

DATA ANALYSIS OF THE LOGISTICS COMPANY'S DATA BY MEANS OF BUSINESS INTELLIGENCE

Miroslava MUCHOVÁ, Ján PARALIČ, Barbora NAGYOVÁ

Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic, tel. +421 55 602 5122,
e-mail: miroslava.muchova@tuke.sk, jan.paralic@tuke.sk

ABSTRACT

The aim of this article is to present how we processed and analysed data from a logistics company using various Business Intelligence tools. The theoretical part of the article is therefore focused on defining Business Intelligence concepts and data warehouses that are relevant to the issue. The practical part of the article focuses on editing data, creating dimension tables and facts. The data were collected from the Dynafleet system and originated from a shipping company. The data provided are for the period from 2013 to 2016. We design user scenarios to help the company's manager in making an efficient assignment of drivers to planned delivery routes. The research is focused on design and creation of a logistics system based on data analytics that can continuously analyze the incoming data and generate current decision support reports. The created user scenarios have a wide range of uses and can also be helpful in assessing the performance of individual drivers and their workloads. Using a logistics system, the logistics manager can get the valuable and useful information needed to effectively operate the business.

Keywords: Business Intelligence, CRISP-DM, Oracle, Data mining, logistics

1. INTRODUCTION

At this time, information and data are all around us. Each person produces an enormous amount of data every day without realizing it. These data are often very valuable and therefore need to be stored and further processed. In the last years, various research studies have highlighted the benefits of using data in the logistics and supply chain. Data helps businesses to create new opportunities for growth, as they gain the ability to collect and analyze data directly from the industry. This way, it is possible to capture and analyze data about products and services, buyers and suppliers, consumer preferences and intentions, and to inspire changes in the organization.

Business data analysis was completely different from today. Most of the information was archived in paper form and many activities were performed manually. Today, these processes are automated and the data is stored on the cloud, disk, or data centres. The data analysis process is carried out by different means according to the needs of individual companies. One of the tools for processing and visualizing data is Business Intelligence (BI).

In this article, we focus on a specific problem that faces logistics company manager when assigning drivers to particular routes. Through information derived by proper data analysis, the user can monitor not only the consumption of vehicles, the driver or the entire fleet but also to observe other factors that influence fuel consumption. Based on this information, timely and accurate decisions can be made.

2. BUSINESS INTELLIGENCE

BI provides decision makers with reliable data in an appropriate context that refers to processes, knowledge, technologies and applications that facilitate decision-making data [1] for qualified decisions and business processes. BI technology sets the business goals based on

initial key performance indicators. BI is also able to perceive the interrelationship of the presented facts [2] to achieve the desired business. One of the most often used BI tools are data warehouses.

Features of quality business intelligence technology:

- expanding options – ensures direct usability,
- speed – responds to requirements,
- topicality – is available,
- accuracy – reliable quality,
- usefulness – provides value.

In business intelligence systems can we find these types of data storage:

- data warehouse,
- data marts,
- temporary data storage,
- operational data storage,
- personal data storage.

2.1. Data warehouse

Data Warehouse is (by Bill Inmon) a subject-oriented, integrated, time-variant and non-volatile collection of data to support the decision-making process. Subject-oriented means that data warehouse can be used to analyse a particular subject area. These subjects need to be specified during the design phase of the data warehouse.

2.2. Data marts

Data mart is a subset of the data warehouse oriented to a specific part of the business. Typically, they include a smaller set or one integrated data area [3] for a specific user group. They may be independent or be related to some data warehouse.

2.3. Temporary data storage

In some BI systems, the data is first copied to a temporary data storage before it is moved to the data warehouse. Data that has been inserted, altered or deleted in the production system is copied to this temporary storage as quickly as possible. During the copying process, some changes in the data structure can be made.

2.4. Operational data storage

Operational Data Storage is a simple data repository that provides assembled, integrated views of transient transaction data from many operating systems. Includes subjective, integrated data that is current or near-current to support day-to-day decision-making. An operational data storage [6] contains analyzed data that has a low granularity or a short retention time to maintain the data warehouse size. ODS allows organizations to perform OLAP analyzes.

2.5. Personal data storage

Personal data storages are designed specifically for the needs of a single user and represent a set of capabilities on a software or service platform that allow an individual to manage and maintain it or its digital information, artefacts or assets.

3. RELATED WORK

We found two very relevant publications presenting analyses of similar types of data that we had available. In the case study 0 authors describe the analysis of data from GPS and CAN bus systems. Data that provides the vehicle information (e.g., current location, speed), vehicle status information, weather data and other related data are available. Based on these data, the authors created a data warehouse. The logical model for the created data warehouse is based on the star schema. Logical model contains one fact table and 13 dimension tables. Data extraction from untreated files is not a trivial task, because the data comes from different formats and in different quality. The system is created as plug-in architecture and concurrently supports 16 different data sources. Some data sources have only a few value, but others have more than 25 values in one row. Data transformation and data integration include detection and manipulation with missing data, quality detection, data characteristics and their integration with various data sources.

In this article are used different rules of data cleaning. The data are stored in a temporary partition and waiting for the final migration process to be put into the data warehouse. Traffic jams can affect the time between two addresses. Using a data warehouse, it is possible to calculate the duration of a trip between two locations at a given time. Based on the created data warehouse, it is possible to identify the lowest fuel consumption by using information about where the vehicle was and what was its current consumption.

The purpose of the case study was to create a system that effectively responds to a wide range of real world demands from transport planners, managers and researchers. The system has been in operation since 2011

and is used for research and commercial purposes. GPS data from many sources is integrated into a single data warehouse model.

Another case study 0 focused on the functional analysis of truck fuel consumption. The data were collected using the Low Voltage Directive (LVD). Recorded vehicle data was collected from sensors inside the vehicle. The authors in this case study used methods such as principal component analysis, hierarchical clustering, the validation method, or functional data analysis. The first part of the data analysis on trucks was focused on the determination of the results using a basic multidimensional analysis and their comparison with the results of the functional analysis. This section shows the difficulties in applying the standard multidimensional method. Using the functional data analysis, the authors point to the problem of applying this analysis to the data. The main task was to apply the PACE algorithm (Preflight Analysis and Correction Engine). The aim was to determine the impact of seasonal changes on fuel consumption. The authors came up with the discovery that fuel consumption is difficult to predict because of the rapidly changing environment.

Over the last decades, transport companies have been trying to reduce fuel consumption through a variety of efficient driving programs. In these programs, motorists have to apply different specific techniques. Driver's performance and energy-efficient driving are gaining ever greater attention. Article [6] deals with applied ecological driving research, which is based on a huge data set. Through OLAP techniques and knowledge discovery techniques, the authors identified the main factors affecting average fuel consumption. Most of the related work on the Effective Driving Review focuses on measuring fuel consumption. The main task is to monitor the behaviour of the driver in order to maximize the efficient use of fuel. The authors in the article [7] created a support system that provides drivers with information on ecological driving to minimize fuel consumption.

4. ANALYSIS OF THE SELECTED SET OF DATA

CRISP-DM is one of the first industrial processes of data mining and knowledge discovery. It consists of the six phases that have been applied to the practical part of the article. We gradually passed through the phases that formed the result of the practical part.

The data used in this article come from a shipping company. It is a company whose business activity is the operation of freight road transport. The available data represent collected information about 4 vehicles and 18 drivers.

For the communication with vehicles, the company uses the Dynafleet Online system - Volvo Truck Corporation [10]. Information downloaded from vehicles is stored in the database. For the purposes of our work, we had two statements, namely the Fuel consumption assessment report, which contained 26 attributes as the driver's name, date, average speed, average fuel consumption, and other attributes that provide information about the overall rating of the driver's driving style - cruise control, economy time, above economy time, coasting, etc. The second report was the Tracking Report that contained 24 attributes, e.g. event

time, travelled distance, fuel, location, weight of cargo, and so on. All data was saved in the Excel spreadsheet format.

The data used in our study was collected between years 2013 and 2016. At this stage of CRISP-DM it is necessary to insert the data into the tables of facts and dimension tables [11], [12]. But first, we had to remove some attributes and select data that we will analyse. Within this data, there were not many values that had to be removed. Since the data provided are not compatible, we have selected attributes from the fuel consumption report and some attributes from the tracking set. An essential attribute for further analysis from the tracking set was the name of the driver of the truck and the country where he was driving the vehicle.

4.1. Creation of data model

Oracle SQL DataModeler was used to create the data model. With this tool, we created 12 dimension tables and 2 tables of facts. Dimension tables were created first. Afterward, we created tables of facts that had to be linked to dimension tables through their primary keys. The created data model consists of two fact sheets:

AAEF1_AVERAGE CONSUMPTION – the fuel consumption attribute is probably the most interesting for the user. It can monitor the consumption of the given truck, change of fuel consumption at higher speeds, how the weight of the vehicle affects the consumption, and so on.

AAEF2_TOTAL DISTANCE – is expressed in kilometers. Based on the total distance, we can find out which driver has passed the most kilometers for what time. At the same time, it is also possible to make other analyzes based on this fact.

The dimension tables in the data models were created from the selected attributes or by their appropriate merging into one dimension table:

AAE_DATE – this table contains the day, month, id, and year attributes. Date is a dimension that is relevant in all subsequently created analyzes.

AAE_STAT1 - AAE_STAT4 – there are tables containing ID and state attributes. These tables allow you to specify where the vehicle was at a given time or date.

AAE_VEHICLE – represents the vehicle designation. This attribute is linked to all of the tracked facts, otherwise it would not be possible to determine the average consumption of the analyzed vehicle or what vehicle was on the road.

AAE_DRIVER – this dimension contains the identification of the driver who was driving the vehicle. This dimension is well-usable in analyzes, where it is possible to see the exact fuel consumption and travelled distance for each driver.

AAE_PERFORMANCE_EVALUATION – is the most comprehensive dimension table because it contains information about the driver's driving style rating.

AAE_VEHICLE_USE – is a dimension that indicates the truck usage in percentage.

AAE_SPEED – this dimension contains the id attributes and the average speed in kilometers per hour. It expresses the speed of the given truck at a given time.

AAE_OVERALL_FATIGUE – this dimension shows

the amount of exhaust fumes produced by the truck during its journey at a given time.

AAE_TOTAL_TIME – represents the time period in hours and minutes, which indicates how long it takes for the driver to travel a certain distance.

4.2. Created user scenario

The goal of this article is to test the data from a logistics company using Business Intelligence tools. It focuses mainly on data preparation, efficient raw data processing, storage in the data warehouse, creation of user scenarios and analyzes that are interesting for the user. In an interview with the data owner, a logistics manager, the most important types of analyzes were identified that would help him make decisions. Analyzes were conducted in the form of user scenarios. After consulting with the manager and the owner of the company, we came to the conclusion that the first scenario could be an average vehicle consumption analysis. That should provide an overview of which vehicles have the best fuel consumption, what factors affect the fuel consumption of individual vehicles and how these factors can be influenced to reduce consumption.

Another scenario can be focused on tracking the consumption of individual drivers. Drivers who give the best performance while driving and how their work is divided. What factors can be improved to keep their fuel consumption as low as possible and what mistakes they are committed when driving. Consequently, the loading of vehicles could be analyzed in terms of the distance traveled over the years. Determine which vehicles are the most used and vice versa, which vehicle is the least used.

In this user scenario, we focused on how workers worked in provided time period of 4 years on the given trucks. Driver_G worked the most and Driver_M worked the least (Fig. 1). In the next step, we analysed which vehicles were used by which driver.

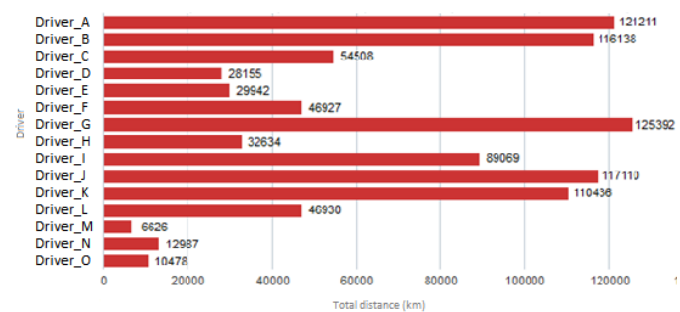


Fig. 1 Total distance travelled over 4 years

The number of kilometres that Driver_G passed on individual vehicles during the period of four years varies (Fig. 2). Vehicle_B and Vehicle_D were used the most, Vehicle_H was used the least. Vehicle_B was used through all 4 years and Vehicle_H was put into use in 2016 only. The interesting fact is that Vehicle_D and Vehicle_B were used for a similar number of kilometres despite the fact that Vehicle_D was in use for 2 years only.

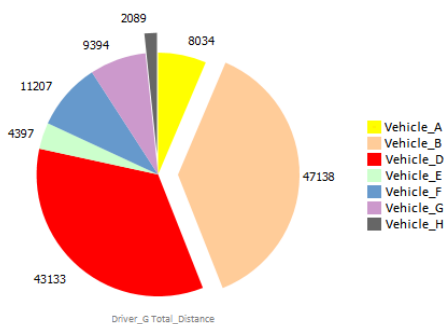


Fig. 2 Distance travelled by Driver_G in different vehicles

4.3. Adding a country attribute to find out in which country Driver_G drove the most

Vehicle_B and Vehicle_D were used the most. Therefore, we decided to analyse these two vehicles. The most visited countries when using the Vehicle_B are Norway and Germany (Fig. 3).

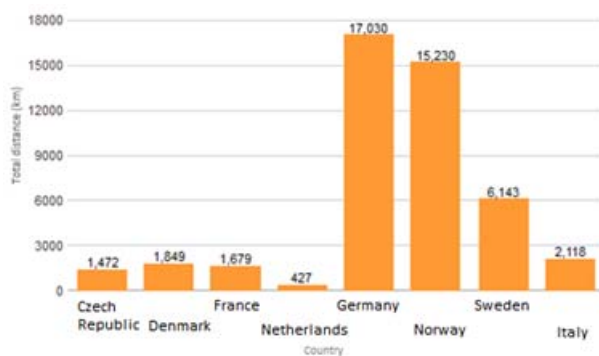


Fig. 3 The most visited country by Driver_G with Vehicle_B

The same results are when using Vehicle_D. Switzerland and Holland are visited the least (Fig. 4).

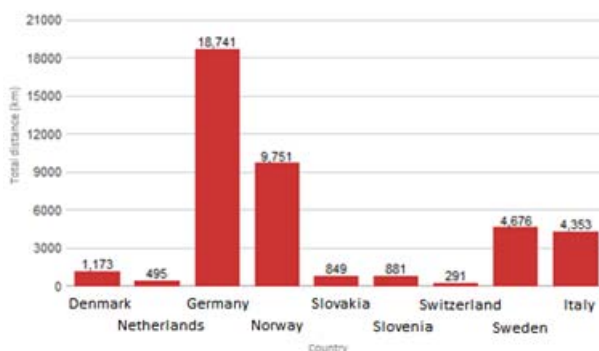


Fig. 4 The most visited country by Driver_G with Vehicle_D

In the next analysis, we decided to focus on these 2 the most visited countries. We wanted to find out what is the difference in fuel consumption in these 2 countries. The result is that the average consumption of Driver_G is a lot higher in Norway than in Germany.



Fig. 5 Fuel consumption by Driver_G with Vehicle_B and Vehicle_D in Norway and Germany

4.4. Focus on Norway and deeper analysis of the fuel consumption

In year 2016 Vehicle_B has higher average fuel consumption in January than in February. Only in these 2 months, the vehicle Vehicle_B was active (Fig. 6).

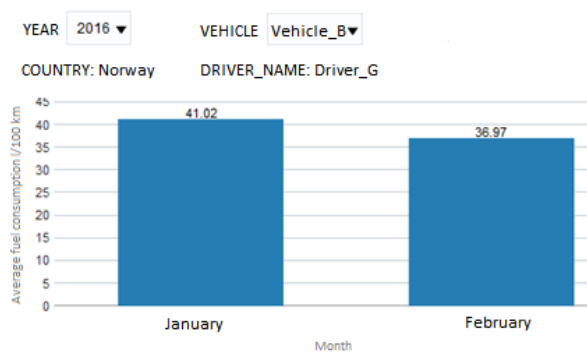


Fig. 6 Fuel consumption by Driver_G with Vehicle_B in Norway 2016

Vehicle Driver_D has the highest average fuel consumption in May. Vehicle_D was used in 5 months of 2016 (Fig. 7).

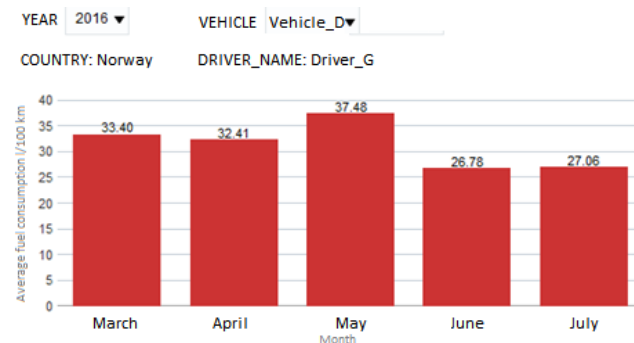


Fig. 7 Fuel consumption by Driver_G with Vehicle_D in Norway 2016

As the highest consumption of Vehicle_B was in January 2016, we decided to add another attribute - day and find out which days were critical for fuel consumption in January 2016. We found out critical days were January 9 and January 12. It is possible that the fuel consumption was affected by cold weather, which is usual in Norway during the winter.

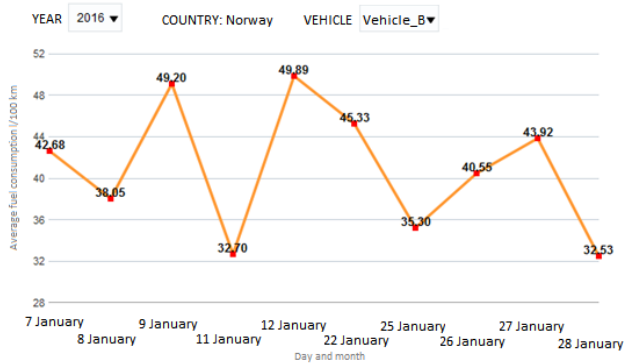


Fig. 8 Fuel consumption of Vehicle_B in January 2016

The results of this analysis are that driver Driver_G worked the most and Driver_M worked the least. Driver_G used mostly vehicles Vehicle_B and Vehicle_D. Vehicle_B, which was used for 4 years, drove only 4000km more than Vehicle_D which was utilized only in years 2015 and 2016. These vehicles are mostly used for transport within the countries of Norway and Germany. The rest of the countries are not visited as much as these, so there is a potential for expansion of truck transportation.

5. CONCLUSIONS

Data and information are a very important aspect of the business. The processing and use of data that is created in the company should be a daily part of the company's business. There are already a lot of tools for data processing, and one of them is Business Intelligence.

We also used this tool to analyse the provided data and make user scenarios. The findings described in user scenario are something new compared to existing data analysis solutions because none of the already existing solutions use BI to visualize the data.

We have proposed several user scenarios (analysis of average driver consumption in individual countries, analysis of fuel consumption in the annual period of individual countries). Based on these scenarios, the business manager can analyze drivers, route, average fuel consumption, and more.

Created user scenarios help to track the consumption of individual vehicles, where it is possible to see which of the vehicles has the highest consumption and why. It is also interesting to monitor the consumption of individual drivers and also to use them in time.

In the future, it will be possible to extend the solution with other attributes. Create a weather-enhanced data model that greatly affects the fuel economy of trucks. Another attribute that could be added is the current traffic situation, which also affects the average fuel consumption

ACKNOWLEDGMENTS

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-16-0213 and by the Research Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/0493/16.

REFERENCES

- [1] LABERGE, R.: Data Warehouses: Agile Methods and Business Intelligence (in Czech). Brno: Computer Press, 2012. pp. 352, ISBN 987-80-251-3729-1
- [2] VAN DER LANS, R.: Data Virtualization for Business Intelligence Systems: Revolutionizing Data Integration for Data Warehouses. United States of America: Morgan Kaufmann, 2012. pp. 296, ISBN 987-0-12-394425-2.
- [3] WESTERMAN, P.: Data Warehousing: Using the Wal-Mart Model. United States of America: Academic Press. 2001. pp. 297, ISBN 1-55860-684-X.
- [4] FERREIRA, J. – ALMEIDA, J. – ALBERTO S.: The Impact of Driving Styles on Fuel Consumption: A Data Warehouse and Data Mining based Discovery Process. In: IEEE Transactions on Intelligent Transportation Systems, 2015. pp. 2653-2662, ISSN: 1524-9050
- [5] PAREKH, A.: Introduction on Data Warehouse with OLTP and OLAP. In: International Journal Of Engineering And Computer Science, 2013. pp. 2569-2573, ISSN: 2319-7242
- [6] LAURENT, B. – HAO, J.: Simultaneous vehicle and driver scheduling: A case study in a limousine rental company. Computers & Industrial Engineering, 53:542-558, 2007
- [7] STRÖMBERG, H. K. – KARLSSON, I. C. M.: Comparative effects of eco-driving initiatives aimed at urban bus drivers – results from a field trial. Transp. Res. Part D: Transp. Environ. 22, 28–33. 2013
- [8] VENKATESWARA, R.: Data Warehouse Building Using Standard Query Language Server. Lulu Press, Inc. 2005. ISBN 1329404718.
- [9] VOGETSEDER, G: Functional Analysis of Real World Truck Fuel Consumption Data. Technical Report, 2008
- [10] Vehicle communication system Dynafleet Online – Volvo Truck Corporation, available on the Internet: <http://www.volvotrucks.com/trucks/dynafleet-help/splash/Pages/splash.aspx>
- [11] BHANSALI, N.: Strategic Data Warehousing: Achieving Alignment with Business. United States of America: Taylor and Francis Group, LLC. 2010. pp. 224, ISBN 978-1-4200-8391-1.
- [12] CHAPMAN, P. a kol.: CRISP-DM 1.0., available on the Internet: <https://www.the-modeling-agency.com/crisp-dm.pdf>

Received December 21, 2017 , accepted October 8, 2018

BIOGRAPHIES

Miroslava Muchová was born on 22.3.1990. In 2014 she graduated (MSc) with distinction at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at Technical University in Košice. Since September 2014 she works as PhD. student at the Department of Cybernetics and Artificial Intelligence. Her scientific research is focusing on data analysis in logistics process (identify the key factors that affect fuel consumption of vehicles and also to identify best practices and driving styles of drivers). She also participates in various research projects.

Ján Paralič received his Master degree in 1992, his Ph.D.

degree in 1998 and became associate professor in 2004 at the Technical University in Košice. Since 2012, he is full professor and deputy head at the Department of Cybernetics and Informatics, Technical University in Košice. His research interests are in the areas of knowledge discovery, text mining, big data analytics, and knowledge-based approaches in information systems.

Barbora Nagyová was born 17.9.1993. In 2017 she graduated (MSc) with distinction at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at Technical University in Košice. Her thesis title was “Data analysis of the logistics company's data by means of business intelligence”.