

IFC-Bank of Italy Workshop on “Data Science in Central Banking: Applications and tools”

14-17 February 2022

Spot the flaw – using power BI for quality control: an application to non-financial corporations’ data¹

José Alexandre Neves, Tiago Pinho Pereira and Ana Bárbara Pinto,
Banco de Portugal

¹ This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

Spot the flaw – Using Power BI for quality control: an application to non-financial corporations' data

Ana Bárbara Pinto | Banco de Portugal

José Alexandre Neves | Banco de Portugal

Tiago Pinho Pereira | Banco de Portugal

Abstract

This paper shows the experience of the Central Balance Sheet Data Office (CBSO) of Banco de Portugal with Power BI dashboards, which represent a powerful tool to perform and monitor the quality control of data. CBSO databases contain quarterly data from the balance sheet and the income statement of firms and yearly data from all the financial statements and notes. Linking Power BI to CBSO databases allows to execute deep analyses and instantly through the dashboards.

As a powerful data visualisation tool, Power BI gives the opportunity to get information directly from the charts and drilling down data. This is of utmost importance in terms of quality control as it allows to quickly identify the situations that need to be tackled urgently to improve quality in non-financial corporations' statistics.

Keywords: firm-level databases; non-financial corporations; quality control; data visualisation; Power BI

JEL classification: C81; C88

Acknowledgements: We would like to thank Paula Casimiro and Nuno Azevedo for their valuable suggestions and comments, Duarte Santos for providing information and training on Power BI and Ana Teresa Oliveira and Joana Gonçalves for developing some of the dashboards presented in this article. The views expressed are those of the authors and do not reflect those of the Banco de Portugal or the Eurosystem.

Contents

1. Introduction.....	3
2. Power BI for quality control purposes at the CBSO of Banco de Portugal	3
3. Conclusion	13
References	14

1. Introduction

We are in the golden era of data visualisation. Translating data into meaningful and appealing charts is essential to not only understand and communicate statistics better, but also to manage organisations.

Non-financial corporations' data available at Banco de Portugal come from two different main sources. The annual survey (IES, in the Portuguese acronym, standing for Informação Empresarial Simplificada), which is mandatory and covers all the non-financial corporations (NFCs) in Portugal, and the quarterly survey to NFCs (ITENF, in the Portuguese acronym, standing for Inquérito Trimestral às Empresas Não Financeiras), which is a sample-based survey. ITENF only contains a summarised balance sheet and income statement, while IES provides complete financial statements comprising the balance sheet, the income statement, the cash flow statement, the statement of changes in equity and detailed quantitative and qualitative data presented as notes to financial statements. Once at Banco de Portugal, data is subject to automatic and manual quality control procedures.

To perform and measure the quality control at the Central Balance Sheet Data Office (CBSO) and assess the quality control that remains to be done at any given moment, several Power BI dashboards were created. Setting up the data connections is easy, refreshing them is not demanding and it is possible to drill down the data by clicking in the visual and quickly check which firms are responsible for the result that is being shown, which is fundamental to boost and redefine the target of the manual validation, if it is necessary. Furthermore, because it is a visualisation tool, we can share the dashboards and deliver a presentation of the most relevant issues with no additional effort. In fact, Power BI dashboards allow performing, measuring and managing quality control and this is key.

The remainder of this work is as follows: Section 2 briefly describes the quality control of data received by the CBSO of Banco de Portugal and shows how Power BI is used for the quality control of CBSO data. Section 3 concludes.

2. Power BI for quality control purposes at the CBSO of Banco de Portugal

The Central Balance Sheet Database (CBSD) of Banco de Portugal is an economic and financial database on Portuguese NFCs. Its information is mostly based on annual and quarterly accounting data.

Annual data is obtained from the annual accounts of corporations reported under IES (Simplified Corporate Information – SCI in the English acronym), while quarterly data is drawn from the quarterly survey to NFCs (ITENF).

IES is a single electronic submission of accounting and fiscal information that meets the statistical needs of Banco de Portugal and the National Statistics Institute and is used by the Portuguese Ministries of Finance and Justice for fiscal purposes and legal deposit of accounts (Banco de Portugal, 2008). It comprises not only the detailed financial statements (balance sheet, income statement, statement of changes in equity and cash flow statement), but also the notes to them. It is mandatory and covers virtually all companies in Portugal (more than 450 000). ITENF, on the other hand, only includes a sample of NFCs (more than 4 500) and it covers a summary of the balance sheet and income statement.

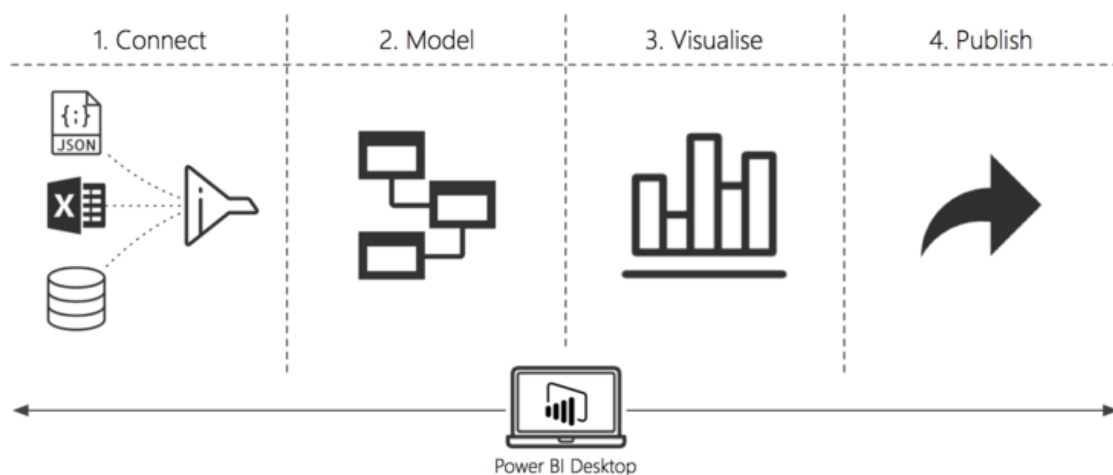
Both IES and ITENF are subject to quality control at the CBSO and Power BI is a key tool to perform and monitor quality control, given that it can rapidly show what is already accomplished, what has to be done and what are the priorities until the end of the process.

For each firm, quality control is performed to ensure that information is horizontally (across different years) and vertically (within financial statements) consistent¹. During quality control, data is usually compared with other internal and external data sources (Casimiro et al., 2017), and it is also possible to query NFCs for additional explanations. Automatic validation procedures are implemented and, for quality control purposes, only a sample of firms from IES and ITENF that violate the consistency checks implemented² are selected for manual validation.

Data from ITENF is received within 2/3 months after the end of the reference quarter, while IES is due 6.5 months after the end of the fiscal year, which corresponds to July 15 for the majority of companies resident in Portugal. Usually, around 700 NFCs are selected every quarter for manual quality control of ITENF and the CBSO staff validates all of them. Regarding IES, around 4 000 to 6 000 NFCs are selected for manual quality control. To validate them, the CBSO hires a group of undergraduate students³ from July to September, given that the period in which IES is available for quality control matches the students' summer break. It is a win-win situation as students develop their financial analysis skills and the CBSO assures the validation of a higher number of firms in a limited timeframe, that otherwise would not be possible.

Power BI allows connecting to data, transforming and cleaning them, as well as creating interactive and rich dashboards and reports. Premium versions of Power BI also enable publishing the dashboards (in websites, for example) and scheduling automatic updates (Figure 1). In the case of the CBSO of Banco de Portugal, Power BI dashboards are published in the integrated information system of the CBSO, which is an internal website that allows the analysis of individual firms' data and the insertion of any manual correction that is deemed necessary. Everyone working at Banco de Portugal without a Power BI software, but with an internet connection can access the dashboards, which is an advantage.

Figure 1: The life cycle of a dashboard created through Power BI



Power BI is a product from Microsoft and, thus, resembles Microsoft Excel very much. Reports may have several sheets and the way variables are selected to build the visuals is similar to the way pivot tables are built. You can add filters and switch from rows to columns to customise the visual.

Figure 2 displays a finished dashboard, designed for the quality control of annual data. Above, one can pick the year, size, institutional sector and economic activity of the firms and the charts will update accordingly, showing the validation and reporting status of the firms, with light green being the firms

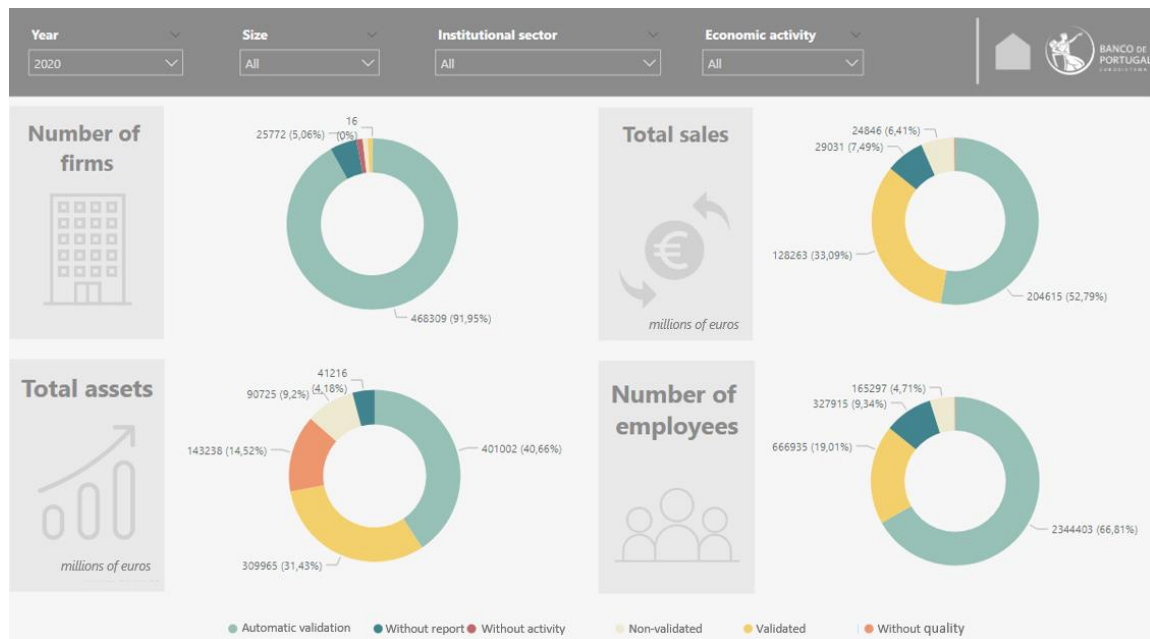
¹ For more details regarding the quality control that is performed at the CBSO of Banco de Portugal, please see Brites (2013)

² Some of the consistency checks implemented include absolute and relative annual/quarterly changes above a given threshold and differences between data submitted by NFCs and internal and/or external data sources above a given threshold.

³ In the area of economics, management, finance and accounting

that have been validated automatically, yellow the firms validated manually, beige the firms awaiting validation and dark green the firms that haven't reported yet. By changing the size and economic activity one can assess the firms that are manually validated or waiting validation for a subset of the universe, which constitutes relevant information for quality control. Power BI drilldown features also allow us to see the most relevant companies without report, which is useful so we can contact them in order to submit the missing information, or the most relevant companies that are awaiting validation in order to give them priority, if we are close to our deadlines and need to check more relevant firms first.

Figure 2: Status of validation and reporting of firms by number, total sales, total assets and number of employees



When one needs to focus in a certain timeframe, the range slicer, which is represented in Figure 3, referring to the number of manually validated firms is a very helpful visual, too. One can choose a specific year and, in addition, move the Days' range slicer to concentrate the analysis in the first days of the quality control, for example. This is particularly useful given that the analysts could be interested in a specific period instead of the whole period. At the beginning of the quality control process, the pace is slower, increasing over time as undergraduate students become more experienced and autonomous. It is always very important to bear this pace in mind to assess the level of manual validation that we can achieve at our deadlines and if we need to change the analysis strategy.

Figure 3: Number of validated firms

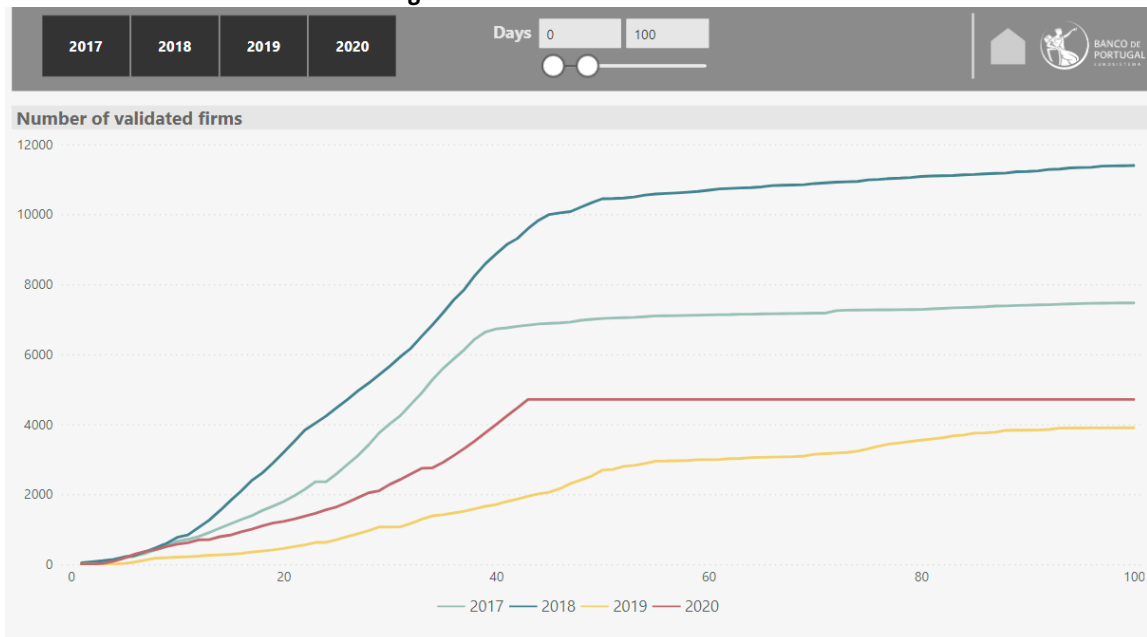


Figure 4 provides an overview of the validation of the quarterly survey of 2021Q1. This picture was taken before the ending of the quality control, which explains the decreases observed in the charts in the last period (2021Q1). The decrease is the first impression one has when looking to the dashboard and this is why visualisation tools like Power BI are very good to quickly identify the state of art at any given moment. The upper three charts show the difference between the total sample of ITENF (in blue) and the answers actually received from firms (in white), while the lower three charts focus on the manual validation of the firms that are selected through the consistency checks. The dashboard clearly shows that the validation level was still below in comparison to the previous quarters and that information is useful to give an idea of how far away of the goals one is and assign more resources to the task, if needed.

Figure 4: Overview of the quarterly survey validation

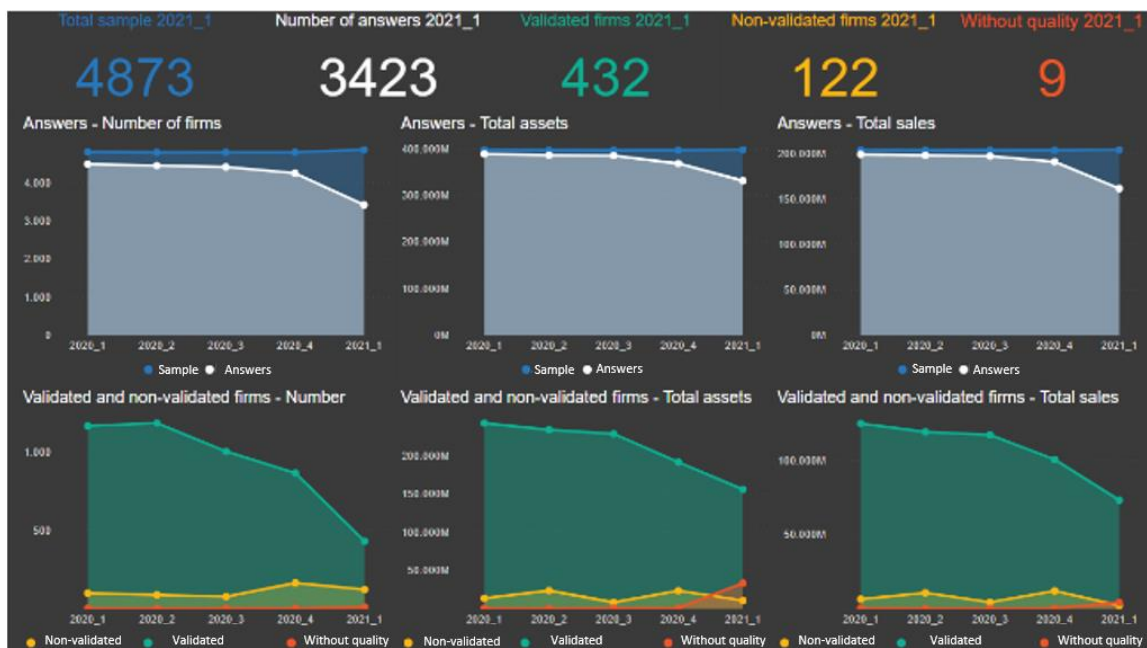


Figure 5 shows the percentage of validated firms by consistency check (in green) for the quarterly survey of 2021Q1. In yellow, it is displayed the portion of firms not validated yet. At the beginning of the quality control, the yellow portion of the bar prevails, of course, but the evolution of the green portion gives important information for monitoring the overall validation status. In order to meet the expected deadlines, this information is crucial to management. If the yellow portion is significant enough close to the deadlines, it is time to assign more resources to the task. Each bar represents a consistency check. As stressed before, the consistency checks refer to changes regarding the previous periods above a given threshold or to differences relative to other data sources above a given threshold, but also to coherency issues across the survey.

For example, if NFCs report meaningful sales or purchases to non-resident counterparts, it is expected that they will have trade credits respecting non-resident counterparts, unless we have a note indicating that the company actually does not have trade credits with non-resident entities. If it is not the case and firms do not report any value in trade credits neither there is a note explaining that, they violate the consistency check *External Trade Credits* and consequently are selected for manual validation. Another example of coherency issues across the survey is consistency check *Cost of Debt*. If a firm reports significant values in loans, it is very likely that it pays interest. If not, it is something to check.

Besides presenting the portion of validated and non-validated firms, Power BI is flexible enough to drill down data and show which NFCs are already validated (and those whose validation is missing). To see more details on the firms not validated yet (e.g., which ones and the username of the person in charge of the validation), it is only needed to click in the right button of the mouse above the yellow part of the bar and choose the option "Show data point as a table". This is valid for every dashboard in Power BI. It is always possible to click in the charts and access the firms that are contributing to that chart.

Figure 5: Validated firms by consistency check - quarterly survey



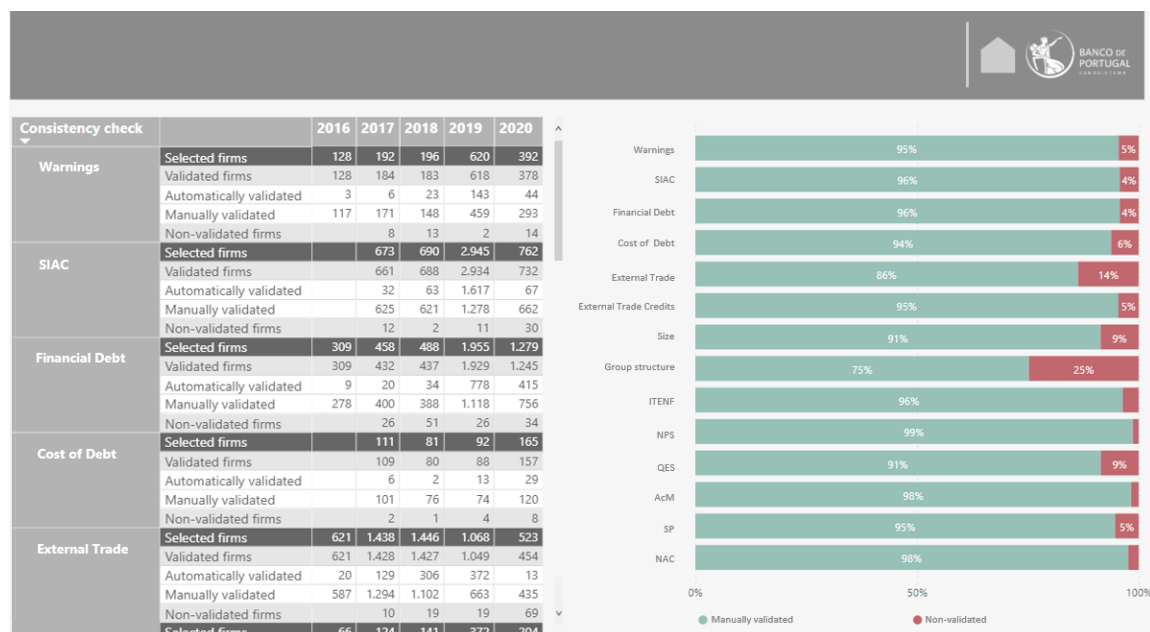
The quality control framework of the annual survey (IES) is similar to the framework of the quarterly survey (ITENF). However, given that the data collected by IES is more comprehensive, the timeframe to perform the quality control is larger and there are more human resources available, namely undergraduate students, as mentioned earlier, the number of validated companies is higher and the consistency checks to analyse are broader. In addition, because IES is submitted after the firms' annual accounts are approved and is available for the universe of Portuguese NFCs, it also contributes to

complete and correct data from ITENF and is essential to depict the performance of the NFCs sector as a whole.

Figure 6 shows the fraction of firms by consistency check already validated. The template is similar to the quarterly survey and, although the number of consistency checks is approximately the same as in Figure 5, their scope is broader. For example, in the quarterly survey, changes in the size class or in the group structure of the firms are not evaluated, while in the annual survey they are (consistency checks *Size* and *Group structure*, respectively).

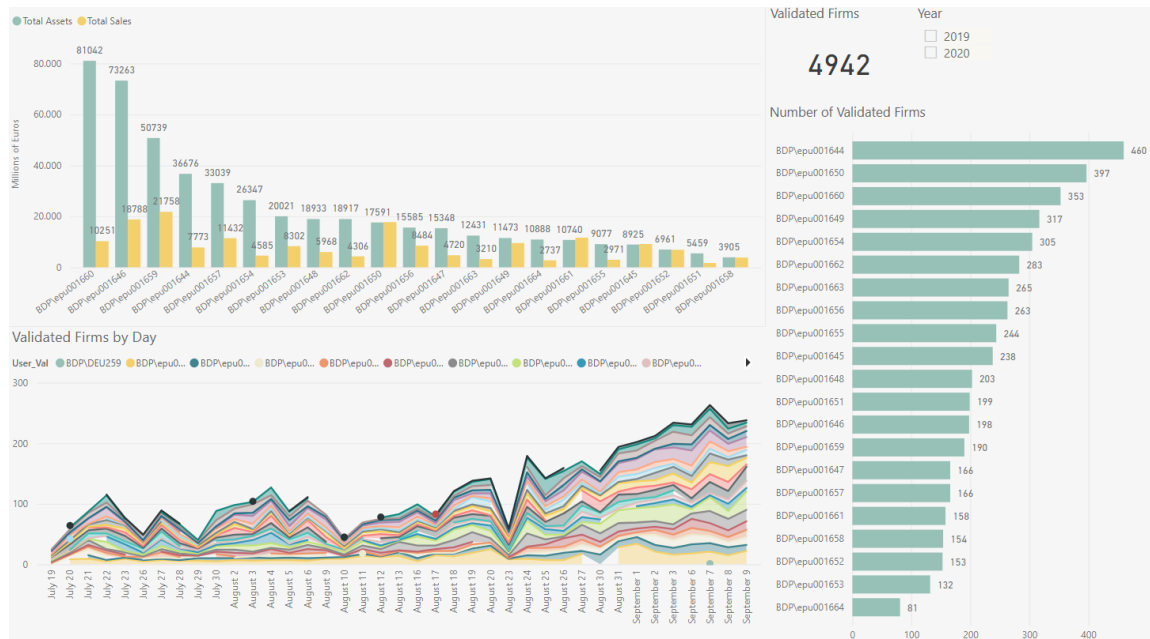
Again, it is possible to drill down to the firms that are behind each one of the consistency checks. This turns easier and faster to implement any change of priority that is needed during the quality control process.

Figure 6: Validated firms by consistency check – annual survey



The number of validated firms by user is shown in Figure 7 and gives very important information to management. This dashboard presents the total number of validated firms by day and user and it enables assessing whether performance is improving or not with time. It is expected that the pace of validation by the undergraduate students increases over time, as they gain more experience and knowledge and become more autonomous. One can see that the number of validated firms by day reaches a maximum in the first days of September, one and half months after the beginning of their work. Decreases in the number of validated firms by day are usually related to the assignment of more (or more complex) firms for validation. When a bar of a given user is selected, only information for that user is shown. With this dashboard we can early detect some issue with any student, and thus provide additional individual training in a timely manner in order to improve her/his contribution, as well as to follow closely the daily pace of validation as a whole.

Figure 7: Validated firms by user and details – annual survey

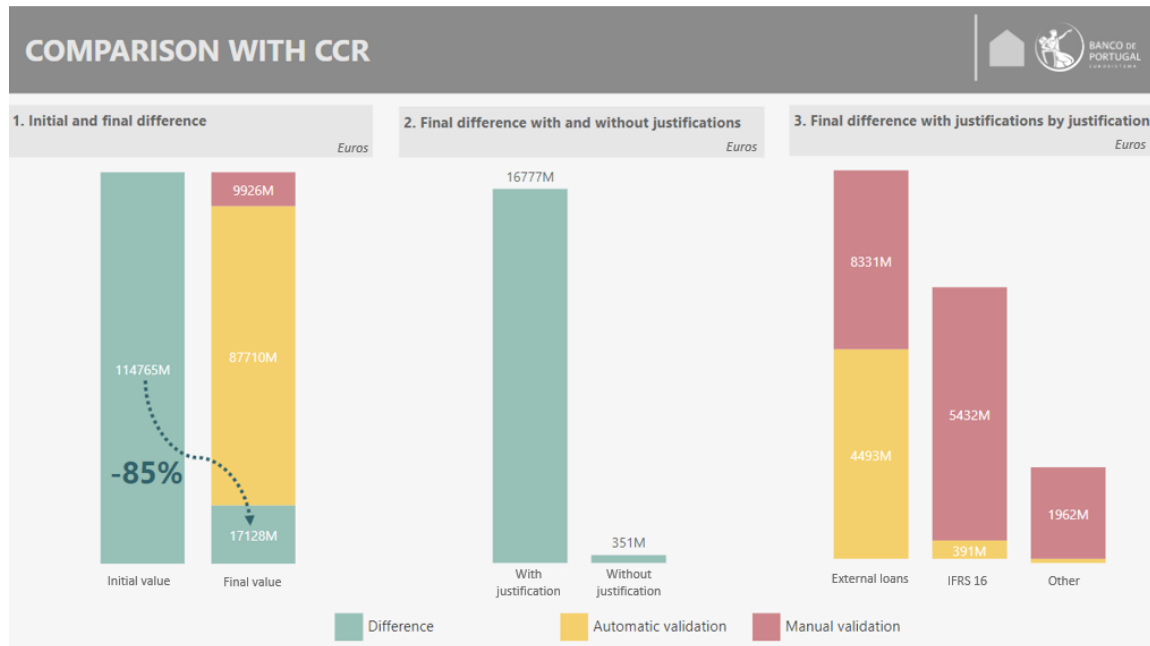


Comparison with other sources of information besides IES or ITENF is also an important part of quality control procedures. Of course, information can differ between data sources, but it should be consistent.

Figure 8 shows the difference between bank loans that are reported by firms through IES and the bank loans that are available from the Portuguese Central Credit Register (CCR) managed by Banco de Portugal. At the beginning of the quality control, the initial difference between sources is large, but as quality control evolves, the difference decreases and, at the end of the quality control, only 15% of the initial differences (green bar in Chart 1. *Initial and final difference*) remain. The majority of the differences are eliminated automatically by our algorithms (yellow bar in Chart 1. *Initial and final difference*) and a lower fraction is solved manually (red bar in Chart 1. *Initial and final difference*). Bank loans are overvalued in IES because many firms tend to misreport all their financial debt as bank loans. Once at Banco de Portugal, data are automatically or manually corrected. Most of the differences that remain have a justification (Chart 2. *Final difference with and without justifications*), with the most frequent being the existence of external bank loans, which are not covered by the CCR, and the adoption of the IFRS 16⁴ (Chart 3. *Final difference with justification by justification*), also not reflected in the CCR.

⁴ Affects data from 2019 onwards.

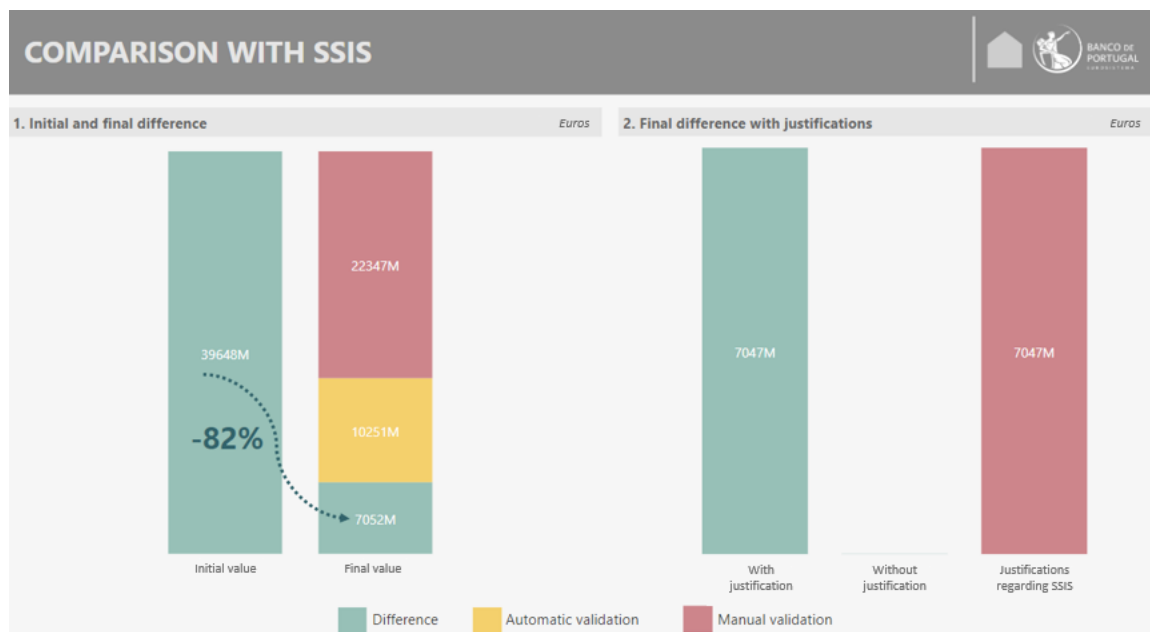
Figure 8: Comparison between CBSD (annual survey) and CCR



This dashboard is regularly updated at the time of the quality control process and allows monitoring the evolution of the differences regarding other sources. Given that one of the aims of the quality control process is to minimise differences regarding other sources, it is extremely helpful. Again, drilling down data is possible and one can check the firms with larger differences and quickly address this issue.

Differences between the amount of debt securities reported through IES and the amount of debt securities issued, which is available from the Securities Statistics Integrated System (SSIS) of Banco de Portugal, are also treated during the quality control process and the way they are monitored is similar to the comparison with the CCR (Figure 9).

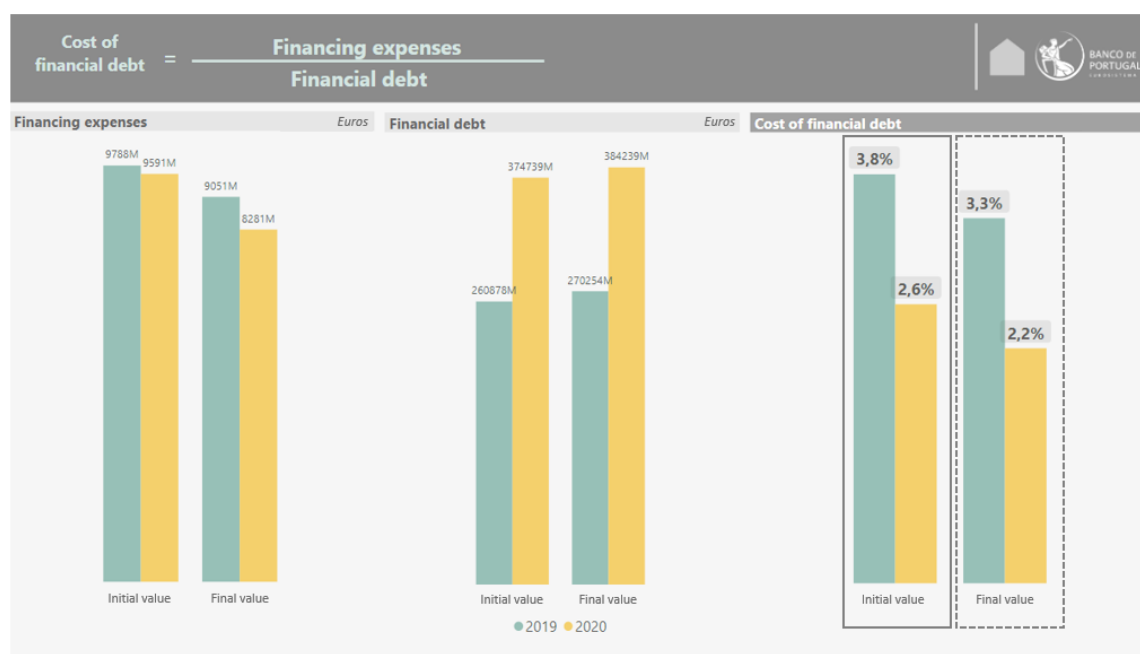
Figure 9: Comparison between CBSD (annual survey) and SSIS



Again, with this visual we can easily see that the initial difference between the two data sources decreases substantially by the end of the quality control process and that any remaining differences are explained and accounted for.

Another kind of analysis that we perform through Power BI is selecting some economic and financial indicators and monitor their evolution vis-à-vis the previous period in order to detect abnormal values. Figure 10 presents the cost of financial debt at the beginning of the quality control (bars on the left, inside the solid line rectangle, in the Chart *Cost of financial debt*) and at the end of the quality control (bars on the right, inside the dashed line rectangle, in the Chart *Cost of financial debt*). At the beginning of the quality control, the cost of financial debt is usually higher because firms incorrectly report expenses other than financing expenses as financing expenses.

Figure 10: Evolution of the cost of financial debt during the quality control process



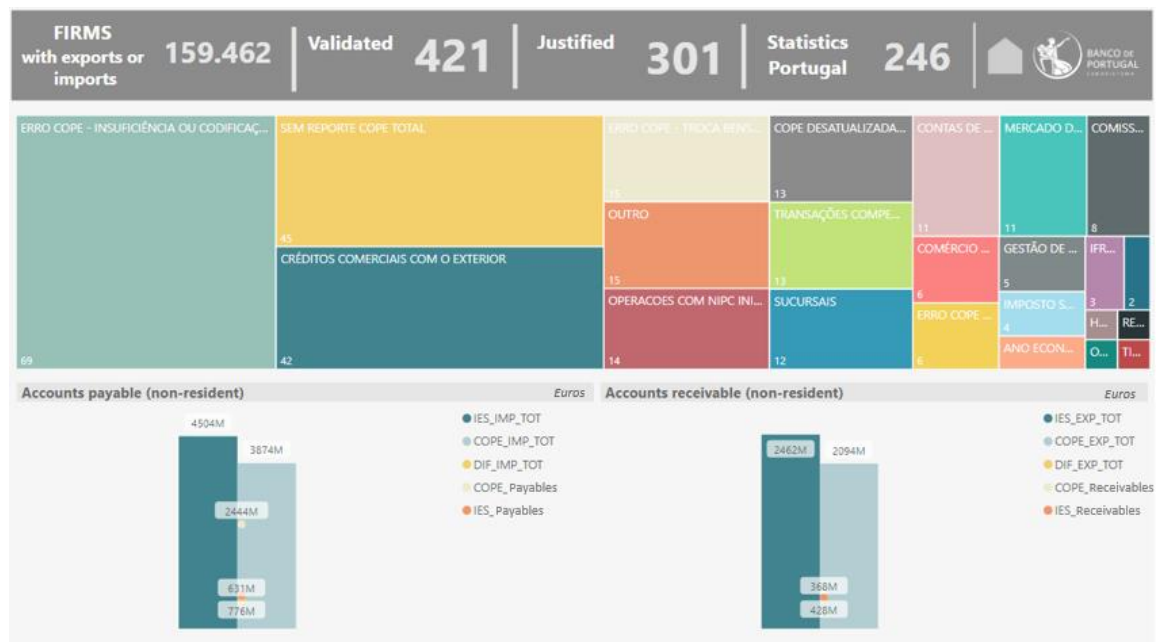
With this dashboard and the possibility of drilling down data, we can easily verify the firms that are influencing the cost of financial debt in a larger extent and correct them if data proves to be inaccurate. Conversely, we can also verify which firms had the larger corrections and double check these adjustments, which is especially important during the work of the undergraduate students.

Besides comparing data from IES with the CCR and the SSIS, exports and imports of goods and services communicated are also compared with Balance of Payments (BoP) data collected by Banco de Portugal. As in the case of CCR and the SSIS, differences arise and one should eliminate or justify them during the quality control process. Figure 11 presents a dashboard that is used to monitor the quality control regarding exports and imports of goods and services, namely the difference between the amounts reported in IES and those available from BoP that violate the consistency checks. The treemap shows the relevance of each justification for that difference. For example, from the 301 firms whose difference between data sources was justified, 69 are attributable to mistakes made by firms when reporting the BoP survey (the so-called COPE report⁵, as mentioned in Casimiro et al., 2017) and 45 to the absence of the COPE report for those firms. BoP figures not only contain exports and imports of

⁵ COPE is the Portuguese acronym for *Comunicação de Operações e Posições com o Exterior*. It is the main source of information for BoP statistics compiled by Banco de Portugal and it is a monthly report where entities in Portugal communicate their transactions and outstanding positions with the rest of the world.

goods and services, but also accounts payable and accounts receivable regarding non-resident counterparts and these data are also used for the quality control process at the CBSO.

Figure 11: Differences between IES and BoP in exports and imports of goods and services



In the course of the quality control of IES, information on the business group structure of firms is also analysed and corrected. In particular, data on the ultimate controlling institutional units (UCI) of NFCs and their location. To examine this issue, we use maps. Instead of presenting statistics for countries in bar charts, it is more appropriate showing a map. They are perfect to communicate when geography and space are important. Figure 12 presents the country of the UCI of firms in Portugal, split by categories, where dark orange represents the countries that control most of firms and light orange those that control the little. This visual rapidly shows that most of the firms in Portugal are controlled by domestic entities, followed by Spanish entities. Dark orange tones in countries that we don't expect to have such large influence or light orange tones in countries that we expect to have large influence represent a quick warning for quality control.

Figure 12: Location of ultimate controlling institutions, by number of firms controlled



3. Conclusion

This paper provides some examples to describe the experience of Banco de Portugal with Power BI for monitoring the quality control of quarterly and yearly accounting data at the CBSO.

Data from the CBSO of Banco de Portugal mostly relies on two sources: the quarterly survey (ITENF, in the Portuguese acronym) and the annual survey (IES, in the Portuguese acronym), which are subject to quality control. To assure the coherence and the consistency of information, data is matched with other data sources managed by the Statistics Department of Banco de Portugal, namely the CCR, the SSIS and the BoP. In fact, the extent of the quality control performed is only possible due to the availability and scope of these micro databases.

In order to assess the magnitude of quality control issues that arise after the implementation of the consistency checks, prioritise them and observe how the work of quality control is evolving, we use Power BI dashboards, which are a powerful visualisation tool that allows connecting quickly to the different data sources. Once the dashboard is created, one only needs to refresh data for evaluating the work in progress, which speeds up decision-making and ignites immediate action over the most urgent matters. Slight changes to visuals and customisation are also easier and faster to perform. Moreover, as a visualisation tool, Power BI not only meets the needs of performing and monitoring the quality control as can be used for delivering presentations. Power BI dashboards are also embedded in our quality control web application, and thus accessible to everyone who performs quality control. Putting all this together, the quality control of data becomes more effective as you can quickly look to the Power BI dashboards and clearly identify where the main problems to be tackled are and how long it takes to solve them. Indeed, Power BI dashboards contributed to fill some data gaps not solved before and to redefine priorities, as well as to decrease time spent doing presentations as they can be presented by themselves.

References

Banco de Portugal (2008), "Simplified reporting: inclusion of the Simplified Corporate Information in the Statistics on Non-Financial Corporations from the Central Balance-Sheet Database", Supplement to the Statistical Bulletin, 1/2008, May 2008.

Available at: <https://www.bportugal.pt/sites/default/files/anexos/pdf-boletim/sup-be-1-2008-en.pdf>

Brites, M. (2013), "Efficient ways of dealing with accounting data from enterprises", Workshop on Integrated Management of Micro-databases, Porto, June 2013

Casimiro, Paula, Pinto, Ana Bárbara and Pereira, Tiago Pinho (2017), Matching firm-level data sources at the Statistics Department of Banco de Portugal, *IFC Bulletin*, 45

Spot the flaw – Using Power BI for quality control: an application to non-financial corporations' data

**IFC Workshop on Data Science in Central Banking:
Applications and tools**

Ana Bárbara Pinto • José Alexandre Neves

• Tiago Pinho Pereira
Statistics Department

February 2022



BANCO DE
PORTUGAL
EUROSYSTEM

QUALITY CONTROL AT THE CBSO OF BANCO DE PORTUGAL

QUARTERLY SURVEY

- ☐ **SAMPLE BASED**
- ☐ **MORE THAN 4 500 FIRMS**
- ☐ **SUMMARISED BALANCE-SHEET AND INCOME STATEMENT**

ANNUAL SURVEY

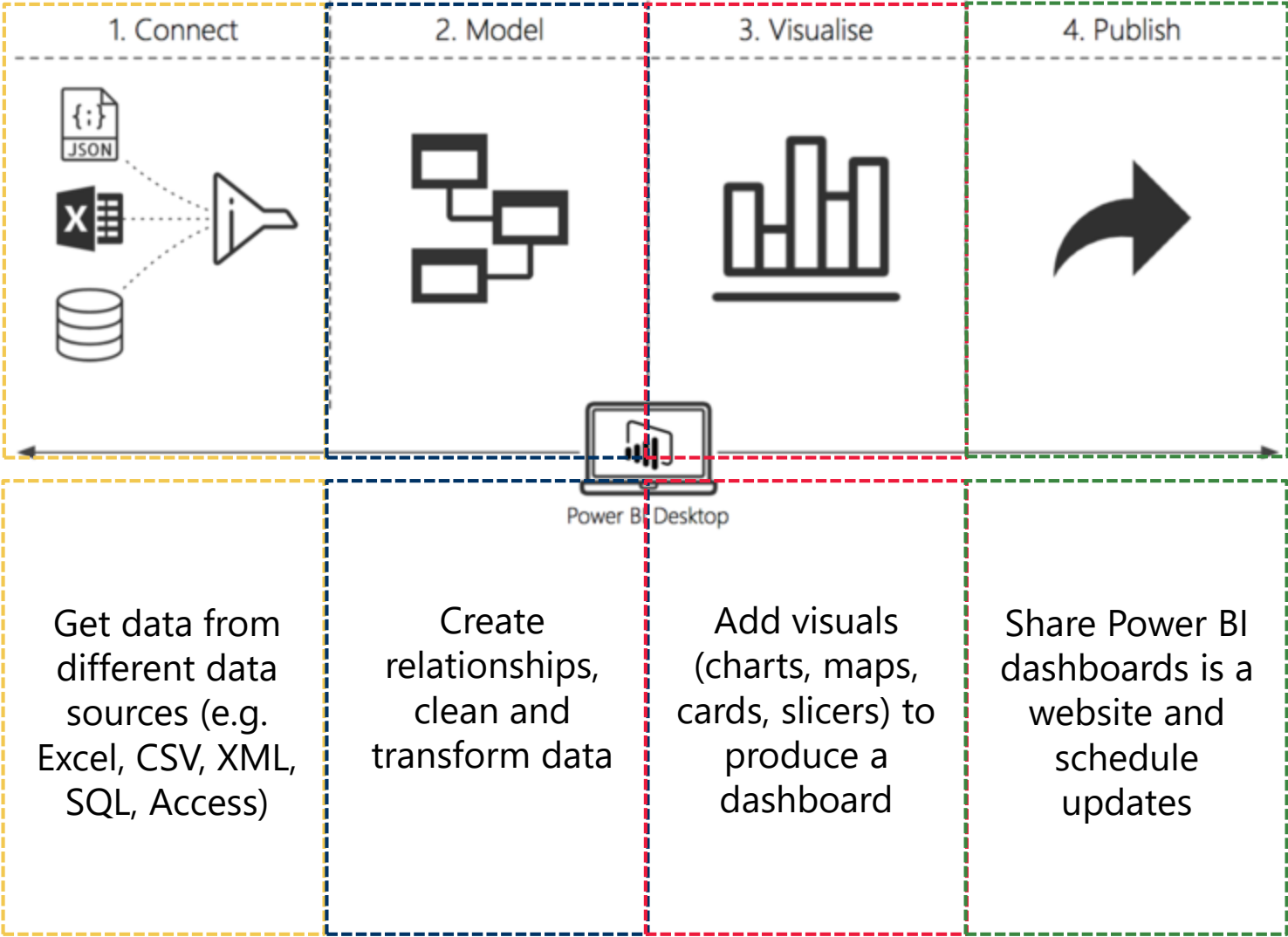
- ☐ **MANDATORY**
- ☐ **MORE THAN 450 000 FIRMS**
- ☐ **DETAILED BALANCE-SHEET AND INCOME STATEMENT**
- ☐ **STATEMENT OF CHANGES IN EQUITY**
- ☐ **CASH FLOW STATEMENT**
- ☐ **NOTES TO THE FINANCIAL STATEMENTS**

QUALITY CONTROL

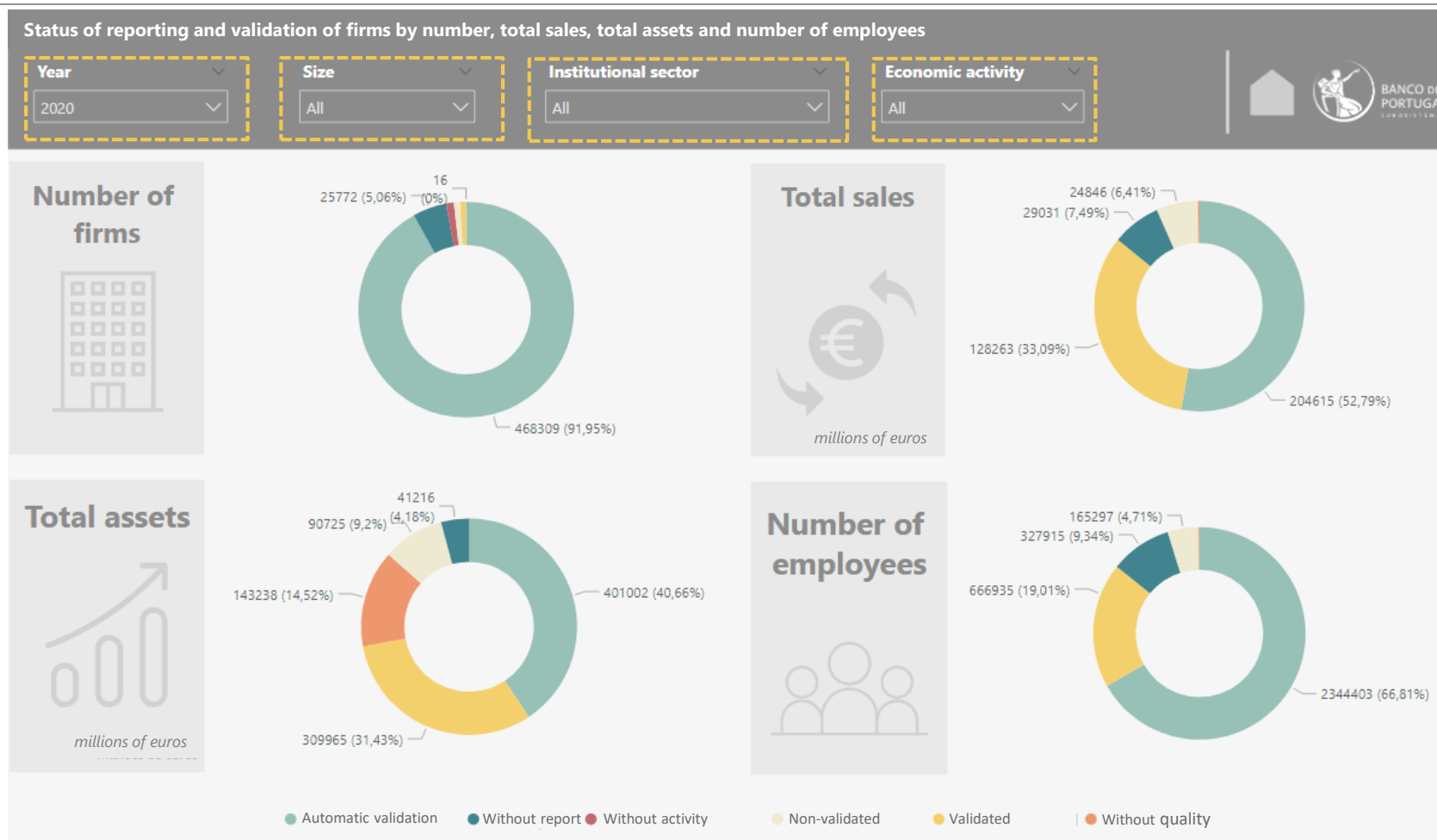
- ☐ **ENSURES THAT INFORMATION IS HORIZONTALLY AND VERTICALLY CONSISTENT**
- ☐ **COMPARISON WITH INTERNAL AND EXTERNAL DATA SOURCES**
- ☐ **AUTOMATIC CORRECTIONS APPLY**
- ☐ **FIRMS WITH LARGE ABSOLUTE OR RELATIVE DIFFERENCES ARE SELECTED FOR MANUAL VALIDATION**



POWER BI FOR QUALITY CONTROL PURPOSES AT THE CBSO OF BANCO DE PORTUGAL



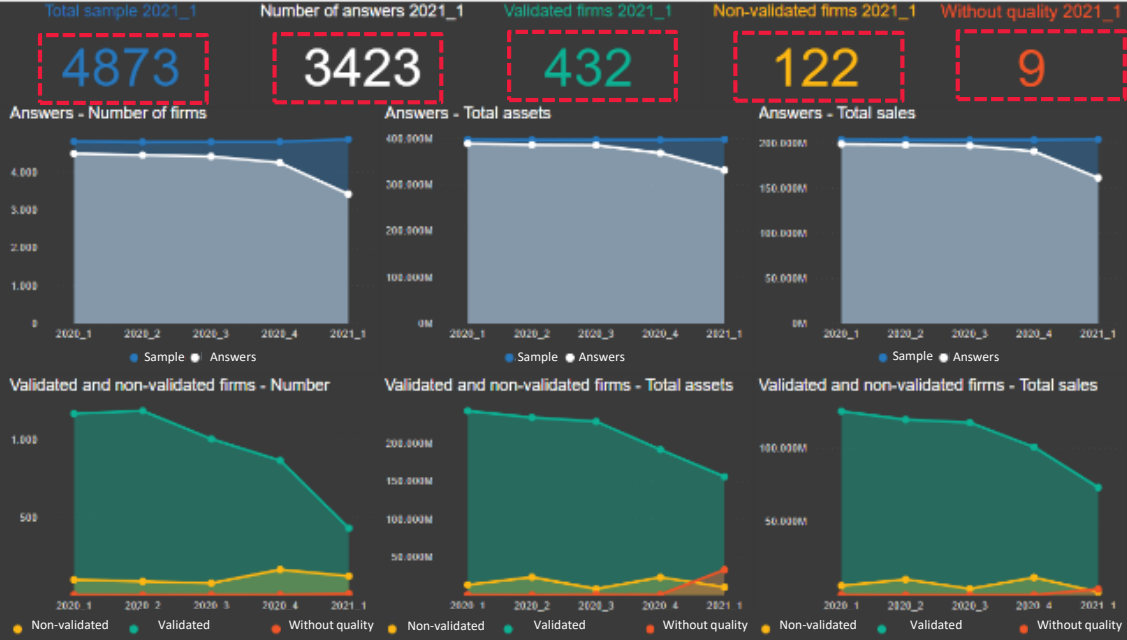
POWER BI FOR QUALITY CONTROL PURPOSES AT THE CBSO OF BANCO DE PORTUGAL



EVALUATE THE STATUS OF VALIDATION AND REPORTING OF FIRMS



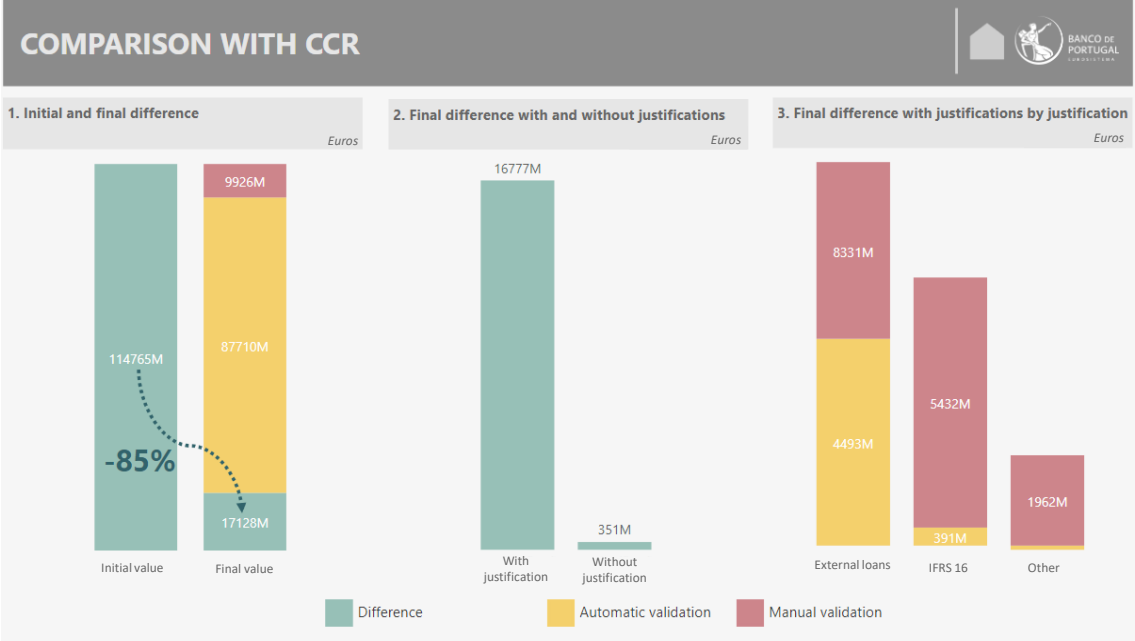
POWER BI FOR QUALITY CONTROL PURPOSES AT THE CBSO OF BANCO DE PORTUGAL



MEASURE THE EVOLUTION OF QUALITY CONTROL



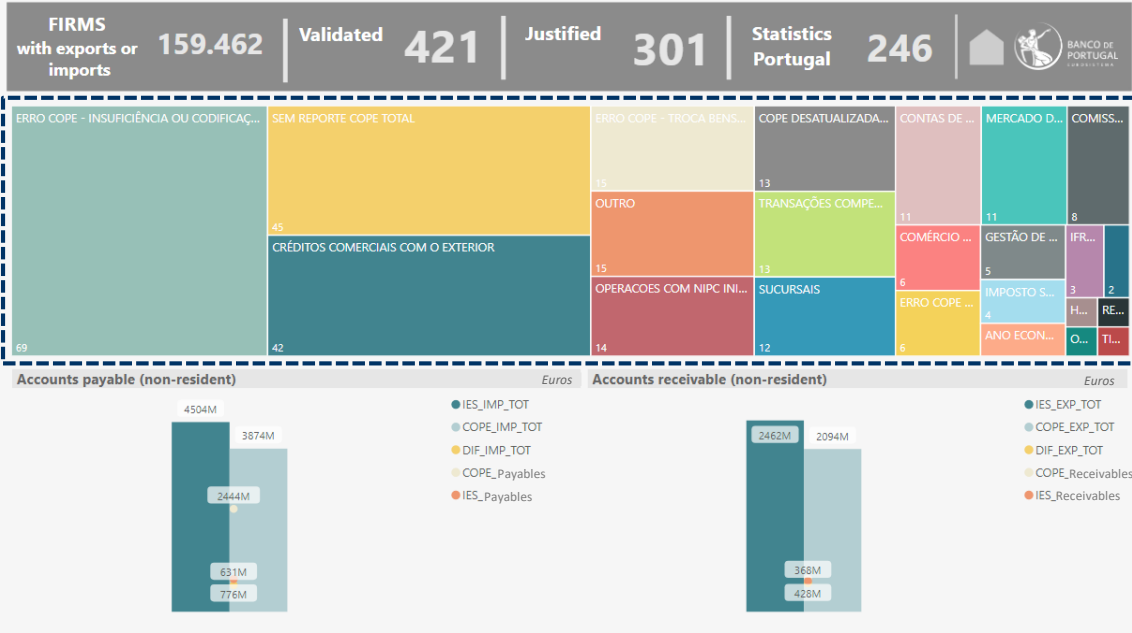
POWER BI FOR QUALITY CONTROL PURPOSES AT THE CBSO OF BANCO DE PORTUGAL



CHECK AND EXPLAIN THE IMPACT OF QUALITY CONTROL ON DATA



POWER BI FOR QUALITY CONTROL PURPOSES AT THE CBSO OF BANCO DE PORTUGAL



HIGHLIGHT RELEVANCE



MAIN CONCLUSIONS

TO SUM UP

- ❑ **ALLOWS CONNECTING QUICKLY TO THE DIFFERENT DATA SOURCES**
- ❑ **EMBEDDED IN OUR QUALITY CONTROL WEB APPLICATION**
- ❑ **DECREASED TIME SPENT DOING PRESENTATIONS AS POWER BI DASHBOARDS CAN BE PRESENTED BY THEMSELVES**
- ❑ **USED TO PERFORM AND MONITOR QUALITY CONTROL OF DATA AT THE CBSO OF BANCO DE PORTUGAL**
 - ✓ **QUICKLY SHOWS TRENDS AND DATA GAPS NOT SOLVED YET**
 - ✓ **HELPS REDEFINING PRIORITIES IN TERMS OF FIRMS AND CONSISTENCY CHECKS**

