

# Analysis of the cars

Data Curation by Mike Silva

## General Dataset Information:

File Name:	cars_data
Description:	A large collection of cars with their characteristic
Dataset Details:	19,238 Rows & 18 Columns
Size:	2,531KB ( 2.531MB )
Source:	Kaggle - <a href="#">Dataset Link</a>

## Data Profile:

- Import Dataset to tableau
- Examine dataset through tableau
- Examine the characteristic and choose which manufacturers are worth working
- Export data into SQL
- Apply SQL filter to the dataset
- Examine the cars prices and choosing the correct models
- Notable Features ( Prior to Export )
  - 10 unique manufacture
  - Between the manufacturers are thousands of cars models
  - The cars chosen has a price that is affordable
  - Contains the top 10 most popular cars
- Recorded Inconsistency:
  - Even if the cars have a huge mileage the price was high
  - Outliers from low-volume manufacturers are distorting price averages
  - No all manufacturers has the same ratio of cars models
  - No a single car have a review, being hard to determine which one is good
- Data Wrangling:
  - Using SQL, to determine how many models did a manufacturer car have and

the total

- Via sql I created a table with the list of the top 10 cars manufacturer being a total of 15,934 car models
- Hyundai, Toyota, and Mercedes-Benz have the most cars, so they show multiple models. The other manufacturers only show one model each due to fewer cars
- The filter selected models with the highest car counts.
- It then displayed the years with the most cars for each model.
- After isolating all cars the filter did got the result from 15,934 models it got reduced to 1,722 cars

- **Data Table Schema:**

Field	Type
• ID	• INTEGER
• PRICE	• REAL
• LEVY	• REAL
• MANUFACTERER	• STRING
• MODEL	• STRING
• PROD. YEAR	• INTEGER
• CATEGORY	• REAL
• LEATHER INTERIOR	• BOOLEAN
• FUEL TYPE	• STRING
• ENGINE VOLUME	• REAL
• MILEAGE	• REAL
• CYLINDERS	• INTEGER
• GEAR BOX TYPE	• STRING
• DRIVE WHEELS	• STRING
• DOORS	• INTEGER
• WHEEL	• STRING
• COLOR	• STRING
• AIRBAGS	• INTEGER

- Data Table:

	Manufacturer	Model	Prod._year	total_cars
1	HYUNDAI	Sonata	2015	258
2	HYUNDAI	Sonata	2016	205
3	HYUNDAI	Sonata	2014	166
4	TOYOTA	Camry	2014	219
5	TOYOTA	Prius	2010	218
6	MERCEDES-BENZ	E 350	2011	121
7	MERCEDES-BENZ	E 350	2014	119
8	FORD	Fusion	2013	92
9	CHEVROLET	Cruze	2014	54
10	BMW	X5	2012	64
11	LEXUS	RX 450	2010	79
12	HONDA	FIT	2013	52
13	NISSAN	Juke	2013	42
14	VOLKSWAGEN	Jetta	2014	33

- USED SQL CODE:

```
-- Top 3 models for Hyundai
SELECT * FROM (
    SELECT 'HYUNDAI' AS Manufacturer, Model, "Prod._year", COUNT(*) AS total_cars
    FROM cars_data
    WHERE Manufacturer = 'HYUNDAI'
    GROUP BY Model, "Prod._year"
    ORDER BY total_cars DESC
    LIMIT 3
) AS hyundai_models

UNION ALL

-- Top 2 models for Toyota
SELECT * FROM (
```

```
SELECT 'TOYOTA' AS Manufacturer, Model, "Prod._year", COUNT(*) AS total_cars
FROM cars_data
WHERE Manufacturer = 'TOYOTA'
GROUP BY Model, "Prod._year"
ORDER BY total_cars DESC
LIMIT 2
) AS toyota_models
```

UNION ALL

-- Top 2 models for Mercedes-Benz

```
SELECT * FROM (
  SELECT 'MERCEDES-BENZ' AS Manufacturer, Model, "Prod._year", COUNT(*) AS
total_cars
  FROM cars_data
  WHERE Manufacturer = 'MERCEDES-BENZ'
  GROUP BY Model, "Prod._year"
  ORDER BY total_cars DESC
  LIMIT 2
) AS mercedes_models
```

UNION ALL

-- Top 1 model for Ford

```
SELECT * FROM (
  SELECT 'FORD' AS Manufacturer, Model, "Prod._year", COUNT(*) AS total_cars
  FROM cars_data
  WHERE Manufacturer = 'FORD'
  GROUP BY Model, "Prod._year"
  ORDER BY total_cars DESC
  LIMIT 1
) AS ford_models
```

UNION ALL

-- Top 1 model for Chevrolet

```
SELECT * FROM (
  SELECT 'CHEVROLET' AS Manufacturer, Model, "Prod._year", COUNT(*) AS
total_cars
  FROM cars_data
  WHERE Manufacturer = 'CHEVROLET'
  GROUP BY Model, "Prod._year"
  ORDER BY total_cars DESC
  LIMIT 1
```

```
) AS chevrolet_models
```

```
UNION ALL
```

```
-- Top 1 model for BMW
```

```
SELECT * FROM (  
    SELECT 'BMW' AS Manufacturer, Model, "Prod._year", COUNT(*) AS total_cars  
    FROM cars_data  
    WHERE Manufacturer = 'BMW'  
    GROUP BY Model, "Prod._year"  
    ORDER BY total_cars DESC  
    LIMIT 1
```

```
) AS bmw_models
```

```
UNION ALL
```

```
-- Top 1 model for Lexus
```

```
SELECT * FROM (  
    SELECT 'LEXUS' AS Manufacturer, Model, "Prod._year", COUNT(*) AS total_cars  
    FROM cars_data  
    WHERE Manufacturer = 'LEXUS'  
    GROUP BY Model, "Prod._year"  
    ORDER BY total_cars DESC  
    LIMIT 1
```

```
) AS lexus_models
```

```
UNION ALL
```

```
-- Top 1 model for Honda
```

```
SELECT * FROM (  
    SELECT 'HONDA' AS Manufacturer, Model, "Prod._year", COUNT(*) AS total_cars  
    FROM cars_data  
    WHERE Manufacturer = 'HONDA'  
    GROUP BY Model, "Prod._year"  
    ORDER BY total_cars DESC  
    LIMIT 1
```

```
) AS honda_models
```

```
UNION ALL
```

```
-- Top 1 model for Nissan
```

```
SELECT * FROM (  
    SELECT 'NISSAN' AS Manufacturer, Model, "Prod._year", COUNT(*) AS total_cars  
    FROM cars_data
```

```
WHERE Manufacturer = 'NISSAN'  
GROUP BY Model, "Prod._year"  
ORDER BY total_cars DESC  
LIMIT 1  
) AS nissan_models
```

UNION ALL

```
-- Top 1 model for Volkswagen  
SELECT * FROM (  
    SELECT 'VOLKSWAGEN' AS Manufacturer, Model, "Prod._year", COUNT(*) AS  
total_cars  
    FROM cars_data  
    WHERE Manufacturer = 'VOLKSWAGEN'  
    GROUP BY Model, "Prod._year"  
    ORDER BY total_cars DESC  
    LIMIT 1  
) AS volkswagen_models;
```

```
-- Optional: You can also apply an ORDER BY clause at the end if needed  
-- ORDER BY total_cars DESC;
```