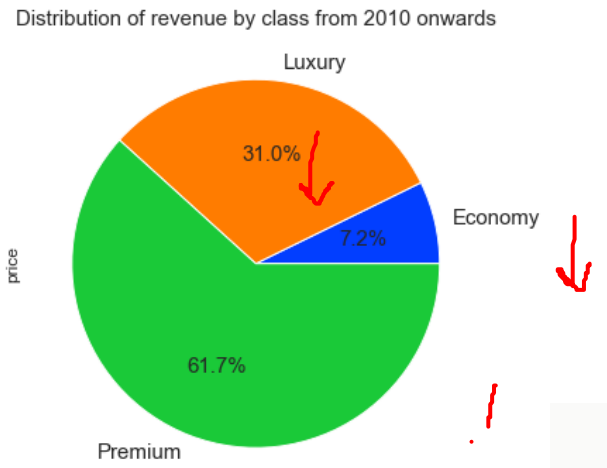
Overall



From 2010 onwards

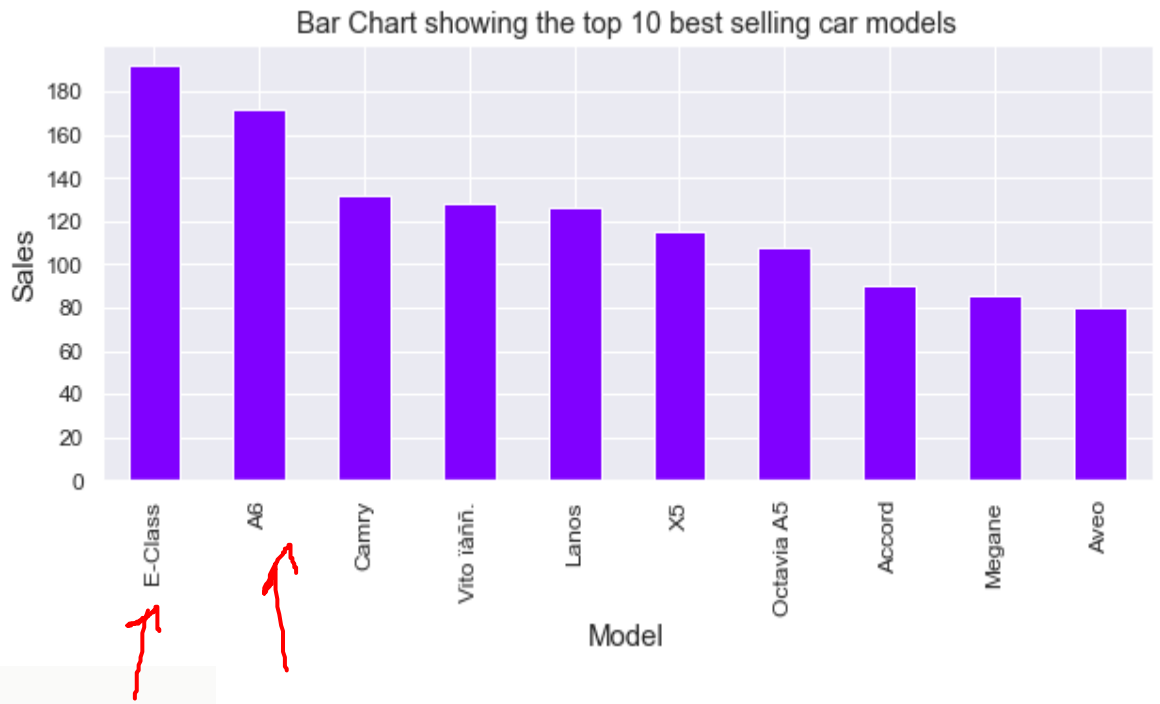
Revenue

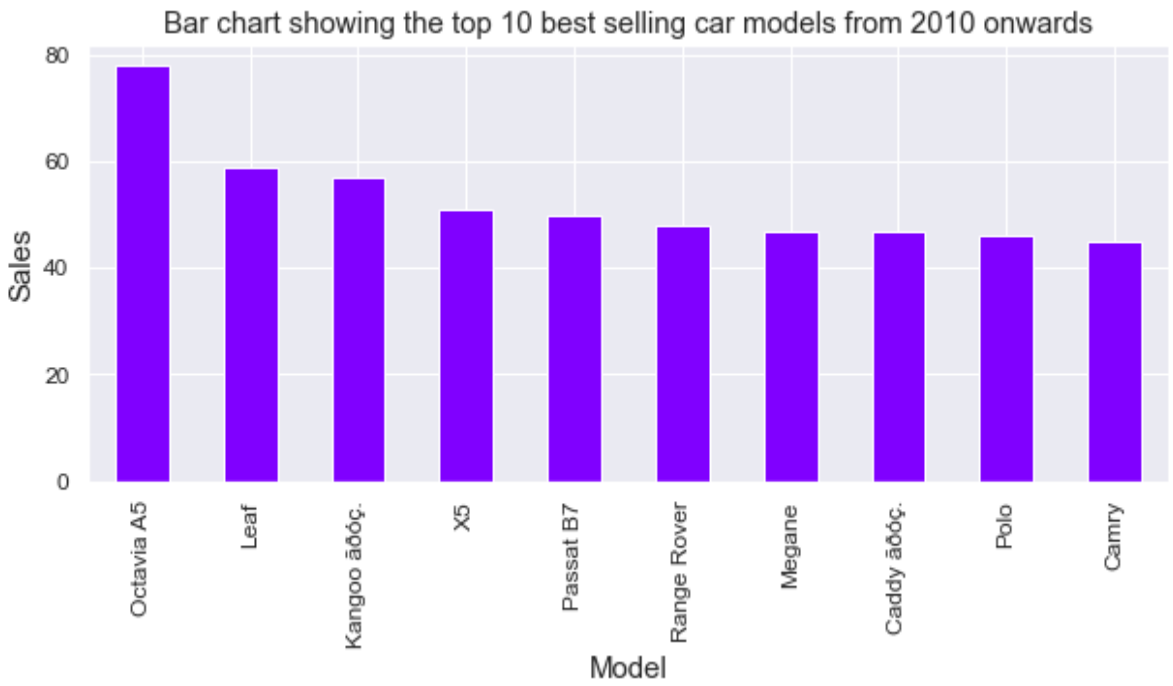




Data Exploration

Sales based on Model





- From the overall sales list of top 10 models, only 4 were able to list down in top 10 from year 2010 onwards. They are **Camry, X5, Octavia A5 and Megane**.
- The top model cars **E-Class and A6** are not in the top 10 list of best selling cars from 2010 onwards.
- Customers are not preferring Economy Class cars in recent years.

Data Exploration

Top 10 best price sold cars

	car	price	body	mileage	engV	engType	registration	year	model	drive
7621	Bentley	547800.0	sedan	129	6.75	Petrol	yes	2016	Mulsanne	rear
1611	Bentley	499999.0	crossover	129	6.00	Petrol	yes	2016	Bentayga	full
4134	Bentley	449999.0	crossover	129	6.00	Petrol	yes	2016	Bentayga	full
4325	Mercedes-Benz	300000.0	sedan	68	6.00	Petrol	yes	2011	S 600	front
5849	Mercedes-Benz	300000.0	other	37	5.00	Petrol	yes	2012	G 500	full
1891	Mercedes-Benz	295000.0	sedan	29	6.00	Petrol	yes	2011	S 600	rear
2165	Mercedes-Benz	295000.0	sedan	29	6.00	Petrol	yes	2011	S-Guard	rear
8205	Land Rover	285000.0	crossover	129	5.00	Petrol	yes	2016	Range Rover	full
1478	Bentley	259000.0	sedan	129	6.00	Petrol	yes	2014	Flying Spur	full
564	Mercedes-Benz	250000.0	other	6	5.50	Petrol	yes	2016	S 63 AMG	full

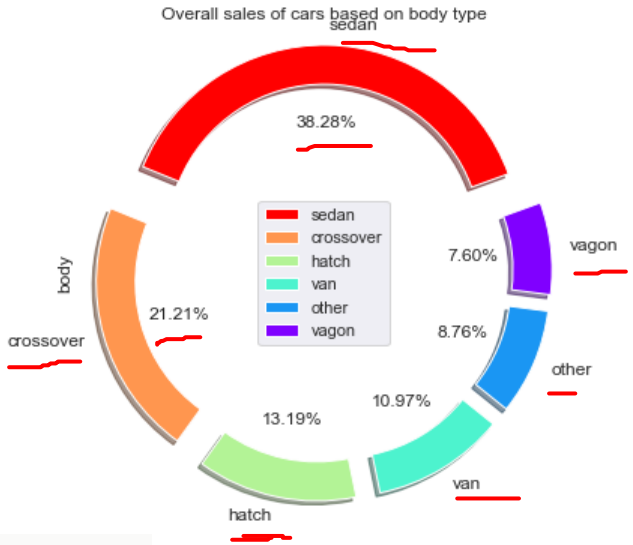
Top 10 best sold car brands and models listed below:

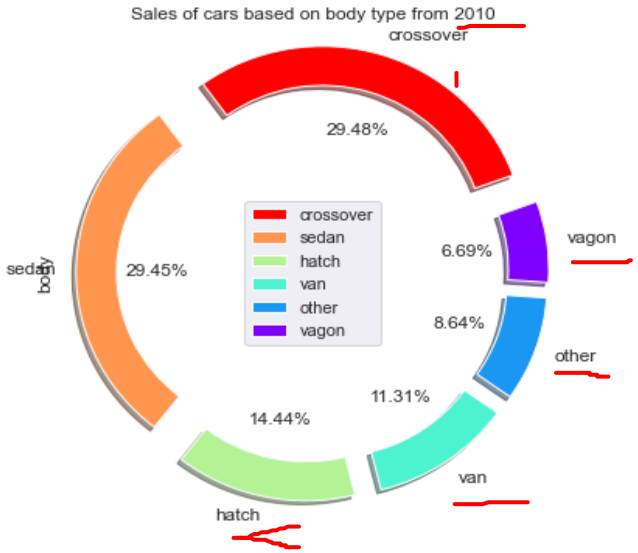
- ***Bently***
  - Mulsanne
  - Bentayga
  - Flying Spur
- **Mercedes-Benz**
  - S 600
  - G 500
  - S-Guard
- **Land Rover**
  - Range Rover

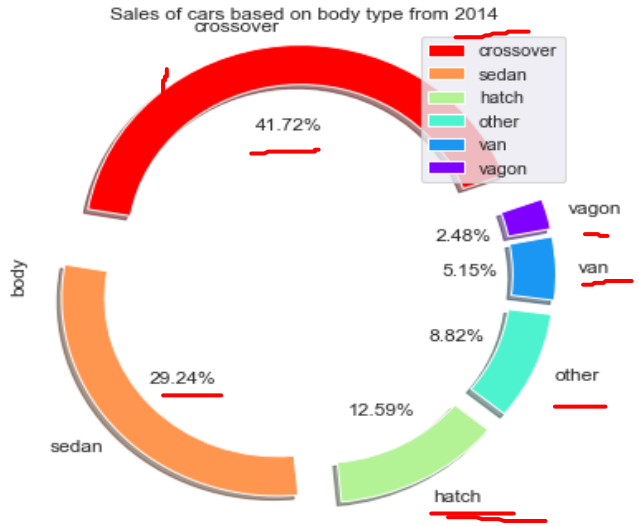
- Though **Mercedes-Benz** and **Land Rover** are proven strength in sales under Premium sector, nevertheless **Bentley** dominating recently in this segment.

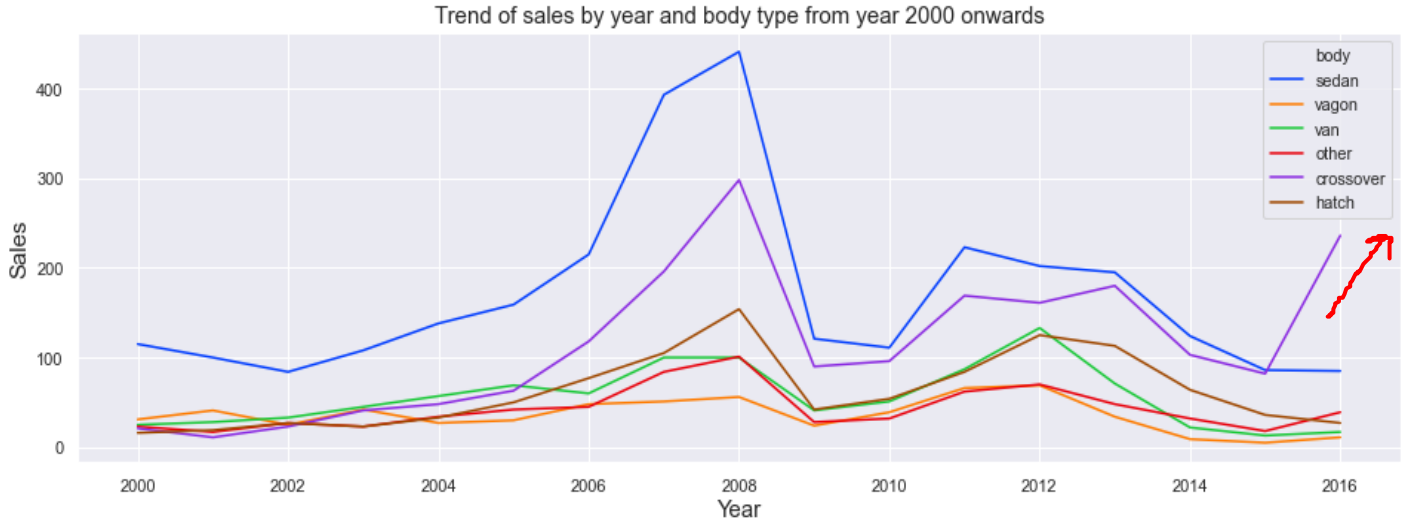
Data Exploration

Sales based on body type





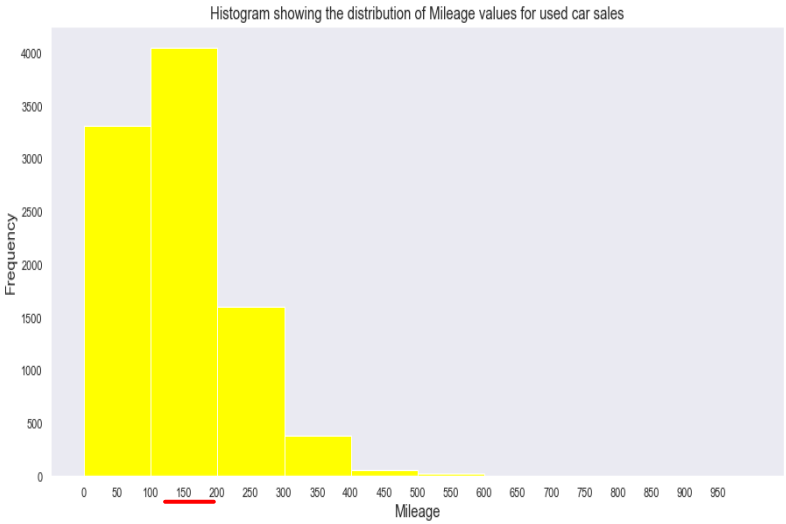


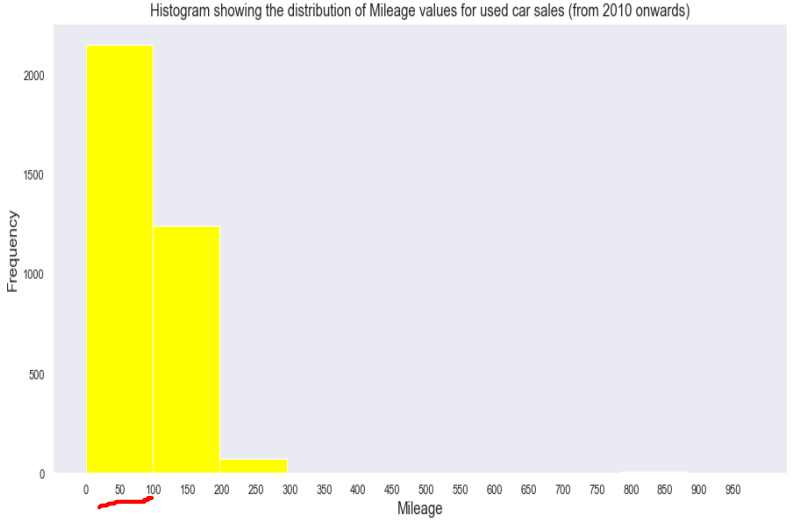


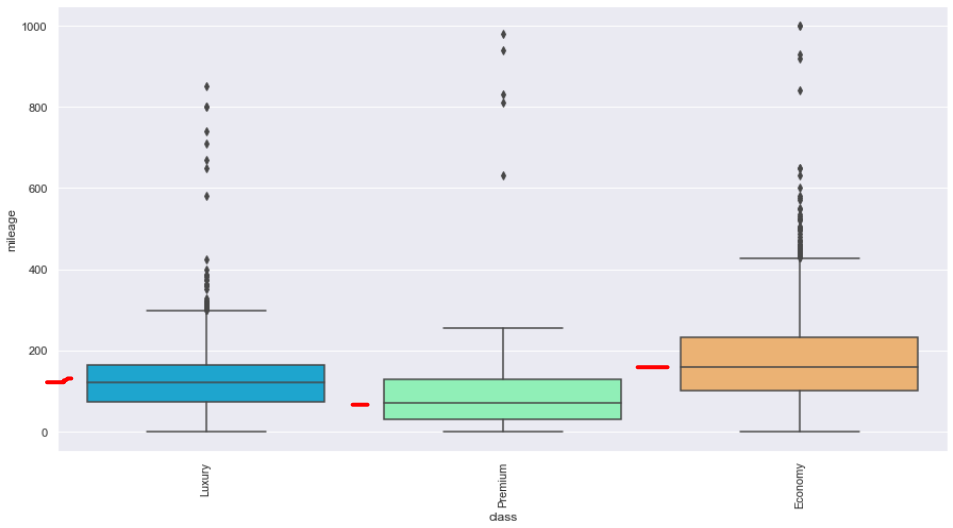
- **Sedan** is **highest sold body** type followed by crossover, hatch, van, other and vagon.
- 2010 onwards, body type **crossover sales increased from overall 21% to 41% and sedan car sales decreased from overall 38% to 30%. Hatch and Other sales are consistent whereas Van and Vagon sales are dropped**.
- So possible reason of declining sales are **Van & Vagon type cars are loosing market**.

Data Exploration

Car sales based on mileage



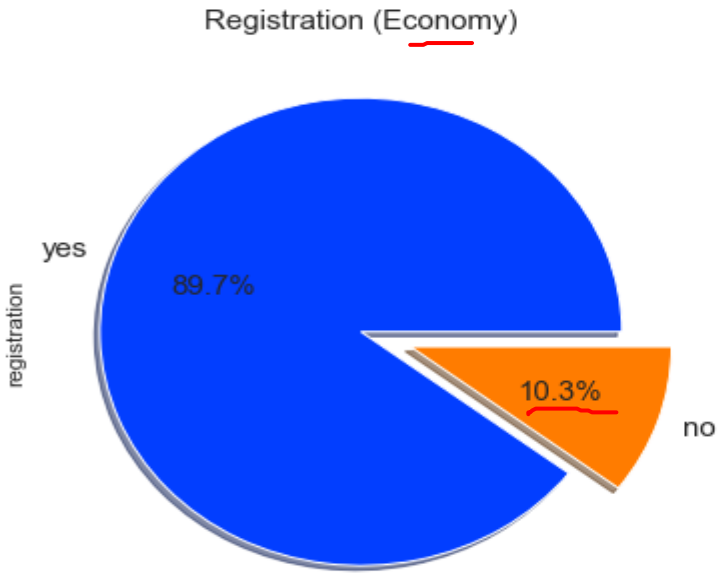


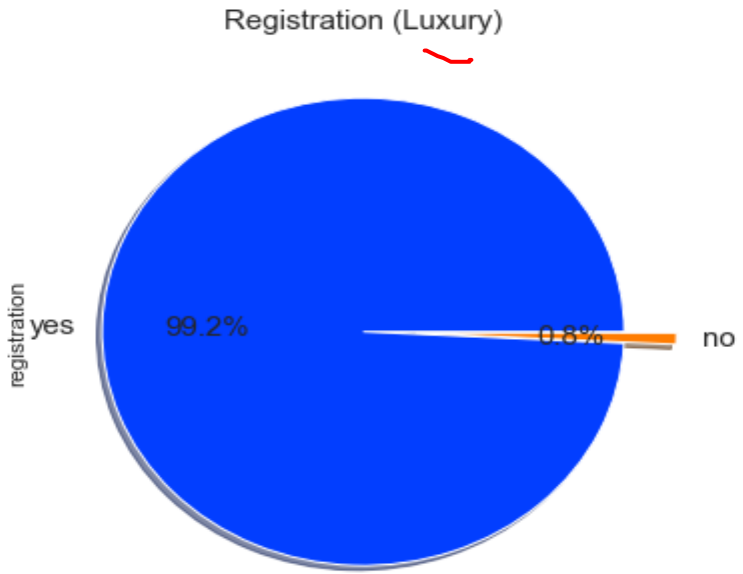


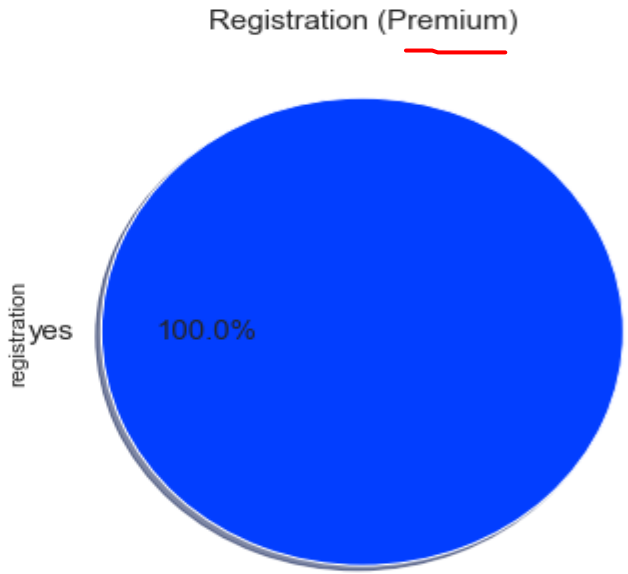
- **Economy class have high mileage vehicles, followed by Luxury and Premium segment vehicles.**
- The average mileage of Economy cars 138, Luxury cars 102 and Premium cars 71.
- In overall, more vehicles sold with the mileage in the range of 100 - 200.
- From year 2010 onwards, cars sales with mileage 0 to 100 range leading significantly compared to 100 to 200 range.

Data Exploration

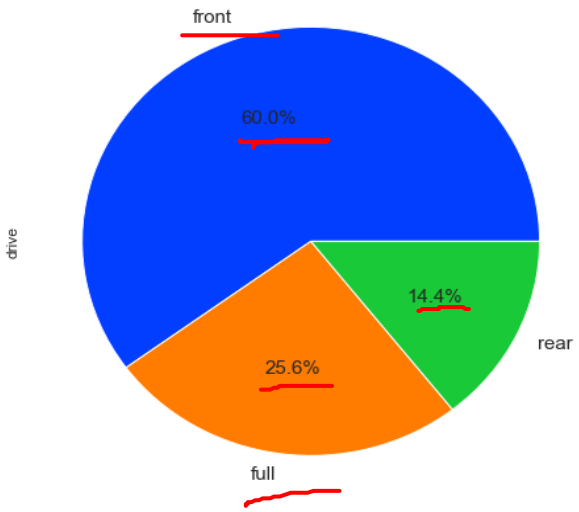
Sales based on registration and drive type



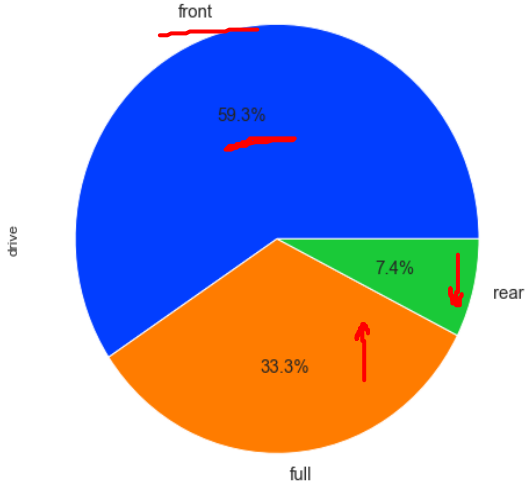




Pie chart showing the distribution of drive type and car sales

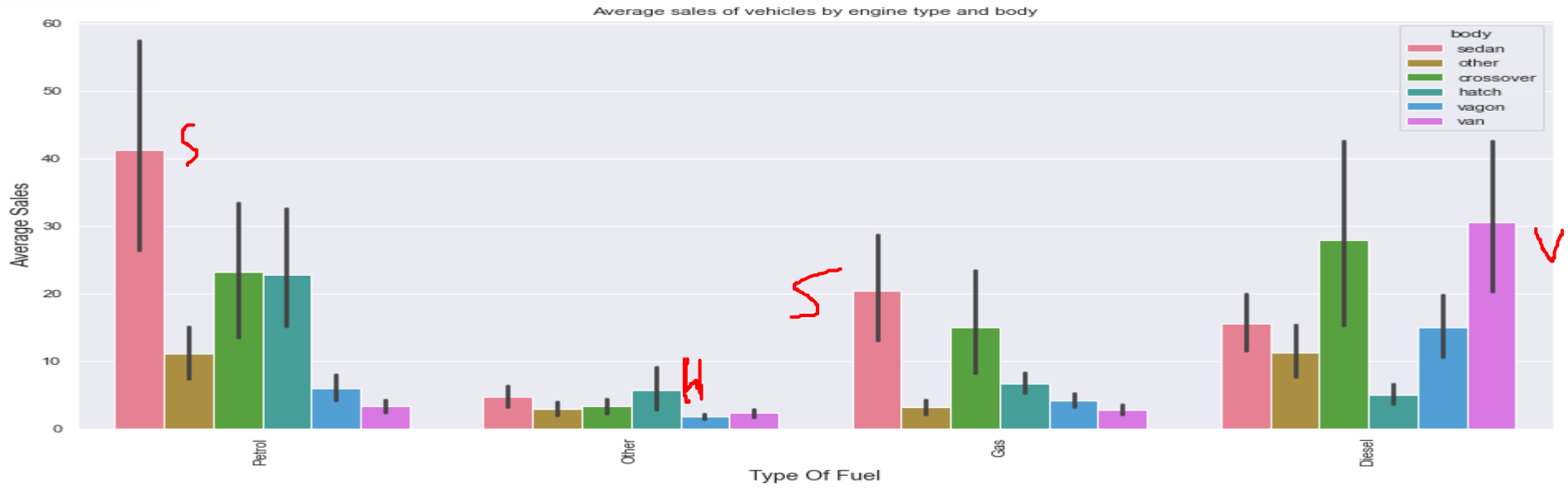


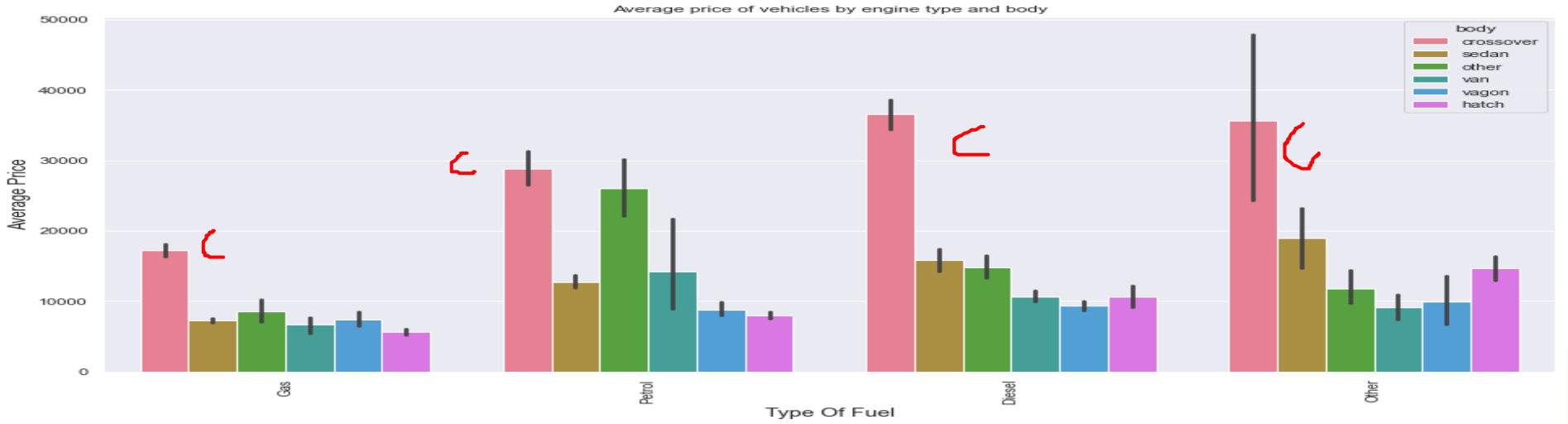
Pie chart showing the distribution of drive type and car sales from 2010 onwards



Data Exploration

Sales based on engine type and body

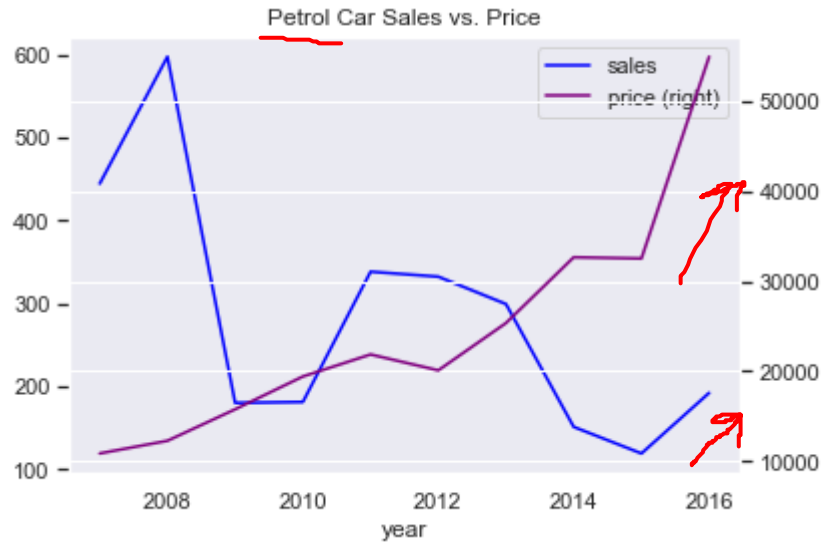


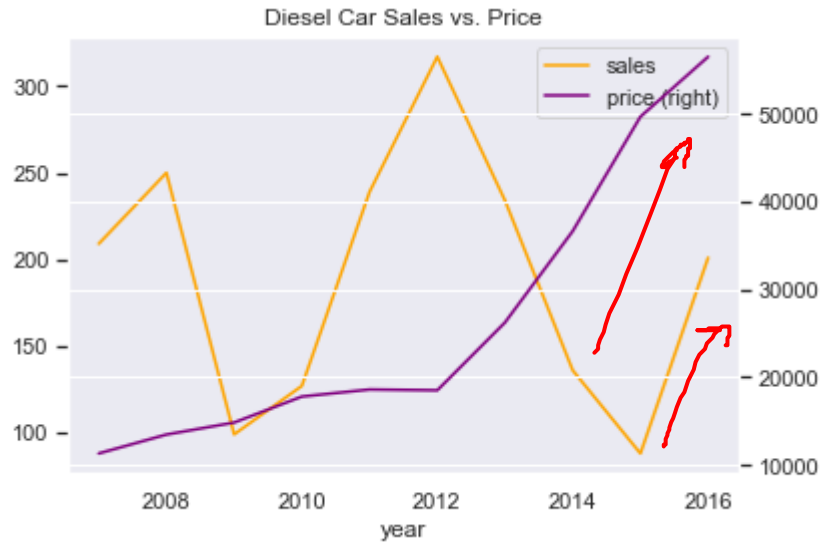


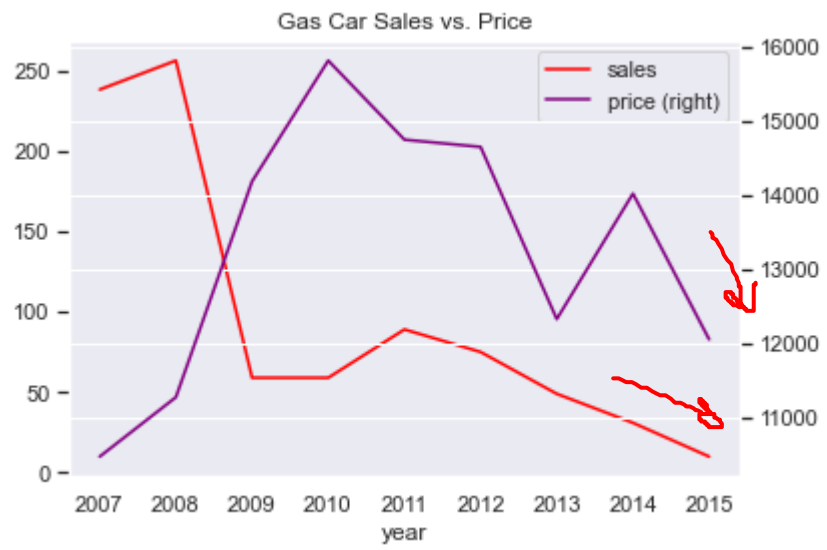
- Sedan is highest sold body type in Petrol and Gas fuel type segments, followed by cross over.
- Van is highest sold one under Diesel type followed by cross over.
- Hatch is highest sold one under Other fuel type followed by Sedan.
- Overall, cross over body type prices are highest in all fuel segments.

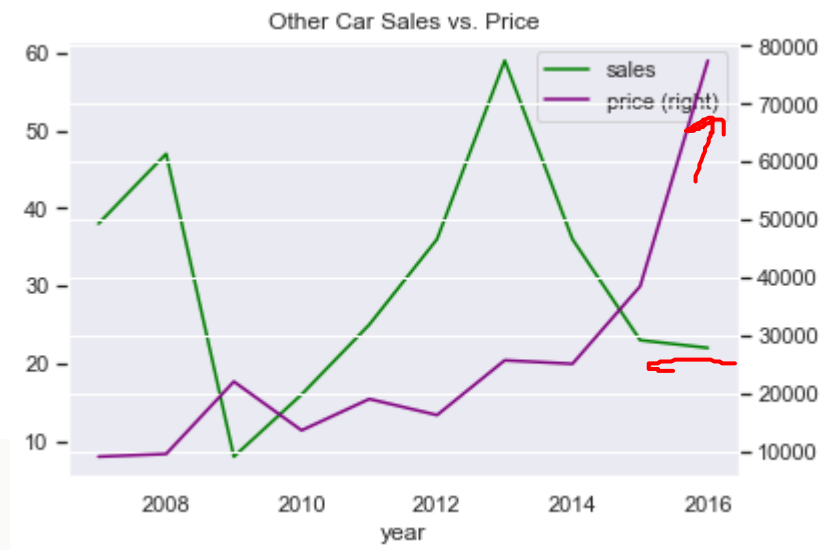
Data Exploration

Sales and Prices based on fuel type



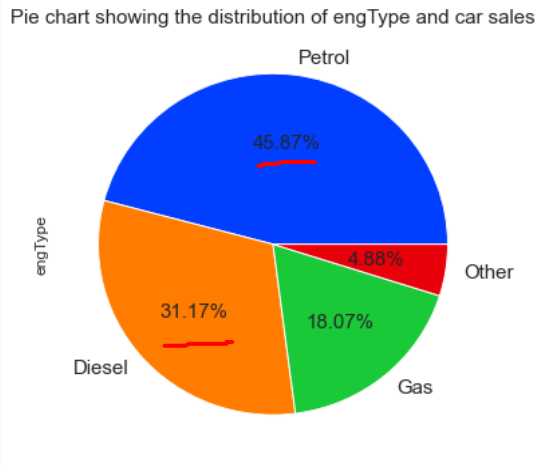


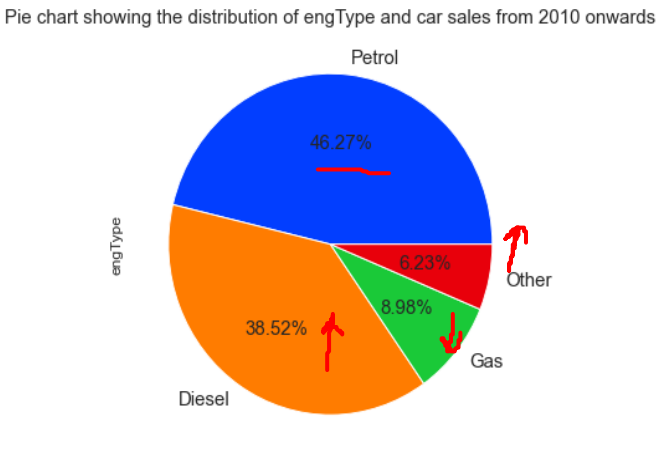


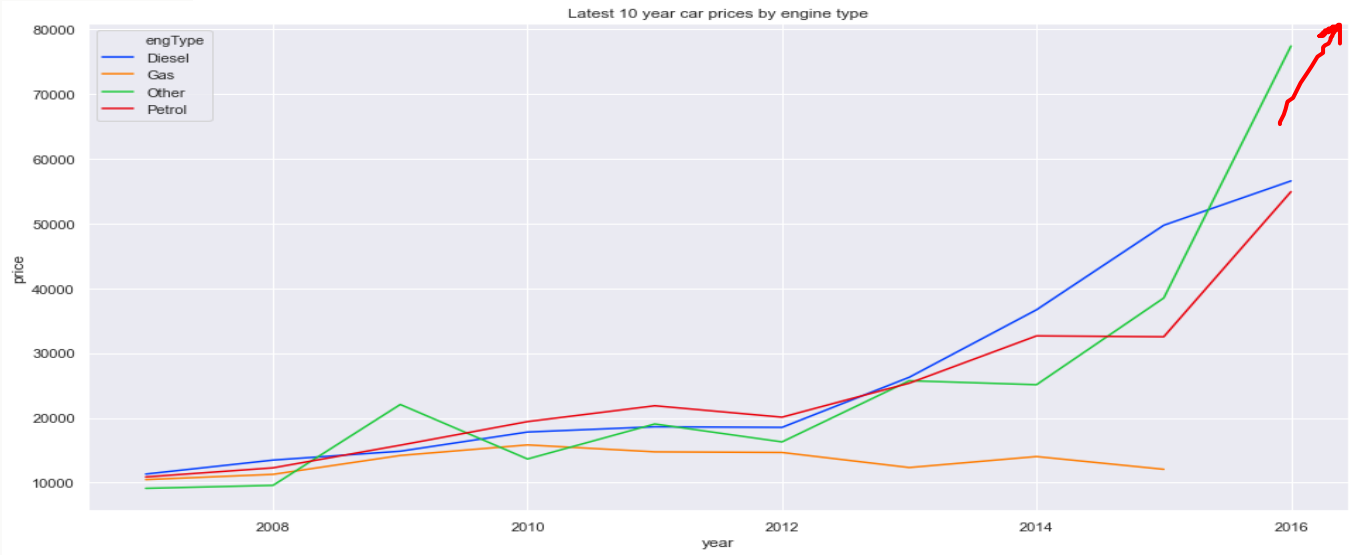


Data Exploration

Sales based on Engine Type



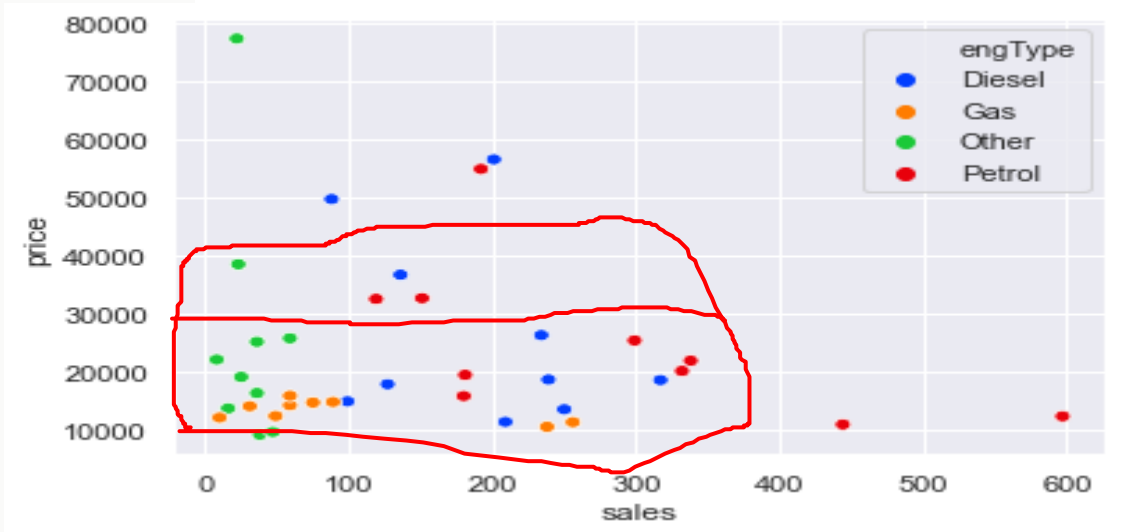


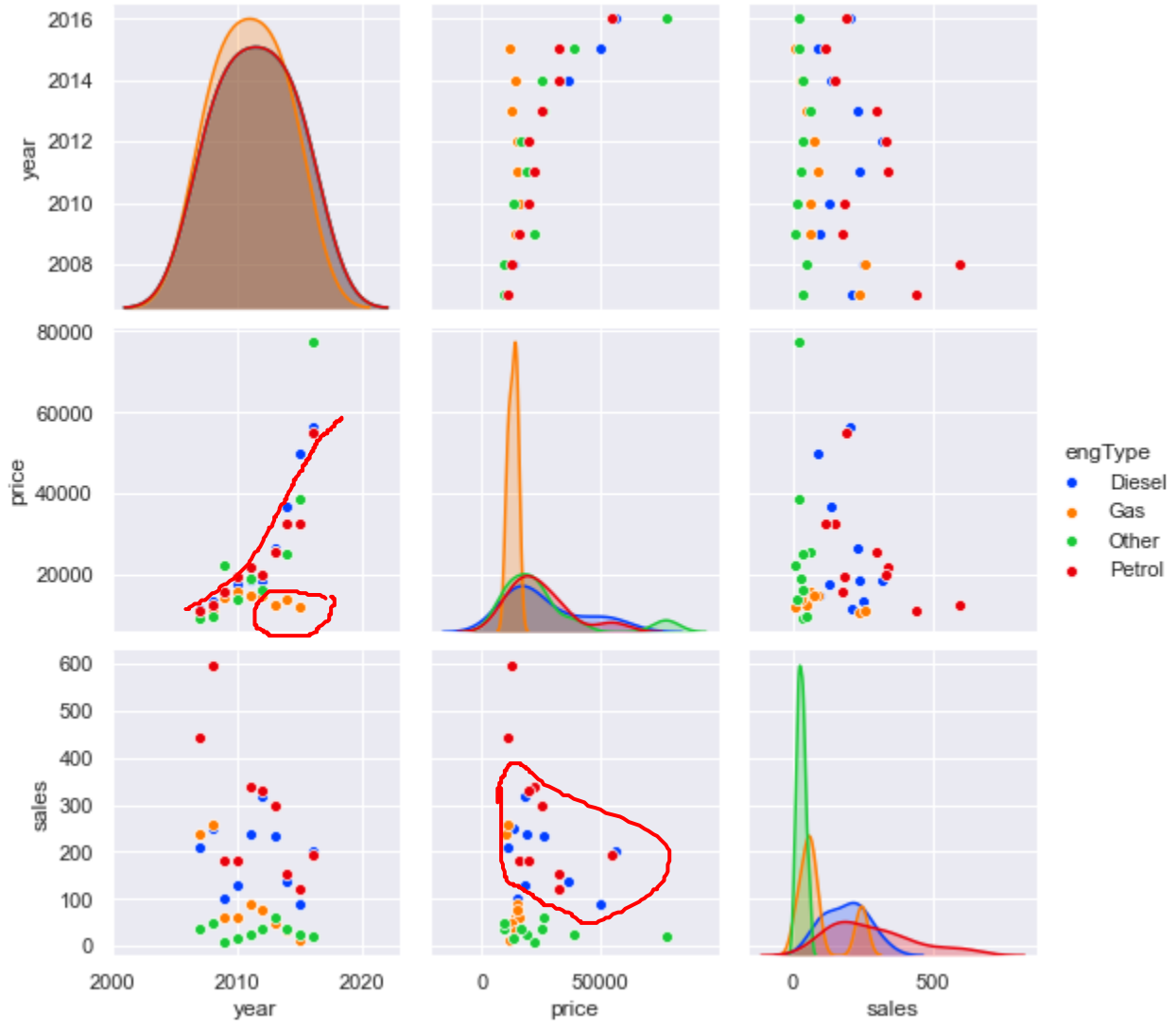


- In overall sales, **the highest sold engine types are Petrol** followed by Diesel, Gas and Other.
- From year 2010 onwards, Petrol car sales are consistent around 46% and increased slightly by 0.4%.
- The **Diesel car sales increased** significantly to 7.4% and the **Gas car sales decreased significantly** by 9% and **Other car sales increased** by 1.3%.
- **The Other cars (Hybrid, Electric, ..etc) prices are increasing continuously**.

Data Exploration

Correlation between Sales and Prices over the years





Observations and Conclusions

1.Most **sales happening price range** is between **10000 to 40000 *I-***
2.There are **very few cars sold** in **20000 to 36000** price bracket, which is best competitive price range
3.Though the cars with **Gas** engine type sold at most economical price, in the **recent days customers are not preferred**.
4.None of the cars sold from **Economy** segment in year **2016**.
5.From the recent past the car sales are stagnated on Gas type. In other hand

- Petrol car sales are consistent
- Diesel car sales are gaining momentum
- Other car sales are gradually increasing

6.Car registrations are most considered in Luxury and Premium cars
7.Even with affordable price range, gas and other engine type cars are least preferred.
8.Most preferred

1. engine types are Petrol and Diesel
2. body types are cross-overs and sedan
3. Other type cars are consistent and gradually increasing in sales

9.Irrespective of tenure, Sedans are always on demand follows with Cross-over and hatch.
10.Tremendous sales reported in **2016** for these super premium cars by **Bentley, Land Rover, Toyota, Tesla, Lexus, Cardillac & BMW**
11.Recently **Crossovers** are leading in sales followed by **sedans**
12.Most of the recent premium cars engine type is **Petrol**

Actionable Insights

Cars should be available from all price ranges.

As 10000 to 40000 /- is the most sales happening price range and there are no cars available in 20000 to 36000 /- price bracket, procuring these price ranged cars should potentially improves the sales.

Procure cars based on customer trends.

- Petrol and Diesel cars are most popular.
- Full wheel drive is most preferred in Crossovers and Front wheel drive in Sedan cars.
- Luxury and Premium used cars booming sales due to hike in new car prices.

Tactical & Strategic Decision

- Less preferred old cars should be announced with great deals and clear the stock. In that way we can save money, renting space for storage
- Procure more demanded cars like Mercedes-Benz, BMW, Volkswagen, Toyota, Nissan, Ford, Audi ..etc
- Procure more Economy price range cars which is of Petrol or Diesel engine types.

THANK YOU