Comparison of Power BI reports for data migration check

Ilija Tavchioski, Pavle Gomboc, Špela Zavodnik

Abstract

The goal of this project was to use data science methods in order to compare two *Power BI* reports and extract the differences between them in order to obtain a knowledge of whether the databases that are used for rendering the reports are equivalent. For this interim report we used basic methods for their comparison, and we achieved decent performance as initial results.

Keywords

Power BI, Database, Data comparison

Introduction

Data migration is a common procedure for companies that uses a large amount of data where in order to gain some new advantages they decide to change the database and thus the need to migrate the whole data to a new server. But, since he amount of data that the companies are migrating is generally at large amount, mistakes can occur during the process of migration.

There are several methods to check whether the migration is done correctly, and one of the computationally simplest one is to compare two *Power BI* reports, rendered on the same queries and with the same structure and check whether the data is the same.

The goal of this project is to research some techniques for extracting the differences between two reports and to conclude whether they are rendered from the same database.

Data

The data was provided by the *In5ight* company [1], was composed of 4 pairs of reports which are about a data related to the sales of vegetables between two years in terms of value and quantity, which were structured into the following structure.

- The difference between sales quantity
- Sales quantity by customer
- Sales quantity by country
- Sales quantity by month
- Sales quantity by vegetable

And, the same structure was for the total value of the vegetables.

Methodology

Data extraction

The data of the reports was extracted manually for each visualization of the report.

Data processing

In addition the data with missing values was removed from the comparison.

Data comparison

The comparison that was performed was divided into two parts. **Difference in years** – this part is used to find out whether there are difference in years between the reports. Since the data that are showing the reports are form two consecutive years, if the current year in the first and the second report are differentiating by 2 or more years we can conclude that we can not compare these reports. Other wise we compare the columns that are compatible. If the current year is the same, we compare the reports for the current and the previous year, if the difference is only one year, (the first report is 2022 and the second at 2021), we compare just the data from the previous year from the first report and the data from the current year of the second report. **Comparison** – our current method for comparison is fairly simple, we iterate over all countries, customers, dates and veggies and try to find out whether they match or not. If they match we can be certain that the visual objects are rendered from the same data which

Pair	Sales quantity	Sales value
1	Match	Match
2	Match	Not match
3	Match	Match
4	Not match	Not match

Table 1. The difference between quantity and sales reports.

means the data is successfully migrated. Otherwise, we are still trying to find a method that will be certain that if there is a difference between the visual objects, we can conclude that the migration of the data is unsuccessful, since there can be some cases where there is some difference in the data but the it is most probably the same data as can be shown in Table 1.

Results

In this section we will present you a brief results that we currently obtained by our methods for comparison. As can be shown in the following Table 1, we can conclude that the first and third pair that the data is the same. But, we can not be sure for the second one since it is matched in the quantity report but not in the value report. This can be due to several factors and its worth investing further.

Conclusion

In conclusion, since our current methods are the baselines methods, although for the 4 pairs of reports they are achieving the expected outcome we still need to improve them.

Future work

It is important to acknowledge here that the data from the reports was not extracted automatically, which is one of the most important part of this project. The first objective for the following month will be to automate the process of extracting the data from the *Power BI* reports.

And the second important part is that we will need to improve our methods for comparisons, maybe use other data to predict, and definitely we will visualize the difference between the reports and try to find some patterns.

Acknowledgments

References

[1] In5ight company. Visited on: 14/04/2023.