Studies in Engineering and Technology
Vol. 7, No. 1; August 2020
ISSN 2330-2038 E-ISSN 2330-2046
Published by Redfame Publishing
URL: http://set.redfame.com

Data Analysis for the Preventive Maintenance of Machinery

Ricardo Piqueras[1], Jose Maria Fernandez-Crehuet[1]

[1]Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, Calle José Gutiérrez Abascal nº2, 28006 Madrid, Spain

Correspondence: Jose Maria Fernandez-Crehuet, Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, Calle José Gutiérrez Abascal nº2, 28006 Madrid, Spain.

Received: November 23, 2019 Accepted: August 18, 2019 Online Published: October 15, 2019

doi:10.11114/set.v7i1.2814 URL: https://doi.org/10.11114/set.v7i1.2814

Abstract

This article shows the importance of data for the resolution of current problems, through a case study of preventive maintenance of a company's machines. From a given inspection of the condition of the machines, enough data is obtained to simulate the operation of the machines under specific conditions using Python programming environment, for a long enough period of time to obtain reliable information about the machines' lifetime. By using "R", the statistical analysis of the data is performed to determine the optimal period between repairs, schedule the company's preventive maintenance and show the possibility of solving complex problems from a simple data set.

Keywords: simulation, data, linear regression, statistical analysis, maintenance

1. Introduction

In recent times, the maintenance of machinery and equipment has experienced great technical breakthroughs that have generated a new mentality around this activity (Rivas, 2020). During the beginnings of the use of machinery for industrial purposes, it was operated in such a way that its repair was carried out after the failure of any of its components (Muthanandan, 2019). This produced chain breakdowns that caused additional costs, unnecessary increases in repairs' complexity and permanent failures in components which could no longer be used (Kruczek, 2019). This type of maintenance is known as corrective maintenance and today is still used for some high complexity equipment in which it is not possible to monitor each and every one of its components (Castillo, 2014). However, in most cases, the upkeep of the equipment has evolved enormouslyand is currently based on failure prediction to try to minimize their impact on the system (Fernandes, 2020). Nowadays, there are several types of maintenance with very different functions and characteristics. In this article, reference will be made, in particular, to preventive maintenance, which can be defined as a set of scheduled activities, such as tests, inspections and repairs, which aim to mitigate the impact of the failures produced on a certain piece of equipment (Muñoz, 2015).

In order to explain in detail the methodology and the process followed to carry out a preventive maintenance plan, the data analysis of a company's machinery will be accomplished. Specifically, the study will focus on "Maderas S.A.", a fictitious Spanish company devoted to the manufacturing of wood pieces and interested in the implementation of a maintenance system that improves the efficiency of its lathes. Nowadays, data are going through a period of revolution driven, not only by their abundance, but also by the continuous development of new technology based on their analysis, treatment and transformation (Emovon, 2018). This article presents a practical example of the power of data and the great support they provide for problem solving.

They have infinite applications and are one of the fundamental pillars for companies' productive development (Business Software Alliance, 2015).

2. Theory

Goods and services received by customers go through a productive cycle that consists of different phases. During the operation stage, the system can be subject to failures that interrupt the productive activity and obstruct the proper manufacturing of goods. To prevent these issues, equipment maintenance is used, which can be defined as the set of activities applicable to goods and services that aim to prevent breakdowns, as well as ensure the regular operation and good condition of the machinery (Cuartas, 2008).

The maintenance tasks are applicable to any type of good susceptible to failing during its period of activity, so that they are focused on the preservation of both equipment and machinery, as well as industrial buildings, among other examples.

The objectives pursued by the maintenance can be synthesized in several points:

- To avoid and reduce the breaks, proceeding to their repair in the event of this taking place.
- Reduce the impact of failures and extend the useful life of the assets.
- Avoid accidents and ensure people's safety.
- Reduce production costs.

There are several types of maintenance, briefly mentioned in the introduction, that are grouped according to their different characteristics and applications. These are: corrective maintenance, total productive maintenance, predictive maintenance and preventive maintenance (Muñoz, 2015).

The previously mentioned corrective maintenance represents the set of activities that involve repair and replacement of deteriorated components, once the failure has occurred. It has obvious drawbacks such as the impossibility of knowing the moment of failure, which involves the risk of occurring at a critical moment in the production process, or chain faults, previously described in the introduction, that cause errors in other components as a consequence of the first fault.

The total productive maintenance is based on the permanent approach towards the improvement of the efficiency, through the full implication of all the people who participate in the productive process. The role of the maintenance department is assumed by the entire company, contributing together to maintenance optimization.

Predictive maintenance represents the set of monitoring activities of a given system, which allow corrections to be made and interventions in goods when a fault symptom is detected. It is characterized by its great complexity and difficult application, since it implies the monitoring of the whole data of all the components that conform a determined system. From a given set of historical data and the monitoring of the different phases that lead to the failure of each of the components, a causal relationship can be developed between small deviations in the performance of the element and subsequent failure of the same, allowing the intervention and repair of the fault before it takes place.

Finally, preventive maintenance, in which the present article focuses, refers to maintenance scheduling based on studies with the aim of mitigating the impact of errors. Good performance of this maintenance can be compared to a healthy habit in terms of exercise and nutrition. It does not represent a health guarantee but increases its probability. Its use gives many advantages to the organization, some of them being mentioned in the following points:

- Increase in the security of all the elements involved in the production cycle.
- Reduction of downtime, with the consequent increase in the availability of the equipment.
- Reduction of costs.
- Decrease in the number and scale of the repairs.
- Undoubtedly increase in the useful life of the equipment.
- Reduction of exposure to risks.

In spite of all the advantages listed, their use involves the risk of making unnecessary changes in components whose use could be extended over time, as well as the increase in inventories and staff available (García, 2012).

Through a correct treatment and interpretation of the data by using the appropriate tools, it is intended to generate a repair schedule that optimizes the useful life of the machinery used by the company "Maderas S.A.". The development of the procedure is explained in the following sections (Levitt, 2011).

3. Obtaining Data

The determination of the period between repairs of the different machines in order to properly characterize and implement the preventive maintenance program can be carried out in several ways, and we have chosen for the present case the analysis of the equipment’s average life. The operating period of the different lathes of the company varies according to two parameters: the machine’s supplier and its maintenance team. The company has 4 different suppliers, who will be named using numbers 1 to 4; and 3 maintenance teams, which will be named with the letters A, B and C. Depending on the assigned parameters, the machines will have different values of their average life, which will need to be reasonably estimated in order to program the periods between maintenances.

“Maderas S.A.” has a total of 300 lathes. With the objective of obtaining reliable data of the machines’ lifetime and to be able to perform its statistical treatment to reach logical conclusions, an inspection of each of the machine’s condition is realized, determining the period elapsed since their last repair, as well as if they are in operation or in a fault state. This information is collected in the following table:

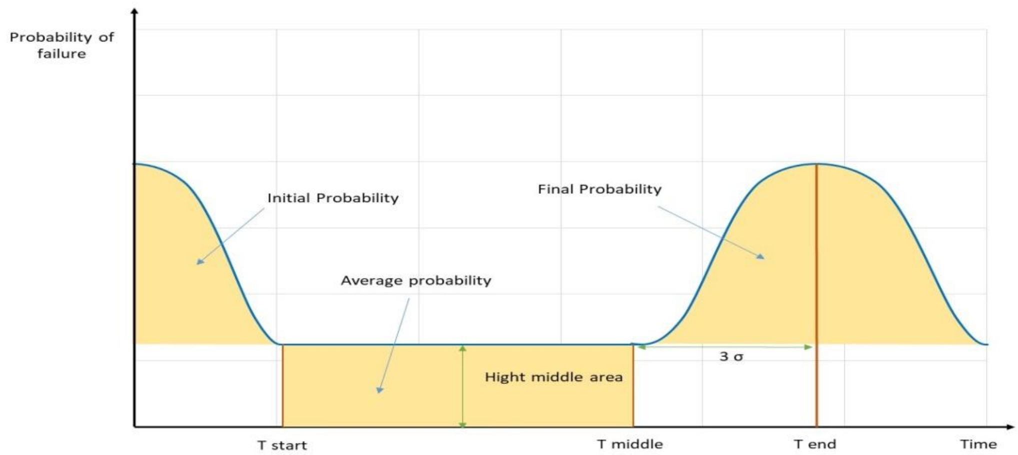
Table 1. Data of 10 o fthe 300 lathes

Lifetime	**Broken**	**Team**	**Provider**
56	0	TeamA	Provider4
81	1	TeamC	Provider4
60	0	TeamA	Provider1
86	1	TeamC	Provider2
34	0	TeamB	Provider1
30	0	TeamA	Provider1
68	0	TeamB	Provider2
65	1	TeamB	Provider3
23	0	TeamB	Provider2
81	1	TeamC	Provider4

The most useful information is offered by the machines that are already broken when the inspection takes place. Since the period of time from the last repair to the fault coincides with their total useful life the total time they are able to operate without failing can be known. From the complete data set of the 300 machines, which is not included in this article, it can be seen that the equipment purchased from supplier 1 has an average useful life that ranges from 85 to 93 days, depending on the assigned maintenance team. The machines belonging to the second supplier last between 85 and 93 days, those of the third between 60 and 65, while those obtained from the fourth provider vary between 81 and 88 days of average lifetime.

This information, although valuable, is not sufficiently representative since most of the lathes are in correct working condition at the moment of the inspection and the lathes' period of operation does not provide conclusive results concerning average lifetime. In addition, it should be mentioned that industrial equipment obeys, in most cases, a failure curve that is known as the bathtub curve. It illustrates the probability of a certain piece of equipment failing during its lifetime. During the early life of the component, called the infant mortality zone, the fault probability begins with a high value and decreases as time goes by, reflecting the possibility of the component not having been produced or repaired correctly. After this initial zone, a phase with a constant and much smaller failure probability is presented, called random zone. Finally, when approaching the end of its useful life, the breakdown probability increases drastically until it occurs, identifying this period with the name of wear zone.

In the few data acquired from the company's lathes, the running time of the machines was grouped only in the wear zone, given the low probability of failure in the infant mortality period compared with thefinal zone of its useful life. This means that the set of data obtained does not represent the failure of the machines in a realistic way, besides being insufficient to be able to take decisions and to draw satisfactory conclusions. Therefore, an interesting alternative to generate a larger volume of data is simulation. It is done through the Python program, and it allows us to observe the behavior of the machines during a year or even more time, while the real simulation time takes a few milliseconds (Rossum, 2009). The code sheet is not included in the article, since it is not necessary for the understanding of the programming because it has been carried out using a considerably simple method, detailed below.



Ilustration 1. Designed failure curve

The graph represents the failure curve of the lathes using normal curves to facilitate the assignment of the failure probabilities. The highlighted time instants have been determined from the original data for each of the parameter combinations, dividing the different zones of the fault curve and allowing the program to assign probabilities depending on the phase and the specific constant of time. To simulate the failures, a procedure that constitutes an application of the Monte Carlo Method has been employed, assigning discrete probabilities to each one of the time intervals, which in this case are days. In each of these days, a random number between 0 and 1 is launched, and if it coincides with the interval assigned to the probability of failure, the machine fails, whereas if it does not, it is passed to the next iteration and the loop is repeated until the end of the simulation period that has been set in 365 days.

Applying the above procedure, since the machines must be repaired several times throughout the year, enough data of the useful life can be obtained for each one of the pieces of equipment, that, collected in a table are shown in the following way:

Table 2. Simulated data of the first 2 machines

Machine number	**Maintenance equipment**	**Supplier**	**Time (days)**	**Faults**
0	C	3	61	S í
0	C	3	61	S í
0	C	3	63	S í
0	C	3	63	S í
0	C	3	60	S í
0	C	3	57	No
1	B	3	65	S í
1	B	3	66	S í
1	B	3	69	S í
1	B	3	67	S í
1	B	3	65	S í
1	B	3	33	No

Only 6 different data are registered for the first machine. The first five indicate data on their useful life, while the latter reflects the state of the machine at the end of the simulation, which is useful for subsequent estimations of the remaining life expectancy.

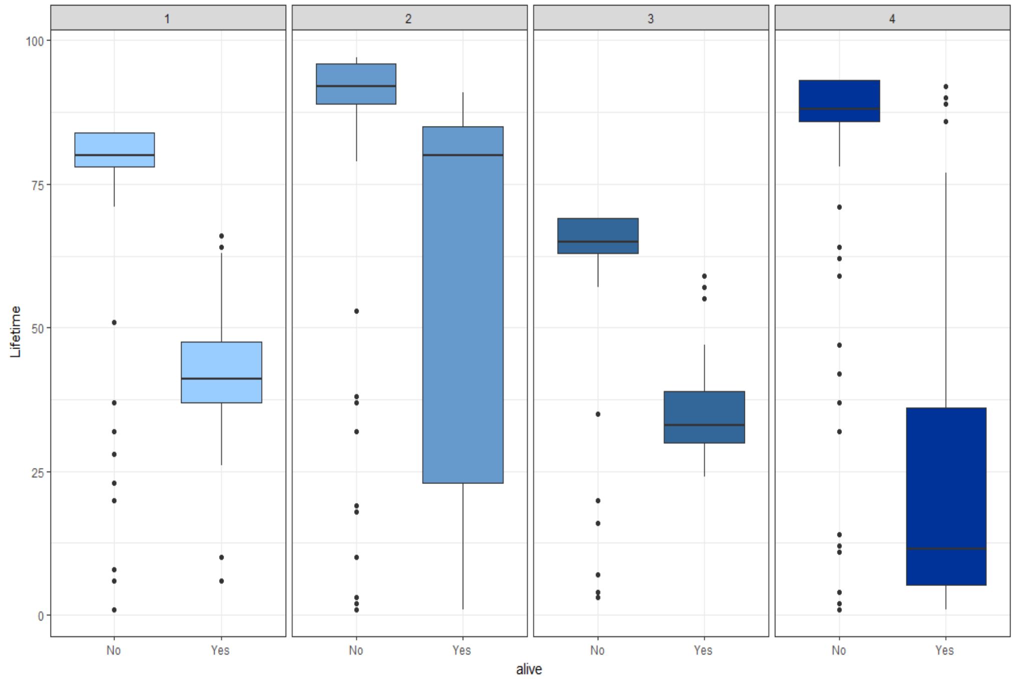
With the information obtained from the simulation, a data analysis which allows us to deduce the best combination of intervals between repairs for the lathes of the company can be performed.

4. Data Analysis

To perform the statistical analysis, each stored article of data is taken as independent so that the machine number to which they belong has little relevance, the only relevant one being the combination of parameters that are shown (supplier and maintenance equipment). In this way the data are classified into 12 groups corresponding to all possible combinations between the 3 suppliers and the 4 maintenance teams, which should be analyzed independently for the possibility of requiring different intervals between repairs.

The "R" programming environment is used for the statistical treatment of the simulated data set (Ram ŕez, 2013). The code used is not included but throughout the article the diagrams and representations derived from its execution, which provide very useful information on the distribution of the data over time, will be introduced.

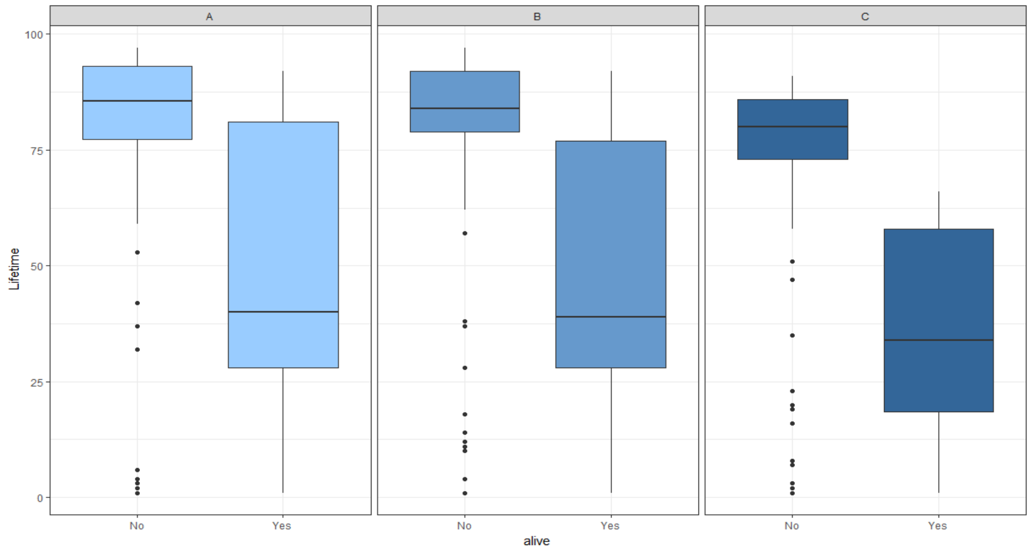
To obtain a concrete idea of the situation of the data according to the supplier, a box diagram is elaborated. When operated with all data, the final states of the machines are also included at the end of the simulation time and do not provide the same information as the failure states, so that they separate from each other in the graph offered.



Ilustration 2. Box diagram according to supplier

In all cases, the lifetime values are concentrated around a small time interval, which can be deduced from the small box width in the diagrams of the four suppliers. This is because most of the failure probability accumulates in the wear zone, which results in a logical pooling of data in this interval. This is a realistic situation as unexpected failures in the youth and random mortality zone occur very infrequently. In addition, it is clearly reflected that the simulation has faithfully respected the failure curve designed through the existence of scattered points outside the boundaries of the box, which correspond to atypical values corresponding to the random stages and infant mortality. Analyzing these atypical failure states, we observe the greater accumulation of data in the first moments of the operation of the equipment, due to the greater relative probability of occurrence of the fault in this zone.

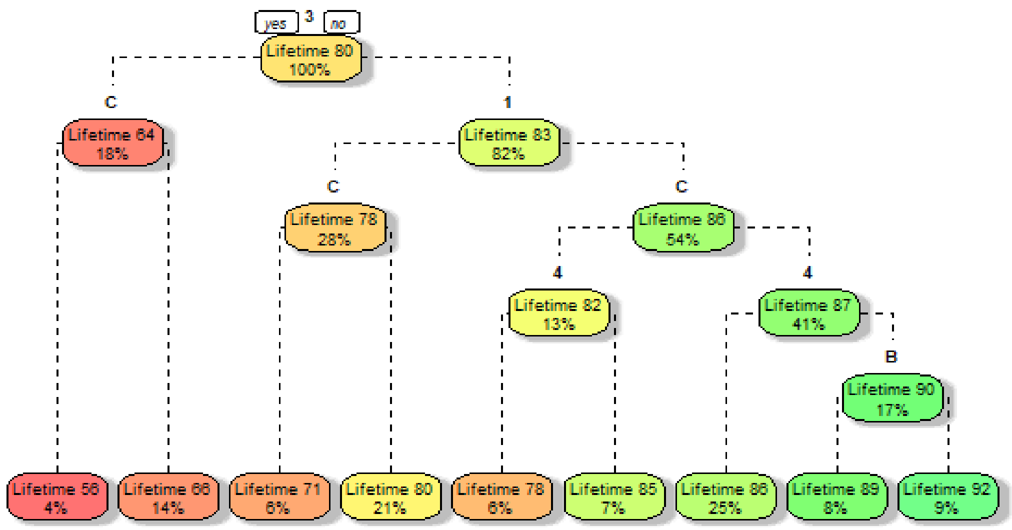
When comparing the graphs of the four vendors, it should be highlighted that the machines purchased from the vendor 3 have a considerably shorter service life than the rest, while on the other hand, vendor 2 manufactures the winches with the longest running time. This confirms the existence of clear differences in the operation of the machinery on the basis of the parameters analyzed.



Ilustration 3. Box diagram according to maintenance team

This new diagram clusters the values according to the maintenance teams, observing much more similar records than in the previous case of the suppliers. Even so, the shorter service life of the machinery supervised by the maintenance team is emphasized.

In order to correctly classify all combinations of parameters simultaneously, and not in isolation as in the previous cases, a tree diagram which divides the data into successive groups according to the half-life, assigning percentages to each of the divisions is made.



Ilustration 4. Tree diagram

Starting from the upper part of the graph, the average life of the machines, which has a value of 80 days, appears first. Each element of the diagram has associated a percentage, which represents the fraction over the total of that half-life. The branches are linked to certain numbers and symbols that define and limit the parameters of the element until a particular group of lathes is reached at the lowest level.

Analyzing in detail the previous representation, it is observed that the machines with a lower half-life are those obtained from supplier number 3 and maintenance equipment C, with an average of 56 days of operation, whereas the ones with the longest operating time are those acquired from the supplier number 2 and maintained by the equipment A, with an average interval between failures of 92 days. The large difference between the two groups mentioned above is observed, which confirms the high influence of the parameters on the values recorded for half-life.

Since the tree diagram has been able to acquire a very specific idea about the distribution of the data of the different groups of parameters, it is possible to find the life expectancy of the machines from their final state of simulation and the average of its operating time with some margin of error.

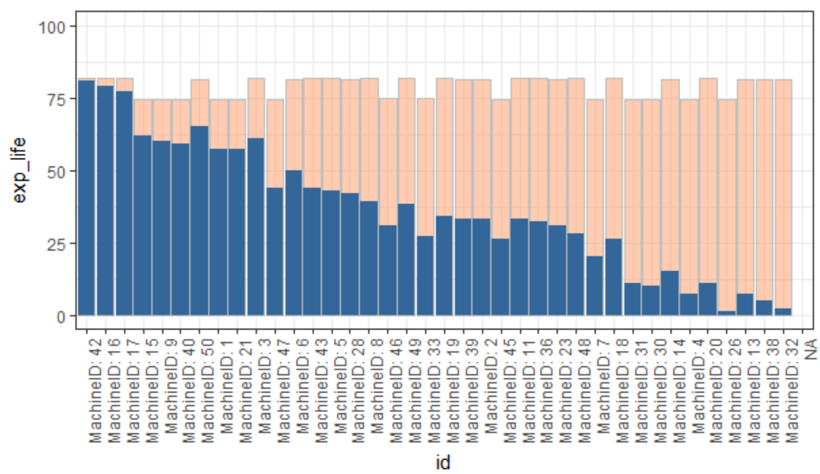


Illustration 5. Machine's life expectancy

This type of graphics is really representative and offers the company's maintenance department a great deal of information about machines that require immediate attention due to the existence of a high risk of a failure.

Finally, in order to understand the extent to which the different parameters affect the results obtained, it is wise to elaborate a linear regression that results in the coefficients with which the different parameters influence the final value of the half-life.

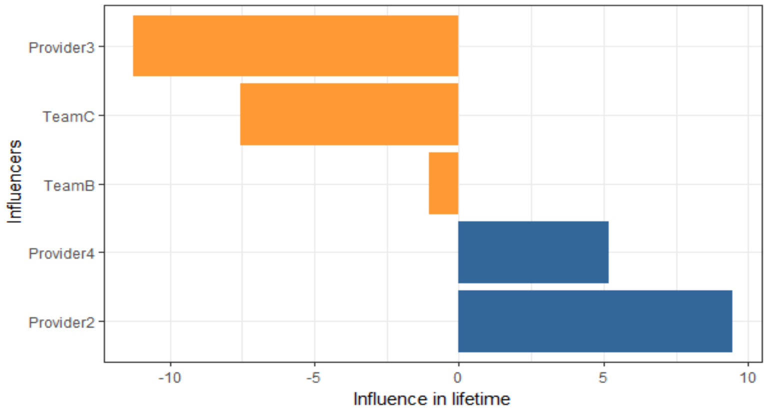


Illustration 6. Influence of the parameters

It should be highlighted that the coefficient of influence is very negative in the cases of the supplier 3 and the maintenance equipment C, so that the machines whose parameters match those mentioned will be the ones with the shortest time between failures, as previously reflected in the tree diagram. It is important to highlight the increase in the results produced by the supplier 2, whose machines must be taken as an example for the selection of future suppliers for their long duration and stability in operation.

5. Results and Conclusions

Once the data have been analyzed, the preventive maintenance planning of "MADERAS S.A." lathes should be organised. From the statistical data of the life of the machines it is possible to estimate a time between repairs to prevent the machines failing and guarantee the good state of its components.

However, before carrying out these procedures, it is necessary to assess whether the company should continue to rely on the same suppliers and maintenance teams. In view of the information obtained through the above diagrams, it can be determined that two of the parameters considerably decrease the equipment's operating time. This causes more frequent revisions with their corresponding increase in costs. Supplier 3 sells winches that fail earlier than other vendors, which only creates detrimental effects for the purchasing company. In this way, change of supplier is a good alternative to raise the overall average life of machines. As in the case of suppliers, maintenance equipment C has a lower performance than other equipment under the same operating conditions, damaging the productive activity of the company. Therefore, it is necessary to advise the recruiting of a new maintenance team with a better training and skills.

After discarding both supplier 3 and maintenance equipment C, the operating time to the fault must be predicted to correctly schedule the preventive maintenance. The programming language "R" used for the statistical treatment allows us to obtain information relative to the machines belonging to each combination of parameters. The statistical data provided, for example, for the machines of the supplier 1 and maintenance equipment A, are presented in the following table:

Table 2. Statistical results

Supplier 1 and maintenance team A					
Minimum	1st Quartile	Median	Average	3rd Quartile	Maximum
1	78	80	71,47	83,50	84

The mean is clearly affected by failures in infant mortality zones, which must be taken as atypical data, so that the data in the box diagram, such as quartiles and median, can be a good alternative to calculate the period between maintenances. The data are very compressed between the first quartile and the maximum number, so that 75% of the data is located in an interval of 6 days. From the model built for these machines, the start time of the wear zone, at which point it increases the possibility of failure, is 76 days, and it is expected that most of the data will be located between this number and the maximum, except atypical data of premature mortality. Therefore, to establish as a time between repairs a number sufficiently below the first quartile so that all the values of the wear zone are situated outside the range, but not in excess to avoid wastage of useful life, is presented as the opttimal alternative for the choice of interval. Taking 95% of the first quartile, i. e 74 days, an acceptable value is reached so that the systems of supplier 1 and maintenance equipment A only fail in exceptional cases. For machines from the same supplier and maintenance equipment B, the result is the same, since the maintenance equipments A and B can be taken as equal by their statistical similarity.

For the machines of vendor 2 and maintenance equipment A and B, the chosen result is 84 days following the same procedure as in the previous case.

Finally, for the machines of supplier 4 and the same maintenance equipment the result will be of 82 days.

Once the periods between repairs are obtained, the maintenance department can have an efficient schedule to carry out all the activities associated with the maintenance of equipment throughout the future.

It is possible to show through the development of this case study an example of the great value of data processing. The ability to process data and transform observations into knowledge, is the one that allows solutions to be obtained of real meaning to the considerable challenges of the present (Business Software Alliance, 2015).

References

Business Software Alliance. (2015). *¿Por qué son tan importantes los datos?*, Washington: s.n. Retrieved from https://data.bsa.org/wp-content/uploads/2015/10/BSADataStudy_es.pdf

Castillo, Y. A. (2014). *El Mantenimiento Industrial.* Retrieved from http://www.monografias.com/trabajos101/el-mantenimiento-industrial/el-mantenimiento-industrial.shtml

Cuartas, P., & Luis, A. (2008). Retrieved from https://es.scribd.com/document/185618755/MANTENIMIENTO-MECANICO

Emovon, I., Norman, R. A., & Murphy, A. J. (2018). *J Intell Manuf*, *29*, 519. Retrieved from https://doi.org/10.1007/s10845-015-1133-6

Fernandes, M., Canito, A., Corchado, J. M., & Marreiros, G. (2020). Fault Detection Mechanism of a Predictive Maintenance System Based on Autoregressive Integrated Moving Average Models. In: Herrera F., Matsui K., Rodríguez-González S. (eds) Distributed Computing and Artificial Intelligence, 16th International Conference. DCAI 2019. *Advances in Intelligent Systems and Computing*, *1003*. Springer, Cham. Retrieved from https://doi.org/10.1007/978-3-030-23887-2_20

García, G. S. (2012). *Mantenimiento Petroquimica.* Retrieved from http://www.mantenimientopetroquimica.com/tiposdemantenimiento.html

Kruczek, P., Gomolla, N., Hebda-Sobkowicz, J., Michalak, A., Śliwiński, P., Wodecki, J., ... Zimroz, R. (2019). Predictive Maintenance of Mining Machines Using Advanced Data Analysis System Based on the Cloud Technology. In: Widzyk-Capehart E., Hekmat A., Singhal R. (eds) Proceedings of the 27th *International Symposium on Mine Planning and Equipment Selection* - MPES 2018. Springer, Cham. Retrieved from https://doi.org/10.1007/978-3-319-99220-4_38

Levitt, J. (2011). *Complete Guide to Preventive and Predictive Maintenance.* s.l.:Industrial Press.

Muñoz, A., & María, B. (2015). Retrieved from http://ocw.uc3m.es/ingenieria-mecanica/teoria-de-maquinas/lecturas/MantenimientoIndustrial.pdf

Muthanandan, S., & Nor, K. A. B. M. (2019). Condition Monitoring and Assessment for Rotating Machinery. In: Sulaiman S. (eds) Rotating Machineries. *SpringerBriefs in Applied Sciences and Technology*. Springer, Singapore. Retrieved from https://doi.org/10.1007/978-981-13-2357-7_1

Ramírez Pérez de Inestrosa, Javier. (2013). *Universidad de Granada.* Retrieved from http://www.ugr.es/~javierrp/master_files/Tutorial%20de%20R.pdf

Rivas, A., Fraile, J. M., Chamoso, P., González-Briones, A., Sittón, I., & Corchado, J. M. (2020). A Predictive Maintenance Model Using Recurrent Neural Networks. In: Martínez Álvarez F., Troncoso Lora A., Sáez Muñoz J., Quintián H., Corchado E. (eds) 14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019). *Advances in Intelligent Systems and Computing*, *950*, Springer, Cham. Retrieved from https://doi.org/10.1007/978-3-030-20055-8_25

Rossum, G. V. (2009). *Python.* Retrieved from http://docs.python.org.ar/tutorial/pdfs/TutorialPython2.pdf