melissa

A **MELISSA** WHITEPAPER

Data Quality for Part Numbers

TABLE OF CONTENTS

Part Numbers are an ubiquitous part of almost every e-tail or retail business. Most companies employ hundreds, if not thousands, of part numbers. Every time a part number is added, or changed, or moved to another system, it is another opportunity for errors, truncation, or corruption to take place.

This white paper shows some techniques that may be applied to maintain data quality within your part number system; to help you organize and manage part numbers. To do this, we will use data quality profiling techniques. Profiling data impacts areas of data quality far beyond simple part numbers, which is beneficial as you may apply data discovery techniques to other areas of your business operations. There are a number of profiling tools available on the market intended to help you discover patterns and issues with your data and help you apply a fix. For this paper, we will show examples and techniques utilizing the Melissa Profiler Object.

DATA PROFILING VS. BILL OF MATERIAL (BOM) SOFTWARE

Note that the Profiler Object is not a full-fledged BOM software. If your company maintains tens of thousands of part numbers or more, Profiler Object can help you with data discovery, but it is not a replacement for BOM software to help you maintain all your part numbers.

However, if you maintain your part numbers in Microsoft® Excel® or text/flat files, or if you use a relational database like Microsoft Access® or SQL Server®, then Profiler Object can be a valuable asset and you will find these techniques useful.

Intelligent part number systems impart knowledge and descriptive details inside the part numbers themselves. This concept applies widely outside of part numbers, like the United States Postal Service® use of IMb® (Intelligent Mail Barcode). Under this scheme, a person should be able to deduce at least certain categorical information and often, other types of information as well. An example "GAR-B-034-012" could be where GAR stands for a department, B stands for the shape of the package, 034 stands for the aisle, and 012 stands for the shelf. This is an extremely descriptive part numbering scheme, and that may not necessarily suit every company's need.

Within an intelligent part number scheme, there are a lot of meaningful and referential data, which means the possibility for errors are higher but the opportunities for data quality are also greater.

FORMAT INTEGRITY

With profiling, you can typically see all the different formats present in your data, usually called pattern analysis. Ideally, taking our intelligent part number example, you would only want to see one format:

Format	Count
AAA-A-###-###	5000

However, you can very possibly see multiple formats in your data:

Format	Count
AAA-A-###-###	4278
AAA-#-###-###	64
A-A-#-#	412
AAA-A-#####	246

BOUNDS CHECK

Since our intelligent part number system contains references to real-life metrics like location, we can apply some real-life sanity checks to make sure the numbers are not obviously incorrect. For instance, if the warehouse employing our theoretical intelligent part number system only contains 55 aisles and 25 shelves, then we can check that the part number is between the following bounds:

AAA-A-[001-055]-[001-024]

This will allow us to flag parts like GAR-B-000-999 immediately as not possible.

NON-INTELLIGENT PART NUMBERS

Non-Intelligent Part numbers use a simple scheme. In most cases, it is simply a number, pulled in numerical order, regardless of category. Part number 139231 and part 139232 could conceivably be completely different and non-related parts.

RANGE CHECK

With a numbers scheme, it is simple to see if a part number is valid as long as it is within a certain range. With profiling, it is easy to see the minimum and maximum range in your part numbers:

MIN: 1

MAX: 72671

DATA TYPE CHECK

With a number scheme, your data is not necessarily stored as a number, it is often stored as a text. This means that nonnumbers can be added into the system without an error or notice. With a data type check, you can see that all the part numbers are, indeed, a number type and there are no alphabets or nulls.

InferredDataType: Numeric

NumericOnlyCount: 451

AlphaOnlyCount: 0

AlphaNumericCount: 1

EmptyCount: 12

NullCount: 0

SEMI-INTELLIGENT PART NUMBERS

Under a Semi-Intelligent part number scheme, you typically employ a limited amount of intelligence into the part number.

An example could be:

1000 – Category 1

2000 – Category 2

3000 – Category 3

Under this scheme, the same type of data quality can apply.

RANGE CHECK

MIN: 0001 (theoretically impossible if there is no Category 0)

MAX: 9044

DATA TYPE CHECK

InferredDataType: Numeric

NumericOnlyCount: 341

AlphaOnlyCount: 12

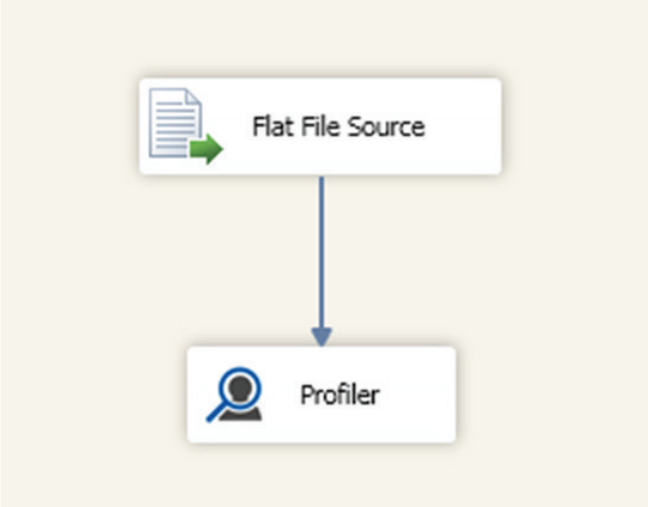
AlphaNumericCount: 0

EmptyCount: 0

NullCount: 5

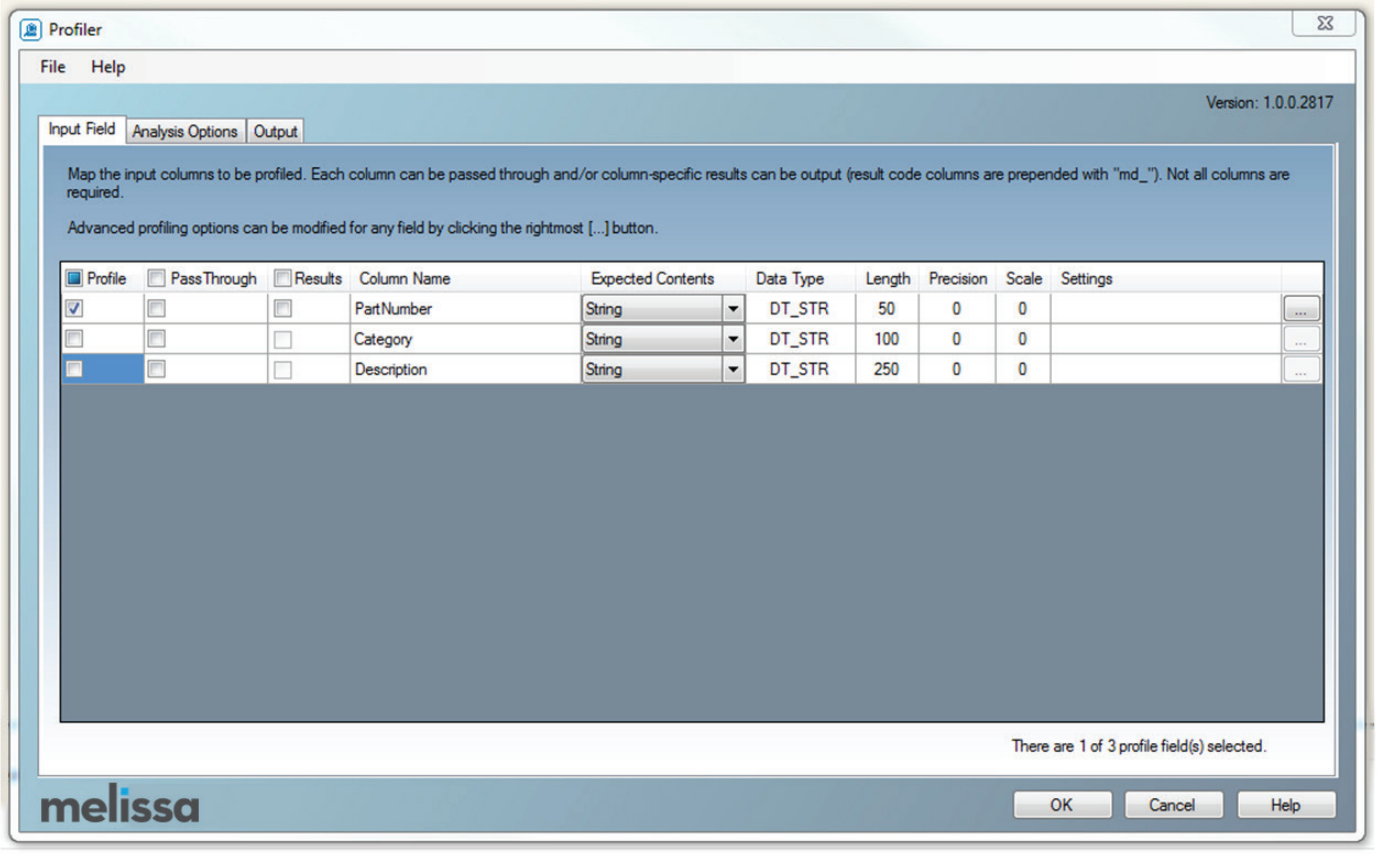
MELISSA PROFILING IN ACTION

Now that we have talked about some of the benefits of profiling, we can see what it would look like to use the Melissa Profiler component for SSIS to profile part numbers. The profiling component was built for much more than profiling part numbers. In actuality, part numbers are usually a small side project. This component allows you to profile any sort of data with all the features and capabilities available in all the leading profilers in the market today. Additionally, we have built in our contact expertise to allow you to perform special proprietary profiling jobs. You can analyze and find duplicate names, phones, emails, and addresses as you profile.

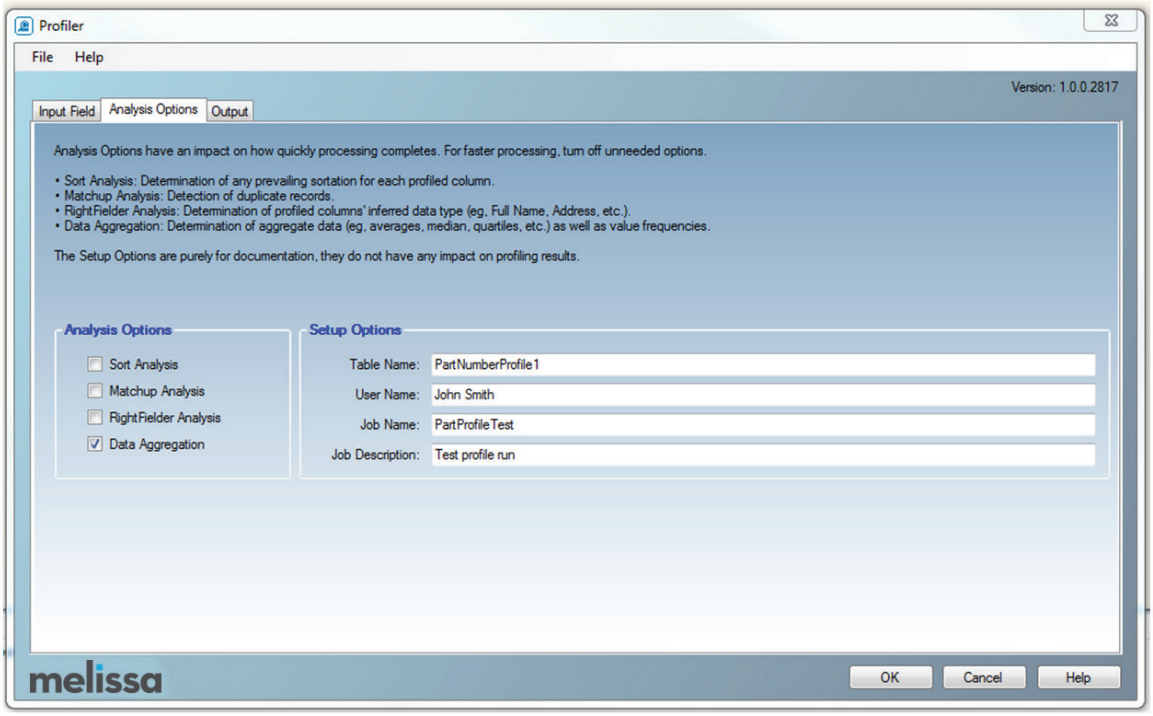


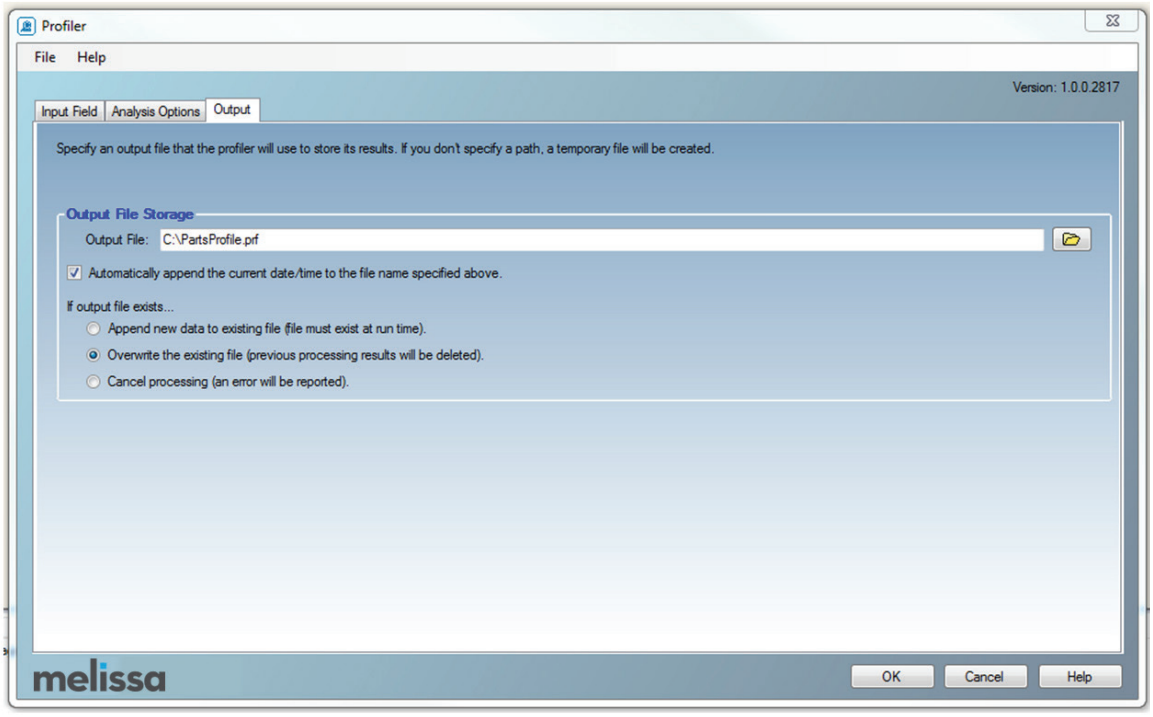
Back to part numbers, we will open up a simple text file in SSIS. The good thing with SSIS is that Microsoft® has built-in connectors to almost all the popular data sources, so the source can just as easily be Excel® or a database like SQL Server®.

Inside the profiling component, we will select the part number column we want to profile.

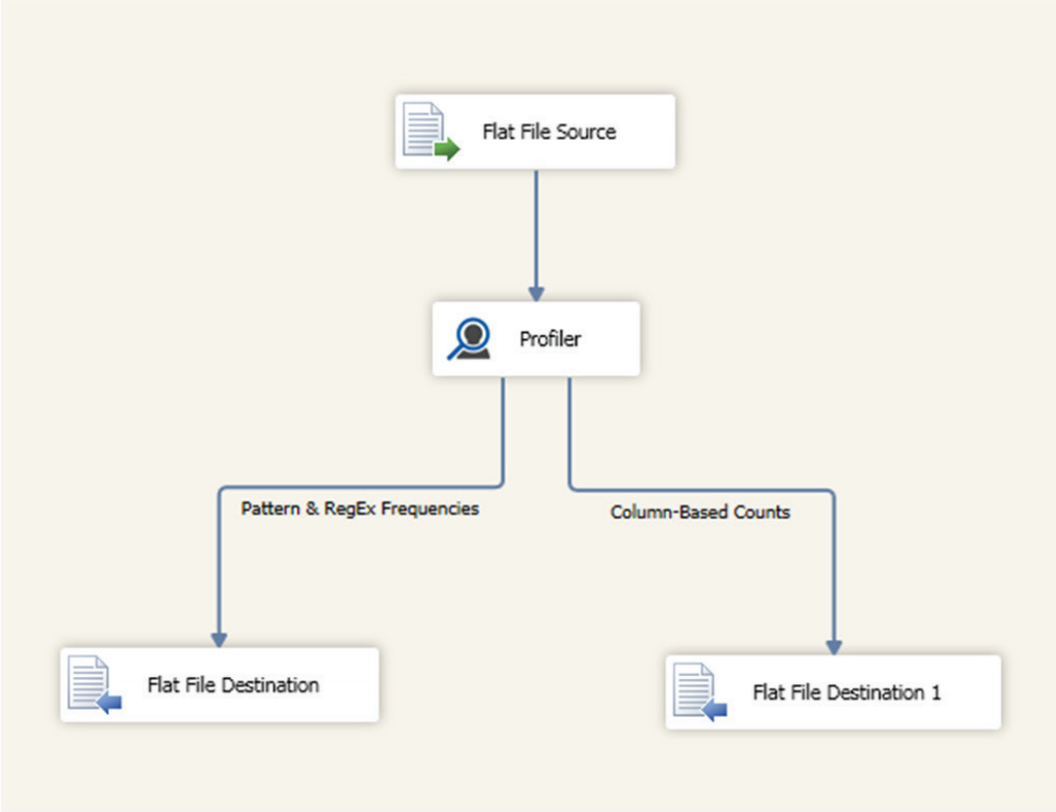


In the next tab, we will specify the types of profiling needed. Part numbers are fairly simplistic, so we only need the data aggregation option and we can save time by turning the other options off. If we had other pieces of data like contact data, the other options would become more useful.





Lastly, we will specify where to write the outputs. The Pattern & RegEx Frequencies output will give you the Format Integrity information shown in this paper. The Column-Based Counts output will give you the Data Type and Range Check information.



Conclusion

If you use part numbers, they are most likely a critical part of your organization. Any errors or corruption in your part number data will cost you time and money. With some basic data quality concepts and a profiling tool, you can detect these issues and fix them before they cause any major problems. Beyond part numbers, Melissa Profiler can help with many other pieces of data. Once you get used to the power and usefulness of our profiling tool, you will likely use it with any data sets you can get your hands on.

melissa

www.melissa.com

About Melissa

Since 1985, Melissa has specialized in global intelligence solutions to help organizations unlock accurate data for a more compelling customer view. Our breadth of data and flexible API technology integrates with numerous third-party platforms, so it works for you and makes sense for your business. More than 10,000 clients worldwide in key industries like insurance, finance, healthcare, retail, education, and government, rely on Melissa for full spectrum data quality and identity verification software, including data profiling, cleansing, matching, and enhancement services, to gain critical insight and drive meaningful customer relationships.

US

22382 Avenida Empresa
Rancho Santa Margarita, CA 92688-2112

800.MELISSA (635.4772)

UK

+44 (0)20 7718 0070

GERMANY

+49 (0) 221 97 58 92 40

INDIA

+91 (0)80 4854 0142

AUSTRALIA

+61 02 8091 6000