

# CARS IN GERMANY

Sevde Yanik  
Aleksandra Pasheva  
Anna Campi  
Rami Bader  
Assylbek Bugybay

THU  
Computer Science  
89073, Ulm

## ABSTRACT

The car industry is one of the most dynamic markets out there. The constant technology development, environmental concerns, cost pressures and customers demands are only a few of the factors that influence the nonstop need for development.

In order to maximize profits a business should have an overall understanding of customers evolving wishes when wanting to buy a car. We gathered some data points in order to create a Data Warehouse and extract information about changes in car features and preferences between 2013 and 2020.

**Keywords:** cars, fuel, consumption, driving, analysis.

## 1. INTRODUCTION

A person wanting to enter the automotive industry needs a good understanding of the current market demands as well as 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 take off.

### A. Scenario

The outcome of the analysis will show the evolving demands of car features throughout the years by using existing data of German drivers purchases.

We want to give the business person a good perspective of which cars he should be more inclined to sell.

### B. Structure of paper

We structured the paper firstly by describing how

the data was obtained. After that, we show how the Data warehouse was built and what analysis was done on it.

Lastly we conclude our projects by showing and commenting on our results.

## 2. API

We obtained the data from <sup>[1]</sup>[kaggle.com](https://www.kaggle.com).

The data set contains information about car features from 2013 up to 2020. Each row represents a separate car purchase and includes: the car model and price, what type of fuel and gear is used, the year of purchase, the horse power and who produced the car.

### 3. CONCEPT FOR PROBLEM

We want to be able to solve the following problems:

- 1) Which cars should the business person decide to sell?
- 2) How have environmental concerns changed the sales of electric cars throughout the years?

We used the following tools to conduct our research:

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

### 4. SNOWFLAKE SCHEMA

#### A. Approach

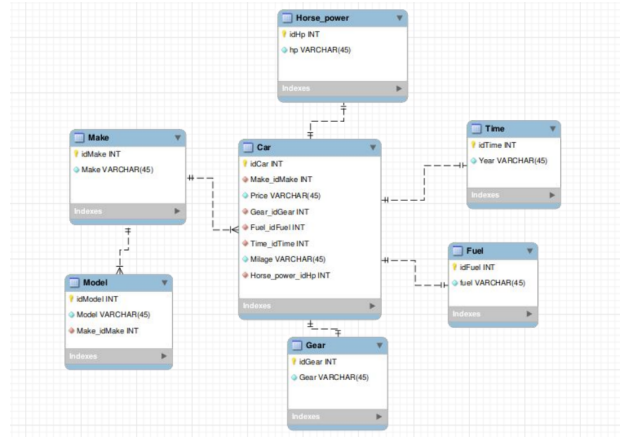
First of all, we modelled our ER model in MySQL Workbench. Dimension Tables as Make, Model, Time, Gear, Fuel and Horse\_Power were created first. After we had Dimension tables we created our Fact table, which includes data columns as well as foreign keys of Dimension tables. Foreign keys are created 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.

Then we used Ktime Analytical Platform and created a workflow to import data from our csv file into our Database. Then using MySQL, we wrote queries to make analysis on the data.

Last step was in order to provide visualization in our report, we used Excel and created pivot tables and charts.

The Car Snowflake schema is provided below:



#### 1. Make:

- idMake: Primary key of Make
- Make: Producer company of car
- Model: Model of the car

#### 2. Fuel:

- idFuel: Primary key of Fuel table
- Fuel: Type of fuel consumed by car

#### 3. Horse\_Power:

- idHp: Primary key of Horse\_power
- hp: Horse Power of car

#### 4. Time:

- idTime: Primary key of Time
- Year: year of car was produced

#### 5. Gear:

- idGear: Primary key of Gear
- Gear: type of transmission

#### 6. Car:

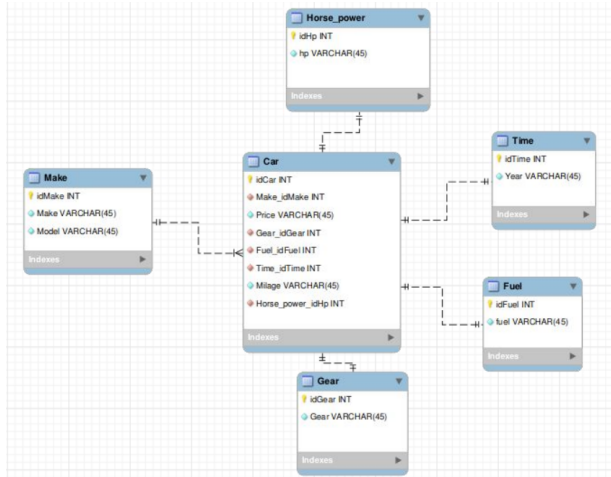
- idCar: Primary key of Car
- Price: Price of the car
- Milage: Distance travelled by Car

### 5. DATA MART

The Data Mart schema was derived from the Snowflake schema by joining sub-dimensions. We merged the Model table with the Make table and we used SQL queries to create a new schema. We copied the Snowflake schema into the new Database with PHP My Admin.

The purpose of Data Mart is to make reporting and analysis of data easier.

The Car Data Mart schema is provided below:



HP	Count of HP	Companies	Count of make
60	1708	Audi	474
69	2576	BMW	134
75	1758	Ford	551
90	2000	Mazda	93
101	1337	Mitsubishi	49
105	1581	Opel	102
110	1762	SEAT	296
116	1461	Skoda	435
140	1520	Volkswagen	1168
150	3708	Volvo	76
<b>Grand Total</b>	<b>19411</b>	<b>Grand Total</b>	<b>3378</b>

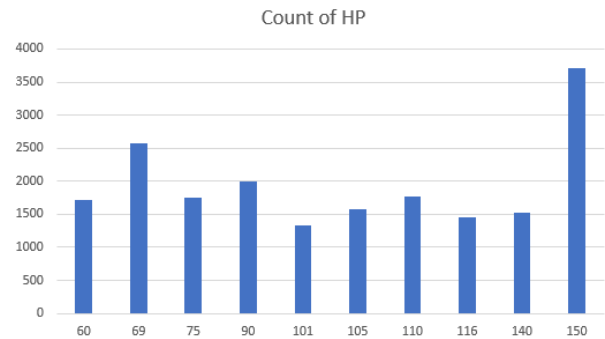
Table 1.1 and 1.2

## 6. ANALYSIS

After deciding on points to investigate, firstly the sql queries were written for each analysis question. The second step was to create pivot tables to see the results in a more structured way and for visualization purposes charts were created for these pivot tables.

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

First using SQL we wrote the query to find out about the most popular HP and which companies offered models with this HP. Then using Excel the tables that show the results were created and 150 was the level of HP that occurred most. And the companies that have this level of HP and the number of them were found using filtering. (see tables 1.1 and 1.2)

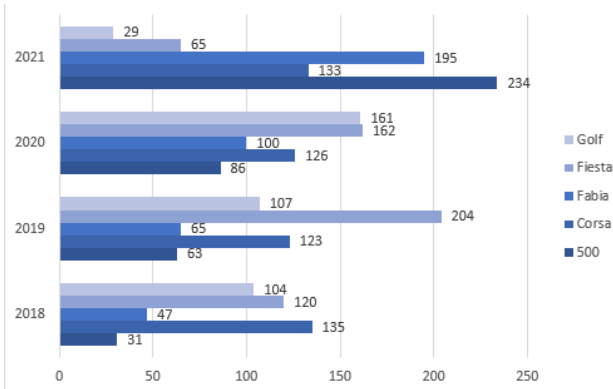


### 6.2 Which are the most popular car models according to the years the cars were produced?

The stored procedure for investigating the question is written in SQL and then using Excel the pivot table is created filtering by the years, because the years between 2011-2021 are in the database and last years are more relevant we pick the top 5 models that were produced in the years between 2018 to 2021 and see the results in table 2.1.

Years	500	Corsa	Fabia	Fiesta	Golf	Grand Total
2018	31	135	47	120	104	437
2019	63	123	65	204	107	562
2020	86	126	100	162	161	635
2021	234	133	195	65	29	656
<b>Grand Total</b>	<b>414</b>	<b>517</b>	<b>407</b>	<b>551</b>	<b>401</b>	<b>2290</b>

Table 2.1



### 6.3 Which company sells the most cars produced in 2020? What is their average price for cars?

The queries to find the companies that have sold more cars produced in 2020 than others and their average price of cars were written in SQL. Later the results are presented in tables having the top 10 companies and their average prices using Excel.

Companies	Number of productions	Companies	Average of price
Audi	377	Audi	46,526.80 €
BMW	181	BMW	45,779.87 €
Fiat	183	Fiat	14,673.45 €
Ford	680	Ford	23,185.40 €
Mercedes-Benz	187	Mercedes-Benz	42,662.71 €
Opel	238	Opel	18,543.95 €
Renault	316	Renault	18,838.76 €
SEAT	205	SEAT	24,882.54 €
Skoda	297	Skoda	22,164.08 €
Volkswagen	721	Volkswagen	29,998.16 €
<b>Grand Total</b>	<b>3385</b>	<b>Grand Total</b>	<b>28,341.16 €</b>

Table 3.1 and 3.2

### 6.4 Which gear type is more popular? What is the average price for different gear types?

Using Excel the number of times different gear types occur in the database and their average price are shown in table 4.1.

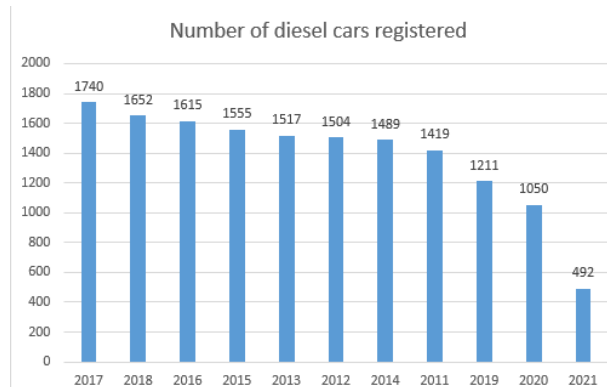
Gear Type	Count of make	Average of price
Automatic	15787	28,216.00 €
Manual	30380	10,569.03 €
Semi-automatic	56	8,424.98 €
<b>Grand Total</b>	<b>46223</b>	<b>16,593.58 €</b>

Table 4.1

### 6.5 When were most diesel cars registered? When were the least?

Using SQL query we calculated how many diesel cars were sold each year. In Excel we created a table, which was sorted from most to least. As the last step a chart is presented, so the years and the diesel cars registered are visible.

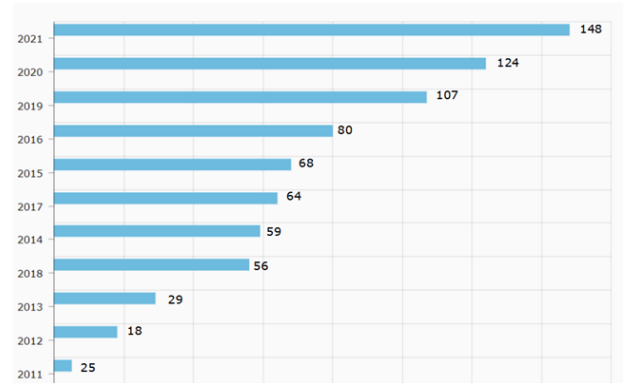
Year	Number of diesel cars registered
2017	1740
2018	1652
2016	1615
2015	1555
2013	1517
2012	1504
2014	1489
2011	1419
2019	1211
2020	1050
2021	492



### 6.6 Which is the most expensive car registered per year?

Pivot table was created with filters - year; rows - model,make; values - max of price. The result price is filtered so in the visualisation we see top 10 results in descending order. As the filter is year we are able to select each year and see the appropriate result of the most expensive car. Below is the pivot table with all years filter.

year	(All)
Row Labels	Max of price
812	439900
Ferrari	439900
991	399911
Porsche	399911
720S	309900
McLaren	309900
Aventador	449900
Lamborghini	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



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

Queries were used in SQL to create a table with the average price for electric cars and how that compares to Diesel and Petrol cars in percentage. The results of the query are presented in a table. Electric cars average price is 12% higher than the average Diesel car price and 15% higher than the average Gasoline car price.

## 6.7 Which companies produce electric cars?

SQL query was created, so it selects the fuel type and the company name and then filters with category 'electric'. The result is sorted in ascending alphabetical order.

## 6.8 In which years were most electric cars registered?

SQL was used to write the queries to obtain the wanted results. Data visualization is done with a vertical bar chart. 2021 is the year in which most electric cars were purchased.

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

## 7. CONCLUSION

As a recap, with our Data Warehouse we analyzed the following points:

- Most desired horsepower by drivers and which companies were fulfilling that desire.
- Popular cars according to production year.
- Best selling companies having production year 2020 and their average price.
- Most desired gear type and average price for different gear types.
- Drivers' behaviour towards diesel cars, when did

they get them the most and when the least.

- Most expensive car each year.
- Companies producing E-Cars.

- Drivers' behaviour towards electric cars, when did they get them the most.
- Average price of E-Cars and we compared that to both gasoline and diesel.

And we found out that:

150hp is the most desired hp and VW was the company to produce that the most. Fiesta was the most desired model collectively through the years 2018-2021, followed by Corsa. In 2020 VW had the most sales with about 30k € as an average price. The manual gear was the most desired gear with an average price of 10.5k€.

Most diesel cars were registered in 2017 and the least were registered in 2021. The result shows that for all the years the most expensive car is Ferrari 812 with price 439900€ followed by Porsche 991 with price 399911€.

Most electric cars were registered in 2021 and the average price for an electric car was about 23k €. Many companies now produce electric cars for example: Ford, Mercedes-Benz, Tesla, Honda, BMW, etc.

Having those results, we hope that we provide future car dealers all the understanding they need to

open their successful cars business. Even those who are simply willing to get a new car can also have a better understanding of the car market in case they got broke in future and needed to sell their new car so they can find a buyer.

## **8. REFERENCES**

- [1] <https://www.kaggle.com/ander289386/cars-germany?select=autoscout24-germany-dataset.csv>
- [2] <https://support.microsoft.com/en-us/office/create-a-pivottable-to-analyze-worksheet-data-a9a84538-bfe9-40a9-a8e9-f99134456576>
- [3] <https://www.knime.com/knime-self-paced-courses>