Case Study of Fuel Consumption by Vehicles Utilizing the Postulates of Bounded Rationality

Joseph E. Mullat *

Abstract

The article presents a new method called "Blind Data Analysis", which simplifies the examination of numerous statistical indicators with an unknown distribution. It highlights the practical context as well as the need for a rational categorization to improve the reliability of forecasts. It seems that this method makes it more accessible to consider indicators, following the principle of parsimony, similar to Ockham's Razor. It explores and discusses the basic postulates of bounded rationality in decision-making process. Arguments based on postulates include a direct justification for the reliability of the method based on the additional postulate of a "constraint" of reality; this seems to be a more dynamic method. The study uses this method with data from the Spritmonitor.de database, which focuses on vehicle information, including gas and electricity consumption, and vehicle mileage. The database helps users track their fuel savings and related costs by providing thousands of car users with real-world cost per gallon or liter per 100 km. The postulates of rationality were tested against this database using the Excel macro program.
Keywords: data analysis; decision-making; vehicle; fuel consumption; monotonic system

Concise Glossary of Mathematical Notations

We consider fuel consumption indicators $p_k \in A$, $|A| = n$ of n car models or labels/issues, $k = \overline{1,n}$. Indicators $\langle \overline{p} \rangle = \overline{p}_1 \geq \overline{p}_2 \geq \ldots \geq \overline{p}_j \geq \ldots \geq \overline{p}_n$, $j = \overline{1,n}$, in contrast to the original list p_k, are necessarily descending. A sequence $\overline{\pi} = \langle \pi_j \rangle = \pi_1, \pi_2, \ldots, \pi_j, \ldots, \pi_n$ arrange so-called moments, where $\pi_j = \overline{p}_j \times j$. The choice consists of an act of selecting several car models from X, Y or $H \subseteq A$ (as described by Ma et al., 2015) or as multiple options $C(X) \subseteq X$, $C(Y) \subseteq Y$ or $C(H) \subseteq H$ according to certain rules also outlined by Strzalecki (2011). A totality of lists of all 2^n samples or issues $H \subseteq A$ is denoted by $2^A = \{H\}$. In accord with the descending sequence $\langle \overline{p} \rangle$, samples X, Y can be reordered into segments $X = [x\_1 > x\_r]$, $Y = [y\_1 > y\_r]$. Thus, segments X, Y correspond to choices $C(X) = [c(x)\_1 > c(x)\_r]$ and $C(Y) = [c(y)\_1 > c(y)\_r]$, etc. In this case, we refer to the segmentation operators $C(X)$ and $C(Y)$ as a choice made from reordered segments $S(A)$.

* mjoosep@gmail.com, Residence: Nygårdsvej 10, 2 sal, Nr.13, 2100 Østerbro, Denmark.
Dr. Joseph E. Mullat holed a docent position at the Faculty of Economics at Tallinn University of Technology

1. Introduction

Usually, a theoretical or practical contribution to the theory and reality consists in expanding existing categories, concepts, models, simplifications, etc. towards obtaining new theoretical facts or solving unsolved problems. However, there is another approach to extracting new knowledge from old and well-established categories, which consists in recognizing new relationships or links hidden between old fundamental categories. This innovation taking two things that already exist and putting them together in a new way is the main motive of this article – a comparison, or rather, an interpretation the data analysis in terms of decision-making process. The relationship between data analysis and decision-making is nuanced, particularly in the contexts of automotive market where precise assessment of issues, such as fuel consumption, may pose challenges. While it may seem apparent, rigorous scientific validation ensures that connections between data analysis and decision-making align with reality, preventing assumptions from leading astray.

Car models are commonly classified into categories based on their fuel consumption or economy, which helps consumers make informed choices about efficiency and environmental impact. These classes often include designations like "compact," "midsize," or "luxury," each reflecting different overall performance and levels of fuel efficiency. We are analyzing a data set of 2927 car models from various manufacturers, focusing on fuel consumption, measured in Miles Per Gallon or Liters per 100 km for traditional cars and, additionally, in kWh/100 km for electric and hybrid cars.

Data on well-established car models tend to provide relatively accurate forecasts, although, many factors, which cannot be easily accounted for by even the most popular models, can affect the car market, such as lifestyle, technological advances and rising fuel prices. Publications of Society of Automotive Engineers (SAE) journals, the International Journal of Automotive Technology, and articles from authoritative sources of the International Council on Clean Transportation (ICCT) usually address topics related to vehicle fuel efficiency and vehicle labeling and data analysis procedures. Thus, a simple data analysis procedure, although not ideal, may be preferable to a very complex probabilistic Case Study of fuel characteristics.

Data analysis and categorization techniques play a crucial role in various fields, from machine learning to information organization. In the realm of machine learning, algorithms like decision trees, support vector machines, and neural networks are commonly used for classification tasks. These techniques aim to assign predefined labels or categories to input data based on patterns and features.

In data analysis and classification, two opposing approaches can be distinguished based on their direction, as one starts with subjective knowledge to arrive at objective knowledge, while the other does the opposite. In the first approach, specialists in a relevant field, such as physicians, biologists, astronomers, market practitioners, and data analysts, apply categorizations techniques to objective statements about the obtained estimates of experimental data, observations, etc. Accordingly, they may employ artificial intelligence" (Mitchell, 1997; Seiffarth et al., 2021), statistical learning" (Hastie et al., 2009), pattern recognition" (Bishop, 2016), machine learning" (Domingos, 2010, which provides insights into the quest for a universal learner), or any other suitable probabilistic method (Murphy, 2012; Koller and Friedman, 2009) in their data analysis, which would also include data mining techniques (Nan and Kamber, 2011, their work covers various aspects of data mining), parameter estimation (Walter and Pronzato, 1997), management and decision-making problems (Narula and Weistroffer, 1989), continuous modeling, multiple time series (Voelkle et al., 2012) and computational methods (Mirkin et al., 1995). While these approaches necessitate the knowledge of the distribution of judgments about the object under analysis, in practice this is not always the case.

The purpose of subjective assessment methods (Frey, Võhandu, 1966) seems to be their application to multi-dimensional indicators division into two classes, which, at first glance, may turn out to be contradictory. It may also be that specialized knowledge is required, although this, as already stated, is not always the case since knowledge of the distribution of numerical parameters, also known as indicators, may be redundant.

These arguments will be demonstrated by considering the newly proposed objective-to-subjective data analysis called "Blind Data Analysis", referred as ***BDA***, which solely considers whether one number is "less/greater" than another. If common sense is achieved, then the well-known law or "Ockham's Razor" principle of parsimony, which states that simpler theories are better than theories that are more complex, will come into force. Accordingly, as such, a procedure that requires fewer assumptions about reality can be considered the most reliable.

2. Bounded rationality postulates

Rational choice theory is a framework that attempts to explain how individuals make decisions based on their preferences and constraints (e.g., Arrow, 1948; Jamison, 1973). Several of its postulates fall under the umbrella of so-called bounded rationality, including the assumption that individuals have well-defined preferences, make choices based on their expected utility, and make rational decisions based on the available information. If the potential car buyers apply these assumptions to the dynamics of vehicle evaluations, the introduction of an additional postulate or constraint of monotonicity is required to account for changing consumer preferences and assessments.

The postulate of monotonicity suggests that as the list of models for a proposed purchase or sale of cars is narrowed down, consumers' assessments or subjective utilities (referred to here as moments) will consistently and monotonously decrease. In other words, as the options become more limited, individual preferences or the perceived value of the remaining models will decrease in a predictable and continuous manner. The constraint refers to the spontaneous or impulsive judgments and evaluations potential buyers make during the car selection process. The argument suggests that as buyers eliminate certain car models, their impulsiveness decreases, indicating a shift towards more thoughtful and considered evaluations as the selection process progresses.

This perspective could be relevant for various decision-making and data analysiss scenarios, including the evaluation and comparison of different products before making a purchase. It is based on the premise, as decision-makers eliminate alternatives; adjust their preferences, classifications and pairwise comparisons and evaluations that will consistently decline rather than fluctuating or exhibiting non-monotonic behavior.

However, it is important to note that the concept of monotonic constraint may not apply to all decision-making processes, as individual preferences and subjective assessments can vary significantly. As different people may have highly divergent valuation criteria or impulses, their evaluations might not always exhibit a monotonic decrease with a narrowing range of options. Therefore, while constraint can provide a useful framework for understanding certain decision dynamics, they should be applied with caution and should be considered in conjunction with other factors that may influence individual choices. For this purpose, in this work, Arrow's (1959) strict consistency postulate is slightly modified to ensure the validity of the basic postulates of rational choice. This modification to the standard rational choice framework accounts for the dynamics of the automotive market and the behavior of car buyers. Specifically, the act of choice consists of selecting several issues X (as described by Ma et al., 2015) or as multiple options $\mathrm{C(X)}$ according to certain rules also outlined by Strzaletski (2011).

Let us recall in a Boolean—that is, in a more formal—form the bounded rationality canonical postulates (cited by Aizerman and Malishevski, 1981, pp. 65-83, English version translated from Russian, p. 189). Here, they are presented in connection with rational choice in the automotive market that involves factors such as fuel efficiency, cost, environmental impact, and personal needs. Evaluating these elements helps customers make an informed decision that aligns with their preferences and priorities, as outlined below.

- *Independence with respect to removing rejected alternatives (or, for brevity, elimination of options),*
  Postulate 5 (Chernoff, 1954*, pp. 422−443) or Axiom 2 (Jamison and Lau,* 1973*, pp. 901−912):*
  From $\mathrm{C(Y)} \subset \mathrm{X} \subset \mathrm{Y}$ *it follows that* $\mathrm{C(X)} = \mathrm{C(Y)}$;
- *Compatibility, the same as Postulate 10 of Chernoff and property* γ *of Sen:*
  From $\mathrm{X} \cup \mathrm{Y}$ *it follows that* $\mathrm{C(X)} \cap \mathrm{C(Y)} \subset \mathrm{C(X \cup Y)}$
- *Non-strict Consistency, which is the same as Postulate 4 (Chernoff), or property* α *(Sen,* 1971*,*
  pp. 307−317) or the axiom C2 of Arrow-Uzawa (Arrow, 1959*, pp. 121−127):*
  From $\mathrm{X} \subset \mathrm{Y}$ *it follows that* $\mathrm{X} \setminus \mathrm{C(X)} \subset \mathrm{Y} \setminus \mathrm{C(Y)}$ *or equivalent to*
  $\mathrm{X} \cap \mathrm{C(Y)} \subset \mathrm{C(X)}$;
- *Strict Consistency or constant residual choice, which is the same as Postulate 6 (Chernoff,* 1954*) and one of*
  the forms of the "weak axiom of revealed preference" of Samuelson, i.e., the axiom C4 (Arrow, 1959*, pp. 121−127):*
  From $\mathrm{X} \subset \mathrm{Y}$ *and* $\mathrm{X} \cap \mathrm{C(Y)} \neq \varnothing$ *it follows that* $\mathrm{X} \cap \mathrm{C(Y)} = \mathrm{C(X)}$.

The strict consistency was confirmed in experiments with the correlation matrix, see the Appendix. In fact, preferences for pairwise comparisons of indicators in the set X of narrowed list $X \subset Y$ remain in the same "less/greater" relation $x \leq y$ or $y \leq x$ as in Y. However, pairwise indicator preferences may shift disproportionately based on narrowed list X, potentially causing unselected indicators in Y to become irrationally selected in X, akin to our Pedagogical Scenario. This shortcoming of the strict consistency postulate C4 can easily be corrected to ensure the validity of the rational choice postulate C4 remains unchanged, i.e.:

From $X \subset Y$ and $X \cap C(Y) \neq \varnothing$ it follows that $X \cap C(Y) = C(X) \cap C(Y)$.

and classification environment what is valuable for reliability and In this slightly changed form, the postulate still operates in the same way as the canonical Arrow's strict consistency postulate, even though the founders of rational choice theory (Simon, 1978) did not consider such dynamics (including Arrow in 1948 and 1959; Chernoff in 1954, and Sen in 1971). These postulates ensure consistent outcomes with repeated decisions in the same categorization predictability in the decision-making process. Namely, the relations in the part of the objects remain in exactly the same state as the complete visual representation of the objects. We are referring to the idea of self-similarity in the Fibonacci principle, where the characteristics of a part reflect the characteristics of the whole. This concept is often observed in various natural patterns and structures.

3. PARSIMONIOUS APPROACHES

To validate the postulates of rational choice in the context of fuel consumption classes, among various car manufacturers, we can examine how consumers make decisions when selecting a car based on fuel efficiency and analyze whether their choices align with the assumptions of rational choice theory.

There are several statistical methods that could be applied to test rational choice postulates, including econometric models to estimate the parameters of a utility function that describes how car owners make decisions and machine-learning algorithms to identify patterns in the data and test whether they are consistent with rational choice theory. At this juncture, it should be noted that the following discussion of the exhibition scenario is only an introduction to the main topic about the results and experiments carried out using the Excel spreadsheet of information and interactive computer services, which are provided on https://www.spritmonitor.de/en/. The spreadsheet was subjected to a validity test of the independence postulates of the rejected alternative and the postulates of non-strict and strict consistency with the established car models in the market. The obtained findings indicate that car buyers may prioritize factors such as engine power or fuel efficiency, which may be rational when postulating rejected alternatives. The presented postulates of consistency according to Ockham's Razor procedure seem to confirm the experimental results. Nevertheless, a rigorous proof of these assertions is beyond the scope of this article and will be left for further research. On the other hand, the proof of independence from the rejected alternatives follows from Proposition I given in Section 3.3. Proposition I has a long history since publication in the Proceedings of Tallinn Polytechnic Institute (Mullat, 1971, pp. 37-44).

3.1. Pedagogical Scenario

The postulates of consistency and compatibility are specific theoretical foundations used in economics to analyze decision-making behavior. These postulates suggest that individuals make decisions based on issues of consistent preferences that do not change over time. With regards to the postulate of strict consistency C4, it is useful to paraphrase Arrow's intuitive interpretation to consider the following. In terms close to the data analysis scenario, this suggests that if some car models are labeled in the context of fuel consumption from the range of models available for sale, then narrowing the range of labels should not change the status of previously labeled or unlabeled to selected or unselected to designate models cars. While this is part of the rationality criteria that Arrow explored in his work on social choice theory and the impossibility of creating a perfect voting system, it does assume some stability in the labeling system, which is often useful when implementing data analytics tools.

Let's start the scenario with a "hypothetical" or "pedagogical exhibition" based on the choice of a car at an exhibition when analyzing decision-making classification phenomena. Suppose further that, after accepting an offer to purchase a car, the salesman tells the customer that some of his preferred options are not available, potentially causing irrational behavior on the part of the customer or the salesperson. From a customer's point of view, it might be wiser to try fuel-efficient cars that were initially overlooked. On the other hand, the seller may offer more stylish and luxurious cars, even though there are economical and equally good options. If a customer preferred fuel-efficient vehicles but learned they were out of stock, they would likely include more fuel-efficient models to expand the list of alternatives that were initially overlooked. However, if the buyer wanted to buy an economical car and his choice was limited for some reason, buyers, contrary to their original intention, may welcome the seller's alternative offering more stylish models.

3.2. Significance

Returning to the cars pedagogical scenario, the cars would be listed linearly in descending order based on fuel/electricity, where the highest fuel or electricity consumption per 100 km is multiplied by **1**, the next item in the list is multiplied by **2**, and so on. Here these figures are interpreted as *"fuel consumption credentials"* or "moments". The local maximum fuel consumption is selected when the moment's maximum is reached. Some details of the car selection procedure just outlined are also relevant for analyzing automobile market data.

Let us assume that the client decides to accept the car fuel consumption at the local moment maximum as an acceptable level of significance when choosing cars with a higher or equal fuel consumption level, e.g., the list $\mathbf{10^2}$, $\mathbf{9^2}$, $\mathbf{8^2}$, $\mathbf{7^2}$, $\mathbf{6^2}$, $\mathbf{5^2}$,… indicates that the peak of this sequence is located at $\mathbf{7^2 = 49}$. Define a list of fuel consumption indicators $p_k \in A$, $|A| = n$ of n car models, $k = \overline{1,n}$. In particular, suppose that in the sample denoted by the letter H, the prospective car buyer selects some potential cars as viable candidates according to the reasonable fuel consumption. We can further define a totality of lists $\{H\}$ of all 2^n samples or issues $H \subseteq A$. Accordingly, $\pi(p_k, H) = p_k \cdot |H|$ moments as monotone system (in terms of Mullat, 1971) , or as monotone linkage clustering (Kempner et al., 1997,) will evaluate so already called *"credentials"* of fuel consumption. The procedure for finding the significance level of fuel consumption commences with sorting all the fuel consumption indicators p_k, constituting (as in the price sticker list) the vehicle fuel/electricity indicators permutation $\langle \overline{p} \rangle = \overline{p}_1 \geq \overline{p}_2 \geq ... \geq \overline{p}_j \geq ... \geq \overline{p}_n$ in descending order. Next, a sequence $\overline{\pi} = \langle \pi_j \rangle = \overline{p}_1 \times 1, \overline{p}_2 \times 2, ... \overline{p}_j \times j ... \overline{p}_n \times n$, which components π_j we called moments, $j = \overline{1,n}$, is constructed. Hereby, the list of fuel consumption indicators $\langle \overline{p} \rangle$, in contrast to the original list p_k, is necessarily descending., We called such sequences $\overline{\pi}$ as defining (Mullat 1971).

3.3. Internal Personal Stability

Internal or intrinsic personal stability refers to an individual's ability to maintain a sense of balance, composure, and well being within themselves, irrespective of external circumstances or challenges. It involves emotional resilience, self-awareness, and a capacity to navigate life's ups and downs with a steady and grounded mindset.

When we talk about internal personal/intrinsic stability in terms of being "interpersonally incompatible" or "impossible to match through a monotone transformation," it means that the relative economic preferences or classifications of individuals cannot be reconciled using a simple, consistent scaling or transformation.

In the context of Narens and Luce's work from 1983, monotone transformation refers to a mathematical function that preserves the order of preferences but might change the classification scale. If two individuals have inherently incompatible preferences that cannot be aligned through such transformation, it implies that there is no single, uniform way to compare or match their classification preferences or evaluations. This concept underscores the complexity of ensuring stability in interpersonal classification framework, as certain inherent classes or categories in individual preferences may resist easy standardization or comparison.

Our experiments show that when categorizing vehicles according to specific utility functions or monotonic transformations based on fuel consumption using ***BDA***, the assumption of intrinsic personal stability is generally not satisfied. Simply put, this highlights the complexity of the process, as individual preferences and ratings, such as the rationale for predicting vehicles' prices based on fuel consumption, in some cases cannot be consistently compared and categorized, regardless of the specific scaling or units of measurement used.

3.4. The reasonable level

The moments $\langle \pi_k \rangle = \pi_1, \pi_2, ... \pi_j ... \pi_n$ are single peaked, where the peak denotes the kernel issues H^* (Mullat, 1971−1995) of a monotone system. The list H^* constitutes the rational, i.e., the monotone linkage clustering implemented in our findings. At the location k^* from the top of the moments $\overline{\pi} = \langle \pi_k \rangle$, i.e., from the top of the defining sequence of models, $j = \overline{1,n}$, where the local maximum $u = \max_{j=\overline{1,n}} \pi_j$ is reached, the peak, denoted by u, will be called the level of significance.

Proposition I*. Among the totality of all samples $H \subseteq A$, i.e., among all the lists $\{H\}$ of all 2^n samples, the kernel H^* guarantees reaching the global maximum of the moment function $F(H)$ of samples H; $F(H) = u$ is equal to $u = \min_{p_k \in H} \pi(p_k, H)$: $H^* = \arg\max_{H \subseteq A} F(H)$.*

Proposition I confirms the postulate of independence from rejected alternatives in two-person games, which was originally studied by John F. Nash in the 1950s, when he developed a solution to the bargaining problem. With regard to the market for the purchasing and production of cars, Proposition I states that any final decisions made or based on statistics should not be affected by the removal of any parts of those statistics that are not reliable or represent a very small number of cases in which, for example, statistics have been collected into a database and selected for review.

A transformation using some monotonic function $tr(x)$ of indicators $\overline{p}_j \geq \overline{p}_{j+1}$, $j = \overline{1, n-1}$, preserving the validity of $tr(\overline{p}_j) \geq tr(\overline{p}_{j+1})$, can still shift the original H^*-kernel to $H^* \neq tr(H^*)$, what happens, for example, when transforming by $tr(x) = x^2$. However, the H^* kernel, which takes into account fuel consumption under the personal/intrinsic guidance of Narens and Luce, remains unbiased for the proportional mapping $tr(x) = \alpha \cdot x$, for example, in the case of converting liters to gallons.

3.5. Threshold-based Time-Series indicators

Time-Series data typically refers to data that is collected or updated periodically or at regular intervals over time. Series are commonly used in data mining and other areas where the objective is to detect outliers or changes in system behavior. This could include information such as sales figures, stock prices, weather data, or any other type of data that is recorded over a specific time period. The frequency of data collection can vary depending on the needs and requirements of the specific use case or analysis.

A threshold value can be used in the fuel consumption series to determine the reasonable significance of cases where consumption level exceeds a certain positive threshold or falls below a certain negative threshold (an illustration of this scenarios are presented in the Appendix). In the specific case considered, the indicators called fuel moments create a dynamic system since the previous state of the consumtion determines its subsequent state. It is worth noting that the postulates of strict and non-strict consistency emphasize the rational behavior of car buyers when new models expand the list of available alternatives. In the event that prospective car buyers have chosen some of the best cars in the past, these postulates state that they will still be inclined to consider old models, in accordance with the "old love does not rust" adage.

Accordingly, if we look at our methodology for determining the significance level u of car fuel consumption indicators p_j, one may get the impression that the procedure is applicable only to positive numbers. Generally speaking, the same procedure is obviously valid for a negative series of numbers. In this sense, the procedure can be used by analogy with what is called a "confidence interval" in statistics. Indeed, if we apply the procedure to a positive series of numbers, then as a result we will obtain a level of significance u in the form of a positive number. Now, based on the found level u of significance, we can create a sequence of deviations around this level, both $\pm\Delta$ towards less and towards excess. Now it will be possible to apply the procedure again, but this time in relation to deviations from the original significance level u. As a result, the "confidence interval" $[-\Delta_1 + u, u + \Delta_2]$ will be determined. By considering this interval and observing whether the dynamic indicators cross the threshold u, one can make significant decisions. Indeed, customers interested in economy cars would likely make a purchase if the indicators consistently cross below $-\Delta_1 + u$, while this will be unlikely if the dynamic indicators intersect above the $u + \Delta_2$ level. Looking at this interval and observing whether the dynamic indicators cross the threshold value u customers can make a decision on how to cluster the relevant data. Since market actual monitoring for the automobiles data includes 2927 vehicle models, a fragmented version of it is presented in Table 1 (screenshot from an Excel spreadsheet).

Negative significance level →				-0,61
Positive significance level →			6,97	3,19
	Count	Fuel type	l/100km	
Alfa Romeo	2053	Gasoline	9,28	2,31
Aston Martin	24	Gasoline	13,22	6,25
Bentley	12	Gasoline	15,56	8,59
BMW	29508	Gasoline	8,86	1,89
Bugatti	2	Gasoline	12,38	5,41
Chevrolet	1677	Gasoline	9,76	2,79
Cadillac	135	Gasoline	13,66	6,69
Chrysler	811	Gasoline	10,84	3,87
Daewoo	366	Gasoline	7,53	0,56
Citroen	5793	Gasoline	7,14	0,17
Daihatsu	1200	Gasoline	5,86	-1,11
Datsun	4	Gasoline	10,89	3,92
Ford	20799	Gasoline	7,99	1,02

Negative significance level →				-5,21
Positive significance level →			15,39	3,87
	Count	Fuel type	kWh/100km	
BMW	315	Electricity	16,55	1,16
Bugatti	1	Electricity	10,18	-5,21
Citroen	72	Electricity	15,94	0,55
Ford	25	Electricity	21,65	6,26
Ferrari	2	Electricity	41,45	26,06
Fiat	131	Electricity	17,11	1,72
Honda	13	Electricity	19,65	4,26
Hyundai	532	Electricity	15,93	0,54
Jaguar	5	Electricity	20,90	5,51
Kia	249	Electricity	17,34	1,95
Mazda	40	Electricity	19,43	4,04
Mercedes-Benz	90	Electricity	24,26	8,87
Mitsubishi	31	Electricity	14,23	-1,16

Table 1. Screen dump from Excel spreadsheet:

$\geq u=+6.97$, $\leq \Delta_1=-61$, $\geq \Delta_2=+3.19$ **$\geq u=+15.39$, $\geq \Delta_2=+3,87$, $\leq \Delta_1=-5.21$,**

4. AUTOMOTIVE MARKET DATA

To demonstrate the effectiveness of the proposed approach, the standard mechanisms and techniques of the MS Windows platform were utilized to view the database related to thousands of car models in Excel spreadsheets. The list includes cars that are not only economical and reasonably inexpensive but also even expensive stylish or luxury and cars of all available models. This information has been extracted and recompiled from the interactive computer services provided on the Spritmonitor.de website and includes vehicle fuel data, significant volumes and other relevant variables.

Some comments are needed to clarify the implementation of our Ockham's Razor "procedure" for analyzing the car fuel consumption dynamics. Specifically, it should be noted that the reliability of data on the lease or purchase cars with regard to fuel consumption, where all fuel consumption data have been available to everyone, is given by the fact that the MPG (mileage or mile per gallon) data are guaranteed by the Cost Calculator and Tracker at the date-to-date basic activity at the Spritmonitor.de database. The spreadsheet https://www.spritmonitor.de/en/search.html was compiled using domain (Accessed July 10, 2023).

An overview of common internal combustion fuels or hybrid/pure electric vehicles is available (*http://www.datalaundering.com/download/MPG-MileAge-Data.xls,* August 22, 2023) in the database used in the experiment is presented here solely for the purpose of illustrating the data collected so that the article is well suited to the layperson of interest. Clearly, each fuel type has advantages and disadvantages in terms of efficiency, emissions, availability and infrastructure. Analyzing fuel consumption across these categories can provide valuable insights into the efficiency and environmental impact of different car models. As seen from the tabulated columns containing fuel consumption, blue and yellow cells differ from others in certain patterns and frames, identified using the macro—Ctrl+s. In accordance with Proposition I given earlier, an analysis of the significance levels of the negative/positive (yellow/blue) values of the car indicator dynamics has been conducted. Using the macro in columns, (selected or ″pasted″) areas X of spreadsheet in their entirety may consist of negative/positive numbers that are distributed without any special purchase. However, the standard EXCEL data sorting options allow the content of selected areas to be sorted in ascending or descending order depending on the specified columns or rows. Thus, relevant cells can be redistributed into ″contiguous areas″ of negative or positive values in the column or row patterns to satisfy the necessary conditions. Such contiguous areas can be used in experiments featuring the Case Study results. The $C(X)$ operator was compiled into the Ctrl+s macro, using automotive market share fuels X as the initial data table below in column format of alternative X.

5. Ockham's Razor Procedure Guide and Properties

The novel procedure proposed here, as previously noted, was called ***BDA***, as it involves finding the simplest explanation or a most parsimonious models that fit the data based on the premise that simpler explanations are more likely to be true than complex explanations. It is important to note that the ***BDA*** procedure is not necessarily equivalent to other known statistical hypothesis testing, such as the null hypothesis (often denoted H_0), which is the statement that there is no significant difference or effect. Researchers seek to test this hypothesis against the alternative hypothesis (H_1), which, in contrast, suggests that there is a significant difference or effect. The goal is to determine, based on statistical analysis of the data, whether there is sufficient evidence to reject the null hypothesis in favor of the alternative. However, it is important to remember that our ***BDA*** procedure is just one tool in a broader set of statistical methods and may not be suitable for every situation.

5.1. Arranging indicators in Excel

In the Excel spreadsheet http://www.datalaundering.com/download/MPG-MileAge-Data.xls you can find the ***BDA*** Ctrl+s macro. Those wishing to use it can copy the base spreadsheet code into their own spreadsheet. In the properties of this macro, you must also indicate that the macro can be executed using the Ctrl+s command. However, when tabulating information, the first two rows of the table must be blank, so users must insert at least two blank rows at the top of the table.

You can also arrange the indicators p yourself. For a column of p_k indicators located somewhere on the spreadsheet, select a block of rows k, which point at the p_k indicators in the located column that you want to sort, $k=\overline{1,n}$. Go to the Data tab and use the Sort option in descending order based on the p_k values of the indicators. The p_k indicators will rearrange itself into a descending sequence of indicators $\langle\overline{p}\rangle=\overline{p}_1\geq\overline{p}_2\geq...\geq\overline{p}_n$. Then create an extra column for the moments $\pi_j=\overline{p}_j\times j$. In this case, those indicators p_j, the moments of which in π–cells exceed the significance threshold $u=\arg\max\pi_j$, should be painted in a different color in contrast to those cells that are smaller.

5.2. Validation of consistency postulates

From the information presented in the main part of the article, it is clear that we are talking about a ″moment indicator″, which was used to classify fuel consumption in the context of choosing the optimal car. The indicator was calculated as the product of the item number and the fuel consumption of the option in a descending linearly ordered list of indicators. It serves to measure the desirability of each option based on its fuel consumption. This measurement involves applying Ockham's Razor procedure to select the optimal option using fuel moment as a scalar criterion. From the perspective of Ockham's Razor, when choosing between competing options, simpler explanations or models are preferred.

Proof of Proposition 1.

To prove or verify the truth of this proposition, we can revisit the article. "On a Maximum Principle for Certain Functions of Set Functions", (Mullat, 1971). This is not necessary, however, since the proof in our particular case is much simpler thanks to the following lemma.

Lemma. *In any subset* $H \subseteq A$ *of indicators* A*, the order of moments* $\pi(p_j, H) = p_j \cdot |H|$ *corresponds to the grand order of moments* $\pi(p_j, A) = p_j \cdot |A|$*,* $|A| = n$ *on the set* A.

Simply put, the lemma states that if we take a pair of indicators $p_i \le p_j$, then no matter in which subset X or Y we consider this pair of indicators $\langle p_i, p_j \rangle$, the moment inequalities of $\pi(p_i, X) \le \pi(p_j, X)$ or $\pi_i(p_i, Y) \le \pi_j(p_j, Y)$ will remain in force.

Now we can begin to prove the Proposition 1. The proof will be carried out by contradiction. So, by construction, the set of indicators p_j, $j = \overline{1, n}$, is ordered in descending order from largest to smallest. In this case, it is obvious that the sequence of moments $\langle \pi_j \rangle$, constructed according to the rules of our procedure by multiplying the indicators by 1, by 2, etc., starting with the largest indicator by multiplying by 1, etc., the sequence $\langle \pi_j \rangle$ is single-peaked. Let this peak be reached at a certain index p_*. This indicator indicates a certain maximum achieved at the local level. Now suppose that, contrary to the achieved local maximum, we can find a certain subset H' of indicators on which the set function $\min \pi_{j \in H}(H) > \pi_*$; this means that on some set H' the global maximum is greater than the achieved local maximum π_*. If we now supplement the list of indicators H' appending H' to $H_k \supseteq H'$—to the list of all indicators in A starting from some indicator $p_k = \arg\min \pi_{j \in H'}(H_k)$, $k = |H_k|$—then, according to the lemma, it turns out that in our single-peaked sequence $\langle \pi_j \rangle$ we encountered an indicator p_k with a moment $p_k \cdot k$, $k = \pi' > \pi_*$. This is not possible due to our construction method of the single peaked sequence $\langle \pi_j \rangle$. ■

The "Strict Consistency Postulate", which has been validated in experiments, has been modified to suit the decision-making process, based on the premise that, along with these modified postulates, a reliable and reasonable way to statistically analyze data is provided. Somehow, however, the theorem of Aizerman and Maliszewski (Theorem I, 1981) may be useful, which states that the scalar condition of utility functions is necessary and sufficient for the truth of the modified strict consistency postulate. A thorough analysis or evaluation of this claim is beyond the scope of this work. If necessary, the list A of indicators must be presented in the form of linearly descending order of fuel consumption or other economic values related to vehicles.

This means that the choice operator $C(X)$ on the issues X of the list A of alternatives/indicators acts on a certain list of segments $S(A)$, which can be either open or closed by resembling the set of all sub-lists 2^A. Thus, we can call this dual terminology by choice or by segmentation/classification. The alternatives A can be identified by special issues, now denoted already as segments $X = [x\_l,\ x\_r]$ of the indicators under consideration. Narrowing a segment $Y \subseteq S(A)$ to a segment $X \subseteq S(A)$ is an action of narrowing the segment $Y = [y\_l,\ y\_r]$ to $X = [x\_l,\ x\_r]$. In view of this understanding that indicators are linearly descending, the situation $x\_r \geq y\_r$ with segments can preserve the original ordering of choice operators $C(X)$ nomenclature. To do this, in the notation just introduced, a set function is defined (hereinafter referred to as the function $f(X)$ of the segment X): $f(X) = x\_r$ or $f(Y) = y\_r$. With this function $f(X)$ notification, we are ready to prove Proposition II given below.

***Proposition II**. When shrinking the segment $Y \in S(A)$ to the segment $X \in S(A)$ as an extent of the segments of indicators of the common grand segment A, the condition $f(C(Y)) > f(C(X))$ is necessary and sufficient for the fulfillment of the non-strict consistency postulate.*

Proof.

Necessity. Suppose $X \subset Y$ and condition $f(C(Y) > f(C(X)))$ are satisfied, or in equivalent form between segments X and Y the situation results in $[c(y)\_r > c(x)\_r]$. Note the validity of $c(y)\_r = \{y \setminus c(y)\}\_l$ and $c(x)\_r = \{x \setminus c(x)\}\_l$. Thus, the $f(C(Y) > f(C(X)))$ condition results in $\{y \setminus c(y)\}\_l > \{x \setminus c(x)\}\_l$. Given that $X \subset Y$, we can rewrite the last inequality in set-theoretic notation as $X \setminus C(X) \subset Y \setminus C(Y)$, which indicates the validity of the non-strict consistency. ■

Sufficiency. Let us assume that the postulate of consistency is not satisfied for some segments $X \subset Y$ in the form of segments $S(A)$: i.e., contrary to the postulate of consistency, the condition $f(C(Y) > f(C(X)))$ is violated. Given the violation, we consider only the opposite case $f(C(X) > f(C(Y)))$, excluding the case $f(C(Y) = f(C(X)))$. The opposite case $c(x)\_r > c(y)\_r$ and $\{x \setminus c(x)\}\_l > \{c(y)\_r\}$ are equivalent. From this we conclude that it is possible to find an indicator $p_* = \{x \setminus c(x)\}\_l$ such that $p_* \in X \setminus C(X)$ in contrast to $p_* \notin Y \setminus C(Y)$. The last statement contradicts the consistency postulate, namely the violation $X \setminus C(X) \not\subset Y \setminus C(Y)$ of Proposition II. ■

6. Discussion, Findings and Conclusions

The rational choice postulates have led to intriguing results. The independence of rejected alternatives can explain how certain car brands are favored for dynamic options, while consistency plays a role in stable fuel consumption decisions. Factors like options available, preferences, and situations influence customer choices. Understanding buying behavior and context is crucial in the automotive industry. Manufacturers should align strategies with customer needs, address biases, and base offerings on objective data. Implementing a data analytics strategy like ***BDA*** can benefit customers in the automotive market.

With all this in mind, it is crucial to delve deeper into the intricate process of data analysis and categorization of data. It's imperative to recognize that this process transcends mere organization; it is a nuanced journey from objective depiction to subjective discernment. As emphasized repeatedly, the insights gleaned from data case study manifest not only in factual descriptions but also in subjective evaluations. These evaluations, often articulated as "interpretations, " serve as invaluable aids for experts across various domains, aiding them in navigating the complexities unearthed during the data case study endeavor. To facilitate this comprehension and preempt potential pitfalls, it becomes imperative to outline certain "situations" or "traps" those motorists may encounter. By elucidating these scenarios, specialists can leverage our comprehensive analysis methodology to preemptively address and mitigate the adverse ramifications of such situations, thereby fostering informed decision-making and proactive risk management.

6.1. Pitfalls interpretation

Potential car owners may prioritize dynamic options such as powerful engines and sportier styling over consistent fuel economy for reasons such as performance preference, driving experience or a desire for a more engaging and responsive ride. These people may prioritize the excitement and thrill of driving, valuing the dynamic aspects of the car over fuel efficiency.

Car buyers can become fixated on the starting price presented by the seller or on the sticker. They may find it difficult to negotiate or deviate from this anchor point, even if it is not the best offer. Some customers may prioritize the social status associated with owning a particular make or model of car over its practicality or affordability. They may be willing to spend more than they can afford simply to maintain or improve their social image.

When buying a car, impulsive behavior is prevalent as many customers make quick decisions without doing thorough research or thinking about the long-term consequences. They may fall in love with a particular car at first sight and rush into the purchase without evaluating the alternatives. Customers can be influenced by the opinions and actions of others, leading to a herd mentality. They may buy a car simply because their friends, family or colleagues have one, without properly evaluating their own needs and preferences. Emotional attachment to a particular make, model, or even color of car can also cloud their judgment.

Customers may overlook practical aspects such as fuel efficiency, maintenance costs or resale value, instead prioritizing their emotional connection. Some customers may be overconfident in their negotiation skills or car knowledge, leading them to make irrational decisions. They may be reluctant to seek expert advice and instead rely solely on their own judgment, which may result in increased fees or sub optimal choices. Customers may have a strong bias in favor of buying brand new cars, believing that new models are inherently superior, even though a used car with similar features could meet their needs at a lower price. This bias can lead to cost overruns and financial stress.

Some clients may be overly concerned about the fear of missing out or losing a perceived opportunity. This fear can lead them to make impulsive decisions or agree to unfavorable terms, driven by the desire to get a deal done quickly, even if it is not the best option available. However, it is important to note that while such behavior may be irrational from a purely logical perspective, it often stems from human psychology and the complex interplay of emotions, biases and social factors.

***APPENDIX* 1.** In data mining, separation refers to the ability to distinguish distinct patterns or classes within a dataset. Techniques like clustering, data analysis, or anomaly detection aim to separate data into meaningful groups based on similarities or differences. This separation helps uncover hidden patterns, trends, or outliers, contributing to better data understanding and decision-making. A well-known categorization in this direction is a Finite Closer System $C(X)$ of sub-lists $X \in 2^A$ of alternatives. Equivalent to a more precise definition provided by Seiffarth et al. (2021), our nomenclature will include the choice operator $C(X)$, which is given by Ctrl-s or a C-macro representing the *Fixed Point* $X = C(X)$ of the macro. Thus, based on the concept of a fixed point, the search problem of closed lists turns into a search for a system of sub-lists from a list A of alternatives such that each sub-list from this system is a fixed point of the operator Ctrl-s. From the database https://www.spritmonitor.de/en/ we have extracted a list A of almost all known gasoline-powered cars, where A includes well-established car models. The experiment shows that the separation consists of three segments of gasoline consumption per 100 km: **X=[26.59,…,9.80]**, **Y=[9.78,…6.16]** and **Z=[6.13,…, 4.75]**. The procedure for finding this separation is simple. First, the Ctrl-s macro is applied to all vehicles under investigation, resulting in the extraction of the first fixed point $X = C(X)$. In the remaining list $A \setminus X$, the Ctrl-s macro is implemented again, resulting in the next fixed point $Y = C(Y)$. Then, you need to extract the third one in the same way (accessed August 22, http://www.datalaundering.com/download/MPG-MileAge-Data.xls, , 2023).

***APPENDIX* 2.** Based on the information provided in the database, it appears that the segment **[5.47−9.80]** liters per **100 km** have been determined as a "reliable" range for gasoline consumption for all gasoline-fueled car models. Within this range, models to the left are considered more fuel-efficient, while those outside the range to the right are deemed to consume fuel more excessively. Particularly, based on the experiment conducted using models **A3** and **A4**, these two models fall within a range of no more than **3.22 l/100 km** to the right of the significance value u = **6.58 l/100 km**. On the other hand, the **A5**, **A6**, **A7** and **A8** models exceed the significance level u by more than **3.22 l/100 km**, indicating a noticeable increase in fuel consumption. We also observed that the segment **[5.47–9.80]** and the significance level **6.58** enable data analysis of all considered car models into four fuel consumption classes. The most economical Audi **A2** model is in the yellow class, as opposed to the blue class, which is quite economical, to which model **A1** belongs, and which is below the significant fuel consumption value **u** = **6.58**.

APPENDIX* 3.** The purpose of this piece of writing is to illustrate what we have called a procedure of "Blind Data Analysis" or ***BDA, based on a popular data containing information about various car models. The data includes indicators such as miles per gallon (mpg), horsepower and other attributes for 32 different car models. We invite the reader to check or at least review the fact that our visual representation of the correlation matrix in the example known to many analysts largely comes down to the same motives that constitute the essence of this study. Many data analysts often use this example (Henderson and Velleman, 1981).

	model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
1	Mazda RX4	21,00	6,00	160,00	110,00	3,90	2,62	16,46	0,00	1,00	4,00	4,00
2	Mazda RX4 Wag	21,00	6,00	160,00	110,00	3,90	2,88	17,02	0,00	1,00	4,00	4,00
3	Datsun 710	22,80	4,00	108,00	93,00	3,85	2,32	18,61	1,00	1,00	4,00	1,00
4	Hornet 4 Drive	21,40	6,00	258,00	110,00	3,08	3,22	19,44	1,00	0,00	3,00	1,00
5	Hornet Sportabout	18,70	8,00	360,00	175,00	3,15	3,44	17,02	0,00	0,00	3,00	2,00
6	Valiant	18,10	6,00	225,00	105,00	2,76	3,46	20,22	1,00	0,00	3,00	1,00
7	Duster 360	14,30	8,00	360,00	245,00	3,21	3,57	15,84	0,00	0,00	3,00	4,00
8	Merc 240D	24,40	4,00	146,70	62,00	3,69	3,19	20,00	1,00	0,00	4,00	2,00
9	Merc 230	22,80	4,00	140,80	95,00	3,92	3,15	22,90	1,00	0,00	4,00	2,00
10	Merc 280	19,20	6,00	167,60	123,00	3,92	3,44	18,30	1,00	0,00	4,00	4,00
11	Merc 280C	17,80	6,00	167,60	123,00	3,92	3,44	18,90	1,00	0,00	4,00	4,00
12	Merc 450SE	16,40	8,00	275,80	180,00	3,07	4,07	17,40	0,00	0,00	3,00	3,00
13	Merc 450SL	17,30	8,00	275,80	180,00	3,07	3,73	17,60	0,00	0,00	3,00	3,00
14	Merc 450SLC	15,20	8,00	275,80	180,00	3,07	3,78	18,00	0,00	0,00	3,00	3,00
15	Cadillac Fleetwood	10,40	8,00	472,00	205,00	2,93	5,25	17,98	0,00	0,00	3,00	4,00
16	Lincoln Continental	10,40	8,00	460,00	215,00	3,00	5,42	17,82	0,00	0,00	3,00	4,00
17	Chrysler Imperial	14,70	8,00	440,00	230,00	3,23	5,35	17,42	0,00	0,00	3,00	4,00
18	Fiat 128	32,40	4,00	78,70	66,00	4,08	2,20	19,47	1,00	1,00	4,00	1,00
19	Honda Civic	30,40	4,00	75,70	52,00	4,93	1,62	18,52	1,00	1,00	4,00	2,00
20	Toyota Corolla	33,90	4,00	71,10	65,00	4,22	1,84	19,90	1,00	1,00	4,00	1,00
21	Toyota Corona	21,50	4,00	120,10	97,00	3,70	2,47	20,01	1,00	0,00	3,00	1,00
22	Dodge Challenger	15,50	8,00	318,00	150,00	2,76	3,52	16,87	0,00	0,00	3,00	2,00
23	AMC Javelin	15,20	8,00	304,00	150,00	3,15	3,44	17,30	0,00	0,00	3,00	2,00
24	Camaro Z28	13,30	8,00	350,00	245,00	3,73	3,84	15,41	0,00	0,00	3,00	4,00
25	Pontiac Firebird	19,20	8,00	400,00	175,00	3,08	3,85	17,05	0,00	0,00	3,00	2,00
26	Fiat X1-9	27,30	4,00	79,00	66,00	4,08	1,94	18,90	1,00	1,00	4,00	1,00
27	Porsche 914-2	26,00	4,00	120,30	91,00	4,43	2,14	16,70	0,00	1,00	5,00	2,00
28	Lotus Europa	30,40	4,00	95,10	113,00	3,77	1,51	16,90	1,00	1,00	5,00	2,00
29	Ford Pantera L	15,80	8,00	351,00	264,00	4,22	3,17	14,50	0,00	1,00	5,00	4,00
30	Ferrari Dino	19,70	6,00	145,00	175,00	3,62	2,77	15,50	0,00	1,00	5,00	6,00
31	Maserati Bora	15,00	8,00	301,00	335,00	3,54	3,57	14,60	0,00	1,00	5,00	8,00
32	Volvo 142E	21,40	4,00	121,00	109,00	4,11	2,78	18,60	1,00	1,00	4,00	2,00

Table 2. The MTcars data frame with 32 observations on 11 (numeric) indicators

mpg	"**Miles/(US) gallon**" represents the fuel efficiency of different car models, specifically the number of miles they can travel on one gallon of fuel.
cyl	Represents the **count of cylinders** in the engine of each car model.
disp	"**Displacement (cu.in.)**" attribute refers to the engine displacement of each car model, typically measured in cubic inches (cu.in.).
hpc	**Gross horsepower** is a measure of the engine's power output before accounting for various losses, such as those from the transmission and accessories.
drat	"**Rear axle ratio**" attribute refers to the ratio of the number of revolutions the drive shaft makes to one revolution of the rear axle.
wt	"**Weight (1000 lbs)**" attribute represents the weight of each car model in thousands of pounds.
qsec	"**1/4 mile time**" is often used as a measure of a car's acceleration and performance, particularly in drag racing.
vs	Engine (0 = **V-shaped**, 1 = straight) is a binary indicator that categorizes the type of engine in each car model. A value of 0 typically represents a V-shaped (V6 or V8) engine, while a value of 1 represents a straight (inline) engine.
am	"**Transmission** (0 = automatic, 1 = manual)" attribute in the `categorizes the type of transmission used in each car model.
gear	"**Number of forward gears**" indicates the count of forward gears available in the transmission of each car model.
carb	"**Number of carburetor**" represents the count of carburetors in the engine of each car model. Carburetors are devices that mix air with a fine spray of liquid fuel for internal combustion engines.

For data analysis in practice, correlation matrices are usually calculated and visualized. Correlation matrix analysis involves examining the relationships between multiple variables by calculating and visualizing their correlations. Each cell in the correlation matrix displays the correlation coefficient, which indicates the strength and direction of the relationship between two variables. Positive values suggest a positive correlation, negative values indicate a negative correlation, and values close to zero suggest a weak or no correlation. This analysis helps in understanding patterns, dependencies, and potential multi-co-linearity of variables recorded in the database.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	1	-0.85	-0.85	-0.78	0.68	-0.87	0.42	0.66	0.6	0.48	-0.55
cyl	-0.85	1	0.9	0.83	-0.7	0.78	-0.59	-0.81	-0.52	-0.49	0.53
disp	-0.85	0.9	1	0.79	-0.71	0.89	-0.43	-0.71	-0.59	-0.56	0.39
hp	-0.78	0.83	0.79	1	-0.45	0.66	-0.71	-0.72	-0.24	-0.13	0.75
drat	0.68	-0.7	-0.71	-0.45	1	-0.71	0.09	0.44	0.71	0.7	-0.09
wt	-0.87	0.78	0.89	0.66	-0.71	1	-0.17	-0.55	-0.69	-0.58	0.43
qsec	0.42	-0.59	-0.43	-0.71	0.09	-0.17	1	0.74	-0.23	-0.21	-0.66
vs	0.66	-0.81	-0.71	-0.72	0.44	-0.55	0.74	1	0.17	0.21	-0.57
am	0.6	-0.52	-0.59	-0.24	0.71	-0.69	-0.23	0.17	1	0.79	0.06
gear	0.48	-0.49	-0.56	-0.13	0.7	-0.58	-0.21	0.21	0.79	1	0.27
carb	-0.55	0.53	0.39	0.75	-0.09	0.43	-0.66	-0.57	0.06	0.27	1

Table 3

Visualization of the MTcars correlation matrix in the form presented is available to everyone in the public domain.
https://miro.medium.com/v2/resize:fit:1400/format:webp/1*UJvgUROXv07GQsQCzukMAw.png

	-0,55	←Negative significance level									
	0,66	←Positive significance level									
	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	**1**	**-0,85**	**-0,85**	**-0,78**	**0,68**	**-0,87**	0,42	**0,66**	0,60	0,48	**-0,55**
cyl	**-0,85**	**1**	**0,90**	**0,83**	**-0,70**	**0,78**	**-0,59**	**-0,81**	-0,52	-0,49	0,53
disp	**-0,85**	**0,90**	**1**	**0,79**	**-0,71**	**0,89**	-0,43	**-0,71**	**-0,59**	**-0,56**	0,39
hp	**-0,78**	**0,83**	**0,79**	**1**	-0,45	**0,66**	**-0,71**	**-0,72**	-0,24	-0,13	**0,75**
drat	**0,68**	**-0,70**	**-0,71**	-0,45	**1**	**-0,71**	0,09	0,44	**0,71**	**0,70**	-0,09
wt	**-0,87**	**0,78**	**0,89**	**0,66**	**-0,71**	**1**	-0,17	**-0,55**	**-0,69**	**-0,58**	0,43
qsec	0,42	**-0,59**	-0,43	**-0,71**	0,09	-0,17	**1**	**0,74**	-0,23	-0,21	**-0,66**
vs	**0,66**	**-0,81**	**-0,71**	**-0,72**	0,44	**-0,55**	**0,74**	**1**	0,17	0,21	**-0,57**
am	0,60	-0,52	**-0,59**	-0,24	**0,71**	**-0,69**	-0,23	0,17	**1**	**0,79**	0,06
gear	0,48	-0,49	**-0,56**	-0,13	**0,70**	**-0,58**	-0,21	0,21	**0,79**	**1**	0,27
carb	**-0,55**	0,53	0,39	**0,75**	-0,09	0,43	**-0,66**	**-0,57**	0,06	0,27	**1**

Table 4

Visualization of the **MTcars** correlation matrix as it appears by applying the ***BDA*** method using the Ctrl+s macro
http://www.datalaundering.com/download/mtcars.xls

Comparing Table 3 from Table 4, it is easy to notice the almost complete similarity of the tables. The only difference is that the significance levels of the coefficients in Correlation Table 3 is a visualization associated with practical or common sense judgment based on experience and traditional reasoning, in particular with the color scheme, while those in Table 4 are based on the postulates of rational choice.

***APPENDIX* 4.** The goal of a graph classification problem is to assign labels to specific nodes or edges in the graph and to learn patterns and features that help make accurate decisions. The challenge is to efficiently aggregate and process information from a graph structure. When graphs are viewed as sets of edges, labels are often used for entire sub-graphs or individual edges. In our example, we have a graph representing correlations. Thus, a classification task may involve labeling specific groups of our **11** parameters from **32** car models forming certain relationships, representing sub-graphs with strong positive correlation "within a sub-graph" or with strong negative correlation "between sub-graphs", each edge of which is associated with a blue or red label. where the correlation is greater than **+.66** within or consistently less than **-0.55** between sub-graphs.

In the example below, Table 5 represents the $A*A$ multiplication, according to standard algebraic rules, obtained from the (0,1)-adjacency matrix A. The (0,1)-cells in A denote by 1 the correlation coefficients (Table 3 or Table 4) between the 11 parameters with a positive correlation threshold above +.66. Then Table 5, where diagonal cells contain 0-s, is converted to Table 6. Table 6 corresponds to vectors R and B outer-product $R \times B$ of the "total" column R to the right of Table 5 by the "total" row B at the bottom of Table 5. In Table 6 we leave only graphically adjacent vertices A, denoting with a 0-value those cells of Table 6 that do not indicate at adjacent vertices in A.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb	R
mpg	0	0	0	0	0	0	1	0	1	1	0	3
cyl	0	0	2	1	0	1	0	0	0	0	1	5
disp	0	2	0	1	0	1	0	0	0	0	1	5
hp	0	1	1	0	0	2	0	0	0	0	0	4
drat	0	0	0	0	0	0	0	1	1	1	0	3
wt	0	1	1	2	0	0	0	0	0	0	0	4
qsec	1	0	0	0	0	0	0	0	0	0	0	1
vs	0	0	0	0	1	0	0	0	0	0	0	1
am	1	0	0	0	1	0	0	0	0	1	0	3
gear	1	0	0	0	1	0	0	0	1	0	0	3
carb	0	1	1	0	0	0	0	0	0	0	0	2
B	3	5	5	4	3	4	1	1	3	3	2	34

Table 5; The product $\mathrm{A} \times \mathrm{A}$ of adjacency **(0,1)**-matrix A

		mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
1	mpg	0	0	0	0	9	0	0	3	0	0	0
2	cyl	0	0	25	20	0	20	0	0	0	0	0
3	disp	0	25	0	20	0	20	0	0	0	0	0
4	hp	0	20	20	0	0	0	0	0	0	0	8
5	drat	9	0	0	0	0	0	0	0	9	9	0
6	wt	0	20	20	0	0	0	0	0	0	0	0
7	qsec	0	0	0	0	0	0	0	1	0	0	0
8	vs	3	0	0	0	0	0	1	0	0	0	0
9	am	0	0	0	0	9	0	0	0	0	9	0
10	gear	0	0	0	0	9	0	0	0	9	0	0
11	carb	0	0	0	8	0	0	0	0	0	0	0
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*

Table 6; The R and B outer-vector product $\mathrm{R} \otimes \mathrm{B}$

We can now apply our ***BDA*** technique to Table 6, the result of which is shown in Figure 1. There are many classification methods on graphs, for example, Viswanathan, et al., 2010. We have also contributed to this field by using the so-called method of "Monotonic Systems" algorithm (simplified in our ***BDA***) to visualize the results of data analysis (Mullat, 1977). The correlation matrix visualization below is only an addition to ***BDA*** technique. As shown in Figure 1, there are two different classes that can be distinguished by dividing our 11 parameters given in Table 2.

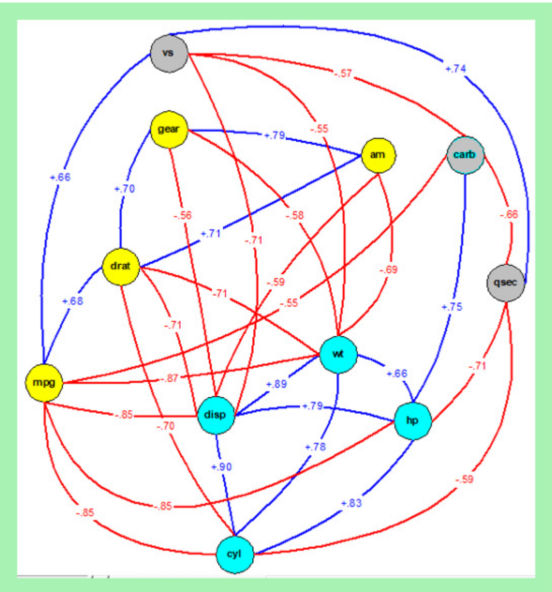

Figure 1; Visualization of ***BDA*** results using correlation of 11 parameters of Table 2.

Indeed: 1) **{cyl, disp, hp, drat, wt}**; 2} **{am, gear, carb}** and separately group of parameters 3) **{mpg, vs, qsec}**. We do not intentionally use any classification method, but simply use common sense, which we hope is sufficient to visualize the effectiveness of our ″Blind Data Analysis Procedure″. However, it can be proven that the first two classes visualize the separation of correlation coefficients when ***BDA*** is applied separately: initially by a block (outer-vector) $\overline{1,...,7} \otimes \overline{1,...,7}$, and then by $\overline{5,...,11} \otimes \overline{5,...,11}$ of rows and columns. The phenomenon of separation has already been discussed in Appendix 1.

References

Aizerman, M.A. and Malishevski, A.V. (1981). Some Aspects of the General Theory of Best Option Choice, Translated from Avtomatika i Telemekhanika, No. 2, pp. 65–83, February, 1981. Original article submitted September 5, 1980.

Arrow, K.J. a). (1948). The possibility of a Universal Welfare Function, Project RAND, RAD(L)-289, 26 October, Santa Monica, California (hectographed); b). (1959). Rational Choice functions and orderings, Economica, 26, No. 102, 121–127.

Bishop, C.M. (2016). "Pattern Recognition and Machine Learning, " Springer-Verlag, New York Inc., First Edition, ISBN 9781493938438.

Chernoff, H. (1954). Rational selection of decision functions, Econometrica, 22, No. 3, pp. 422–443.

Domingos, P. (2016). "The Master Algorithm:" How the Quest for the Ultimate Learning Machine, Ingram Publisher Services, ISBN-13: 9780141979243.

Frey, T. and Võhandu, L.K. (1966). Uus Meetod Klassifikatsiooniühikute Püstitatimiseks, Eesti NSV Akadeemia Toimetised, XV Kõide, Bioloogiline Seeria, Nr. 4, 565-576, *http://www.datalaundering.com/download/Frey-Vohandu.pdf* .

Han, J. and Kamber, M. (2011). "Data Mining: Concepts and Techniques," The Morgan Kaufmann Series in Data Management Systems, 3rd Edition, ISBN 978-9380931913.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). "The Elements of Statistical Learning", SAXO, Data Mining, Inference, and Prediction, Book, Second edition, Springer, ISBN : 978-0-387-84857-0.

Henderson, H.V. and Velleman, P.F. (1981). Building multiple regression models interactively for statistical analysis and data visualization. Biometrics, 37, 391-411.

Jamison, D.T. and Lau, L.J. (1973). Semiorders and the theory of choice, Econometrica, 41, No. 5, 901–912.

Kempner, Y., Mirkin, B. and Muchnik, I.B. (1997). Monotone Linkage Clustering and Quasi-Concave Set Functions, Appl. Math. Lett. Vol. 10, No. 4, pp. 19-24.

Koller, D. and Friedman, N. (2009). Probabilistic Graphical Models, Principles and Techniques. The MIT Press, Cambridge, Massachusetts, London, England, ISBN-13: 978-0262013192.

Ma, J., Li, W. and Dubois, D. (2015). Rational Partial Choice Functions and Their Application to Belief Revision. In Knowledge Science, Engineering and Management - Proceedings of 8th International Conference, KSEM 2015, (pp. 128-140). (Lecture Notes in Artificial Intelligence; Vol. 9403). Springer.

Mirkin, B.G., Muchnik, I.B. and Smith, T. F. (1995). Journal of Comp. Biology, pp. 493-507.

Mitchell, T.M. (1997). "Machine Learning", Publisher: McGraw-Hill Science/Engineering/Math; ISBN-10: 0070428077.

Mullat, J.E. a). (1971). On a Maximum Principle for Certain Functions of Sets, in “Notes on Data Processing and Functional Analysis”, Proceedings of the Tallinn Polytechnic Institute Series A, No. 313, Tallinn Polytechnic Institute, pp. 37–44, (in Russian), *http://www.datalaundering.com/download/RG-193-201-Modular.pdf*; b). (1976). Extremal Subsystems of Monotonic Systems, 1, Avtomatika & Telemekhanika, No. 5, pp. 130–139 (in Russian); *http://datalaundering.com/download/RG-268-284-extrem01.pdf*; c). (1995). A fast algorithm for finding matching responses in a survey data table, Mathematical Social Sciences 30, pp. 195–205, *https://www.sciencedirect.com/science/article/abs/pii/016548969500780P*; d). Data Structure Opening Method: Methodological Guide, *http://www.datalaundering.com/download/RG-211-244-method-guide.pdf*..

Murphy, K.P. (2012). "Machine Learning: A Probabilistic Perspective," ISBN-13: 978-0262018029, The MIT Press.
Narens, L. and Luce, R.D. (1983). How we may have been misled into believing in the Interpersonal Comparability of Utility. Theory and Decisions, 15, 247–260.
Narula, S.C. and Weistroffer, H.R. (1989). A flexible method for non-linear multicriteria decision-making problems. In IEEE Transactions on Systems, Man and Cybernetics, vol. 19, No. 4, pp. 883-887.
Nash, J.F. (1950). The Bargaining Problem, Econometrica 18, pp. 155-162.
Seiffarth, F., Horváth, T. and Wrobel, S. (2021). Maximal Closed Set and Half-Space Separations in Finite Closure Systems, Lecture Notes in Computer Science (including sub-series Lecture Notes in Artificial Intelligence and Lecture Notes in Bio-informatics), Vol.12457, p.3-18, https://arxiv.org/pdf/2001.04417.pdf (Accessed 5.08.2023).
Sen, A.K. (1971). Choice functions and revealed preferences, Rev. Econ. Stud., 38, No, 3 (115), pp. 307-317.
Simon, H.A. (1978). Assar Lindbeck (ed.). Nobel Lectures, Economics 1969–1980. Singapore: World Scientific.
Strzalecki, T. (2011). Axiomatic foundations of multiplier preferences, Econometrica, 79(1), 47–73.
Viswanathan, S.V.N., Schroudolf, N.N., Condor, R. and Borgwardt, K.M. (2010). Graphs Kernels, John Lafferty, ed., Journal of Machine Learning Research 11, 1201–1242.
Voelkle, M.C., Oud, J.H.L., Davidov, E. and Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomie. Psychological Methods, 17(2), pp. 176–192; Correction to Voelkle, Oud, Davidov and Schmidt (2012). Psychological Methods, 17(3), p. 384.
Walter, E. and Pronzato, L. (1997). Identification of Parametric Models from Experimental Data. London, England: Springer-Verlag.