

# Data Analysis on Cars

Ganguru, Rohith Sai, Gorantla, Manoj, Yarlagadda, Akhil Babu, Makkena,  
Ajay, Chundu, Avinash, Smith, Carter

Northwest Missouri State University, Maryville MO 64468, USA

## 1 Introduction

There are many theories about how the market decides the prices of used cars offered for sale by private parties, particularly in comparison to new cars. According to one theory, new car dealers can generate more demand by making a greater effort to sell a car. This dataset describes about the Cars and the make model of the car manufactured. Full detailed information about the car is included in the data set for example the Mileage of the car and body type and the estimated cost of the used car according to the market.

## 2 A detailed explanation of your proposed methodology with respect to 5 Vs

**Volume:** Big data can be used to describe a set of data that is sufficiently massive. However, the definition of big data is subjective and subject to change depending on the market's supply of computer power. Since our dataset is related about cars. How many cars are being buy/sold in particular time and based on the specific brands.

**Velocity:** For businesses who require their data to flow fast so that it is available when needed to make the best business decisions, this is a crucial factor. Since there will be many models of cars in the dataset it contains the data about mileage of the car that have been driven where we need to retrieve the results soon based on the type users search.

**Variety:** Data might originate both inside and outside of an organization. The standardization and distribution of all the data being gathered pose a problem in terms of variety. Since this dataset might contain unstructured data like the NA, NULL, NO DATA FIELDS of the particular film.

**Veracity:** Veracity, overall, refers to the level of trust there is in the collected data. since the data format overall, all cars are going to be the same we may not notice veracity.

**Value:** It is necessary to be able to extract value from big data because the value of big data greatly depends on the insights that can be obtained from them. The value of the data set is typical going cost of the cars, that actually depends on the model type, the condition of the car and the mileage on the car, by considering all these aspects the value of the car can be decided in the market.

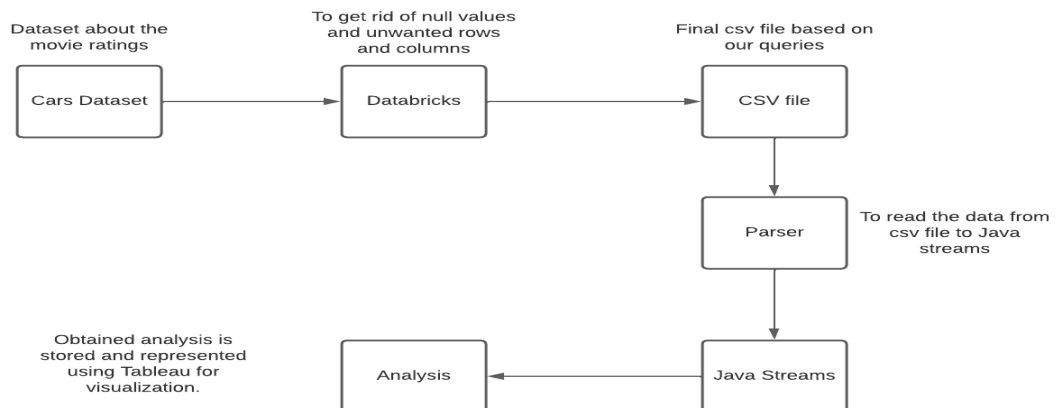
### 3 Goals of your project

- 1. List all Maruthi Brand With Model Alto?
- 2. What are the car Brands Manufactured between 2011 to 2015?
- 3. What are the brands which run on diesel fuel?
- 4. Specify all the Car Brands, Models, years, and prices between 3L to 6L?
- 5. List Brand with Model and year which are in the location Delhi?
- 6. List out all the brands, models, and counts with BMW?
- 7. List SUV body type Brand and Models?
- 8. List all first-owner car brands with Models?
- 9. List the car model, Brand, Year, and Km that are driven less than 10000 kilometers.
- 10. Name the honda cars with the model, Body type, and year manufactured after 2015.

### 4 Tools and Technologies you plan to use:

- Data Bricks
- IO Streams
- Java
- Spring Boot
- Overleaf
- To segregate the data we use data bricks i.e, to remove the null values and duplicate data. Also, we will remove unwanted rows and columns from that static dataset by developing queries. We will take the final output CSV file and using IO streams we will do some analysis and display the data in text format.

### 5 A detailed methodology in the form of a block diagram



## 6 A thorough and detailed explanation of all the steps of your implementation

- Initially we need to find the data set to display the set of goals depend on the data.
- By using the data bricks tool to clean the unstructured data and extracting the csv file from the data bricks.
- After that created a project with name BigDataFinalProject using spring boot Application.
- After creating project we need to create a package, then we need to import the csv data set file into resource folder.
- We need to create a main class and fetch the csv file using "path" library, then we created an data object class named as Car and we need to initiate the required attributes with corresponding data types and using those attributes created a constructor.
- In main class,from the data set the each row will be read and inserted in the Car object and store the object into the list by using collections and Map.
- Using the Car object List we are created the Goal Analysis by using the IOStreams we are analysed the result of the goal.
- The result output is storing into the text file by using PrintWritable Library.
- we reading the time stamp while executing the each goal and calculating the time elapsed to executing the goal.

## 7 Discussion of your results of each goal in detail by providing screenshots

### 7.1 List all Maruthi Brand With Model Alto?

```

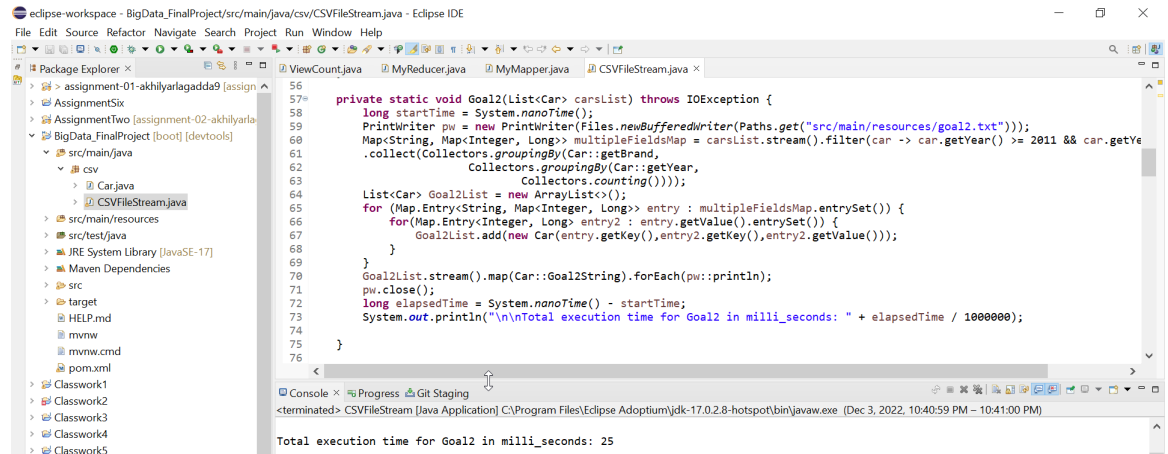
43
44 private static void Goal1(List<Car> carsList) throws IOException {
45     long startTime = System.nanoTime();
46     PrintWriter pw = new PrintWriter(Files.newBufferedWriter(Paths.get("src/main/resources/goal1.txt")));
47     carsList.stream()
48         .filter(car -> car.getBrand().equalsIgnoreCase("Maruti") && car.getModel().equalsIgnoreCase("Alto"))
49         .map( Car :: toString1 )
50         .forEach(pw::println);
51     pw.close();
52     long elapsedTime = System.nanoTime() - startTime;
53     System.out.println("\n\nTotal execution time for Goal1 in milli_seconds: " + elapsedTime / 1000000);
54 }
55
56

```

Console × Progress Git Staging  
 <terminated> CSVFileStream [Java Application] C:\Program Files\Eclipse Adoptium\jdk-17.0.2.8-hotspot\bin\javaw.exe (Dec 3, 2022, 10:40:59 PM – 10:41:00 PM)  
 Total execution time for Goal1 in milli\_seconds: 22

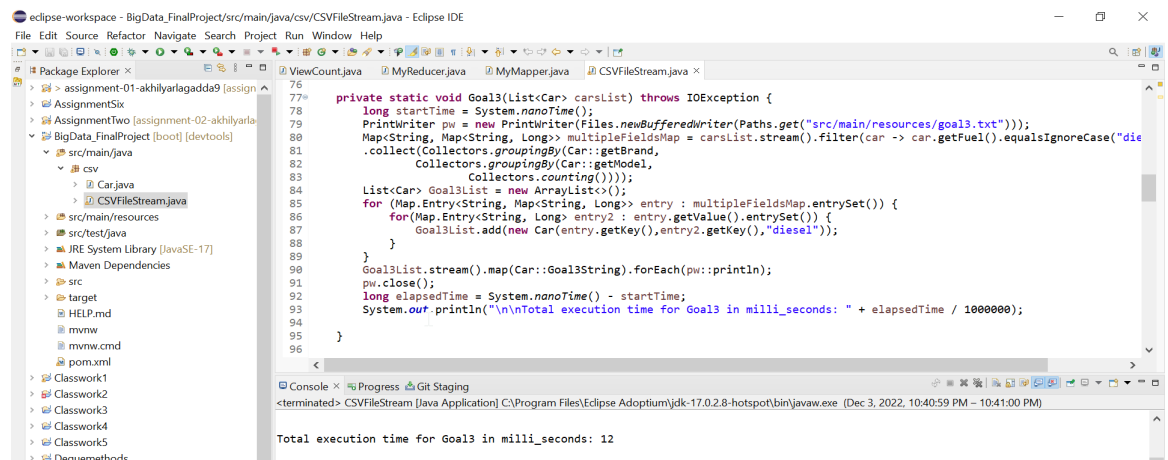
- To display the cars based on the specific brand like "Maruthi" and specific model "Alto".
- No difficulty find while we executing this goal.

## 7.2 What are the car Brands Manufactured between 2011 to 2015?



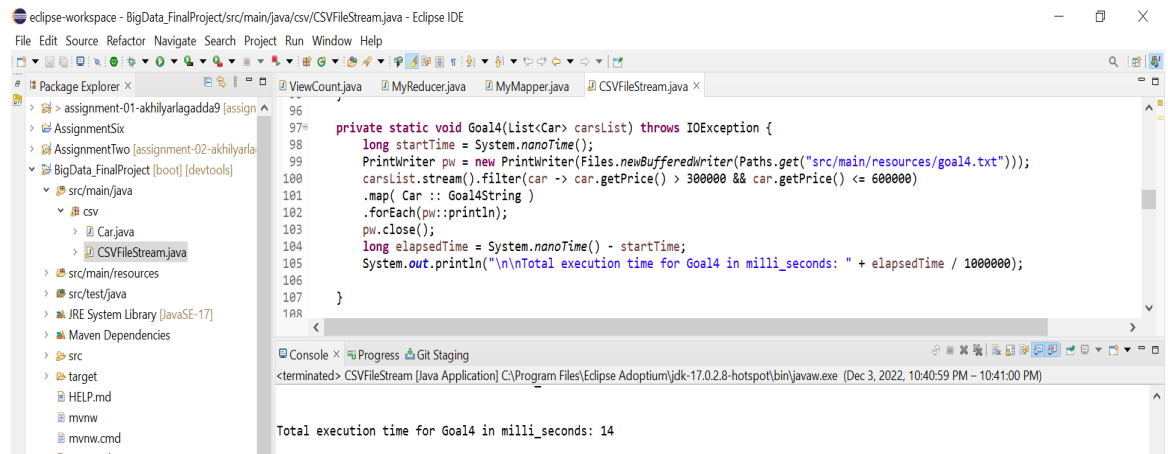
- In this goal we are displaying the cars that are manufactured from the year 2011 to 2015.
- No difficulty find while we executing this goal.

## 7.3 What are the brands which run on diesel fuel?



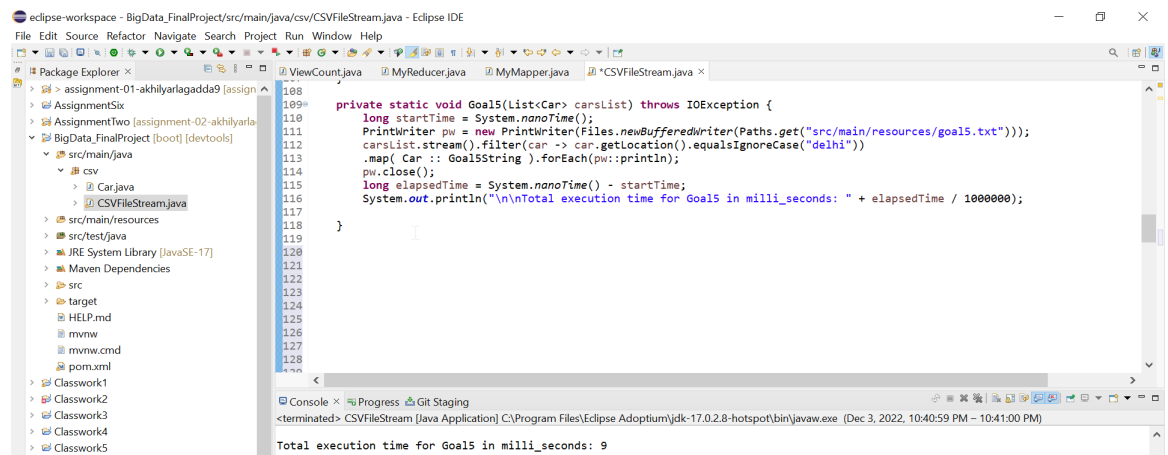
- Here we are presenting the cars that runs on using the fuel type either it is using diesel or fuel.
- No difficulty find while we executing this goal.

## 7.4 Specify all the Car Brands, Models, years, and price between 3L to 6L?



- To specify the cars that are ranging from the price 3 lakhs to 6 lakhs depends on all brands of cars.
- No difficulty find while we executing this goal.

## 7.5 List Brand with Model and year which are in the location Delhi?



- Here we are displaying all the cars that are in the Delhi location with all the details like Brand of the car with the model and the year.
- No difficulty find while we executing this goal.

## 7.6 List out all the brand, model and count with BMW?

```

131
132
133 private static void Goal6(List<Car> carsList) throws IOException {
134     long startTime = System.nanoTime();
135     Map<String, Map<String, Long>> multipleFieldsMap = carsList.stream().filter(car -> (car.getBrand().equalsIgnoreCase("BMW")))
136     .collect(Collectors.groupingBy(Car::getBrand,
137     Collectors.groupingBy(Car::getModel,
138     Collectors.counting())));
139     List<Car> Goal6List = new ArrayList<>();
140     PrintWriter pw = new PrintWriter(Files.newBufferedWriter(Paths.get("src/main/resources/goal6.txt")));
141     for (Map.Entry<String, Map<String, Long>> entry : multipleFieldsMap.entrySet()) {
142         for (Map.Entry<String, Long> entry2 : entry.getValue().entrySet()) {
143             Goal6List.add(new Car(entry.getKey(), entry2.getKey(), entry2.getValue()));
144         }
145     }
146     Goal6List.stream().map(Car::Goal6String).forEach(pw::println);
147     pw.close();
148     long elapsedTime = System.nanoTime() - startTime;
149     System.out.println("\n\nTotal execution time for Goal6 in milli_seconds: " + elapsedTime / 1000000);
150 }
151
152

```

Console Output:

```

<terminated> CSVFileStream [Java Application] C:\Program Files\Eclipse Adoptium\jdk-17.0.2.8-hotspot\bin\javaw.exe (Dec 3, 2022, 10:40:59 PM - 10:41:00 PM)
Total execution time for Goal6 in milli_seconds: 6

```

- Calculating the number of manufactured cars for each model of the BMW cars.
- No difficulty find while we executing this goal.

## 7.7 List SUV body type Brand and Models?

```

150
151
152
153 private static void Goal7(List<Car> carsList) throws IOException {
154     long startTime = System.nanoTime();
155     PrintWriter pw = new PrintWriter(Files.newBufferedWriter(Paths.get("src/main/resources/goal7.txt")));
156     Map<String, Map<String, Long>> multipleFieldsMap = carsList.stream().filter(car -> car.getBodyType().equalsIgnoreCase("SUV"))
157     .collect(Collectors.groupingBy(Car::getBrand,
158     Collectors.groupingBy(Car::getModel,
159     Collectors.counting())));
160     List<Car> Goal7List = new ArrayList<>();
161     for (Map.Entry<String, Map<String, Long>> entry : multipleFieldsMap.entrySet()) {
162         for (Map.Entry<String, Long> entry2 : entry.getValue().entrySet()) {
163             Goal7List.add(new Car(entry.getKey(), entry2.getKey(), "SUV"));
164         }
165     }
166     Goal7List.stream().map(Car::Goal7String).forEach(pw::println);
167     pw.close();
168     long elapsedTime = System.nanoTime() - startTime;
169     System.out.println("\n\nTotal execution time for Goal7 in milli_seconds: " + elapsedTime / 1000000);
170 }
171
172

```

Console Output:

```

<terminated> CSVFileStream [Java Application] C:\Program Files\Eclipse Adoptium\jdk-17.0.2.8-hotspot\bin\javaw.exe (Dec 3, 2022, 10:40:59 PM - 10:41:00 PM)
Total execution time for Goal7 in milli_seconds: 5

```

- Providing the data of the Brand and Models if the SUV body typed cars.
- No difficulty find while we executing this goal.

## 7.8 List the all first owner car brand with Models?

```

170
171
172
173
174 private static void Goal8(List<Car> carList) throws IOException {
175     long startTime = System.nanoTime();
176     PrintWriter pw = new PrintWriter(Files.newBufferedWriter(Paths.get("src/main/resources/goal8.txt")));
177     carList.stream().filter(car -> car.getOwnerType().equalsIgnoreCase("first owner")).forEach(pw::println);
178     pw.close();
179     long elapsedTime = System.nanoTime() - startTime;
180     System.out.println("\n\nTotal execution time for Goal8 in milli_seconds: " + elapsedTime / 1000000);
181 }
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196

```

Console Output: Total execution time for Goal8 in milli\_seconds: 52

- Providing the data of the First owner who used particular brand and model of the cars.
- No difficulty find while we executing this goal.

## 7.9 List the car model, Brand, Year, and Km that are driven less than 10000 kilometers?

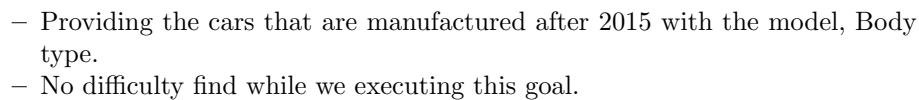
```

197
198
199 private static void Goal9(List<Car> carList) throws IOException {
200     long startTime = System.nanoTime();
201     PrintWriter pw = new PrintWriter(Files.newBufferedWriter(Paths.get("src/main/resources/goal9.txt")));
202     carList.stream().filter(car -> car.getDrivenKilometers() <= 5000).
203     .forEach(pw::println);
204     pw.close();
205     long elapsedTime = System.nanoTime() - startTime;
206     System.out.println("\n\nTotal execution time for Goal9 in milli_seconds: " + elapsedTime / 1000000);
207 }
208
209
210
211
212
213
214
215
216
217
218
219
220

```

Console Output: Total execution time for Goal9 in milli\_seconds: 4

- To display the cars that are driven 10000 kilometers with specific Car model, Brand and the Year.
- No difficulty find while we executing this goal.



To Finalize, the data we have taken is to show the numbers of the cars in the market are being sold based on the Make Model of the cars by its body type and the location of the buyer and seller with the price of the car based on the market price.

## GITHUB LINK