

DATA MINING WITH MICROSOFT TECHNOLOGY FAVORS ANALYSIS AND PREDICTIONS IN THE AUTOMOTIVE SUPPLIER INDUSTRY

DATA MINING WITH MICROSOFT TECHNOLOGY ENABLES ANALYTICS AND FORECASTING FOR AUTOMOTIVE SUPPLIERS

Martínez Ramírez Violeta¹, Susana Martínez Rabanales², Teresa Luciano Machorro³,
Efrén Armando Osorio Ramírez⁴, Francisco Manuel Pérez Sánchez⁵

<https://doi.org/10.61117/ipsumtec.v7i2.333>

¹Doctorate in Education. Tecnológico Nacional de México, Campus Instituto Tecnológico de Puebla. Department of Systems and Computing. Address violeta.martinez@puebla.tecnm.mx, Av. Tecnológico 420 Col. Maravillas, C.P. 72220, Puebla, Puebla, México. <https://orcid.org/0000-0003-1518-786X>

²Master's Degree in University Teaching. Tecnológico Nacional de México, Campus Instituto Tecnológico de Puebla. Department of Systems and Computing. Address susana.martinez@puebla.tecnm.mx, Av. Tecnológico 420 Col. Maravillas, C.P. 72220. Puebla, Puebla, México. <https://orcid.org/0009-0002-8510-5380>

³Master's Degree in Engineering. Tecnológico Nacional de México, Campus Instituto Tecnológico de Puebla. Address: teresa.luciano@puebla.tecnm.mx, Av. Tecnológico 420 Col. Maravillas, C.P. 72220. Puebla, Puebla, México. <https://orcid.org/0000-0003-3979-9369>

⁴Doctorate in Science in Business Planning and Regional Development. Tecnológico Nacional de México, Campus Instituto Tecnológico de Puebla. efrén.osorio@puebla.tecnm.mx, Av. Tecnológico 420 Col. Maravillas, C.P. 72220. Puebla, Puebla, México. <https://orcid.org/0000-0002-6902-1841>

⁵9th semester student of the Information and Communications Technology Engineering program. Tecnológico Nacional México, Campus Instituto Tecnológico de Puebla. Address: i18221794.19@puebla.tecnm.mx, Av. Tecnológico 420 Col. Maravillas, C.P. 72220 Puebla, Puebla, México, <https://orcid.org/0000-0002-3861-8909>

Abstract - Microsoft SSIS is a platform for data transformations and data integration aimed at providing complex business solutions, whose focus is directed to data cleansing and data mining [1]. paper intends to show the use of Microsoft SSIS as a tool that favored the analysis of data and predictions generated on the sales made in a business unit focused on providing items within the automotive sector in the region, managing information and data objects in SQL Server.

Keywords: Analysis, mining, predictions, SQL, Server.

Abstract -- Microsoft SSIS is a data transformation and data integration platform aimed at providing complex business solutions, whose focus is on data cleaning and mining. The present work intends to show the use of Microsoft SSIS as a tool that favored the analysis of data and predictions generated on the sales made in a business unit focused on providing items within the automotive sector in the region, managing the information objects. and data in SQL Server.

Key words - Analysis, mining, predictions, SQL, Server.

INTRODUCTION

The region of Puebla has two automotive plants of great national and international relevance and with them a large number of companies dedicated to supplying articles and products for their auto parts which, due to their own inertia of quality, are in constant technological evolution at the same time that they are in constant technological evolution.

the demands of the plants. Innovating in computer technology demands an urgency to implement digital technologies.

See the importance, KNAUF Industries, reports is its portal the year 2018, the highest number of patent applications in the technological area of autonomous or intelligent cars lies in Europe with 33.3% of the world total. Ranking above the United States with 30.1%. The third place goes to Japan with 11.5% and the fourth to South Korea with 10.5% [1].

According to eng publica's "Driving Innovation for Automotive Suppliers" portal, the automotive sector is applying severe pressure on automotive parts and systems suppliers (defined as Tier 1 manufacturing) to prioritize investment in next-generation technologies to reduce costs and margins. Solutions may include, among others, the use of AI and MES (man woman, machine, method, measurement, material), which enable process transformation and achieve innovation, for automotive suppliers to successfully meet these demands [2].

Data analysis for data mining requires applications that facilitate its goal, that is why Microsoft's SSIS with its Integration Services capability extracts and transforms data from various sources such as XML data files, flat files and relational data sources and, loading the data into one or more destinations is a wise choice [3].

Microsoft applications offer the visualization of data stored on servers for subsequent debugging, dynamic presentation by means of simple interaction panels.

Problem

One of the main problems of the current server is the support because it uses Microsoft technologies of 2012, the increase of clients in the different branches of the country produces an increase of data, in turn increases the costs in the support and maintenance of the server, another problem is the accessibility since currently all the departments have access to their respective information of the production server, this produces that the performance is affected when using obsolete technology.

In the sales area of the company all company information reaches the BigData area where analysis, estimations, data mining and creation of processes for the different departments are performed, therefore the server receives a very high workload every day, in addition to the requests it receives from applications and services, causing delays when consulting or updating information. For data extraction and data mining currently the production server is very slow in response times data analysts working in the company reducing efficiency of employees, in addition to data engineers are responsible for performing SQL transactions to display the desired information from each department for it limits the resources of other employees who use the server.

Finally, some processes, tasks and services within the server have not been technologically updated since their creation, resulting in poor performance by not taking advantage of the server's hardware.

Considering the problems, it was concluded that it is necessary to change the server in order to have the latest technologies and take advantage of the tools offered by this cloud computing service.

General Objective

Migrate the company's processes and services to a cloud environment with Microsoft SSIS and Power BI tools to obtain the latest technologies, as a result achieve a predictive analysis of the behavior in the sales area.

In order to achieve the above, the following specific objectives must be established:

- Analyze and debug information from all sales departments.

- Analyze, restructure and upgrade databases to Microsoft SQL Server 2022.
- Optimize SQL queries for each of processes, if necessary.
- Migrate Jobs, Tasks, Web Services, Programs, SFTP Folders.
- Upgrade ETL processes to Microsoft SSIS 2022 to achieve the highest performance on the new server.

Scopes

Possibility of contracting services or tools for the server that in the future will help the departments to achieve their future goals or objectives.

Sales department will be able to open new branches without worrying about the increase of information in future openings.

Improved accuracy in future calculations for sales estimates and projections.

Once the server is deployed, control rules can be implemented to the server to avoid the excess charges that are incurred when systems come online after a sharp increase in demand is completed.

Long-term productivity gains will be achieved in each of the departments.

State of the Art

Recent publications in Mexico, show how the Microsoft tool applied in the processing and analysis of Power BI data, helped in the historical predictive analysis of maritime mercantile traffic in Campeche experienced from 1969 to 1973 where it is possible to visualize in graphs, maps and timelines the maritime mercantile activity of Yucatan, where the documents are dated [4].

In Peru, the César Vallejo University of Lima, disseminates the thesis work called "Implementation of business intelligence using data mining for decision making in sales", which applies data mining to obtain a predictive analysis sales using Microsoft SQL Server and Power BI tool under a qualitative approach, so sales were improved by having the study [5].

In Ecuador, a paper published by the Universidad Técnica Particular de Loja entitled "Data mining model of profitability of corporate customers of Produbanco", uses a template within the Power BI application for modeling, facilitating the relevant information to generate a projection of customer profitability, significantly reducing response times in decision making [6].

The Technical University of Babahoyo in Babahoyo, Ecuador, publishes its work "Application of data mining for decision making in SMEs", which concludes that Power BI is an effective tool for visualizing data in reports and interactive panels, which favors the identification of patterns and trends. Its ease of use and integration with other Microsoft tools make it an excellent choice for any SME seeking to optimize its decision making in a timely manner [7].

A recent publication entitled "Effective approach to data analysis and visualization using Microsoft Power BI in agribusiness" defines Power BI as a visual and intuitive tool, due to its agile report visualization panel, generation of statistical reports, sales, new clients, income, in addition to facilitating analysis according to the regional situation in Latin America and the Caribbean; In business, it optimizes the time for the implementation of commercial strategies that allows to improve its management, concluding that the business analysis solution substantially improves competitiveness in the market, positioning itself as an effective tool for decision making [8].

DEVELOPMENT

Next, the method for data extraction and data mining is shown, implementing technology that allows us to automate processes for different departments that benefit the business model within the automotive sector.

Methodology, creation of enterprise data integration solutions with Microsoft Technology.

ETL allows you to connect to a variety of data sources, including SQL databases, flat files, web services and other data sources.

In today's business model, efficient data extraction is crucial for success. Microsoft Integration Services (SSIS) is a powerful tool that allows the automotive company to effectively integrate, transform and manage data from various sources. It is worth mentioning that the company has the technological infrastructure to use these information extraction methodologies.

To use this technology it is essential to have the following infrastructure such as a server can be in the cloud or a physical server but it is necessary that it is managed by Microsoft services that is to say that it has installed Windows Server (Virtual Machine), SQL Server, SQL Server Management, Visual Studio and the complement of integration services services, this so that it can be deployed to the server and allow us to schedule the time of extraction of information if necessary.



Figure 1. Remote connection to server virtual machine with Kerberos security.

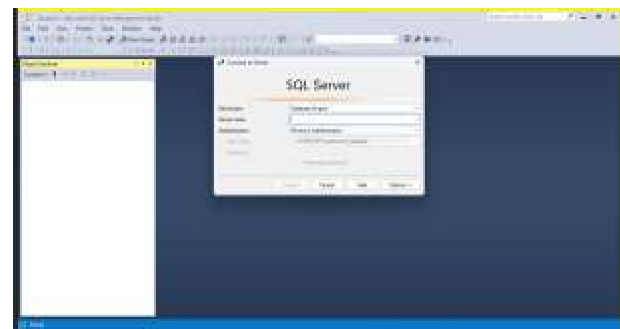


Figure 2. SQL Server Management Studio connection.

For the creation of data integration solutions it is necessary that the Big Data department team of Data Engineers and Data Analysts maintain constant communication, as it is necessary for the creation of the ETL.

The procedure used for the creation of solutions consists of the following steps:

This allows us to maintain a workflow and if the information will be transformed and how it will be returned to the target source, the following illustration shows an example of the programming of an ETL from the requirements of the Sales department.



Figure 3. ETL programming according to requirements.

The illustration above shows the programming of how it extracts information and transforms it from

certain conditions, calculations if it is the case and debug certain information that is not necessary to the target source, it is necessary to specify if the target source is a file that will be generated as new or is an existing file.

Another point to consider in this step is that when programming the ETL is to automate the sending of emails with the transformed information, this is achieved thanks to the Visual Basic programming language.

The second step is to automate the extraction of the information once the ETL is deployed, which allows specifying the time and date of execution, i.e. the extraction concurrency, as shown in the following illustration.

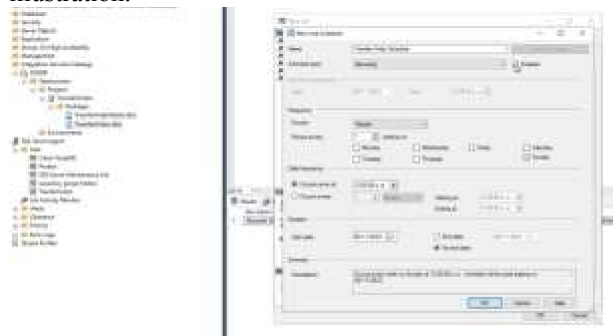


Figure 4. Automation of data extraction.

Figure 4 shows how the extraction of information from the company is automated, this allows us to reduce the time and effort needed to collect data, allowing employees to concentrate on more strategic tasks.

The last step is to validate if the information will be visualized with business intelligence tools that allow us to interact with the information, this ensures that the extracted information is more accurate and consistent, the tool used by the company is Power BI that allows us to collect and analyze information through filters and pivot tables of the data already transformed by the ETL as shown in the following figure.



Figure 5. Power BI dashboard with information extracted by ETL.

It is important to mention that the sales department is in charge of manipulating the extracted information and this information is again compiled by an ETL, so that it can later be converted into historical information.

Historical information is of utmost importance, data analysts are responsible for creating algorithms and business cubes that allow creating future sales predictions for all or certain automotive products, again an ETL is responsible for collecting this information to transform them into reports and then send them to the sales department, so it can generate indicators and sales targets are achieved, repeating an automated cycle of information and data.

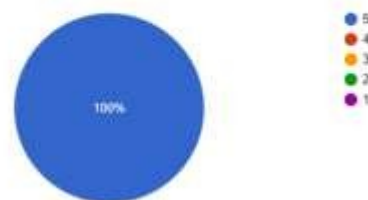
DISCUSSION AND ANALYSIS OF RESULTS

Likewise, a customer satisfaction survey was conducted, that is to say, to each one of the managers of the 7 departments, which will allow us to use it as a satisfaction indicator that will allow the creation of graphs, where the highest degree of satisfaction is 5 and 1 is the lowest degree of satisfaction.

Graph No. 1 addresses the need to know about the visual presentation of the screens with which the end user constantly interacts, rating their visual quality as UX user experience, 100% fully satisfied.

¿La información histórica de la empresa se visualiza correctamente para los departamentos que requieren realizar análisis de datos?

7 respuestas

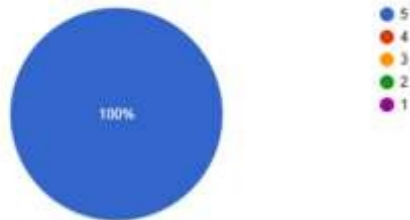


Graph 1. Question No. 1 with 100% excellent satisfaction.

Graph No. 2 shows the relevant information of the operations on the information management process, this image validates that the actions corresponding to the recurring transactions on the stored information have covered the requirements at 100%.

¿Los procesos están clasificados para una mejor gestión?

7 respuestas

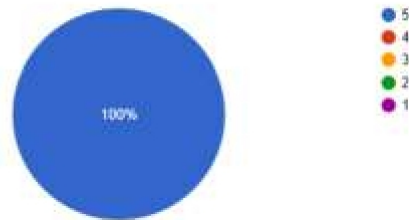


Graph 2. Question No.2 with 100% excellent satisfaction.

Graph No.3 shows the importance of displaying the list of actions that the user executes efficiently with a high satisfaction value.

¿Todos los procesos se encuentran dentro de un catalog?

7 respuestas

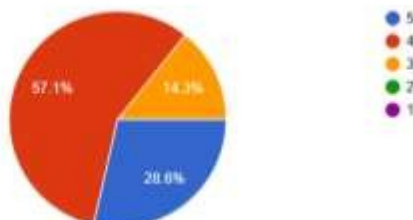


Graph 3. Question No. 3 with 100% excellent satisfaction.

Graph No.4 shows the evaluation of the agreed delivery dates within the established schedule, and most users consider the delivery period to be acceptable.

¿El proyecto fue terminado en las fechas acordadas?

7 respuestas

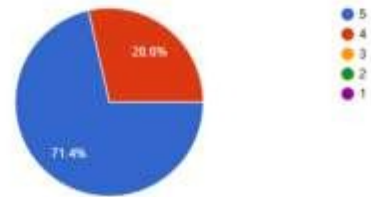


Graph 4. Question No. 4 with a good satisfaction rate of 57.1%.

In the following graph No. 5 focuses strictly on cybersecurity implemented in the implemented software, more than 70% express that they perceive the server configuration to be very secure.

¿El nuevo servidor cumple con las medidas de seguridad pertinentes para la ejecución Jobs?

7 respuestas

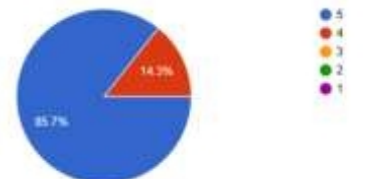


Graph 5. Question No. 5 with an excellent satisfaction rate of 71.4%.

Regarding the control of information and its consistency, the numbers shown in Figure 6 reveal that users are highly satisfied with the measures taken to extract information stored in the database, as shown in Figure 6.

¿Se utilizaron usuarios de servicio para un mejor control de las bases de datos?

7 respuestas



Graph 6. Question No. 6 with an excellent satisfaction rate of 85.7%.

The evaluation of the explanatory data of the reports elaborated by means of the application created shows an overwhelming satisfaction of them. See graph 7.

¿Los reportes de los Jobs son suficientemente detallados?

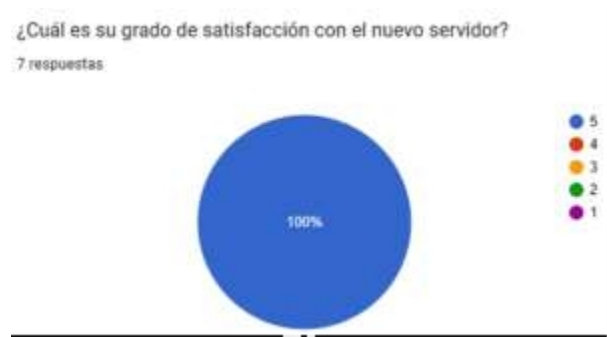
7 respuestas



Graph 7. Question No. 7 with 100% excellent satisfaction.

Once the entire cybersecurity system has been implemented and the information has been controlled for analysis, it is important to know whether the implementation of the new server provides storage confidence, resulting in the

graph that fully assesses end-user satisfaction. See graph No. 8.



Graph 8. Question No. 8 with 100% excellent satisfaction.

The results obtained by automating information extraction are the following company processes:

Sending and receiving mailings that are useful for the company and its different departments.
Create routines that collect information from the ETLs that were scheduled at the time of deployment. This allows us to create reports for employees that generate sales calculations for a previous day.
The Power Bi dashboards are automatically fed by the information extracted by the ETL.

It allows you to automate algorithms through Python with the data it provides in real time, improving your ability to make informed and timely sales decisions.

The development department should only be concerned with automating tasks or activities that require manual data entry, since the information they require is automated.

Discussion.

Authors Riquelme, Ruiz and Gilbert of the article "Data Mining: Concepts and Trends" state that, "sooner rather than later data mining will be used by society, at least with the same weight that Statistics currently has", its greatest contribution will be the ability to extract useful information, decision making or exploration, since raw data is of little use, hence the importance of having the right tool for its cleaning. The use of computers to apply mathematical methodologies in statistics gives rise to Machine Learning of today and the future [8].

The challenges for the future are the volume of historical information that will be stored on gigantic servers, methodologies that Big Data will address.

In this regard, the INNOWISE portal states that traditional data analysis technologies cannot handle Big Data, non-traditional techniques such as machine learning and artificial intelligence that are currently in demand must be used [9].

For Endeavor in Mexico, positions within the 6 Megatrends in technologies artificial intelligence (AI), which allows the development of systems capable of performing tasks that require human intelligence, including decision making to streamline the emerging responses of human behavior to preferences in terms of sales and purchases to place a company in sight of its competitors [10]. So the future catches up with us, AI, Big Data, Data Mining are related to provide human mapping of historical behaviors for better predictive analysis and informed decision making.

CONCLUSIONS

The migration to the new server with Windows Server 2022 and its respective data science tools such as Power BI desktop, Visual Studio 2022, SQL Sever 2022 and SQL Server Management Studio 2022 provides improvements in performance and efficiency as opposed to the old server, in addition to providing greater functionality for the processes currently in place.

According to the new features of SSMS us with the functionality of a detailed tracking of each ETL entered in the catalog, in addition to a better stability of execution, finally, by adding service users provides better management and control of solutions in the different databases and tables that point.

As a recommendation when creating a new ETL it is necessary to specify a name that gives us a little more context on how the Job works, also specify a name for each package so that they can be managed and analyzed in a better way in future migrations.

Recommendations

When updating this work in the medium term, an ETL would be deployed using functionality provided by Microsoft that is creating the catalog which will allow us to track quickly and details possible errors in the execution of ETL, so it is not recommended by file system as it is limited in functions.

If new ETLs are made it is necessary to document them in order to have a reference point when making modifications or if it is necessary to debug the processes again, the documentation provides us with better

context and will help in making decisions if the process is no longer useful or needs to be updated.

Within the server, machine learning can be created to learn and provide recommendations with customer data for future applications.

Test new algorithms with Python and Visual Studio with the ability to identify patterns in data and enable future predictions.

The research to be carried out in a period of no more than 1 year would be to focus on data quality assurance after a simple real-time search on data mining hand in hand with the theory of big data will offer the exploitation of prospective data that will provide cutting-edge solutions in various service and technology sectors, with simple and intuitive interfaces. [12]

As the most precious value in this company, the second stage of the technological research presented, considers essential the future inclusion to obtain greater increase in data mining by efficiently managing the entire life cycle of data mining, from data exploration to the production of new data, as an example is seen businesses in the insurance sector, who require the integration of a data mining method because of the relevance of the data stored in their computer systems. [13]

It should be considered that, in a time no longer than 6 months, inserting predictive processes on implicit customer behavior within the system will be able to obtain a high level of prediction of their behavior, thus favoring marketing. [14]

Another strategy to be taken into account should be included in the technological research presented in this work, focused on implementing prediction. With the added predictive mining it is possible to study the real impact that decision making has on the future of the company and to choose the most effective options.[15] The following is an example of a strategy that should be taken into account.

BIBLIOGRAPHY.

- [1] knaufIndustries (2024). Main lines and trends in innovation and development of the automotive industry in Europe. Available at <https://knaufautomotive.com/es/innovacion-en-el-sector-del-automovil/>
- [2] eng. (2024). Driving innovation for automotive suppliers. automotive. Available at <https://www.indx.com/es/posts/driving-innovation-for-automotive-suppliers>
- [3] Microsoft (2024). SQL Server Integration Services Available at <https://learn.microsoft.com/es-en/sql/integration-services/sql-server-integration-services?view=sql-server-ver16>
- [4] Einstein, M. (2024). Data mining...: Paper. Available at

https://scholar.google.es/scholar?hl=es&as_sdt=0%2C5&q=miner%C3%ADa+of+data+with+power+bi&btnG=&btnG=&btnG=

- [5] Tantalean Salazar, B. B. (2023). Implementing business intelligence using data mining for sales decision making. Available at <https://repositorio.ucv.edu.pe/handle/20.500.12692/131155>
- [6] González Eras, Alexandra Cristina, Garcés Poveda, Cristian Andrés (2020). Profitability data mining model of Produbanco's corporate clients. Available at <https://dspace.utpl.edu.ec/handle/20.500.11962/25253>
- [7] León, T., & Edigno, A. (2023). Application of data mining for decision making in SMEs (Bachelor's thesis, Babahoyo: UTB-FAFI. 2023). Available en <http://190.15.129.146/handle/49000/14253>
- [8] PA Barahona-Morales, E Prado-Carpio, J Quezada-, E Prado-Carpio, J Quezada-, J Quezada-, J Quezada-, J Quezada-. Campoverde, VJ Garzón-Montealegre (2023) Effective approach to data analysis and visualization using Microsoft Power BI in agribusiness. Multidisciplinary Latin American Journal (MLAJ) mlaj-magazine.org. Available at <https://mlaj-revista.org/index.php/journal/article/view/23>
- [9] Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Data Mining: Concepts and Trends. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, 10(29), 11-18. Available at <https://www.redalyc.org/comocitar.oa?id=92502902>
- [10] INNOWISE (2024). The future of big data - Predictions and research| Data analytics and its impact on the enterprise. Available at <https://innowise.com/es/blog/the-future-of-big-data-predictions-and-researches-data-analytics-and-its-business-impacts/>
- [11] Megatrends: That will revolutionize 2024. (2024). Technologies that Available at https://mexico.endeavor.org/megatendencias/?gad_source=1&gclid=EAIaIQobChMI5rbMqtjtjtiAMVdAitBh0uEA3QEAAAYASAAEgKmSvD_BwE
- [12] LOGICALIS (2014). Trends in data mining. Available at <https://blog.es.logicalis.com/analytics/tendencias-en-data-mining>
- [13] LISA Insurtech (2024): What will the future of mining look like? of mining? Available at <https://lisainsurtech.com/como-sera-el-futuro-de-la-data-mining/>
- [14] iberdrola (2024). Discover how data mining will predict our behavior. behavior. Available at <https://www.iberdrola.com/innovacion/data-mining-definition-examples-and-applications>
- [15] AWS amazon (2024) What is data mining? Available at <https://aws.amazon.com/es/what-is/data-mining/>

CONTRIBUTION ROLES

Role	Author(s)

Conceptualization	Francisco Manuel Pérez Sánchez
Data Curation	Francisco Manuel Pérez Sánchez
Methodology	Susana Martínez Rabanales
Project management	Susana Martínez Rabanales
Resources	Proofreading and editing - Violeta Martínez Ramírez
Software	Francisco Manuel Pérez Sánchez
Supervision	Efren Armando Osorio Ramirez
Validation	Francisco Manuel Pérez Sánchez
Visualization	Teresa Luciano Machorro
Editorial staff	Original draft - Violeta Martínez Ramírez
Editorial staff	Proofreading and editing - Violeta Martínez Ramírez



This work is
 licensed under an
 international license
 Creative Commons Attribution 4.0.