**Report On a Used Car’s Business**

By Paresh Patil

Date:- 18-03-2022

**Table of content**

1. Abstract
2. Introduction
3. Analysis Report
4. References
5. Appendix

**Abstract**

This report is on the used cars to help an investment firm to make a profitable business decision. Thus, the firm may not have significant financial effects or irreparable damage to its brand and reputation. There is some analysis in the report on the used cars using the cars datasets, on which categories of the car the firm could invest. Which categories are better for the business? This analysis has been performed in HIVE, it’s a platform for writing SQL-style scripts that perform MapReduce processes. The result of this analysis will help suggest to users which brands they should invest in for used cars and which cars the company should buy according to price range.

**Introduction**

The analysis of the used cars is based on the car’s dataset, which is having specific details like Makers, Models, Mileage, Colours, Price, etc. This includes implementing several logical questions in the form of SQL queries to get useful results. The need of this analysis is to show organizations some good sections in used cars where they can invest. In the report, an investigation is performed on cars data, actions like cleaning the data mean removing the invalid data, finding the lowest price cars and the brand and vice-versa, best in mileage, good economy and bad economy, Luxuries cars and affordable cars of intermediate or middle class. Performing several questions on the data to get answers in terms of suggestions to make a good business decision. The platform used for this analysis is HIVE which is a data warehouse software, in which logical queries can be implemented to get the end results of our questions. This analysis is possible to implement in regular tool excel but the amount of data in terms of rows is more than 15 lakhs and it’s merely possible to handle this huge amount of data in excel whereas Hive is invented to handle the big data.

**Analysis Report**

The used cars are divided into some categories in the analysis they are Economic segment, the intermediate segment, and the luxuries segment.

**Economic segment**:- The term "economy car" refers to vehicles that are designed to be inexpensive to buy and operate. Small lightweight, and affordable to build and purchase, economy cars are the most common. In this segment, analysis is showing the 5 most fuel-efficient used cars are

* Volvo v40
* Peugeot rcz
* Skoda 130
* Lexus gs
* Volkswagen beetle

These are some most fuel-efficient cars for economic segment customers with a price range of 3000$ to 20,000$.

**Intermediate segment**:- The term “intermediate segment” means the customer who purchases these cars for resale. It will either be consumed by the ultimate customer or passed on to another person to consume.

The 5 used cars for the intermediate segment are

* bmw z8
* Tesla roadster
* Tesla model-x
* Bentley brooklands
* Rolls-royce wraith

These cars are having a range between 20,000$ to 3,00,000$ based on top average price.

**Luxury Segment**:- A luxury Segment Customer is a customer that looks for higher levels of comfort, equipment, amenities, quality, performance, and status than conventional cars and doesn’t care about higher prices.

* Lamborghini aventador
* Porsche Carrera-gt

These are the most comfortable and luxurious cars with a price range of 300,000$ to 2000,000$.

Furthermore, focusing on the customer who wants the cheapest used cars only and doesn’t investigate features.

The cheapest used cars makers and models are listed below:

* Skoda galaxy
* Rover streetwise
* Kia retona
* BMW transit
* Chevrolet alero
* Hyundai santamo
* Opel kadett
* Fiat 128
* Nissan frontier
* Seat inca

These are top 10 cheapest used cars with the price range of 3000$ to 4000$.

The customer who wants the costliest used car can look into the luxury segment. Moreover, focusing on the customer who needs **high mileage** used cars. The top five average high mileage used cars are

* lexus rx-330
* lexus 400
* rover 200
* subaru tribeca
* volvo 460

average mileage range between 800,000 to 4000000.

Some customers want high engine power used cars, Here firms can buy these cars for this high engine power segment.

* jeep scrambler
* lamborghini aventador
* rolls-royce wraith
* dodge viper
* bentley continental-flying-spur

The engine power is between 400 HP to 900 HP.

If the customer wants less car age and lowest price then these are top five use cars.

* audi a3
* volkswagen polo
* 0volvo v70 6
* toyota yaris
* skoda superb

These cars are used 6 years and have price lowest than 4000$.

To put it into a nutshell these are some used cars and categories in which firm can think to invest and enhance the business.

**References**

* *Analytical report outline: Formal analytical report outline*. iWriteEssays. (n.d.). Retrieved March 14, 2022, from [https://www.iwriteessays.com/outline/analytical-report-outline](https://www.iwriteessays.com/outline/analytical-report-outline%20)
* Zoricak, M. (2017, March 16). *Classified ads for Cars*. Kaggle. Retrieved March 14, 2022, from [https://www.kaggle.com/mirosval/personal-cars-classifieds](https://www.kaggle.com/mirosval/personal-cars-classifieds%20)
* *Apa @ conestoga: Appendix / appendices*. Library Services. (n.d.). Retrieved March 14, 2022, from [https://lib.conestogac.on.ca/visuals-appendices/appendix-appendices](https://lib.conestogac.on.ca/visuals-appendices/appendix-appendices%20)
* Hive tutorial. (n.d.). Retrieved March 14, 2022, from <https://www.tutorialspoint.com/hive/index.html>

**Appendix**

* **Hive** is a Hadoop data warehouse infrastructure solution that allows you to process structured data. It sits atop Hadoop to summarise Big Data and facilitate searching and analysis.
* The **Hive Query Language** (HiveQL) is a query language for processing and analysing structured data in Apache Hive. It shields users from Map Reduce programming's complexities. To make learning easier, it borrows notions from relational databases, such as tables, rows, columns, and schema. Hive provides a command-line interface for writing Hive queries using the Hive Query Language.
* A **data set** is a collection or set of data. This information is usually presented in a tabular format. Each column denotes a distinct variable.
* **Kaggle** is a data science and machine learning community on the internet. Users can use Kaggle to search and publish data sets, explore, and construct models in a web-based data science environment, collaborate with other data scientists and machine learning experts, and compete in data science challenges.
* **SQL** is a programming language that is used to access and manipulate databases.
* A **data warehouse** is a sort of data management system that is intended to facilitate and assist business intelligence and analytics activities.