Analytical Lab Project

CARS IN GERMANY

Group A:
Assylbek Bugybay
Sevde Yanik
Aleksandra Pasheva
Anna Campi
Rami Bader

CONTENT

- How the data was obtained.
- How did we build database and what analysis was done.
- Results we got.

PURPOSE:

- By analysing data of Autoscout in Germany, we investigate evolving demands of car features throughout the years.
- We want to give the business person a good perspective of how customer preferences change throughout time in order to maximize profits.
- Our business person needs information about the most popular brands, fuels, and the preferred amount of money German drivers are willing to spend if they want their business to takeoff.

TOOLS USED:

- MySQL Workbench – we created the Snowflake schema with forwards engineering and solved queries in order to find information and make analysis.
- Knime for Data Insertion.
- Excel – we created Pivot tables and charts in order to visualize our results.

ABOUT THE DATA:

- Data was obtained from Kaggle data Science community;
- Data shows cars in AutoScout24 car market. AutoScout24 is the online vehicle market, for the purchase and sale of all types of vehicle;
- Dataset represents collection of data from 2011 to 2021.
- Each row represents a separate car purchase.
- Columns show the car model, price, type of fuel and transmission, the year of production, the horse power and producer of the car.

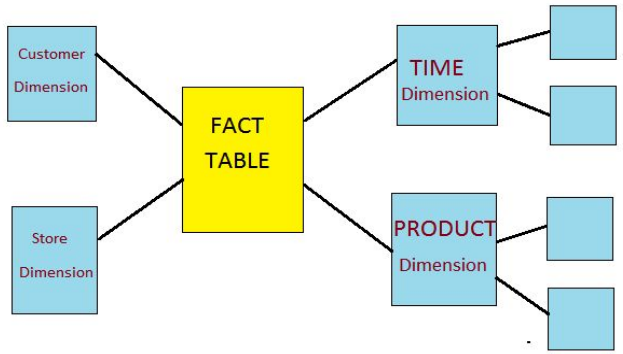
SNOWFLAKE SCHEMA

Necessary steps to produce the schema:

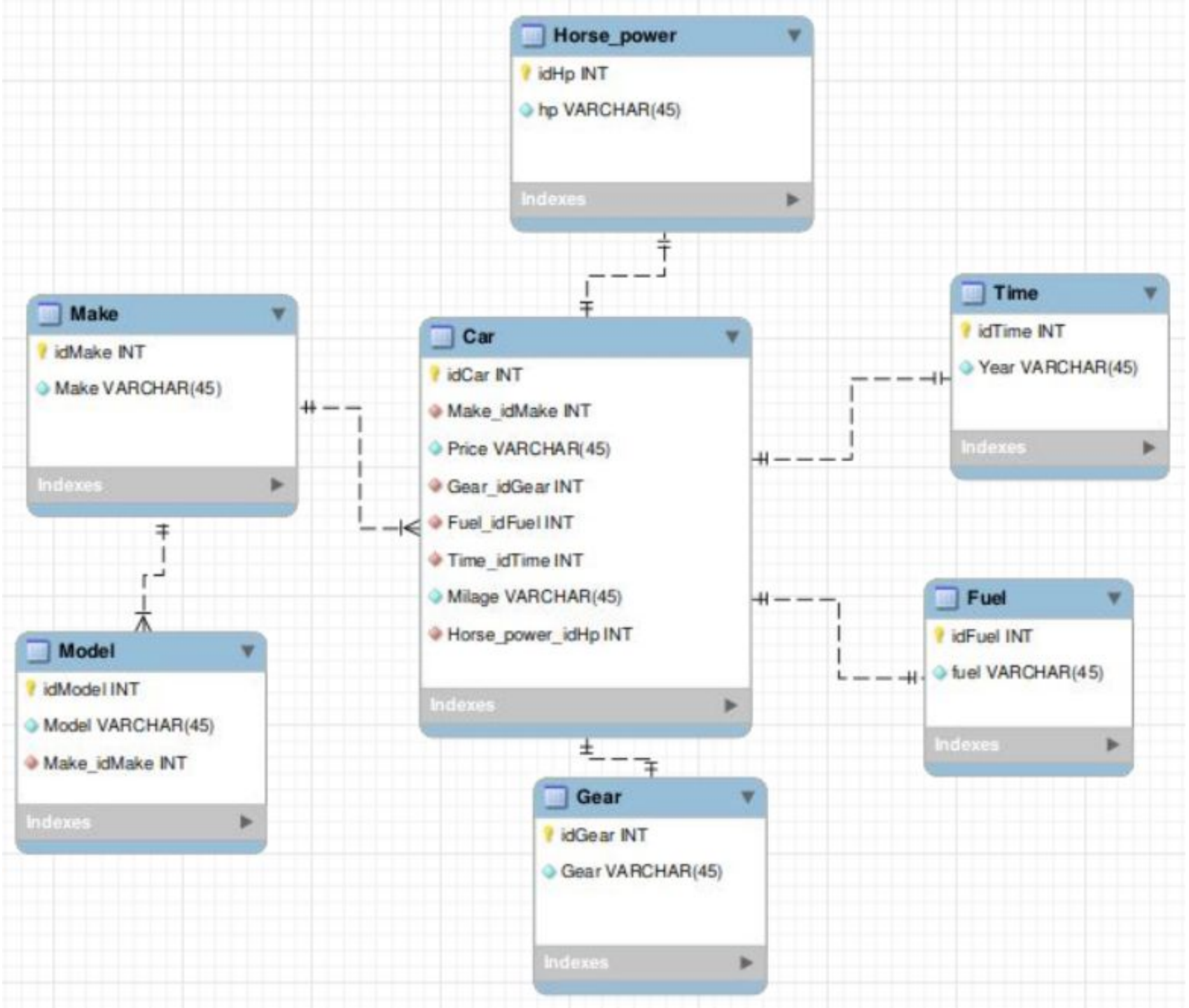
- Modelling the ER model in MySQl Workbench.
- Dimension Tables as Make, Model, Time, Gear, Fuel and Horse_Power were created first.
- After we had Dimension tables we created our Fact table, which includes data columns as well as foreign keys of Dimension tables.
- Foreign keys are created automatically, while configuring Relation types between tables.
- After having ER model we used Forward Engineering to automatically generate sql queries to produce tables and set Relation Types.

CHARACTERISTICS OF THE SNOWFLAKE SCHEMA:

- Data Split into different Dimension Tables.
- Hierarchies are divided into separate tables.
- One fact table surrounded by dimension table which are in turn surrounded by other dimension tables
- It is normalized Data Structure.
- The centralized fact table is connected with multiple dimensions. The fact table holds the main data. It includes a large amount of aggregated data.



FINAL SCHEMA PRODUCED:



PURPOSE OF SNOWFLAKE SCHEMA:

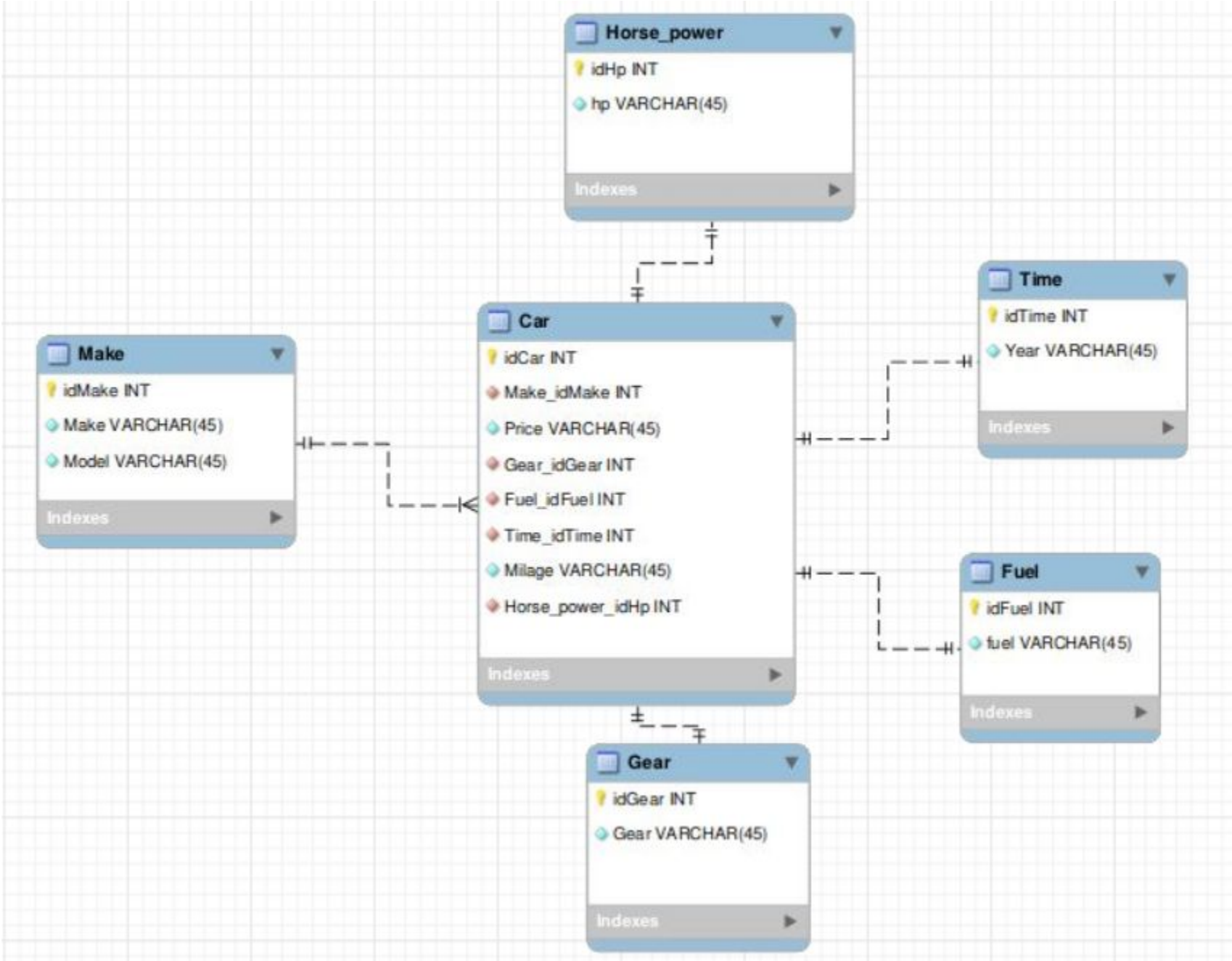
- It makes easier to implement, further dimensions. They are simply added to the Schema.
- Data is more structured.
- Better to storing data.

DATA MART SCHEMA

Necessary steps to produce the schema:

- The Data Mart schema was derived from the Snowflake schema by joining sub-dimensions.
- New schema was created using SQL queries.
- Snowflake schema was copied into the new Database using PHP My Admin.
- In our case Model table with the Make table was joined.

DATA MART SCHEMA

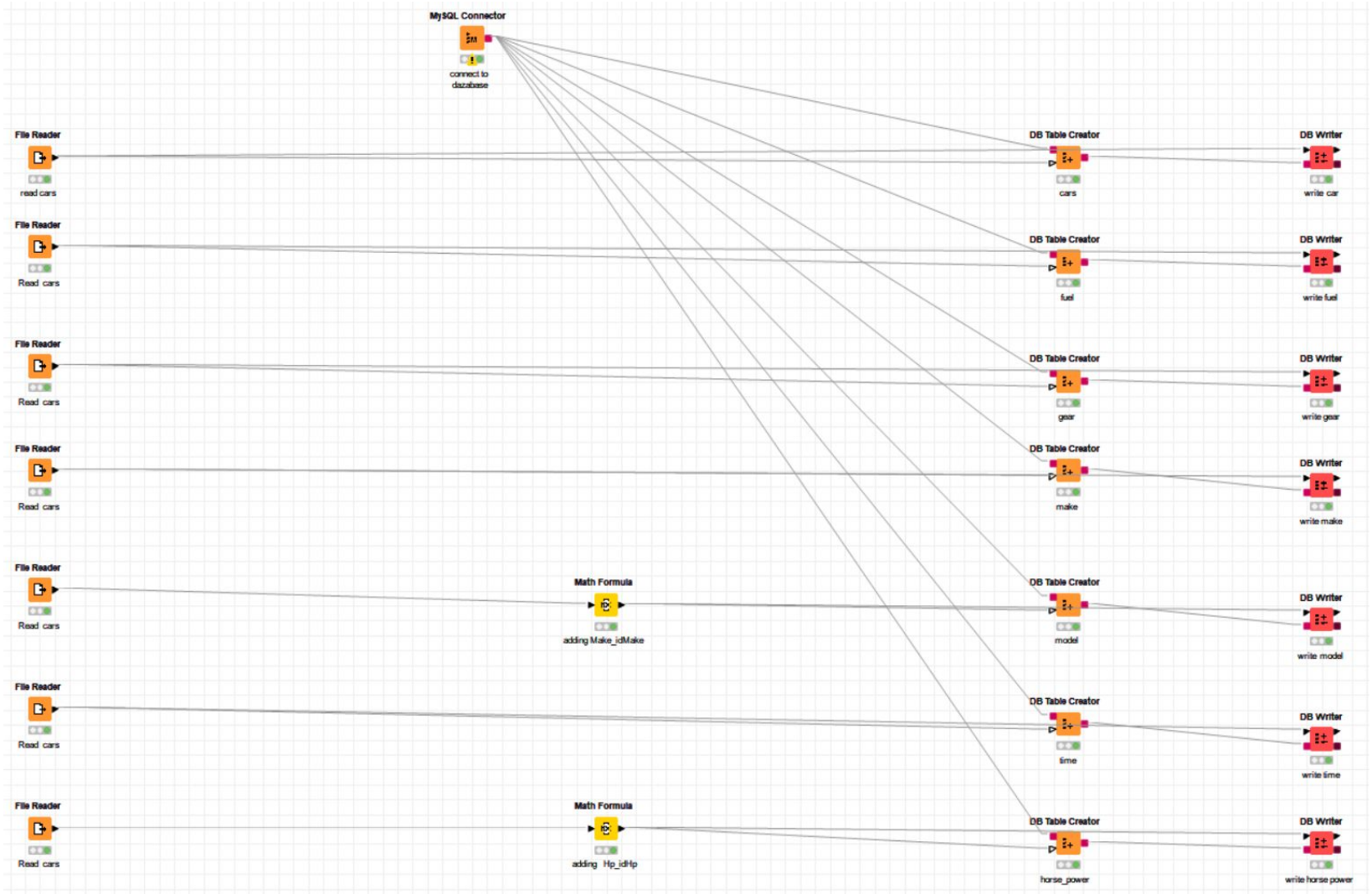


PURPOSE OF DATA MART SCHEMA:

- Fewer joins are needed when writing queries.
- Makes analysis much easier.
- Takes less space to store dimension tables, but it has more data, so that can be difficult to maintain.

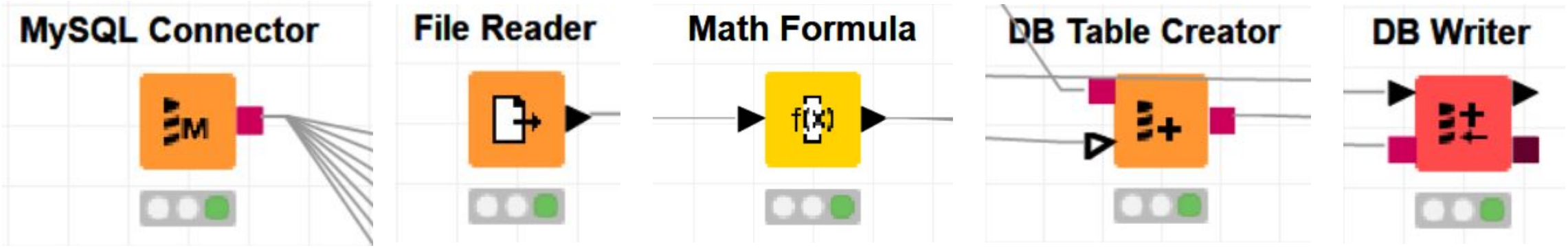
Data insertion with Knime

Workflow:



Data insertion with Knime

Used nodes:

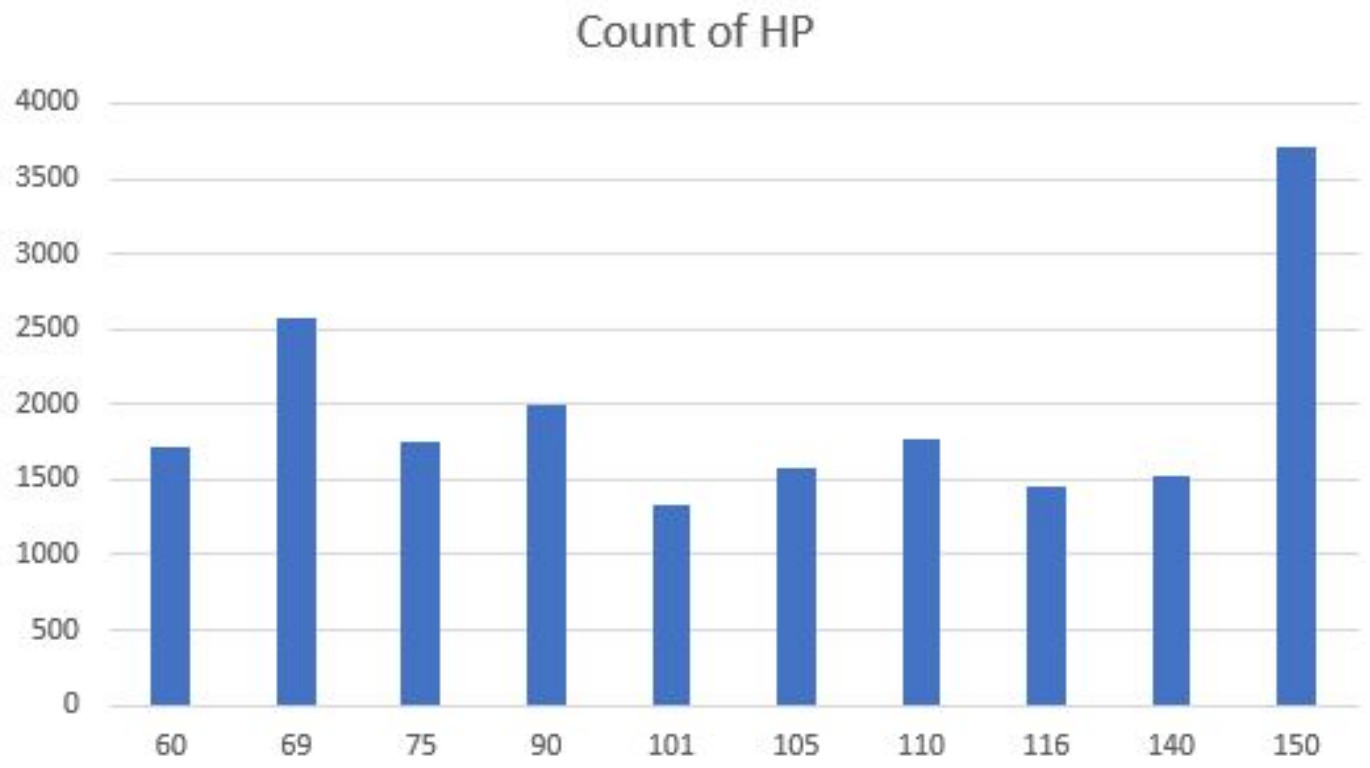


DATA ANALYSIS

Since the data analysis is a process of inspection, transforming, and modeling the data with the goal of discovering useful information we created questions in the following categories in order to make conclusions about the database:

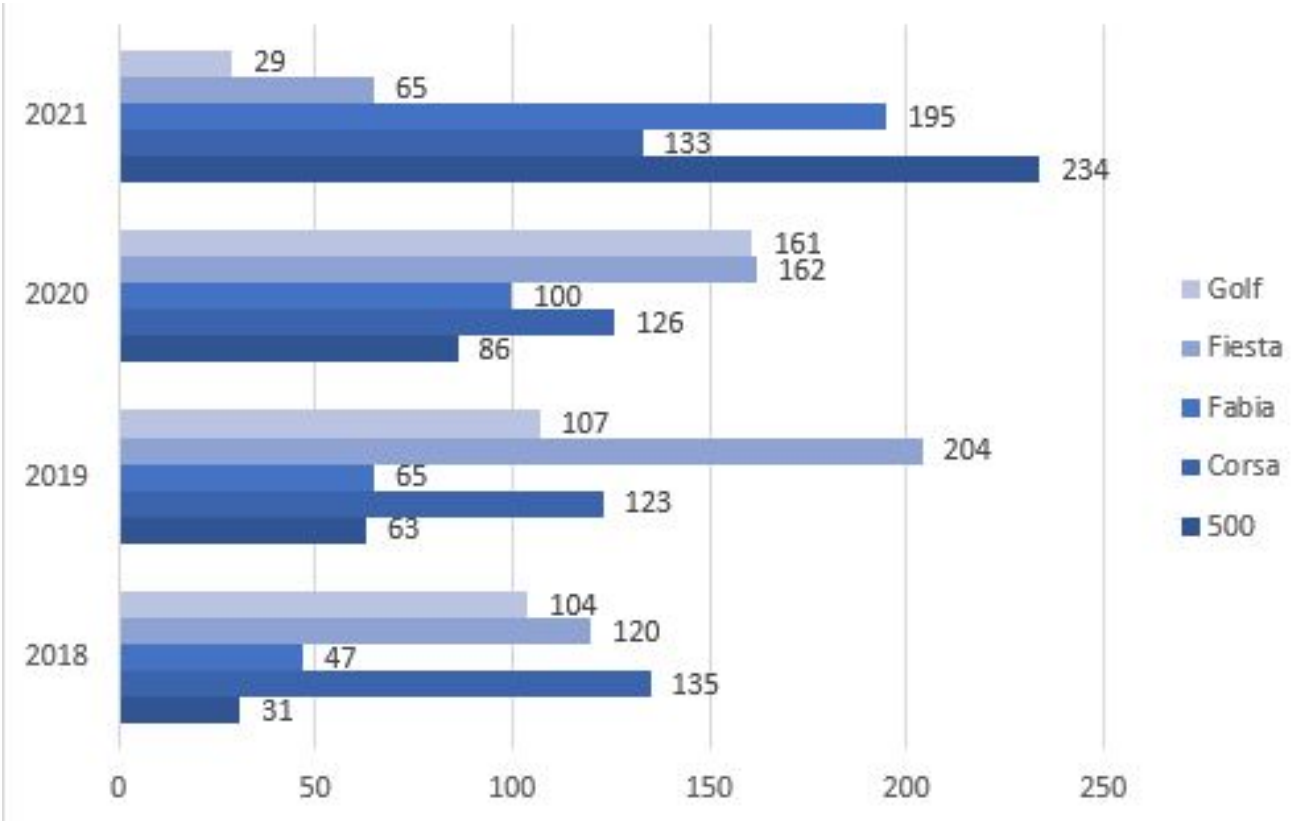
- **Environmental** companies offering electric cars, registered vehicles, etc.
- **Customer demands** most desired HP through the production years, most popular car models, etc.
- **Money** most expensive car, average price for different type of cars, etc.

What is the most desired HP by German drivers? Which companies offer this level of HP?



Companies	Count of make
Audi	474
BMW	134
Ford	551
Mazda	93
Mitsubishi	49
Opel	102
SEAT	296
Skoda	435
Volkswagen	1168
Volvo	76
Grand Total	**3378**

Which are the most popular car models according to the years the cars were produced?



What are the average prices of cars produced in 2020?

```
SELECT DISTINCT ma.make, t.year, COUNT(ma.make) oc,
                ROUND(AVG(Price)) avarage_price
FROM makes ma
LEFT JOIN cars c ON ma.Make_idMake = c.Make_idMake
LEFT JOIN times t ON c.Time_idTime = t.Time_idTime
WHERE t.year = 2020
GROUP BY ma.make, t.year;
```

Companies	Average of price
Audi	46,526.80 €
BMW	45,779.87 €
Fiat	14,673.45 €
Ford	23,185.40 €
Mercedes-Benz	42,662.71 €
Opel	18,543.95 €
Renault	18,838.76 €
SEAT	24,882.54 €
Skoda	22,164.08 €
Volkswagen	29,998.16 €
Grand Total	**28,341.16 €**

When were most diesel cars registered? When were least?

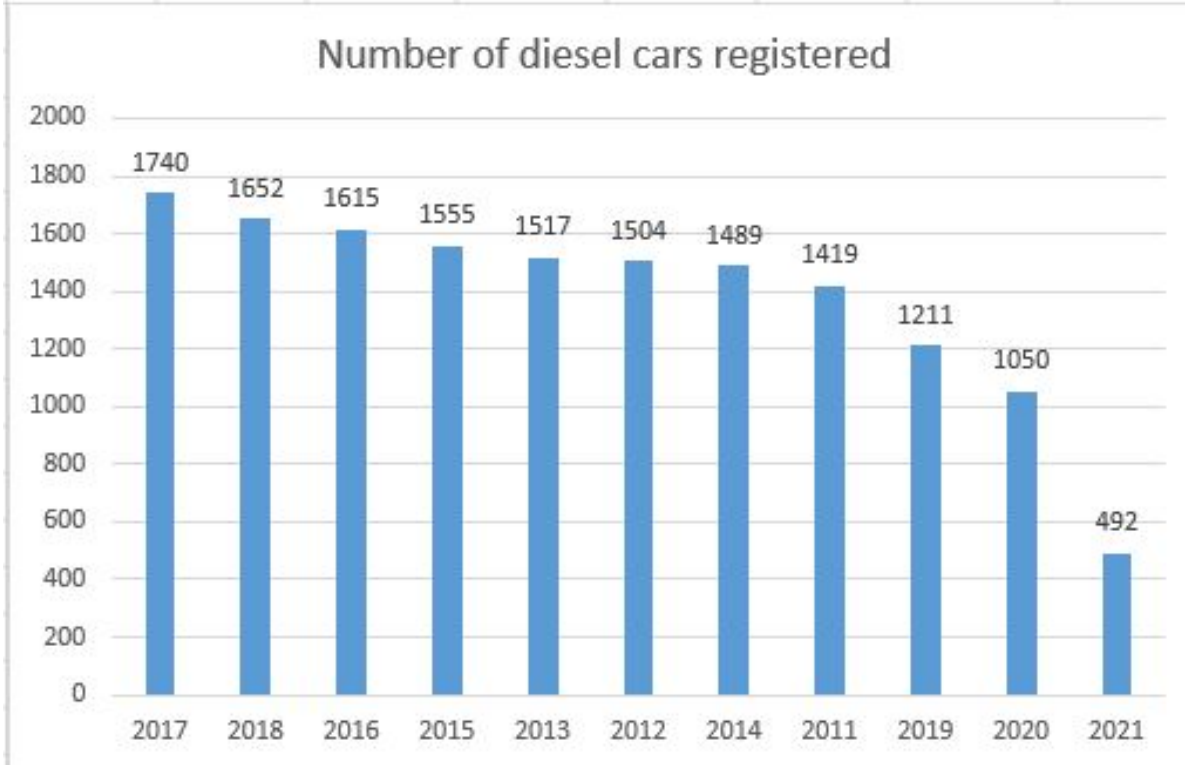
- The concept of this solution was created using sql query for calculations of each year's registration of diesel cars. After gathering all the information needed, the following Excel table was created:

```
SELECT  count(T.year) as 'Diesel cars registered'
FROM times T
JOIN fuels F on  F.Fuel_idFuel = T.Time_idTime
WHERE F.fuel = 'diesel'  && T.year = '2011'
```

Year	Number of diesel cars registered
2017	1740
2018	1652
2016	1615
2015	1555
2013	1517
2012	1504
2014	1489
2011	1419
2019	1211
2020	1050
2021	492

When were most diesel cars registered? When were least?

- After the sorted table was created, the following chart visualizes all the important information:



Which is the most expensive car registered per year?

- The concept of the solution was created using a pivot table in Excel.

year	(All)

Row Labels	Max of price
⊟812	439900
Ferrari	439900
⊟991	399911
Porsche	399911
⊟720S	309900
McLaren	309900
⊞Aventador	449900
⊟F12	1199900
Ferrari	1199900
⊟F8 Tributo	304900
Ferrari	304900
⊟Martin	398000
Aston	398000
⊟Pullman	717078
Maybach	717078
⊟S 650	717078
Mercedes-Benz	717078
⊟SLS	465000
Mercedes-Benz	465000
Grand Total	1199900

year	2019

Row Labels	Max of price
⊟600LT	226100
McLaren	226100
⊟720S	269890
McLaren	269890
⊟Cayenne	167300
Porsche	167300
⊟Continental	219800
Bentley	219800
⊟G 500	199900
Mercedes-Benz	199900
⊟G 63 AMG	252500
Mercedes-Benz	252500
⊟GT	169900
McLaren	169900
⊟Pullman	717078
Maybach	717078
⊟S 650	717078
Mercedes-Benz	717078
⊟Urus	244000
Lamborghini	244000
Grand Total	717078

year	2020

Row Labels	Max of price
⊟991	349000
Porsche	349000
⊟Bentayga	235900
Bentley	235900
⊟Continental	244777
Bentley	244777
⊟Flying Spur	248900
Bentley	248900
⊟Martin DB11	199900
Aston	199900
⊟Martin DBX	187900
Aston	187900
⊟Martin V8	155900
Aston	155900
⊟Panamera	162900
Porsche	162900
⊟RS Q8	149890
Audi	149890
⊟S 500	153400
Mercedes-Benz	153400
Grand Total	349000

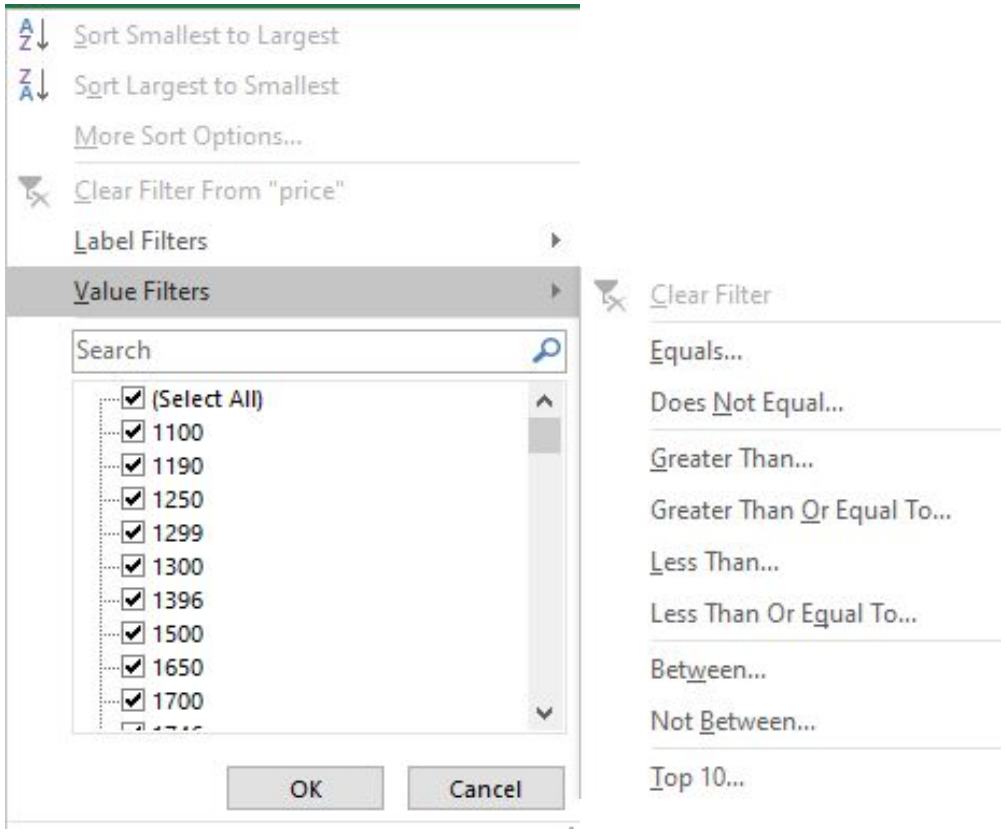
year	2021

Row Labels	Max of price
⊟812	439900
Ferrari	439900
⊟911	237880
Porsche	237880
⊟720S	309900
McLaren	309900
⊞Aventador	449900
⊟F8 Tributo	304900
Ferrari	304900
⊟Flying Spur	257900
Bentley	257900
⊟GLS 600	238900
Mercedes-Benz	238900
⊟Martin DB11	225007
Aston	225007
⊟Martin DBX	224900
Aston	224900
⊟RS Q8	249000
Audi	249000
Grand Total	449900

Which is the most expensive car registered per year?

- Explanation of the pivot table:
- Filters: year
- Rows: model, make
- Values: Max of price:

  The shown result is filtered, so only top 10 models appear



Which companies produce electric cars?

- The concept for this solution was created using the following SQL query:

```
SELECT DISTINCT M.Make, F.fuel
FROM fuels F
JOIN  makes M ON F.Fuel_idFuel = M.Make_idMake
WHERE F.fuel = 'electric'
ORDER BY M.Make ASC
```

- The result is:



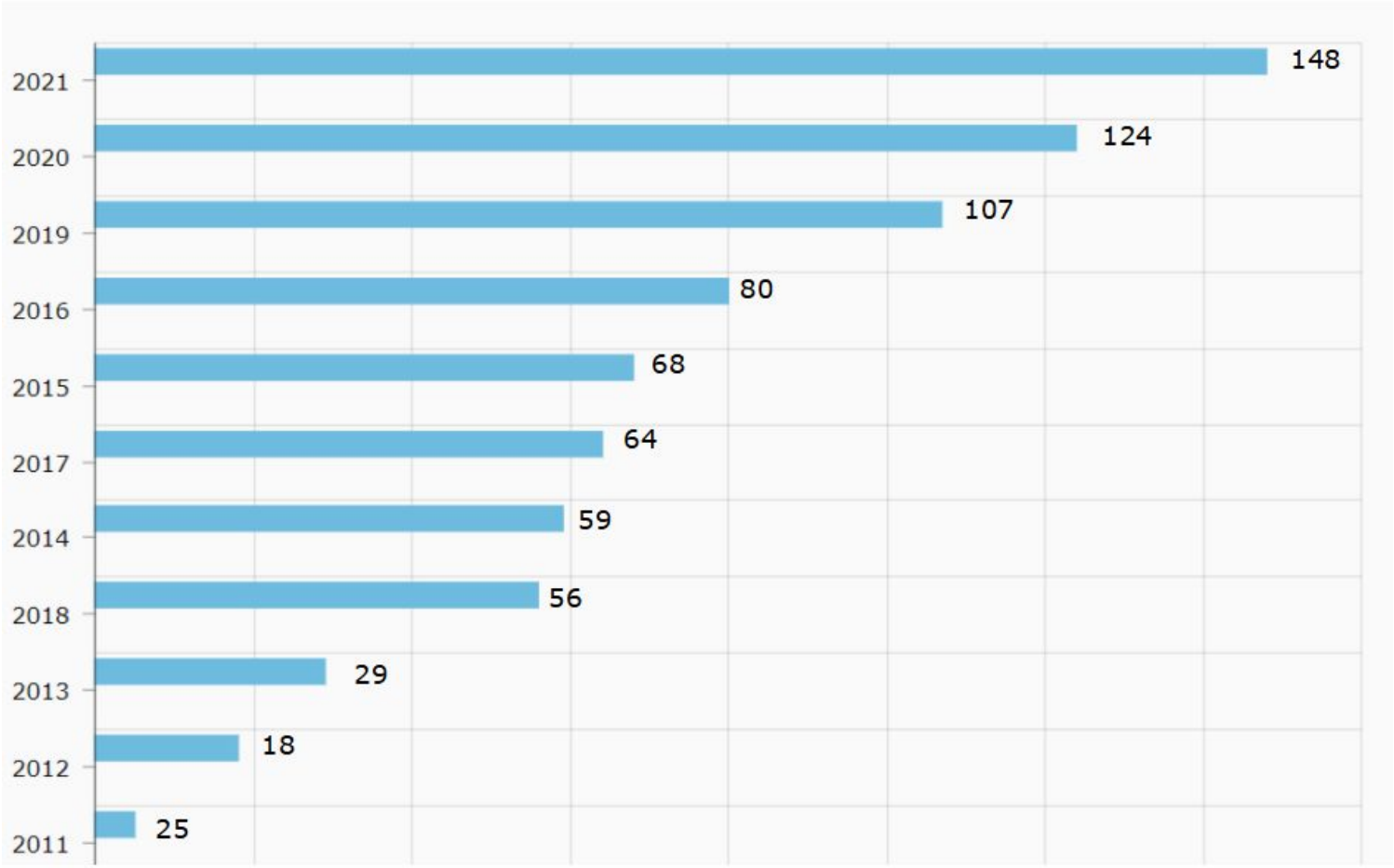

	Make	fuel
▶	Aixam	Electric
	Audi	Electric
	BMW	Electric
	Citroen	Electric
	DS	Electric
	Estrima	Electric
	Fiat	Electric
	Ford	Electric
	Honda	Electric
	Hyundai	Electric
	Jaguar	Electric
	Kia	Electric

	Make	fuel
	Mazda	Electric
	Merced...	Electric
	MINI	Electric
	Mitsubishi	Electric
	Nissan	Electric
	Opel	Electric
	Others	Electric
	Peugeot	Electric
	Polestar	Electric
	Porsche	Electric
	Renault	Electric
	SEAT	Electric

	Skoda	Electric
	smart	Electric
	Tazzari	Electric
	Tesla	Electric
	Toyota	Electric
	Volksw...	Electric
	Volvo	Electric
	Zhidou	Electric

In which years were most electric cars registered?

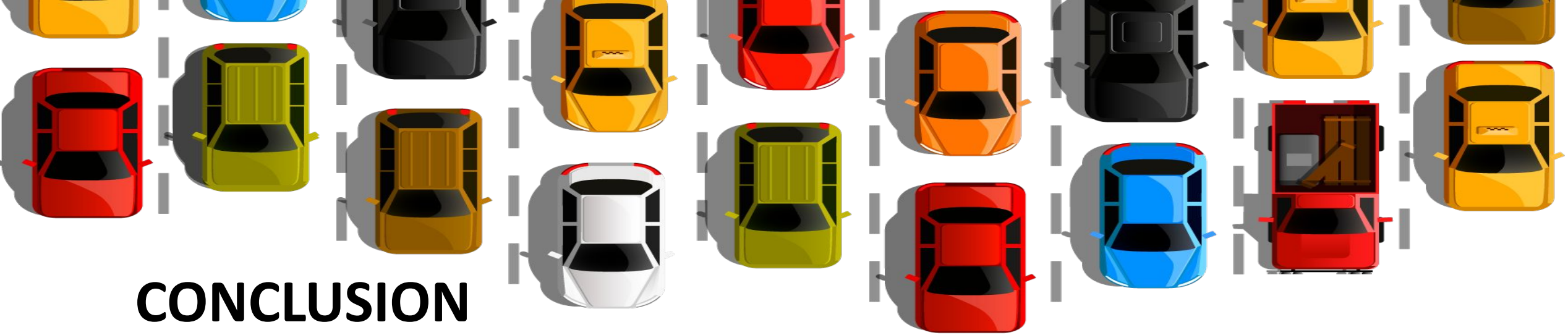
- The following chart allows us to observe the incremental growth of sales of electric cars through the years.



What is the average price for electric cars? How does it compare to gasoline/diesel cars?

- Electric cars average price is 27% higher than the average Diesel car price and 52% higher than the average Gasoline car price.

Average Electric Cars Price	Electric / Diesel	Electric / Gasoline
23004.94	127.01	152.62



CONCLUSION

- With the results provided, we hope we gave to future car dealers all the understanding they need for opening a successful car business. Even those who are simply willing to get a new car also can have a better understanding of the car market.

ANY QUESTIONS? IF NOT, WE: